Meta measurement of intelligence with crowd network

Zheming Yang

*Institute of Computing Technology, Chinese Academy of Sciences,
University of Chinese Academy of Sciences, Beijing, China, and*

Wen Ji

*Beijing Key Laboratory of Mobile Computing and Pervasive Device,
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China*

Abstract

Purpose – The multiple factors of intelligence measurement are critical in intelligent science. The intelligence measurement is typically built as a model based on multiple factors. The different agent is generally difficult to measure because of the uncertainty between multiple factors. The purpose of this paper is to solve the problem of uncertainty between multiple factors and propose an effective method for universal intelligence measurement for the different agents.

Design/methodology/approach – In this paper, the authors propose a universal intelligence measurement method based on meta-analysis for crowd network. First, the authors get study data through keywords in the database and delete the low-quality data. Second, they compute the effect value by odds ratio, relative risk and risk difference. Then, they test the homogeneity by Q-test and analyze the bias by funnel plots. Third, they select the fixed effect and random effect as a statistical model. Finally, through the meta-analysis of time, complexity and reward, the weight of each factor in the intelligence measurement is obtained and then the meta measurement model is constructed.

Findings – This paper studies the relationship among time, complexity and reward through meta-analysis and effectively combines the measurement of heterogeneous agents such as human, machine, enterprise, government and institution.

Originality/value – This paper provides a universal intelligence measurement model for crowd network. And it can provide a theoretical basis for the research of crowd science.

Keywords Citizen science, Models and methods for crowd science and engineering, Crowd network, Meta-analysis, Intelligence, Measurement

Paper type Research paper

1. Introduction

1.1 Background

With the rapid development of crowd science, the crowd network gradually enters people's vision. Crowd network is the relationship network of many intelligent agents, their relatives,

International Journal of Crowd Science
Vol. 4 No. 3, 2020
pp. 295-307
Emerald Publishing Limited
2398-7294
DOI 10.1108/IJCS-03-2020-0008

friends, business and government affairs. Different from swarm intelligence, crowd network can support more interactive modes of different depth and breadth between intelligent individuals. The crowd network is in a three-dimensional superposition space of deep integration of physical space, consciousness space and information space. It conforms to the interaction and influence of matter, information and consciousness of different laws of motion, which makes the behavior of the agent show a broader uniform and opposite characteristics, as shown in Figure 1. Therefore, the intelligent measurement of crowd network is very important. The applications of intelligence measurement become more and more extensive. For example, the authors Gignac and Bates (2017) found that the intelligence measurement can moderate the effect between brain volume and intelligence. The authors in Vamsi and Bose (2018) measure the business intelligence by adopting the IT-based performance measurement systems (PMS) to evaluate the performance of the organization. And the authors Kahraman *et al.* (2018) measure the collective intelligence to evaluate the performance in energy systems. Therefore, intelligence measurement has brought many changes to our lives.

Now intelligence measurement methods can be divided into human intelligence quotient (IQ) test, machine intelligence measurement and universal intelligence measurement. In IQ test, it measures individual intelligence mainly through their perception and understanding of knowledge, words and graphics. At present, the two mainstream IQ tests in the world are the Binet-Simon intelligence scale and the Wechsler intelligence scale. They both measure people's intelligence by answering many questions. The work by McGrath (2011) defined standard score as a type of normally distributed standard score (with a mean of 100 and a standard deviation of 15) that represented level of performance on tests of cognitive ability. In the machine intelligence measurement, it is mainly based on Turing test. Turing (1950) adopted the mode of "question" and "answer" in 1950, that is, the observer talks to two test subjects by controlling the typewriter, one is a person, the other is a machine. He measures the intelligence of the machine by the questions that the observer constantly raises. Cochrane (2010) assumed an entropic measure able to account for the reduction or increase in the system information or state change, before and after the application of intelligence. Then he defined the machine intelligence as the change in entropy. The work by Legg and Hutter (2006) takes many well-known informal definitions of human intelligence that have been given by experts and extracts their essential features. These are then mathematically formalized to produce a general measure of intelligence for arbitrary machines. The authors Bien *et al.* (2002) analyzed those engineering systems or products that are said to be intelligent and have extracted four common constructs. Then they adopted the Sugeno fuzzy integral and the Choquet fuzzy integral to find a number called machine intelligence quotient. In the universal intelligence measurement, C-test was proposed in 2000
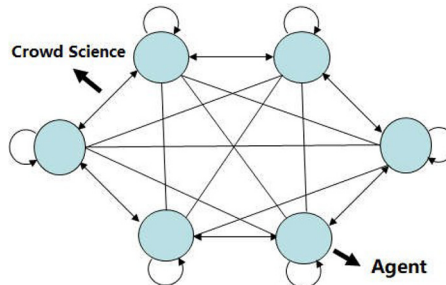

