Stream Data Mining

Hebah H. O. Nasereddin
Department of computer Information system
Faculty of IT
Middle East University
Amman – Jordan
Hebah66@yahoo.com

ABSTRACT: *Data mining is a part of a process called KDD-knowledge discovery in databases. 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. Association rule techniques are used for data mining if the goal is to detect relationships or associations between specific values of categorical variables in large data sets. There may be thousands or millions of records that have to be read and to extract the rules for, but the question is what will happen if there is new data, or there is a need to modify or delete some or all the existing set of data during the process of data mining. In the past user would repeat the whole procedure, 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 purpose of this study is to find solution for dynamic data mining process that is able to take into considerations all updates (insert, update, and delete problems) into account.*

Key words: Static data mining process, dynamic data, data mining, data mining process, dynamic data mining process

Received: 5 February 2011, Revised 28 March 2011, Accepted 2 April 2011

© 2011 DLINE. All rights reserved

1. Introduction

Data mining is the task of discovering interesting and hidden patterns from large amounts of data where the data can be stored 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, 10, 11, 12, 13].
- **Classification**: [3, 5, 8].
- **Regression**: [3, 5].
- **Cluster analysis**: [6, 8, and 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.

2. Static Data Mining Process

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] modelled semi-structured data and patterns by labelling 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 labelled 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, 5) Model description, and validation as seen in Figure.1

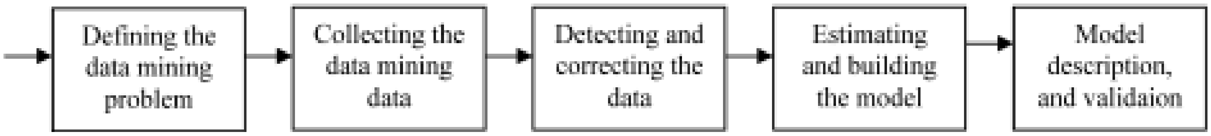

Figure 1. Data mining process

2.1 Defining the Data Mining Problem

Most data-based modelling studies are performed for a particular application domain. Hence, domain-specific knowledge and experience are usually necessary in order to come up with a meaningful problem statement. Unfortunately, many application studies are likely to be focused on the data mining technique at the cost of a clear problem statement. In this step, a modeller usually specifies a set of variables for the unknown dependency and, if possible, a general form of this dependency as an initial hypothesis. There may be several hypothesis formulated for a single problem at this stage [19]. The first step requires the combined expertise of an application domain and a data mining model. In successful data mining applications, this cooperation does not stop in the initial phase; it continues during the entire data mining process, the requirement to knowledge discovery is to understand data and business [9]. Without this understanding, no algorithm, regardless of complexity, is able to provide result that can be confident.

2.2 Collecting the Data Mining Data

This process is concerned with the collection of data from different sources and locations. The current methods used to collect data are:

- Internal Data: data are usually collected from existing databases, data warehouses, and OLAP. Actual transactions recorded by individuals are the richest source of information, and at the same time, the most challenging to be useful

- External Data: data items can be collected from Demographics, psychographics and web graphics. In addition to data shared within a company.

2.3 Detecting and Correcting the Data

All raw data sets which are initially prepared for data mining are often large; many are related to humans and have the potential for being messy [19]. Real-world databases are subject to noise, missing, and inconsistent data due to their typically huge size, often several gigabytes or more.

Data preprocessing is commonly used as a preliminary data mining practice. It transforms the data into a format that will be easily and effectively processed by the users. There are a number of data preprocessing techniques which include: Data cleaning; that can be applied to remove noise and correct inconsistencies, outliers and missing values. Data integration; merges data from multiple sources into a coherent data store, such as a data warehouse or a data cube [19]. Data transformations, such as normalization, may be applied; normalization improves the accuracy and efficiency of mining algorithms involving distance measurements. Data reduction; can reduce the data size by aggregating, eliminating redundant features. The data processing techniques, when applied prior to mining, can significantly improve the overall data mining results. [3].

Since multiple data sets may be used in various transactional formats, extensive data preparation may be required [20]. There are various commercial software products that are specifically designed for data preparation, which can facilitate the task of organizing the data prior to importing it into a data mining tool.

