Crowd science and engineering: concept and research framework

Yueting Chai
National Engineering Laboratory for E-commerce Technology, Tsinghua University, Beijing, China

Chunyan Miao
Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, Nanyang Technological University, Singapore, Singapore and School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

Baowen Sun
Central University of Finance and Economics, Beijing, China, and

Yongqing Zheng and Qingzhong Li
Shangdong University, Jinan, China

Abstract

Purpose – The synthetic application and interaction of the internet, Internet of Things, cloud computing, big data, Industry 4.0 and other new patterns and new technologies shall breed future Web-based industrial operation system and social operation management patterns, manifesting as a crowd cyber ecosystem composed of multiple interconnected intelligent agents (enterprises, individuals and governmental agencies) and its dynamic behaviors. This paper aims to explore the basic principles and laws of such a system and its behavior.

Design/methodology/approach – The authors propose the concepts of crowd science and engineering (CSE) and expound its main content, thus forming a research framework of theories and methodologies of crowd science.

Findings – CSE is expected to substantially promote the formation and development of crowd science and thus lay a foundation for the advancement of Web-based industrial operation system and social operation management patterns.

Originality/value – This paper is the first one to propose the concepts of CSE, which lights the beacon for the future research in this area.

Keywords Transaction, Intelligence, Crowd science, Stability, Ecology and evolution, Modelling and simulation

Paper type Conceptual paper

© Yueting Chai, Chunyan Miao, Baowen Sun, Yongqing Zheng and Qingzhong Li. Published in the International Journal of Crowd Science. Published by Emerald Publishing. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode
1. Introduction
Following the first industrial revolution marked by the invention of steam engine and the second industrial revolution marked by the invention of electric power, man has ushered in the third industrial revolution with the internet as the symbol. Compared with the first two, the third one is transforming the industrial forms and social operation patterns of the human society in a more comprehensive and fundamental way. The underlying reason is that the internet has ignited man’s demand and unleashed potential needs of humans, thus having formed the motive power for internet, changing industry and social operation mode.

With a great prospect for integrated development in various fields, the internet is exerting overall and strategic impact on the economic and social development of various countries. For this reason, major developed economies have drawn up corresponding measures to gain the power of initiative in future competition, for instance, Industry 4.0 in Germany and Industry Internet in the USA. Meanwhile, the internet-based economy represented by E-commerce has sprung up in China, extending from consumption to industry, taking on robust power in ensuring economic growth, rebuilding economic structure, stabilizing employment and promoting innovations. Chinese government has in a timely manner made a series of policies and strategies to promote E-commerce, Internet Plus, Internet of Things (IOT), big data, cloud computing, mass entrepreneurship and innovation, crowd making, crowd sourcing, crowd supporting and crowd funding, China Manufacturing 2025 Development Program, etc.

A comprehensive survey of the policies, measures and action plans of various nations finds that most of them are advancing the work through the perspectives of the internet, IOT, cloud computing, big data, Industry 4.0, etc. As a matter of fact, the synthetic application and interaction of/between the above-mentioned new patterns and new technologies shall breed future Web-based industrial operation systems and social operation management patterns. See Figure 1.

In comparison with traditional industrial operation systems, future Web-based industrial operation systems take on the following essential features: initiative, personalized consumption, immediacy, centralized circulation, decentralization, smart production, self-organization and ecologicalized system.

![Figure 1. Future Web-based industrial operation system](image-url)
2. The concept of crowd science and engineering

A deep analysis of general types and substantive characteristics of future Web-based industrial operation systems can easily identify a smart interconnected network system that takes on the features of large scale, openness, self-organization and ecologicalization. Hence, the system is called “Crowd cyber System” in short, as shown in Figure 2. In terms of components, this kind of network systems take the forms of smart entities composed of a great many intelligently interconnected individuals, enterprises, governmental agencies and articles, with sort of supply–demand relationship between them. In terms of major behaviors on the network, the interaction between any smart entities boils down to some sort of transaction or behavior. From the perspective of system, the network system is a ternary system of information, physics and society. Openness, interconnection, cooperation and sharing constitute the basic logic of the network system while stability, efficiency, innovation and development are the targets the system aims to fulfill.

The numerous intelligent agents (enterprises, institutions, individuals and articles, etc.) in the crowd cyber system can give full play to their wisdom and potentiality, interact with each other and produce various kinds of unexpected nonlinear behaviors and effects, which cannot be explained by an existing discipline, but rather need the basic theories and frontier achievements of multiple disciplines. Therefore, an emerging interdiscipline is needed.

Therefore, we propose the concept of “crowd science and engineering” (CSE). Using knowledge of system theory, information theory, cybernetics, computer science and engineering, management, economics, sociology, psychology, as well as the new technological means of IOT, cloud computing, big data, artificial intelligence, CSE aims to explore the basic principles and laws of the swarm intelligence activities of the ternary system of information, physics and society in the context of large-scale online interconnection. In so doing, it seeks to establish related methods and tools, give full play to individual and group wisdom of mankind, tap into the potentials and effectively advance a new Web-based industrial operation system and social operation management patterns.

With crowd cyber system as the object of study, crowd science aims to establish the system of concepts of the crowd cyber system featuring stability, efficiency, innovation and
development, the basic principles, behavior patterns and related methodologies, thus laying a solid foundation for the Web-based industry. The essential questions to be addressed include:

Q1. How to construct the crowd cyber system to realize a synchronic mapping in structure, information, behavior and consciousness between individuals, enterprises, governmental agencies and articles of the physical world and various corresponding intelligent agents in the information world, and support various kinds of intelligent transaction behaviors between intelligent agents.

Q2. What kind of eco-structure and behavior mode will the multiple intelligent agents take on to ensure the efficient operation of the network system?

Q3. The originality of the crowd cyber system rests with the multitude of its intelligence. What is intelligence? How to measure it? What is the relationship between group intelligence and individual intelligence of various intelligent agents? How to make use of the group intelligence of the crowd cyber system to measure the originality of the network system?

Q4. The sustainability of the crowd cyber system rests with the evolution of various intelligent agents. What are the mechanisms, ways, orientations and conditions of various intelligent agents in the system? How to maintain the rational evolution of various intelligent agents so as to arrive at the sustainable development of the network system?

Q5. How to maintain the relative stability of the network system and rule out breakdowns? This, in the real world, means to avoid destructive “bubble economy”, “economic crisis” and “subversive revolution”.

3. The research framework of crowd science and engineering
To answer the basic questions mentioned above, the research on the theories and methodologies of CSE mainly concerns issues including:

- research on the modeling and simulation of crowd cyber system;
- theoretical research on the structure of crowd cyber eco-system;
- theoretical research on the intelligent transactions of crowd cyber system;
- research on the intelligence measurement methods of crowd cyber system;
- theoretical research on the evolution of crowd cyber system; and
- theoretical research on the stability of crowd cyber system.

3.1 Research on the modeling and simulation of crowd cyber
Mainly dealing with the construction of crowd network, these types of research aim to build the mental models of network for various intelligent agents (individuals, enterprises, governments and things), so as to enable all intelligent agents in cyber space to reflect individuals, enterprises, governments and things in the real world comprehensively, authentically, correctly and simultaneously. Such models refer to analytic models (for computing), static structure models (for building different kinds of systems), dynamic behavior models (for studying the operational rules of systems), presentation models (for
presenting real individuals, enterprises, governments and things in cyber space) and other intelligent interconnected models and algorithms between all intelligent agents.

3.2 Theoretical research on the structure of crowd cyber eco-system

In the future Web-based industrial systems, varied properties and features of industries or services will definitely result in a large number of operation platforms for interconnected web-based industries and thus form an industrial eco-structure. Then, how many platforms are needed to sustain efficient running of specific industrial systems? Should they be comprehensive, semi-professional or professional? These actual questions indicate the necessity of research on the crowd cyber eco-system.

The eco-structure of crowd network is the relationship of division and collaboration between intelligent agents that form the network, while in terms of future Web-based industries, it is industrial organization form. Eco-structure is one of the key factors affecting the efficiency of network operation.

In a certain industrial organization ecological system, each intelligent agent (each market entity) has its corresponding ecological niche, which refers to the sum of an intelligent agent's particular position in time and space and its functional relationship with other agents. The ecological niche and such relationship will change as transaction efficiency changes, leading to eco-structure changes. Specific transaction efficiency goes with its own eco-structure which is the most efficient structure for the operation of crowd cyber or industrial systems. Transaction efficiency is the main inducement to the evolution of eco-structure. That an intelligent agent's ecological niche and its relationship with other agents change after the changing transaction efficiency is in fact a strategy adopted by the agent according to its ecological niche and environmental variation. Therefore, in eco-structure theories, the focuses of research lie in the evolutionary dynamics, mechanisms and paths of the eco-structure.

3.3 Theoretical research on the intelligent transactions of crowd network

The interaction between various intelligent agents that constitute crowd network boils down to some transaction activities or behaviors. A transaction activity involves the supply side, demand side, objects and rules of the transaction. Major problems in intelligent transactions to be tackled include:

- to describe the needs of demand sides accurately and choose the most suitable out of many supply sides;
- to accurately describe the needs of supply sides and choose the most suitable out of many demand sides; and
- to quickly conclude transactions.

To cope with the above problems, the theoretical research on intelligent transactions should focus on the following aspects:

- intelligent methods of acquiring the law of demand and accurate demands;
- intelligent methods of acquiring the law of supply and accurate supplies;
- methods of intelligent matching transactions;
- mechanisms and methods of forming dynamic pricing; and
- methods of evaluating intelligent transactions.
3.4 Research on the intelligence measurement methods of crowd cyber

The purposes of research on the intelligence measurement methods of crowd network are to evaluate the intelligence of individuals and groups and to offer support for evaluating the innovation potential of individuals and groups. The research focuses on the concept of standardized intelligence and the measurement methods of intelligence of individuals and groups.

There is no standardized concept of intelligence at present. Intelligence and its nature have been studied by lots of philosophers and brain scientists all around the world since ancient times but still remain so mysterious that the occurrence of intelligence is among the four secrets of nature with other three being the nature of matter, the origin of the universe and the nature of life. We should, on the basis of existing research results, attach a standardized concept to intelligence, which can contribute to quantitative evaluation of intelligence.

Generally speaking, intelligence may be high or low, but to what degree? Related research is short of fruits. We should build a measurement model of individual intelligence to calculate individual intelligence quantitatively. Meanwhile, group intelligence is higher than individual intelligence in most cases. But there is no research explaining the exact difference and the relationships between them. We should build a crowd cyber-based measurement model of intelligence to calculate group intelligence quantitatively.

3.5 Research on the evolution of crowd network

The purpose of researching the co-evolution theories of crowd network is to support the sustainable development of Web-based industries in the future with scientific evidence. To achieve a sustainable development, an industry must, first, create customers’ demands; second, satisfy these demands with low cost and high efficiency; and third, get the most out of all resources, particularly physical resources. All of these require continuous innovation on products and services, improvement in manufacturing or service process and reduction in wastes of resources. Such processes of innovation or improvement are manifested as a continuous evolution of specialized knowledge, business process, ways of resource utilization in an environment of market competition, in other words, to sift out those lagging behind and promote those going ahead. Therefore, research on the co-evolution theories of crowd cyber should focus on the following aspects:

- evolution methods of specialized knowledge;
- co-evolution mechanisms and algorithms of business process; and
- evolution mechanisms and algorithms of resource utilization methods.

3.6 Theoretical study on crowd network stability

Stability is the foundation of everything, including the future network-based industry. Of course, the stability in this case is dynamic, as the industry itself is developing. It can be derived from existing research results, and a stable network cannot be established without the following factors:

- enough network nodes; otherwise, a network with every few network nodes could be vulnerable to external interferences, which would ruin the network;
- excellent network connectivity; otherwise, a loose connection among network nodes would result in a quick network partition under external interferences; and
controllable differences among network nodes; otherwise, the internal imbalance and insoluble contradiction intensification would cause the collapse of the network.

Given the above-mentioned factors, the concept of crowd cyber stability should be raised and regularized, while the crowd cyber stability quantification methods are demonstrated, which provide the basis of network stability evaluation and thus facilitate the establishment of crowd cyber stability model, through which the solution comes out with network stability status and relevant conditions.

Further reading

About the authors
Yueting Chai is a Professor of Department of Automation at Tsinghua University, China and the Director of National Engineering Laboratory for E-Commerce Technologies (NELET). He graduated from Department of Automation at Tsinghua University in 1991. His research focuses on e-commerce theory, model, technologies and applications, crowd science and engineering, trust model for e-commerce, etc. Yueting Chai is the corresponding author and can be contacted at: chaiyt@tsinghua.edu.cn

Chunyan Miao is a Professor in the School of Computer Science and Engineering (SCSE), Nanyang Technological University. Prof. Miao is currently serving as the Founding Director of the Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY). LILY is one of the first research centers focused on artificial intelligence (AI) technologies for help the elderly lead an active, healthy and dignified lifestyle.

Baowen Sun is a Professor of Central University of Finance and Economics. He graduated from Northeast University, China in 1986 and received PhD from Central University of Finance and Economics in 2004. His research focuses on E-Commerce, economics, crowd science and engineering, etc.

Yongqing Zheng is a Research Fellow of School of Computer Science and Technology at Shandong University, China and the Deputy Director of National Engineering Laboratory for E-Commerce Technologies (NELET). He graduated from the Department of Computer Science at Shandong University in 1991. His research focuses on software architecture and business modelling, big data, cloud computing, e-commerce, crowd science and engineering, etc.

Qingzhong Li is a Professor of School of Computer Science and Technology at Shandong University, China. He graduated from the Department of Computer Science at Shandong University in 1986 and received PhD from Institute of Computing Technology (ICT), the Chinese Academy of Sciences in 2000. His research focuses on computer system, e-government and office automation, etc.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm Or contact us for further details: permissions@emeraldinsight.com
A cyber-anima-based model of material conscious information network

Jianping Shen and Yadong Huang
Department of Automation, Tsinghua University, Beijing, China, and
Yueting Chai
National Engineering Laboratory for E-commerce Technology, Tsinghua University, Beijing, China

Abstract

Purpose – This paper aims to study the node modeling, multi-agent architecture and addressing method for the material conscious information network (MCIN), which is a large-scaled, open-styled, self-organized and ecological intelligent network of supply–demand relationships.

Design/methodology/approach – This study models the MCIN by node model definition, multi-agent architecture design and addressing method presentation.

Findings – The prototype of novel E-commerce platform based on the MCIN shows the effectiveness and soundness of the MCIN modeling. By comparing to current internet, the authors also find that the MCIN has the advantages of socialization, information integration, collective intelligence, traceability, high robustness, unification of producing and consuming, high scalability and decentralization.

Research limitations/implications – Leveraging the dimensions of structure, character, knowledge and experience, a modeling approach of the basic information can fit all kinds of the MCIN nodes. With the double chain structure for both basic and supply–demand information, the MCIN nodes can be modeled comprehensively. The anima-desire-intention-based multi-agent architecture makes the federated agents of the MCIN nodes self-organized and intelligent. The MCIN nodes can be efficiently addressed by the supply–demand-oriented method. However, the implementation of the MCIN is still in process.

Practical implications – This paper lays the theoretical foundation for the future networked system of supply–demand relationship and the novel E-commerce platform.

Originality/value – The authors believe that the MCIN, first proposed in this paper, is a transformational innovation which facilitates the infrastructure of the future networked system of supply–demand relationship.

Keywords E-commerce, Cyber physics social system, Multi-agent, Supply-demand relationship

1. Introduction

In a new round of global revolution in technology and industry, the fusion of internet and other domains has produced a strategical and holistic impact on the development of economic society in various countries of the world. New models, technologies and applications are jointly breeding a future networked system of supply–demand relationship. Compared to
traditional systems, the future networked system of supply–demand relationship has four essential characteristics: proactive personalized consumption, direct centralized distribution, intelligent decentralized manufacture and ecological self-organized system. To facilitate such characteristics, a novel network, which evolves from traditional internet to a large-scaled, open-styled, self-organized and ecological intelligent network, called material conscious information network (MCIN), is required by the aforementioned networked system.

The MCIN has essential logics of openness, federation, cooperation and sharing, and characteristics of consistency, efficiency, innovation and sustainability. In the MCIN, the nodes are information counterparts of person, enterprise, administrative department and thing in both material and conscious worlds. Any relationship between the MCIN nodes is a supply–demand relationship. From a behavioral perspective, interactions between any nodes of the MCIN are behaviors or activities of trading in some extent. Each node of the MCIN is facilitated with a set of intelligent federated agents which enable the nodes to perceive, think, interact and evolve adaptively.

Using a counterpart to reflect material-conscious individuals on information world mostly emerges in the research of cyber-physical system, e.g. Cyberself, Cyber-Individual, etc. Cyberself (Robinson, 2007) is proposed to be an embodiment of an individual as a means of identity signaling and as a medium of online interaction, whereas Cyber-Individual (Wen et al., 2009), with a short term as “Cyber-I”, is the unique counterpart of a real person on cyberspace and also a digital counterpart going with one’s whole life. However, studies on the counterpart of a realistic individual remain in the conceptual phase and is incapable to model the MCIN.

Modeling of the MCIN is a two-step process. In the first step, each node of the MCIN is modeled in a generic way. Second, an addressing approach for each node in the MCIN is defined. There are three main challenges in the modeling of the MCIN:

1. how can we model the basic and supply–demand information of each kind of nodes generically?
2. how can we build the architecture for the intelligent federated agents of each node to achieve supply–demand interactions spontaneously?
3. how can we define a supply–demand-oriented addressing method for each node?

In our previous work, we proposed cyber-anima (Shen, 2015), a semantic model to reflect individuals in the cyberspace and reason their desires and intentions for recommender systems. Cyber-anima has six dimensions, namely, physiology, belief, character, knowledge, experience and context, which correspond to the intrinsic and extrinsic factors of humanity. On the one hand, cyber-anima possesses the generalization which can develop to various disciplines, e.g. biology, medical science, psychology, sociology, statistics, economics, management science, politics, philosophy, anthropology, etc. On the other hand, a four-unitary semantic structure of dimension, concept, relation and axiom empowers cyber-anima to reflect any realistic participants to cyberspace. Based on cyber-anima, an anima-desire-intention-based (ADI-based) reasoning approach is proposed for recommendation interactions, which can also be extended to the scenario of trading-styled interactions.

In this work, we present an approach based on cyber-anima to model the MCIN. Leveraging and refining the dimensions of structure, character, knowledge and experience, we extend cyber-anima to fit the basic information of all kinds of the MCIN nodes, e.g. persons, enterprises, departments of government and thing, and present a double-chain structure for both basic and supply–demand information of each node. We propose an ADI-based architecture for the intelligent federated agents of each node in the MCIN. We propose a six-degrees-of-separation blockchain for the MCIN node addressing. We develop a
prototype of novel E-commerce platform based on the MCIN to prove the effectiveness and soundness of the MCIN modeling.

The main contributions of this paper are as follows:

- design a generic model for the basic and supply–demand information of the MCIN nodes based on cyber-anima;
- propose an ADI-based multi-agent architecture for the MCIN nodes to drive their spontaneous trading-styled interactions; and
- present a novel node addressing method based on six-degrees-of-separation blockchain for the MCIN.

The rest of the paper is organized as follows: Section 2 discusses existing related work and highlights the key contributions of this study of work. Section 3 provides a detailed elucidation of the generic model for the MCIN nodes. Section 4 focuses on the multi-agent architecture of the MCIN nodes. Section 5 presents the node addressing method in the MCIN. Section 6 develops a prototype system based on the modeled MCIN and discusses its efficiency. Section 7 concludes the work and provides a summary of future work.

2. Literature review

The research on cyber-physical system has been quite popular recently. Using counterparts to reflect realistic individuals is the first step to modeling a cyber-physical system. To this end, several concepts have been proposed. Robinson proposed an embodiment of individual as a means of identity signaling and as a medium of online interaction, called Cyberself (Robinson, 2007). Besides Cyberself, Robinson also presents Cyberpersona and Cyberbody as the counterparts of persona and body in cyberspace. Wen et al. proposes Cyber-Individual, with a short term “Cyber-I”, to create a unique, digital, comprehensive description for every real person existing in the cyberspace including human’s social context, mood, temper, physical status and so on (Wen et al., 2009). Cyber-I is a real individual’s counterpart in the cyberspace. However, research of this kind is still in its conceptual phase and cannot be applied to the modeling of the cyber-physical system directly.

Eskins et al. presents a framework of modeling Cyber-Human Systems (CHS) which divides CHS elements into four types: components, participants, processes and tasks (Eskins et al., 2001). Each type of CHS elements is related to a set of properties. The framework of modeling CHS provides a structured and quantitative method of analyzing cyber security problems whose outcomes are influenced by human–system interactions. Hence, such a framework can only be applied to a specific domain and its modeling of CHS lacks of generality and semantization.

In our previous work, we proposed a semantic model, called cyber-anima, to reflect individuals in the cyberspace and reason their desires and intentions for recommender systems (Shen, 2015). Cyber-anima composing of physiology, belief, character, knowledge, experience and context represents the intrinsic and extrinsic factors of user humanity. We performed a study on the semantic modeling approach based on the cyber-anima and proposed a multi-phase, multi-task and multi-technology mechanism for learning and populating the terms, concepts, taxonomy relations, non-taxonomy relations and axioms in the cyber-anima. We presented the cyber-anima-based reasoning approaches to derive individual user desire and intention, which leverage the existing or past individual desires and individual intentions with a template to reason over with. The cyber-anima can produce individual desire-oriented user profile and individual intention-oriented recommendation based on cyber-anima model. Albeit cyber-anima was proposed for enhancing the...
performance of the recommender system. Its generic and semantic modeling approach, model learning mechanism and ADI-based reasoning framework can be extended to model the novel network of supply–demand relationships, especially for the MCIN.

3. A cyber-anima-based model of the material conscious information network node

The first main challenge in modeling of the MCIN is: how can we model the basic and supply–demand information of each kind of nodes generically? In this section, we introduce a generic model for both basic and supply–demand information of the MCIN nodes based on cyber-anima.

3.1 Model the basic information

The MCIN is a large-scaled, open-styled, self-organized and ecological intelligent network. The nodes of the MCIN are information counterparts of supply and demand sides in realistic trading activities or interactions, which can be divided into four categories: person, enterprise, administrative department and thing. Each of the MCIN node has basic and supply–demand information. The basic information reflects the material world of a node, whereas the supply–demand information represents the history of the conscious world of a node.

In the light of the aforementioned characteristics, a model which can generically represent the basic information of all four categories of the MCIN nodes should be designed from a cross-disciplinary perspective. In particular, each category of the MCIN nodes relates to both general and specific disciplines, shown in Table I.

In our previous work (Shen, 2015), we proposed a semantic model, called cyber-anima, to reflect individuals in the cyberspace and reason their desires and intentions for recommender systems. In light of the equivalence between “individual” and “person” in our previous work and this work, respectively, the basic information of the MCIN person node can be modeled by cyber-anima. Without the loss of generality, we formalize the basic information of The MCIN nodes based on cyber-anima as tuple $I_B$ with four attributes, as explained below.

**Definition 1** (Model of the basic information of the MCIN node). The basic information of the MCIN node is modeled based on cyber-anima as tuple $I_B$ with four attributes:

$$I_B : = < D_{ST}, D_{CH}, D_{EX}, D_{KN} >$$

where:

- $D_{ST}$ is the structure dimension;
- $D_{CH}$ is the character dimension;
- $D_{EX}$ is the experience dimension; and
- $D_{KN}$ is the knowledge dimension.

<table>
<thead>
<tr>
<th>Category</th>
<th>General discipline</th>
<th>Specific discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>Economics, natural science</td>
<td>Biology, medical science, psychology, sociology, anthropology</td>
</tr>
<tr>
<td>Enterprise</td>
<td>Organizational behavior, management science, operations research, statistics, sociology, anthropology</td>
<td></td>
</tr>
<tr>
<td>Administrative department</td>
<td>Public administration, politics, management science, statistics, sociology, anthropology</td>
<td></td>
</tr>
<tr>
<td>Thing</td>
<td>Engineering science, design science, statistics</td>
<td></td>
</tr>
</tbody>
</table>
In accordance with Definition 1, the dimensions of the basic information of the MCIN person node are designed with cross-disciplinary perspective shown in Table II. The basic information of the MCIN person node has four dimensions including structure, character, experience and knowledge.

For the sake of self-organized, the basic information of The MCIN nodes are modeled semantically. Dimension is the maximum granularity composition of the semantic model. Formally, we define a dimension of the basic information of the MCIN nodes as tuple $D$ with two attributes, as explained below.

**Definition 2 (Dimension).** A dimension of the basic information of the MCIN node is modeled as tuple $D$ with two attributes:

$$D :: = < C, R >$$

where:

$C$ is the set of concepts, and $C :: = < MC, AXC >$, where $MC$ is the marker of concept, and $AXC$ is the set of concept axioms; and

$R$ is the set of relations, and $R :: = < MR, CP, AXR >$, where $MR$ is the marker of relation, $CP$ is the pair of concept, and $AXR$ is the set of relation axioms.

To clarify the representation of dimensions, we give each dimension several classifications to distinguish their concepts with different perspectives. For example, the structure dimension of the basic information of the MCIN person node has two classifications of DNA and physique distinguished in perspectives of biology and medical science, respectively. Accordingly, we can extend the MCIN person node to model the other three categories of the MCIN nodes with identical dimensions but different classifications of concept. Tables III-V show the dimensions of the basic information of the MCIN enterprise node, the MCIN administrative department node and the MCIN thing node with their classifications and perspectives.

Based on cyber-anima, the basic information of all four categories of the MCIN nodes has been modeled generically and semantically. In maximum granularity level, such models

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Classification</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure (D&lt;sub&gt;ST&lt;/sub&gt;)</td>
<td>DNA, physique</td>
<td>Biology, medical science, natural science</td>
</tr>
<tr>
<td>Character (D&lt;sub&gt;CH&lt;/sub&gt;)</td>
<td>16PF, Big Five, psychogenic needs</td>
<td>Psychology, sociology</td>
</tr>
<tr>
<td>Experience (D&lt;sub&gt;EX&lt;/sub&gt;)</td>
<td>Life, work, study, trade</td>
<td>Psychology, sociology</td>
</tr>
<tr>
<td>Knowledge (D&lt;sub&gt;KN&lt;/sub&gt;)</td>
<td>Discipline, skill</td>
<td>Sociology, psychology</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Classification</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure (D&lt;sub&gt;ST&lt;/sub&gt;)</td>
<td>Finance, ownership, organizational structure</td>
<td>Organizational behavior, economics, management science, statistics, natural science</td>
</tr>
<tr>
<td>Character (D&lt;sub&gt;CH&lt;/sub&gt;)</td>
<td>Brand, corporate culture</td>
<td>Organizational behavior, management science, sociology, philosophy</td>
</tr>
<tr>
<td>Experience (D&lt;sub&gt;EX&lt;/sub&gt;)</td>
<td>R&amp;D, marketing, management, production, trade</td>
<td>Organizational behavior, management science, operations research, sociology</td>
</tr>
<tr>
<td>Knowledge (D&lt;sub&gt;KN&lt;/sub&gt;)</td>
<td>Production management, operation</td>
<td>Economics, management science, operations research, statistics</td>
</tr>
</tbody>
</table>
have uniform dimensions: structure, character, experience and knowledge. Albeit the dimensions of the basic information of the MCIN nodes are identical in semantic structure, they are distinguished by different classifications generated in different perspectives. We use these classifications as an explanatory tool, so these classifications of dimensions illustrated above are not exclusive and varied with weights of blended perspectives.