Figure 1.
Agent of meta-analysis

(Hernandez-Orallo, 2000), which can calculate many useful test problems. And these questions have been proved to be related to real IQ test score (Legg and Hutter, 2007; Insacabrera *et al.*, 2011). On the basis of Kolmogorov's complexity, C-test, and compression enhanced Turing test. The authors in Hernandez-Orallo and Dowe (2013) proposed a universal idea of intelligence measurement in anywhere and anytime and defined the universal. Then a new measurement of intelligence for general reinforcement learning agents is proposed (Gavane, 2013). And it is based on the notion that an agent's environment can change at any step of execution of the agent. And the resulting intelligence measurement is more general than the universal intelligence measurement (Legg and Hutter, 2007) and the anytime universal intelligence test (Hernandez-Orallo and Dowe, 2013). The work by Mesiar *et al.* (2006) proposed the concepts of generated universal fuzzy measures and basic generated universal fuzzy measures and discussed the special classes and properties of generated universal fuzzy measures.

However, according to the results of these papers, all the proposed methods have some drawbacks. Although they draw into environmental complexity and time, they do not consider the relationship between the multiple factors. So, they cannot combine the different agents. Therefore, the multiple factors and heterogeneity are the main difficulties in universal intelligence measurement. We propose a universal intelligence measurement method based on meta-analysis to solve the problem. Meta-analysis originates from statistics. It is a statistical method that integrates multiple research data. It can conduct a unified integrated analysis of the existing conclusions and objectively evaluate the existing research data to draw more valuable conclusions. At present, meta-analysis has been widely used in medical field (Gavin *et al.*, 2018; Lundh *et al.*, 2018), social science field (Braga *et al.*, 2017; Azucar *et al.*, 2018) and library information science field (Saxton, 2006; Ke and Cheng, 2015). Myszkowski *et al.* (2018) analyzed the relationship between intelligence and visual measurement by meta-analysis. On the basis of these, we solve the problem of multiple factors. And it combines human, machine, company, government and institution at the same time, as shown in Figure 2.

In this paper, we propose a universal intelligence measurement method based on meta-analysis for crowd network. In this paper, we make the following contributions:

- We consider the relationship between the multiple factors by meta-analysis.
- Our method solved the heterogeneity by studying many different data. And it can combine the different agents, especially for the intelligence of human, machine, company, government and institution.
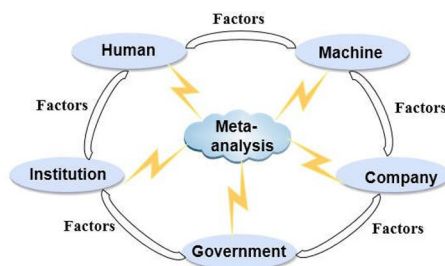- We first apply meta-analysis for intelligent science. It provides a great idea for other scholars.

