



A Survey on Contribution of Data Mining Techniques and Graph Reading Algorithms in Concept Map Generation

B. Lavanya*, **A. Auxilia Princy**

Department of Computer Science, University of Madras, Chennai-25, India

*Email: lavanmu@gmail.com

ARTICLE INFORMATION

DOI: 10.15415/jotitt.2018.62009

Keywords: Concept maps, big data, graphs

ABSTRACT

Concept maps are a pictorial representation of concepts found in data and it shows relationship between concepts. These Concept map help us to understand the whole data content, makes it easily readable and memorable. They are used to deliver complex data in an understandable form (map, tree, graph, etc), which is used for a better understanding and decision making for researchers and business, etc. This paper discusses the recent researches about concept maps and data mining techniques, and graph reading algorithms used for concept map generation.

1. Introduction

Concept map was originally developed by Joseph D. Novak and his analysis team at Cornell University in the year 1970. The basic idea was to make complex and scientific part of studies easy and understandable. The automatic development of concept mapping is still a working research area. The concept maps identify the relationship in the context but its accuracy is still below expected percentage. However this survey paper depicts that using different mining algorithms one can automatically generate concept map for any big data. This work focuses on concept map generation using frequent mining algorithms and graph reading algorithms.

2. Literature work

2.1. Concept map

A systematic concept mapping study says that, assistance in teaching and learning and knowledge organization are the main purpose for concept maps. Computer science has a vast and many sub areas to explore by using a concept map, with which one can easily learn and understand the concept. [1] This mapping involves three main phases, first is planning which includes a review protocol, inclusion and exclusion criteria, the second one is conducting ,that is searches and select the studies in order to extract and synthesize data and the third one is reporting the final phase that aims at writing up the result.

Using concept mapping as tool for conducting research, a study [2] says that three

different approaches were used and they are word frequency, relational and cluster approach. All these approaches collected data, analyzed data for finding interconnection between concepts and finally presented. This data presentation consists of illustrating concepts, present findings, highlighting connections, and framework of research or research process.

To identify concepts in data, different association rule mining technique are used to construct concept map automatically [3] A grade fuzzification and fuzzy data mining is used and then by mining association rules we can construct a map and it is run for anomaly diagnosis where the data redundancy can be reduced.

J. Villalon *et al.*[4] have proposed an automatic system for generating concept

maps, Concept map miner which has three phase of work, first it identifies the concepts and second to find its relationship (syntactic meaning) between the concepts and finally summarize the concept Fig 1. Keeping this paper as a base paper [4], [5] proposed that automatic generation of maps, that they find the dependency between the word and the domain. If null hypotheses between word and domain there is no dependence between them, an alternative hypothesis is dependence between the word and domain, therefore there is a positive set (A) and negative set(B) and test is conducted between A and B. The result value is compared to a threshold value, considered as the concept and then the map is generated.

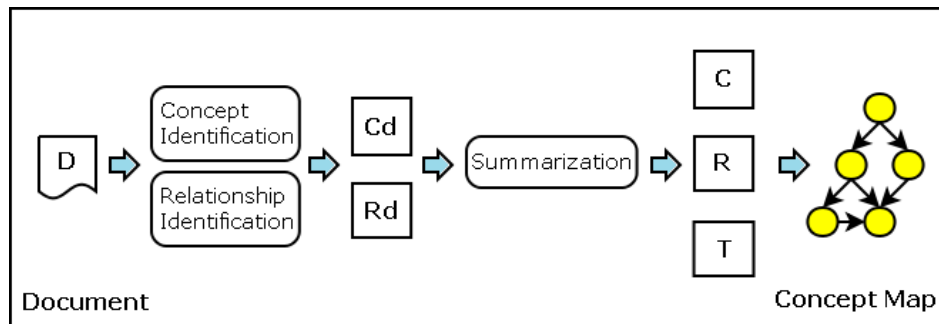


Figure 1: A Concept map miner process [5]

A CM is outlined as a triplet $CM = (C, R, T)$ where C could be a set of ideas, R a group of relationships between ideas, and T is that the maps topology or spatial distribution of the concepts [5].

