International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

Enrichment on High Utility Sub Graph Mining on Transactional Database

[1]Akshatha H M,[2]Geethika M M,[3]Suresh K

ABSTRACT*— Enrichment on High utility item set that produces a great deal of benefit for the seller. High utility item set mining is an effective technique for dynamic huge data. Fundamentally the database that contains products with amounts and mining ability is limited to transaction data consist of items. Graph mining is a non-trivial graph structure from a confounded system. In the proposed technique that we are going to change over the transaction database into sub graph structure. By utilizing high utility sub graph mining calculation and it produces yield as a graph structure.*

Keywords— *Data mining, High utility, Graph mining, Frequent mining.*

I. INTRODUCTION

Data mining is the method toward finding designs in vast informational collections with techniques at the measurements, and database structures. Data mining is an interdisciplinary sub field of software manufacturing and visions. The general impartial to remove data from an informational group and change the transaction data into an graph assembly for further use. The distinction between data analysis and data mining is that information is applied to test replicas and speculations on the data set. Data mining utilizes artificial intelligence and truthful models to reveal furtive or concealed. This examine the properties of genuine graphs. And predict how the construction and assets of a given illustration may impact. Graph mining application create replicas that can yield reasonable diagrams that match the samples create in real world graph. In the proposed technique that we are going to change over the transaction database into sub graph structure. By utilizing high utility sub graph mining calculation and it produces yield as a graph structure. Data removal from a given information vault to decide the conduct of a specific framework, or to decide the predictive result of a specific issue articulation for the instance of an obscure stateor contribution procedures **a** utilization of varied distance in facilitating the lifetime of individuals, by its infiltration into spaces going from online business to healthcare. Achieved knowledge algorithms have been generally utilized in expectation issues to gauge the result of a changed issue from a basic information mirroring the genuine results of comparative issues. A main worry that emerges out of the upstairs systems of data sources for data mining utilizing regulated knowledge procedures for working of arrangement directions is the security and classification of the data, particularly in protecting the character of the topics to whom

[1]Department of Computer Science, Amrita School of Arts and Sciences, Mysuru,Amrita Vishwa Vidyapeetham, India, akshathahm1397@gmail.com
[2] Department of Computer Science, Amrita School of Arts and Sciences, Mysuru,Amrita Vishwa Vidyapeetham, India, geethikamm@gmail.com
[3]Department of Computer Science,Amrita School of Arts and Sciences,Mysuru,Amrita Vishwa Vidyapeetham, India, suresh.kalaimani@gmail.com

Received: 02 Jan 2020 | Revised: 12 Feb 2020 | Accepted: 17 Mar 2020

the data relates to. Many confidentiality topics could emerge due of the mining of such delicate individual information, and abuse of the information by break of security can cause lawful and moral issues past the space of information mining. The successive and high utility itemsets discovers request in varied assortment of spaces, for example, trade, stockroom or circulation focus, electronic trade, banking, protection and social insurance. In the stockroom or appropriation setting, a chief is regularly keen on streamlining material developments and extra room. High Utility Itemsets can benefit a director adequately break down the client requesting or request designs and progress the operational productivity of a distribution center (Chen and Wu, 2005; Chen et al., 2005). In the electric business setting, mining continuous or High Utility Itemsets can help find intriguing examples of client snap or buy designs and create reasonable item suggestions. These items have particular qualities as far as their buy frequency (clothing versus bread), value (precious stone rings versus shades), edges (gems versus kitchen things, etc. A retail supervisor is frequently intrigued to investigate various blends of things, find significant examples and settle on key choices identified with estimating, advancements and item arrangements. While the utilization of HUIs can be helpful to address such issues, it doesn't represent various varieties in item attributes. In addition, the utilization of standard HUI mining systems involve determination and tuning of a solitary least utility edge esteem. Picking besides little a base value limit worth (to deal with unique thing attributes) prompts combinatorial fit of exploration space and the quantity of examples produced can be extremely huge.