3.2 Construct information for the material conscious information network node

The basic information of The MCIN nodes has been generically and semantically modeled with four dimensions including structure, character, experience and knowledge. As we all know, anything in the material world will be changed by the environment eventually with the time. The dimensions of the basic information of the MCIN nodes have to be synchronized with these changes. As aforementioned, the history of the conscious world of the MCIN node can be represented by its supply–demand information. In a temporal perspective, the two worlds of a the MCIN node interact as both cause and effect. For the sake of modeling, the MCIN nodes, both basic and supply–demand information have to be constructed in a temporal chain styled formation. Formally, we define the model of the MCIN nodes as tuple $A$ with four attributes, as explained below.

**Definition 3** (Model of the MCIN node). An MCIN node is modeled based on cyber-anima as tuple $A$ with four attributes:

$$A := < I_B, I_S, I_D, T >$$

where:

- $I_B$ is the basic information of a MCIN node;
- $I_S$ is the supply information of a MCIN node;
- $I_D$ is the demand information of a MCIN node; and
- $T$ is the set of time points.

### Table IV.
Dimensions of the basic information of the MCIN administrative department node

<table>
<thead>
<tr>
<th>Dimension (D&lt;sub&gt;ST&lt;/sub&gt;)</th>
<th>Classification</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>Finance, system</td>
<td>Public administration, politics, management science, statistics, natural science</td>
</tr>
<tr>
<td>Character (D&lt;sub&gt;CH&lt;/sub&gt;)</td>
<td>Style of work, administrative culture</td>
<td>Public administration, politics, management science, sociology</td>
</tr>
<tr>
<td>Experience (D&lt;sub&gt;EX&lt;/sub&gt;)</td>
<td>Administrative affair, function adjustment, trade</td>
<td>Public administration, politics, economics, sociology</td>
</tr>
<tr>
<td>Knowledge (D&lt;sub&gt;KN&lt;/sub&gt;)</td>
<td>Administration, service, control</td>
<td>Public administration, politics, economics, management science</td>
</tr>
</tbody>
</table>

### Table V.
Dimensions of the basic information of the MCIN thing node

<table>
<thead>
<tr>
<th>Dimension (D&lt;sub&gt;ST&lt;/sub&gt;)</th>
<th>Classification</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>Material, composition</td>
<td>Natural science, engineering science, design science, statistics</td>
</tr>
<tr>
<td>Character (D&lt;sub&gt;CH&lt;/sub&gt;)</td>
<td>Function, quality, size, manufacture cost</td>
<td>Economics, natural science, engineering science, design science, statistics</td>
</tr>
<tr>
<td>Experience (D&lt;sub&gt;EX&lt;/sub&gt;)</td>
<td>Design, manufacture, logistics, install</td>
<td>Economics, engineering science, design science, statistics</td>
</tr>
<tr>
<td>Knowledge (D&lt;sub&gt;KN&lt;/sub&gt;)</td>
<td>Technology, patent</td>
<td>Natural science, engineering science, design science</td>
</tr>
</tbody>
</table>
The model of the MCIN node has a DNA-styled double-chain structure. As shown in Figure 1, there are two chains in the model which represent the supply information and demand information, respectively. The connector of these two chains stands for the basic information of the MCIN node. The whole chain is growing with time, and each minimum segment of the double chain refers to the basic and supply–demand information of the MCIN nodes in a particular time point.

4. An anima-desire-intention-based multi-agent architecture for the material conscious information network nodes

The second main challenge in modeling of the MCIN is: how can we build the architecture for the intelligent federated agents of each node to achieve supply–demand interactions spontaneously? After modeling the MCIN nodes based on cyber-anima, we explore the architecture of a multi-agent for the MCIN nodes to enable the spontaneous interactions in this section.

4.1 Spontaneous interactions required for the material conscious information network nodes

The MCIN is a ternary system of material, conscious and information, which networks any supply–demand relationships. Figure 2 illustrates the interactions between the MCIN nodes which are in a trading style. Services and information are demanded and supplied by the MCIN nodes mutually.

No matter which kind of the MCIN nodes, they have the same pattern of interactions including creating supply–demand information, searching supply–demand information, conforming supply–demand information, delivering service, searching particular nodes, linking with other nodes, etc. Though persons, enterprises, administrative departments and things can interact among themselves, most of such time-consuming work has better been efficiently accomplished by intelligent federated agents facilitated by the MCIN nodes.

These intelligent federated agents perceive, think, interact and evolve as a whole through cooperating and collaborating. Each agent has its particular intelligence and generates outputs by handling different inputs. As an example, some of the agents can handle vision (graphics), auditory (audio) and touch (pressure, vibration, temperature...
and humidity) inputs and generate manageable outputs, and others can think of the subsequent reactions toward such outputs. Different MCIN nodes have different levels of intelligence on account of its agents and can evolve to a higher level via interacting mutually.

4.2 Build anima-desire-intention-based multi-agent architecture
For the purpose of perceiving, thinking, behaving and evolving as a whole, the federated agents of the MCIN nodes require an efficient multi-agent architecture to self-organize. In our previous work (Shen, 2015), we proposed an ADI-based multi-agent architecture for recommender system. As intelligent interactions between the MCIN nodes have a similar scenario of recommending items between users, we can extend and use the architecture for the agents of the MCIN nodes. To this end, we formalize the spontaneous interaction generation problem as follows:

Definition 4 (Spontaneous interaction generation problem). For each MCIN node, \( n \in N \) has a desire \( d_w^n \in D_n \) and an interaction \( i_w^n \in I_n \) on time point \( t \in T \) in possible world \( w \in W \) according to its Cyber-Anima \( A_n \), we want to choose such interaction \( i_w^n \) that

\[
\hat{i}_w^n = \delta (\text{maxexpval}, d_w^n).
\]

where:

- \( W \) is a set of possible worlds, and for each \( w \in W \), we have \( w = < T_w, A_w, E_w > \).
- \( T_w \subseteq T \) is a set of time points in possible world \( w \); \( A_w \) is a binary relation of time points, namely, \( \forall i(t_i, t_{i+1}) \in A_w ; E_w \) is a set of event type in possible world \( w \);
- \( D_n \) is the possible desire world of The MCIN nodes \( n \in N \), and \( D_n \subseteq W \times T \times W \);
- \( I_n \) is the possible interaction world of The MCIN nodes \( n \in N \), and \( I_n \subseteq W \times T \times W \);
- \( \delta \) is the deliberation function;
- \( A_n \) is the Cyber-Anima model of The MCIN nodes \( n \in N \); and
- \( \text{maxexpval} \) is the maximizing expected value.
In light of Definition 4, generating an optimum interaction for a particular MCIN node in a specific time point is a twofold process shown in Figure 3. First, we reason the desire $d_{ti}^{we}$ of the MCIN node based on its cyber-anima model. Second, we reason and find such an interaction $x_{ti}^{we}$ that $x_{ti}^{we} = \delta (\text{maxexpval}, d_{ti}^{we})$.

For the purpose of reasoning the desire of the MCIN node, we acquire the model of the MCIN node model $A_n$ on current time point $t_i$, then drive the agents for perceiving and thinking to reason each event type $\epsilon_{ti}^{we}$ of node’s desire $d_{ti}^{we}$ along with its payoff value $PAYOFF(\epsilon_{ti}^{we})$ on current time point $t_i$ based on the MCIN node model $A_n$. The node’s desire $d_{ti}^{we}$ illustrates who it expects to interact with and what service or information it expects to interact for. Next, we drive the agents for perceiving and thinking again to reason the node’s interaction $x_{ti}^{we}$ which illustrates the probability $PROB(\epsilon_{ti}^{we})$ and payoff value $PAYOFF(\epsilon_{ti}^{we})$ toward each candidate to be interacted on current time point $t_i$ based on node’s desire $d_{ti}^{we}$. Finally, we sum up the utility value $PAYOFF(\epsilon_{ti}^{we}) \cdot PROB(\epsilon_{ti}^{we})$ of each candidate in the node’s interaction $x_{ti}^{we}$ and sort for the optimal interactions. As the model, desire and interaction of the MCIN node are built on material world, conscious world and information world, respectively, the process of generating spontaneous interaction based on ADI reasoning is an exploration from unary world to ternary world via binary world.

Furthermore, the process of evolving intelligence level of the MCIN nodes is just a remix of ADI reasoning process, namely ADI evolving. In ADI evolving, we can explore from ternary world to unary world via binary world, to train the agents with given sets of model, desire and interaction of the MCIN node.

5. An addressing method based on six-degrees-of-separation blockchain
The last main challenge in modeling of the MCIN is: how can we define a supply-demand oriented addressing method for each node? As modeling of single MCIN nodes has been discussed, this section explains the networking of the MCIN nodes and addressing method in the MCIN.

5.1 Build the material conscious information network based on six-degrees-of-separation blockchain
As a future networked system of supply-demand relationship, the MCIN has the characteristics of coevolving intelligence, decentralized organization and efficient addressing. Coevolving intelligence relies on the multi-agent architecture of the intelligent federated agents of the MCIN nodes, which has been discussed in the last sections. Once an MCIN node is created, it would like to link to other nodes, evolve constantly and never destroy. For the purpose of linking, these MCIN nodes together and building the decentralized network, we propose a six-degrees-of-separation blockchain.
Blockchain (Nakamoto, 2008) is a kind of distributed database, which serves as the public ledger for all transactions of Bitcoin, a decentralized digital currency, after conceptualization. We use blockchain to enable the decentralization of the MCIN and make the following specifications of the MCIN blockchain:

- each MCIN person, enterprise or administrative department node must join a blockchain with at least $N$ different The MCIN nodes, where $N = \sqrt{\text{total amount of MCIN person, enterprise and administrative department nodes}}$;
- each MCIN thing node must join a blockchain with at least five different the MCIN nodes; and
- each MCIN blockchain must have at least one MCIN administrative department node.

The joining of the MCIN blockchain is driven by a supply–demand relationship. Whenever the MCIN nodes construct new supply–demand relationships with other the MCIN nodes, they may join or create new MCIN blockchain. As one of the interactions, the selection of blockchain is accomplished by the agents of the MCIN nodes spontaneously according to abovementioned specifications. Therefore, in each MCIN blockchain, we have the following truths:

- each member of the blockchain has a supply–demand relationship with at least one other member;
- all members of a blockchain maintain an equilibrium of supply–demand relationship; once the equilibrium is broken, new members may join the blockchain; otherwise, the old members may join or create new blockchain; and
- via at most six MCIN blockchains, any two MCIN nodes can link with each other. It means that the MCIN blockchain satisfies six-degrees-of-seperation (Milgram, 1967).

Leveraging the MCIN blockchain, each MCIN node can acquire a unique, immutable address. Formally, we define the address of the MCIN nodes as follows:

**Definition 5** (the MCIN blockchain and node address). For each MCIN blockchain $bc \in BC$ has a member set of the MCIN nodes $N$. Then, we have $ADDR_{bc} = Hash(I_{B_1}, I_{B_2}, I_{B_3}, \ldots, I_{B_n})$. For each MCIN node $n \in N$ in the MCIN blockchain $bc$, we have $ADDR_n = ADDR_{bc} + Hash(I_{B_n})$.

where:

$I_B$ is the basic information of a MCIN node;
$ADDR_{bc}$ is the address of the MCIN blockchain $bc \in BC$;
$ADDR_n$ is the address of the MCIN nodes $n \in N$; and
$Hash$ is the hash function to generate the address code.

Figure 4 illustrates the process of generating the MCIN blockchain and node address. As an example, China government, Beijing city government and Haidian district government are the MCIN administrative department nodes, which are linked to the same MCIN blockchain. In the light of Definition 5, the address of the MCIN blockchain can be generated by hashing all basic information of these three MCIN administrative department nodes, which is a unique, immutable address. Meanwhile, the address of a single MCIN node, e.g. China government, can be generated by hashing its basic information concatenating the address of the MCIN blockchain.
5.2 Supply–demand-oriented addressing

As a decentralized network, the MCIN requires an efficient addressing method. Compared to a centralized network, the addressing method of the MCIN has the characteristics of socialization, information integration, collective intelligence, traceability, high robustness, unification of producing and consuming, high scalability and decentralization. As the MCIN is a future networked system of supply–demand relationship, we propose a supply–demand-oriented addressing method to locate target nodes efficiently via supply–demand relationships.

In light of Definition 3, the MCIN nodes are modeled with supply and demand chains, i.e. the supply information $I_S$ and the demand information $I_D$, which represent the history of supply and demand respectively. The supply and demand information is materialization of conscious world of the MCIN nodes and can be modeled in the same way. Formally, we define the model of the supply and demand information as tuple with two attributes, as explained below:

**Definition 6 (Model of the supply and demand information of the MCIN node).** The supply and demand information of the MCIN nodes are modeled as tuple $I_S$ and $I_D$ with two attributes:

$\quad I_S = I_D : = < R, P >$

where:

- $R$ is a set of supply and demand request, we have $R : = < C_{\text{object}}, C_{\text{attribute}} >$, where $C_{\text{object}}$ is the set of concepts of request object, and $C_{\text{attribute}}$ is the set of concepts of the attributes correspondent to the object; and

- $P$ is a set of addressing path correspondent to the request. For each addressing path $p \in P$, we have the partially ordered set $p = (ADDR_1, \ldots, ADDR_n)$. 

---

**Figure 4.** Process of generating the MCIN blockchain and node address.
The request of the MCIN node addressing is a segment of the supply and demand information. The supply–demand-oriented addressing method begins from the MCIN node of requester and traverses a certain number (not all) of blockchains to locate the MCIN nodes which match the request. Leveraging the supply and demand information of the MCIN nodes, the performance of addressing can be improved with the amount of similar requests. Figure 5 illustrates an example of supply–demand-oriented addressing in the MCIN. In Figure 5(a), Jerry launches a request, and the addressing of his request first starts in his directly linked blockchain A to find matched requests in supply and demand information. Unfortunately, no such requests matches. Then, in Figure 5(b), the addressing extends to adjacent blockchain A and B via Haidian District Government. This time, a matched supply request is found in the supply information of Supermarket. As Jerry and Supermarket have a supply–demand relationship after the addressing, a new blockchain D is created, and the path of Jerry’s request is from Jerry to Supermarket via blockchain D. In Figure 5(c), Angela launches a similar request, and the addressing of her request first starts in blockchain A as well. Obviously, Jerry’s demand request matches. Then in Figure 5(d), a matched supply request is found in the supply information of Supermarket by leveraging the path of Jerry's previous demand request.

Figure 5. Process of supply–demand-oriented addressing
As Angela and Supermarket have a supply–demand relationship after the addressing, a new blockchain E is created, and the path of Angela’s request is from Angela to Supermarket via blockchain E. The first addressing traverses 3 MCIN blockchains and 7 MCIN nodes, whereas the second addressing for the similar request traverses 1 MCIN blockchain and 4 MCIN nodes. With the supply–demand-oriented addressing method, the same type of similar requests in others’ supply–demand information can efficiently improve the addressing performance.

6. Prototype implementation and discussion
The MCIN is proposed for the purpose of implementing the future networked system of supply–demand relationship. Compared to traditional systems, the future networked system of supply–demand relationship has four essential characteristics: proactive personalized consumption, direct centralized distribution, intelligent decentralized manufacture and ecological self-organized system. The future networked system of supply–demand relationship has to build its own operation platform, a novel E-commerce platform.

In the last sections, we have generically modeled the MCIN nodes based on cyber-anima, presented an ADI-based multi-agent architecture for intelligent federated agents of the MCIN nodes and proposed an addressing method for the MCIN. In this section, we implement the prototype system of the novel E-commerce platform relied on the modeling of the MCIN. The prototype system composes of four function modules: Who am I; My demand; My supply; and My space.

The novel E-commerce platform is a personalized portal for every unique MCIN node, which means that any the MCIN node has its own independent portal. Intuitively, we call each MCIN node the owner of its platform. The model of the MCIN node drives the platform for the information counterpart of its owner in the MCIN autonomously, and meanwhile, the owner can operate the platform manually. The platform shown in Figure 6 shows the 3D virtual image of its owner and entrances of the four function modules at a first glance.

![Figure 6. The novel e-commerce platform at a first glance](image)
The “Who am I” function module is a snapshot of the owner’s model of the MCIN node. In this function module, we can overview interested basic and supply–demand information of owner’s model of the MCIN node to know who it is. Figure 7 shows the “Who am I” function module of the MCIN person node, including her 3D virtual image, physique-related concepts of structure dimension, values-related concepts of character dimension, trade-related concepts of experience dimension, etc.

The “My demand” function module tackles all demands of its owner rely on owner’s model of the MCIN node and supply–demand-oriented addressing method, covering pull-styled demand and push-styled demand. Any interaction in the MCIN is trade-styled and associated with a demand part and a supply part, so the owner is a demand part in “My demand” function module. The owner’s demand can proactively be launched by itself or passively be elicited by the supply side, which are so-called pull-styled and push-styled, respectively. No matter who launches the demand, an optimal matching of demand and supply sides can be generated by the supply-demand oriented addressing. Figure 8 illustrates the supply-demand oriented addressing snapshot for a pull-styled demand of automobile.

Relatively, “My supply” function module handles all supplies of its owner relying on an owner’s model of the MCIN node and the supply-demand oriented addressing method, covering pull-styled supply and push-styled supply. As the supply part in the “My supply” function module, the owner’s supply can proactively be launched by itself or passively be elicited by the demand side, which are so-called pull-styled and push-styled respectively. No matter who launches the supply, an optimal matching of demand and supply sides can be generated by supply–demand-oriented addressing. Figure 9 shows the supply interface.

The “My space” function module is a visualization of the MCIN from the owners’ perspective. Figure 10 presents a snapshot of a “My space” function module. The node in the center represents the owner of the platform, while other nodes around represent persons, enterprises, administrative departments and things directly and indirectly linking to the owner via six-degrees-of-separation blockchain. Driven by the owner’s model of the MCIN node and the supply-demand oriented addressing method, the owner’s MCIN node would be linked in more and more six-degrees-of-separation blockchain.

Figure 7.
“Who am I” function module
Compared to traditional ones, the novel E-commerce platform enhanced by the MCIN has the advantages in:

- directly connection with consumer end and producer end directly;
- intelligent trade matching; and
- personalized portal.

The modeling of nodes based on cyber-anima, the multi-agent architecture based on ADI and the addressing method oriented to supply and demand make the main differences between the MCIN and current internet, as shown in Table VI.
7. Conclusion and future work
The MCIN is defined to be a large-scaled, open-styled, self-organized and ecological intelligent network, which evolves to the next generation of internet. The MCIN nodes are information counterparts of person, enterprise, administrative department and thing in both material and conscious worlds. Interactions between any MCIN node are trading-styled interactions to some extent. Each MCIN node equips a set of intelligent federated agents which enable the nodes perceive, think, interact and evolve adaptively. To model the MCIN, we have several main challenges to tackle.

To this end, we present an approach based on cyber-anima to model the MCIN. We extend the cyber-anima model with the dimensions of structure, character, knowledge and experience to fit the basic information of all kinds of the MCIN nodes, e.g. persons, enterprises, departments of government and thing. We present a double-chain structure for both basic and supply–demand information of each node. We propose an ADI-based
architecture for the intelligent federated agents of each node in the MCIN. We present a six-
degrees-of-separation blockchain for the MCIN node addressing. We develop a prototype of
novel E-commerce platform based on the MCIN to prove the effectiveness and soundness of
the MCIN modeling.

We intend to continue our research in the following two directions:

(1) implement the ADI-based multi-agent architecture of the MCIN; and
(2) enhance the function of the prototype system and deploy it to industrial
application.

References
January 2017).
Shen, J. (2015), A Study on Cyber-Anima Model and Its Application in Intelligent Recommendation: [D],
Fudan University.
age”, New Media & Society, Vol. 9, pp. 93-110.

Further reading
modeling cyber-human systems”, Eighth International Conference on Quantitative Evaluation of

Corresponding author
Jianping Shen can be contacted at: shenjianping@tsinghua.edu.cn
Experimental results on large-scale cyber-physical hybrid discussion support

Takayuki Ito, Takanobu Otsuka, Satoshi Kawase, Akihisa Sengoku, Shun Shiramatsu, Takanori Ito and Eizo Hideshima
Nagoya Institute of Technology, Nagoya, Japan

Tokuro Matsuo, Tetsuya Oishi and Rieko Fujita
Advanced Institute of Industrial Technology, Tokyo, Japan

Naoki Fukuta
Shizuoka University, Hamamatsu, Japan, and
Katsuhide Fujita
Tokyo University of Agriculture and Technology, Tokyo, Japan

Abstract

Purpose – This paper aims to present a preliminary experimental result on a large-scale experiment on a cyber-physical hybrid discussion support environment in a panel discussion session in an international conference.

Design/methodology/approach – In this paper, the authors propose a hybrid (cyber-physical) environment in which people can discuss online and also offline simultaneously. The authors conducted a large-scale experiment in a panel discussion session in an international conference where participants can discuss by using their online discussion support system and by physical communications as usual.

Findings – The authors analyzed the obtained data from the following three viewpoints: participants’ cyber-physical attention, keywords cyber-physical linkage and cyber-physical discussion flow. These three viewpoints indicate that the methodology of the authors can be effective to support hybrid large-scale discussions.

Originality/value – Online large-scale discussion has been focused as a new methodology that enable people to discuss, argue and make consensus in terms of political issues, social complex problems (like climate change), city planning and so on. In several cases, the authors found that online discussions are very effective to gather people opinions and discussions so far. Moreover, this paper proposes a hybrid (cyber-physical) environment in which people can discuss online and also offline simultaneously.

Keywords Gamification, Citizen science, Crowd AI, Crowd as a service, Harnessing the crowd in human-computer interaction

Paper type Research paper

© Takayuki Ito, Takanobu Otsuka, Satoshi Kawase, Akihisa Sengoku, Shun Shiramatsu, Takanori Ito, Eizo Hideshima, Tokuro Matsuo, Tetsuya Oishi, Rieko Fujita, Naoki Fukuta and Katsuhide Fujita. Published in the International Journal of Crowd Science. Published by Emerald Publishing. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

This work is partially supported by JST CREST.
1. Introduction
Much attention has recently been focused on the experiments that gather large-scale opinion gathering (Malone et al., 2009; Klein, 2012). Research interest continues to increase in online crowd decision-making, which might become one of the next generation methods for open and public forums.

This paper presents preliminary experimental results on a large-scale experiment on a cyber-physical hybrid discussion support environment in a panel discussion session at an international conference. Our research group has been studying supporting technologies for online large-scale discussions. Online large-scale discussion has been focused as a new methodology that enable people to discuss, argue and make consensus in terms of political issues, social complex problems (like climate change), city planning and so on. By social experiments that collaborate with some town meetings introducing the Web-based forum system, we found that online discussions are very effective to gather opinions from the participants and discussions so far. Moreover, in this paper, we propose a hybrid (cyber-physical) environment in which people can discuss online and also offline simultaneously.

2. Background: large-scale discussion support system
To harness large-scale discussion intelligently, there are several critical factors including facilitation, incentives and understanding. These factors can make the entire discussion be held in fruitful ways and avoid negative behaviors that encourage “flaming”. “Flaming” means a hostile and insulting interaction by Wikipedia.

An open Web-based forum system called COLLAGREE (Ito et al., 2014; Sengoku et al., 2016a) has facilitator support functions and an incentive mechanism for the large-scale opinion gathering. They held a two-week long online town meeting, Nagoya Next Generation Total City Planning, where people in Nagoya City, Japan, used COLLAGREE to discuss city-planning to operate the municipal administration of Nagoya from fiscal years 2014 to 2018. In the two weeks, COLLAGREE gathered 266 total registered participants, 1,151 opinions, 3,072 visits and 18,466 views. The results demonstrated that COLLAGREE succeeded in gathering many opinions, while people understood the importance of facilitators.

Figure 1 shows a typical user-interface used by both facilitators and participants. The following are its typical functions, and we especially adopted ①, ② and ③ to support facilitators. ① shows agreement or disagreement analysis for a comment is shown. Facilitators can understand whether a discussion thread is positive or negative. ② shows...
keywords are highlighted so that facilitators can understand what keywords are being focused on and which are important. ③ shows facilitation tab from which facilitators can input their instructions to participants. ④ shows searching and reordering opinions and discussions. ⑤ displays issue tags that participants can add to each opinion and comment so that they can search for it afterwards. ⑥ is e-mail reminders for participants as well as reminders when related events happen.

Nagoya in Aichi Prefecture has over three million people. After three months of preparation with its city officers, they created an internet-based town meeting about the Nagoya city planning. Mayor Takashi Kawamura announced this project in newspapers and on TV as one actual town meeting of the Nagoya Next Generation Total City Planning for 2014-2018.

The experiment ran on COLLAGREE system during a two-week period from 12.00 on November 19, 2013 to 12.00 on December 3, 2013 with nine expert facilitators from the Facilitators Association of Japan. The participants discussed about their ideal city based on the Nagoya Next Generation Total City Planning 2014-2018.

As preliminary results over the two weeks, COLLAGREE gathered 266 registered participants, 1,151 opinions, 3,072 visits and 18,466 views. The total of 1,151 opinions greatly exceeded the 463 opinions obtained by previous real-world town meetings. From the questionnaires, both participants and facilitators realized the importance. However, facilitators had difficulty managing such large-scale discussions because this was their first experience (Ito et al., 201).

In the work (Takahashi et al., 2016), they have proposed an incentive mechanism for large-scale collective discussions, where the discussion activities of each participant are rewarded based on their effectiveness. With these incentives, we encourage both the active and passive actions of participants. Active actions include posting opinions, replying and agreeing and should be done for warming up discussions. Passive actions, which include getting replies and gaining agreement from others, are more highly rewarded in our system. Such passive actions suggest that one’s opinions have received interest or are supported by others. In other words, they submitted opinions that did not lead to impassioned responses from other participants.

Further, they extended their incentive mechanism so that the mechanism can take the quality of opinions into account (Takahashi et al., 2016) by using a natural language processing technique called BM25. By measuring the quality of opinions, we successfully incentivized participants to submit different opinions at the different phases in a discussion.

Discussion Tree (Sengoku et al., 2016a, 2016b) is a tree diagram that visualizes the flow of a discussion on the basis of the reply relationships in the conversations to make the discussion more efficient. A major difference of Discussion Tree from the argumentation map used in Deliberatrium (Gurkan et al., 2010) is that the Discussion Tree is generated automatically from chunk texts submitted freely by participants on a discussion forum. In addition, our Discussion Tree uses text-mining techniques to present the important keywords in discussion contents. These features avoid imposing a load on participants, while the argumentation map requests participants to manually create a logical argumentation structure.

3. Cyber-physical discussion support and metrics
The experimental results show the online discussion support worked well. Moreover, we found that a hybrid approach to support discussion seems also work well. In the experiment in the Aichi design league in 2015 explained above, we found that people were very excited to discuss online and also offline simultaneously.
Thus, as one methodology to support large-scale discussion, we propose the cyber-physical discussion support methodology. This approach could influence each other between the physical world and the online world.

In discussion, while some people can say their opinions physically, the other people tend to hesitate to say their own opinions. If the discussion is large, such silent people might be in majority. Our hybrid approach becomes a possibility to resolve that silent majority can say something online.

