Scientific Map of Papers Related to Data Mining in Civilica Database Based on Co-Word Analysis

Fateme Shahrabi Farahani, Meysam Alavi, Mina Ghasemi, Babak Teimourpour*
Information Technology Management dept., Tarbiat Modares, Tehran, Iran
{fateme.shahrabi, meysamalavi, Minaghasemi, B.teimourpour}@modares.ac.ir

Received: 2020/08/08 Revised: 2020/09/11 Accepted: 2020/09/27

***Abstract*—Today, due to the large volume of data and the high speed of data production, it is practically impossible to analyze data using traditional methods. Meanwhile, data mining, as one of the most popular topics in the present century, has contributed to the advancement of science and technology in a number of areas. In the recent decade, researchers have made extensive use of data mining to analyze data. One of the most important issues for researchers in this field is to identify common mainstreams in the fields of data mining and to find active research fields in this area for future research. On the other hand, the analysis of social networks in recent years as a suitable tool to study the present and future relationships between the entities of a network structure has attracted the researcher's scrutiny. In this paper, using the method of co-occurrence analysis of words and analysis of social networks, the scientific structure and map of data mining issues in Iran based on papers indexed during the years 1388 to 1398 in the Civilica database is drawn, and the thematic trend governing research in this area has been reviewed. The results of the analysis show that in the category of data mining, concepts such as clustering, classification, decision tree, and neural network include the largest volume of applications such as data mining in medicine, fraud detection, and customer relationship management have had the greatest use of data mining techniques.**

Keywords*— *Data Mining; Scientific Map; Co-Word Analysis; Social Network Analysis.

1. INTRODUCTION

Today, due to the increasing volume and complexity of data, a suitable tool is necessary to have to analyze the existing data and gain knowledge within them. These tendencies to acquire knowledge hidden in data have led to significant growth in data mining [1]. Data mining, as one of the most popular topics in the present century, contributes to the advancement of science and technology in a significant number of fields and disciplines [2]. With the advancement of data mining, research has gradually moved away from statistical approaches having long been considered a standard method [3]. In recent years, data mining has made a tremendous impact on academic settings and has found many applications in various fields and in various sciences examples of which include its applications in business, management and deception detection, sports, text mining and web mining [1]. The data mining process consists of two main stages. The first stage is data preprocessing and the second stage is pattern recognition, which is the main task of data mining [4].

1-1. Co-Word

Due to the increasing volume of information and the increasing speed of science production, it is difficult to observe scientific trends and identify the knowledge hidden in the working texts. Therefore, today, different methods, including text mining, are used to discover the hidden layers of knowledge in different sciences. Researchers have been able to use these methods to discover hidden knowledge in scientific texts. While achieving such knowledge through traditional methods takes a long time and effort. One of these techniques, having been of great popularity in recent years for experts in various fields, is the co-occurrence analysis of words, introduced as a new indicator to the scientific community. This indicator is a suitable way to monitor the evolution of various sciences. The co-occurrence of words is a content analysis technique that expresses both the frequency of topics and the relationship between them [5]. The analysis of word co-occurrence, relying on high-frequency words, can identify the most important research topics in any scientific field; that is, the incidence of a word indicates its importance in the field of science. First used in 1980 in France at the Center for Sociological Innovation, this method was then used to draw a scientific map of research areas. Calvin was the first to describe this method in 1983 and used it in his research [6]. The word co-occurrence analysis can help researchers to identify key patterns and trends that lead to a particular change in a research topic and emerging or declining research topics of a field [7]. The method of co-word has been considered notably in recent years by the researchers of different countries. Many researchers have studied certain areas with this method to analyze and draw its structure. In this analysis, the co-occurrence of keywords in the title, abstract or the main text of papers is examined. Using the method and the co-word, it is possible to extract scientific topics and the relationship between them directly from the thematic content [8]. The word co-occurrence analysis is one of the scientometric methods that identify the conceptual structure of a field of research by analyzing the content of texts. The purpose of scientometrics is to reveal the characteristics of scientific phenomena and processes in scientific research for more effective management of science; therefore, with scientometric methods, science and in general different scientific fields can be better managed and the way can be paved for future research [9].

1-2. Scientific Map

The scientific map provides a new perspective for revealing scientific boundaries and its dynamic structure using illustrative methods [10]. The concept was coined in 1974 by Henry Small and recaptured in 1980 by Howard White[11]. In

the definition of scientific maps, it is believed that a scientific map is a representation of scientific fields that are prepared by quantitative analysis of bibliographic information. The constituent elements of scientific maps are the outputs of research fields. In these maps, the scientific fields that have a stronger conceptual connection are placed next to each other and the fields that have a weaker connection are located at a farther distance from each other [12].

