Data Mining: Application, Handling and Future

Junaid Khan, Amin ul Haq, Muhammad Ishfaq Ahmad Khan, Waseem Khan, Ihsan Ilahi Jan and Fayyaz Gul
Gandahara University of Sciences, Peshawar, Pakistan

***Abstract*– Data mining is a chunk of a process/procedure called KDD stands for knowledge discovery in current databases. This practice entails fundamentally of phases that are accomplished earlier/before performing data mining, such as data assortment, data clean-up, pre-processing, and data renovation. Association rule methods are utilized for data mining if the objective is to discover relationships/associations between specific/particular values of uncompromising variables in bulky data sets. There might be thousands or millions of records that have to be comprehended and to take out the procedures for, but the issue is what will ensue if there is novel data, or there is a necessity to amend or obliterate some or all the prevailing set of data during the process of data mining. Bygone consumer/user would reprise the whole/entire procedure/practice, which is time-consuming in addition to its lack of efficiency. From this, the importance of dynamic data mining process appears and for this reason this problem is going to be the main topic of this paper. Therefore the intention of this revision is to uncover clarification for dynamic/active data mining process that is capable of to take into considerations all updates (insert, update, and delete problems) into account.**

Keywords **–** Applications, Handling, Association and Data Mining

I. INTRODUCTION

Data mining is an assignment of discerning fascinating and concealed patterns/shapes from hefty amounts of data where the data can be accumulated in databases, data warehouses, OLAP (on line analytical process) or other repository information [1]. It is also defined as knowledge discovery in databases (KDD) [2]. Data mining involves an integration of techniques from multiple disciplines such as database technology, statistics, machine learning, neural networks, information retrieval, etc [3]. According [4]: "Data mining is the process of discovering meaningful patterns and relationships that lie hidden within very large databases". Also [5] defines Data mining as "the analysis of observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner". Data mining is a part of a process called KDD-knowledge discovery in databases [3].

This process consists basically of steps that are performed before carrying out data mining, such as data selection, data cleaning, pre-processing, and data transformation [6]. The architecture of a typical data mining system may have the following major components [3]: database, data warehouse, or other information repository; a server which is responsible for fetching the relevant data based on the user's data mining request, knowledge base which is used to guide the search. Data mining engine consists of a set of functional modules,

Pattern evaluation module which interacts with the data mining modules so as to focus the search towards interesting patterns and graphical user interface which communicates between users and the data mining system, allowing the user interaction with system. The basic Data Mining Tasks consists of a number of processes:

- Time series analysis: [5], [7], [8].
- Association analysis: [9] - [13].
- Classification: [3], [5], [8].
- Regression: [3], [5].
- Cluster analysis: [6], [8], [10].
- Summarization: [5].

Data mining is one of the most important research fields that are due to the expansion of both computer hardware and software technologies, which has imposed organizations to depend heavily on these technologies [14]. Data is considered as the number one asset of any organization, it is obvious that this asset should be used to predict future decisions [15]. Consequently, and since organizations are continuously growing, their relative databases will grow as well; as a result their current data mining techniques will fail to cope up with large databases which are dynamic by nature [16]. Data mining is the way to help organization make full use of the data stored in their databases [35], and when it comes to decision making, this is true in all fields, and is also true in all different types of organizations. Databases tend to be large and dynamic thus their contents usually do change; new information might need to be inserted, current data might need to be updated and/or deleted.

The problem with this from the data mining perspective is how to ensure that the rules are up-to-date and consistent with the most current information. Also the learning system has to be time-sensitive as some data values vary over time and the discovery system is affected by the correctness of the data. In this respect and in view of what have been introduced regarding dynamic data mining and its importance and its effects on decision making. It is our intention to put forward a solution in order to run data mining without the need to restart the whole process every time there are changes on the data being used, in other words the running process should focus solely on the amendments taking into consideration that the mining run is held constant.