This paper proposes the following three metrics which represent how discussion has been supported physically and virtually in our hybrid environment:

1. **Participants’ cyber-physical attention**: This metric represents how participants can participate in cyber discussion and also in physical discussion simultaneously by measuring how participants’ attention relates the number of views and postings in online discussion.

2. **Keywords cyber-physical linkage**: This metric measures how contents are interrelated between virtual and physical discussions by measuring how keywords appeared in both discussions.

3. **Cyber-physical discussion flow**: This metric measures how discussion flows online and offline by measuring relations temporal behaviors between virtual and physical discussions.

These metrics are currently preliminary and need to be discussed and improved. However, as an initial attempt, it is quite new to propose this kind of metrics as far as we know.

4. **A large-scale experiment and results**

4.1 **Setting**

We conducted an experiment in the panel session in the international congress on advanced applied informatics (AAI 2016):

- Conference name: 5th International Congress on Advanced Applied Informatics (AAI 2016).
- Session name: International Forum on Collective Intelligence and ICT Future
- Date: 2016.07.12, 2.30 p.m.-4.30 p.m.
- Location: Kumamoto City International Center, Kumamoto, Japan.

We have one facilitator who is in charge of facilitating physical discussion and four panelists who discuss about the following themes. The third theme was not discussed due to the time limitation:

1. Artificial intelligence (AI) is taking the place of human intelligence, e.g. AlphaGo. how does AI impact human intelligence?
2. AI is used in economy and government administration. How does AI impact the social evolution?
3. AI is applied to our infrastructures, i.e. control of distributing electricity. Is AI robust enough? What are the conditions for AI’s robustness? (this theme was not discussed due to the time limitation.)

A commentary participant was encouraged to make postings to online discussion. He is a kind of the leading participants who lead the others’ discussion.

Table I shows the actual timeline of this panel discussion.
4.2 Three metrics for supporting hybrid discussions

We conducted an experiment to validate the efficiency of cyber-physical discussion support by using the proposed metrics: participants’ cyber-physical attention, keywords cyber-physical linkage and cyber-physical discussion flow.

Participants’ cyber-physical attention: This metric represents how participants can participate in cyber discussion and also in physical discussion simultaneously by measuring how participants’ attention relates to the number of views and postings in online discussion.

We will compare the participation of the real-world discussion with the number of views and postings in the virtual discussion, and found that there is correlation between them. We will show the details of the results in the experimental results session. Here, we explain the experimental settings.

To measure the attention of participants, we installed several high-quality video cameras in the discussion room so that we can record the whole participants’ behaviors. Figure 2 shows the concrete arrangement of the cameras. We installed one camera for recording the stage and three cameras for recording participants.

We combined these three videos recorded by the three cameras with the software Final Cut Pro X by Apple so that we can easily recognize the situations both of panelists and participants. Also, we put time stamps to enable temporal analysis and comparisons between posting/viewing in the virtual world and discussion in the physical world. Figure 3 shows a one-shot of the combined movie-file.

We extracted pictures for each 20 min from this movie-file.

The participant is defined as the person who gives attention, that is, is attending, to real-world discussion if he/she satisfies one of the following conditions:

<table>
<thead>
<tr>
<th>Time</th>
<th>Duration</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.48-15.59</td>
<td>12 mins</td>
<td>System explanation</td>
</tr>
<tr>
<td>15.59-15.15</td>
<td>15 mins</td>
<td>Asked Theme 1 and responded by a panelist (Katsuhide)</td>
</tr>
<tr>
<td>15.15-15.21</td>
<td>6 mins</td>
<td>Asked Theme 2 and responded by a panelist (Andrew)</td>
</tr>
<tr>
<td>15.21-15.33</td>
<td>12 mins</td>
<td>Asked Theme 3</td>
</tr>
<tr>
<td>15.33-15.41</td>
<td>8 mins</td>
<td>QA by participants, question from the participants and facilitator encouraged to write questions on Collagree</td>
</tr>
</tbody>
</table>

Table I. Timeline of the panel discussion

<table>
<thead>
<tr>
<th>Time</th>
<th>Duration</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.41-16.04</td>
<td>15 mins</td>
<td>Panelists discussed about the opinions posted in Collagree</td>
</tr>
<tr>
<td>16.04-16.06</td>
<td>2 mins</td>
<td>Explanation of discussion</td>
</tr>
<tr>
<td>16.06-16.15</td>
<td>9 mins</td>
<td>Panelists’ final comments and facilitator wrapped up</td>
</tr>
</tbody>
</table>

Figure 2. Camera arrangement
he/she is looking ahead on the stage where the facilitator or panelists are there; and

he/she is looking at the questioner when there is a person who is asking a question.

Also, we assume the participant is participating in the virtual world discussion except for the above situations. We counted the above situations for each 20 s in the video, and sum up for each 5 min. Figure 4 shows the rate of the number of participants who are attending the real-world discussion, namely, looking forward or making a comment to discussions in the real world.

We compared the above participation of the real-world discussion with the number of views and postings in the virtual discussion and found that there is correlation between them. We will show these results in the experimental results session.

4.3 Keywords cyber-physical linkage

This metric measures how contents are interrelated between virtual and physical discussions by measuring how keywords appeared in both discussions in this paper. Ideally, this interrelation should be moderate, and different ideas from different perspectives should be generated online and offline.

We counted the frequency of the appeared keywords from the discussion among panelists recorded as texts, and also from the contents in the online (virtual) discussion. We extracted keywords manually while ignoring non-sense words and same-meaning words and ranked top 50 keywords by using BM25 algorithm (Robertson and Zaragoz, 2009). Table II shows the ranking of the keywords.

Based on the above scores, we found that there is efficient correlation between real-world and online discussions. The details will be shown in the experimental result session.
4.4 Cyber-physical discussion flow
This metric measures how discussion-flows interconnected online and offline by measuring relations temporal behaviors between virtual and physical discussions. In this paper, this is called “Cyber-physical discussion flow”. We have been analyzed several types of relations between real-world and virtual discussion. Then, we found that there is some correlation between the number of people who are looking ahead, and the number of views after 5 min after 5 min. Figure 5 shows the temporal data about the number of participants who are looking ahead and the number of views online. The details of the analyzed results will be shown in the experimental result session.

4.5 Evaluation and analysis
We evaluate the proposed three metrics: participants’ cyber-physical attention, keywords cyber-physical linkage and cyber-physical discussion flow shown in Section 4.2, 4.3 and 4.4, respectively, by calculating correlations based on the gathered data.

The parameters for calculating correlation coefficients are described as follows:

- Looking ahead: The number of participants who are looking ahead in the real-world.
- No. of views: The number of views online.
- No. of postings: The number of postings online.
- Length of No. of characters per post: The average number of characters per a post online.

Table III shows the Person correlation coefficients and significance probabilities (both sides) for each pair of the above parameters.

Based on the calculated results in Table III, we analyze the three metrics as follows.

4.6 Participants’ cyber-physical attention
This metric represents how participants can participate in cyber discussion and also in physical discussion simultaneously by measuring how participants’ attention relate to the number of views and postings in online discussion.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Physical space</th>
<th>Value</th>
<th>Cyber space</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Question</td>
<td>8.081766</td>
<td>Emotion</td>
<td>0.128389</td>
</tr>
<tr>
<td>2</td>
<td>Example</td>
<td>7.271388</td>
<td>Name</td>
<td>0.128389</td>
</tr>
<tr>
<td>3</td>
<td>Opinion</td>
<td>7.232191</td>
<td>Being</td>
<td>0.125664</td>
</tr>
<tr>
<td>4</td>
<td>Point</td>
<td>7.02837</td>
<td>Human</td>
<td>0.111402</td>
</tr>
<tr>
<td>5</td>
<td>People</td>
<td>6.513154</td>
<td>Problem</td>
<td>0.110833</td>
</tr>
<tr>
<td>6</td>
<td>Being</td>
<td>6.360155</td>
<td>Idea</td>
<td>0.110833</td>
</tr>
<tr>
<td>7</td>
<td>Issue</td>
<td>6.307973</td>
<td>People</td>
<td>0.095043</td>
</tr>
<tr>
<td>8</td>
<td>Emotion</td>
<td>5.853828</td>
<td>System</td>
<td>0.09154</td>
</tr>
<tr>
<td>9</td>
<td>Robot</td>
<td>5.210628</td>
<td>Something</td>
<td>0.09154</td>
</tr>
<tr>
<td>10</td>
<td>Intelligence</td>
<td>5.194714</td>
<td>Robot</td>
<td>0.085709</td>
</tr>
<tr>
<td>11</td>
<td>Future</td>
<td>4.600353</td>
<td>Use</td>
<td>0.085709</td>
</tr>
<tr>
<td>12</td>
<td>Responsibility</td>
<td>4.554622</td>
<td>Art</td>
<td>0.084179</td>
</tr>
<tr>
<td>13</td>
<td>Answer</td>
<td>4.508041</td>
<td>Person</td>
<td>0.077811</td>
</tr>
<tr>
<td>14</td>
<td>System</td>
<td>4.420725</td>
<td>Post</td>
<td>0.077811</td>
</tr>
<tr>
<td>15</td>
<td>Thing</td>
<td>4.392325</td>
<td>Rule</td>
<td>0.072855</td>
</tr>
<tr>
<td>16</td>
<td>Job</td>
<td>4.365512</td>
<td>Point</td>
<td>0.072855</td>
</tr>
<tr>
<td>17</td>
<td>Research</td>
<td>4.127652</td>
<td>Company</td>
<td>0.069357</td>
</tr>
<tr>
<td>18</td>
<td>Government</td>
<td>3.988191</td>
<td>Job</td>
<td>0.063857</td>
</tr>
<tr>
<td>19</td>
<td>Brain</td>
<td>3.911277</td>
<td>Law</td>
<td>0.063857</td>
</tr>
<tr>
<td>20</td>
<td>Infrastructure</td>
<td>3.784719</td>
<td>Work</td>
<td>0.063857</td>
</tr>
<tr>
<td>21</td>
<td>Problem</td>
<td>3.606552</td>
<td>Ethic</td>
<td>0.063857</td>
</tr>
<tr>
<td>22</td>
<td>Something</td>
<td>3.606552</td>
<td>student</td>
<td>0.063857</td>
</tr>
<tr>
<td>23</td>
<td>Function</td>
<td>3.549693</td>
<td>Future</td>
<td>0.063857</td>
</tr>
<tr>
<td>24</td>
<td>Comment</td>
<td>3.549693</td>
<td>Research</td>
<td>0.059806</td>
</tr>
<tr>
<td>25</td>
<td>Country</td>
<td>3.504786</td>
<td>Economy</td>
<td>0.059806</td>
</tr>
<tr>
<td>26</td>
<td>View</td>
<td>3.495502</td>
<td>Responsibility</td>
<td>0.059806</td>
</tr>
<tr>
<td>27</td>
<td>Thread</td>
<td>3.49565</td>
<td>Situation</td>
<td>0.059806</td>
</tr>
<tr>
<td>28</td>
<td>Law</td>
<td>3.46965</td>
<td>Brain</td>
<td>0.059806</td>
</tr>
<tr>
<td>29</td>
<td>Human</td>
<td>3.461004</td>
<td>Thing</td>
<td>0.054063</td>
</tr>
<tr>
<td>30</td>
<td>Sort</td>
<td>3.385105</td>
<td>Issue</td>
<td>0.054063</td>
</tr>
<tr>
<td>31</td>
<td>Datum</td>
<td>3.34411</td>
<td>Advantage</td>
<td>0.054063</td>
</tr>
<tr>
<td>32</td>
<td>Course</td>
<td>3.310993</td>
<td>Government</td>
<td>0.054063</td>
</tr>
<tr>
<td>33</td>
<td>Word</td>
<td>3.306047</td>
<td>Account</td>
<td>0.054063</td>
</tr>
<tr>
<td>34</td>
<td>Machine</td>
<td>3.19705</td>
<td>Partner</td>
<td>0.054063</td>
</tr>
<tr>
<td>35</td>
<td>Panelist</td>
<td>3.086104</td>
<td>Utility</td>
<td>0.054063</td>
</tr>
<tr>
<td>36</td>
<td>Information</td>
<td>3.033029</td>
<td>AI</td>
<td>0.054063</td>
</tr>
<tr>
<td>37</td>
<td>Nation</td>
<td>3.033029</td>
<td>Talk</td>
<td>0.054063</td>
</tr>
<tr>
<td>38</td>
<td>Feeling</td>
<td>3.033029</td>
<td>Topic</td>
<td>0.054063</td>
</tr>
<tr>
<td>39</td>
<td>Accident</td>
<td>3.033029</td>
<td>Electricity</td>
<td>0.054063</td>
</tr>
<tr>
<td>40</td>
<td>Floor</td>
<td>2.987705</td>
<td>technology</td>
<td>0.049837</td>
</tr>
<tr>
<td>41</td>
<td>Place</td>
<td>2.975786</td>
<td>Software</td>
<td>0.049837</td>
</tr>
<tr>
<td>42</td>
<td>Impact</td>
<td>2.906097</td>
<td>Creation</td>
<td>0.049837</td>
</tr>
<tr>
<td>43</td>
<td>Application</td>
<td>2.906097</td>
<td>Imagination</td>
<td>0.049837</td>
</tr>
<tr>
<td>44</td>
<td>Knowledge</td>
<td>2.906097</td>
<td>Novel</td>
<td>0.049837</td>
</tr>
<tr>
<td>45</td>
<td>University</td>
<td>2.906097</td>
<td>Profit</td>
<td>0.049837</td>
</tr>
<tr>
<td>46</td>
<td>Purpose</td>
<td>2.799884</td>
<td>Function</td>
<td>0.042917</td>
</tr>
<tr>
<td>47</td>
<td>Life</td>
<td>2.799884</td>
<td>Threat</td>
<td>0.042917</td>
</tr>
<tr>
<td>48</td>
<td>Factory</td>
<td>2.742671</td>
<td>World</td>
<td>0.042917</td>
</tr>
<tr>
<td>49</td>
<td>Technology</td>
<td>2.710613</td>
<td>Drive</td>
<td>0.042917</td>
</tr>
<tr>
<td>50</td>
<td>Game</td>
<td>2.61704</td>
<td>Car</td>
<td>0.042917</td>
</tr>
</tbody>
</table>

Table II. Scoring result of keywords
In Table III, the correlation between looking ahead and views is negatively significant. This means the number of views of Collagree system increases when the participants do not give any attention to real-world discussion, i.e. they do not look ahead, or vice versa. Namely, the participants always gave attention to real-world discussion or online discussion. From this result, we can conclude that the participants continuously attended the real-world or online discussion.

In the classic style panel discussion, i.e. only physical discussion, participants tend to be difficult to keep their attention or incentive to participate in the discussion if the discussion theme does not fit to their interest. Our methodology can overcome this situation and succeed to keep the participants’ motivation and attentive during this discussion session.

4.6.1 Keywords cyber-physical linkage. This metric measures how contents are interrelated between virtual and physical discussions by measuring how keywords appeared in both discussions. The correlation value of top 52 keywords between online and real world is $r = 0.339$ ($p = 0.024$), and it is significantly correlated. Further, the top 54
Keywords in online and real-world keywords is $r = 0.342$ ($p = 0.045$), and it is also significantly correlated.

These results show that the keywords in online and real-world are correlated. But the value of correlation coefficient is not higher. This means that discussion contents were somehow related but not completely the same. Namely, we can say that the discussion contents in virtual world were different from that in the real world. This contributed to the above participants’ cyber-physical attention as well.

4.6.2 Cyber-physical discussion flow. This metric measures how discussion flows online and offline by measuring relations temporal behaviors between virtual and physical discussions. Table III shows the temporal changes of the number of looking ahead (real world) and the number of views (online). We can say that the number of views increases 5 min after the number of looking ahead decreases. But, we cannot find the opposite situation. Namely, there is not the case that the number of looking ahead increases 5 min after the number of views increases. This implies the following story: when the participants are interested in the real-world discussion, they look ahead. And then, they tended to look into the virtual world discussion (the number of looking ahead decreases). Then, after 5 min, the number of views (online) increases.

Also, we analyzed the relation between the number of looking ahead (real-world) and the number of posting (online). The correlation value of the number of looking ahead and the number of posting at the same time is $-0.275$ (no significance). The correlation value of the number of looking ahead and the number of posting before 5 min is positively significant ($r = 0.504$, $p = 0.055$). There is no correlation between the number of looking ahead and the number of posting after 5 min ($r = -0.329$, n.s.). Namely, it can be said that the participants give attention to the real-world 5 min after posting online. But, there is no relation in the opposite case. This implies that the participants have interest to see how their posting make effect to the real-world discussion. Also, it implies that real-world discussion did not incentivize posting activities online. This could imply that the cyber-physical discussion flow would be asymmetric relation, and further investigation would be required. Also, we found that looking ahead activity often happens after posting online. These preliminary results demonstrate the possibility that there are cyber-physical discussion inter-connected flows.

5. Conclusion

In this paper, we proposed a hybrid (cyber-physical) environment in which people can discuss online and also offline simultaneously. We conducted a large-scale experiment in a panel discussion session in an international conference where participants can discuss by using our online discussion support system and by physical communications as usual. We analyzed the obtained date from the following three proposed metrics: participants’ cyber-physical attention, keywords cyber-physical linkage and cyber-physical discussion flow.

We found that our methodology succeeded to keep the participants’ attention active and continuous during this discussion session by measuring the participants’ cyber-physical attention. Also by measuring keywords cyber-physical linkage, we found that the keywords in online and real-world are correlated and somehow linked. But discussion contents were somehow related but not completely the same. Namely, we can say that the discussion contents in virtual world were different from that in the real world. By measuring cyber-physical discussion flow, we found that the number of views increases 5 min after the number of looking ahead decreases. A possible explanation would be that when the participants are interested in the real-world discussion, they look ahead. Then, they tended to look into the virtual world discussion (the number of looking ahead decreases). And, then,
after 5 min, the number of views (online) increases. We found that looking-ahead activity often happens after posting online as well. These preliminary results demonstrate the possibility that there is cyber-physical discussion inter-connected flows. These are the preliminary results, and we need to do more investigations as future work.

References


About the authors
Dr Takayuki Ito is a Professor of Nagoya Institute of Technology. He received the Doctor of Engineering from the Nagoya Institute of Technology in 2000. He was a JSPS Research fellow, an Associate Professor of JAIST and a Visiting Scholar at USC/ISI, Harvard University, and MIT twice. He was a board member of IFAAMAS, the PC-chair of AAMAS2013, PRIMA2009, General-Chair of PRIMA2014, IEEE ICA2016, is the Local Arrangements Chair of IJCAI2020 and was a SPC/PC member in many top-level conferences (IJCAI, AAMAS, ECAI, AAAI, etc.). He received the JSJ Achievement Award, the JSJ Prize, the Fundamental Research Award of JSSST, the Prize for Science and Technology of the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology (MEXT), the Young Scientists’ Prize of the Commendation for Science and Technology by the MEXT, the Nagao Special Research Award of IPSJ, the Best Paper Award of AAMAS2006, the 2005 Best Paper Award of JSSST and the Super Creator Award of 2004 IPA Exploratory Software Creation Project. He was a JST PREST Researcher and a Principal Investigator of the Japan Cabinet Funding Program for Next Generation World-Leading Researchers. He is currently a Principal Investigator of JST CREST project. Takayuki Ito is the corresponding author and can be contacted at: ito.takayuki@nitech.ac.jp

Dr Takanobu Otsuka, PhD, is a Research Assistant Professor, Nagoya Institute of Technology, Japan. He received Doctor of Engineering from Nagoya Institute of Technology in 2016. Since April 2012. His main research interests include wireless sensor networks, IoT system, knowledge-based
software engineering, embedded programming and IoT-driven intelligent systems. He is a member of ACM (Association for Computing Machinery), IEEE-CS (IEEE Computer Society), JSAI (Japanese Society for Artificial Intelligence), IPSJ (Information Processing Society of Japan) and IEICE (Institute of Electronics, Information and Communication Engineers).

Dr Satoshi Kawase, PhD, is a Researcher at Nagoya Institute of Technology. He received Doctor of Human Science from Osaka University in 2007. His research interests include human interaction, nonverbal communication, music psychology, social psychology and developmental psychology. He is a Member of the Japanese Psychological Association, the Japanese Cognitive Science Society and European Society for the Cognitive Sciences of Music.

Akihisa Sengoku received the BE degrees in Computer Science from Nagoya Institute of Technology, Japan, in 2016. He is now a master course student of Nagoya Institute of Technology.

Dr Shun Shiramatsu, PhD, is an Associate Professor of Nagoya Institute of Technology, Japan. He received Doctor of Informatics from Kyoto University in 2008. From 2007 to 2008, he was a Research Fellow of the Japan Society for the Promotion of Science (JSPS). From 2009 to 2014, he was an Assistant Professor of Nagoya Institute of Technology. His research interests include civic tech, natural language processing and linked open data. He won the 2nd Prize at Dataset Track of the LOD Challenge Japan 2013. He is also one of the “Captains Emeritus” of Code for Nagoya, which is a local civic tech organization in Nagoya city.

Takanori Ito, PhD, is an Associate Professor at Graduate School of Engineering, Nagoya Institute of Technology. He received Doctor of The Graduate School of Design from Nagoya City University in 2008. His main research interests include environmental design, architectural design, urban planning and social branding. He is a member of AIJ (Architectural Institute of Japan), JSSD (Japanese Society for the Science of Design) and some academic associations. He is an author of books including Smart Direction, Kajima Publishing, 2013. He also works for national and local governments as a member or the chief of committee on urban renewal, city promotion and other administrative plans.

Eizo Hideshima, PhD, is Full Professor at Graduate School of Engineering, Nagoya Institute of Technology. He received Master and Doctor of Engineering from Kyoto University in 1992 and 1996, respectively. His main research interests include urban planning, policy science and social infrastructure management. He is a member of RSAI (Regional Science Association International), JSCE (Japan Society for Civil Engineers) and some academic associations. He was a Research Associate at Kyoto University during 1992-1996, the Chief Advisor of Japan International Cooperation Agency Brazil Urban Transportation Human Resource Development Project during 1999-2000 and a Visiting Researcher at Stanford University during 2004-2005. He is an author of books including Transportation, Traffic Safety and Health, Springer Verlag, 2000. He also works for national and local governments as a member or the chief of committee on urban renewal, disaster mitigation and other administrative plans.

Dr Tokuro Matsuo, PhD, is a Full Professor at Advanced Institute of Industrial Technology since 2012. He received the Doctor’s degree from the Department of Computer Science at Nagoya Institute of Technology in 2006. Currently, he is a Visiting Professor at University of Nevada at Las Vegas, USA; a Guest Professor at Bina Nusantara University, Indonesia; a Project Professor, Nagoya Institute of Technology, Japan; and a Research Fellow of SEITI in Central Michigan University, USA. He was a Visiting Researcher at University of California at Irvine, USA in 2010-2011; a Research Fellow at Shanghai University, China between 2010 to 2013; and a Research Project Professor of Green Computing Research Center at Nagoya Institute of Technology, Japan between 2011 to 2014. His current research interests include electronic commerce and business, service science and marketing, business management, artificial intelligence, material informatics, tourism informatics, convention and event management research and incentive design on e-services.

Dr Tetsuya Oishi, PhD, is a Research Associate Professor at Advanced Institute of Industrial Technology. He received his PhD from Graduate School of Information Science and Electrical Engineering, Kyushu University. He worked as a system engineer at NTT DATA KYUSHU Corporation from 2004; a Post-Doctoral Fellow and a Technical Support Staff in Faculty of Information Science and Electrical Engineering at Kyushu University from 2009; an Assistant Professor in the office of institutional research at Kyushu University from 2013; and the present post
from 2016. His academic interests include data mining, information retrieval, institutional research and several related areas.

Ms Rieko Fujita is a Research Assistant Professor at Advanced Institute of Industrial Technology, Japan from 2016. She received MA degree of political science from Meiji University at 2013, studied social psychology and cultural anthropology. Her areas of research include analysis of negotiation and persuasion, lobbyist psychology, computer-based social dynamics analysis and analysis between organizational motto and its effects.

Naoki Fukuta, PhD, is an Associate Professor of informatics at Shizuoka University. He received BE and ME from Nagoya Institute of Technology in 1997 and 1999, respectively. He received Doctor of Engineering from Nagoya Institute of Technology in 2002. His main research interests include multi-agent systems, mobile agents, e-commerce, applications on auction mechanisms, semantic Web, knowledge-based software engineering, logic programming and WWW-based intelligent agent systems. He is a member of ACM (Association for Computing Machinery), IEEE-CS (IEEE Computer Society), JSAI (Japanese Society for Artificial Intelligence), IPSJ (Information Processing Society of Japan), IEICE (Institute of Electronics, Information and Communication Engineers), JSSST (Japan Society of Software Science and Technology) and ISSJ (Information Systems Society of Japan).

Dr Katsuhide Fujita, PhD, is an Associate Professor of Faculty of Engineering, Tokyo University of Agriculture and Technology. He received the BE, ME and Doctor of Engineering from the Nagoya Institute of Technology in 2008, 2010 and 2011, respectively. From 2010 to 2011, he was a Research Fellow of the Japan Society for the Promotion of Science (JSPS). From 2010 to 2011, he was a Visiting Researcher at MIT Sloan School of Management. From 2011 to 2012, he was a Project Researcher of School of Engineering, the University of Tokyo. He is an Associate Professor of Faculty of Engineering, Tokyo University of Agriculture and Technology since 2012. His main research interests include multi-issue negotiation, multi-agent systems, intelligent agents and decision support systems.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Crowd wisdom drives intelligent manufacturing
Jiaqi Lu, Shijun Liu, Lizhen Cui, Li Pan and Lei Wu
Shandong University, Jinan, China

Abstract
Purpose – A fundamental problem for intelligent manufacturing is to equip the agents with the ability to automatically make judgments and decisions. This paper aims to introduce the basic principle for intelligent crowds in an attempt to show that crowd wisdom could help in making accurate judgments and proper decisions. This further shows the positive effects that crowd wisdom could bring to the entire manufacturing process.

Design/methodology/approach – Efforts to support the critical role of crowd wisdom in intelligent manufacturing involve theoretical explanation, including a discussion of several prevailing concepts, such as consumer-to-business (C2B), crowdfunding and an interpretation of the contemporary Big Data mania. In addition, an empirical study with three business cases was conducted to prove the conclusion that our ideas could well explain the current business phenomena and guide the future of manufacturing.

Findings – This paper shows that crowd wisdom could help make accurate judgments and proper decisions. It further shows the positive effects that crowd wisdom could bring to the entire manufacturing process.

Originality/value – The paper highlights the importance of crowd wisdom in manufacturing with sufficient theoretical and empirical analysis, potentially providing a guideline for future industry.

Keywords Big Data, Crowd science, Crowd wisdom, Intelligent manufacturing

Paper type Conceptual paper

1. Introduction
Intelligent manufacturing is the development and implementation of intelligent machine in manufacturing. It refers to the use of production process technology that can automatically adapt to changing environments and varying process requirements, with the capability of manufacturing various products with minimal supervision and assistance from operators (Yang and Ding, 1992; Qu and Liu, 2009).