A study of semantic knowledge [6], lists some notable tools for creating concept maps or mind mapping. CMAP Tools, Coggle, Compendium, Docear, Free Mind, Freeplane, MindMup, SciPlore MindMapping, WikkaWiki, VUE, Xmind, these are some freeware that can create a tree like image or any pictorial format and some software turns maps into pdf format also.

Divya *et al.* [7] Proposed an idea that mind mapping tools using data mining techniques such as classification, clustering, regression and association rule mining, will help the user in deep understanding of information and its association and helps for strategizing the information in more accurate manner.

A Study on Predictive Analysis on Concept Maps concludes that neural network and decision tree are widely used technique for predicting student performance. Here B. Lavanya *et al.* [8] says predicting student performance is important that helps to improve the learning and teaching process.

2.2 Frequent pattern mining algorithms

Dina Fawzy *et al.* [9], has proposed that, the knowledge mining technique to big data analysis, big data volume are often resolved by k nearest neighbor, parallel processing and sample modeling are done by decision tree, k means, neural network, bagging, random forest and apriori and big data velocity by decision tree and FP growth and Apriori and big data veracity by some technique like variant of k nearest neighbor.

M. Nagalakshmi *et al.* [10] Proposed an implementation of Apriori in big data sets, using hadoop, that is used for big voluminous data in such cases when Apriori algorithm is used, it helps to search the information from the data.

A comparative study of tools and techniques used in big data [11] says

that the technologies used by big data application to handle massive data are Hadoop, Map Reduce, Apache Hive, No SQL and HPCC. Tool has four classification algorithms implemented which is taken from WEKA's machine learning library, namely: Decision Trees, Naïve Bayes, Random Forest and Support Vector Machines (SVM).

M. Sinthuja *et al.* [12] Proposed a research of improved FP growth (IFP) algorithm in association rule mining, usually FP growth has item name and count as two attributes but in proposed system it is suggested to have four attributes, item name, count, node link and flag that makes lot easier to evaluate and it is proved by experimental research that IFP is better than FP growth algorithm Table 1 Popular data mining techniques used in big data analytics.

Table1: Popular data mining techniques used in big data analytics

| Data Mining Task | Most used Techniques |
|----------------------|--|
| Classification | k-nearest Neighbour Decision Trees Support Vector Machines ID3 ID4 |
| Association Mining | Apriori |
| Clustering | K-means Clustering |
| Optimization | Genetic Algorithms |
| Classifier Ensembles | Random Forests |
| Regression | Logistic Regression Linear Regression |

Arwa Altameem *et al.* [13], proposed a hybrid approach for improving efficiency of Apriori algorithm for frequent item set mining. The approaches of improved Apriori may be of hash based, transaction reduction, partitioning and sampling. The time comparison tested by experiments has proved that improved Apriori is faster than the basic Apriori.

Lior Shabtay *et al.* [14] proposed a guided FP growth (GFP) algorithm for targeted mining, this GFP uses minority report algorithm with normal FP growth and this paper proves numerical tests by census income dataset that GFP is better than FP growth algorithm.

An improved pre post algorithm for frequent pattern mining with hadoop was

proposed by Sanket Thakarea *et al.* [15] discussed about Pre Post algorithm like PPC tree is generated before the FP tree will help us to avoid multiple data set scan. A PPC uses post order traversal and pre order traversal. Repost algorithm is implemented on hadoop architecture in the map reduce phase.

A survey of periodic pattern mining in spatiotemporal database [16] proposes that two algorithms that are EFPMA (Extended Regular Model Detection Algorithm) used to find frequent sequential patterns from the spatiotemporal dataset and the ETMA (Enhanced Tree-based Mining Algorithm) for detecting effective cyclic models with symbolic database representation.

K.A.Baffour *et al.* [17] proposes a modified Apriori algorithm (MAA) which follows six steps, this MAA is tested and compared with all other improved Apriori algorithms and proved MAA is more efficient than the other and it overcame the drawback of classical Apriori algorithm.

A survey paper [18] has done a comparative study of decision tree algorithms for classification in data processing. Decision tree algorithm like ID3, C4.5, J48, CART, are

analyzed and algorithms compare using various parameters like Advantages, Disadvantages, Measure, Procedure, Pruning and Approach.