2.4 Estimating and Building the Model

This process includes four parts: 1) select data mining task, 2) select data mining method, 3) select suitable algorithm 4) extract knowledge as can be seen in Figure.2

Figure.2: show that this process is divided into four parts these are:

1-Select Data mining task (s)
Selecting which task to use depends on the model whether it is predictive or descriptive [3, 5, 9].predictive models predict the values of data using known results and/or information found large data sets, historical data, or using some variables or fields in the data set to predict unknown, classification, regressions, time series analysis, prediction, or estimation are tasks for predictive model [2]. A descriptive model identifies patterns or relationships in data and serves as a way to explore the properties of the data examined. Clustering, summarization, association rules and sequence discovery are usually viewed as descriptive [5]. The relative importance of prediction and description for particular data mining applications can vary considerably. That means selecting which task to use depends on the model whether it is predictive or descriptive.

2-Select Data mining method (s)
After selecting which task we can choose the method and assuming we have a predictive model and the task is classification while the method is Rule Induction, with Decision tree or Neural Network. In most research in this area; researchers estimates the relevant model this model to produce acceptable results. There are number of methods for model estimation includes these but not limited to neural networks, Decision trees, Association Rules, Genetic algorithms, Cluster Detection, Fuzzy Logic.

3-Select suitable algorithm
The next step is to construct a specific algorithm that implements the general methods. All data mining algorithm include three primary components these are:
(1) Model representation, (2) model evaluation, and (3) search. [2].

4-Extracting knowledge
This is the last step in building the model which is the results (or the answers for the problem solved in data mining) after making the simulation for the algorithm. This can be best explained by presenting an example of Auction Fraud [20]

2.5 Model Description, Validation
In all cases, data mining models should assist users in decision making. Hence, such models need to be interpretable in order to be useful because humans are not likely to base their decisions on complex "black-box" models; the goals of the accuracy of the model and accuracy of its interpretation are somewhat contradictory. Modern data mining methods are expected to yield highly accurate results using high dimensional models [5]. The problem of interpreting these models, are very important and is considered as a separate task with specific techniques to validate the results [19].
Model validity is a necessary but insufficient condition for the credibility and acceptability of data mining results [21]. If, for example, the initial objectives are incorrectly identified or the data set is improperly specified, the data mining results expressed through the model will not be useful; however, we may still find the model valid [20]. One always has to keep in mind, that a problem correctly formulated is a problem half-solved. The ultimate goal of a data mining process should not be just to produce a model for a problem at hand, but to provide one that is sufficiently credible, acceptable and implemented by the decision-makers; this type would need to consider all the data i.e. using a dynamic database.

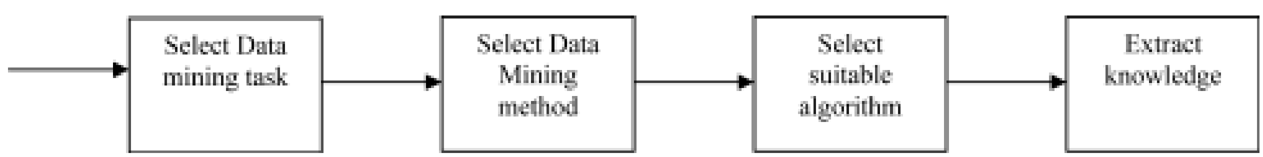

Figure 2. Estimating and building the model

3. Dynamic Data Mining Process

Due to the continuous, unbounded, and high speed characteristics of dynamic data, there is a huge amount of data in both offline and online data streams, and thus, there is not enough time to rescan the whole database or perform a rescan as in traditional data mining algorithms whenever an update occurs. Furthermore, there is not enough space to store all the stream data for online processing.

As mentioned earlier many researchers and developers have specified a process model designed to guide the user through a sequence of steps that will lead to good results. Many have been reported for data mining process. Some of these assumed that this is possible for Dynamic data mining process.