To achieve the goal of intelligent decision-making in the process of manufacture, an urgent need is to find the source of intelligence. Thanks to the internet era, there never has been a time like this when enterprises can access such a massive crowd, constituted by
individuals with various backgrounds, coming from different countries and districts, and thus, it is reasonable to come up with the idea to exploit the crowd of diverse strangers, which is a desired intelligent crowd. In fact, consumer-to-business (C2B) mode is calling for industry to make good use of crowd wisdom.

In this paper, we propose to use the concept of crowd wisdom for working as the brain of manufacturing. Many examples (Surowiecki, 2004) prove that a qualified crowd would make wise decisions and accurate judgments with the abundant information collected by the consisting individuals. Thus, we apply the theory onto manufacturing, pointing out that with various data contributed by the crowd, including customers especially, manufacturing decision could be obtained, and we try to explicate the reason why Big Data is key to intelligent manufacturing in the perspective of crowd wisdom.

The rest of the paper is organized as follows. Section 2 explicates the base of crowd wisdom, including the qualification for an intelligent crowd and the source of its wisdom. Section 3 introduces the crowd participation in detail and how it contributes to intelligent manufacturing. Section 4 strengthens our points with several cases where enterprises utilize crowd wisdom, which is of great help. Section 5 concludes the paper.

2. The principle of crowd wisdom
Predecessors’ works revealed an amazing fact: the wisdom of crowds is incredibly surprising. In 1906, a British scientist, Francis Galton invited 800 participants to evaluate the weight of an ox. Some of the participants are experts on agriculture, while the others are sheer laymen. The average evaluation of the crowd is only one pound less than the real weight (Surowiecki, 2004).

Individuals make their own judgments on the same topic independently. When the judgments aggregate, the crowd’s judgment would be incredibly close to the fact. In fact, the aggregate judgment is often accurate even in the case of predicting the future. First, we analyze the key principles of crowd wisdom.

2.1 Characteristics of intelligent crowd
After aggregating individuals’ judgment, we will obtain a good judgment. The approach is seemingly too naïve to be real.

But unfortunately, not all crowds possess such sharp judgment. J. Surowiecki elaborated the necessary characteristics of a crowd with great wisdom in his work, that is, under decentralization system, a diverse crowd constituted of independent individuals would make good judgment (Surowiecki, 2004). Figure 1 shows the characteristics and how they are related to each other.

The base of crowd wisdom is the private information within an individual’s judgment. The information from each individual partly depicts the fact more or less. When a large amount of information from the individuals is aggregated, those pieces of information could counteract as well as complement each other. Thus, mistakes cancel out and the crowd possesses the entire information about the truth. This guarantees the authenticity of the crowd’s judgment and the decision derived.

The wisdom comes from the crowd, as the crowd owns abundant information. Ultimately, the wisdom comes from the information. And the reason why we request the diversity of crowd is that we have to ensure that the information from the crowd is truly comprehensive rather than duplicates of the same information in diverse expressions.
2.2 Diversity of the crowd

Generally speaking, making a crowd judgment is a two-stage process. The first stage is searching for a judgment pool, which is the collection of each individual’s own judgment using his or her private information. And the second stage is obtaining a crowd judgment, by means of aggregating the judgment of every individual in the crowd with a particular technique.

It is obvious that the more extensive the judgment pool, the more preferable it is, as a good judgment is more likely to be included, enter the second-stage and impact the final judgment rather than being ignored regretfully.

Hence, now our problem is how to extend our judgment pool?

Because judgment originates from individuals’ analysis and thinking, we need people with unlike thought patterns, experts in different domains, owning diverse experience and background to be involved. Individual differences ensure the diversity of information and judgment.

2.3 Independent individuals

Independence guarantees that a wrong judgment from an individual would cause no impact on other individuals in the crowd. Mistaken judgments can counteract in the process of aggregating; therefore, the accumulated information is still correct.

Besides, independence encourages individuals to explore new private information, rather than be sedentarily satisfied with the public information. Apparently, new information is advantageous for extending our judgment pool.

Here we arrive at a paradox. If the crowd’s judgment is intelligent, why do we not just simply adopt the crowd’s judgment instead of contemplating our own judgment, which is likely to be a wrong one?

The answer may sound contradictory: the mistakes of individuals make the crowd smart, thus leading to a good crowd judgment. But when individuals begin adopting the crowd’s judgment, the crowd becomes sluggish, while the individual is gaining wisdom.

Making a judgment is never a short-term process, neither is the evaluation of a judgment. A good process should be like this: individuals make judgments independently, then the crowd wisdom works and a good crowd judgment is obtained. After a period when a large number of individuals adopt the crowd judgment, defects of the judgment would gradually emerge or someone would figure out a new judgment. Then, the massive followers would stop adopting the existing judgment and a new round of independent exploration ensues.
Perhaps as time goes by, the defects would be exposed to the public, so the process proceeds. Nevertheless, generally, a promoter, the one who figures out a new judgment when most of his or her peers adopt the crowd’s judgment, plays a role. Those insistent independent thinkers propose a new judgment, break the original stagnant state of the crowd and weaken the authenticity of the prevailing crowd’s judgment. More individuals are encouraged to cease following and pay more attention on private information and personal judgment. The promoters help us from drowning in the crowd wisdom.

2.4 Decentralization system
The concept of crowd is always related to system design. After all, the art of management is designing an appropriate system, one that vitalizes the crowd.

In our context, we expect a system that can motivate the individuals to search for diverse judgments, maintain individual independence and ensure that the final crowd’s judgment is an aggregated one, instead of coming from a dominant individual.

Many precedents prove the superiority of a decentralized system in the aspect of ensuring the three very critical points above. A decentralized system is one that is absent of the center of power.

In a system with distinct powerful members, for example, in a centralized system, various influences from dominant individuals can be a problem that cannot be neglected. Individuals may judge no longer based on facts, but reluctantly, indifferently or blindly follow the decision of superiors, experts or elites.

The undesirable situation defies the demand for independence. When dominant individuals spread their influence, they publicize their private information, which subtly influences the choice of other individuals. Even worse, a wrong judgment from the dominant individual would lead the whole crowd into an abyss.

Moreover, a blind crowd can neither be tagged as diverse nor be provided with a diverse judgment pool, as individuals can hardly be expected to give unique or distinct judgments. It is predictable that the judgment pool would be partial.

A centralized system stifles the diversity of the crowd. The crowd may come to an agreement promptly; meanwhile, they may enter the dilemma of group think (Zhang et al., 2010), where everyone tries to obey the majority and avoids proposing new angles or objection. Such a crowd may extremely believe in the correctness of crowd judgment and strongly object to any criticism or advice from the outsiders, as they deem that the final judgment is a real crowd wisdom while actually it is not. A decentralized system helps us to avert such a plight. At least, during the first stage, a decentralized system ensures that the judgment from an individual is independently made, thus guaranteeing the diversity of the judgment pool. Therefore, the final judgment will not be greatly influenced by some individuals in the crowd.

3. How crowd wisdom participates in manufacturing?
Figure 2 gives an overview of how crowd wisdom participates in the manufacturing process. The manufacturing process could be generally divided into requirement eliciting, design, trial, production, marketing and upgrade. Each phase is confronted with a main judgment and decision, which is specific to this phase and critical to the entire manufacturing procedure. A system accomplishes the decision-making tasks with the help of a large amount of data that was contributed by the crowds. Thus, crowd participation is vital for accurately judging and making smart decisions, and the influence exists throughout the whole process.
3.1 Manufacturing from customers, for customers

The concept of intelligent manufacturing is tremendously extensive, among which product innovation is an important proposition. Normally, user requirements point out the direction for enterprises. Nevertheless, it is not easy for enterprises to fully understand users’ real demands. Therefore, it is common that contrary to its over-rhetoric advertisement, a new product is of no novelty at all, but full of trivial so-called new functionality.

To solve the dilemma, a C2B mode was proposed. The core of C2B is that customers bring value to the company, involving consumers co-creating ideas, product or service concepts and solutions with a company through the social media (Hart, 1991). Customers put forward their dream product and then businesses produce accordingly. In the generation of intelligent manufacturing, the traditional business-to-customer (B2C) mode would naturally transform into the bottom-to-top C2B mode, where users’ desires substantially dominate manufacturing. There is no need for businesses to devise new products painstakingly with its own innovation team and then persuade the customers to buy their stuff, as they only produce what customers desire for. Obviously, it is a much more efficient mode, and it promotes personalization and customization, which are also key words in the theory of intelligent manufacturing.

A distinction between B2C and C2B is the involvement of customers’ opinion. It is reasonable to ascribe the advantage of C2B to its positive employment of the power of customer group, in other words, the wisdom of crowd.

3.2 The intelligent crowd

The explanation above may create an illusion: intelligent design is a really simple task for enterprises, as all they would do is to ask customers, and then adopt customer’s choice. It seems that innovation departments can be disbanded forever.

There is a possibility that the design process of some products is that simple. For example, many brands had encouraged their customers to design and vote, and they
produced the most popular design. But in most cases, the process is more complex, and the intelligent crowd involves more individuals than customer group solely.

Innovation departments are needed. Most of the enterprises cling to their own characteristics, such as the front design of a BMW series. Internal staff’s judgment would bring in such consideration and thus the final judgment would be a balance between customer needs and the enterprise style.

Another group of important participants includes investors, or the management of enterprises. Judgment from investors would take business factors into consideration, reflecting a new angle in the final crowd judgment.

A relative concept is crowdfunding (Jussila et al., 2016), that is, a presented idea appeals for various investors’ funding on internet, and only with adequate funding can the project proceed. This is similar to the concept of “voting with feet” in the discipline of corporate finance, where you express your judgment by buying or selling stocks. Similarly, investors provide their personal judgment in a financial way.

Besides, potential contributors may include various seemingly irrelevant individuals not belonging to the groups above. Consider an example of car design: mechanics and washers can provide useful information about car design. For example, the existing design with superfluous details causes inconvenience for maintenance.

3.3 Data reveal wisdom, but crowd creates data

Now we have to answer a question: When aggregating the judgment from individuals, what are we actually aggregating? Are we only aggregating the judgments?

We mentioned that “the base of crowd wisdom is the private information within individuals’ judgments”. What we actually aggregate is the information conveyed in the form of judgment. However, when individuals are unwilling to express their judgments or inconvenience in using their private information, can we skip the step of an individual making judgment but aggregate information directly? Can we still obtain crowd wisdom in this way? Can we paraphrase “the wisdom comes from the crowd, as the crowd owns abundant information” as “the wisdom comes from abundant information”?

Sure, we can. To some extent, we can simply ignore individuals’ judgments. Individuals’ judgment ability cannot explain the superior performance of the crowd’s judgment. The good performance is the result of the quality of input information, precisely, whether the information is abundant and diverse enough to depict the fact wholly.

And, thus, we can explain the importance and attention the public puts on Big Data, which is a hit in the past years. The industry widely highlights the necessity of Big Data in the on-going revolution to intelligent manufacturing (Krumeich et al., 2014). A more common expression is “Big Data drives Intelligent Manufacturing” rather than what we propose here. But we must understand the reason why Big Data could provide us with smart judgment and decision. In fact, Big Data is the abundant information generated by individuals from the crowd, including the netizens (internet citizens) crowd. And thus, as discussed earlier, the truth is depicted and the reliability of an aggregated judgment based on the data is well guaranteed. Therefore, we can conclude that the wisdom of Big Data is the wisdom of crowd.

It seems somewhat inconsistent with what we described earlier. An individual’s judgment is no longer necessary, but we still rely on the individuals to explore abundant and diverse information for the crowd. And, thus, the final crowd judgment still reveals crowd wisdom.

3.4 Data collection and crowd collaboration

Information is far more comprehensive than user requirements solely. And, thus, what we ask for from the individuals in the crowd extends beyond their own judgment. Particularly,
customers are expected to contribute personal habits of consumption, evaluation of the purchased products and so on. Various, ordinary, even seemingly unrelated information may tell an incredible business secret when they are aggregated. And that is why many enterprises are eager for user information, even when the information is of no relevance to its business.

The process can be described as collaboration. Information from a single individual is partial, but when information is obtained from all the individuals working as a whole, it is complete enough for an accurate description of the truth, assisting manufacturing decision further than providing inspiration and design. For example, with abundant feedback, decisions such increasing or reducing production can be made automatically, and defects’ feedback could be collected and used to guide better next-generation products. Therefore, the production process would improve and management technique would get upgraded.

In a word, the more individuals in a crowd contribute, the more powerful of the crowd wisdom, and thus, the better intelligent manufacturing would function.

3.5 Aggregation: the emergence of wisdom

So far, we blur the boundary between judgment and decision. The explanation for this vagueness is that a smart decision is usually derived from an accurate judgment to the target and thus they are the same in essence. Also, we have not explicitly explained how a crowd makes a judgment.

In fact, we would not and could not explore how the crowd decision is made. This is a question of how to deal with the pool. We guarantee you that wisdom does exist in the pool. But as for how to extract the wisdom, the techniques vary.

In free markets, price is the method, while in diverse evaluation experiments, which J. Surowiecki mentioned in his book, statistical methods such as counting votes are recommended. A point worth noting is that the method should ensure that all the information would be made good use of and unified to work as a component of the whole (Surowiecki, 2004).

For some problems, the technique is naïve and explicit, just as what we mentioned above, while for others, it involves more advanced technologies, which becomes the subject of another discipline, that is, artificial intelligence (Meziane et al., 2000), an art of extracting information. Artificial intelligence, along with all of us, would be indispensable in the trends of intelligent manufacturing.

4. Crowd wisdom drives manufacturing: several case studies

In the following paragraphs, we will study several cases, where enterprises successfully use crowd wisdom.

4.1 Crowd wisdom inspires design

In February 2016, Haier launched a water heater with purification functionality. The sales of the new heater surpassed 10,000 in only 21 days (Xu, 2016).

Not long ago, Haier received complaints from over 300,000 netizens about unclean bathwater. After consulting with the netizens, Haier was inspired to develop a new series of heaters. The new function of these heaters includes full purification capacity, not only eliminating the bacteria but also removing the residual chlorine in tap water and sediment (Su, 2015).

Complaining customers must be far more than 300,000 and the contents must be diverse. Entering Haier’s official website, you will notice that some complain about the price, while some are dissatisfied with the inconstant temperature. But the 300,000 must be the majority
and thus in this way statistics work. A simple aggregate of individual judgment leads to a
crowd judgment: purification functionality is a good idea.

Crowd wisdom efficiently facilitated Haier on the issue of novelty, thus researchers and
developers can focus on these techniques.

4.2 Crowd wisdom streamlines production

In Zara stores, there are cameras hidden in every corner and the staff carry personal digital
assistants. The purpose is to collect customers’ every single piece of advice, like preference
toward patterns or the size of buttons. Besides, Zara’s online stores contribute more data.
Zara has opened six online stores in Europe. Every day, the collected information would be
aggregated to Zara’s internal global network to facilitate the designer. Once the design is
determined, it would be immediately sent to the production line.

The data help a lot. Making production decision based on collected data greatly reduces
inventory, and thus reduces costs. Also, an analysis of the data reveals domestic prevailing
color and other information which help discover target markets and design.

The Inditex Group, the parent company of Zara, prides itself to deliver quality merchandise
in as little as three weeks from its own factories. Designers develop new models daily,
sometimes three or four a day, which are then reviewed and put into production (Chen, 2009).

Valuable customer information is not only applied in the manufacturing but also in
all departments throughout the Inditex Group, including customer service centers,
marketing departments, design teams and production lines, etc. These massive data greatly
help on the application of the vertical integration principle inside Zara and also contribute to
the measurement of KPI (Key Performance Indicators) for all departments.

Further, the optimization based on positive utility of massive data brings substantial
profits. In fiscal 2014 (ended January 31, 2015), the company had sales of $19.7bn compared
to Uniqlo $16.6bn, Gap $16.4bn, Mango $2.1bn. Inditex sales increased 8%, far stronger than
its competitors (Loeb, 2016).

4.3 Crowd wisdom creates business opportunity

A challenge for enterprises is that not all the customers are willing to provide their
information. Obtaining a piece of advice may be easy, but for more private information, such
as birth, age, daily routine, many people would regard it as an invasion of privacy and
strongly oppose it.

But because enterprises are desperate for the information revealed from customers’ data,
they have to resort to tricks and strategies. The strong motivation drives enterprises to
create new business modes, which enable them to establish intimate relationships with their
customers.

Nike transformed its sports shoes into wearable devices, and connected user data to
Nike + fitness social network via shoes or the wristband sensor. The shoes would
automatically record users’ training data while the iPod stores and displays the data,
including motion date, time, distance, calories burned, etc. Users upload data to the Nike
community, share and discuss with friends. Members in the community can compare,
analyze and set targets for their own training (Wu and He, 2014).

Nike accesses users’ data by means of encouraging them to upload their training records.
With an increasing number of people using its products, Nike learns about the best jogging
routine in major cities, and even gets a list of songs for those who love to jog.

These data, if Nike plans to organize a marathon, to advertise its new products, will
prove to be of great help. Besides, the abundant data play an important role in the effort to
understand user habits, for improving products, etc.
5. Conclusion
Wisdom is derived from crowds, or, to be precise, from the information contributed by the crowd. Information, in the forms of various data, may be seemingly unrelated or useless, but actually it tells a lot when aggregated, which we proved with several real cases. Crowd wisdom assists, or even partly substitutes, humans to make judgments, which could be well applied to manufacturing, where plenty of judgments and decisions are made. In the blueprint of future industries, crowd wisdom would surely drive the development of intelligent manufacturing.

References

Corresponding author
Shijun Liu can be contacted at: lsj@sdu.edu.cn

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Application of keyword extraction on MOOC resources

Zhuoxuan Jiang  
*Peking University, Beijing, China*

Chunyan Miao  
*Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, Nanyang Technological University, Singapore, Singapore and School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore, and*

Xiaoming Li  
*Peking University, Beijing, China*

Abstract

**Purpose** – Recent years have witnessed the rapid development of massive open online courses (MOOCs). With more and more courses being produced by instructors and being participated by learners all over the world, unprecedented massive educational resources are aggregated. The educational resources include videos, subtitles, lecture notes, quizzes, etc., on the teaching side, and forum contents, Wiki, log of learning behavior, log of homework, etc., on the learning side. However, the data are both unstructured and diverse. To facilitate knowledge management and mining on MOOCs, extracting keywords from the resources is important. This paper aims to adapt the state-of-the-art techniques to MOOC settings and evaluate the effectiveness on real data. In terms of practice, this paper also tries to answer the questions for the first time that to what extend can the MOOC resources support keyword extraction models, and how many human efforts are required to make the models work well.

**Design/methodology/approach** – Based on which side generates the data, i.e. instructors or learners, the data are classified to teaching resources and learning resources, respectively. The approach used on teaching resources is based on machine learning models with labels, while the approach used on learning resources is based on graph model without labels.

**Findings** – From the teaching resources, the methods used by the authors can accurately extract keywords with only 10 per cent labeled data. The authors find a characteristic of the data that the resources of various forms, e.g. subtitles and PPTs, should be separately considered because they have the different model ability. From the learning resources, the keywords extracted from MOOC forums are not as domain-specific as those extracted from teaching resources, but they can reflect the topics which are lively discussed in forums. Then instructors can get feedback from the indication. The authors implement two applications with the extracted keywords: generating concept map and generating learning path. The visual demos show they have the potential to improve learning efficiency when they are integrated into a real MOOC platform.

**Research limitations/implications** – Conducting keyword extraction on MOOC resources is quite difficult because teaching resources are hard to be obtained due to copyrights. Also, getting labeled data is tough because usually expertise of the corresponding domain is required.
**Practical implications** – The experiment results support that MOOC resources are good enough for building models of keyword extraction, and an acceptable balance between human efforts and model accuracy can be achieved.

**Originality/value** – This paper presents a pioneer study on keyword extraction on MOOC resources and obtains some new findings.

**Keywords**  Concept map, Graph model, Keyword extraction, Learning path, Massive Open Online Courses (MOOCs)

**Paper type**  Research paper

---

1. **Introduction**

In recent years, massive open online courses (MOOCs) have benefited tens of millions of students all over the world. A very important characteristic of MOOC is it provides a one-stop online learning environment which consists of lecture videos, assignments, email notifications, discussion forum, and quizzes and examinations. Along with the popularity of MOOCs, a large amount of online educational resources of various subject areas, ranging from humanity to science, are unprecedentedly produced. Not only instructors can provide videos, subtitles, lecture notes, questions, etc., but also learners can generate forum contents, Wiki posts, log of homework submissions, etc. In fact, each MOOCs platform is a large-scale “knowledge base” where the educational resources can be regarded as the outcome of crowd intelligence (from both instructors and learners). However, those resources are unstructured and diverse. For example, subtitles are well-organized and formal, as they are usually produced by instructors, whereas the contents of posts are written by different learners; thus, they are colloquial and informal. As Figure 1 shows, MOOC resources can be considered as teaching resources and learning resources. By proposing proper models to discover knowledge from the crowd intelligence, it is promising to implement knowledge management, knowledge mining and even smart education for MOOCs.

This paper explores the task of keyword extraction and its applications on MOOC resources. The reason for conducting this task is that in most work of knowledge engineering, e.g. construction of knowledge graph and knowledge management, entity extraction is the first step. As for our task, we call the “entity” as keyword. The meaning of keyword in the educational setting is regarded general and intuitively can include concept, terminology, named entity and so on. Keyword extraction from MOOC resources may face several difficulties:

![Figure 1. Overview of MOOC resources](image-url)
MOOCs are of different subject areas, any domain-specific method should not help much, and as such, the method we use should be instructor- and course-agnostic; obtaining labeled training data set is extremely expensive, as usually domain expertise is required; and the volume of data is usually large and the textual styles are various.

Despite those difficulties, once keywords are well extracted, many subsequent applications are feasible, e.g. construction of course-specific and domain-specific concept map, management of cross-domain concepts, knowledge discovery from crowd and even personalized learning by mining learners’ behaviors.

Based on the partition of who generates the MOOC resources, i.e. instructors and learners, the research design of this paper is composed by three parts:

1. keyword extraction on resources generated by instructors;
2. keyword extraction on resources generated by learners; and
3. applications with keywords in MOOC settings.

As to the first part, it is difficult to collect entire instructor-generated resources of many courses. Also, labeling the data requires expertise in the corresponding subject area. Even so, we invite the instructors and teaching assistants (TAs) to help label the teaching resources of one course, as we expect to use human knowledge to learn a classifier by supervised machine learning methods. Moreover, we design a semi-supervised learning framework to test whether using less labeled data is practical. We regard this task as a problem of natural language processing, i.e. word sequence labeling. Sutton and McCallum (2011) believe that the probabilistic graphical models, especially conditional random fields (CRFs), can obtain the state-of-the-art performance in many sequence labeling tasks like part of speech (POS), named entity recognition (NER) and word segment, so we leverage this kind of model to extract keywords on MOOC teaching resources.

As to the second part of keyword extraction on resources generated by learners, i.e. discussion forum contents, it is relatively convenient to collect the contents of many courses. However, the number of posts may be quite large, e.g. over ten thousands, so it is difficult to use human knowledge through labeled data. On the other hand, as a kind of social media, forums have relational information between learners and contents. By referring to many methods of keyword extraction for social media, we model the MOOC forum to a heterogeneous network for each course. Then through graph-based random walk algorithm, keywords are extracted by ranking the importance of each word. We regard the top words in the ranking list as keywords.

After keywords are extracted, lots of novel educational applications can be developed within the MOOC settings. In the third part of this paper, we introduce two preliminary applications: generation of concept map and generation of learning paths. Romero and Ventura (2010) proposes that in the educational field, concept map is useful to organize, design and manage the course resources for instructors. We propose a new concept map which is called semantic concept map (SCM). The main difference with traditional concept maps is that the edge, i.e. relationship between keywords, is defined as semantic similarity. This kind of concept map can be easily extended to various courses. Then based on the SCM, we propose a method to automatically generate learning paths which have the potential for personalized learning.
In what follows, we review the related work in Section 2. Section 3 introduces data sets used in this paper. Section 4 introduces the method of keyword extraction on the side of teaching resources. The corresponding method on the learning side, i.e. forum contents, is introduced in Section 5. Then in Section 6, we report the experiment results obtained from both sides of resources respectively. In Section 7, we state the two demo applications with extracted keywords. Finally, we conclude this paper in Section 8.

### 2. Related work

Keyword is the word which people regard as important in a text. In different situations, keyword can be named entity, proper noun, terminology or concept. In this paper, the meaning of keyword is general. So if not otherwise specified, their differences are neglected.

In the past decades, Finkel et al. (2005), Nadeau and Sekine (2007) and Ratinov and Roth (2009) have studied the tasks of keyword extraction by machine learning methods, e.g. NER, terminology extraction and key phrases extraction. NER methods focus on named nouns, such as person name, location, time and address, and they are for constructing knowledge base, as seen from the papers by Dong et al. (2014) and Nickel et al. (2015). Terminology extraction methods are developed to extract domain-specific words. Recently, Nojiri and Manning (2015) and Qin et al. (2013) propose the methods based on machine learning for keyword extraction. However, methods for one kind of keywords extraction may not be used to another kind. For example, Nojiri and Manning (2015) exhibit that directly applying existing methods of NER to terminology extraction will not perform well. It is different to our task that we have labeled data.

Apart from supervised machine learning methods with human knowledge, another perspective for solving keyword extraction is the unsupervised approach. For example, Justesona and Katza (1995) propose the rule-based method, Frantzi et al. (2000) and Bin and Shichao (2011) propose the statistical methods. In this paper, we leverage a graph-based method proposed by Sonawane and Kulkarni (2014), which can model the social relationship between words to a network and then rank all the words in accordance with their importance.

To our knowledge, a large number of studies of data analytics on MOOC data have been proposed in recent years. For example, Anderson et al. (2014) try to classify MOOC learners after analyzing their behavior patterns. It also studies how to use a badge system to produce incentives based on learners’ activity and contribution in the forum. Huang et al. (2014) analyze the behaviors of superposters in 44 MOOCs forums and finds that MOOCs forums are mostly healthy. Wen et al. (2014) study the sentiment analysis in MOOCs discussion forums and find that no positive correlation exists between the sentiment of posts and the course dropout. Wang et al. (2015) study the learning gain reflected through forum discussions. Jiang et al. (2015) conduct an analysis from the perspective of influence by modeling the MOOC forum to a heterogeneous network. Kizilcec et al. (2013) conduct a research on the behavior of learner disengagement. Moreover, some statistical reports and case study papers analyze behavior of MOOC learners, such as Ho et al. (2013) and Breslow et al. (2013). However, few studies of keyword extraction have been conducted on MOOC data.

Romero and Ventura (2010) and Novak and Cañas (2006) define that a concept map is a connected graph that shows relationships between concepts and expresses the hierarchical structure of knowledge. To our knowledge, plenty of work of automatically constructed concept map has been studied with data mining techniques. For example, Tseng et al. (2007), Lee et al. (2009) and Qasim et al. (2013) leverage association-rule mining; Chen et al. (2008), Lau et al. (2009) and Huang et al. (2006, 2015) base on text mining; and Marian and...
Maria (2009) and Chu et al. (2007) design specific algorithms. However, the majority of those methods are domain-specific, e.g. for specific courses or specific learning settings. We expect to explore new methods by reducing their dependency on domains, so a new kind of semantic relationship is leveraged in this paper.

3. Overview of a data sets
Also based on the partition of sources of generated resources, i.e. instructors and learners, we introduce the available MOOC data we have respectively.

