2

A Novel Theory in Risk-Management by Numerical Pattern Analysis in Data-Mining

Masoud Kargar, Farzaneh Fartash and Taha Saderi
Islamic Azad University, Tabriz Branch
Iran

1. Introduction

By technology development and highly affect of computer usage in people's daily life, the way of gathering and saving information has differed a lot. Deep change in storing and retrieving data - raw information - from paper based to computer based has brought up a condition needing new techniques for data management necessarily. Increasing need for gathering information in different social systems and on the other hand, decreasing time periods from asking to get a reply, made computer engineers think about ways for speed process on stored data and get useful information as soon as possible. As it is said, "necessity is the mother of invention," from late 1980s data mining has been developed and become warmly accepted amongst software engineers and database designers (Larose, 2005),(Pei, 2004).
Representing local features of data in the forms called patterns is the main task of data mining (Bloedorn, 2000). In this paper, a new pattern in data mining is represented by using probability-trees to have a process on different type of data stored in databases and give familiar information to system users.

1.1 Terminology

Data mining is a popular technique in searching for interesting and unusual patterns in data, has also been enabled by the construction of data warehouses, and there are claims of enhanced sales through exploitation of patterns discovered in this way. In other words, data mining is extraction of interesting (non-trivial, implicit, previously unknown and unexpected and potentially useful) information or patterns from data sources, e.g. databases, texts, web, images, etc (Kuonen, 2004), (Karasova, 2005), (Laxman, 2006).
Data mining is not data warehousing, (deductive) query processing, SQL/reporting, software agents, expert systems, Online Analytical Processing (OLAP), statistical analysis tool, statistical programs or data visualization. Further definitions are available in (Aflori, 2004), (Piatetsky-Shapiro, 2006), (Alvarez, 1994), (McGrail, 2002), (Ordieres Meré, 2002).
Pattern is a local feature of the data, departure from general run of data, a group of records that always score the same on some variables, a pair of variables with very high correlation and unusual combination of products purchased together. Patterns must be valid, novel, potentially useful and understandable; validity is holding on new data with some certainty;

novelty is to be non-obvious to the system; usefulness is to be possible to act on the item and ability to understand is being interpretable by humans (Hand, 2001).
In data-bases relation has the same meaning with table. In this paper relation is used instead of using the word table.

1.2 Outlines

This chapter is organized in eleven sections. In Section 1 we started an introduction to the subject. In section 2 and 3 the problem and an idea solution to the problem are set forth. In section 4 we present a brief survey of the previous work, their strength and weaknesses, and the areas, requiring further improvements. From section 5 to 9 we present our work under the title of Probability-Based Patterns and describe how the job can be done. In section 10 the steps of risk management is explained. Finally we conclude the paper in Section 11 with an evaluation of our work and some future topics of research.

2. Concepts

Nowadays, entrance of software and IT in all commercial systems has caused a special sensitiveness in the comparison world between different companies and different ideas. Maybe we can now claim that good knowledge about past, present and future can lead us to success and the company having confidence to such information, can easily present proper strategies and protocols, establish proper connections and start to upgrade their commercial systems.
They can easily and frequently do reengineering and management in proper manner and guide commercial processes towards development. It is impossible to reach such a guidance and leading stage in development and growth unless the operations lean to the reliable knowledge and information.
These days, business and commerce processes amongst companies and customers have become so important and comporting the relation between them needs reliable and precise analysis. For this aim, information and knowledge producing process seams to be helpful and beneficial. In trademark process, all the process gather next to each other in the way that they lead the system to the stage of improvement which can come to existence in different paths: its structure and organism, its processes and conducts with others, its feedbacks and results.

3. Different Paths

3.1 Structure and Organism

If a part of an organisation or society reaches to improvement, surely the commerce process will cause less overhead and cost upon the system. On the other hand, the external conditions of the organisation will be known easily and the organisation can make itself ready to comport with them. The important point here is that the external conditions depending on time, place and culture are in various types. Therefore the situation adjustment of company and its structure with the external conditions of the system will be gained only if the change in every part be distinguished easily and quickly; otherwise expecting improvement and productivity will be impossible.

