IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.12 ⁑ Vol. 10, Issue 1, January 2023

DOI: 10.17148/IARJSET.2023.10114

A Novel Framework for Effective Information Data Mining in Big Data Domain by Using Machine Learning Techniques

Dr. Bodla Kishor[1], E Mounika Reddy[2], Pirangi Hymavathi[3]

Assistant Professor, Department of CSE, CMR Engineering College, Hyderabad[1]

Assistant professor, Department of CSE (AIML), Sri Indu College of Engineering and Technology, Hyderabad[2]

Assistant Professor, CSE Department, Sri Indu College of Engineering and Technology, Hyderabad[3]

Abstract: Traditionally, education assets are shared insufficiently, and up to date slowly; the education records aren't applied adequately. What is worse, the traditional information filtering technique can't successfully mine preferred information, if the huge records has a heavy noise. This article offers an information mining technique from education huge records, on the premise of aid vector machine (SVM), and cleans the sampled odd records via records integration and conversion. Besides, the authors supplied a technique that mechanically builds education expertise image. Based at the filtered and mined education records, a neural community turned into designed to retrieve the subject matters of lecture room expertise, and the education correlations among those notions had been diagnosed from the assessment records via way of means of opportunity correlation rules. The outcomes display our technique carried out fantastic outcomes on coaching belief retrieval and education correlation recognition.

Keywords: Knowledge mining, neural network, knowledge image, education big data.

I. INTRODUCTION

The usage of learning management systems in education has been increasing in the last few years. Students have started using mobile phones, primarily smart phones that have become a part of their daily life, to access online content. Student's online activities generate enormous amount of unused data that are wasted as traditional learning analytics are not capable of processing them. This has resulted in the penetration of Big Data technologies and tools into education, to process the large amount of data involved. This study looks into the recent applications of Big Data technologies in education and presents a review of literature available on Educational Data Mining and Learning Analytics.

In the 21st century, social progress is mainly driven by the Internet and education. Thanks to the rich education resources on the Internet, online learning and education have been integrated with various education notions into various education models, namely, massive open online course (MOOC) and computer supported collaborative learning (CSCL), forming numerous knowledge images [1]. In the field of education, knowledge images are often adopted in course teaching. Many popular MOOCs platforms apply knowledge images visualize notions and recommend education resources. These knowledge images are usually prepared artificially by field experts [4]. However, the artificial preparation consumes too much time, and cannot be extended to many notions and correlations. Due to the explosive growth of courses and themes on MOOC platforms, it is extremely hard to artificially plot a knowledge image for each new course [5].

Besides, artificial preparation has another huge problem: Teaching research shows that every expert has his/her blind spot, that is, the same notion could be perceived differently by experts and learners. Learning that initially started in the class room was based on three models namely behavioral, cognitive and constructivist models [2] The behavioral models rely on observable changes in the behavior of the student to assess the learning outcome. The cognitive models are based on the active involvement of teacher in the learning which helps in guided learning. In the constructivist models, the students have to learn on their own from the knowledge available to them. Siemens (2004) [4] proposed a new model termed “Connectivism” which was characterized as the “amplification of learning, knowledge and understanding through the extension of personal network”. According to this model, learning is no longer an internal activity [5] Connectivism proposed learning in a network of nodes which improved the learning experience of students and reduced the need for the direct involvement of an instructor. Since then, traditional learning environments have gradually mutated into community-based learning environments.

DOI: 10.17148/IARJSET.2023.10114

Relying on particle filters to screen and mine information, this traditional information filtering method sets a high requirement on the initial particle trajectory. Thus, this paper firstly puts forward a data mining algorithm from education big data based on support vector machine (SVM), providing good data support to the establishment of knowledge images.

Considering the growing demand and limited creation method of knowledge images in the education field, this paper also designs a system that automatically builds knowledge images fit for the teaching of school courses and online courses [8]. In education knowledge images, the expected nodes represent the teaching notions of a subject or course; the education notions refer to the basic notions (e.g., the physical notion of acceleration and geographical notion of landmass formation) that must be fully understood and understood by learners, the data need to be retrieved from the education field and new entity markers (e.g., people, locations, and organizations), instead of traditional markers.

II. RELATED WORK

Therefore, this paper specially chooses to identify education correlations from the data on learning evaluations and activities [11]. This paper proposes an SVM-based data filtering algorithm from education big data. The algorithm cleans the abnormal entries in the collected data, and integrates and transforms the abnormal data, providing good data support to the creation of knowledge images.