Ganti et al [22] examine mining of data streams. A block evolution model is introduced where a data set is updated periodically through insertions and deletions. In this model, the data set consists of conceptually infinite sequence of data blocks D1, D2, ... that arrive at times 1, 2, ... where each block has a set of records. Some applications require mining all of the data encountered thus far (unrestricted window scenario), while others require mining only the most recent part (restricted window scenario) and updating the models accordingly. The authors highlight two challenges in mining evolving blocks of data: change detection and data mining model maintenance. In change detection, the differences between two data blocks are determined. Next, a data mining model should be maintained under the insertions and deletions of blocks of the data according to a specified data span and block selection sequence.

Crespoa, and Weberb [6] presented a methodology for dynamic data mining using fuzzy clustering that assigns static objects to dynamic classes. Changes that they have studied are movement, creation, and elimination of classes and any of their combination.

Chung and Mcleod [23] proposed mining framework that supports the identification of useful patterns based on incremental data clustering, they focused their attention on news stream mining, they presented a sophisticated incremental hierarchical document clustering algorithm using a neighborhood search.

Zaslavsky et al [8] discussed data stream mining and the importance of its applications; the proposed techniques have their roots in statistics and theoretical computer science. Data-based techniques and task-based techniques are the two categories of data stream mining algorithms. Based on these two categories, a number of clustering, classification, and frequency counting and time series analysis have been developed.

Babu et al [24] focused on the problem of query processing, specifically on how to define and evaluate continuous queries over data streams, address semantic issues as well as efficiency concerns; they specified a general and flexible architecture for query processing in the presence of data streams. They also used their basic architecture as tool to clarify alternative semantics and processing techniques for continuous queries. They mapped out research topics in the area of query processing over data streams.

Reigrotzki et al [25] presented the application of several process control-related methods applied in the context of monitoring and controlling data quality in financial databases. They showed that the quality control process can be considered as a classical control loop which can be measured via application of quality tests which exploit data redundancy defined by meta-information or extracted from data by statistical models. Appropriate processing and visualization of the tests results enable human or automatic detection and diagnosis of data quality problems. Moreover, the model-based methods give an insight into business-related information contained in the data. The methods have been applied to the data quality monitoring of a real financial database at a customer site, delivering business benefits, such as improvements of the modeling quality, a reduction in the number of the modeling cycles, and better data understanding. These benefits in turn lead to financial savings.

In many situations, new information is more important than old information, such as in publication database, stock transactions, grocery markets, or web-log records. Consequently, a frequent itemset in the dynamic database is also important even if it is infrequent in the updated database [27].

Once a data mining system is installed and is being used in daily operations, the user has to be concerned with the system's future performance because the extracted knowledge is based on past behavior of the analyzed objects [6]. If future performance is very similar to past performance (e.g. if company customers files do not change their files over time) using the initial data mining system could be justified. If, however, performance changes over time (e.g. if hospital patients do not change their files over time), the continued use of the early system could lead to an unsuitable results and (as an effect) to an unacceptable decisions based on these results [5]. Here is where dynamic data mining comes into play by offering logical suitable techniques for "updating". In practice, and looking to empirical cases, dynamic data mining could be extremely helpful in making the right decision in the right time and affects the efficiency of the decision as well [1].

It becomes obvious, that something has to be done if a user is to keep applying his/her data mining system in a changing environment. In this case basically, there are three strategies [6]:

1-The user can neglect changes in the environment and keeps on applying the initial system without any further updates. It has the advantage of being "computationally cheap" since no update to data mining system is performed. Also it does not require changes in subsequent processes. Its disadvantage is that current updates could not be detected.

2-Every certain period of time, depending on the application, a new system is developed using all the available data. The advantage in this case is the user has always a system "up-to-date" due to the use of current data. Disadvantages of this strategy are the computational costs of creating a new system every time from scratch.

3-Based on the initial system and "new data" an update of data is performed. This will be shown to be available method in this dissertation.

In the area of data mining various methods have been developed in order to find useful information patterns within data. Among the most important methods are association rules, clustering, and decision trees methods [5]. For each of the above data mining methods, updating has different aspects and some updating approaches have been proposed:

- **Decision trees**: Various techniques for incremental learning and tree restructuring
- **Neural networks**: Updating is often used in the sense of re-learning or improving the net's performance by learning with new examples presented to the network.
- **Clustering**: [23, 28] describes in more detailed approaches for dynamic data mining using clustering techniques.
- **Association rules** [13, 29 30, 31, 32, 33, and 34]: Raghavan et al. developed a system for dynamic data mining for association rules [11].