3.1 Resources on teaching side
We collect the resources of an interdisciplinary course conducted in the fall of 2013 on Coursera. The course involves computer science, social science and economics. Textual content includes video subtitles, PPTs, questions and forum contents (i.e. threads, posts and comments). Table I shows the statistics of resources. We invited the instructor and two TAs to help label the data. As seen in Table I, the number of keywords in questions and PPTs are much smaller than that in subtitles. Based on our observation during labeling the data, the instructor and TAs would still spend much time on understanding each sentence, even though they should be more familiar with the contents than any others. We guess it is because the resources are composed by different people. During the activity of labeling data, everyone would spend about 8 h on labeling 3,000 sentences (in average 10 s per sentence).

A preprocessing step of word segment for Chinese may be necessary. We adopt the Stanford Word Segmenter[1] proposed by Chang et al. (2008). All data are randomly shuffled before they are processed late.

3.2 Resources on learning side
We collect data from 12 courses offered by Peking University from Coursera. They were offered in Fall Semester of 2013 and Spring Semester of 2014. There are totally over 4,000 threads and over 24,000 posts. For convenience later in the paper, Table II lists the pairs of course codes and course titles. Table III shows the statistics of the data sets per course. The “posts” denotes both posts and comments.

4. Keyword extraction on teaching side
The resources generated by instructors in MOOCs mainly include lecture notes, subtitles, PPTs and questions. In order to extract keywords from the teaching resources, we regard this task as a sequence labeling problem. It is similar to other sequence labeling tasks, e.g. NER and part-of-speech annotation. So probabilistic graphical models are the natural solution to this kind of tasks. Sutton and McCallum (2011) exhibits CRFs can achieve the state-of-the art performance. And we define instructor- and course-agnostic features in order to reduce the domain dependency. Moreover, we propose a semi-supervised learning framework to reduce human efforts of labeling data.

<table>
<thead>
<tr>
<th>Source</th>
<th># Sentence</th>
<th># Word</th>
<th># Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtitles</td>
<td>3,036</td>
<td>69,437</td>
<td>402</td>
</tr>
<tr>
<td>PPTs</td>
<td>2,823</td>
<td>22,334</td>
<td>249</td>
</tr>
<tr>
<td>Questions</td>
<td>268</td>
<td>7,138</td>
<td>95</td>
</tr>
</tbody>
</table>
4.1 Conditional random field’s model

The problem of keyword extraction can be formally described as solving the conditional probability $P(Y|X)$. The random variable $X$ refers to features of each sentence which follows a word sequence $x = \{x_1, x_2, \ldots, x_T\}$, and the random variable $Y$ is a label sequence of the sentence $y = \{y_1, y_2, \ldots, y_T\}$. The label of a word is defined as three classes: $\text{NO}$, $\text{ST}$ and $\text{IN}$. They respectively mean not a keyword, the beginning word of a keyword and the middle word of a keyword. So the label variable is $Y \in \{\text{NO}, \text{ST}, \text{IN}\}$.

We consider the conditional probability of labeling sequence $Y$, i.e. $p(Y|X)$, rather than their joint probability $p(Y, X)$, so linear chain CRFs framework proposed by Lafferty et al. (2001) is the natural choice. The conditional distribution over label sequence $y$, given an observation word sequence $x$, can be defined as:

$$P(Y|X) = \frac{1}{Z} \exp \left( \sum_{t=1}^{T} \sum_{i=1}^{3} \theta_i y_{t,i} \phi_t(x) + \theta_0 y_{t,0} \right)$$

where $Z$ is the normalizing constant, $\theta_i$ are the parameters for each class, and $\phi_t(x)$ is the feature function of the sentence.

### Table II.
Pairs of course code and course title

<table>
<thead>
<tr>
<th>Course code</th>
<th>Course title</th>
</tr>
</thead>
<tbody>
<tr>
<td>peopleandnetworks-001</td>
<td>People and networks</td>
</tr>
<tr>
<td>arthistory-001</td>
<td>Art history</td>
</tr>
<tr>
<td>dsalgo-001</td>
<td>Data structures and algorithms A</td>
</tr>
<tr>
<td>pkuc-001</td>
<td>Introduction to computing</td>
</tr>
<tr>
<td>ao-001</td>
<td>Advanced object-oriented technology</td>
</tr>
<tr>
<td>bdsalgo-001</td>
<td>Data structures and algorithms B</td>
</tr>
<tr>
<td>criminallaw-001</td>
<td>Criminal law</td>
</tr>
<tr>
<td>pkupop-001</td>
<td>Practice on programming</td>
</tr>
<tr>
<td>chemistry-001</td>
<td>General chemistry (Session 1)</td>
</tr>
<tr>
<td>chemistry-002</td>
<td>General chemistry (Session 2)</td>
</tr>
<tr>
<td>pkubioinfo-001</td>
<td>Bioinformatics: Introduction and methods (Session 1)</td>
</tr>
<tr>
<td>pkubioinfo-002</td>
<td>Bioinformatics: Introduction and methods (Session 2)</td>
</tr>
</tbody>
</table>

### Table III.
Statistics of resources generated by learners

| Course code               | # registrants | # threads | # posts | Maximum # posts per thread | Average # posts per thread | Average # votes | # overall certificates (ratio) | $P(C|F)$ | $P(F|C)$ |
|---------------------------|---------------|-----------|---------|-----------------------------|---------------------------|-----------------|---------------------------------|---------|---------|
| peopleandnetworks-001     | 10,807        | 219       | 1,206   | 38                          | 5.5                       | 304             | 149 (0.013)                     | 0.271   | 0.584   |
| arthistory-001            | 16,395        | 273       | 2,181   | 124                         | 8.0                       | 1,541           | 237 (0.014)                     | 0.272   | 0.620   |
| dsalgo-001                | 13,197        | 283       | 1,221   | 36                          | 4.3                       | 266             | 57 (0.004)                      | 0.180   | 0.930   |
| pkuc-001                  | 14,462        | 1,029     | 5,942   | 141                         | 5.8                       | 595             | 285 (0.020)                     | 0.290   | 0.782   |
| ao-001                    | 9,563         | 97        | 515     | 63                          | 5.3                       | 204             | 53 (0.006)                      | 0.160   | 0.528   |
| bdsalgo-001               | 852           | 319       | 1,299   | 48                          | 4.1                       | 132             | 248 (0.320)                     | 0.853   | 0.774   |
| criminallaw-001           | 8,190         | 118       | 763     | 58                          | 6.5                       | 648             | –                              | –       | –       |
| orgchem-001               | 4,374         | 28        | 85      | 12                          | 3.0                       | 11              | 12 (0.004)                      | 0.091   | 0.250   |
| pkupop-001                | 18,410        | 1,085     | 6,443   | 92                          | 5.9                       | 977             | 205 (0.012)                     | 0.255   | 0.780   |
| chemistry-001             | 9,124         | 110       | 591     | 40                          | 5.4                       | 65              | 116 (0.013)                     | 0.400   | 0.448   |
| chemistry-002             | 6,782         | 167       | 715     | 92                          | 4.3                       | 678             | 125 (0.020)                     | 0.336   | 0.336   |
| pkubioinfo-001            | 18,367        | 361       | 2,139   | 201                         | 5.9                       | 1,474           | 581 (0.032)                     | 0.362   | 0.370   |
| pkubioinfo-002            | 16,714        | 170       | 942     | 51                          | 5.5                       | 235             | 510 (0.032)                     | 0.571   | 0.212   |
| Overall                   | 147,237       | 4,259     | 24,042  | –                           | –                         | –               | –                               | 0.309   | 0.508   |

**Notes:** $P(C|F)$ represents the ratio of certificated forum learners to overall forum learners; $P(F|C)$ represents the ratio of certificated forum learners to overall certificated learners of the course.
\[ p(y|x) = \frac{1}{Z(x)} \exp \left( \sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_k f_k(y_{t-1}, y_t, \bar{x}_t) \right) \]  

(1)

where \( Z(x) = \sum_y \exp \left( \sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_k f_k(y_{t-1}, y_t, \bar{x}_t) \right) \) and \( \mathcal{F} = \{f_k(y_{t-1}, y_t, \bar{x}_t)\}_{k=1}^{K} \) are the set of feature functions defined on given \( x \); \( \Theta = \{\lambda_k\} \in \mathbb{R}^K \) are parameter vector. \( N \) is the length of sentence and \( K \) is the number of features.

Given a training data set, the model \( \Theta = \{\lambda_k\}_{k=1}^{K} \) could be learned by maximum likelihood estimation. To avoid overfitting, we add a regularized term to the function. Then the log-likelihood function of \( p(y|x, \lambda) \) based on the Euclidean norm of \( \lambda \sim (0, \sigma^2) \) is represented as:

\[ L(\Theta) = \sum_{x,y} \log p(y|x, \Theta) - \sum_{k=1}^{K} \frac{\lambda_k^2}{2\sigma^2} \]  

(2)

So the gradient function is:

\[ \frac{\partial L}{\partial \lambda_k} = \sum_{x,y} \sum_{t=1}^{T} f_k(y_{t-1}, y_t, x_t) - \sum_{x,y} \sum_{t=1}^{T} f_k(y, y', x_t) p(y, y'|x) - \frac{\lambda_k}{\sigma^2} \]  

(3)

The detail of learning the CRFs model can be referred to Sutton and McCallum (2011). Then, given a new word sequence \( x^* \) and a learned model \( \Theta = \{\lambda_k\}_{k=1}^{K} \), the optimal label sequence \( y^* \) could be calculated by:

\[ y^* = \arg \max_{y \in \mathcal{Y}} p(y|x^*, \Theta) \]  

(4)

where \( \mathcal{Y} \) is the set of all possible label sequences for the given sentence \( x^* \). We use L-BFGS algorithm to learn the model and Viterbi algorithm to infer the optimal label sequence \( y^* \).

### 4.2 Feature engineering

A crucial part of CRFs framework is the definition of feature functions. Based on our observation, we define five kinds of features which are adapted to our educational data. All the features are course-agnostic and make our framework flexible for scalability.

#### 4.2.1 Text style features

- whether the target word is English;
- whether the two neighbor words are English;
- whether the word is the first word in a sentence;
- whether the word is the last word in a sentence; and
- whether the target word is in a quotation.

Text style features capture the stylistic characteristics. Some keywords usually appear at the beginning or the last of a sentence in instructor’s language, e.g. “Network means[...]'” or “[..]This is the definition of Network”. Because our data are from a Chinese MOOC, we
regard whether the word is English as a feature. Obviously, when it comes to English MOOCs, capitalization is the key feature of English keywords. So this kind of features are flexible to different situations.

4.2.2 Structure features
- POS tag of the target word;
- POS tag of the previous word; and
- POS tag of the next word.

We treat the POS as a feature because fixed combination of POS, e.g. adjective + noun or noun + noun, may indicate keyword phrases. We use the Stanford Log-linear POS Tagger [2] proposed by Toutanova et al. (2003) to assigns POS to each word. Note that as to the corresponding feature functions, we adopt binary value, 0 or 1, to every POS. For example, there is a function to capture whether the target word is a noun and so on.

4.2.3 Context features
- term frequency and inverted document frequency (TF-IDF) value of the target word and two neighbor words;
- normalized uni-gram BM25 score of the target word;
- normalized bi-gram BM25 score of the target word; and
- normalized bi-gram BM25 score of the two neighbor words.

Context features capture the importance of words and word-level information within the whole document. The training set is partitioned to documents based on video clips. Statistical metric of normalized bi-grams BM25 scores proposed by Robertson et al. (2004) is used to quantify word relevance by default parameters.

4.2.4 Semantic features
- semantic similarity of the target word with the previous two words; and
- semantic similarity of the target word with the next two words.

Some frequent-co-occurrence words may be keywords. Also, close words in the semantic space may be keywords. So by learning the word semantics, features of adjacent words can be captured. The similarity of two adjacent words in semantic space is calculated with the corresponding word vectors trained by Word2Vec[3] proposed by Mikolov et al. (2013). All textual contents are used to learn the word embeddings. The corpus size is 145,232 words and the vector dimension is set as 100 by default.

4.2.5 Dictionary features
- whether the target word and two neighbor words are in the dictionary; and
- whether the two neighbor words are in the dictionary.

As in most tasks about natural language processing, a dictionary is useful. We therefore design a run-time dictionary which is just a set of keywords in training data set.

4.3 Semi-supervised learning framework
Because the effort for labeling training data is extremely expensive, we propose the semi-supervised framework. We leverage the ideas of self training proposed by Liu et al. (2009) and k nearest neighbors (KNN). The intuition is that if an unlabeled sample is similar to a labeled sample in semantic space, the unlabeled sample is very probable to be successfully
inferred by the model which is learned from all the current labeled data. Then, the unlabeled sample is turned to a labeled one and can be added into the labeled dataset with model-inferred labels. A new model can be learned. The new thing proposed here is that we use the word embeddings learned by Word2Vec to calculate the similarity between two sentences. Sentence vector is denoted as:

\[
\text{VecSentence}_i = \frac{1}{T} \sum_{t=1}^{T} \text{VecWord}_t
\]

(5)

where \( \text{VecWord} \) is the word vector. Algorithm 1 is the details of the semi-supervised version of training process.

The time complexity of Algorithm 1 is \( O(NM^2) + M O(\text{TrainCRF}) \) where \( N \) and \( M \) are the sizes of labeled set and unlabeled set, respectively, and \( c \) is the number of unlabeled data which are selected to be inferred in each loop. The additional computing cost is rewarding, as human effort can be largely reduced, especially when \( N \) and \( M \) is not large.

Algorithm 1 Semi-supervised learning based on KNN self-training

**INPUT**: labeled data set \( X_L = \{(x,y)\} \), unlabeled data set \( X_U = \{x\} \), number of candidates \( c \)

**OUTPUT**: model \( \Theta \)

1: \textbf{repeat}
2: \hspace{1em} \( \Theta = \text{TrainCRF}(X_L) \)
3: \hspace{1em} \( X_{c-nearest} = \emptyset \)
4: \hspace{1em} \textbf{for} \( i = 1 : c \)
5: \hspace{2em} \( x = \arg\min_{x \in X_U} \text{Cosine\_distance}(x,X_L) \)
6: \hspace{2em} \( X_U = X_U - \{x\} \)
7: \hspace{2em} \( X_{c-nearest} = X_{c-nearest} \cup \{x\} \)
8: \hspace{1em} \( Y_{c-nearest} = \text{InferCRF}(X_{c-nearest}, \Theta) \)
9: \hspace{1em} \( X_L = X_L \cup \{(X_{c-nearest}, Y_{c-nearest})\} \)
10: \textbf{until} \( X_U = \emptyset \)
11: \( \Theta = \text{TrainCRF}(X_L) \)
12: \textbf{return} \( \Theta \)

5. **Keyword extraction on learning side**

Due to the difficulty and complexity of labeling massive data of MOOC forums, we leverage unsupervised approaches to extract keywords from contents generated by learners. As the discussion forums are a kind of social media, we build a graph to model the relationship of post-reply. Then, a random walk algorithm is proposed to rank the importance of words. Finally, we regard the top words as keywords.

The intuition to build a graph model is that the more words are replied to, the more important the word is, and the more important word A is when word A replies to word B, the more important word B would be. This is similar to the algorithms for ranking Web pages, e.g. PageRank proposed by Brin and Page (1998).

5.1 **Data model of massive open online course forum**

To better model the importance of keywords, we design a heterogeneous network. Two kinds of entities are involved, learners and words. In the following, we introduce definition of the data model. Then, we explain the intuition for designing such a network.
Definition 1. Heterogeneous network with learners and words. Given all the learners’ records of a MOOC forum, heterogeneous network \( G = (V, E, W) = (V_L \cup V_D, E_L \cup E_D \cup E_{LD}, W_L \cup W_D \cup W_{LD}) \) where \( V_L = \{v_L^1, v_L^2, \ldots, v_L^n\} \) and \( V_D = \{v_D^1, v_D^2, \ldots, v_D^m\} \) are sets of learners and words, respectively. \( E_L \) is the set of directed edges which mean the co-occurrence of two learners in the same thread. The learner who posts later points to the other. \( E_D \) is the set of directed and bidirectional edges which mean the co-occurrence of two words in the same thread. Directed edges mean the two words belong to different posts. And the one which appears later points to the other. The bidirectional edges mean the two words belong to the same post. \( E_{LD} \) is the set of bidirectional edges which mean a learner’s contents contain the word and in reverse a word appears in the learner’s contents. \( W_L, W_D \) and \( W_{LD} \) are the sets of weight values which mean the times of co-occurrence of two entities on corresponding edges. Self co-occurrence is meaningless and is consistently ignored.

Figure 2 is a demo of the heterogeneous network with learners and words of a MOOC forum. By the way, we denote \( G_L = (V_L, E_L, W_L) \) as a weighted directed graph of learners, \( G_D = (V_D, E_D, W_D) \) as a weighted directed and bidirectional graph of words and \( G_{LD} = (V_{LD}, E_{LD}, W_{LD}) \) as the weighted bipartite graph of authorship between students and keywords. Denote \( n_L = |V_L| \) and \( n_D = |V_D| \) are the numbers of entities in \( V_L \) and \( V_D \), respectively.

Such a heterogeneous network can embody the latent post-reply relationship between learners and words. In \( G_L \) and \( G_D \), the more edges point to an entity, the more important it is. Moreover, if more important entities point to a specific entity, the entity would be more important. Similarly seeing from \( G_{LD} \), the more edges point to a word, the more popular it is while also more important, if an important learner points to it. All the weight values can capture the importance degree of relationship. The transmission of importance between learners (in \( G_L \)) can be transited to \( G_D \). It is a process of mutual reinforcement between the two subnetworks.

5.2 Jump random walk algorithm

We design an algorithm for co-ranking learners and words, named Jump-Random-Walk (JRW) which simulates two random surfers jumping and walking between different types of entities. Figure 3 shows the framework of JRW algorithm. \( G_L \) is the subnetwork of learners and \( G_D \) is the subnetwork of words. \( G_{LD} \) is the subnetwork of authorship. \( \beta \) is the probability of walking along an edge within the homogeneous subnetwork. \( \lambda \) is the probability for jumping to the other subnetwork. \( \lambda = 0 \) means the two random surfers are independent to jump and walk within respective homogeneous subnetworks. We assume the probabilities of jump and walk are consistent.

Denote \( l \in \mathbb{R}^{n_L} \) and \( d \in \mathbb{R}^{n_D} \) are the ranking result vectors, also probability distributions, whose entries are corresponding to entities of \( V_L \) and \( V_D \), subject to \( \|l\|_1 \leq 1 \) and \( \|d\|_1 \leq 1 \) respectively.
and $||d||_1 \leq 1$ due to existence of no-out-degree entities. Denote four transition matrixes of $G_L$, $G_D$, $G_{LD}$ and $G_{DL}$ as $L \in \mathbb{R}^{n_L \times n_L}$, $D \in \mathbb{R}^{n_D \times n_D}$, $LD \in \mathbb{R}^{n_L \times n_D}$ and $DL \in \mathbb{R}^{n_D \times n_L}$, respectively. Adding the probability of random jumping for avoiding trapped in small set of entities or no-out-degree entities, the iteration functions are:

$$l = (1 - \lambda)(\beta L\tilde{l} + (1 - \beta)e_{n_L}/n_L) + \lambda LD\tilde{d}$$  \hspace{1cm} (6)$$

$$d = (1 - \lambda)(\beta D\tilde{d} + (1 - \beta)e_{n_D}/n_D) + \lambda DL\tilde{l}$$  \hspace{1cm} (7)$$

where the former terms right the equal signs are iteration functions within a homogeneous subnetwork and the latter are across the two homogeneous subnetworks. $\lambda$ is the probability of jumping to the subnetwork. $\beta$ is the probability of walking along an edge within a homogeneous subnetwork. $e_{n_L} \in \mathbb{R}^{n_L}$ and $e_{n_D} \in \mathbb{R}^{n_D}$ are the vectors whose all entries are 1. The four transition matrixes are:

$$L_{i,j} = \frac{w^L_{i,j}}{\sum_i w^L_{i,j}} \text{ where } \sum_i w^L_{i,j} \neq 0,$$  \hspace{1cm} (8)$$

$$D_{i,j} = \frac{w^D_{i,j}}{\sum_i w^D_{i,j}} \text{ where } \sum_i w^D_{i,j} \neq 0,$$  \hspace{1cm} (9)$$

$$LD_{i,j} = \sum_i w^{LD}_{i,j},$$  \hspace{1cm} (10)$$

$$DL_{i,j} = \sum_i w^{DL}_{i,j} \text{ where } \sum_i w^{DL}_{i,j} \neq 0.$$  \hspace{1cm} (11)$$

$w^L_{i,j}$ is the weight of the edge from $V^L_i$ to $V^L_j$, $w^D_{i,j}$ is the weight of the edge between $V^D_i$ and $V^D_j$, $w^{LD}_{i,j}$ is the weight of the edge between $V^L_i$ and $V^D_j$ and $w^{DL}_{i,j}$ is the weight of the edge between $V^D_i$ and $V^L_j$. Actually, $w^{LD}_{i,j} = w^{DL}_{i,j}$. When $\sum_i w^L_{i,j} = 0$, it means the student $V^L_j$ always posts the last in a thread. If $\sum_i w^D_{i,j} = 0$, it means the keyword $V^D_j$ always has no peer in a thread. Actually, this situation almost never happens in our filtered words. $\sum_i w^{LD}_{i,j} = 0$ is also impossible, which means every word would have at least one author.

Figure 3.
The overview of JRW algorithm for co-ranking learners and words in the heterogeneous network.
(learner). On the contrary, it does not make sure that every student would post at least one keyword because maybe someone’s post has nothing valuable and contains no keyword. All the transition matrixes are non-negative. Algorithm 1 is the detail of JRW algorithm in the heterogeneous network.

Algorithm 2 JRW on G

```plaintext
INPUT \( L, D, LD, DL, \beta, \lambda, \epsilon \)
1: \( l \leftarrow e/\eta_L \)
2: \( d \leftarrow e/\eta_D \)
3: \textbf{repeat}
4: \( l \leftarrow l \)
5: \( d \leftarrow d \)
6: \( l = (1 - \lambda)(\beta Ll + (1 - \beta)e_{nL}/\eta_L) + \lambda L\Delta d \)
7: \( d = (1 - \lambda)(\beta Dd + (1 - \beta)e_{nD}/\eta_D) + \lambda DLl \)
8: \textbf{until} \(|d - d| \leq \epsilon\)
9: \textbf{return} \( l, d \)
```

Finally, we can actually get two ranking lists of learners and words, but we only consider the ranking list of words within this paper.

6. Experiment

Again, based on the partition of two kinds of resources, as well as an extra experiment, this section consists of three parts.

6.1 On teaching side

In this subsection, we use teaching resources, i.e. subtitles, PPTs and questions, to evaluate the supervised learning model. We introduce several baselines to extract keywords for comparison:

- **Term frequency (TF):** Words are ranked by their term frequency. If a word is a keyword, the instructor may say it repeatedly in lecture.
- **Bootstraping (BT):** Instructors may have personal language styles to give talks. So we design the rule-based algorithm by giving several patterns containing keywords. This method is actually course- and instructor-dependent.
- **Stanford Chinese NER (S-NER):** This is an exiting tool developed for NER, whose model is already trained, and we just use it to infer keywords in our educational data sets[4] proposed by Nadeau and Sekine (2007).
- **Terminology extraction (TermExtractor):** This is an exiting tool for terminology extraction[5]. The well-trained model is also only used to infer keywords in our data sets.
- **Supervised keyword-CRF (SK-CRF):** This is a method of supervised learning based CRFs with all features as defined before.
- **Semi-supervised keyword-CRF (SSK-CRF):** This is the semi-supervised version for keyword extraction. The parameter of \( c \), number of candidates, is empirically set as 20.

We adopt three metrics, precision, recall and F1-value, to measure the results.

6.1.1 Results and analysis. Table IV shows the comparison of performance between baselines. We use 30 per cent data of subtitles as training data for SK-CRF and SSK-CRF,
and the rest are for evaluation. Especially for SSK-CRF, half of the training data are unlabeled. The statistic-based methods (TF@500 and TF@1000) are unreliable because many stopwords may degrade the performance. The rule-base method (BT) is highly dependent on human experience, and the low precision means plenty of subsequent work for filtering the outputs is required. On the other hand, Stanford Chinese NER and TermExtractor do not perform well maybe because of two reasons, namely, named entity and terminology are actually different from the keywords in our data, and the models are not learned from our data set. The semi-supervised CRF is comparable to the supervised version.

Figure 4 manifests that the semi-supervised learning would be comparable to the supervised version, especially when less than 20 per cent data are used for training. Half of training data is identically regarded as unlabeled by SSK-CRF. Note that the amount of labeled data when using 10 per cent training data by SK-CRF is equivalent to that of using 20 per cent training data by SSK-CRF, but SSK-CRF performs better than SK-CRF. This result means the semi-supervised framework can obtain satisfactory performance by only labeling a handful of data.

Now, we evaluate the different model abilities among various MOOC textual content. As shown in Table V, the items in row are training data set, while those in column are testing data set. This table can explain some common situations of educational settings. Subtitles can cover almost all the keywords. They are ideal to be regarded as the training data. PPTs is also decent to be as training data seeing from the precisions, but the recalls are low. Maybe due to usually in PDF format, PPTs may cause incomplete sentences when being converted to text. Questions could lead to lower recalls than PPTs because not all keywords are present in questions as shown in Table I. In summary, different kinds of MOOC textual content have different model ability, so they should be separately considered.

6.1.2 Feature contribution. We analyze how the different kinds of features contribute to the model. The result is shown in Table VI. Dictionary feature has a predominant influence on the final results, and structure feature is the second important. Other features are also contributive, but the difference is small. Even so, every kind of features contribute to the model positively.

6.2 On learning side
After building a heterogenous network for each course, Table VII shows the parameters of the network per course.

The important of keywords ranked at top is hard to evaluate. Table VIII lists the top ten high-frequency words and top ten keywords ranked by JRW, respectively. We can see the

<table>
<thead>
<tr>
<th>Baselines</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF@500</td>
<td>0.402</td>
<td>0.500</td>
<td>0.446</td>
</tr>
<tr>
<td>TF@1000</td>
<td>0.600</td>
<td>0.746</td>
<td>0.665</td>
</tr>
<tr>
<td>BT</td>
<td>0.099</td>
<td>0.627</td>
<td>0.171</td>
</tr>
<tr>
<td>S-NER</td>
<td>0.131</td>
<td>0.080</td>
<td>0.099</td>
</tr>
<tr>
<td>TermExtractor</td>
<td>0.202</td>
<td>0.107</td>
<td>0.140</td>
</tr>
<tr>
<td>SK-CRF</td>
<td>0.914</td>
<td>0.897</td>
<td>0.905</td>
</tr>
<tr>
<td>SSK-CRF</td>
<td>0.889</td>
<td>0.825</td>
<td>0.856</td>
</tr>
</tbody>
</table>

Table IV. Performance of baselines

Notes: SK-CRF and SSK-CRF use 30% data of subtitles for training; half of the training data as unlabeled for SSK-CRF; the italic data mean they are the best results among all the baselines.
two kinds are highly overlapped, but the order is slightly different. The bold keywords are related to course content, and the italic ones are mainly about the course quiz, assignment, video and other course stuff.