II. LITERATURE REVIEW

Cristobal´ Romero et al. [2010] This paper discloses about how Educational data mining is developing and building up a coordinated data that manages the advancement of information association in an educational context. Educational data mining particularly used to break down the instructive information. This paper reviews the applicable information of educations approved out in this field. The methodologies are Educational data mining contemplates that utilization regularly data mining procedures by using typically data mining techniques, such as organization, association rule mining, sequential mining, clustering, text mining. Psychometrics and statistical methods have been applied to educational mining, similar to understand conduct/execution, educational program, and so forth that was accumulated in study hall situations to determine student's performance in various field. The Educational data mining procedure changes over huge information originating from instructive frameworks into helpful data that might greatly affect instructive research and practice. The target of the examination and representation of data is to feature valuable data and support decision mining.

Tin Truong et al. [2015] proposed mining high normal utility thing sets in a measurable database is an expansion of the customary issue of continuous thing set mining. High Average Utility Itemset is more testing than mining incessant thing sets utilizing the customary help model. This paper proposes close-fitting normal utility upper limits dependent on vertical database portrayal and three proficient pruning procedures. High utility thing set mining High Utility Itemset Mining and pruning are the two calculations they utilized. And afterward descending conclusion property (DCP) produces exchange weighted use on the grounds that DCP doesn't hold for the utility measures. Conventional structure for assessing the pruning capacity of UBs dependent on anti to drone-like criteria is anticipated. And in this paper, three pruning plans are proposed to remove gloomy candidates early, which dramatically decrease the search space. At long last, in this paper they have shown exact proof that High

Average Utility Itemset Mining beats the best in class calculations as far as runtime and amount of connections on both reality and engineered databases.

Yifan Chen et al. [2016] uncertain graphs are mainly seen in bioinformatics, social networks and also to solve many problems related to graphs. This paper examines about researched frequent subgraph mining on single questionable diagrams. Propose a powerful calculation to expand mining execution. In uncertain graphs uses effective algorithm to achieve high performance using #P-hard. The #P-hard help calculation by changing over the genuine incentive into accuracy assurance and they have devised calculation sharing strategies to accomplish better digging execution for high proficiency calculation sharing is utilized. #P algorithm is mainly used for solving the probabilistic value. For the better solution some of the computation technique are used for the better performance. Pruning and validation technique are combined together for giving good results. #P hard mainly support the computation for the good mining performance. Propose a powerful calculation to expand mining execution.

Buczak et al. [2016] Depicts about Arificial intellegience and data mining procedure digital security in help with interruption recognition. The pre-owned strategies are peculiarity identification, signature-based and half and half techniques. Another bit of leeway which is utilized right now profiles of typical action are tweaked for each framework, application, or system. A metric maxPr has been created to findout the upper bound of expSupport of an itemset. Right now potential False Alarm Rates is utilized. Digital security frameworks are made out of both system security frameworks and PC security frameworks. They are particularly successful for recognizing known sort of assaults without producing a bogus caution. This paper is for the most part valuable for perusers who wish to start look into in the field of Machine leraning for digital interruption discovery. A mechanism knowledge method usually contains of two stages: training and testing. Often, some stages are performed:

• Identifying class qualities and modules from training data.

• Classify a subsection of the characteristics necessary for association.

• Learn the perfect using preparation data.

• Use the skilled model to classify the unidentified data.

To determine the effectiveness of the approaches, there is not only one standard but they are numerous criterion to this. Such as complexity, accuracy time for finding an unknown instance.

Xixianchan et al. [2016] Paper is about efficient mining which is one of the huge action in data mining. PFIM count is used on massive data it has two areas, gigantic old table for taking care of bona fide data and minimal new table for as of late delivered data. It has various central focuses when we

Differentiate and various computations that is PFIM (pre -computation visit itemset mining) runs twice snappier than other figuring. Existing relentless itemsets contains up-and-comer age and model based count anyway they are not appropriate in enormous data. Pruning movement is moreover discussed right presently reduce the execution cost which accelerates various errands. Figuring which are existing isn't useful for finding the enormous data. To beat this PFIM (precomputation visit itemset mining) count is being used. PFIM is one of the fastly used estimation used for count. Pruning essentially used for reducing the size of itemsets. By using this PFIM computation three pruning rules are used which quicken the customary itemsets execution.