Figure 2.
Agent of meta-analysis

The rest of the paper is organized as follows. In Section 2, we introduce the construction of the data set. In Section 3, we introduce meta-analysis and the method of merger effect value, Q-test and bias analysis. Meta measurement model is proposed in Section 4. The experimental results are provided in Section 5. Finally, the conclusion and future work are presented in Section 6.

2. Construction of data set

2.1 Acquisition of data
Retrieving data of meta-analysis is different from the traditional retrieval method. It should retrieve as much research data as possible related to intelligent measurement. It is necessary to provide a large number of keywords and a retrieval database for meta-analysis. Then we retrieve the keywords in the database to get data set. By researching the current academic progress of intelligence measurement, we determined the keywords and database. Keywords were intelligence, measurement, universal, increment, crowd, level, digital, physical, crowd network, entropy, machine and artificial. Database was Google Scholar. Finally, we get a total of 42 papers that cover all fields related to intelligence measurement.

2.2 Data filter
There may be some low-quality data in the data set. Therefore, we established a data-filtering standard to delete low-quality data. The data-filtering standard is dependent on the research subject and the research data. In this paper, we determine the data-filtering standard as follows: if the data title contains any one of the keyword "Intelligence" or "Measurement," we regard it as high-quality data. And if it does not contain the abovementioned two keywords, but contains more than two other arbitrary keywords, we also regard it as high-quality data. Besides, the filtering standard is not fixed and can be adjusted according to the actual situation. For example, it can also be regarded as high-quality data as long as the research is highly relevant to the title. In addition, they are all low-quality data. Finally, to ensure the reliability of the result, we selected eight papers as the data set of meta-analysis, as shown in Figure 3.
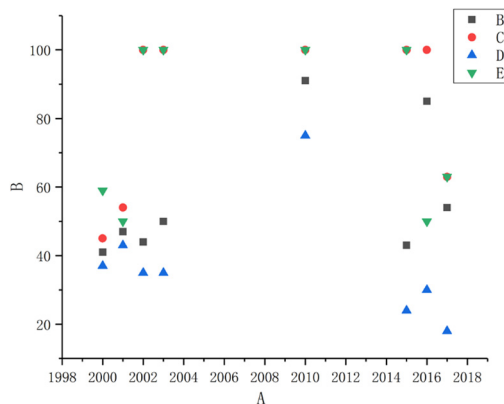

Figure 3.
Distribution of
data set

2.3 Coding of data
We encode the data set for the statistical analysis. The encoding format is as follows:
Number-Author-Time. As shown in Table 1. Then the coded papers are put into the data set
in turn. Besides, the size of the number only represents the order of coding.
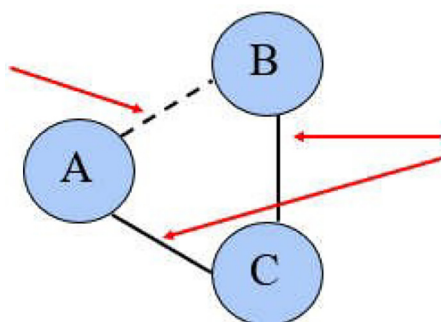
3. Meta-analysis
Meta-analysis is a statistical method that integrates multiple research data. As far as its
application is concerned, it is a new method of literature review. As shown in Figure 4, it can
be seen that both A and B have a direct relationship with C. There is no direct relationship
between A and B, but the relationship between A and B can be indirectly known through C.
Meta-analysis focuses on this indirect evidence mainly through statistical methods. It can
carry out a unified integration analysis for the existing conclusions and objectively evaluate
the existing research data to draw more valuable conclusions.