II. MINING PROCESS OF STATIC DATA

Data mining process is a step in Knowledge Discovery Process consisting of methods that produce useful patterns or models from the data [3]. In some cases when the problem is

known, correct data is available as well, and there is an attempts to find the models or tools which will be used, some problems might occur because of duplicate, missing, incorrect, outliers values and sometimes a need to make some statistical methods might arise as well. Kenji et al [17] modeled semi-structured data and patterns by labeling ordered trees and studied the problem of discovering all frequent tree-like patterns that have at least min-sup support in a given collection of semi-structured data. They represented an efficient pattern mining algorithm FREQT for discovering all frequent tree patterns from large collection labeled ordered trees. Raedt and Kersting [18] identified some of the key concepts and techniques underlying probabilistic logic learning. And explained the differences between the various approaches and at the same time provide insight into some of the remaining challenges in probabilistic logic learning.

The techniques of probabilistic logic learning were analyzed starting from a logical (or inductive logic programming) perspective. Furthermore, principles of both statistical learning and inductive logic programming (or multi-relational data mining) are employed for learning the parameters and structure of the probabilistic logics considered. Jin and Ag awal [10] presented the design and initial performance evaluation of a middleware, enabling rapid development of the parallel data mining applications, which can help exploit parallelism on both shared memory and distributed memory configurations. They studied parallel versions of methods, in each of these methods; parallelization can be achieved by dividing the data instances (or records or transactions) among the nodes.

The computation on each node involves reading the data instances in an arbitrary order, processing each data instance, and performing a local reduction. The KDD procedures are explained bellow in a way to help us focus on data mining process. It includes five processes: 1) Defining the data mining problem, 2) Collecting the data mining data, 3) Detecting and correcting the data, 4) Estimating and building the model and 5) Model description, and validation as seen in Fig. 1.

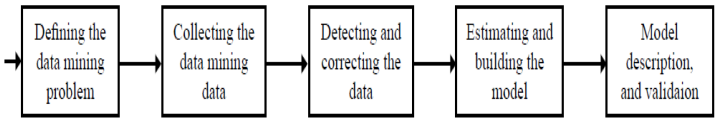

Fig. 1: Data mining process

A. Defining Data Mining Problem

Most data-based modeling/sculpting studies/research is carried/performed out just for a specific/particular application domain. So from this place, domain-specific acquaintance and involvement are frequently compulsory just for the reason to come up with an eloquent problem/issue statement. But unfortunately, numerous application amendments are projected to be engrossed on the concern data mining method/procedure at the charge/cost of a vibrant/clear problem/issue statement. In this step/phase, a modeler generally stipulates a set/series of variables for the unrevealed enslavement and, if possible, a common form of this enslavement as a preliminary hypothesis. There might be numerous hypothesis articulated for a solitary problem/issue at this stage/phase [19]. The principal step/phase entails the pooled capability of an application domain and a data mining model. In efficacious data mining applications, this collaboration does not halt/stop in the early phase; it endures throughout the all-inclusive data mining process, the compulsion/obligation to knowledge discovery is to comprehend data and business [9]. Deprived of this understanding, no algorithm, irrespective of complexity, is capable or able to offer/provide result/outcome that can be confident and clear.

B. Data Mining Data Collection

This process is apprehensive with the assortment of data from diverse sources and locations/places. The contemporary methods/procedures used/exercised for collection data are:

•*Interior/Internal Data:* data are customarily collected/gathered from pre-existing databases, data warehouses, and OLAP. Genuine transactions logged by individuals are the richest source of information, and at that time, the most perplexing to be worthwhile

• *Exterior/external Data:* data items can be collected/gathered from psychographics, Demographics, and web graphics. In addition to data shared/pooled within a company.

C. Data Detection and Correction

All raw data groups which are primarily equipped for data mining are frequently large/bulky; countless are interrelated to humans and also have the potential/capability for being cluttered [19]. Real-world databases are subject to missing, noise, and fickle data due to their stereotypically mammoth size, habitually quite a lot of gigabytes or more. Data preprocessing is generally used/utilized as a maiden data mining practice/work. It renovates the data into such kind of a layout which will be effortlessly and meritoriously handled by the users/consumers. There are a number of data preprocessing techniques/procedures which include: Data cleaning; that can be smeared to eradicate noise and spot-on inconsistencies, outliers and missing values. Data integration; amalgamates data from manifold sources into a lucid data store, such as a data warehouse or a data cube [19].