1-3. Knowledge Visualization

Knowledge visualization is a computer-aided information processing technology that can reveal the visual appearance of data objects in scientific texts (such as authors, keywords), and the relationships between them. To realize the visualization of thinking structure of areas of knowledge, the relationships between objects are expressed in two-dimensional or three-dimensional aspect, which indicates the perspective of knowledge [13] and [14]. Visualizing can effectively strengthen human`s knowledge to understand the huge amount of information and general structure of a scientific field. Co-word analysis is a powerful technic to visualize scientific fields. [15] The findings of the research have shown that visualization through co-word analysis provides a clearer picture of that field because vocabulary analysis provides the researcher with the basic topics of that field in an overview. Giving a clear picture of the research taken and determining the connection of the various fields is one of the aims of the scientific map. Since scientific maps have a structure similar to that of social networks, for illustration, analysis and interpretation, social network analysis techniques are used [16].

1-4. Social Network Analysis

Social media analysis is a research tool that has received a lot of attention in recent years. [17] A social network consists of nodes that are connected by relationships. There are various indicators in the analysis of these networks that can be used in scientific maps. Network size, density, centrality are important indicators in the analysis of this type of network [16]. Social network analysis is an interdisciplinary subject between sociology, mathematics, and computer science that is used in various sciences such as sociology, economics, communication sciences, psychology, physics, and computer [18].

Many researchers have already tried to draw a scientific map of different fields of science to get acquainted with its hidden dimensions, which will be discussed in detail in the literature review section. A review of the background shows that due to the significant increase in research related to data mining and its techniques, it is necessary to analyze the process of this research and determine the structure of its scientific map. As no independent research has been done in Iran on the application of co- words in clarifying the structure and scientific map of the data mining field, the main task of the present study is to discover and determine a scientific map of data mining topics in Iran based on indexed papers in the Civilica database, in the 10-year period from 2009 to 2017. The main objectives of this research are to increase the knowledge of common grounds in the subfields of data mining and to find active research fields in this area for researchers, knowing the growth rate of using techniques and applications related to data mining and finding which areas have been worked on less and have the potential to work.

The general structure of this paper is as follows. The second part reviews the literature. In the third part, the research methodology will be described. The fourth section analyzes and evaluates the results. The fifth section deals with conclusions and future work.

2. LITERATURE REVIEW

In this section, the work done in the field of drawing scientific maps of various fields is reviewed.

Lin explored innovation in research through a corpus-based approach. According to his research, innovation and entrepreneurship are among the most frequently mentioned words in 1460 abstracts of the Journal of Business Research from 1973 to 2015[19]. In [20], there is found a study aimed at identifying the areas of financial marketing research and the relationship between them, analyzing papers in this area from 1961 to 2010 using the co-words. In another study, Lee and Su analyzed 223 highly cited papers in the field of electrically conductive nanocomposites. They combined co-words analysis and social network analysis to depict the scientific map and determined the emergent fields in this area [21]. Gan and Wang studied the characteristics and status of social media in China between 2006 and 2013 using the Co-Word Analysis method [22]. In [23] using co-word analysis to map the structure of the Internet of Things, the development between 2001 and 2014, focusing 758 papers from the WoS database was examined. In their study, Hu et al. examined library and information science in China between 2008 and 2012 using the Co-Words Analysis method. [24] In another study, Chang et al. used the co-words analysis to recognize the weakness of research on clinical competence of medical students on 588 papers. [25] Also, in [26], the structural process and conceptual evolution of research on genetically modified organisms using the scientific map method was examined during the 27-year period from 1990 to 2016. In another study based on papers in the period from 2015 to 2019, the concepts of business, innovation and related concepts were examined in the scientific map [27]. In another study, 400 papers from the WoS database in the field of epidemiology were examined, which analyzed issues related to epidemiology using co-words analysis [28]. Another study in [29] is a comprehensive review of academic research between 1975 and 1975 on the degradation of polymers after gamma irradiation based on papers published in the WoS database. Another study in the field of child-computer interaction has been conducted based on the review of 1059 papers from the ACM database related to the period 2003 to 2018 on the method of co-words analysis [30]. Of other research on the topic is environmental crisis management based on data from 2005 to 2018 WoS database [31], the review of research trends and patterns on the subject of recommending systems based on academic journals in China from 2004 to 2013 [32], and the knowledge structure of non-medical methods in the period 1987 to 2017 based on the co-words method [33].