Effect value is one of the most important factors in meta-analysis. Meta-analysis needs to
turn multiple results into a unified statistical factor of effect value because they are
heterogeneous. To solve the problem that the coefficients of factors are different, we select
some statistical variables according to the particularity of intelligence measurement. In this
paper, we select the odds ratio (OR), relative risk (RR) and risk difference (RD) as effect
values. The homogeneity test is to test the rationality of merging results in data sets. It is
mainly to check whether the results of every data can be merged or not. In this paper, we use
the Q-test to test the homogeneity. The Q-test obeys the chi-square distribution with degree
of freedom $k-1$, where k is the number of effect values. If Q is statistically significant,
means the effects values are heterogeneous distributions. We should adopt random effect
model because it can consider the variation between studies and estimate the average of

No.	Author	Time
1	John Duncan	2000
2	Hee-Jun Park	2001
3	Zeungnam Bien	2002
4	Jacob W. Crandall	2003
5	José Hernández-Orallo	2010
6	Hao Zhong	2015
7	Jose Hernandez-Orallo	2016
8	Monireh Dabaghchian	2017

Table 1.
Coded data

Figure 4.
Meta-analysis

effects distribution at the same time. Then it avoids underestimating the weight of small samples or overestimating the weight of large samples. It can also get a larger confidence interval and then obtain a better conclusion. If Q is not statistically significant, the results of fixed effect model and random effect model are similar. But if the statistical factor of the Q-test is near the critical value, two models should be used simultaneously. Finally, we compare the difference in parameter estimation. In this paper, we select the method of funnel plotting to analyze the bias. The bias analysis is mainly the accuracy of each effect value increase with the sample size. We take the effect value as abscissa and the standard error as ordinate to plot. If there is no bias, it should be an inverted funnel. And the points on the funnel plot are symmetrically dispersed around the real value of the point estimate of the effect value. The standard errors of small samples are large and scattered at the bottom of the funnel plot. With the increase of sample size, the accuracy is also increased and the scatter points are more concentrated. On the contrary, there are bias problems.

4. Meta measurement model

Time, complexity and reward are three factors that have a great impact on the intelligence level of the agent, so this paper chooses these three main factors for analysis and modeling, as shown in Figure 5. In the traditional measurement methods, although these three factors are considered at the same time, they are all treated equally and there is no comparison of their influence and the relationship between these three factors, which will seriously affect the accuracy of the final measurement results. Therefore, we can get the weight of time, complexity and reward by meta-analysis.

According to Hernandez-Orallo and Dowe (2013), the reward of the agent is defined as follows:

$$V_\mu^\pi = E\left(\sum_{i=1}^\infty r_i^{\mu,\pi}\right) \tag{1}$$

The complexity is then defined as follows:

$$Kt_U(x) = \min\{l(p) + \log a \cdot time(U,p,x)\} \tag{2}$$

where $U(p) = x$, $l(p)$ represents the bit length of p and $U(p)$ represents the result of executing p on U. Time (U, p, x) is the time when U executes p to generate x. The relevance of U selection depends on the size of x. As any machine can simulate another machine, there is a constant $c(U, V)$ for every two machines U and V, which only depends on U and V and does not depend on x.
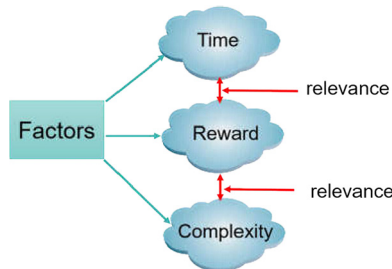
Based on the calculation methods and the results of meta-analysis of time, complexity and reward, this paper constructs a universal intelligence measurement model as follows:

$$I(\pi, U) = \sum_{\mu=i}^{\infty} b \cdot Kt_U(x) \cdot c \cdot V_\mu^\pi$$
$$= \sum_{\mu=i}^{\infty} b \cdot \left(l(p) + \log a \cdot time(U, p, x)\right) \cdot c \cdot E\left(\sum_{i=1}^{\infty} r_i^{\mu, \pi}\right)$$

(3)

where μ is any environment encoded on the universal machine U and π is the agent to be evaluated. In this paper, probabilities are assigned to each environment by $p(\mu)$, although these probabilities will not increase to 1. Among them, a, b and c are the weights of time, complexity and reward, As shown in Figure 6.