Data transformations, such as normalization, may be applied/smeared; normalization rallies the exactitude and adeptness of mining algorithms encompassing aloofness measurements. Data reduction; can diminish the data size by aggregating, jettisoning fired features. The data Processing techniques/methods, when smeared erstwhile to mining, can expressively mend the overall data mining results/outcomes. [3]. Since, manifold data groups may be used/utilized in innumerable transactional formats, extensive data groundwork might be compulsory [20]. There are innumerable profitable software products that are explicitly premeditated for data preparation, which can assist the mission of organizing the data erstwhile to importing it into a data mining tool.

D. Estimating and Building the Model

This process embraces four main parts: 1) select/choose data mining task/job, 2) select/choose data mining method/procedure, 3) select/choose appropriate algorithm, and 4) excerpt knowledge as can be seen in Fig. 2. The Fig. 2 shows that this development is alienated into four parts these are:

Data mining task (s) selection: Selecting/choosing which chore to use/utilize be contingent on the model whether it is extrapolative or evocative [3], [5], [9].extrapolative models envisage the values of data exhausting known results and/or information originated hefty data sets, historical data, or exhausting some variables or fields in the data set to envisage unknown, classification, regressions, time series analysis, prediction, or estimation are tasks/duty for predictive model [2]. A evocative model recognizes shapes or relationships in data and obliges as a way to sightsee the belongings of the data examined/scrutinized. Clustering, summarization, association rules and structure discovery are habitually observed as evocative [5]. The comparative prominence of prophecy and portrayal for exact/precise data mining applications can differ considerably. That means picking which job to use/utilize is contingent on the model whether it is extrapolative or descriptive.

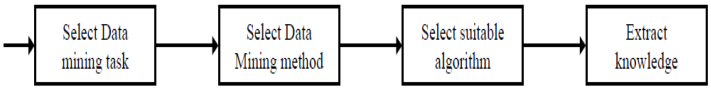

Fig. 2: Estimating and building the model

Data mining method (s) selection: After choosing which task/duty we can select the method or manner and assuming/supposing we have a prognostic model and the task/job is classification or cataloging whilst the apprehension routine is Rule Induction, with Neural Network or Decision tree. In most exploration in this filed; intellectuals appraises the appropriate exemplary. This model to vintage/yield acceptable/applicable results. There are quite a few number of methods/plans for model calculation encompasses these but not only limited to neural networks, Decision trees, Association Rules, Genetic algorithms, Cluster Detection, Fuzzy Logic.

Algorithm selection: Next phase is to build/construct an exacting algorithm which implements/outfits the general or common methods. Each and every data mining algorithm embraces/comprises three major components these are: i) Model representation, ii) model evaluation, and iii) search [2].

Knowledge Extraction: This is the last/final stride in constructing the model which is the results or outcome (or the answers/solutions for the problem solved in data mining) after fashioning the simulation for the algorithm. This can be preeminent elucidated by bestowing an example of Auction Fraud [20].

E. Validation, Description of Model

In each and every case, data mining models ought to assist/help customers in making the conclusions. From this place, such models need/require to be interpretable in order to be beneficial because humans are not anticipated to sordid their verdicts on multifaceted black-box models; the penalty area of the perfection of the model and exactness of its interpretation are marginally contradictory. Contemporary data mining methods/plans are liable to yield/grow exceptionally clear-cut results/outcomes exhausting high dimensional models [5]. The problem/difficulty of interpreting of taking into mean these models, are exceedingly imperative and is considered as a diverse task/job with certain/vague techniques/methods to authenticate the results [19]. Model legitimacy is an indispensable but inadequate ailment for the trustworthiness and satisfactoriness of data mining results [21]. If, for example, the maiden purposes are erroneously renowned or the data set is indecorously specified, the data mining results/outcomes articulated through the model will not be useful; however, we might still ascertain the model valid [20]. One constantly has to preserve in mind, that a delinquent appropriately uttered is an unruly half-solved. The absolute objective of a data mining process should not be just to yield/develop a model for a difficulty/problem at hand, but also to deliver one that is adequately credible, suitable and implemented by the decision-makers; this type would need to consider all the data i.e. using a dynamic database.