2.3 Graph clustering algorithms

A study of algorithms for Extraction of Sub trees of a Sentence Dependency Parse Tree [19] says that using parse tree the syntactic n grams can be used to extract the sub trees. The syntactic gram can find the internal meaning of the sentence where each parse tree has a grammar lying within. They are suitable, because they explore directly the syntactic information and allow introducing into machine learning methods, for example, identifying more accurate patterns of how a writer uses the language.

Reena Mishra *et al.* [20] compares graph clustering algorithm via random graph, they proposed comparison of RNSC (Restricted Neighbourhood Search Clustering) and MCL (Markov Clustering) algorithms based on Erdos-Renyi and Power-Law Distribution graphs and concluded that in case of Erdos-Renyi graphs run time of RNSC algorithm is better as compared to MCL and RNSC is better than MCL in case of sparse graphs.

Table 2: Popular techniques in graph mining

| Graph Mining Techniques | Popular Algorithms | Used For |
|---------------------------------------|--------------------------------|--|
| Graph reading algorithm | Erdos-Reni | Reading Random Graph |
| Graph clustering algorithm | MCL(Markov Clustering) | Clustering, Random walks |
| Frequent sub graph mining | SUBDUE, Gspan, jpmminer, mspan | Mining sub graphs, mining with Apriori,FP growth |
| Graph Based Structured pattern mining | Gspan, Subdue, sleuth | DFS and BFS to identify patterns |

Yu-Feng Li *et al.* [21] proposed a new a meaningful graph representation for clustering graph algorithm, to develop data, where each resulting sub-graph

corresponds to a cluster with highly similar objects connected by edges. Numerical evidences show that algorithm can provide a very good clustering accuracy for a number of benchmark data. In addition, it has a relatively low time complexity in comparison with two sophisticated clustering methods kernel K-means and HCS.

Hongzhi chen *et al.* [22] proposed and experimented, graph miner (G-miner) architecture, for general graph mining. G-Miner adopts a unified programming framework for implementing a wide range of graph mining algorithms. G-Miner, which provides an expressive API and achieves outstanding performance with its novel task pipeline that removes the synchronization barrier and hides the overheads of network and disk I/O.

A survey of algorithms for dense sub graph discovery [23], gives a brief explanation about the graph terminologies and components and the different algorithms used and which is mostly used algorithm in graph mining Table 2 Popular techniques in graph mining.

3. Conclusion and future work

Concept mapping is useful and helps in simple understanding; our study says that concept mapping is immensely used in several fields. Exploitation mining algorithms combined with graph algorithms and neural network, we will determine the concepts in big data and can be easily mapped. In future automatic concept map generator will be designed, where it has a varied spectrum of applications.

References

- [1] V. D. Santos, É. F. D. Souza, K. R. Felizardo and N. L. Vijaykumar, "Analyzing the use of concept maps

in computer science: A systematic mapping study", *Informatics in Education*, vol. 16, no. 2, pp. 257–288. 2017.

- [2] S. C. O. Conceição, A. Samuel and S. M. Y. Biniecki, "Using concept mapping as a tool for conducting research: An analysis of three approaches", *University of Wisconsin Milwaukee UWM Digital Commons*, vol. 3, no. 1, pp. 11-18, , 2017.
- [3] S. S. Tseng , P. C. Sue, J. F. Weng, Jui-Feng and J. M. Su, "A new approach for constructing the concept map: conference and education", *IEEE International Conference on Advanced Learning Technologies, 2004*.
- [4] A. Nugumanova, M. Mansurova, E. Alimzhanov, D. Zyryanov and K. Apayev "Automatic generation of concept maps based on collection of teaching materials", *4th International Confernece on Data Management Techologies and Applications*, January 2015.
- [5] J. Villalon and A. C. Rafael, "Concept maps as cognitive visualizations of writing assignments", *Educational Technology & Society*, vol. 14, no. 3, pp. 16–27, 2011.
- [6] R. Katagall, R. Dadde, R. H. Goudar and S. Rao, "Concept mapping in education and semantic knowledge representation", An illustrative survey: iccc, *procedia computer science*, vol. 48, pp. 638–643. 2015.
- [7] D. Premchandran and B. V. V. S. Chalamayya, "Retrieving information through mind mapping tools for mining data", *International Journal of Scientific Research Engineering & Technology*, vol. 7. no. 1, pp. 10-13, 2018.