The authors put forward a novel and practical automatic generation system of education knowledge images. This system can retrieve teaching notions and identify important education correlations from heterogeneous data, which usually include teaching data and learning evaluation data.

Given the education purpose of the teaching notions, the authors proposed to apply neural network to the teaching data, such as to complete the retrieval of classroom notions.

Possibility correlation rules were mined from the notion-based learner evaluation data to derive the required correlations.

Knowledge image is essentially an atlas of semantic network and related knowledge. Through the new round of technology change, a series of well-known knowledge images have emerged, such as Google, DBpedia, NELL, SSCO, Baidu, etc. These knowledge images intuitively present the information about knowledge background and development history of various subjects. Early knowledge images are mainly used in scientific research [12]. Knowledge image as a teaching method takes shape with the continuous development of online information. For example, Zhang et al. [16] produced a tutorial video on an outsourcing website, teaching complete mathematical notions to ordinary people. Wang et al. [17] introduced notional graphics into the teaching process.

Based on the learning notion map of education data, Chaplot et al. [7] established a directed notion graph of the correlation degrees between given courses, and applied the graph into teaching to reveal the implicit correlations between courses. Through graph analysis, Chen et al. [18] discussed the scientific learning views of learners, predicted the learning situation of learners by the activities, emotions, and attitudes of learner portraits, and proposed a new automatic learning method to qualify the contribution to the knowledge base. Nickel et al. [9] developed a textbook-based notional graph, which is superior to supervised learning baseline. Based on the retrieved entity correlations, Xie et al. [2] built up an education knowledge image, and constructed a visualization analysis platform called EduVis. With the aid of graph mining, Bordes et al. [2] modeled the network graphs of different types of learners, lecturers, and subjects in public education, providing a good reference for similar research.

Camacho et al. [4] explained how to design a knowledge management system in a rural low knowledge school in Costa Rica, and how the participatory approach used in the design process creates a learning culture that encourages the use of knowledge management systems (KMSes). This paper contributes to the methodology of KMSes, and details three novel methods that enable KMSes developers to handle he social and technical dimensions of KMSes. Bouton et al. [3] described that, although social network technology (SNT) helps to build knowledge collaboratively, recent studies in secondary schools indicate that students mainly use these tools to learn about knowledge sharing of related artifacts. Here, this discovery is extended to higher education, and two surveys are reported, which respectively summarize the SNT learning features of college students in an undergraduate program (N D 264) and a normal college (N D 449).

Despite their achievements, the previous studies have not effectively extracted the key information of education data (e.g., the knowledge points of each class hour), summarized after-class learning factors (e.g., personality, and diligence), or constructed the logic relationship between different knowledge points. To solve these problems, this paper introduces the possibility correlation rule mining to recognize the education correlations, establishes knowledge images, and experimentally verifies the effectiveness of such images.

III. DATA MINING ALGORITHM

This paper mines data with the help of the SVM algorithm [6]. First, the abnormal data were cleaned from the education big data. Then, the clean data were converted and classified, followed by integration, regularization, and transform. After that, data mining rules were set up to filter the uncorrelated information, and mine the effective information from the data, thereby enhancing the education efficiency and performance of the neural network [9]. The data gathering tree was adopted for data conversion and classification. The tree can discover the correlations and dependence between data. To filter the information in the education dataset [7], the expectation of data X, exp SN(x), can be defined as the support to the correlation degree of the data gathering tree of the dataset. Then, the unbiased risk of the data gathering tree can be estimated as

$$\exp \mathrm{SN}(X) = \sum_{T_d X \wedge T_d \in D} P(X, T_d) \qquad (1)$$

$$\min_{0\le\alpha_i\le c} W = \frac{1}{2}\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad - \sum_{i=1}^{l} \alpha_i + b\left(\sum_{i=1}^{l} y_j \alpha\right) \qquad (2)$$

In each round, the SVM was adopted to mine key features of education big data, which dominates the data transmission between nodes on the knowledge image.