A review of the literature on the subject shows that the co-word is one of the methods used in most scientometric studies to draw a thematic structure in various scientific fields. The present study seeks to identify and plot the data fields of data mining with the help of the co-word method with an analytical view. This research seeks to answer the following questions:

- What is the trend of Iranian scientific productions in the field of data mining over a period of 10 years?
- What are the key and high-frequency words in Iranian scientific productions in the field of data mining?
- What are the most important and widely used sub-clusters in the field of data mining?
- Which fields are active in the field of data mining?

3. RESEARCH METHODOLOGY

In this research, the co-words analysis, which is one of the scientometric methods, has been used to obtain quantifiable quantitative information from the data mining domain. The research population is 3057 papers in the field of data mining that are indexed in the Civilica database in the period of 2009 to 2019. The review of these papers has been done in two periods of the first 5 years (1388-1393) and the second (1398-1393). For this research, the word data mining is considered as the core. All keywords related to the 3057 papers from the Civilica site that included this title in their keyword collection were collected using the Selenium Library in Python. Each keyword is then considered as a node in the network. Network edges are also defined as connections between words. There will be an edge between the two nodes if the two nodes (words) are shared in the list of keywords of a paper. After collecting all the words, the pairwise relationship between all the words was examined and if the two words had a common paper, an edge was defined between them. Given that researchers do not consider a certain standard for the selection of keywords, and the selection of these words is done by interest, in the collected data set, many keywords, despite being the same in concept and meaning, are considered in different papers, such as "neural network", "neural networks", "artificial neural networks" and "MLP neural networks" that all convey the same concept. This causes a node to be considered for each keyword in the network resulting from the keywords of the papers, and as a result the network has duplicate nodes in terms of concept. In addition to creating isolated nodes in the network and forming very small communities with a maximum of 3 or 4 nodes, this makes the results of the analysis not reflect the reality of the problem. To solve this problem, all words are examined in terms of form and concept, and words with the same meaning are considered as a single node. This is done as a combination of machine and manual, as Table 1 shows part of the matching of synonyms.

4. ANALYSIS OF RESULTS

As mentioned, in order to study the scientific structure of the data mining field, the articles were examined in two 5-year periods, the results of which can be seen in Table 2. The results show that the number of data mining articles in the 5 years leading up to 1398 increased from 1098 in the first period to 1959 in the second period, which had a growth of 78.4%, and consequently the frequency of keywords rose by 78.6% from 4101 to 7326.

The synonyms having been matched as a part of preprocessing, the corresponding network is plotted in terms of preprocessed data and finally, the analysis was done. In Fig.1, the steps to implement the research are shown.

In Fig.2, the trend of growth of data mining papers is shown from 1388 to 1398.

As can be seen in the figure, the highest number of articles was in 1394. Also, of the total number of articles published during this period, 96% were related to the papers presented at the conference and 4% were the share of articles published in

TABLE 1 .MATCHING SYNONYMS

Synonymous keywords	alternative
Decision tree, Optimal Decision tree, Decision trees,	*Decision tree*
Neural Networks, Artificial Neural Networks, MLP Neural Networks	*Neural network*
Machine learning, Machine learning Algorithm, ML	*Machine learning*
Support Vector Machine, SVM Algorithm, SVM	*SVM*
Classification, Classifier, classification Algorithm	*Classification*

TABLE 2. NUMBER OF PAPERS IN TWO-5-YEAR PERIODS

period	First(1388-1393)	Second(1393-1398)
No. of papers	*1098*	*1959*
No. of keywords	*4101*	*7326*