Nader Aryabarzan et al. [2017] describes itemset is essential data mining task which has various applications. Database structure stores noteworthy information about progressive itemsets. This work Paper fuse capable dataset negNodeset which is set of center points in prefix tree subject to negNodeset data the negFIN is proposed which is one of the successful counts. The exploration paper fuses connection examinations to check the show of negFIN and dFIN. In the assessment it shows that negFIN is faster with least assistance in connection with other. Visit itemset mining is predominantly utilized in datamining and its other undertakings. Data can be put away right now itemsets. Basically utilized proposed work is neg-hub set which fundamentally dependent on bitmap and its portrayal. The proficiency of this neg-hub is for the most part determined by bitwise activity, negNodeset tallying and cheerfulness of negNodeset.

Rahman et al. [2018] Another method has been found to oversee groupings in uncertain databases. The methodology has been here is uW Sequence, iM axP r, and expSupporttop. A capable figuring for mining weighted ordinary itemsets. Some may require only one yield of the database and some may require equivalent dealing with and particular system requires different data structures. The field of progressive itemset mining has an irritated with the disclosure of FP Growth tally, a model headway approach. At this moment, strategy utilizes portion and-vanquish system which engages it to perform speedier all around. The computation looks for after an upper-bound based expected help check to discover the entirety of the strategies correspondingly as various fake movements which are deducted by finding the legitimate anticipated assistance from the database around the end. The proposed structure would be a superior than normal advancement to prepare for weighted questionable groupings. It will by and large be utilized in various works. Rather than gathering and bundling, here proposes a sketchy database considering both weight and reinforce Constraints.

Unil Yun et al. [2018] This paper discloses step by step instructions to successfully mine the ordinary utility on itemsets. The delivered count has high ordinary utility itemsets by significance first-search based mining technique to dodge the augmentation of itemsets pruning methodology is used. Connection rule is used, for instance, course of action, packing for finding information. For the most part used ceaseless itemsets are apriori and FP improvement with make and test approach. To crush hindrance of apriori figuring FP tree structure is used and having better execution appeared differently in relation to apriori. Right now high normal utility itemsets are utilized. It creates profundity first inquiry pruning process. There is a pruning technique for finding the upper limits of the successive itemsets. Calculation utilizes list structures called HAI list for getting the data about mining the itemsets. It isn't as that costly planning to be less expense. To check the exhibition, they have directed numerous examinations with their datasets. It shows better outcome with the calculation.

Jerry Chun-Wei Lin et al. [2018] proposed High Utility Itemset Mining is an augmentation of regular thing set mining. It considers the benefit and nature of thing sets to find high-utility thing sets. this methodology is not withstanding, wrong in true applications since the utility of the thing sets increments. The quantity of things inside it high normal utility thing set mining High Average Utility Itemset Mining was intended to quantify the utility size of the thing sets into account. Incremental high average utility pattern mining algorithm and apriori algorithm are used to handle the incremental database with deal insertion. Direct hash pruning is an updated algorithm used pruning technique used in DHP (direct hashing 7 pruning) algorithm for informing the discovered material. The presentation of the proposed Incremental high average utility pattern mining calculation is contrasted and the best in class PAI, HAUI-tree algorithms on several datasets. At that point embedded to the HAUP-tree as indicated by its right location in the branches.

Rahman et al. [2018] Another technique has been discovered to manage successions in unsure databases. The technique has been here being uW Sequence, iMaxPr, and expSupporttop. A proficient calculation for mining weighted continuous itemsets. Some may require just one sweep of the database and a few may require equal preparing and distinctive strategy requires various information structures. The field of incessant itemset mining has an annoyed with the revelation of FP development computation, a model improvement approach. Right now computation, the methodology uses segment and-survive strategy which empowers it to perform snappier out and out. The computation seeks after an upper-bound based expected assistance check to find all of the courses of action similarly as different false progressions which are deducted by determining the veritable foreseen help from the database at the end. The proposed structure would be a nice development to get ready for weighted uncertain groupings. It will in general be used in different works. Rather than characterization and grouping, here proposes a dubious database thinking about both weight and backing Limitations.

Srikumar Krishnamoorthy et al. [2018] proposed Top-K High Utility Item set mining issue offers more prominent adaptability to a chief in determining their ideas of things. The top-k HUI mining issue, be that as it may, is all the more testing and requires utilization of compelling limit raising techniques. A few limit raising techniques have been future in the writing to recover the general productivity of mining top-k High Utility Itemsets. Top K utility and Reduce Error pruning tree are the two calculations in the writing that embrace a two-stage strategy. In this paper proposed about primary motivation is to enterprise improved threshold raising plans to significantly recover the presentation of top-k High Utility Itemsets mining for dense datasets. In this manner, the calculation develops 1-itemset utility records what's more, investigates the item set search tree to mine the top-k High Utility Itemsets. Two stage strategy for mining top-k High Utility Itemsets.