5. Experiment
The experimental environment of this paper is completed under the RevMan 5.3. After importing the data set, we select the binary variables as data types and select Mantel–Haenszel as analysis methods. And we select OR, RR and RD as effect values and select fixed effect and random effects as statistical models. Finally, we analyze their advantages and disadvantages. Figures 3, 4, 5, 6, 7 and 8 are the results of the experiment. The center of the rectangle represents the point estimate of effect value. The length of the rectangle represents the confidence intervals of effect value. And the larger the confidence interval of the effect value, the less accurate the result is.
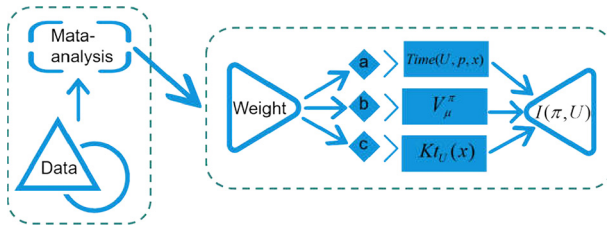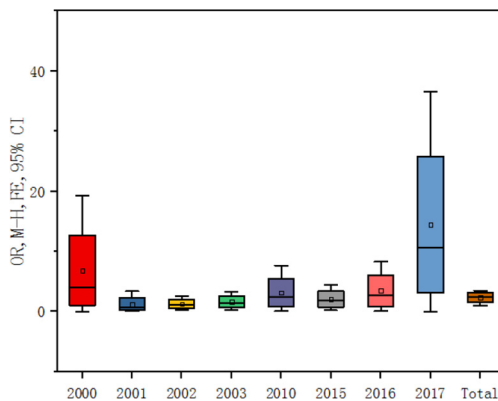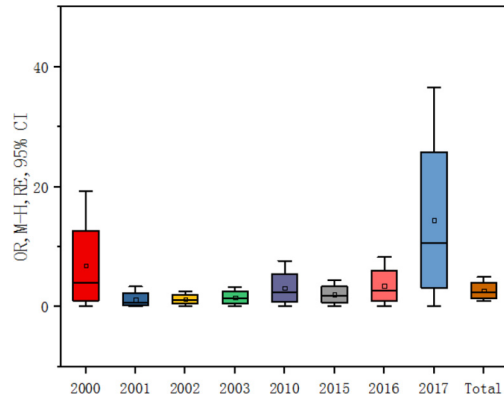