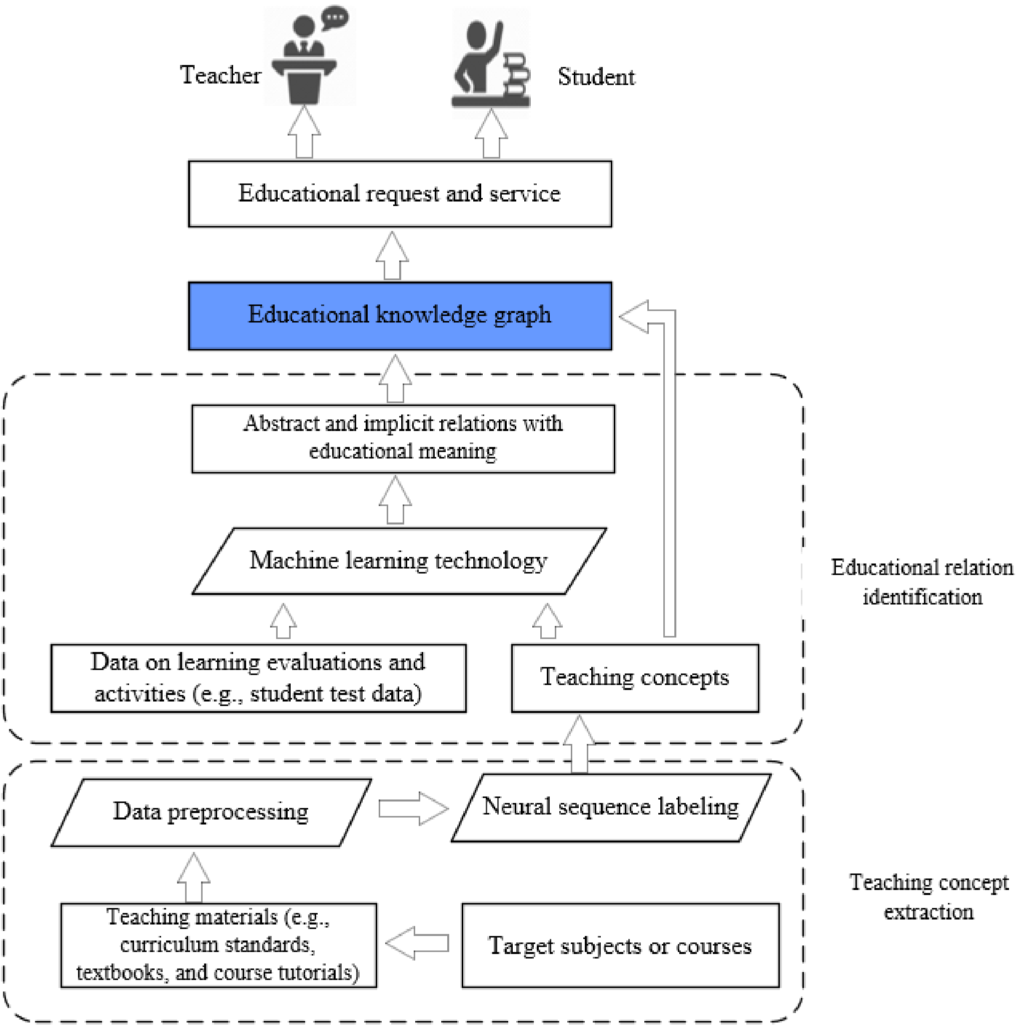

Teacher
Student
Educational request and service
Educational knowledge graph
Abstract and implicit relations with educational meaning
Machine learning technology
Data on learning evaluations and activities (e.g., student test data)
Teaching concepts
Educational relation identification
Data preprocessing
Neural sequence labeling
Teaching materials (e.g., curriculum standards, textbooks, and course tutorials)
Target subjects or courses
Teaching concept extraction

Fig 1. Proposed frame work

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

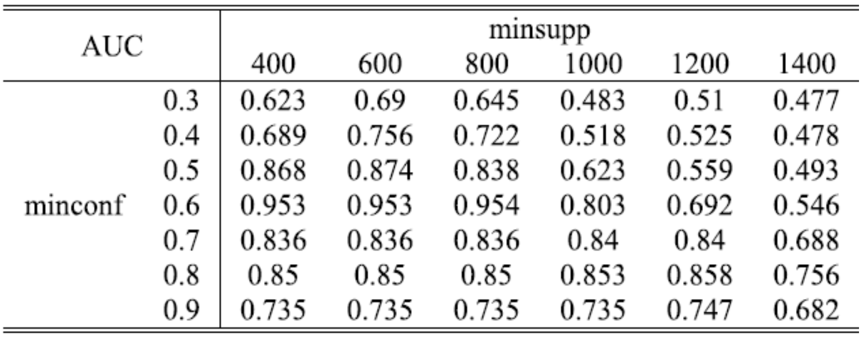
Impact Factor 7.12 ⁂ Vol. 10, Issue 1, January 2023