III. DYNAMIC/ACTIVE DATA MINING PROCESS/PROCEDURE

Just because of continuous, unbounded, and excessive rapidity individualities of dynamic/active data, there is an enormous amount of data in both offline and online data streams, and thus, no sufficient time to rescan the entire database or accomplish a rescan as in outmoded data mining algorithms each and every time an update occurs. In Addition, no adequate space to stock/store each and every stream data for online processing. As stated former many researchers and developers have quantified a procedure model envisioned to escort the customer through a series of steps/phases that will lead/direct to upright results/outcomes. Several have been pronounced for data mining process. Little of them whispered that this is plausible for Dynamic/active data mining process. Ganti et al [22] scrutinize mining of data streams.

A block development model is familiarized where a data set is rationalized/updated intermittently through additions and deletions. In this model, the data set entails abstractly immeasurable sequence/series of data blocks D1, D2, ... that attain at times 1, 2, ... where each block has a set of records. Some applications oblige mining all of the data encountered thus far (unrestricted window scenario), whilst others oblige/requires mining only the most recent part (restricted window scenario) and updating the models accordingly. The authors highlight/spots down two challenges/tests in mining sprouting blocks of data: change detection and data mining model maintenance. In modification detection, the dissimilarities between two data blocks are determined/cleared. Next, a data mining model ought to be

maintained/sustained under the insertions/additions and erasures/deletions of blocks of the data rendering to a specified/particular data span/width and block selection/choosing sequence. Crespoa, and Weberb [6] offered/vacant a methodology/procedure for dynamic/active data mining utilizing fuzzy clustering which consigns static/stagnant objects to dynamic/active classes. Alterations that they have planned/proposed are movement, creation, and abolition of classes and any of their or other combination. Chung and Mcleod [23] recommended mining framework that manacles the citations of controllable outlines purely based on incremental data clustering, they engrossed their gallantry on newsflash stream mining, and also they offered a sophisticated incremental hierarchical/tree document/deed clustering algorithm using/utilizing a neighborhood search. Zaslavsky et al [8] debated data stream mining and the also significance of its applications; the planned/proposed techniques have their roots in statistics and hypothetical computer science.

Data-based techniques/procedures and task-based techniques are the two main classes of data stream mining algorithms. And based on these two main categories/classes, a number of clustering, classification, and frequency counting and time series analysis have been developed/evolved. Babu et al [24] concerted on the problem of query processing, specifically on how to prompt and evaluate incessant queries over data streams, address semantic issues/matters as well as proficiency concerns; they quantified a common and supple architecture for query processing in the existence of data streams. They also used/utilized their rudimentary architecture as tool to elucidate substitute semantics and processing techniques for continuous queries. They plotted out research topics in the area/field of query processing over data streams. Reigrotzki et al [25] offered the application of several process control-related methods smeared in the context of monitoring and governing data quality in financial databases.

They exhibited that the quality control process can be measured as a classical control loop which can be restrained via application of excellence tests which exploit/employs data redundancy defined by meta information or took out from data by statistical models. Applicable processing and visualization of the tests results empower human or automatic detection and verdict/diagnosis of data quality problems. Moreover, the model-based methods provide discernment into business-related information enclosed in the data. The methods have been smeared to the data superiority observing of a real financial database at a customer/user site, supplying business benefits, such as enhancements of the demonstrating/modeling quality, a decline in the number of the modeling cycles, and healthier data understanding. These benefits/pros in turn lead/directs to financial savings.

In numerous situations/cases, new/novel information is much more significant as compared to old information, such as in publication database, stock transactions, grocery markets, or web-log records. Consequently, a recurrent item set in the dynamic/active database is also imperative even if it is intermittent in the updated database [27]. Once a data mining system is mounted/installed and is being used/utilized in everyday operations, the customer has to be concerned with the system's future enactment because the extracted knowledge is based on past/formal behavior of the scrutinized objects [6]. If future enactment is very comparable to past performance/enactment (e.g. if company clients files do not change/modify their files over time) using/utilizing the primary data mining system could be justified. If, however, performance/enactment changes/switch over time (e.g. if hospital patients do not change their files over time), the continued use/utilization of the premature system could lead/directs to an inappropriate results and (as an effect) to an intolerable decisions based on these results [5].