Figure 6.
Meta measurement model

Figure 7.
Box-plot of OR-Fixed
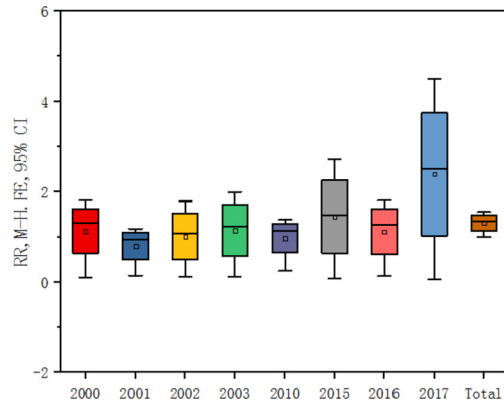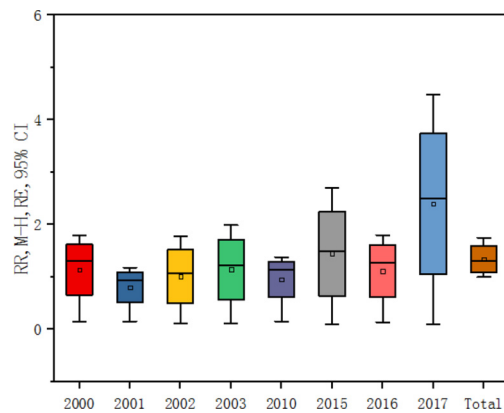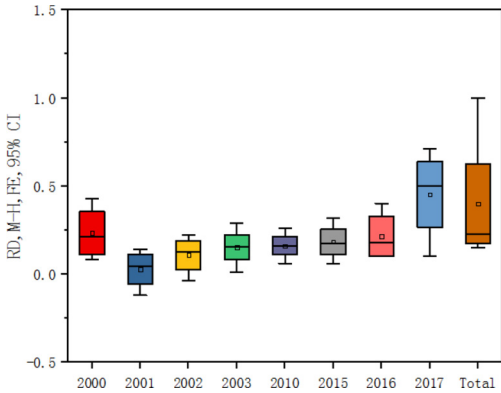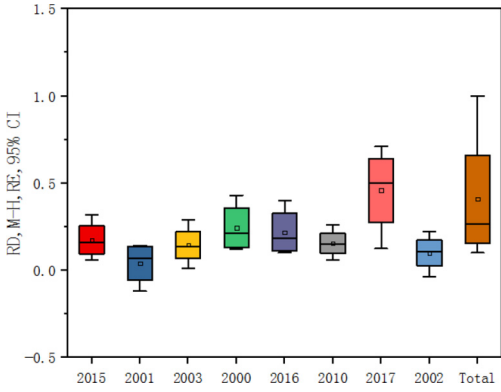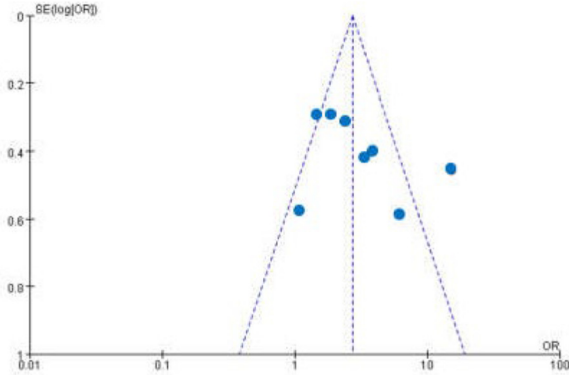
Figure 8.
Box-plot of OR-
Random

Figure 9.
Box-plot of RR-Fixed
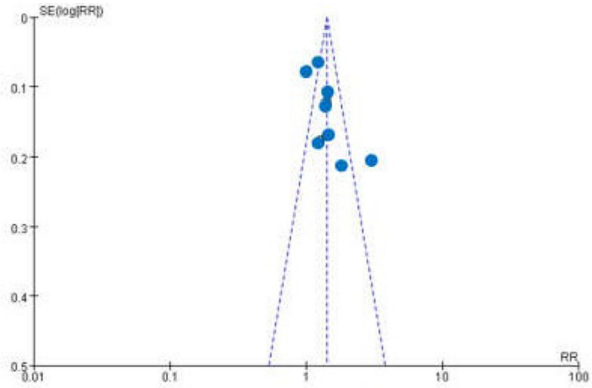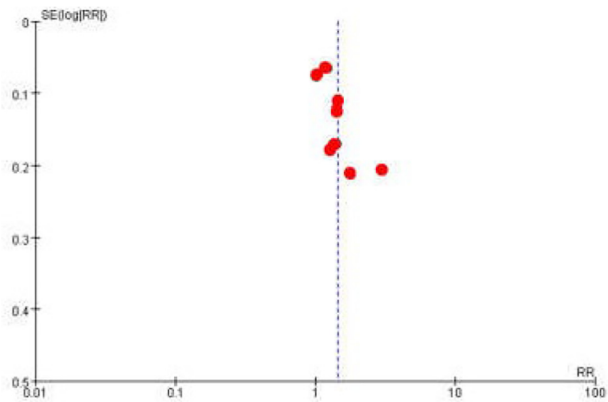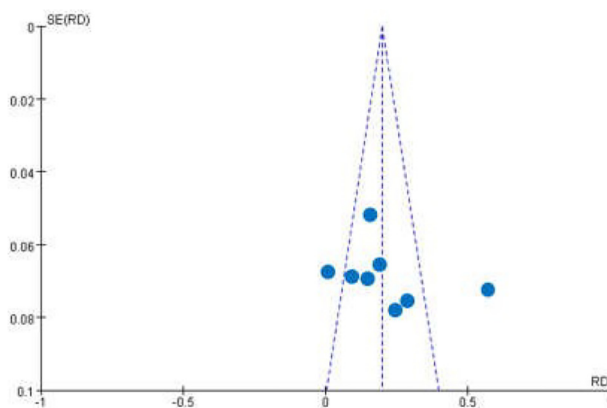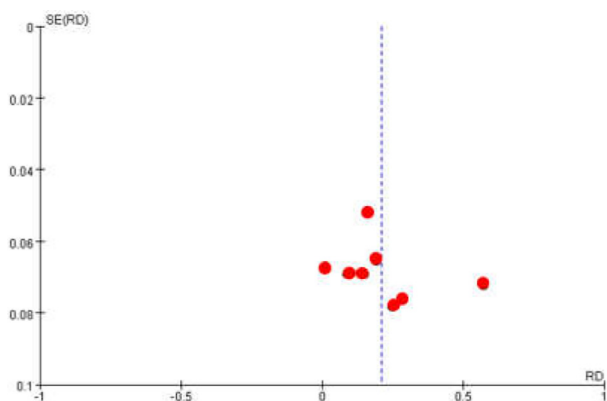
Figure 10.
Box-plot of RR-
Random