Table IX shows the statistics of the top three “important posts”, meaning that the posts contain the top 20 keywords. The more frequency of keywords they contain, the higher they rank. From Table IX, we can first find that the content lengths are mostly long, which is obvious by our definition of “important posts”. From the dimension of vote, we cannot find some insight of the numbers. Author rank means the ranking of the post author in the ranking list of important learners. We find they are truly the “important learners” of each
course. Also, the important posts are mostly at the top of a thread, seeing from position in thread. It means the initial authors in a thread are inclined to express important information. By the way, the lengths of a thread, i.e. # Post in Thread, are significantly correlated to the important posts. Some empirical conclusions can be summarized as below:

6.3 Extra experiment

Considering the available data in our hand, although we do not have labels of forum data, we can learn a classifier from labeled teaching resources and conduct a task of identifying the need of concept comprehension on forum contents. This task can be regarded as a binary classification of forum threads, that is to identify whether a thread is about concept comprehension. So if the question contains keywords of the course, it is much likely to ask for the explanation of some concepts. The result is post-evaluated which means: to each thread, if the score is marked as “1”, two situations are included as the following:

- if no concept is identified and this thread is not about need of concept comprehension; and
- if at least one concept is identified and the definition of identified concepts can answer the question.

Other situations are marked as “0”.

We use 30 per cent of subtitles to learn a classifier by the semi-supervised method. Only threads title and the initial post are involved in this experiment, instead of all the posts. Table X exhibits the result. The accuracy is not bad. The relatively high recall is meaningful because this can accurately remind instructors which threads to intervene. Moreover, this method not only can identify whether a thread is about concept comprehension but also can identify which concept needs to be explained.

7. Application with keywords for massive open online course

After keywords are extracted from teaching resources of one course, we exhibit two intelligent applications with keywords in the MOOC settings: generation of concept map and generation of learning path. We conduct the applications on the course of people and network.

| Course             | $n_L$ | $|E_L|/n_L^2$ | $n_D$ | $|E_D|/n_D^2$ | $|E_LD|/n_{LD}^2$ | $|E_{LD}|/(n_L + n_D)^2$ |
|-------------------|-------|--------------|-------|--------------|------------------|-------------------------|
| peopleandnetworks-001 | 321   | 3.287        | 0.032 | 1.193        | 104,821          | 0.074                   | 4,814                   | 0.002                   |
| arthistory-001    | 540   | 17,022       | 0.058 | 3,376        | 1,019,289        | 0.089                   | 14,195                  | 0.001                   |
| dsalgo-001        | 295   | 1,876        | 0.022 | 1,152        | 124,118          | 0.094                   | 5,099                   | 0.002                   |
| pkuic-001         | 768   | 19,801       | 0.034 | 2,302        | 302,989          | 0.057                   | 14,599                  | 0.002                   |
| aoo-001           | 175   | 1,963        | 0.064 | 783          | 73,208           | 0.119                   | 2,597                   | 0.003                   |
| bdsalgo-001       | 225   | 2,369        | 0.047 | 781          | 23,540           | 0.039                   | 3,133                   | 0.003                   |
| criminallaw-001   | 219   | 2,971        | 0.062 | 1,224        | 123,737          | 0.083                   | 4,577                   | 0.002                   |
| pkupop-001        | 628   | 12,883       | 0.033 | 1,748        | 88,035           | 0.029                   | 13,807                  | 0.002                   |
| chemistry-001     | 130   | 886          | 0.052 | 1,055        | 111,026          | 0.100                   | 2,685                   | 0.002                   |
| chemistry-002     | 125   | 2,341        | 0.150 | 964          | 61,425           | 0.066                   | 2,574                   | 0.002                   |
| pkubioinfo-001    | 594   | 22,275       | 0.063 | 686          | 46,768           | 0.099                   | 1,946                   | 0.001                   |
| pkubioinfo-002    | 189   | 1746         | 0.049 | 380          | 16662            | 0.115                   | 784                     | 0.002                   |

Table VII.
Summary of the constructed heterogeneous network per course
<table>
<thead>
<tr>
<th>Course</th>
<th>Top 10 high-frequency words</th>
<th>Top 10 JRW ranked words</th>
</tr>
</thead>
<tbody>
<tr>
<td>peopleand networks-001</td>
<td>relationship, people, node, homework, question, teacher, course, problem, video, answer</td>
<td>people, relationship, node, homework, teacher, question, problem, video, course, network</td>
</tr>
<tr>
<td>arthistory-001</td>
<td>art, art history, people, question, teacher, course, class, classmate, homework, artistic work</td>
<td>art, art history, teacher, people, course, question, class, classmate, artistic work</td>
</tr>
<tr>
<td>dsalgo-001</td>
<td>teacher, course, question, class, data, problem, homework, code, video, algorithm</td>
<td>teacher, course, class, question, data, people, code, homework, structure, problem</td>
</tr>
<tr>
<td>pkuic-001</td>
<td>question, program, teacher, homework, code, course, array, problem, mistake, result</td>
<td>question, program, teacher, course, computer, homework, code, people, array, video</td>
</tr>
<tr>
<td>aoo-001</td>
<td>object, model, method, software, question, code, system, teacher, graph, video</td>
<td>object, model, method, software, question, code, system, video, graph, teacher</td>
</tr>
<tr>
<td>bdsalgo-001</td>
<td>question, data, problem, code, pointer, occasion, teacher, function, table, algorithm</td>
<td>question, data, problem, table, teacher, occasion, code, algorithm, function, array</td>
</tr>
<tr>
<td>criminallaw-001</td>
<td>penal law, law, behavior, people, judicature, guilt, question, classmate, country, teacher</td>
<td>law, penal law, people, behavior, judicature, guilt, classmate, question, society, Part B</td>
</tr>
<tr>
<td>pkupop-001</td>
<td>function, question, problem, homework, program, code, teacher, object, result, array</td>
<td>function, question, code, program, object, homework, problem, teacher, array, time</td>
</tr>
<tr>
<td>chemistry-001</td>
<td>chemistry, teacher, course, class, question, problem, video, door, people, answer</td>
<td>chemistry, teacher, course, class, question, problem, video, door, people, answer</td>
</tr>
<tr>
<td>chemistry-002</td>
<td>chemistry, electron, teacher, course, chemicalbond, question, atom, orbit, class, radius</td>
<td>chemistry, teacher, electron, course, university, class, student, question, atom, radius</td>
</tr>
<tr>
<td>pkubioinfo-001</td>
<td>sequence, teacher, biology, course, question, class, informatics, video, information, data</td>
<td>biology, teacher, sequence, informatics, question, course, class, data, video, information</td>
</tr>
<tr>
<td>pkubioinfo-002</td>
<td>course, question, teacher, biology, sequence, video, door, classmate, certificate, genome</td>
<td>course, question, sequence, teacher, classmate, biology, door, content, species, video</td>
</tr>
</tbody>
</table>

**Note:** The bold words are course content-related and the italic words are course resource-related.

### 7.1 Concept map

For generating a concept map, we need to define the meaning of nodes, edges and their weights. The nodes are concepts, and the edges are defined as the semantic similarity which is general for every course. We call the new concept map as SCM. Based on our observation, there are two kinds of node weight definitions, i.e. term frequency (TF) and TF-IDF.

It can be observed that the more frequent a concept appears, the more fundamental it is. For example, the top ten concepts, *Node, Network, Reward, Probability, Graph, Game, Edge, Tactic, Hypothesis and Price*, are all the fundamental knowledge points of the course. So the metric of TF can capture the feature of *fundamentality*. The formal definition is:

\[
\text{NodeWeight}_{i}^{(F)} = \sum_{k} f_{ki} \quad (12)
\]
where $f_{ki}$ is the times of the $i$th concept existing in the $k$th document; a document corresponds to a video clip in MOOCs.

However, on the other hand, low-frequency concepts often are the important knowledge points. So TF-IDF is ideal to measure the importance of a concept. For example, the top ten important concepts are PageRank, SignalSequence, Table IX.

<table>
<thead>
<tr>
<th>Course</th>
<th>Rank</th>
<th>Content length</th>
<th>Vote</th>
<th>Author rank</th>
<th>Position in thread</th>
<th># post in thread</th>
</tr>
</thead>
<tbody>
<tr>
<td>peopleandnetworks-001</td>
<td>1</td>
<td>1,515</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>716</td>
<td>1</td>
<td>10</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>664</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>arthistory-001</td>
<td>1</td>
<td>5,146</td>
<td>2</td>
<td>14</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4,670</td>
<td>3</td>
<td>11</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2,380</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>dsalgo-001</td>
<td>1</td>
<td>757</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>757</td>
<td>0</td>
<td>6</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>619</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>pluiac-001</td>
<td>1</td>
<td>1,790</td>
<td>0</td>
<td>39</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1,418</td>
<td>0</td>
<td>23</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>619</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>aoo-001</td>
<td>1</td>
<td>1,991</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1,468</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1,394</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>bdsalgo-001</td>
<td>1</td>
<td>3,501</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>855</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1,257</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>criminallaw-001</td>
<td>1</td>
<td>1,683</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1,967</td>
<td>2</td>
<td>2</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1,480</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>pkupop-001</td>
<td>1</td>
<td>368</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>457</td>
<td>0</td>
<td>22</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>579</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>chemistry-001</td>
<td>1</td>
<td>1,664</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>359</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1,539</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>chemistry-002</td>
<td>1</td>
<td>424</td>
<td>16</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>472</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>306</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>pkuinfo-001</td>
<td>1</td>
<td>769</td>
<td>-2</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1,322</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1,956</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>pkuinfo-002</td>
<td>1</td>
<td>309</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>422</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>515</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Table IX.
Summary of top three important posts which contain the top 20 keywords per course.

Table X.
Result of identifying threads about concept comprehension by SSK-CRF.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.822</td>
<td>0.523</td>
<td>0.784</td>
<td>0.627</td>
</tr>
</tbody>
</table>

where $f_{ki}$ is the times of the $i$th concept existing in the $k$th document; a document corresponds to a video clip in MOOCs.
**PreEstimatedPrice, MixedStrategyEquilibrium, TradingRight, HinderAggregation, Cluster, ConformityBehavior, NashBargaining and Popularity.** The formal definition of TF-IDF weights is:

\[
\text{NodeWeight}_{i}^{(l)} = \frac{1}{n_{i}} \sum_{k} \log(f_{ki} + 1) \cdot \log(N/n_{i})
\]

(13)

where \(N\) is the number of video clips and \(n_{i}\) is the times of video clips in which the \(i\)th concept appears.

Considering the word embeddings learned by Word2Vec have the characteristic that semantically similar words are close in the embedding space, so we use the similarity as the weights of edges between two concepts. For example, the most semantically similar concepts around \(\text{Network}\) are: \(\text{NetworkAnalysis}, \text{SocialNetwork}, \text{ResidualNetwork}, \text{ComplexNetwork}, \text{NetworkSwitch}, \text{ComplexNetworkAnalysis}, \text{SocialNetwork}, \text{TraficNetwork}, \text{SocialNetworkAnalysis}\) and \(\text{NetworkSwitchExperiment}\).

Figure 5 shows the demos of SCM of the course of People and Network for fundamentality and importance, respectively. We find the map can visually reveal the degree of semantic relationships between concepts. This is beneficial for learners to build a “concept map” in their brain and remember concepts easily. We use the tool of Gephi to draw the maps.

### 7.2 Learning path

Based on the SCM, learners can also learn the course in line with their own pace. Here, we propose an algorithm (Algorithm 3) to generate a primary learning path according to the definition of SCM. Then, the learning path can be revised by both the instructors and learners as required.

The basic idea of the algorithm is simple. Every time a current concept is taken, then a candidate set of \(k\) the most semantically similar neighbors of the concept are selected. Among the candidate set, TF or TF-IDF of concepts is calculated. Then, the top concept is selected as a node in the path and as the next current concept. The algorithm can start from any concept. Note that the concepts which are selected in the candidate set should appear later than the current concept along with the course because learners may be confused to learn concepts through the path which does not conform to the instructor’s design.

By taking the concept \(\text{Node}\) as the starting point and setting \(k = 10\), the first ten concepts in the learning path with metric of TF are: \(\text{Node} \rightarrow \text{Edge} \rightarrow \text{Element} \rightarrow \text{Set} \rightarrow \text{Alternative} \rightarrow \text{Vote} \rightarrow \text{MajorityVoting} \rightarrow \text{MajorityVotingRule} \rightarrow \text{IndividualRanking} \rightarrow \text{GroupRanking}\).

By taking the concept with the highest TF-IDF as the starting point, the first ten concepts are: \(\text{PageRank} \rightarrow \text{PageRankAlgorithm} \rightarrow \text{SmallWorld} \rightarrow \text{Balance} \rightarrow \text{NashBalance} \rightarrow \text{StructuralBalance} \rightarrow \text{EquilibriumTheorem} \rightarrow \text{MixedStrategyEquilibrium} \rightarrow \text{NashBargaining} \rightarrow \text{NashBargainingSolution}\). We can see these concepts are all important along the course:

**Algorithm 3** Generation of learning path

**INPUT**: SCM = \(\{C, R\}\), starting concept \(c_{i}\), number of candidates \(k\)

**OUTPUT**: learning path \(p_{i} = \{n_{1}, n_{2}, \ldots, n_{|C|}\}\)

1: \(j = 1\)
2: \(n_{j} = c_{i}\)
3: \(p_{i} = \{n_{j}\}\)
4: \(C' = C - \{n_{j}\}\)
Figure 5.
Two kinds of SCM based on different concepts metrics

Notes: (a) For fundamentality; (b) For importance
5: repeat
6: \( T = \) the \( k \) most semantically similar and later appeared concepts to \( n_j \) in \( C \)
7: \( j = j + 1 \)
8: \( n_j = \) the concept selected by some metric (TF or TF-IDF) in \( T \)
9: \( p_i = p_i \cup \{n_j\} \)
10: \( C = C - \{n_j\} \)
11: until \( C = \emptyset \)
12: return \( p_i \)

Admittedly, the two demo learning paths are very primitive. They cannot support personalized learning and adaptive learning yet. However, by analyzing the learners' behavior and log of homework, the learning paths can be more intelligent. We leave this for the future work.

8. Conclusion
Along with the development of MOOCs, massive online educational resources are unprecedentedly produced from crowd. Instructors can provide videos, subtitles, lecture notes, questions, etc., while learners can generate forum content, Wiki, log of homework, etc. How to process these data from unstructured to structured is a challenging problem. In this paper, we explore the task of keyword extraction on MOOC resources.

Keyword extraction can benefit a lot of subsequential applications. First, it is a kind of annotation for MOOC resources. The annotation can be used for studying machine learning methods for MOOC-related natural language processing tasks, such as information extraction, information retrieval and question answering. Second, keyword extraction can pick up domain-specific or cross-domain knowledge points from complex text. This result can be further processed to build knowledge graph or concept map. With the graph (or the map), instructors can better organize the course, and learners can plan their own learning paths more easily. Then by collecting the feedback from learners, the whole teaching and learning process can be a virtuous cycle. Thus finally, crowd intelligence can lead to intelligent education.

Back to the task of this paper, we are faced with two challenges: MOOCs are cross-domain, labeling training data is extremely expensive. So we propose a flexible framework based on semi-supervised machine learning with domain-agnostic features. Experiments demonstrate the efficacy of our framework. Using a very little labeled data can achieve decent performance. We find that various kinds of MOOC content, e.g. subtitles and PPTs, have different modeling ability for keyword extraction. So they should be separately treated in future work. Our framework also can be applied to the task of concept identification on MOOC forum content. Moreover, unsupervised method based on graph model is proposed by modeling MOOC forum to a heterogeneous network. Although the top keywords in MOOC forums are not as the same as those keywords extracted from teaching resources, they can indicate the concerned topics which are discussed in forums. At least instructors can get feedback from the information.

In the future, methods of transfer learning and deep learning may be better for extracting cross-domain keywords. External resources of knowledge, e.g. Wikipedia, may be helpful. The relationship between keywords is deserved to be paid more attention for building a domain-specific or even cross-domain concept map.
Notes
3. Word2Vec: https://code.google.com/p/word2vec/
5. Terminology extraction by translated labs: http://labs.translated.net/terminology-extraction/

References


**About the authors**

Zhuoxuan Jiang is currently pursuing a PhD degree at the School of Electronics Engineering and Computer Science, Peking University, China. His current research interests include machine learning, MOOC data mining and online education. Zhuoxuan Jiang is the corresponding author and can be contacted at: jzhx@pku.edu.cn

Chunyan Miao is a Professor with the School of Computer Science and Engineering (SCSE) at Nanyang Technological University (NTU), Singapore. She is the Director of the NTU-UBC Joint Research Centre of Excellence in Active Living for the Elderly (LILY). Prior to joining NTU, she was an Instructor and Post-Doctoral Fellow with the School of Computing, Simon Fraser University, Canada. Her research focuses on studying the cognitive and social characteristics of intelligent agents in multi-agent and distributed AI/CI systems, such as trust, emotions, incentives, motivated learning, ecological and organizational behavior. She has worked on new disruptive Artificial intelligence (AI) approaches and theories that synergize human intelligence, artificial intelligence and behavior data analytics (AI powered by humans). Her current research interests include human-agent interaction, cognitive agents, human computation and serious games.

Xiaoming Li is a Professor in Computer Science and Technology and the Director of Institute of Network Computing and Information Systems (NCIS) at Peking University, China. His current research interest is in Web data mining and online education. He led the effort of developing a Chinese search engine (Tianwang) since 1999 and is the founder of the Chinese Web archive (Web InfoMall). Related papers have been published in *WWW, KDD, Computer Networks, Journal of Software and Systems, Journal of Web Engineering*, etc. Under his direction, the Institute is focused on the areas of search engine and Web mining, peer-to-peer computing, distributed systems, mobile computing, high productivity computing and database systems. He serves on the editorial boards of several journals, including *Concurrency and Computation* (John Wiley), *Journal of Web Engineering* (Rinton), etc. He is a Senior Member of IEEE and a Member of Eta Kappa Nu. He also serves as a Vice President for Chinese Computer Federation and is chairing the Advisory Subcommittee for Undergraduate Computing Education in China.

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)
Motivation mechanism of gamification in crowdsourcing projects

Zhiyuan Zeng, Jian Tang and Tianmei Wang
School of Information, Central University of Finance and Economics, Beijing, China

Abstract
Purpose – The purpose of this paper is to study the participation behaviors in the context of crowdsourcing projects from the perspective of gamification.

Design/methodology/approach – This paper first proposed a model to depict the effect of four categories of game elements on three types of motivation based upon several motivation theories, which may, in turn, influence user participation. Then, 5 × 2 between-subject Web experiments were designed for collecting data and validating this model.

Findings – Game elements which provide participants with rewards and recognitions or remind participants of the completion progress of their tasks may positively influence the extrinsic motivation, whereas game elements which can help create a fantasy scene may strengthen intrinsic motivation. Besides, recognition-kind and progress-kind game elements may trigger the internalization of extrinsic motivation. In addition, when a task is of high complexity, the effects from game elements on extrinsic motivation and intrinsic motivation will be less prominent, whereas the internalization of extrinsic motivation may benefit from the increase of task complexity.

Originality/value – This study may uncover the motivation mechanism of several different kinds of game elements, which may help to find which game elements are more effective in enhancing engagement and participation in crowdsourcing projects. Besides, as task complexity is used as a moderator, one may be able to identify whether task complexity is able to influence the effects from game elements on motivations. Last, but not the least, this study will indicate the interrelationship between game elements, individual motivation and user participation, which can be adapted by other scholars.

Keywords Gamification, Crowd behaviour analysis, Crowd-sourced design and engineering, Task-oriented crowdsourcing

1. Introduction
Gamification refers to the use of gameplay mechanisms in non-gaming contexts to encourage desired behaviors (Werbach and Hunter, 2012). Many scholars and practitioners have applied this approach in several contexts such as education, health, project management, environment protection, outsourcing and software development. (Deterding, 2012). Gamification can benefit corporations and companies by extending the market share,
changing the attitude of users and enhancing the motivation effect (Hamari and Koivisto, 2013). The application of game elements helps managers or organizers dismantle the obstacles like the disability to attract users, inadequate participation and low engagement (McGonigal, 2011), and finally, guarantee that projects or activities hosted by those corporations and companies could move forward as expected.

Crowdsourcing is one of the contexts where gamification could be applied, referring to a new outsourcing approach that takes tasks $a \leq x \leq b$ and $y \leq c \leq d$, where $a < b$ and $c < d$.

Suppose that we use $m$ layers, where $m \in \mathbb{N}^+$ an open call to an undefined, large group of people (Howe, 2006). The cooperation and competition traits of crowdsourcing projects allow companies to make full use of knowledge, skills and resources beyond those of their own staff (Neyer et al., 2009; Whitla, 2009), so that they can reduce costs and increase benefits (Howe, 2006).

Crowdsourcing relies on the number of participants who are willing to devote and contribute their time and efforts. Whether crowdsourcing projects could be operated smoothly is highly relevant with the ability of those initiators to attract and obtain adequate participants. Many crowdsourcing projects have failed because they were unable to gain enough attention from the public (McGonigal, 2011). Gamification is considered as an approach to attract participants and improve the engagement. The application of gamification may avoid the situation where a crowdsourcing project may fail owing to the lack of participation. Hence, how to apply gamification appropriately to increase participation and engagement has become a major topic in our research.

Some prior research discussed the application of gamification in the realm of crowdsourcing (Zheng et al., 2011; Mekler et al., 2015; Kavaliova et al., 2016); yet, the extant literature lacks a systematic understanding of motivation mechanism of gamification in crowdsourcing projects. Some scholars who are interested in the application of gamification intended to abstract game elements to provide a more complete structure of gamification (Hunicke et al., 2004; McGonigal, 2011; Robson et al., 2015), but the process or mechanism via which these game elements can have an impact on user participation is barely mentioned in these frameworks; others intended to discuss the motivational influence of some specific game elements in different contexts (Sigala, 2015), but the categories of these game elements having been discussed in these studies are limited. As a result, which kinds of game elements are more effective in increasing participation and engagement in crowdsourcing projects, and whether motivation mechanism of these game elements vary from one kind to another remains uncovered.

To fill these gaps, we intend to investigate the motivation mechanism of gamification in crowdsourcing projects. First, we categorized game elements that have been used widely in crowdsourcing projects. Then, we proposed a research model that depicts the effect of these game elements on human motivation, which may, in turn, influence user participation. In addition, we are interested in identifying the moderating effects of task complexity on user motivation. We designed a series of laboratory experiments to validate our research model. Finally, we discussed expected contributions and implications of our research.

2. Literature review

2.1 Gamification

Game has been considered as the inspiration for designing captivating user interface since 1980 (Deterding, 2012; Malone, 1981). Until now, many scholars and practitioners have applied game dynamics and elements to different contexts to enhance user engagement. The
most frequently mentioned and used game elements within these contexts are narratives, fantasy world, point, level, status, badge, leaderboard, challenge, countdown, deadline, etc.

A common definition of gamification embraces the concept of using game elements and design techniques in a non-game setting, often with the end-goal of shaping user behaviors (Kavaliova et al., 2016). Thanks to the high adaptability of gamification, many non-gaming contexts, such as education, health, project management, environment protection, crowdsourcing, software development, have used the idea of gamification (Deterding, 2012). Many corporations and companies have used the concept of gamification and game elements to help them extend the market share, alter the attitude of users and enhance the motivation effect (Hamari and Koivisto, 2013) to maintain high user engagement and obtain user loyalty. However, most initiators cannot meet their expectations because game elements were not implemented appropriately. Hence, how to select proper game elements under varied contexts has gradually gained attention among scholars and practitioners, so is the motivation mechanism of gamification.

Despite the wide application of gamification, a discussion of the inner mechanism of gamification is still open. One of the most frequently used framework of gamification is Hunicke et al.’s (2004) MDA (mechanic–dynamic–esthetics) framework where game elements have been abstracted into three parts: game mechanics, game dynamics and game esthetics. Mechanics refer to the fundamental settings and rules including point, level, trophy, badge, achievement, leaderboard and virtual gift; dynamics refer to the motives which can lead to some particular emotions, such as reward, status, achievement, self-expression, competition and altruism (Simoes et al., 2013); and esthetic, which is substituted sometimes as emotion (Robson et al., 2015), refers to the feelings and emotions evoked by game dynamics. The MDA framework has brought out a hierarchical picture of game elements. However, the MDA framework emphasized on the abstract of the components of gamification while they left out the part on the motivation process of these components. Besides, while this framework takes gamification as a whole to analyze, the distinct ways in which those game elements may function are not discussed.

Some scholars put their emphasis on the effect of some specific game elements. Hanus and Fox (2015) applied badge and leaderboard to the context of education and discovered that these two kinds of game element may decrease the motives and final scores of students. Mekler et al. (2015) used point, level and leaderboard in a tagging experiment and found that these three game elements have no influence on individuals’ intrinsic motivation, but will change users’ ultimate behaviors by affecting extrinsic motivation. These studies show that the impacts of game elements may vary under different contexts. Hence, we should seek for a more detailed explanation. In addition, game element categories that have been mentioned in these studies are limited. The effect and mechanism of game elements such as narratives and progress bar are not well studied.

To address these issues, we discuss the motivation mechanism and process of different categories of game elements in crowdsourcing projects, and posit a theoretical model that can uncover the relationships among varied game elements, user motivations and user participation.

2.2 Crowdsourcing
Crowdsourcing is one of the contexts where the concept of gamification has been frequently used. It refers to a new outsourcing approach that takes tasks as an open call to an undefined, large group of people (Howe, 2006). Knowledge seekers can post their task requirements in some kind of a platform and then problem solvers can participate in these tasks to gain rewards provided by knowledge seekers after they manage to meet all the
requirements requested by knowledge seekers (Liu et al., 2014). Through crowdsourcing projects, managers and organizers can make better use of knowledge, skills and resources (Neyer et al., 2009; Whitla, 2009) and reduce the cost (Howe, 2006).

However, as the cost for operating a crowdsourcing project is relatively low, and there is an increasing number of platforms supporting the crowdsourcing projects, the competition among those crowdsourcing project initiators to call up participants becomes fierce. Under these circumstances, many crowdsourcing projects failed as an end owing to the initiators’ disability to absorb enough attention from the public (McGonigal, 2011). As user engagement is of great necessity for accomplishing a crowdsourcing project, the motives and incentives for those users to participate in crowdsourcing projects are becoming a topic for relevant scholars.

2.3 Motivation theory
Porter and Lawler’s (1968) intrinsic and extrinsic motivation and Ryan et al.’s (1985) self-determination theory (SDT) are the most applied motivation theories among all the motivation research. Zheng et al. (2011) developed a research model to explain participation in crowdsourcing contexts based on the theory of extrinsic and intrinsic motivation. And by analyzing the case of Threadless, Kavaliova et al. (2016) used the SDT to explore how companies can motivate contributions to a crowdsourcing project.