Here is where dynamic/actives data mining comes into play by offering/providing logical appropriate techniques for "updating". In practice, and beholding to experiential cases, dynamic/active data mining could be tremendously cooperative in constructing/making the exact decision in the veracious time and affects the effectiveness of the decision as well [1]. It becomes obvious/clear, that something has to be done if a user/customer is to keep smearing his/her data mining system in a changing environment. In this case basically, there are three strategies [6]:

i). User can disregard/neglect changes/modifications in the environment and keeps on applying the preliminary system without any additional updates. It has the benefit of being computationally cheap since no update/modification to data mining system is performed. Also it does not entail changes in following processes. Its detriment is that contemporary updates could not be detected.

ii). Each and Every definite period of time, liable on the application, a new/novel system is developed using/utilizing all the accessible data. The benefit in this case is the customer has always a system "up-to-date" due to the usage of contemporary data. Drawbacks of this approach are the computational costs of generating a new/novel system every time from scratch.

iii). Based on the initial/primary system and "new data" an update of data is performed. This will be shown to be obtainable technique in this dissertation. In the field of data mining numerous approaches have been established in order to discover valuable information patterns within data. Amid the most significant approaches are connotation/association rules, clustering, and decision trees methods [5]. For each of the above data mining methods, updating has diverse features and some updating methodologies have been proposed/suggested:

• Decision trees: Innumerable procedures for incremental learning and tree reorganization

• Neural networks: Updating is frequently used/utilized in the nous of re-learning or refining the net's enactment by erudition with new/novel examples offered to the network.

• Clustering: [23, 28] defines in additional comprehensive attitudes for dynamic/active data mining using/utilizing clustering techniques.

• Association rules [13], [29]-[34]: Raghavan et al. established a system for dynamic/active data mining for association rules [11]. This sector will announce the diverse facets that might affect the consequence of data mining run if the data used/utilized for the run are to be changed or removed during this run.

A. Defining Data Mining Problem

Since this procedure is concerned with the description of the problem/matter and the data is not yet built/constructed, no modification will be performed in this regard, except the predictor modifies the unruly goals.

B. Data Mining Data Collection

This practice is concerned with the data gathering from diverse cradles/sources and/or locations, if we assume that an update was carried out on the date after the data was collected by the algorithm, the following will take place:

i). If a new/novel source of data (new database) and the data encompassed in this database is a foremost source, this source will be then used/utilized. This can be attained by reminiscing this data again, and/or substitute prevailing data somewhat or totally/entirely. If part of the data in the new/novel source is more interrelated to the problem/issue in hand the appropriate part in the new/novel source will swap the older part whilst keeping the similar source.

ii). New updated data (insert, update, delete) If the data used/utilized is changed/modified by means of any way during a contemporary run, then the data being used/utilized will be take into account as an unacceptable and the new/novel version of this data ought to be collected, a new/fresh run will be initiated/invoke, with the similar cradle of data, but it ought to be update the data straightaway either by inserting/adding new data, updating current data, or deleting data completely/entirely.

C. Data Detection and Correction

If it is realized that the current data is changed in any way, the following steps will be applicable:

i). If it is a new source of data and there is a need to recollect the data, this takes us back to the Collecting data process.

Or:

ii). If it is a new source of data and there is a need to include this data then, there is no need to go back, from the beginning, it should simply make small database containing the new source of data and decide which of the new data items it needs; it should then carry out data cleaning to remove noise, correcting inconsistencies, integration (merges data from multiple sources into a coherent data store), Data transformations, such as normalization, Data reduction where we can reduce the data size. Afterwards, it simply combines the new set of data (the detected and corrected data) with the main data. Note here some times after combining data one could have another detecting and correcting process i.e. (data reduction or integration) but this will not flow the same procedures as starting from scratch Or:

iii). If it is a new updated version of the data, the steps mentioned above (cleaning, correcting inconsistencies, data transformations) can be used if and only if it is necessary and then combine the new updated data with the main data, taking into consideration that this step is concerned with inserting new record, updating, and/ or deleting an existing record.

D. Estimating and Building the Model

In resembling and constructing the model process in dynamic data mining; we have three main parts: choosing Data Mining task (s), choosing Data Mining method (s), and selecting the suitable algorithm that will not change or switch over if a new/novel version of data is announced (since they are concerned with the delineation of the problem). But here in this case the pulling out knowledge part will change liable on the new/novel version of the data, and whether it is an insert/add, update and/or delete case. If a new/fresh source of data appears during this process one of the steps stated in section 3.3 will be applied conferring to the significance to the data.