What's more, knowing a system relates to different and various parameters. It is also possible that a complicated and hidden relation may exist between different indexes and different parameters which before discovering such a relation, no proper method would be presentable.
For an instance it is important for a well-known company that in an LCD sale, how much gender, nationality, age, financial stage, job and such information should be effective in quality, price and size of it. On the other hand, the company should be aware of the needs of markets and the requirements that haven't been fulfilled by others or a proper way is possible and should be emphasized and be counted and analyzed.
Therefore such information can organize the companies to access different methods in their systems. Such a way in organizing the company policies will be trustable though general organizing systems cannot be so reliable. Generally the organizations are classified according to the process, task, function, operation case and/or objects and are gathered next to each other through definite basics. This job should be followed until the result which is production, business or an operation is gained and improvement is noticeable in the existing cycle.
Usually complex of these methods are not acceptable and usually there is doubt in efficiency and yield of such organizations. Although all methods lead to improvement, but unless the knowledge covers all the parts of an organization and relations and receives proper reply from them, and there is no proper analysis from all the changes, the change in the system and organization would be impossible.

3.2 Processes and Conducts

A system is made up of a set of processes. These processes are set next to each other in special templates and special behaviors. However necessity of the processes themselves or the changes among them and on the other hand revising in the type of connections between them are so important. This subject is so difficult that after passing a period, no effectiveness of a process becomes known or the relation between processes are weaken or strengthened or wholly eliminated or given existence.
The need to the new processes and relation amongst them is a necessity at the pass of time. The orders of customers and the ones who are in connection with us vary frequently and if any delay to wait for a general and major change in processes and relations and in other words establishing a new system will make us late. So, it is clear that along with partial changes in orders, those changes should be known and some changes simultaneously should be done in the system. There is a question: which processes should be changed and what relations?
This subject is so important that the processes and their internal and external connections should change gradually. Otherwise it is expected to here that the system has run down. This precise job that accompanies many risks requires instant, exact and clear recognition. It needs special ability till for each change in its processes or relations their future effects become analyzed and observed properly.

3.3 Feedbacks and Results

Every change, every process and generally every operation does a task and probably success in many cases lonely but when they are set next to each other defeat of system or decrease in

expected success might happen. Basically when the processes have finished and the operations are done, the results come to the system at the end to see if the operation cycle is proper sufficiently or not; what the strong and weak points are.

3.4 Problem

Such an operation result references to the operation cycle would be so efficient but producing such proper results that can give a brief and exact knowledge about the proper stage of the system, is so tenuous and difficult. The problem is how to solve this question though the question is so familiar to general that if the results are gained through feedbacks then the system can improve but how the results should be produced? How the result should be analyzed to clarify the parts that should have change in.

3.5 Solution

Generally while considering a commercial system and operation of a company come to this conclusion that setting data and statistics next to each other, discovering dependencies and correlations between different attributes and values of them and finally finding a relation between those results is not easy. If this process programmed in an engineering way, the processes and organisms of the commercial system will improve reliably and the improvement and upgrade management gain to a confidence that the amount of risks caused by not recognizing them, reach to its lowest level.

Knowledge processing should be instant and precise. Effective and impressible parameters should be ranked according to their priority. Dependence to knowledge process is so efficient in knowledge confidence ability. The aim is not knowledge processing for once but a continuous operation is important. Entering knowledge in system for upgrading system and its efficiency can be part of the end. Knowledge processing cycle should be simple.

4. Previous work

Although the pattern suggested in this paper is new and there is no previous work done on it, but here some information in given about a number of previous models presented in articles.