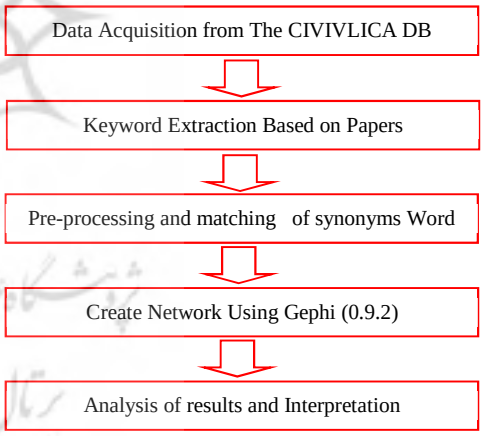

Fig. 1. Workflow of science mapping

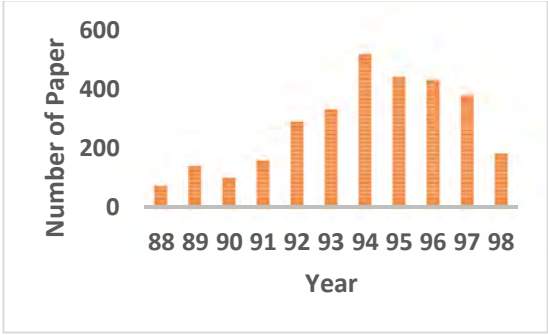

Fig. 2. trend of growth of data mining papers

journals. Table 3 shows the percentage of conference and journal articles based on the two periods ending in 1393 and 1398.

As seen in Fig.3, the methods of data mining are categorized into predictive modelling (supervised) and descriptive modelling (unsupervised).

Based on the keywords of the articles collected during 2009-2010, the results were extracted and evaluated for the two categories. For this purpose, the techniques of predictive and descriptive modeling subsets were compared. The results show that descriptive methods in the first period up to 1393 and in the second period up to 1398, respectively, were used 22% and 14% more than predictive methods. Also, the results indicate that in the whole period of 1388 to 1398, descriptive methods have been used 16% more. The results also show that classification and regression as techniques of predictive modeling and clustering methods and association rules as descriptive modeling techniques have been most used in research. The analysis of the results showed that among the sub-techniques of decision tree classification, neural network, support vector machine and among the clustering techniques, the k-means method has been used the most in solving problems. Table 4 shows the growth rate of using these techniques for the period up to 1398 compared to that up to 1393 and as can be seen, the support vector machine and neural network (in predictive modeling) and k-means (in descriptive modeling) have the highest growth rate used in solving data mining problems.

Regarding the applications of data mining in various fields, the results of the analysis show the application of data mining in medicine, in customer-related issues such as customer relationship services, analysis of customer behavior and finally in issues of fraud detection had the largest share of research in the period under review. Meanwhile, as can be seen in Table 5, medical data mining with 56% and fraud detection with 46% had the highest growth rate in the 5-year period ending 1398 compared to the 5-year period ending 1393 and management of customer relationship with 20% growth is in third place. This indicates that in recent years, data mining in medicine and fraud detection topics in applications such as banking, e-commerce and the insurance industry have become more popular.

In this article, the concept of social network analysis and in particular centrality analysis have been used to identify data mining domains. In centrality analysis, important and influential nodes are identified in terms of various centrality criteria. For the specific applications considered in this paper, the important criteria of centrality are the degree, betweenness, and closeness described in the previous sections. The values of degree centrality, betweenness, and closeness for the respective network are calculated in the first and second 5-year periods. Tables 6 and 7 show the values of 5 nodes that have the greatest centrality. Of the words reviewed in the network, the issues of "decision tree", "classification", "neural network" and "clustering" have the greatest centrality of degree, betweenness and closeness, which indicate the much production of articles and extensive work on these topics. From the point of view of betweenness centrality, the higher the mentioned values suggest that these words play a greater mediating role in the network. From a centrality point of view, closeness also

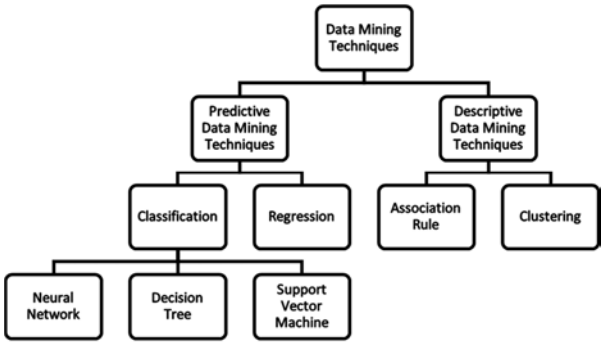

Fig. 3. classifying data mining techniques

TABLE 3. PERCENTAGE OF JOURNAL PAPERS AND CONFERENCE PAPERS PUBLISHED IN TWO PERIODS

period	First(1388-1393)	Second(1393-1398)
Conference papers	***97.5%***	***95%***
Journal papers	***2.5%***	***5%***

TABLE 4 .THE GROWTH RATE OF USING DATA MINING TECHNIQUES

Technique	The growth rate of the second period to the first period
SVM	***46%***
Neural network	***42%***
Decision tree	***32%***
k-means	***30%***