DOI: 10.17148/IARJSET.2023.10114

IV. PROPOSED WORK

Research in education has resulted in several new pedagogical improvements. Community based learning environments have increased in number. In the current learning environments, users learn in online communities like discussion forums, online chats, instant messaging clients and various Learning Management Systems like Moodle. Recent learning methods like Flipped Classroom [6] greatly depend on online activities. Several frameworks [7] and models have been proposed for online learning management systems to improve the learning experience. Entry of open source projects in mobile computing has led to low cost smartphones and smart phones have penetrated much. Students have started using smart phones to access learning content. As the learning environments have become accessible anywhere through the internet, students access their courses anywhere and indulge in learning activities. Students' activities through learning management systems create large amount of data that can be utilized in developing the learning environment, helping the students in learning and improving the overall learning experience.

Figure 1 is the block diagram of the proposed education data-oriented knowledge image system, which covers a teaching notion retrieval module (a) and an education correlation recognition module (b). Module (a) mainly obtains the teaching notions from teaching data collected from the education field [8]. Then, the instructive notions could be retrieved by methods like neural sequence labeling. Module (b) mainly recognizes the education correlations associated with teaching notions. The latest data mining techniques are embedded in this module, such as the mining of possibility correlation rules, due to the implicit and abstract nature of education correlations. In addition, the module uses the data on learning evaluations and activities, which can reflect the learners' cognition and knowledge acquisition process. The correlations mined by this module link up the teaching notions into the knowledge image required for education, providing support to the various requests and services of learners and lecturers.



AUC		minsupp					
		400	600	800	1000	1200	1400
minconf	0.3	0.623	0.69	0.645	0.483	0.51	0.477
	0.4	0.689	0.756	0.722	0.518	0.525	0.478
	0.5	0.868	0.874	0.838	0.623	0.559	0.493
	0.6	0.953	0.953	0.954	0.803	0.692	0.546
	0.7	0.836	0.836	0.836	0.84	0.84	0.688
	0.8	0.85	0.85	0.85	0.853	0.858	0.756
	0.9	0.735	0.735	0.735	0.735	0.747	0.682

Table 2. AUC Values of Different Pairs of Parameters.

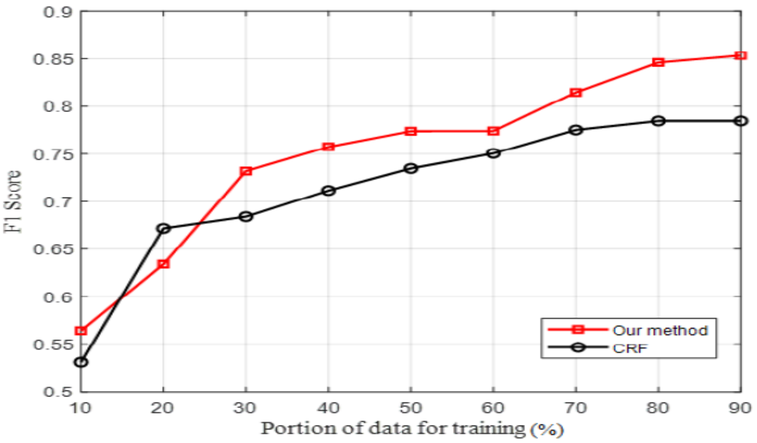
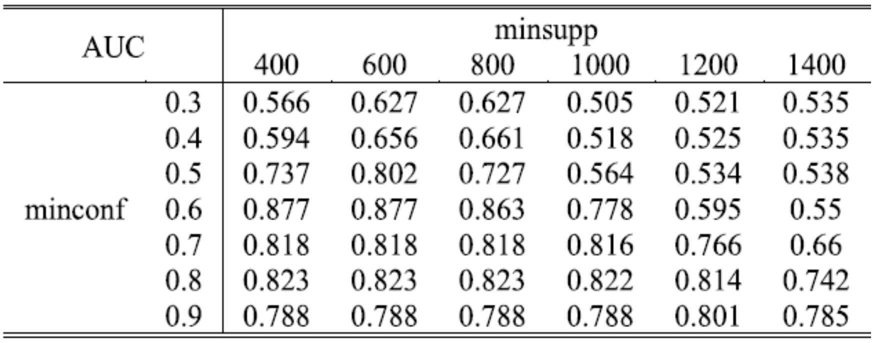