4.1 Decision tree

Many data mining methods generate decision trees—trees whose leaves are classifications and whose branches are conjunctions of features that lead to those classifications. One way to learn a decision tree is to split the example set into subsets based on some attribute value test. The process then repeats recursively on the subsets, with each splitter value becoming a sub-tree root. Splitting stops when a subset gets so small that further splitting is superfluous or a subset contains examples with only one classification.

A good split decreases the percentage of different classifications in a subset, ensuring that subsequent learning generates smaller sub-trees by requiring less further splitting sorting out the subsets. Various schemes for finding good splits exist (Menzies, 2003). For further information about decision trees refer to (Fayyad, 1996).

4.2 Declarative networking

In Declarative Networking, the data and query model proposed for declarative networking are introduced. The language presented is Network Datalog (NDlog), a restricted variant of traditional Datalog intended to be computed in distributed fashion on physical network graphs. In describing the model, the NDlog query which performed distributed computation of shortest paths was used. In that model, one of the novelties of our setting, from a database perspective, is that data is distributed and relations may be partitioned across sites. To ease the generation of efficient query plans in such a system, NDlog gives the query writer explicit control on data placement and movement. Specifically, NDlog uses a special data type and address, to specify a network location. For further information refer to (Thau Loo, 2006).

4.3 Databases and logic

Logic can be used as a data model in databases. In frame of such a model we can distinguish Data Structures, Operators and Integrity Rules.

All these three elements are represented in the same unique way as axioms in the logic language. As a deductive database system we can treat such a system, which has the ability to define deductive rules and can deduce or infer additional information from the facts stored in a database. Deductive rules are often referred to as logic databases. A deductive database (DDB) can be defined as:

DDB= {I, R, F}, where: F – a fact set, R – a deductive rule set, I – an integrity constraints. For further information refer to (Nycz, 2003).

In (Wul, 2004) a data-mining is surveyed in an artificial intelligence perspective that presents a new and interesting view to readers. Avoiding prolongation it is not noted here.

4.4 Probability-based patterns

In the previous work done for representing a pattern in data-mining and risk management, a new step was put forward that some points of them are mentioned here. For the perfect information about the job, refer to (Kargar, 2008).

5. Database Models in Probability Based Patterns

Using probability in the pattern makes it so sensible, understandable and easy to be performed. The description of the pattern is kept on by representing an example:

Data-bases are considered as relations with logic relationships between them. The objective is to concentrate on the special data and to reach to a point that there exists desired amount of them or it grows with a special norm increasingly or decreasingly. In some cases such as profit, revenue, sale amount and etc the increasing growth is requested and in other cases such as withdrawal, warranty time, absolute produce cost and etc the decreasing growth is requested.

Because of the relationship between the relations, fluctuations of the amounts of a field may influence the amounts of other fields. So, by designing an appropriate pattern from the data-base, fluctuations of the amounts of a field could be managed in a way that it goes up and down, in the manner we want.

In the supposed data-base some relations exist as below:

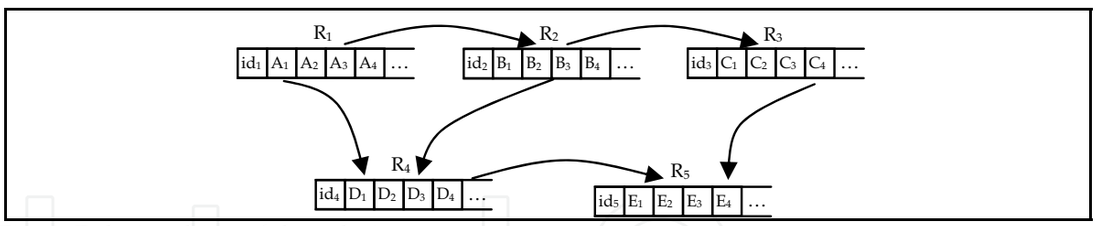

Fig. 1. Relationships of the relations

The relation R1 is supposed as the source and the relation R5 is supposed as the destination. Other relations play the role of interfaces and are the connection terminals between the source and destination. However they can be supposed as independent destinations too.
As it is shown in figure 1, some relations are directly or indirectly related to R1 and a change in their amounts can be affective in the change of the amounts of a field like A1.
Some points are considered:
1- The relations that are related to a relation like R1 directly or with less number of hops can have more affect on it. So the dependency amount of near relations is more than the far ones (from the perspective of hop amounts of relationships).
Although the presupposition 1 is a basic rule in designing the pattern, but it never causes the far relations to be worthless. The exit of a far relation from the set of relations which are affective in the pattern depends on its dominant and related data.
2- The relation which includes more relationships is an important relation and should be set in higher priority in designing the pattern.
3- The R set includes dominant and affective attributes in the pattern. These attributes are selected through the prime and experienced guess which becomes perfect during pattern designing period and possibly some attributes would be less important (not affective).
4- An attribute is considered as a head or a key attribute and pattern is designed according to it.
The following steps explain how the pattern is designed:
1- Firstly a graph is drawn based on the R set;
2- Then importance and priority of the nodes should be computed;
3- The decision tree or the decision graph is drawn based on the graph and the statistics gathered from the data-base, for designing the pattern;
4- Deduction of superior patterns is done from the graph.

6. Creating Probability Based Decision Tree & Decision Graph

The decision tree or graph is used to put together the effective and impressible components existing in the operational environment to deduce the hidden relations among them by this dependency and correlation. Using each of them in special situations can provide special gains. The statistics existing in databases can help with making and creating each of them. The importance of subject is shown by numbers and probabilities which demonstrate themselves as the weight in tree and graph and creates the probability based weighted tree and graph.
In the probability based decision tree, the operation starts from the root and in each step a level is added to the tree and this job is kept on till it reaches to the point that the tree and the number of levels seem sufficient. In this tree, each level relates to an attribute and according to

various statistics the attribute has in the database, the related weight of each edge and branch is produced.
For example, if the supposed attribute and component is gender, it means that the gender is mentioned as an effective parameter and element; therefore the statistics related to it is gathered according to dominant values. As the values of this attribute are precise and significant, then its statistics are resulted easily. Figure bellow demonstrates the subject clearer.

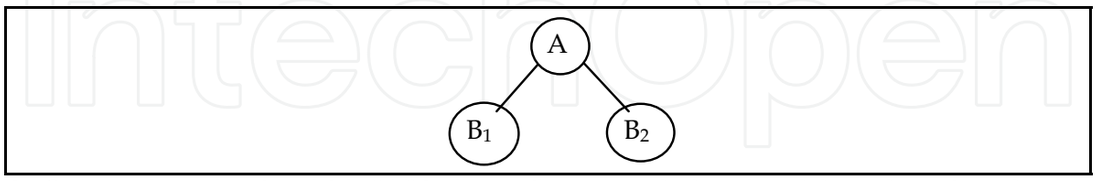

Fig. 2. A tree branches in which "A" is the attribute "gender" and each branch are its values

In this tree branch "A" is considered as gender and each branch as each value. In the branch A-B_1 the weight related to the first value and in the branch A-B_2 the weight related to the second value is taken into consideration. Imagine B_1 as the attribute "male" and B_2 as the attribute "female". If the branch A-B_1 is supposed as 40% and the branch A-B_2 is supposed as 60% then the tree will be formed as the figure bellow:

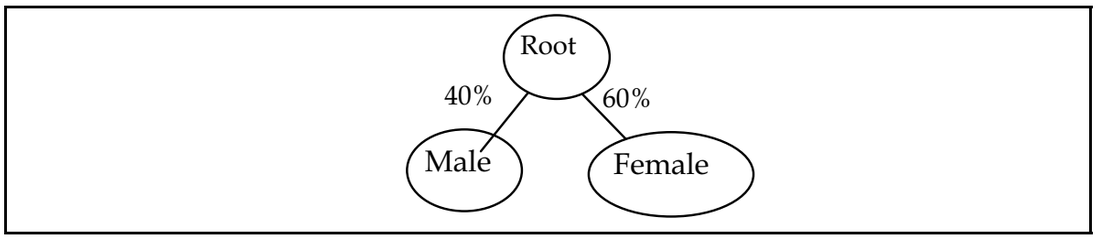

Fig. 3. A tree branches in which the attributes are given values

Now, the whole set is clustered to the two sets of "male" and "female". This operation would be starting point of other component and statistics selection and producing the next weight which operates in intricacy manner.
As an instance, the product type that can be dominant values such as LCD, CRT, Plasma and other types for a TV company, because of considering 4 values for the product type then 4 branches are mentioned and each of them will be evaluated in "male" and "female" sets. The forth value and the statistics and probability related to that consists other products than LCD, CRT and Plasma. In the figure bellow the intricacy of clustering operation is demonstrated clearly.

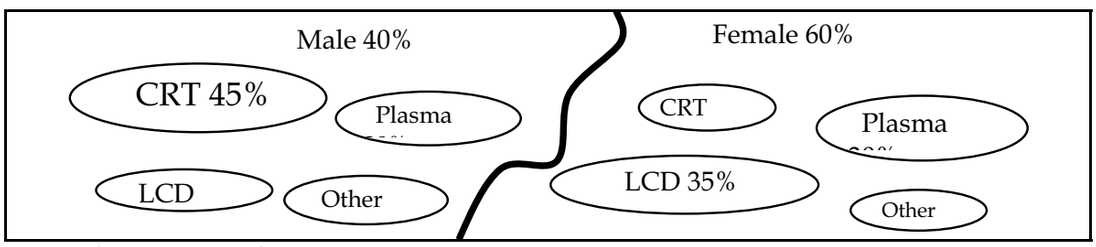

Fig. 4. Clustering attributes

In better words, while computing the purchase statistics of CRT and LCD for "male", this statistics doesn't relate to the whole database and the set. It is observed under the subset of "male." As the first component which was related to gender, had 2 values then both CRT statistics and LCD statistics will be computed twice and will continue like this.

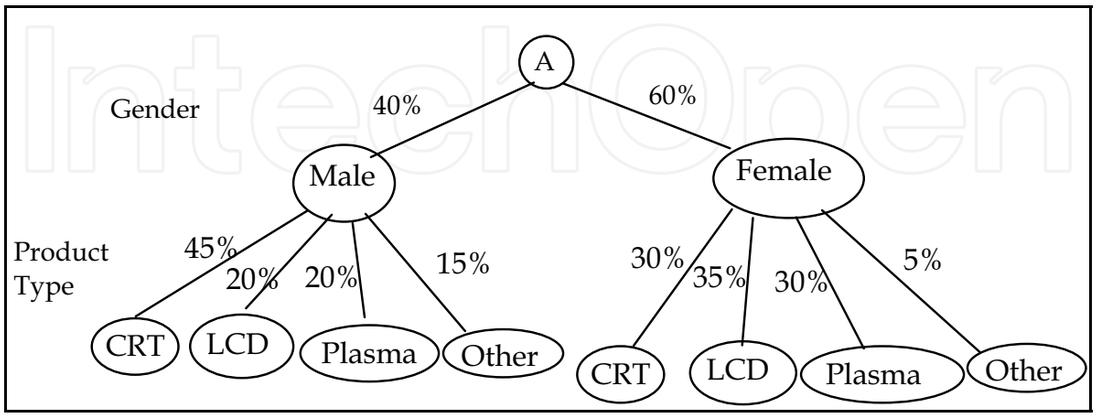

Fig. 5. A tree branches in which the attributes are given values

As it is illustrated in figure 4, the whole sum of the probabilities for children of a parent after computation will be 1 or 100%. The important point that should be paid attention is that weight and importance of each branch relates to the previous branch and the statistics of each branch is analyzed locally not globally. In other words, allocating 45% to CRT doesn't convey that generally the use value of CRT is 45%; this statistics is locally computed for "male".