Intrinsic motivation involves people doing an activity because they find it interesting and derive spontaneous satisfaction from the activity itself (Gagne and Deci, 2005). In contrast, extrinsically motivated individuals perform an activity as a means to achieve some separable objective or personal benefits, so satisfaction comes not from the activity itself but rather from the extrinsic consequences to which the activity leads (Wong-On-Wing et al., 2010). In Ryan et al.’s (1985) SDT, individual motivation is further categorized as an intrinsic motivation, internalized extrinsic motivation and extrinsic motivation. Internalization is defined as people taking in values, attitudes or regulatory structures, such that the external regulation of a behavior is transformed into an internal regulation and thus no longer requires the presence of an external contingency. Besides, Zheng et al. (2011) put forward an idea in their research on the motives of a crowdsourcing project that job characteristics are of great importance in influencing participants’ behaviors, which may act as regulatory variables to enhance or weaken the effect of intrinsic motivation. This idea is also the kernel of Hackman and Oldman’s (1980) job design theory.

2.4 Motives in crowdsourcing
Under the drive of these theories, there are a number of motives on crowdsourcing projects discovered through various studies (see Table I). Among all these motives, “reward” is one of the most important one, which could be found in all motive lists of distinct crowdsourcing projects (Brabham, 2008, 2010; Zheng et al., 2011; Fuller, 2010; Kavaliova et al., 2016; Sun et al., 2015), “Recognition” appeared just as often as “reward”. However, internalized extrinsic motivation and intrinsic motivation is relatively less found in crowdsourcing projects (Fuller, 2010; Zheng et al., 2011; Kavaliova et al., 2016).

Besides, among all these motives listed above, “skills development”, “professional development” and “social reason” (community building, social contribution, etc.) are determined by the context and purpose of the project itself, meaning that these motives are difficult to be enhanced by game elements. “Relationship building” is related with the function of the crowdsourcing platform. If the platform is very powerful in supporting social communication like providing functions such as liking others, commenting, sharing to other SNS, creating a forum/group, such kinds of motives will be enhanced. However, this can also
be difficult for any game element to achieve. Hence, we exclude motives like “skills development”, “professional development”, “social reason” and “relationship building” when we discuss which motives can be enhanced by those game elements (see Table II).

### 3. Research model and hypothesis development

#### 3.1 Game element categorization

Game elements which have been frequently used in crowdsourcing are point/score, leaderboard/ranking, badge/achievement, level, reward, progress bar, mission, feedback, storytelling and virtual territories (Morschheuser et al., 2016). Based on the characteristics of these game elements and the similarities among them, we further categorized these game elements into four categories: reward, recognition, progress and fantasy. Reward refers to game elements that provide rewards for some specific behaviors or achievements of participants, including tangible money and gifts, as well as virtual goods and game properties. Recognition refers to game elements that support mechanism where actions of participants to accomplish the task are converted to some kind of countable values to be compared with, including point (/score), leaderboard (/ranking), badge (/achievement) and level. Progress refers to game elements that provide information on the progression of an ongoing task on the premise that the accomplishment of a specific project is divided into several submissions or stages. Fantasy refers to game elements that create virtual circumstances, such as storytelling or virtual territory.

The research model based on the categorization mentioned above is shown as Figure 1.


3.2 Motivation mechanism

Reward refers not only to some tangible money or gifts, but also to some virtual goods or properties. This kind of game element is unable to improve the enjoyment itself, which means it cannot enhance intrinsic motivation that participants might feel. Once this kind of reward is canceled, individuals will not continue their behaviors any longer just out of their habits (Gagne and Deci, 2005).

However, no matter it is tangible money or gifts, or virtual goods or properties, all these reward-kind game elements can act as external purposes for individuals to take part in a project. Therefore, we hypothesize:

H1. Reward-kind game elements are positively associated with extrinsic motivations.

Humans make ability judgments about self and others via comparisons (Hoorens and van Damme, 2012). Festinger’s (1954) social comparison theory predicts that individuals compare themselves to others to validate opinions, make judgments and reduce uncertainty. As an intuitive comparison result, leaderboard/ranking could help individuals ascertain the level of their competence. The rise of ranking is no doubt a symbol to recognize the improvement of someone’s ability. Compared with ranking, game elements like points/score, level are not able to present individuals’ relative positions in the whole picture. However, as the basis of ranking and leaderboard, the value of points and levels can be a representative of someone’s ability. Badge or achievement is involved when individuals’ score or level reached some extent, which can be used as a certificate for some kind of a “quality change” and recognition for someone. Hence, recognition-kind game elements like ranking, scores, levels, badges and achievements can recognize the behaviors or improvement of individuals from different perspectives, meeting the psychological need of recognition. Topper ranking, higher scores and levels, more badges or achievements could be considered as purpose or incentives for participants to take the activity enthusiastically. In other words, these kinds of game elements can enhance participants’ extrinsic motivation. In addition, the improvement of one’s ranking, the increase of scores and levels, the acquisition of new badges, the fulfillment of a new achievement could provide participants with a sense of achievement, make them realize their accumulation and progress, which can help turn such
a behavior stimulated by outer purpose into an internal regulation. Therefore, by enhancing the sense of achievement, game elements that are supposed to provide recognition could internalize the extrinsic motivation. Based on the reasons above, we hypothesize:

H2a. Recognition-kind game elements are positively associated with extrinsic motivation.

H2b. Recognition-kind game elements are positively associated with internalized extrinsic motivation.

Another kind of game elements that could provide participants with a sense of achievement is the kind offering the participants information on his/her progress. The common game elements of this kind are progress bar and sub-missions. Progress bar can take every effort participants put into account and turn it into a visible bar, so that participants are able to get aware of how much has been done, as well as how much has to be done in real time, while the design of sub-missions or challenges enables participants to acquire a sense of satisfaction from every tiny fulfillment or accomplishment until they finally meet all the acquirements and finish their job.

Self-efficacy is defined as one's belief in one's ability to perform a task (Gist, 1987). Vroom's (1964) expectancy theory has indicated the positive relationship between self-efficacy and the effort they are likely to put. In another word, higher self-efficacy participants feel leads to more efforts that individuals may expend. Despite the fact that most crowdsourcing projects include lack of challenges and difficulties, there is still a possibility that participants feel impatient and unmotivated. Many participants choose to give up owing to the lack of information on how much time and efforts they must put in to reach the finishing line. However, game elements designed to provide such kind of information can assist participants in ascertaining where they are, how much effort and time is still needed and give them the confidence of their ability to accomplish the job, which in other words, can enhance their self-efficacy.

On the one hand, progress-kind game elements are capable of providing goals for participants to achieve and giving constant recognition of their effort, by which it can enhance extrinsic motivation of participants. On the other hand, these game elements can provide information on one's progress that can lead to a sense of achievement, as well as enhance participants' self-efficacy and eventually increase internalized extrinsic motivation. Hence, we hypothesize:

H3a. Progress-kind game elements are positively associated with extrinsic motivation.

H3b. Progress-kind game elements are positively associated with internalized extrinsic motivation.

Csikszentmihalyi (1990) has brought up the theory of flow to explain the acceptance intention and behavior of individuals toward some specific technology. Flow is defined as a state in which people are so involved in something that nothing else seems to matter. What's more important is that when individuals are immersed in the flow provided by some object, they will more easily feel a sense of enjoyment and satisfaction. In other words, intrinsic motivation will be enhanced in this state.

Fantasy-kind game elements such as storytelling and virtual territory can inundate participants with the information flow. Such game elements could meet all four dimensions brought up by Trevino and Webster (1992) which are used to describe the characteristics when individuals are in the state of flow. First, participants could get a new clue or feedback when they have conducted any specific activities, making the participants feel like they are the ones who are controlling the process of the whole project; the constant feedback is more
likely to grab participants' attention; participants' curiosity may be aroused considering they might want to know what will happen next; and ultimately, the constant information flow of the virtual story or territory could provide participants with a sense of enjoyment and satisfaction, while users' intrinsic motivation gets enhanced. Hence, we hypothesize:

**H4.** Fantasy-kind game elements are positively associated with the intrinsic motivation.

### 3.3 Intrinsic motivation, extrinsic motivation and internalized extrinsic motivation

Intrinsic motivation involves people doing an activity because they find it interesting and derive spontaneous satisfaction from the activity itself (Wu and Lu, 2013), which has an influence on whether the application of gamification would meet the expectation. Those companies that figure out how to effectively use gamification to amplify the intrinsic motivations of their employees, fans and customers will have a lasting competitive edge in their markets (Deterding, 2012). Intrinsic motivation is one of the cardinal incentives for participants to take part in crowdsourcing projects, as participants might choose these tasks out of the enjoyment they might gain in the process of overcoming the challenge and solve the problem (Zheng et al., 2011). Thus, the effect of intrinsic motivation has obtained wide acceptance from scholars and practitioners. Therefore, we hypothesize:

**H5.** Intrinsic motivation is positively associated with user participation.

Extrinsic motivation is the motivation to work for something apart from and external to the work itself, such as reward or recognition from other people (Zheng et al., 2011). Extrinsic motivation requires an instrumentality between the activity and some separable consequences, so satisfaction comes not from the activity itself but rather from the extrinsic consequences to which the activity leads (Gagne and Deci, 2005). Reward and recognition are most common extrinsic motivations in the context of crowdsourcing (Brabham, 2008; Fuller, 2010; Kavaliova et al., 2016; Sun et al., 2015; Zheng et al., 2011). They are both capable of becoming a purpose for individuals to motivate themselves to take some specific actions. Hence, we hypothesize:

**H6.** Extrinsic motivation is positively associated with user participation.

The SDT and situational relevance theory advocate that the motivational affordances of the game mechanics are most effective when they work toward personally meaningful goals and when users are empowered to design or set their own goals (Wu, 2011). Deterding et al. (2011) even have a viewpoint that a mechanic can motivate only when there is a match between this mechanic and the user's profile (needs, values and interests). According to the definition of internalization (Ryan et al., 1985), internalized extrinsic motivations could make people “taking in values, attitudes or regulatory structures”, thus could align the personal goals with the context of the project. Hence, internalized extrinsic motivation could motivate the engagement behaviors of participants, even better than extrinsic motivation. Thus, we hypothesize:

**H7.** Internalized extrinsic motivation is positively associated with user participation.

### 3.4 Task complexity as a moderator

According to Wood (1986), task complexity is a function of the number of distinct acts that must be completed and the number of distinct information cues about the attributes of the task-related stimulus object an individual has to process when performing a task. Jiang and
Benbasat (2007) hold the opinion that high task complexity can increase information processing requirements and demand more resources from task executors. An increase in task complexity equals the increase in efforts and time that participants must spend on the tasks; thus, they may find that the awards or recognition from others are not worthy of efforts of participating in the project. As a result, they are less likely to be lured by extrinsic motivation. Thus, we think:

**H8a.** The effect of game elements that could enhance extrinsic motivation will be less prominent when tasks become highly complex.

Besides, as tasks in crowdsourcing projects are relatively simple, repetitive and unchallenging, individuals’ attention to tasks may get diluted when task complexity reaches a certain level (Todd and Benbasat, 1999; Kahneman, 1973). As participants’ attention is no longer paid to the task itself, they may not be attracted by the information flow those tasks provide and consequently get out of the state of flow and obtain no fun from the process of accomplishing the task. Therefore, we hypothesize:

**H8b.** The effect of game elements that could enhance intrinsic motivation will be less prominent when the tasks become highly complex.

Things may become different when referring to the game elements that could enhance internalized extrinsic motivation. The internalized extrinsic motives we have discussed so far are self-efficacy and achievement. As simple tasks that could be done in no time, the progress bar may not play a significant role as a reminder. However, with an increase in task complexity, these kinds of game elements are beginning to make a difference. These game elements can remind participants of how much they have done and how much remains ahead of them, and thus can enhance the self-efficacy and weaken the anxiety led by task complexity. Besides, simple tasks may not be able to provide participants with enough sense of achievement, but complex tasks can make it happen. Thus, we hypothesize:

**H8c.** The effect of game elements that could enhance internalized extrinsic motivation will be more prominent when tasks become highly complex.

4. Research methodology

To test the hypotheses, a 5 (no game element vs reward vs progress bar vs virtual territory vs point and leaderboard) by 2 (different choice sizes of the kinds of clothes) between-subjects’ laboratory experiments will be conducted. We plan to design a crowdsourcing task where we will act as a clothing company that needs to collect opinions and preferences of customers based on their choices of clothes.

In the control group, none of game elements will be applied and a list of clothing will be shown to the participants who are supposed to choose the clothes they prefer from each kind. In each of our experimental groups, a specific game element will be embedded, so that the motivation effect of game elements from different categories will be examined separately. The setting of our control group and experimental groups is shown in Table III. While in the second dimension, the influence of task complexity that acted as a moderator will be tested by using two different choice sizes of the kinds of clothes which participants are requested to choose from.

Data will be collected from two sources, the performance data from system logs and the subjective data from post-survey. The time those participants have spent on choosing clothes and the number of clothes they have chosen will be retrieved from the system and used as the measure for user participation. Measurement items for several major
constructs – intrinsic motivation, two kinds of extrinsic motivation and two kinds of internalized intrinsic motivation – will be adapted from the Work Preference Inventory (Amabile et al., 1994), which is used to assess intrinsic and extrinsic motivation. Two different choice sizes of the kinds of clothes will be taken as the measure of task complexity.

5. Expected contribution
Gamification has attracted increasing attention from researchers and practitioners in varied contexts. Our research can make contributions in several ways.

First, this research uncovers the motivation mechanism of these game elements, enabling us to clarify the latent relationship between game elements and user behavior. Second, we are more likely to find out which game elements are more effective in crowdsourcing projects through a series of experimental groups. Third, by categorizing game elements and selecting representative game elements in each category as our research subjects, we are able to discover motivation processes of game elements. Forth, as we use task complexity as a moderator, we may be able to identify whether task complexity influences the application of game elements, as well as which game elements should be selected in varied extents of task complexity. Last but not the least, as our model is proposed within the context of crowdsourcing, the framework that indicates the interrelationship between game elements, human motivation and user participation could also be referred and adapted by other research investigations that are not limited to crowdsourcing projects.
References


Porter, L.W. and Lawler, E.E.I. (1968), Managerial Attitudes and Performance, Irwin-Dorsey, Homewood, IL.


Corresponding author
Zhiyuan Zeng can be contacted at: zengzy16@163.com

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Computerizing Trail Making Test for long-term cognitive self-assessment

Zhiwei Zeng
Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, Nanyang Technological University, Singapore, Singapore and Interdisciplinary Graduate School, Nanyang Technological University, Singapore, Singapore

Chunyan Miao
Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, Nanyang Technological University, Singapore, Singapore and School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

Cyril Leung
Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, Nanyang Technological University, Singapore, Singapore, and Zhiqi Shen
Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, Nanyang Technological University, Singapore, Singapore and School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

Abstract

Purpose – This paper aims to adapt and computerize the Trail Making Test (TMT) to support long-term self-assessment of cognitive abilities.

Design/methodology/approach – The authors propose a divide-and-combine (DAC) approach for generating different instances of TMT that can be used in repeated assessments with nearly no discernible practice effects. In the DAC approach, partial trails are generated separately in different layers and then combined to form a complete TMT trail.

Findings – The proposed approach was implemented in a computerized test application called iTMT. A pilot study was conducted to evaluate iTMT. The results show that the instances of TMT generated by the

© Zhiwei Zeng, Chunyan Miao, Cyril Leung and Zhiqi Shen. Published in the International Journal of Crowd Science. Published by Emerald Publishing. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore under its IDM Futures Funding Initiative; and the Interdisciplinary Graduate School Research Scholarship. The authors would like to thank Jun Ji for his contributions in application implementation and testing.
DAC approach had an adequate level of difficulty. iTMT also achieved a stronger construct validity, higher test–retest reliability and significantly reduced practice effects than existing computerized tests.

**Originality/value** – The preliminary results suggest that iTMT is suitable for long-term monitoring of cognitive abilities. By supporting self-assessment, iTMT also can help to crowdsource the assessment processes, which need to be administered by healthcare professionals conventionally, to the patients themselves.

**Keywords** Telehealth, Self-assessment, Computerized cognitive assessment, Longitudinal assessment, Practice effect, Trail Making Test

**Paper type** Technical paper

1. Introduction

Both falls and dementia are major health concerns among the elderly. About one-third of the elderly aged over 65 years fall each year (Tinetti et al., 1988). Falls also account for about 10 per cent of visits to hospital emergency departments among the elderly (Sattin, 1992). On the other hand, a new dementia case is reported every 3.2 seconds, and the cost of dementia is equivalent to about 1.1 per cent of global Gross Domestic Product (GDP) in 2015[1]. A decline in cognitive functions has been associated with increased risk of fall (Muir et al., 2012) and identified as a precursor syndrome to dementia (Lyketsos et al., 2002). Thus, monitoring changes in cognitive functions may be helpful for fall prevention, as well as early diagnosis and intervention of dementia.

Repeated assessments are required to track the changes in cognitive functions timely and effectively. The Trail Making Test (TMT) is one of the most frequently used neuropsychological tests for cognitive assessment (Butler et al., 1991) due to its sensitivity, simplicity and ease of administration. The current version of TMT is adapted by Reitan (1955) (thereafter referred to as Reitan’s TMT), which has only one instance and is administered using paper and pencil. However, the repeated use of Reitan’s TMT in longitudinal assessment is severely limited by its high susceptibility to practice effects (Beglinger et al., 2005). Practice effects refer to improvements in test performance that occur when a subject is retested on the same instance, or tested repeatedly on very similar ones. It is hard to segregate performance improvements due to practice effects from meaningful cognitive changes. This affects test accuracy and reliability. Increasing the test–retest interval may attenuate practice effects, but this may obscure the timely detection of meaningful cognitive changes (Buck et al., 2008).

Some researchers have proposed to use alternative forms (Atkinson and Ryan, 2007) and mirror images (Wagner et al., 2011) of TMT serially in consecutive test administrations to reduce practice effects. However, the number of equivalent alternative forms and mirror images of TMT is limited. Other researchers have proposed more systematic and divergent approaches for generating new instances of TMT (Vickers et al., 1996; Vickers and Lee, 1998). Although their approaches could generate theoretically unlimited instances of TMT, the generated instances may be less difficult than Reitan’s version – they may have shorter average trail length and less visual interference than the instance in Reitan’s version. Consequently, the generated instances may have poorer diagnostic efficacy and less discriminating power to distinguish among different cognitive status. Moreover, as a cognitive assessment tool, the generated instances should be assessed for their validity and reliability; yet, to the best of our knowledge, none of these two approaches have been validated in user studies.

To generate TMTs that can be used in repeated assessments, we propose a divide-and-combine (DAC) approach – a systematic approach for generating instances of TMT which:
(1) are sufficiently different from each other to reduce practice effects when used in consecutive test administrations; and
(2) have a similar level of difficulty and thus diagnostic power to Reitan’s TMT.

To achieve (1), the proposed approach uses pseudo-randomized processes to generate different instances. To achieve (2), our approach attempts to reproduce the spatial characteristics of Reitan’s TMT to the greatest extent possible. According to Vickers et al. (1996), trails in Reitan’s TMT are self-avoiding and gradually unwind in a clockwise or anticlockwise direction. To reproduce these characteristics, our DAC approach generates sub-solutions in divided problem spaces and combines sub-solutions to form a complete solution. In the “divide” phase, the test region is divided into several nested and non-overlapping layers. Within each layer, a partial trail is generated with the desired spatial characteristics. Then, in the “combine” phase, the partial trails are joined together to form a complete trail while preserving the desired characteristics.

With the increasing penetration and improved usability of digital devices, there has been a number of attempts to administer TMTs using digital devices, e.g. smartphones, tablets and computers. Computerized tests can facilitate more standardized and accurate data capturing as well as support detailed analysis. Following pre-designed test generation algorithms, they can also implement unlimited instances of TMT using systematic approaches. With these advantages, computerized tests also open the possibility to self-assessment in home environment. Traditionally, cognitive assessments need to be administered by healthcare professionals. With the aid of well-designed computerized tests, the assessment processes can be crowdsourced to the patients themselves.

To evaluate our DAC approach, we created a test application, called iTMT, which implements the DAC approach to generate computerized TMTs. A pilot study was conducted using this application and involving ten participants with different levels of cognitive abilities. The pilot study results suggest that no significant difference exists between the computerized tests generated by our DAC approach and Reitan’s TMT in terms of total segment length and visual interference, indicating that they had a similar level of difficulty. Moreover, iTMT also demonstrated stronger construct validity, higher test–retest reliability and significantly reduced practice effects than existing computerized tests. These preliminary results support the effectiveness of the DAC approach and indicate that iTMT is a promising tool in longitudinal cognitive assessment.

In the following sections, we first introduce the neuropsychological background of TMT. Then, we review the prior efforts made to adapt it for longitudinal assessment and to computerize it. To address the issues identified from prior works, we propose our DAC approach in Section 4. In Section 5, we describe how we computerized our DAC approach and created iTMT. The pilot study results are presented in Section 6. Finally, the main findings are presented in Section 7.

2. Neuropsychological background of Trail Making Test
The TMT is one of the most frequently used neuropsychological tests (Butler et al., 1991) due to its sensitivity, simplicity and ease of administration. The TMT was originally constructed to assess general intelligence in Army Individual Test Battery (1944). It was later adapted by Reitan (1955) and included in the Halstead-Reitan Neuropsychological Battery in its current form. It is often used as a diagnostic tool to detect cognitive
impairments due to brain damage, e.g. dementia, stroke and traumatic brain injuries (Ashendorf et al., 2008; Chen et al., 2015).

During test administration of Reitan’s TMT, the subject is instructed to connect a set of dots as quickly as possible while maintaining accuracy. Reitan’s version only has one test instance and uses fixed dot arrangements. It contains two parts. As shown in Figure 1, Part A involves connecting 25 numbered dots in increasing order. Part B involves connecting 25 dots labelled with numbers and letters, in the alternating sequence “1-A-2-B-3-C[…]”. The test is meant to assess cognitive functions such as cognitive flexibility, executive functioning, mental processing speed, divided attention as well as visual scanning (Sanchez-Cubillo et al., 2009). Part B is more difficult than Part A, possibly due to differences in symbolic complexity and spatial arrangement (Fossum et al., 1992). While Part A only contains numerals, Part B involves two symbol systems, alphabets and numerals, making it a more difficult cognitive task. Besides, segment length and amount of visual interference are also factors affecting the difficulty of TMT. Part B has a longer total trail length and more visual interfering stimuli than Part A, making it more demanding in terms of motor speed and visual scanning (Gaudino et al., 1995).

Reitan’s TMT is performed using paper and pencil and is administered by a healthcare professional. Time to completion and frequency of errors are the most common metrics which are recorded and used to interpret TMT performance of both Parts A and B.

3. Related work

3.1 Adapting Trail Making Test for longitudinal assessment

Although Reitan’s TMT is widely used for assessing cognitive abilities, its use in longitudinal assessment is constrained due to its high susceptibility to practice effects.
Each repeated administration of it uses the same test instance, which decreases its sensitivity.

Atkinson and Ryan proposed to use other validated neuropsychological tests as alternative forms of TMT in longitudinal assessment (Atkinson and Ryan, 2007; Atkinson et al., 2010). They identified three alternative forms that are equivalent and can be used interchangeably in a serial manner with a brief test–retest interval. However, due to their similarity in content and format, the use of alternative forms can only slightly reduce practice effects during longitudinal assessment. Instead of using other validated tests, Wagner et al. (2011) proposed to create new instances of TMT using its mirror images. Similar to the use of alternative forms, the mirror images also exhibited discernible practice effects when used serially in assessment.

The number of equivalent alternative tests and mirror images of TMT is limited. Hence, it is still not practical to adopt the aforementioned two approaches in longitudinal cognitive assessment over a long period. A group of researchers (Vickers et al., 1996; Vickers and Lee, 1998) have proposed more systematic and divergent approaches for generating a theoretically unlimited number of TMT instances. They contended that, although it is not clear whether Reitan’s TMT was constructed according to some implicit principles, it is vastly different from trails that are generated by purely random processes (Vickers et al., 1996). They observed two characteristics of Reitan’s TMT trails. First, for both Parts A and B, the trails gradually uncoil in either a clockwise or an anticlockwise direction from the inner to outer part. Second, the trails are self-avoiding, i.e. the line segments connecting consecutive points have no intersections with one another. However, it is a non-trivial problem to generate new instances of TMT which are endowed with these two characteristics (Vickers and Lee, 1998). Working reversely from the desired characteristics, two approaches to generate self-avoiding trails were proposed in Vickers et al. (1996) and Vickers and Lee (1998).

The first approach is suggested by considering the fractal nature of TMT trails. The problem of generating new TMT instance is transformed into the problem of generating self-avoiding fractal curve (Vickers et al., 1996). Starting with a seed element, fractal curves are generated by repeatedly applying a set of transformations to the seed. The second approach transforms the problem of generating new TMT instance into a travelling salesman problem (TSP) (Vickers and Lee, 1998). Dots in TMT are treated as cities in TSP which can only be visited exactly once in a trail. The solution to TSP, the shortest path, is typically self-avoiding as having intersections in the trail tends to lengthen the path. Thus, given a set of dots, the solution to the corresponding TSP can be converted into a new instance of TMT.

Although both the fractal curve and the TSP approach can generate a theoretically unlimited number of TMT instances, the generated instances may have shorter average trail length and less visual interference than Reitan’s paper-and-pencil TMT. Consequently, these instances may be less difficult than Reitan’s version and less demanding in terms of the cognitive functions assessed. For example, less visual interfering stimuli would make the visual scanning much easier. Moreover, when test performance is measured by time to completion, the total segment length will also affect the test performance, as it affects the drawing time. Due to their reduced level of difficulty, the generated instances may not have enough discriminative power to produce statistically significant performance differences between different cognitive status. Consequently, their sensitivity and diagnostic efficacy may also be compromised.
3.2 Computerized Trail Making Test

Unlike manually administered tests which are susceptible to variations in test administrator and test procedures, computerized tests maintain standardized test procedures and are consistent across subjects (Woods et al., 2015). Smith (2012) developed a computerized touch-screen version of TMT, eTrails, which contains a digital embodiment of the paper-and-pencil TMT and four computerized variants of the standard test. Experimental results suggest that all the five computerized tests have considerably higher test–retest reliability than TMT, possibly due to standardized procedures and less administrative errors. Another benefit brought by computerized tests is the ability to capture more high-fidelity data which increase the accuracy of performance measurements and facilitate in-depth performance analysis. Woods et al. (2015) developed a computerized TMT, C-TMT, which supports segment-by-segment analysis of performance and separates analysis of time spent on different tasks, e.g. dwelling and moving.

More importantly, computerized tests also provide the ideal format for generating a theoretically unlimited number of TMT instances. Following algorithmic approaches, pseudo-randomized dot arrangements can be generated for each test administration. Compared to Reitan’s TMT, the practice effects for both eTrails and C-TMT were significantly attenuated (Smith, 2012; Woods et al., 2015). Yet, it was not clearly indicated that how the instances of TMT were generated in these two computerized tests. The fractal curve approach and the TSP approach introduced in the previous section are both systematic and divergent approaches rooted in sound theoretical basis. However, to the best of our knowledge, neither of them has been formally computerized and validated in user studies.

We have reviewed the prior efforts made to adapt TMT for longitudinal assessment and to computerize TMT, as well as discussed the issues with these attempts. Next, we will propose our solution to the issues.

4. The divide-and-combine approach

In this section, we propose a DAC approach for generating different instances of TMT in a systematic manner. As introduced in Section 1, the generated instances need to be sufficiently different from each other while having a similar level of difficulty to Reitan’s TMT. More specifically, the aim of the DAC approach is to produce different sets of ordered dots in a rectangular test region, so that the trail obtained by connecting dots in each set according to their order exhibits the following characteristics:

- uncoiling in clockwise or anticlockwise direction; and
- self-avoiding.