Yun Sing Kohe et al. [2018] proposed a significant constraint of customary High Utility Item set Mining (HUIM) calculations is that they don't think about that the utility of thing sets may fluctuate after some time. Along these lines, customary HUIM calculations can't discover thing sets that don't yield a high utility when thinking about the entire database. Finding such thing sets is valuable, as an item may sell uncommonly well during explicit timespans yet not during the remainder of the year. Nearby high utility thing set - Miner and Potential high utility thing set- Miner are the calculations proposed to mine these examples. A third calculation named Non-Excess Peak High Utility Item set -Miner is proposed to find a littler arrangement of examples called Non-Excess Peak High Utility Item set. In this paper they propose to find another kind of examples called Local High Utility Itemsets. It comprises of discovering itemsets that yield a utility that is no not exactly a client determined edge during at least one timeframes making some base memories length. An effective calculation called Local High Utility Itemsets - Miner is intended to find Local High Utility Itemsets. It depends on a novel information sssconstruction named Local Utility-list, what's more, expands the fundamental inquiry system and utility-list information structure of the High Utility Itemsets-Miner calculations.

Tin Truong et al. [2018] Proposed finding high normal utility itemsets. High Average Utility Itemsets in a measureable database is a prevalent information removal task, which targets recognizing sets of items bought together that have a high significance or return a high benefit. None the less, a key test of HAUI mining is that the descending conclusion stuff doesn't hold for the normal utility measure. Numerous studies have structured normal utility upper limits, which fulfill the descending conclusion property, to overestimate the normal utilities of thing sets. A few information structures and calculations were proposed dependent on those UBs to find all HAUIs. A key test of High Average Utility Itemset Mining is that the descending conclusion property doesn't hold for the normal utility of item sets. Therefore, to rapidly lessen the pursuit space of High Average Utility Itemset Mining, numerous scientists have structured UBs on the normal utility of item sets, which fulfil the descending conclusion property, to permit pruning unpromising item sets early. In this paper most of its fields are not appropriate for the issue of High Average Utility Itemset Mining. where UBs or then again WUBs should be determined in different manners during the mining process.

Truong et al. [2019] Mining utility-based groupings regularly requires additional time and memory. To beat this issue, this paper proposes two brief portrayals of Frequent High Utility Sequences (FHUS), which have a Cardinality that gives a synopsis of all FHUS. Those portrayals are characterized as two sets, FCHUS and FMaxHUS. To productively mine these exact portrayals, two width and profundity pruning techniques are proposed for taking out low utility groupings early. A thing that makes a high advantage can be considered as increasingly noteworthy. At that point a conspicuous thing that arrival a low advantage the issue of High Utility Sequence Mining is more problematic than the standard issue of FSM. The essential clarification is that the assistance measure used to survey plans in FSM satisfies the Downward Closure Property, which licenses to efficiently decrease the interest space. To get rid of this risk right now, is appealing to use a negative and progressively secure technique for calculating the utility of arrangements. The indistinguishable quality of PDBs expect a noteworthy activity in recognizing and pruning non-shut groupings in front of timetable during the examination of the chase space. Here predominantly two calculations are utilized regular high utility successions (FHUS), FMaxHUS, nearby pruning. Mostly diminish the inquiry space. Proficient as far as time and memory.

Rajshekhar Sunderraman et al. [2019] proposed Chart information mining has been tried in visit mining information set. Then enormous measure of diagram has been created in numerous territories. Versatile diagram information mining is very main stream in light of the fact that because of expanded chart complexities. Frequent sub chart mining implies the succession of digits' rehashes unendingly. A few investigate bunches have end devoured to deal with visit sub diagram mining issue in various ways. Visit design mining and continuous sub diagram mining and graphical preparing units are two calculations they utilized and which has multi-increase development over in-memory while managing enormous database. In this paper they proposed a segment based methodology, ADI-Mine, wherein they made a record structure ADI (contiguousness list). For each edge, they kept up the chart ids in a connected rundown. The base help edge utilized in their methodology is the division of client offered help isolated by k. After neighbourhood mining is finished, a union join system is called to join the results. This area clarifies how the usefulness measure and novel upper-limits on the average utility measure $a\ u$ can be characterized utilizing a vertical structure (sections of the lattice), conflictingly to past examinations that have utilized the level structure.