This section will introduce the different aspects that might affect the outcome of data mining run if the data used for the run are to be changed or removed during this run.

3.1 Defining the Data Mining Problem

Since this process is concerned with the definition of the problem and the data is not yet built, no change will be carried out in this step, unless the analyst changes the problem goals.

3.2 Collecting the Data Mining Data

This process is concerned with the data collection from different 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:

1-If a new source of data (new database) and the data included in this database is a main source, this source will be used. This can be achieved by recollecting this data again, and/or replace existing data partially or totally. If part of the data in the new source is more related to the problem in hand the relevant part in the new source will replace the older part while keeping the same source.

2-New updated data (insert, update, delete) If the data used is changed in any way during a current run, then the data being used will be considered invalid and the new version of this data should be collected, a new run will be initiated, with the same source of data, but it must update the data immediately either by inserting new data, updating current data, or deleting data completely.

3.3 Detecting and Correcting the Data

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

1-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:

2-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 s ome 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:

3- 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.

3.1 Estimating and Building the Model

In estimating and building the model process in dynamic data mining; there are three main parts : select Data Mining task (s), select Data Mining method (s), and selecting the suitable algorithm that will not change if a new version of data is introduced (since they are concerned with the defintion of the problem). But in this case the extraction knowledge part will change depending on the new version of the data, and whether it is an insert, update and/or delete case. If a new source of data appears during this process one of the steps mentioned in section 3.3 will be applied according to the importance to the data.

3.2 Model Description, and Validation

Modern data mining methods are expected to yield highly accurate results [5] using high dimensional models, with new updated version of the data, it will surely change the results; say for example the number of rules in association rules could be changed; it may add a new rules, replace existing number of rules or change the percentage of the rules, and this change could affect the decision making process.

4. Conclusion

It is clear from all that have been said that data mining is still in its infancy, or at the beginning of the road as there are many aspects of data mining that have not been tested[37]. Up-to-date most of the data mining projects have been dealing with verifying the actual data mining concepts. Since this has now been established most researchers will move into solving some of the problems that stand in the way of data mining, this research will deal with such a problem, in this case the research is to concentrate on solving the problem of using data mining dynamic databases.

References

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

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

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

[4] Seidman, Claude (2000). Data Mining with Microsoft SQL Server 2000 Technical Reference, amazon.com/Mining-Microsoft-Server-Technical-Reference/dp/0735612714.

[5] Hand, David., Mannila, Heikki., Smyth, Padhraic (2001). Principles of Data Mining, MIT Press, Cambridge, MA.

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

[7] Papadimitriou, J., Sun, C., Faloutsos, (2005). Streaming Pattern Discovery in Multiple Time-Series, Proceedings of the 31st VLDB Conference, Trondheim, Norway, p. 697-708.

[8] Gaber, Mohamed Medhat., Zaslavsky, Arkady., Krishnaswamy ,Shonali (2005). Mining Data Streams: A Review, VIC3145, Australia, *ACM SIGMOD Record* 34 (2).

[9] Two Crows Corporation. Introduction to Data Mining and knowledge Discovery.

[10] Jin,Ruoming., Ag awal, Gagan (2000). A Middleware for Developing Parallel Data Mining Applications, *In*: Proc. of the 1-stSIAM Conference on Data Mining, 2000 - cs.ubc.ca.

[11] Raghavan, V., Hafez., A. (2000). Dynamic Data Mining, *In*: R. Loganantharaj, G. Palm, M. Ali (Eds.), Intelligent Problem Solving Methodologies and Approaches, *In:* Proc. of Thirteenth International Conference on Industrial Engineering Applications of AI & Expert Systems, Springer, New York, p. 220–229.

[12] Aaron, L., Ceglar and Roddick, J. (2007). Incremental Association Mining using a Closed-Set Lattice, *Journal of Research and Practice in Information Technology*, 39 (1) 35-45.