TABLE 5. THE GROWTH RATE OF USING DATA MINING APPLICATIONS

Application	The growth rate of the second period to the first period
Data mining in medicine	***56%***
Data mining in fraud detection	***46%***
Data mining in customers service relationship	***20%***

TABLE 6. DEGREE, BETWEENNESS, AND CLOSENESS CENTRALITY (FIRST FIVE YEARS LEADING TO 93)

Betweenness Centrality		Closeness Centrality		Degree Centrality	
Node Name	***Value***	***Node Name***	***Value***	***Node Name***	***Value***
Clustering	1736.9	Clustering	0.84	Decision Tree	816
Decision Tree	730.71	Decision Tree	0.75	Clustering	802
Classification	621.09	Classification	0.74	Classification	694
Association Rule	360.70	Neural Network	0.68	Neural Network	548
Prediction	302.13	Prediction	0.65	Prediction	462

TABLE 7. DEGREE, BETWEENNESS, AND CLOSENESS CENTRALITY (SECOND FIVE YEARS LEADING TO 98)

Betweenness Centrality		Closeness Centrality		Degree Centrality	
Node Name	***Value***	***Node Name***	***Value***	***Node Name***	***Value***
Clustering	2045.2	Clustering	0.77	Decision Tree	1668
Decision Tree	1362.9	Neural Network	0.75	Neural Network	1486
Neural Network	1254.5	Classification	0.73	Classification	1262
Classification	1233.7	Decision Tree	0.72	Medical Datamining	1196
Prediction	543.2	Prediction	0.68	Prediction	1196

means that these words can be connected to other nodes in the convenience network and are more closely related to other topics.

Community detection

In this research, Gephi software has been used to draw and analyze network results. Also, the results are analyzed based on Girvan -Newman and Louvain community detection algorithms [34]. According to the Girvan -Newman algorithm, the number of communities in the first five years is 90, and in the second 5- years is 133. The results, according to the Louvain algorithm are five communities in the first 5-years and 6 communities in the second 5-years, which are shown in Figs 4 and 5. This shows that according to the Louvain algorithm, community detection is very general compared to the Girvan -Newman algorithm. The results of both algorithms confirm that in the first 5-year period, unsupervised modeling with focus on clustering ,issues and supervised modeling with focus on classification and issues related to the application of data mining techniques, especially issues related to CRM and medicine, respectively are the most important. The same trend has been repeated in the second 5-year period and in the 10-year period from 1388 to 1398.

The representation of the network of data mining paper analysis in 10 years is presented in Fig.6. As seen in Fig.6, the blue community with focus on clustering includes data mining issues in sales and customer discussions, and the yellow community with focus on classification includes data mining topics in medicine. It can be analyzed that classification is used more in medical data mining topics, but clustering is used more in CRM topics. Our results show that in the field of data mining techniques, the discussion of fuzzy logic has become more prominent in recent years. It can be analyzed that since uncertainty is always present in issues, researchers will turn more to fuzzy topics in the future.

As can be seen in these figures, some communities are less distant from each other than the others, and this means that related research is more common among these communities than others. From this point of view, it can be understood that clusters are more or less related to each other based on the distances they have from each other; Closer communities are basically areas that have more in common and more communication. For example, "clustering", "decision tree", "medical data mining" and "neural network" are among the items in close communities and related fields the number of joint researches among which is more than other cases. Also, as can be seen from these figures, in the second 5-years, the density of subjects such as clustering, neural network, prediction and medical data mining is much higher than the previous period, which indicates the focus of research. is about using these themes.

Regarding the key and high frequency words in Iranian scientific productions in the field of data mining, which is considered as one of the research questions, as seen in Figs 7 and 8, " Decision Tree "," clustering "," neural network

Fig. 5. schema of network(93-98)

Fig. 4. schema of network(88-93)

Fig. 6. schema of network(88-98)

"and" classification " has the highest frequency of repetition in their articles.

Also, words such as medical data mining and innovative and meta-innovative algorithms in the 5-year period ending in 1398, in terms of repetition frequency, have gained higher growth compared to other words. This shows that in recent articles and researches, the category of data mining in medical issues has been more considered by researchers.

Examining the results based on network analysis, it was found that the terms "clustering", "classification", "prediction", "neural network", "medical data mining", "support vector machine", "Innovative algorithms" and "machine learning" have been most used in research during the 10-year period studied in this article. Fig.9 shows the trend of using these keywords in articles in this field.