Fig 2. Beginning of notation



AUC		minsupp					
		400	600	800	1000	1200	1400
minconf	0.3	0.566	0.627	0.627	0.505	0.521	0.535
	0.4	0.594	0.656	0.661	0.518	0.525	0.535
	0.5	0.737	0.802	0.727	0.564	0.534	0.538
	0.6	0.877	0.877	0.863	0.778	0.595	0.55
	0.7	0.818	0.818	0.818	0.816	0.766	0.66
	0.8	0.823	0.823	0.823	0.822	0.814	0.742
	0.9	0.788	0.788	0.788	0.788	0.801	0.785

Table 3. Map Values of Different Pairs Of Parameters.

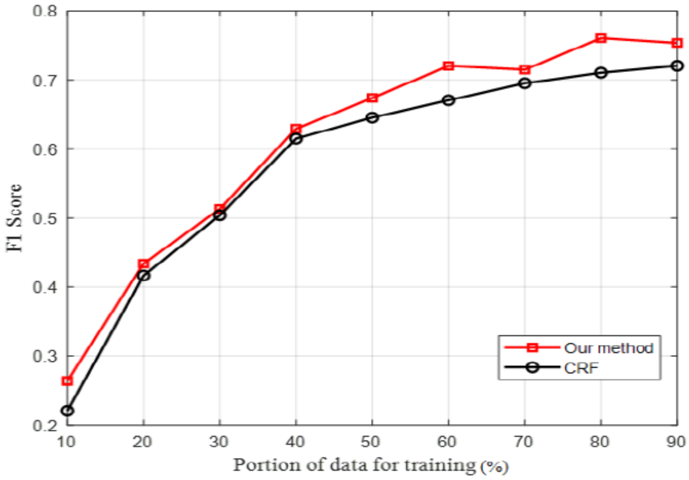


Fig 3. Inside of a notation

The possibility correlation rule mining is an extended approach to mine the correlation rules by processing the uncertainties in data. In a given possibility data, the support and confidence were treated as stochastic variables; the deterministic correlation rule Sj) Si was formulated as P(Sj) Si); if P(Sj) Si) is greater than the given threshold of minimum prob, then the rule is valid.

There is a prerequisite correlation from notion A to notion B called assertive correlation. The two notions are positively correlated in the absence of this correlation, and negatively in the presence of this correlation. Similar to traditional mining of education data, the scoring rate was adopted as the estimated possibility that a learner understands a notion.

Thus, correlation rule mining treats each notion as a project, and the estimated knowledge states of the 6,000 learners as 6,000 deals. For each candidate of every prerequisite condition, the system computes its positive possibility. The key parameters are minsupp and minconf. This paper takes the area under the curve (AUC) of receiver operating characteristic (ROC) and mean average precision (MAP) as the main metrics.

Tables II and III suggest that the AUC and MAP of the minconf of (0.6, 600) and minsupp of (0.6,800) were obviously higher than those of the other pairs of minsupp and minconf values. Figure 6 further presents the constructed knowledge image of math. Each circle represents a notion.

The solid and dotted arrows stand for prerequisite and containment correlations, respectively. In addition, the red circles are level 1 knowledge points, the green circles are level 2 knowledge points, and the yellow circles are level 3 knowledge points. Our strategy mines and identifies the deep correlations between knowledge points on different levels, revealing the potential learning and teaching paths for teachers and students.

DOI: 10.17148/IARJSET.2023.10114

V. CONCLUSION

This paper designs a neural network to retrieve teaching notions, and introduces possibility correlation rule mining to identify education correlations. Besides, an SVM-based data mining algorithm for education big data, which can rapidly process the abnormal entries in the collected data. During experimental evaluation, excellent performance was achieved by the constructed knowledge image. However, the following issues of this work need to be solved urgently, the non-uniform formats of education big data, and the massive number of knowledge notions, the lack of personalized plans for student end, the neglection of the causality between the knowledge notions in the construction of knowledge image. The future research will try to deal with the effectively extracting the key information from education data, further improving the personalized teaching services based on knowledge image for online diagnosis of learning disabilities, and intelligent recommendation of learning resources and analyzing the causality and logic among knowledge points of different subjects and grades in education.