[13] Cheung, D.W., Han, J., Ng, V.T., Wong, C.Y. (1996). Maintenance of discovered association rules in large databases: An incremental updating approach, *In*: The Twelfth IEEE International Conference on Data Engineering, p. 106-114.
[14] Agrawal, R., Imielinksi, T., Swami, A. (1993). Mining association rules between sets of items in large database, The ACM SIGMOD Conference, p. 207-216, Washington DC, USA .
[15] Fukuda, T., Morimoto, Y., Morishita, S., Tokuyama, T. (1996). Mining optimized association rules for numeric attributes, The ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, p. 182-191.
[16] Domingos, P., Hulten, G.. (2000). Mining high-speed data streams, *In:* Proc. of the 2000 ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, p. 71–80, August.
[17] Asai, Tatsuya., Abe, Kenji., Kawasoe, Shinji., Arimura, Hiroki., Sakamoto, Hiroshi., Setsuo Arikawa., (2002). Efficient Substructure Discovery from Large Semi-structured Data, SDM- siam.org.
[18] De Raedt, Luc., Kersting, Kristian (2003). 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.
[20] Mena, Jesus. Investigative Data Mining for Security and Criminal Detection, First Edition, amazon.com/Investigative-Mining-Security-Criminal-Detection/dp.
[21] Daniel T. (2005). Discovering knowledge in data: an introduction to data mining, Larose, p. ,6.3_12—dc22.
[22] Ganti, V., Gehrke, J., Ramakrishnan, R., Loh, W., (2002). Mining data streams under block evolution, *In: ACM SIGKDD Explorations* 3 (2) 1-10.
[23] Chung,Seokkyung., Mcleod, Dennis. Dynamic Pattern Mining: An Incremental Data Clustering Approach, www.sigmod.org/ dblp/db/indices/a-tree/m/McLeod:Dennis.html.
[24] Babu, S., Widom, J., (2001). Continuous Queries over Data Streams, Stanford University, *SIGMOD Record*, 30. 109–120.
[25] Janusz Milek, Martin Reigrotzki, Holger Bosch, Frank Block. (2001). Monitoring and Data Quality Control of Financial Databases from a Process Control Perspectivem, PREDICT AG, Reinach BL, IQ, 2001 - finscore.com.
[27] Bruce,LTC., Caulkins, D., Lee, J., Wang, M (2005). A Dynamic Data Mining Technique for Intrusion Detection Systems, *In:*43rd ACM Southeast Conference, March 18-20, Kennesaw, GA, USA. ACM.
[28] Gionis, A., Mannila, H., Tsaparas, P., (2007). Clustering Aggregation, ACM Transactions on Knowledge Discovery from Data, 1 (1).
[29] Jiang, N., Gruenwald, L., (2006). Research Issues in Data Stream Association Rule Mining, SIGMOD Record, 35 (1) Mar.
[30] Cormode, G . (2007). Fundamentals of Analyzing and Mining Data Streams, Workshop on Data Stream on Data AnalysisANALYSIS – San Leucio, Italy - March, 15-16.
[31] Ceglar, A., Roddick, J. (2006). Association Mining, *ACM Computing Surveys*, 38 (2).
[32] Tzung-Pei Hong, Ching-Yao Wang, Yu-Hui Tao, (2001). A new incremental data mining algorithm using pre-large itemsets, *Intelligent Data Analysis*, 5 (2) 111–129.
[33] Hong, Tzung-Pei., Huang, Tzu-Jung (2007). Maintenance of Generalized Association Rules for Record Deletion Based on the Pre-Large Concept, *In*: Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, February.
[34] Chun-Wei Lin, Tzung., Pei Hong, Wen-Hsiang Lu, Chih-Hung Wu (2007). Maintenance of Fast Updated Frequent Trees for Record Deletion Based on Prelarge Concepts.
[35] Agrawal, R., Srikant, R.(1994) . Fast algorithm for mining association rules, *In*: The International Conference on Very Large Data Bases, p. 487-499.
[36] Nasraoui1, O., Rojas, C., Cardona, C. (2005). 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.
[37] Jiang, N., Gruenwald, L. (2006). Research Issues in Data Stream Association Rule Mining, SIGMOD Record, 35 (1) Mar.