Fig.10 shows the trend of using keywords in percentage in different years.

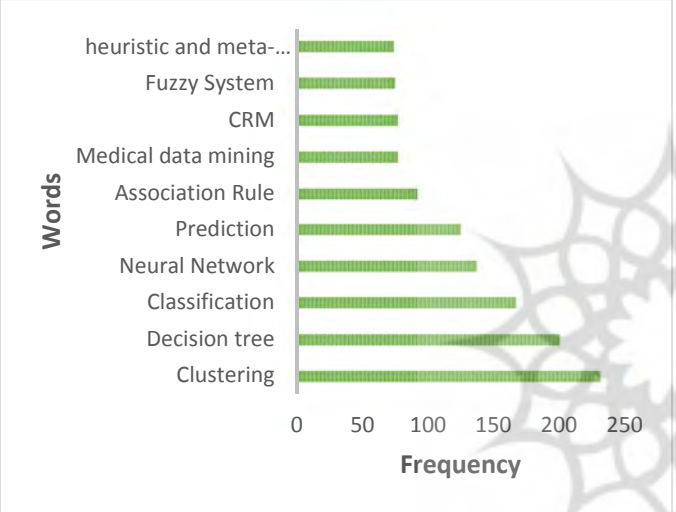

Fig. 7. keywords with the greatest frequency in five years leading to 93

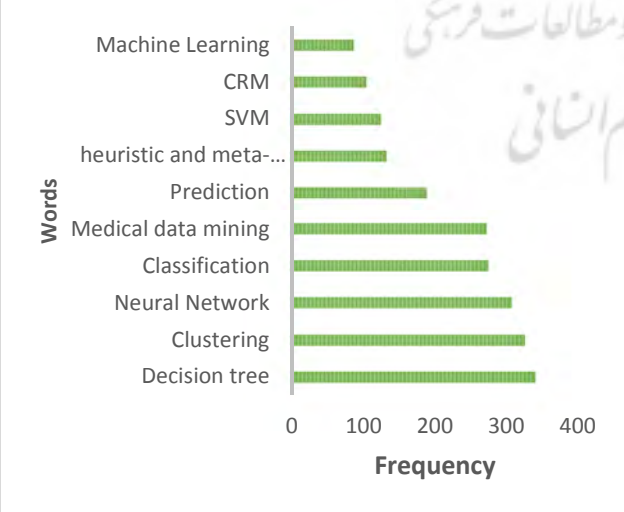

Fig. 8. keywords with the greatest frequency in five years leading to 98

5. Conclusion

Analysis of scientific texts and articles in a specialized field highlights the specific topics of that field and reveals other topics related to the main topic. Today, data mining is widely used in various fields. The need for familiarity with its various fields as well as important topics and updates for research is more and more essential for researchers. In this article, based on the analysis of social networks as well as co- words , the scientific map and the network corresponding to the category of data mining and its sub-domains were drawn. The results of this study showed that data mining and its sub-domains are still considered as interesting topics to researchers in various sciences. Also, based on this research, it was found that in some sub-fields as well as fields related to data mining such as fuzzy logic, recommend systems, text-mining, sentiment analysis, the potential for conducting and continuing new and applied research is on the rise and less attention is paid to other related topics that interested researchers can search for their research topics in these areas. Finally, identifying and predicting sub-domains that may be more likely to be the focus of researchers' research in the future and examining trends in different domains over different periods, and the same issue in valid databases such as Scopus and WoS can be explored as future work in this area.