- [8] B. Lavanya, and V. J. Bai, "A study on predictive analysis on concept maps using data mining techniques", *International Journal of Advanced Research in Science, Engineering and Technology*, vol. 5, no.10, pp. 7141-7144, 2018.
- [9] D. Fawzy, S. Moussa and N. Badr, "The evolution of data mining techniques to big data analytics: an extensive study with application to renewable energy data analytics", *Asian Journal of Applied Sciences*, vol. 4, no. 3, pp. 756-766, 2016.
- [10] M. Nagalakshmi, I. S. Prabha and K. Anil, "Big data implementation of apriori algorithm for handling voluminous data-sets", *International Journal of Engineering and Technology*, vol. 7, no. 1.5, pp. 217-220, (2018).
- [11] B. Thillaieswari, "Comparative study on tools and techniques of big data analysis", *International Journal Of Advanced Networking & Applications*, vol. 8, no. 5, pp. 61-66, 2017.
- [12] M. Sinthuja, N. Puviarasan and P. Aruna, "Research of improved FP-growth (IFP) algorithm in association rules mining", *One Day National Conference On "Internet Of Things - The Current Trend In Connected World" NCIOT*, pp. 24-31, 2018.
- [13] A. Altameem and M. Ykhlef, "Hybrid approach for improving efficiency of apriori algorithm on frequent itemset", *International Journal of Computer Science and Network Security*, vol. 18, no. 5, pp. 151-155, 2018.
- [14] L. Shabtay, R. Yaari and I. Dattner, "A guided FP-growth algorithm for multitude-targeted mining of big data", Cornell University, Arxiv. Org/Abs/1803.06632v2 [July 5, 2018].
- [15] S. Thakare, S. Rathi, R. R. Sedamkar, "An improved prepost algorithm for frequent pattern mining with hadoop on cloud", *Procedia Computer Science*, vol. 79, pp. 207-214, 2016.
- [16] O. Obulesu, A. R. M. Reddy and M. Mahendra, "IOP Conf. Series:A Ccomparative study of frequent and maximal periodic pattern mining algorithms in spatiotemporal databases, *Materials Science and Engineering*, vol. 225, 2017.
- [17] K. A. Baffour, C. Osei-Bonsu and A.F. Adekoya, "A modified apriori algorithm for fast and accurate generation of frequent item sets", *International Journal of Scientific & Technology Research*, vol. 6, no. 8, pp.169-173, 2017.
- [18] S. N. Chary, B. Rama, "A survey on comparative analysis of decision tree algorithms in data mining", *International Conference On Innovative Applications In Engineering and Information Technology (ICIAEIT-2017)*, vol. 3, no. 1, pp. 91-95, 2017.
- [19] J. P. P. Dur'An, et. al, "Algorithm for extraction of subtrees of a sentence dependency parse tree", *Acta Polytechnica Hungarica*, vol. 14, no. 3, 2017.
- [20] R. Mishra, S. Shukla, D. Arora, M. Kumar, "An effective comparison of graph clustering algorithms via random graphs", *International Journal Of Computer Applications*, vol. 22, no. 1, 2011.

-
- [21] Y. Li, L. H. Lu., Y. C. Hung, "A new clustering algorithm based on graph connectivity", *Computing Conference 2018, 10-12 July 2018 J London, UK*.
- [22] H. Chen, M. Liu, Y. Zhao, X. Yan, D. Yan, and J. Cheng, "G-miner: An efficient task-oriented graph mining system", *The Thirteenth Eurosys Conference, Porto, Portugal, 2018*.
- [23] V. E. Lee, N. Ruan, R. Jin and C. Aggarwal, "A Survey of Algorithms for Dense Sub Graph Discovery", *Managing and Mining Graph Data*, 40, 2010, pp. 303-336.