III. PROPOSED METHODOLOGY

In the proposed technique that we are going to change over the transaction database into sub graph structure. By utilizing high utility sub graph mining calculation and mathematically we are going to assign each and every node. It produces yield as a graph structure. Additionally, in proposed framework, we are planning to reduce the time complexity problem and develop structure with various modules to guarantee the better taking care of for uncertain data. The relationship between a sub graph design and a lot of vertices is characterized by its critical enhancement dependent on a graph mining algorithm. This interesting measure requires a committed pruning strategy to confine the quantity of sub graph that must be determined. The presented mining calculation to discover related sub graph designs in enormous diagrams is intended to proficiently navigate the search space. The activity of this strategy by applying it on transaction database and we can associate sub graphs.

Obtaining the Dataset: An appropriate dataset is looked over well-known stores like UCI, Kaggle, Reuters, etc, or the dataset self-generated, in view of the given domain knowledge.

Dataset Cleaning: Generally required for self-gathered or brought forth dataset, it includes blunder treatment of the dataset by trait adequacy assurance and by distinguishing proof and expulsion of the conflicting models from the preparation set. Writing refers to different techniques for information purifying. Anyway datasets from standard stores as a rule don't require this progression, since the information is typically spotless.

Confidentiality protection by removal of direct and indirect characters: This is the chief advance engaged with the disposal of character data identified with the topic to whom the information relates to. The stages elaborate in this are

(a) Determining and Removing Candidate key attribute(s) **-** Includes recognizable proof of those characteristics which have ethics extraordinary to every datum article or greatest, and not a type of a coasting idea information (since many skimming idea numbers be able to exist among two drifting point numbers.). These properties can be evacuated straight forwardly later it may not be a circumstance that these will add to the classifier structure procedure. This deposits with semi – characteristics.

Competition Selection: An unplanned subgroup of the first dataset is on a level plane picked and over and over picked, a method called competition choice, to make ten to fifteen information subsets relying upon the issue area. The produced information subsets are then dependent upon particular averaging including picking a characteristic aimlessly, and be around a subgroup of the estimations of this quality into the other subset, to guarantee extra degree of protection conservation. These information subgroups are then vertically divided to make k information subgroups from every subgroup.

Classifier Group and Classifier Compounding: The found 10k information subsets are then taken care of into an appropriate choice tree producer segment in a reasonable encoding stage, and 10k sub-classifiers made. Classifier Compounding includes joining of choice sub-classifiers produced fittingly dependent on space information and the sub-classifier need.

Algorithm for Proposed methodology

Input: Weighted Database WD

Output: Consistent subgraph and weighted graph (WD)

Proposed name: Organized weighted sub graph tree

1. Scan Weighted Database of the transactions data, WD of sub graph constructions
2. Filter out sub graph branches with conventional mathematic procedure.
3. Scan the second level of construction in sub graph with pre-defined weight values with branch wise.
4. Established the root node of WD-graph = Null
5. For weighted transaction (WT) in WD do
6. T = fix the transaction data into several branches by using threshold
7. Update the tree value which is dynamically changing the weighted value.

IV. CONCLUSIONS

In this paper, describes on applying information on mining which is advances to the graph mining. Frequent sub graph mining, which contains of learning sub graphs looking regularly in a set of charts. This information mining issue has been read for over 15 years, and numerous calculations have been proposed. Graph mining algorithms are precise calculations (will locate the right answer), while some other are inexact calculations (don't ensure to locate the right answer, yet might be quicker). Graph mining algorithms are likewise intended to deal with coordinated or undirected charts, or mine sub charts in a solitary graph or in a graph database, or can do both. In addition, there exists a few different varieties of the sub diagram mining issue, for example, finding continuous ways in a chart, or successive trees in a diagram. Which minimize the time and also Improvising transaction search in transaction data.