The experimental results show that the confidence intervals of the effect values in Figures 7, 8, 9 and 10 are larger, so the results are not accurate. The confidence intervals of the effect values in Figures 11 and 12 are smaller. Therefore, RR is more suitable for this study than other methods. And the confidence interval of the total effect value in Figure 11 is smaller than that in Figure 12. Overall, the confidence intervals of the total effect values in Figures 7, 9 and 11 are smaller than those in Figures 8, 10 and 12. It shows that the fixed effect statistical model is better than random effect.

Figures 13, 14, 15, 16, 17, and 18 are funnel plots of the experiment. It is mainly used for bias analysis. The abscissa is the effect value of the data set and the ordinate is the standard error of the data set. The smaller the sample size, the more dispersed the distribution is. And the larger the sample size, the more concentrated the distribution is. If there is no bias, it will be symmetrical funnel-shaped. On the contrary, if its symmetry is poor, there is bias.

The experimental results show that the distributions of Figures 17 and 18 are more concentrated. And their symmetry is better than others. It shows that the RD is significantly better than other methods. Overall, the symmetry, centralization and standard errors of Figures 13, 15 and 17 are similar to those of Figures 14, 16 and 18. It shows that the fixed

effect statistical model is similar to the random effect. The analysis of the abovementioned experiments shows that RD and fixed effect are better methods for meta-analysis. And the experimental results under these methods have no bias, which proves the correctness of the model.

6. Conclusion

In this paper, we analyze the existing research data of universal intelligence measurement by meta-analysis. The experimental results have no bias and show that the proposed method is effective. It can effectively combine different agents. And it provides a good research idea for the measurement of agents such as human, machine, company, government and institution. The research results of this paper can also promote the development of intelligence science. But this paper has some shortcomings. Because the research quality of different research data are different and all of them are treated equally in meta-analysis. So, there are some deviations in statistical analysis. We hope that a quantitative standard for research data can be proposed in future studies. This is also our future research focus.

References

Azucar, D., Marengo, D. and Settanni, M. (2018), "Predicting the big 5 personality traits from digital footprints on social media: a meta-analysis", *Personality and Individual Differences*, Vol. 124, pp. 150-159.

Braga, T., Gonçalves, L.C., Basto-Pereira, M. and Maia, A. (2017), "Unraveling the link between maltreatment and juvenile antisocial behavior: a meta-analysis of prospective longitudinal studies", *Aggression and Violent Behavior*, Vol. 33, pp. 37-50.