Of course it will be explained that the local effect is itself a defect for probability based decision tree because in pattern production and dependency declaration among the attributes, the rate of 30% for "female" won't have any affect on the last pattern. On the other hand, because of considering various values for an attribute, this variety will increase the degree of each node at each step and it helps making the subject to become more local and the produced pattern to become weaker.

The other point that is important in a tree production and ignoring it, is effective in weakening the pattern is correct selection of attributes, with considering their priority. The attribute selected as the root will have more effect on in the final weight of the pattern because of clustering in the first steps. Therefore if it is recognized that instead of using product type at the second layer of the decision tree, producer country should be used, certainly different effect will be noticed in pattern production.

It is why the attribute importance should be observed. For this job, database should be checked and its design should be considered. The major effective attributes are recognized according to our observation from database and are set from root to leaf sequentially according to their priority.

7. Giving Priority to the Effective Attributes

The target or impressible attribute is chosen while database design. Revenue, sale rate or visit amount can be supposed as an example for this attribute. This attribute is also an

attribute or a component in database; then according to the supposed direction among the database relations and according to the number of relations between them, a degree is computed for each attribute. In fact this degree and priority will illustrate the attribute position among the effective attributes. So, the attribute which has high priority will be nearer in tree production to the root and its efficiency amount will be much more.

7.1 The Important Points in Computing the Attribute Priority

Suppose that the impressible attribute is "A_1" and A_1 depends to the relation "R_1". Then A_1 and R_1 are considered as a base for computing the priority. The attribute indeed loses its importance how much it goes farther than R_1 in database. Hop in database is computed according to the relations and the links among them.

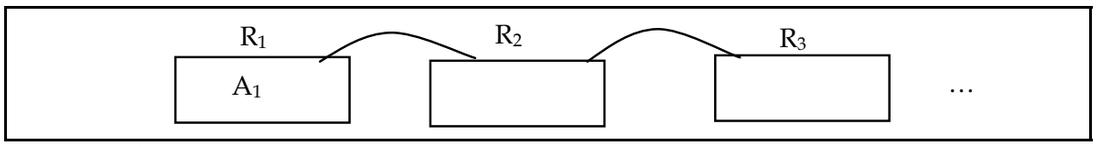

Fig. 6. Relations

As it is shown in the figure above, each "R" is considered as a major subject and "A" is considered as an attribute of "R".

The other point which is important in computing the attribute priority is dependency or becoming dependant. When two relations like R_1 and R_2 have linked to each other, usually one of the two cases may happen: either R_1 is related to R_2 or R_2 is related to R_1. In this relation, too, the roles of parent and child exist and usually the relation of 1 to n is noticed. Among these two relations, one or some attributes have key role and one would give its main key to the other to make relation. The one which gives its main key to the other key plays the role of parent and the one which accepts the key plays the role of child. Now suppose R_1 is parent and R_2 is child. Then R_1 will be more important and more effective than R_2 and as it is shown in the figure a directed edge is drawn from R_1 towards R_2.

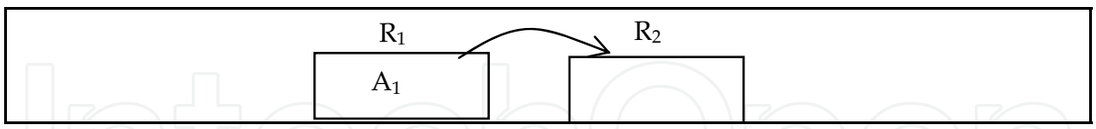

Fig. 7. The relation between R_1 and R_2

For this cause the relation which has high external degree, will be more important than the relation which has high internal degree. This subject is also important in computing attribute priority of R_1 and R_2. On the other hand the amount of being far or near to the target relation or the relation which has impressible attributes, is another cause at the priority value of an attribute.