REFERENCES

1. Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.
2. Liu, D., Yan, Z., Ding, W., & Atiquzzaman, M. (2019). A Survey on Secure Data Analytics in Edge Computing. IEEE Internet of Things Journal.
3. Truong, T., Duong, H., Le, B., & Fournier-Viger, P. (2018). Efficient vertical mining of high average-utility item sets based on novel upper-bounds. IEEE Transactions on Knowledge and Data Engineering, 31(2), 301-314.
4. Jena, B. S., Khan, C., & Sunderraman, R. (2019). High performance frequent subgraph mining on transaction datasets: A survey and performance comparison. Big Data Mining and Analytics, 2(3), 159-180.
5. Lin, J. C. W., Ren, S., Fournier-Viger, P., Pan, J. S., & Hong, T. P.(2018). Efficiently updating the discovered high average-utility item sets with transaction insertion. Engineering Applications of Artificial Intelligence, 72, 136-149.
6. Krishnamoorthy, S. (2019). Mining top-k high utility item sets with effective threshold raising strategies. Expert Systems with Applications, 117, 148-165.
7. Suresh. K and Pattabiraman. V, “An improved utility itemsets mining with respect to positive and negative values using mathematical model”, International Journal of Pure and Applied Mathematics, Volume-101 No-5, Page No-763772, 2015.
8. Anusha, Koyi & C, Yashaswini & Sankar, Manishankar. (2016). Segmentation of Retail Mobile Market Using HMS Algorithm. International Journal of Electrical and Computer Engineering 6(4):1818, August 2016.
9. Sindhushree B, Manishankar S, Dhanushya B P, Cloud Based Healthcare Framework for Criticality Level Analysis, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-1, May 2019
10. P. Rakshith, S. Manishankar and P. Sushmitha, "Enterprise data analytics and processing with an integrated hadoop and R platforms," 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017, pp. 1-5.

11. R. Baker, A. Corbett, and K. Koedinger, “Detecting student misuse of intelligent tutoring systems,” in Proc. Int. Conf. Intell. Tutoring Syst., Alagoas, Brazil, 2004, pp. 531–540.
12. R. Baker, “Modeling and understanding students’ off-task behavior in intelligent tutoring systems,” in Proc. Conf. Hum. Factors Comput. Syst., San Jose, CA, 2007, pp. 1059–1068.
13. R. Baker, J. E. Beck, B. Berendt, A. Kroner, E. Menasalvas, and S. Weibelzahl, “Track on educational data mining,” presented at the 11th Int. Conf. User Model. Workshop Data Mining User Model, Corfu, Greece, 2007.
14. Suresh, K., and Devika Mohan. "Development of High Utility Itemsets In Streaming Database." TEST Engineering & Management 82 (2020): 13052-13056.
15. Suresh, K., and V. Pattabiraman. "Reduction Of Large Database And Identifying Frequent Patterns Using Enhanced High Utility Mining." International Journal of Pure and Applied Mathematics 109.5 (2016): 161-169.
16. R. Baker and K. Yacef, “The state of educational data mining in 2009: A review and future visions,” J. Educ. Data Mining, vol. 1, no. 1, pp. 3–17, 2009.
17. R. Baker, “Data mining for education,” in International Encyclopedia of Education, B. McGaw, P. Peterson, and E. Baker, Eds., 3rd ed. Oxford, U.K.: Elsevier, 2010.
18. R. Baker, A. Merceron, and P. I. Pavilk, presented at the 3rd Int. Conf. Educ. Data Mining, Pittsburgh, PA, 2010.
19. H. Ba-Omar, I. Petrounias, and F. Anwar, “A framework for using web usage mining for personalise e-learning,” in Proc. Int. Conf. Adv. Learn. Technol., Niigata, Japan, 2007, pp. 937–938.
20. D. Barker-Plummer, R. Cox, and R. Dale, “Dimensions of difficulty in translating natural language into fist order logic,” in Proc. Int. Conf. Educ. Data Mining, Cordoba, Spain, 2009, pp. 220–228.
21. T. Barnes, “The q-matrix method: Mining student response data for knowledge,” in Proc. AAAI Workshop Educ. Data Mining, Pittsburgh, PA, 2005, pp. 1–8.
22. T. Barnes and J. Stamper, “Toward automatic hint generation for logic proof tutoring using historical student data,” in Int. Conf. Intell. Tutoring Syst., Montreal, QC, Canada, 2008, pp. 373–382.
23. T. Barnes, M. Desmarais, C. Romero, and S. Ventura, presented at the 2nd Int. Conf. Educ. Data Mining, Cordoba, Spain, 2009.
24. C. B. Baruque, M. A. Amaral, A. Barcellos, J. C. Da Silva Freitas, and C. J. Longo, “Analysing users’ access logs in Moodle to improve e learning,” in Proc. Euro Amer. Conf. Telematics Inf. Syst., Faro, Portugal, 2007, pp.1–4.
25. C. R. Beal and P. R. Cohen, “Temporal data mining for educational applications,” in Proc. 10th Pacific Rim Int. Conf. Artif. Intell.: Trends Artif. Intell., Hanoi, Vietnam, 2008, pp. 66–77.
26. J. E. Beck, presented at the 5th Int. Conf. Intell. Tutoring Syst. (ITS) Workshop Applying Mach. Learning to ITS Design/Construction, Montreal, QC, Canada, 2000.
27. J. E. Beck, and B. P. Woolf, “High-level student modeling with machine learning,” in Proc. 5th Int. Conf. Intell. Tutoring Syst., Alagoas, Brazil, 2000, pp. 584–593.
28. H. Mannila, H. Toivonen, and A. I. Verkamo, Discovery of frequent episodes in event sequences, Data Min. Know. Disc., vol. 1, no. 3, pp. 259–289, 1997.
29. R. Agrawal and R. Srikant, Mining sequential patterns, in Proc. Eleventh Int. Conf. Data Engineering, Washington, DC, USA, 1995, pp. 3–14.