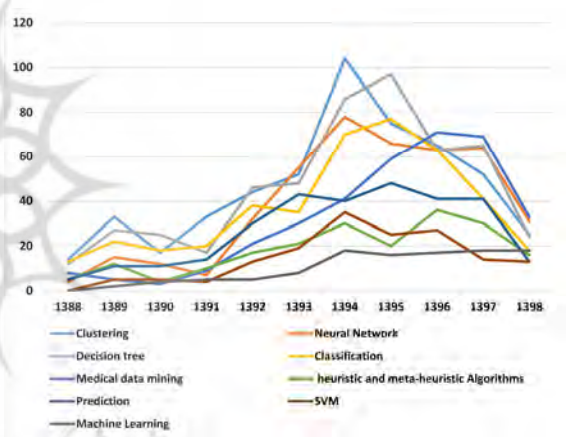

Fig. 9. Trend of using data mining keywords

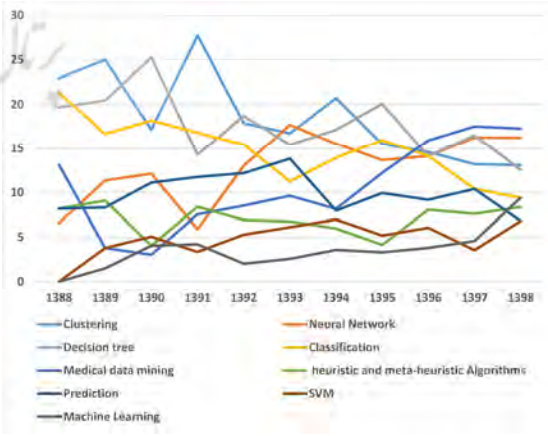

Fig. 10. trend of using data mining keywords

REFERENCES

[1] D. T. Larose, and C. D. Larose. *Discovering knowledge in data: an introduction to data mining*, vol. 4, John Wiley & *Sons*, 2014.

[2] *G. K. Gupta*, *Introduction* to data mining with case studies. PHI Learning Pvt. Ltd., 2014.

[3] I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J. F. Chang, and L. Hua. "Data mining in healthcare and biomedicine: a survey of the literature". *Journal of medical systems*, vol. 36, no. 4, pp. 2431-2448, 2012.

[4] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[5] A. Rip, and J. Courtial, "Co-word maps of biotechnology: An example of cognitive scientometrics". *Scientometrics*, vol. 6, no. 6, pp. 381-400, 1984.

[6] Q. He, "Knowledge discovery through co-word analysis", *Library Trends*, Summer 1999, pp. 133-159, 1999.

[7] M. Giannakos, Z. Papamitsiou, P. Markopoulos, J. Read, and J. Pablo Hourcade, "Mapping child–computer interaction research through co-word analysis". *International Journal of Child-Computer Interaction*, vol. 23-24, p. 100165, 2020.

[8] M. Callon, A. Rip, and J. Law, eds. *Mapping the dynamics of science and technology: Sociology of science in the real world*. Springer, 1986.

[9] P. Vinkler, *The evaluation of research by scientometric indicators*. Elsevier, 2010.

[10] L. Alcaide–Muñoz, M. P. Rodríguez–Bolívar, M. J. Cobo, and E. Herrera–Viedma. "Analysing the scientific evolution of e-Government using a science mapping approach", *Government Information Quarterly*, vol. 34, no. 3, pp. 545-555, 2017.

[11] Chen, C. "Holistic sense-making: conflicting opinions, creative ideas, and collective intelligence", Library Hi Tech, Vol. 25 No. 3, pp. 311-327,2007.

[12] E. C. M. Noyons, *Bibliometric mapping as a science policy and research management tool*. Diss. Centrum voor Wetenschaps-en, Faculty of Science, Leiden University, 1999.

[13] K. Börner, C.Chen, and K. W. Boyack, "Visualizing knowledge domains", *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 179-255, 2003.

[14] Y. Chen, and Z. Y. Liu. "The rise of mapping knowledge domain", *Studies In Science of Science*, vol. 2, 2005.

[15] L. Ames, "Mapping the field of arts and economics", *In Paper 16th International Conference on Cultural Economics, ACEI, Copenhagen*, 2010.

[16] R. Guns, Y. X. Liu, and D. Mahbuba. "Q-measures and betweenness centrality in a collaboration network: a case study of the field of informetrics", *Scientometrics*, vol. 87, no. 1, pp. 133-147, 2011.

[17] S. Dai, X. Duan, and W. Zhang, "Knowledge map of environmental crisis management based on keywords network and co-word analysis, 2005-2018", *Journal of Cleaner Production*, vol. 262, p. 121168, 2020.

[18] B. Furht, *Handbook of Social Network Technologies and Applications*, Springer, 2010, p. 3.

[19] Y. S. Lin, "Topic evolution of innovation academic researches", *Journal of Small Business Strategy*, vol. 26, no. 1, pp. 25-36, 2016.

[20] F. Muñoz-Leiva, J. Sánchez-Fernández, F. J. Liébana-Cabanillas, and A. G. López-Herrera. "Applying an automatic approach for showing up the hidden themes in financial marketing research (1961–2010)", *Expert Systems with Applications*, vol. 39, no. 12 pp. 11055-11065, 2012.