E. Validation, Description of Model

Contemporary data mining approaches are anticipated to crop extremely truthful results/outcomes [5] using/utilizing high dimensional models, with novel reorganized version of the data, it will assuredly change or modify the results/outcomes; say for example the number of rules in association rules might be changed; it might enhance a new rules, substitute prevailing number of rules or change/modify the proportion of the rules, and this modification might disturb the process of decision making.

IV. CONCLUSION

It is clear/obvious from all that have been assumed that data mining is still in its infancy, or at the commencement of the road as there are many facets of data mining that have not been tested/justified [37]. Up-to-date most of the data mining projects have been allocating with authenticating the authentic data mining concepts. Since this has now been recognized most researchers will move into resolving some of the problems that stance in the way of data mining, this research will pact with such a problem, in this case the research is to concentrate on solving the problem of using data mining dynamic databases.

REFERENCES

[1] Maria Halkidi, "Quality assessment and Uncertainty Handling in Data Mining Process" http://www.edbt2000.unikonstanz.de/phd-workshop/papers/Halkidi.pdf.

[2] Fayyad, U. M., G. P. Shapiro, P. Smyth. "From Data Mining to Knowledge Discovery in Databases", 0738-4602-1996, AI Magazine (Fall 1996): 37–53.

[3] Jiawei Han, Micheline Kamber. "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, Champaign: CS497JH, fall 2001, www.cs.sfu.ca/~han/DM_Book.html.

[4] Claude Seidman. "Data Mining with Microsoft SQL Server 2000 Technical Reference", ISBN: 0-7356-1271-4,amazon.com/Mining-Microsoft-Server-Technical-Reference/dp/0735612714.

[5] David Hand, Heikki Mannila, Padhraic Smyth. "Principles of Data Mining", ISBN: 026208290 MIT Press, Cambridge, MA, 2001.

[6] Fernando Crespoa, Richard Weberb. "A methodology for dynamic data mining based on fuzzy clustering", Fuzzy Sets and Systems 150 (2005) 267–284.