30. T. Asai, K. Abe, S. Kawasoe, H. Arimura, H. Sakamoto, and S. Arikawa, Efficient substructure discovery from large semi-structured data, in Proc. 2002 SIAM Int. Conf. Data Mining, Arlington, VA, USA, 2002.
31. M. J. Zaki, efficiently mining frequent trees in a forest, in Proc. Eighth ACM SIGKDD Int. Conf. Knowledge
32. Discovery and Data Mining, 2002, pp. 71–80.
33. R. C. Agarwal, C. C. Aggarwal, and V. V. V. Prasad, A tree projection algorithm for generation of frequent item sets, J. Parall. Distrib. Comput., vol. 61, no. 3, pp. 350–371,2001.
34. M. Kuramochi and G. Karypis, Frequent subgraph discovery, in Proc. IEEE Int. Conf. Data Mining, San Jose,CA, USA, 2001, pp. 313–320.
35. [34] X. Yan, Mining, indexing and similarity search in large graph data sets, PhD dissertation, University of Illinois at Urbana-Champaign, IL, USA, 2006.
36. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to Algorithms. Boston, MA, USA: MIT Press and McGraw-Hill, 2001.
37. Savasere, E. Omiecinski, and S. B. Navathe, An efficient algorithm for mining association rules in large databases, in Proc. 21th Int. Conf. Very Large Data Bases, San Francisco, CA, USA, 1995, pp. 432–444
38. Wang, W. Wang, J. Pei, Y. T. Zhu, and B. L. Shi, Scalable mining of large disk-based graph databases, in Proc. Tenth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, New York, NY, USA, 2004, pp. 316–325.
39. J. M. Wang, W. Hsu, M. L. Lee, and C. Sheng, A partition-based approach to graph mining, in Proc. 22nd Int. Conf. Data Engineering, Atlanta, GA, USA, 2006, p. 74.
40. S. N. Nguyen, M. E. Orlowska, and X. Li, Graph mining based on a data partitioning approach, in Proc. Nineteenth Conf. Australasian Database, Darlinghurst, Australia, 2007, pp. 31–37.
41. S. N. Nguyen and M. E. Orlowska, Improvements in the data partitioning approach for frequent itemsets mining, in Knowledge Discovery in Databases: PKDD 2005, A. M. Jorge, L. Torgo, P. Brazdil, R. Camacho, and J. Gama, eds. Springer, 2005, pp. 625–633.
42. S. Chakravarthy, R. Beera, and R. Balachandran, DB- Subdue: Database approach to graph mining, in Advances in Knowledge Discovery and Data Mining, H. Dai, R. Srikant, and C. Zhang, eds. Springer, 2004, pp. 341–350.
43. J. Cook and L. B. Holder, Graph-based data mining, IEEE Intell. Syst. Their Appl., vol. 15, no. 2, pp. 3241, 2000.