[21] P. C. Lee, and H. N. Su. "Quantitative mapping of scientific research—The case of electrical conducting polymer nanocomposite", *Technological Forecasting and Social Change*, vol. 78, no. 1, pp. 132-151, 2011.

[22] Gan, Chunmei, and Weijun Wang. "Research characteristics and status on social media in China: A bibliometric and co-word analysis", *Scientometrics*, vol. 105, no. 2, pp. 1167-1182, 2015.

[23] B. N. Yan, T. S. Lee, and T. P. Lee, "Mapping the intellectual structure of the Internet of Things (IoT) field (2000–2014): A co-word analysis", *Scientometrics*, vol. 105, no. 2, pp. 1285-1300, 2015.

[24] C. P. Hu, J. M. Hu, S. L. Deng, and Y. Liu, "A co-word analysis of library and information science in China", *Scientometrics*, vol. 97, no. 2, pp. 369-382, 2013.

[25] X. Chang, X. Zhou, L. Luo, C. Yang, H. Pan, and S. Zhang, "Hotspots in research on the measurement of medical students' clinical competence from 2012-2016 based on co-word analysis", *BMC Medical Education*, vol. 17, no. 1, p. 162, 2017.

[26] R. Elisabetta, M. Gabriele, S. Massimo, B. Sofia, and S. M. Giuseppe, "Structural trend and conceptual evolution of research on genetically modified organisms using a science mapping approach", *Journal of Cleaner Production*, vol. 205, pp. 329-338, 2018.

[27] J. Piñeiro-Chousa, M. Á. López-Cabarcos, N. M. Romero-Castro, and A. M. Pérez-Pico, "Innovation, entrepreneurship and knowledge in the business scientific field: Mapping the research front", *Journal of Business Research*, vol. 115, pp. 475-485, 2020.

[28] H. Baziyad, S. Shirazi, S. Hosseini, and R. Norouzi, "Mapping the intellectual structure of epidemiology with use of co-word analysis", *Journal of Biostatistics and Epidemiology*, vol. 5, no. 3, pp. 210-215, 2019.

[29] N. Girard-Perier, S, Dorey, S. R. Marque, and N. Dupuy, "Mapping the scientific research on the gamma irradiated polymers degradation (1975–2018)", *Radiation Physics and Chemistry*, vol. 168, p. 108577, 2020.

[30] M. Giannakos, Z. Papamitsiou, P. Markopoulos, J. Read, and J. P. Hourcade, "Mapping child–computer interaction research through co-word analysis". *International Journal of Child-Computer Interaction*, vol. 23-24, p. 100165, 2020.

[31] S. Dai, X. Duan, and W. Zhang, "Knowledge map of environmental crisis management based on keywords network and co-word analysis, 2005-2018", *Journal of Cleaner Production*, vol. 262, p. 121168, 2020.

[32] J. Hu, and Y. Zhang, "Research patterns and trends of Recommendation System in China using co-word analysis", *Information Processing & Management*, vol. 51, no. 4, pp. 329-339, 2015.

[33] D. Nguyen, "Mapping knowledge domains of non-biomedical modalities: A large-scale co-word analysis of literature 1987–2017", *Social Science & Medicine*, vol. 233, pp. 1-12, 2019.

[34] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and experiment*, no. 10 P10008, 2008.

Babak Teimourpour received his B.Sc in Industrial Engineering, Sharif University, Tehran, Iran in1996 and received his M.Sc. degree in Socio-Economic Systems Engineering, Institute for Research on Planning and Developement, Tehran, Iran in 1998. Also he received his Ph.D. in Industrial Engineering at Tarbiat Modares University, Tehran, Iran in 2010. His research interests include data mining, social network analysis.

Meysam Alavi was born in 1982 in Hamadan, Iran. He received his B.Sc. and an M.Sc. degree in Computer Engineering - Software in 2005 and 2010 respectively. Currently, he is an M.Sc. student in Information Technology at Tarbiat Modares University. His research interests include medical image processing, machine learning, data mining, social network analysis.

Fateme Shahrabi Farahani was born in 1993 in Tehran. She received her B.Sc. in Information Technology Engineering from the Department of Electrical and Computer Engineering, University of Tehran in 2016. Currently, she is an M.Sc. student in Information Technology at Tarbiat Modares University. Her research interests include data science and social network analysis.

Mina Ghasemi was born in 1991. She received her B.Sc. in Computer Engineering from Payam Noor University in 2015 and M.S. degree in Information Technology from Tarbiat Modares University in 2020. Her research interests include machine learning and bioinformatics.