$$H(A_i) = C_1(DI_{RAi}) + C_2(DO_{RAi}) + C_3(H_{RAi}) \quad (1)$$

At the formula above, DI is internal degree, DO is external degree and H is number of hops to the target relation from which the priority of A_i is computed. C_1, C_2, C_3 are constants that

show the importance of internal degree, external degree and hops. Obviously hop has an important role in every case and the external degree is much important and effective than the internal degree. Selection of numbers and constants are also important.

7.2 Evaluating the Probability-based Decision Tree

Although creating tree and documenting statistics and probability are both simple and reliable and in most cases helps us with reaching to our final end which is correct pattern, but the more important matter that should be paid attention, is effective attributes selection and giving priority to them. Priority is simply recognized by internal degree, external degree and hop though clustering is not a fact that could be ignored or maintained easily because nature of tree is based on local clustering.

7.3 Computing the importance and priority

The formula below is used in computing the value of a node in which, Pr is the priority computed by the formula; c1 and c2 are two constants; di and do are respectively input and output degree; h is the hop number or the distance from the vertex A which is the root. While computing h the direction of edges is not considered.

$$Pr = c1 * do + di + c2(n-h) \quad (2)$$

$$n = hmax+1 \quad (3)$$

The result of computing the priorities produces a list which includes the members of R and some that are not members. In that list A is the head and other members are tails. There, fields may be put together with the relations. Non of the fields from the intermediate relations are selected this time, but while designing tree from the existing data in the database and extracting operation, some fields of them might be chosen.
Figure 3 shows a sample of designed tree through the graph and relations. It is supposed in the figure that dominant amount in R3 gained by statistics is the attribute c2 which is called the license.

8. Probability-based Decision Graph

The main points that are mentioned in design and producing of probability based decision tree, are important cases. In order to reach to a reliable model and eliminating intricacy local clustering, decision graph can be used instead of decision tree. All of the points which were used in designing decision tree from the graph related to the database and in giving priority to the attributes are used here either without changing graph to tree.

8.1 Producing Probability-based Decision Graph

The effective attributes that can have influence on the impressible attributes are chosen. Each attribute is supposed as a node in the graph. Various edges are drawn from each node as external to the related attributes at database. Each edge will have 2 values and this ordered pair will consist of (the related probability and statistics, value).

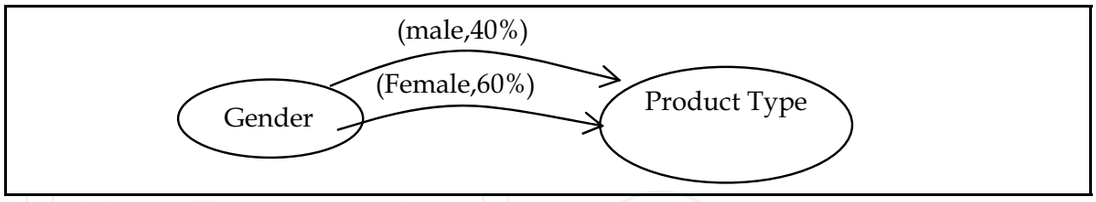

Fig. 8. Relations between two attributes

As it is shown in the figure above, "gender" is selected fromthe customer relation and "product type" is selected from the product relation. As it is supposed that the customer relates to the product type, then if the gender is a member of customer relation and the product type is a member of product, then the direction of the edge in the decision graph will be from gender to product type. This adaptation of database designed model with decision model will have an important role in its correctness.

8.2 The Advantages of Probability-based Decision Graph

The major two advantages that can be listed are:

- Adaptation of designed model with decision model,
- No effect of number of different values in clustering localization.