[7] Papadimitriou, J.Sun, C.Faloutsos, "Streaming Pattern Discovery in Multiple Time-Series", Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005,p697-708.
[8] Mohamed Medhat Gaber, Arkady Zaslavsky and Shonali Krishnaswamy. "Mining Data Streams: A Review", VIC3145, Australia, ACM SIGMOD Record Vol. 34, No. 2; June 2005.
[9] Two Crows Corporation. "Introduction to Data Mining and knowledge Discovery", ISBN: 1-892095-02-5,
[10] Ruoming Jin and Gagan Ag awal. "A Middleware for Developing Parallel Data Mining Applications", Proc. of the 1-stSIAM Conference on Data Mining, 2000 - cs.ubc.ca.
[11] V. Raghavan, A. Hafez. "Dynamic Data Mining", in: R. Loganantharaj, G. Palm, M. Ali (Eds.), Intelligent Problem Solving Methodologies and Approaches: Proc. of Thirteenth International Conference on Industrial Engineering Applications of AI & Expert Systems, Springer, New York, 2000, pp. 220–229.
[12] L. Aaron Ceglar and J.Roddick, "Incremental Association Mining using a Closed-Set Lattice", Journal of Research and Practice in Information Technology, Vol. 39, No. 1, February 2007, 35-45.
[13] D.W. Cheung, J. Han, V.T. Ng, and C.Y. Wong, "Maintenance of discovered association rules in large databases: An incremental updating approach," The Twelfth IEEE International Conference on Data Engineering, pp. 106-114, 1996.
[14] R. Agrawal, T. Imielinksi and A. Swami, "Mining association rules between sets of items in large database," The ACM SIGMOD Conference, pp. 207-216, Washington DC, USA, 1993.
[15] T. Fukuda, Y. Morimoto, S. Morishita and T. Tokuyama, "Mining optimized association rules for numeric attributes," The ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 182-191, 1996.
[16] P. Domingos and G. Hulten. "Mining high-speed data streams". In Proc. of the 2000 ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, pages 71–80, August 2000.
[17] Tatsuya Asai, Kenji Abe, Shinji Kawasoe, Hiroki, Arimura, Hiroshi Sakamoto, and Setsuo Arikawa. "Efficient Substructure Discovery from Large Semi-structured Data", SDM, 2002 - siam.org.
[18] Luc De Raedt, Kristian Kersting. "Probabilistic Logic Learning", ACM-SIGKDD Explorations: 2003 - portal.acm. org.
[19] Mehmed Kantardzic J. B. "Data Mining: Concepts, Models, Methods, and Algorithms", ISBN: 0471228524, IEEE Computer society, Wiley-Interscience, Hoboken, NJ, 2003.
[20] Jesus Mena. "Investigative Data Mining for Security and Criminal Detection", First Edition, amazon.com/Investigative-Mining-Security-Criminal-Detection/dp.
[21] Daniel T. "Discovering knowledge in data: an introduction to data mining", Larose, p. cm, ISBN 0-471-66657-2 (cloth), QA76.9.D343L38 2005 006.3_12—dc22.
[22] V. Ganti, J. Gehrke, R. Ramakrishnan, andW. Loh. "Mining data streams under block evolution". In ACM SIGKDD Explorations 3(2):1-10, 2002.
[23] Seokkyung Chung and Dennis Mcleod. "Dynamic Pattern Mining: An Incremental Data Clustering Approach", www.sigmod.org/ dblp/db/indices/a-tree/m/McLeod:Dennis.html.
[24] Babu, S. and Widom, J. (2001). "Continuous Queries over Data Streams", Stanford University, SIGMOD Record, SIGMOD Record, 30:109–120.
[25] Janusz Milek, Martin Reigrotzki, Holger Bosch, Frank Block. "Monitoring and Data Quality Control of Financial Databases from a Process Control Perspectivem", PREDICT AG, Reinach BL, IQ, 2001 - finscore.com.
[26] LTC Bruce D. Caulkins, J.Lee, M.Wang, "A Dynamic Data Mining Technique for Intrusion Detection Systems", 43rd ACM Southeast Conference, March 18-20, 2005, Kennesaw, GA, USA. Copyright 2005 ACM 1-59593-059- 0/05/0003.
[27] A.GIONIS, H.MANNILA, P.TSAPARAS, "Clustering Aggregation", ACM Transactions on Knowledge Discovery from Data, Vol. 1, No. 1, Article 4, Publication date: March 2007.
[28] N. Jiang, L. Gruenwald. "Research Issues in Data Stream Association Rule Mining" SIGMOD Record, Vol. 35, No. 1, Mar. 2006.
[29] G. Cormode, "Fundamentals of Analyzing and Mining Data Streams", WORKSHOP ON DATA STREAM ANALYSIS – San Leucio, Italy - March, 15-16, 2007.
[30] A. CEGLAR, J. RODDICK, "Association Mining", ACM Computing Surveys, Vol. 38, No. 2, Article 5, Publication date: July 2006.
[31] Tzung-Pei Hong, Ching-Yao Wang, Yu-Hui Tao. "A new incremental data mining algorithm using pre-large itemsets" Intelligent Data Analysis, Issue: Volume 5, Number 2 / 2001, pages: 111–129.
[32] TZUNG-PEI HONG, TZU-JUNG HUANG. "Maintenance of Generalized Association Rules for Record Deletion Based on the Pre-Large Concept", Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, February 2007.
[33] Maintenance of Fast Updated Frequent Trees for Record Deletion Based on Prelarge Concepts Chun-Wei Lin1, Tzung-Pei Hong2, Wen-Hsiang Lu1, and Chih-Hung Wu2 July 2007.
[34] R. Agrawal and R. Srikant, "Fast algorithm for mining association rules," The International Conference on Very Large Data Bases, pp. 487-499, 1994.
[35] O.Nasraoui1, C.Rojas, C.Cardona, "Data Mining Meets Evolutionary Computation: A New Framework for Dynamic and Scalable Evolutionary Data Mining based on Non-Stationary Function Optimization", Journal of Computer Sciences, 56-63, 2005 ISSN, © 2005 Science Publications.
[36] N. Jiang, L. Gruenwald. "Research Issues in Data Stream Association Rule Mining" SIGMOD Record, Vol. 35, No. 1, Mar. 2006