The DAC approach consists of two phases. In the “divide” phase, the test region is divided into several nested and non-overlapping layers. Within each layer, a partial trail is generated with the desired spatial characteristics. Then, in the “combine” phase, the partial trails are combined together to form a complete trail.

Suppose that we want to generate instances of TMT that contains $n$ dots in a rectangular test region on the $x$-$y$ plane. The plane spans across the area $x \in [a, b]$ and $y \in [c, d]$, where $a < b$ and $c < d$.

Suppose that we use $m$ layers, where $m \in \mathbb{N}^+$. There are no constraints on the shape of the layers. However, the layers defined should satisfy two requirements. First, to make the generated trail follow an unwinding pattern, the layers need to be nested. Second, the layers should be non-overlapping to reduce intra-layer intersections.
Generally, the inner most layer can be a solid shape, while outer layers can be nested hollow shapes.

**Definition 1:** A layer $L_i$, where $i = 1, 2, \cdots, m$, is defined as a 4-tuple $L_i = \{d_i, A_i, P_i, S_i\}$:
- $d_i \in \mathbb{R}_{>0}$ represents the relative density of the dots in layer $L_i$ (refer to Definition 2).
- $A_i$ defines the area layer $L_i$ spans;
- $P_i = \{(x_j, y_j) | x_j \in [a, b]; y_j \in [c, d]; j = 1, 2, \cdots, d_i \times n\}$ represents the dots in layer $L_i$; and
- $S_i = \{s(k) | s(k) \in [1, d_i \times n]; k = 1, 2, \cdots, d_i \times n\}$ represents the order of the dots in layer $L_i$. To form a partial trail $T_i$, the dots are connected according to the sequence of $p_{s(1)} \rightarrow p_{s(2)} \rightarrow \cdots \rightarrow p_{s(d_i \times n)}$.

**Definition 2:** The density of dots $D_i$ in a layer $L_i$ can be defined as:

$$d_i = \frac{|P_i|}{n} \quad (0 < d_i \leq 1)$$

where $d_i$ is a measure of relative density and is the ratio of number of dots in layer $L_i$ (the cardinality of $P_i$) to the total number of dots $n$.

The pseudo code of the DAC approach is outlined in Algorithm 1. In the “divide” phase (Lines 3-5 of Algorithm 1), a partial trail $T_i$ is generated within each layer. For each layer $L_i$, the dots are randomly generated within the layer and then sorted to determine their connecting order. The function `randomDots()` is called to generate $d_i \times n$ number of dots in the area defined by $A_i$. Then, the function `sortDots()` is invoked to sort the dots in layer $L_i$ with respect to an anchor point and return their sorted order $S_i$. For the simplicity, the bottom left dot in the layer can be chosen as the anchor point in sorting. The dots are sorted according to their angles with respect to the anchor point. The order of sorting (increasing or decreasing angle) is determined randomly for each layer. A self-avoiding partial trail can be formed by traversing the dots by their sorted order. Figure 2 illustrates sorting by increasing angle.

![Figure 2](image-url)

**Notes:** The bottom left dot in the layer, $P_{anchor}$, is chosen as the anchor point in sorting. The dots are then connected according to the sorted order “$s(1) - s(2) - \cdots - s(6)$”, where $s(1)$ is the anchor point.

![An illustration of sorting six dots in a layer by calling the function sortDots()](image-url)
Algorithm 1: The DAC approach

1: \texttt{var} \(P_{\text{regen}}\) \Comment{The dots that need to be regenerated}
2: \texttt{var} \(p_{\text{pivot}}\) \Comment{The pivot dot for combining two sub trails}

3: \texttt{for each} \(L_i\) \Comment{The “divide” phase}
4: \(P_i \leftarrow \text{randomDots}(A_i, d_i \times n)\)
5: \(S_i \leftarrow \text{sortDots}(P_i)\)

6: \texttt{for} \(i = 1\) \texttt{to} \(m - 1\) \Comment{The “combine” phase}
7: \texttt{while} \(\text{intersect}(T_i, T_{i+1}) \neq \text{NULL}\) \Comment{Dealing with intersections}
8: \(P_{\text{regen}} \leftarrow \text{intersect}(T_i, T_{i+1})\)
9: \(P_{i+1} \leftarrow \text{regenerateDots}(P_{i+1}, P_{\text{regen}})\)
10: \(S_{i+1} \leftarrow \text{sortDots}(P_{i+1})\)
11: \texttt{for} \(i = 1\) \texttt{to} \(m - 1\) \Comment{Connecting adjacent layers}
12: \(p_{\text{pivot}} \leftarrow \text{pickPivot}(S_i, P_i, S_{i+1}, P_{i+1})\)
13: \(S_{i+1} \leftarrow \text{connect}(S_i, S_{i+1}, p_{\text{pivot}})\)
14: \texttt{for each} \(P_i\) \texttt{do}
15: \(\text{assignLabel}(P_i, S_i)\)

In the “combine” phase (Lines 6-15 of Algorithm 1), the partial trails in adjacent layers are combined together to form a complete trail. Before connecting the partial trails, a round of checking is performed to ensure that there are no intersections among them (Lines 6-10). The function \(\text{intersect}(T_i, T_{i+1})\) is used to detect intersections between the partial trails in two adjacent layers, \(L_i\) and \(L_{i+1}\). It returns the end points of line segments in the outer layer \(T_{i+1}\) that intersects with \(T_i\). In the case when two trails intersect, i.e. \(\text{intersect}(T_i, T_{i+1}) \neq \text{NULL}\), the trail in the outer layer will be adjusted to eliminate the intersection. The end points of the intersecting segments in \(T_{i+1}\) (or \(P_{\text{regen}}\)) are regenerated by calling the function \(\text{regenerateDots}()\). With the newly generated dots, the dots in layer \(L_{i+1}\) are re-sorted by calling \(\text{sortDots}()\) again. When no intersections are found, all the partial trails are connected together to form a complete trail (Lines 11-13). By calling \(\text{connect}()\), each two partial trails in adjacent layers are joined through a pivot dot selected from the outer layer \(L_{i+1}\). The function \(\text{pickPivot}()\) is called to choose a pivot from \(P_{i+1}\) to connect \(T_i\) and \(T_{i+1}\). The choice of pivot dot should guarantee that no intersection will be resulted from joining it and the last dot in the inner layer. Figure 3 illustrates connecting two trails \(L_{i-1}\) and \(L_i\), where \(p_{(3)}\) in \(L_i\) is chosen as the pivot dot. Finally, alphanumerical labels are assigned to the dots based on their order in the complete trail.

With the algorithm described above, the DAC approach is able to generate TMT instances that reproduce the two desired spatial characteristics: unwinding clockwise or counter-clockwise and self-avoiding. In the following part of this section, we will explain how the design of the DAC approach help to reproduce these characteristics.

### 4.1 Unwinding clockwise or counter-clockwise

The DAC approach uses both intra-layer and inter-layer designs to reproduce this characteristic. The intra-layer strategy ensures the partial trails exhibits the desired pattern locally within a layer, while the inter-layer strategy ensures that the pattern is embodied globally by a trail.

Within a layer (intra-layer), sorting and connecting the dots according to their angles with respect to the anchor point endows a partial trail with the desired unwinding pattern. As shown in Figure 2, the dots in a layer can also be viewed in a polar coordinate system with the anchor point as pole. Each dot can be represented by a radial coordinate \((r, \theta)\),
where $\theta \in [0, 2\pi]$. Connecting the dots according to the value of $\theta$ would form a trail that starts from and then zigzags around the anchor point. Visually, it can be viewed as a trail gradually uncoiling from the anchor point. The direction of unwinding can be controlled by the sorting order. If the angles are sorted in ascending order, the partial trail unwinds counter-clockwise. Conversely, the partial trail unwinds clockwise.

At the inter-layer level, the definition of the layers help to create a complete trail in the desired unwinding pattern. The layers defined must be nested. This requirement ensures that the area of an inner layer is surrounded by its outer layers. The trail formed by joining partial trails within nested layers also appears to be nested, which gradually unfolds when traversing from an inner layer to an outer layers.

4.2 Self-avoiding

To generate self-avoiding trails, the DAC approach also uses intra- and inter-layer strategies. The intra-layer strategy ensures each partial trail formed is self-avoiding. The inter-layer strategy generates a self-avoiding trail by connecting the partial trails while avoiding intersections in the connecting process.

At the intra-layer level, sorting and connecting dots by angles ensures that the resultant partial trail is self-avoiding. Referring to Figure 2 again, the area that spans across $[0, 2\pi]$ is divided into sectors by the dotted lines connecting the anchor point to other dots. These sectors are contiguous but non-overlapping. Each line segment of a partial trail spans within exactly one of the sectors. Thus, the line segments have no intersections with each other, except for the joints.

At the inter-layer level, using non-overlapping layers can help to reduce intersections among partial trails in different layers. However, a partial trail formed by traversing the dots within a layer may still cut through other layers. Intersection checking is performed to detect and eliminate intersections in such cases, as described by Algorithm 1 Lines 6 to 10. Moreover, the choice of pivot dot while connecting adjacent trails also avoids creating intersections in the combined trail.

The DAC approach is a generalized approach that is suitable for generating trails for both TMT Parts A and B. It maintains the modelling flexibility to accommodate the differences between Parts A and B. The dots in Part A scatter more evenly over the rectangular test region, while the dots in Part B are more skewed towards the rim of the

**Before connecting** $L_{(i-1)}$, $L_i$

**Inner Layer**

$P_{s(1)} \rightarrow P_{s(2)} \rightarrow P_{s(3)} \rightarrow \ldots \rightarrow P_{s(d(i-1) \times n)}$

**Outer Layer**

$P_{s(1)} \rightarrow P_{s(2)} \rightarrow P_{s(3)} \rightarrow P_{s(4)} \rightarrow \ldots \rightarrow P_{s(d \times n)}$

**After connecting** $L_{(i-1)}$, $L_i$

**Inner Layer**

$P_{s(3)} \rightarrow P_{s(4)} \rightarrow \ldots \rightarrow P_{s(d \times n)} \rightarrow P_{s(1)} \rightarrow P_{s(2)}$

**Outer Layer**

$P_{s(1)} \rightarrow P_{s(2)} \rightarrow P_{s(3)} \rightarrow \ldots \rightarrow P_{s(d(i-1) \times n)}$

**Pivot**

Figure 3. An illustration of connecting partial trails in two adjacent layers, $L_{(i-1)}$ and $L_i$. $P_{s(3)}$ in the outer layer is chosen as the pivot point for connecting the two trails; the last point in the inner layer ($L_{(i-1)}$) is connected to $P_{s(3)}$. 

Trail Making Test

91
area. There are a number of strategies (Table I) to manipulate the distribution of the dots to cater to such differences.

In this section, we proposed the DAC approach for generating different instances of TMT. In the following sections, we will describe how this algorithmic approach can be implemented and built into computerized TMTs.

5. Computerizing Trail Making Test using the divide-and-combine approach

We created a test application called iTMT, which implements the DAC approach with the parameters determined from Reitan’s paper-and-pencil TMT.

After examining the dot arrangement in Reitan’s TMT, we chose to use 25 dots and three layers for both Part A and Part B, i.e. \( n = 25 \), \( m = 3 \). For the ease of modelling, we used three concentric layers, with each of their centres at the origin. As illustrated in Figure 4, \( L_1 \) is a rectangle, while \( L_2 \) and \( L_3 \) are hollow rectangles. Collectively, the three non-overlapping layers cover the test region exhaustively. The values of \( d_i \) are presented in Table II. Compared to Part A, Part B has smaller relative density in \( L_1 \) and \( L_2 \), but much greater density in \( L_3 \), which is designed in accordance with the observations made from the paper-and-pencil TMT.

<table>
<thead>
<tr>
<th>Parameter to adjust</th>
<th>Part A</th>
<th>Part B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative density of dots in a layer, ( d_i )</td>
<td>Assign larger ( d_i ) to inner layers, smaller ( d_i ) to outer layers</td>
<td>Assign smaller ( d_i ) to inner layers, larger ( d_i ) to outer layers</td>
</tr>
<tr>
<td>Area of a layer, ( A_i )</td>
<td>Define smaller ( A_i ) for inner layers, larger ( A_i ) for outer layers</td>
<td>Define larger ( A_i ) for inner layers, smaller ( A_i ) for outer layers</td>
</tr>
<tr>
<td>Total number of layers, ( m )</td>
<td>Use larger ( m ) for Part A</td>
<td>Use smaller ( m )</td>
</tr>
</tbody>
</table>

**Figure 4.**
The three layers used in the implementation of iTMT \((m = 3)\)

**Notes:** \( L_1 \) is a rectangle, while \( L_2 \) and \( L_3 \) are hollow rectangles. Collectively, the three non-overlapping layers cover the test region exhaustively.
The area each layer spans can be defined as following:

\[
A_i \text{ spans the area } \begin{cases} 
|x| \leq x_1, |y| \leq y_1 & i = 1 \\
 x_1 < |x| \leq x_2, y_1 < |y| \leq y_2 & i = 2 \\
 x_2 < |x| \leq x_3, y_2 < |y| \leq y_3 & i = 3 
\end{cases}
\]

Following the strategies in Table I, \( A_i \) is defined differently for Parts A and B. Comparatively, for Part B, \( A_1 \) spans across larger area, while \( A_2 \) and \( A_3 \) are made smaller so that more dots would be placed near the edge of the test region.

Using the aforementioned parameters, we built the test application, \( iTMT \). It is a touchscreen application that can administer computerized tests on tablets. Instead of using a pencil, the subjects need to connect the dots using their fingers. During each administration, \( iTMT \) generates a new instance of TMT using the DAC approach. Figure 5 shows a test screenshot of \( iTMT \), where Part A was being administered.

### 6. Evaluating computerized Trail Making Test

To evaluate the DAC approach, a pilot study was conducted for \( iTMT \). The instances of computerized TMT generated with the DAC approach was evaluated for the following aspects:

<table>
<thead>
<tr>
<th>Density</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
<th>( d_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part A</td>
<td>0.24</td>
<td>0.36</td>
<td>0.4</td>
</tr>
<tr>
<td>Part B</td>
<td>0.2</td>
<td>0.2</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Note:** The trail formed by connecting the dots in sequence exhibits the two desired characteristics: uncoiling in clockwise or anticlockwise direction, and self-avoiding.
• **Level of difficulty**: Whether the generated instances are equivalent to Reitan’s paper-and-pencil TMT in terms of test difficulty (as measured by total segment length and visual interference).

• **Construct validity**: Whether the generated instances are able to produce test results that are comparable to Reitan’s paper-and-pencil TMT (as measured by time to completion).

• **Test–retest reliability**: Whether the test results are consistent across different instances (as measured by time to completion) and whether there are significant practice effects.

The study involved ten participants aged from 33 to 84 years, with 8 above 50 and 4 above 65. As age is an important factor which affects cognitive abilities, we recruited participants from a broad age range to include people with different levels of cognitive abilities. Considering that some elderly participants are illiterate or only had limited education, only **Part A** of TMT was administered. Each subject was administered three tests consecutively, including one paper-and-pencil test and two computerized tests on a tablet. The computerized tests were generated by iTMT using the DAC approach. As the size of test regions may affect the difficulty metrics, we used an adapted version of Part A in Reitan’s TMT for the paper-and-pencil test to control this factor. Reitan’s Part A is compressed proportionally into a smaller test region which has an equal area as the screen of the tablet used, with the relative positions of the dots preserved.

Before conducting the tests, the participant was shown a sample TMT with eight dots so as to become familiar with the connecting rules and test procedures. Then, the paper-and-pencil test was administered to the participant. The time taken to complete the trail, i.e. time to completion, was recorded with a timer. After an interval of 30 s, two computerized tests were administered to the participant consecutively. The test statistics were recorded by iTMT for later analysis.

6.1 **Level of difficulty**

As proposed in Gaudino et al. (1995), we use total segment length and visual interference as two metrics for the difficulty of TMTs. Total segment length is defined as the summed length of the shortest line segments connecting successive dots. Visual interference is quantified by summing the number of dots that lie within 3 cm of each line segment. The two metrics of the paper-and-pencil TMT administered in the study were calculated and used as normative values for evaluating the computerized tests. The paper-and-pencil TMT has a total segment length of 135.9 cm and visual interference of 49. For each of the 20 computerized tests (2 per participant × 10) administered during the study, its total segment length and visual interference was calculated. Two t-tests were performed to determine whether the computerized tests generated by iTMT have a similar level of difficulty to the paper-and-pencil test (Table III). One t-test compared the mean of total segment length (μ\textsubscript{seg}) of computerized tests and the total segment length of the paper-and-pencil test with the null hypothesis μ\textsubscript{seg} = 135.9. As p = 0.33 > α = 0.05 (t\textsubscript{0} = 0.98 < t\textsubscript{0.05,19} = 2.093), we accept the null hypothesis. The other t-test compared the mean of visual interference (μ\textsubscript{vis}) of computerized tests and the visual interference of the paper-and-pencil test with the null hypothesis μ\textsubscript{vis} = 49. As p = 0.31 > α = 0.05 (t\textsubscript{0} = 1.04 < t\textsubscript{0.05,19} = 2.093), we also accept H\textsubscript{0}.

The test results suggest that there is no significant difference between the computerized tests and the paper-and-pencil test in terms of total segment length and visual interference.
Hence, the results further indicate that, as measured by these two metrics, computerized tests and the paper-and-pencil test can be considered to have a similar level of difficulty.

6.2 Construct validity

The construct validity of the computerized tests generated by iTMT was measured by correlating the time taken by the same participant to complete a computerized test and the paper-and-pencil test. The Pearson product-moment correlations between time to completion are shown in Figure 6. The time taken by a participant to complete the first computerized test was positively correlated with the time he/she needs to complete the paper-and-pencil test \((r = 0.89, p = 0.0006)\). Positive correlation was also reported between the time to completion of the second computerized test and the paper-and-pencil test, with a higher correlation coefficient \((r = 0.97, p = 0)\). Both correlations reached statistical significance \((p < 0.01)\), indicating significant linear relationships between the time to completion of computerized tests and paper-and-pencil test. Compared to eTrails (Smith, 2012) (Table IV), an existing computerized test that achieved moderate correlation with the paper-and-pencil test (highest \(r = 0.668\)), iTMT demonstrated a much stronger correlation with the paper-and-pencil test \((r = 0.89\) and 0.97). When test performance was measured by time to completion, the computerized tests generated by iTMT were able to produce test scores that are highly correlated with the paper-and-pencil test score of the same participant, suggesting high construct validity of the computerized tests.

<table>
<thead>
<tr>
<th>(H_0)</th>
<th>(\bar{x})</th>
<th>(p)-value</th>
<th>(t) score</th>
<th>(t_0)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment length (\mu_{seg} = 135.9)</td>
<td>141</td>
<td>0.33</td>
<td>0.98</td>
<td>2.093</td>
<td>Accept (H_0)</td>
</tr>
<tr>
<td>Visual interference (\mu_{vis} = 49)</td>
<td>51.19</td>
<td>0.31</td>
<td>1.04</td>
<td>2.093</td>
<td>Accept (H_0)</td>
</tr>
</tbody>
</table>

Table III. Hypothesis tests for segment length and visual interference \((\alpha = 0.05)\)

**Figure 6.** The correlation between time to completion of computerized tests and paper-and-pencil test

**Notes:** Left: positive correlation between first iTMT test and paper-and-pencil test \((r = 0.89, p = 0.0006)\). Right: positive correlation between second iTMT test and paper-and-pencil test \((r = 0.97, p = 0)\)
6.3 Test–retest reliability
The test-retest reliability of the computerized tests generated by iTMT was calculated by correlating the time taken to complete the first and the second computerized test. The Pearson product-moment correlation between the time to completion of two computerized tests is shown in Figure 7. Significant linear relationship was observed between the time to completion of the two computerized tests \((r = 0.93)\). iTMT achieved considerably higher test–retest reliability than eTrails (highest \(r = 0.62\)) (Smith, 2012) and C-TMT (ICC = 0.87) (Woods et al., 2015). Moreover, iTMT was found to have almost no discernible practice effects. As shown in Figure 7, some participants took longer to complete the first test, while others took longer to complete the second test on iTMT. From the first to the second administration, the average time to completion was only reduced by 5.58 per cent. While the reduction in average time to completion for eTrail (Smith, 2012) was 10.33 and 12 per cent for C-TMT (Woods et al., 2015) (Table IV). Comparatively, iTMT was less susceptible to practice effects, providing further evidence to support its high reliability when administered repeatedly.

7. Conclusion
In this paper, we proposed a DAC approach to generate instances of TMT that can be used in longitudinal cognitive assessment. Our proposed approach is able to generated a theoretically unlimited number of different TMT instances which can be used in consecutive test administrations. Moreover, the instances generated by the proposed approach have a similar

Table IV.
Comparing the validity and reliability of iTMT and other existing computerized TMTs

<table>
<thead>
<tr>
<th>Name</th>
<th>Construct validity (measured by (r))</th>
<th>Test–retest reliability (measured by (r))</th>
<th>Test–retest reliability (measured by % reduction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>iTMT</td>
<td>0.89, 0.97</td>
<td>0.93</td>
<td>5.58</td>
</tr>
<tr>
<td>eTrails</td>
<td>0.668</td>
<td>0.62</td>
<td>10.33</td>
</tr>
<tr>
<td>C-TMT</td>
<td>–</td>
<td>0.87</td>
<td>12</td>
</tr>
</tbody>
</table>

Figure 7.
The correlation between time to completion of the first iTMT test and the second iTMT test \((r = 0.93)\)
level of difficulty to Reitan's paper-and-pencil TMT by reproducing its spatial characteristics. We also created a test application, iTMT, which implements the DAC approach to generate computerized TMTs. The preliminary results from the pilot study support the effectiveness of our DAC approach. Compared to existing computerized tests, the instances of TMT generated by the approach produced test results that were significantly more correlated with results of Reitan's TMT. Similar difficulty and highly correlated results suggest that these instances possess similar diagnostic power to the Reitan's TMT and are able to better distinguish subjects with different levels of cognitive abilities. iTMT also demonstrated higher test–retest reliability and significantly reduced practice effects than existing computerized tests. Due to the illiteracy of some participants, only TMT Part A and its computerized versions were tested in this study. In the future, we plan to study Part B in a similar way.

By supporting self-assessment, iTMT also can help to crowdsource the assessment processes, which need to be administered by healthcare professionals conventionally, to the patients themselves. To validate the feasibility of using iTMT in self-assessment over a long time period, it is worthwhile to study how test performance of it would change beyond the 2nd administration. Further experiments need to be conducted to find out whether the test performance of iTMT is convergent and how does the test performance converge.

Note


References

Army Individual Test Battery (1944), Manual of Directions and Scoring, War Department, Adjutant Generals Office, Washington, DC.


Further reading


About the authors
Zhiwei Zeng received the BEng (Hons) degree in Computer Science and BBus (Hons) degree from Nanyang Technological University (NTU) in 2015, where she is currently pursuing a PhD degree at the NTU-UBC Joint Research Centre of Excellence in Active Living for the Elderly (LILY), Interdisciplinary Graduate School. Her current research interests include artificial persuasion, its
computational modelling and applications of intelligent agents in healthcare systems. Zhiwei Zeng is the corresponding author and can be contacted at: i160001@e.ntu.edu.sg

Chunyan Miao is a Professor with the School of Computer Science and Engineering (SCSE) at Nanyang Technological University (NTU), Singapore. She is the Director of the NTU-UBC Joint Research Centre of Excellence in Active Living for the Elderly (LILY). Prior to joining NTU, she was an Instructor and a Post-Doctoral Fellow with the School of Computing, Simon Fraser University, Canada. Her research focuses on studying the cognitive and social characteristics of intelligent agents in multi-agent and distributed AI/CI systems, such as trust, emotions, incentives, motivated learning, ecological and organizational behavior. She has worked on new disruptive Artificial intelligence (AI) approaches and theories that synergize human intelligence, artificial intelligence and behavior data analytics (AI powered by humans). Her current research interests include human–agent interaction, cognitive agents, human computation and serious games.

Cyril Leung is a Member of the IEEE and the IEEE Computer Society. He received the BSc (Hons) degree from Imperial College, University of London, England, UK, in 1973, and the MS and PhD degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1974 and 1976, respectively. From 1976 to 1979, he was an Assistant Professor with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. From 1979 to 1980, he was with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada, where he is a Professor and currently holds the PMC-Sierra Professorship in Networking and Communications. He is the Deputy Director of the NTU-UBC Joint Research Centre of Excellence in Active Living for the Elderly (LILY). His current research interests include wireless communications systems. He is a Member of the Association of Professional Engineers and Geoscientists of British Columbia, Canada.

Zhiqi Shen is currently with the School of Computer Science and Engineering (SCSE), Nanyang Technological University, Singapore. He received the BSc in Computer Science and Technology from Peking University, Beijing, China, the MEng in Computer Engineering with the Beijing University of Technology, Beijing, and the PhD with Nanyang Technological University, Singapore. His current research interests include artificial intelligence, software agents, multi-agent systems (MAS); goal oriented modelling, agent oriented software engineering; internet of things, crowdsourcing, e-learning, agent augmented interactive media, game design and interactive story telling.
Emerald is excited to announce a recent partnership with Peerwith, a platform that provides authors with a variety of services.

The Emerald Peerwith site can be found here: https://authorservices.emeraldpublishing.com/

Peerwith connects academics seeking support for their work with a relevant expert to get their research submission-ready. Peerwith experts can help with the following: language editing, copy editing, scientific editing, translation services, statistical support, funding application support, visuals, video, publication support, literature search, peer review and indexing services. Authors post their assignments on the Peerwith site, experts provide a quote, and the fee and conditions are then agreed upon directly between the author and the expert.

While we are not, of course, guaranteeing publication upon use of Peerwith, we hope that being able to direct academics to this resource either before submission or during the peer review process will help authors further improve the quality of their papers and increase their chances of positive reviews and acceptance.

Academics with relevant expertise can sign up as an expert on the Peerwith system here: https://www.peerwith.com/services/offer
International Journal of Crowd Science

Number 1

1  Editorial advisory board

2  Crowd science and engineering: concept and research framework
Yueting Chai, Chunyan Miao, Baowen Sun, Yongqing Zheng and Qingzhong Li

9  A cyber-anima-based model of material conscious information network
Jianping Shen, Yadong Huang and Yueting Chai

26 Experimental results on large-scale cyber-physical hybrid discussion support
Takayuki Ito, Takanobu Otsuka, Satoshi Kawase, Akihisa Sengoku, Shun Shiramatsu, Takanori Ito,
Eizo Hideshima, Tokuro Matsuo, Tetsuya Oishi, Rieko Fujita, Naoki Fukuta and Katsuhide Fujita

39 Crowd wisdom drives intelligent manufacturing
Jiaqi Lu, Shijun Liu, Lizhen Cui, Li Pan and Lei Wu

48 Application of keyword extraction on MOOC resources
Zhuoxuan Jiang, Chunyan Miao and Xiaoming Li

71 Motivation mechanism of gamification in crowdsourcing projects
Zhiyuan Zeng, Jian Tang and Tianmei Wang

83 Computerizing Trail Making Test for long-term cognitive self-assessment
Zhiwei Zeng, Chunyan Miao, Cyril Leung and Zhiqi Shen