Bien, Z., Bang, W.C., Kim, D.Y. and Han, J.S. (2002), "Machine intelligence quotient: its measurements and applications", *Fuzzy Sets and Systems*, Vol. 127 No. 1, pp. 3-16.

Cochrane, P. (2010), "A measure of machine intelligence (point of view)", *Proceedings of the IEEE*, Vol. 98 No. 9, pp. 1543-1545.

Gavane, V. (2013), "A measure of real-time intelligence", *Journal of Artificial General Intelligence*, Vol. 4 No. 1, pp. 31-48.

Gignac, G.E. and Bates, T.C. (2017), "Brain volume and intelligence: the moderating role of intelligence measurement quality", *Intelligence*, Vol. 64, pp. 18-29.

Gavin, A., Pim, C. and Craske, M.G. (2018), "Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis", *Plos One*, Vol. 5 No. 10, p. e13196.

Insacabrera, J. Dowe, D.L. and Sergio, E. (2011), "Comparing humans and AI agents", *Artificial General Intelligence*, pp. 122-132.

Hernandez-Orallo, J. (2000), "Beyond the Turing test", *Journal of Logic, Language and Information*, Vol. 9 No. 4, pp. 447-466.

Hernandez-Orallo, J. and Dowe, D.L. (2013), "Measuring universal intelligence: towards an anytime intelligence test", *Artificial Intelligence*, Vol. 174 No. 18, pp. 1508-1539.

Kahraman, C. Onar, S.Ç. and Oztaysi, B. (2018), "Fuzzy collective intelligence for performance measurement in energy systems", *Energy Management Collective and Computational Intelligence with Theory and Applications*, pp. 497-517.

Ke, Q. and Cheng, Y. (2015), "Applications of meta-analysis to library and information science research: content analysis", *Library and Information Science Research*, Vol. 37 No. 4, pp. 370-382.

Legg, S. and Hutter, M. (2006), "A formal measure of machine intelligence", arXiv preprint cs/0605024.

Legg, S. and Hutter, M. (2007), "Universal intelligence: a definition of machine intelligence", *Minds and Machines*, Vol. 17 No. 4, pp. 391-444.

Lundh, A., Lexchin, J., Mintzes, B., Schroll, J.B. and Bero, L. (2018), "Industry sponsorship and research outcome: systematic review with meta-analysis", *Intensive Care Medicine*, Vol. 44 No. 10, pp. 1603-1612.

McGrath, M.C. (2011), "Deviation IQ".

Myszkowski, N., Celik, P. and Storme, M. (2018), "A meta-analysis of the relationship between intelligence and visual 'taste' measures", *Psychology of Aesthetics Creativity and the Arts*, Vol. 12 No. 1, p. 24.

Mesiar, R., Mesiarová, A. and Valášková, L.U. (2006), "Generated universal fuzzy measures", *International Conference on Modeling Decisions for Artificial Intelligence*, Springer, Berlin, Heidelberg, pp. 191-202.

Saxton, M.L. (2006), "Meta-analysis in library and information science: method, history, and recommendations for reporting research", *Library Trends*, Vol. 55 No. 1, pp. 158-170.

Turing, A.M. (1950), "Computing machinery and intelligence", *Mind (New Series)*, Vol. 59 No. 236, pp. 433-460.

Vamsi, V. and Bose, I. (2018), "Business intelligence for performance measurement: a case based analysis", *Decision Support Systems*, Vol. 111, pp. 72-85.

Further reading

Hajovsky, D.B. (2014), "Deviation IQ", *Encyclopedia of Special Education*, John Wiley & Sons.

Prpic, J. and Shukla, P. (2016), "Crowd science: measurements, models, and methods", *HI International Conference on System Sciences*, IEEE.

Corresponding author

Wen Ji can be contacted at: jiwen@ict.ac.cn