REFERENCES

[1] Q. Guo, ``Detection of head raising rate of students in classroom based on head posture recognition," Traitement du Signal, vol. 37, no. 5, pp. 823-830, Nov. 2020.

[2] J. Whitehill and M. Seltzer, ``A crowdsourcing approach to collecting tutorial videos-toward personalized learning-at-scale," in Proc. 4th Annu. ACM Conf. Learn., 2017, pp. 157-160.

[3] R. D. Senthilkumar, ``Concept maps in teaching physics concepts applied to engineering education: An explorative study at the middle east college, sultanate of oman," in Proc. IEEE Global Eng. Educ. Conf. (EDUCON), Apr. 2017, pp. 263-270.

[4] Published a paper titled "Data Mining Challenges With Big Data" International Journal for Research in Applied Science and Engineering Technology (IJRASET)" with Impact Factor 1.241, ISSN: 2321-9653,Volume 3 Issue VI, June 2015.

[5] D. S. Chaplot, Y. Yang, J. Carbonell, and K. R. Koedinger, ``Data-driven automated induction of prerequisite structure graphs," in Proc. 9th Int. Conf. Educ. Data Mining, Raleigh, NC, USA, Jun. 2016, pp. 318-323.

[6] Ravindra Changala, "A Survey on Development of Pattern Evolving Model for Discovery of Patterns in Text Mining Using Data Mining Techniques" in Journal of Theoretical and Applied Information Technology, in 31st August 2017. Vol.95. No.16, ISSN: 1817-3195, pp.3974-3987.

[7] P. G. Wolf, J. Manero, K. B. Harold, M. Chojnacki, J. Kaczmarek, C. Liguori, and A. Arthur, ``Educational video intervention improves knowledge and self-efficiency in identifying malnutrition among healthcare providers in a cancer center: A pilot study," Supportive Care Cancer, vol. 28, no. 2, pp. 683-689, Feb. 2020.

[8] Published a paper titled "Evaluation and Analysis of Discovered Patterns Using Pattern Classification Methods in Text Mining" In ARPN Journal Of Engineering and Applied Sciences Volume 13, Issue 11, Pages 3706-3717 with ISSN:1819-6608 in June 2018.

[9] S. Wang, C. Liang, Z. Wu, K. Williams, B. Pursel, B. Brautigam, S. Saul, H. Williams, K. Bowen, and C. L. Giles, ``Concept hierarchy extraction from textbooks," in Proc. ACM Symp. Document Eng., Sep. 2015, pp. 147-156.

[10] Ravindra Changala, "Retrieval of Valid Information from Clustered and Distributed Databases" in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1,Pages 21-25, September 2016.ISSN: 2455-3506.

[11] M. Nickel, V. Tresp, and H. P. Kriegel, ``A three-way model for collective learning on multi-relational data," in Proc. ICML, vol. 11, 2011, pp. 809-816.

[12] T. Xie, C. Zhang, Z. Zhang, and K. Yang, ``Utilizing active sensor nodes in smart environments for optimal communication coverage," IEEE Access, vol. 7, pp. 1133811348, 2019.

[13] Y. Yang, H. Liu, J. Carbonell, and W. Ma, ``Concept graph learning from educational data," in Proc. 8th ACM Int. Conf. Web Search Data Mining. Shanghai, China: ACM, Feb. 2015, pp. 1059-1090.

[14] Z. Zhang, C. Zhang, M. Li, and T. Xie, ``Target positioning based on particle centroid drift in large-scale WSNs," IEEE Access, vol. 8, pp. 127709-127719, 2020.

[15] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, ``KnowEdu: A system to construct knowledge graph for education," IEEE Access, vol. 6, pp. 31553-31563, 2018.

[16] C. H. Tan, E. Agichtein, P. Ipeirotis, and E. Gabrilovich, ``Trust, but verify: Predicting contribution quality for knowledge base construction and curation," in Proc. 7th ACMInt. Conf.Web Search Data Mining. NewYork, NY, USA: ACM, Feb. 2014, pp. 553-562.