One thing that had influence on the pattern and decision tree and weakened them, was variety of values for an attribute. This point was so efficient in the amount of localization and shrinking of the final clusters, which that impressibility is solved by this appraoch.

9. Producing Pattern and its Concepts

A pattern is a set of attributes and the values related to each other which can cover a concept together. On the other hand, these attributes and the related to each other values show the amount of effectiveness. Every pattern consists of two parts: head and tail.

Head is that target or impressible attribute which can be supposed as different instances such as revenue rate, sale rate, satisfactory rate, relation rate and so on. The important thing is that usually one attribute is supposed for this instance. For considering both revenue rate and sale rate, as an example, then an independent and different pattern will be introduced for each of them. At the tail part, a set of effective attributes exists along with the related probabilities. Therefore the pattern can be illustrated as a matrix.

Head	Tail		
Impressible Attribute	Effective Attribute 1	…	Effective Attribute n
Related Weight	Related Weight		Related Weight

Fig. 9. Head and tail of a pattern

In the figure above, n effectivve attributes are shown on one impressible attribute. It is illustrated symbolistic like the figure bellow.

A_H	A_{t1}	A_{t2}	…	A_{tn}
P_P	P_1	P_2		P_n

Fig. 10. Head and tail of a pattern in symbolic form

In the figure above, P_P is that probability of the pattern or the final weight which relates to the attribute A_H. This probability is computed from the multiplication of effective attribute probability.

$$P_P = \prod_{i=1}^{n} P_i \tag{4}$$

The number P_p is between 0 and 1 that how much it gets closer to 1, shows that the pattern is strong and useful and how much it gets closer to 0 shows its weakness. Obviously how much the number of n or that effective attributes is more, this number gets closer to 0 and how much there are efficiency attribute which its value is less than average of probabilities of n attributes then the effect will be negative or weakening. If this probability is greater than average of probabilities of n attributes then the effect will be positive or strengthening the final weight of the pattern. Meanwhile, how much the number of attributes is lower it is a sign of pattern weakness.

9.1 Pattern Production in Probability-based Decision Tree

We start from the root of the tree and consider all the paths ending to leaves. Every path that has greater value in the final state, that path is accepted as a pattern. The considerable point is that how much the optimum path or dominant pattern is selected or in other words the path with greater numbers is selected it shows the correctness choosing the attributes closer to the root. The figure bellow gives an example.

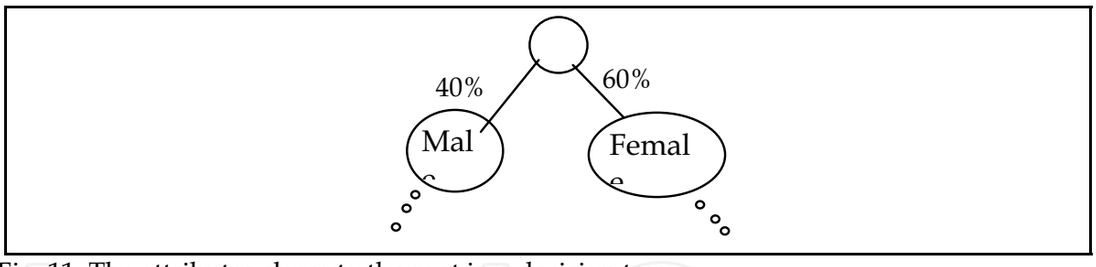

Fig. 11. The attributes closer to the root in a decision tree

In the final pattern, if "female" selected as an effective attribute with the weight of 60% then the attribute selection of gender has been done correctly; otherwise the priority, position and the statistics relater to that attribute should be checked and observed precisely.

9.2 Pattern Production in Probability-based Decision Graph

In the decision graph, at first we suppose the graph is connected because the database model is connected. By an algorithm such as Kruskal, Prim, Sollin or every Dijkstra algorithm this graph is changed to spanning tree. The tree itself will be the result. In that tree every edge, consisting the related value and probability, is selected as a member of the tail part.

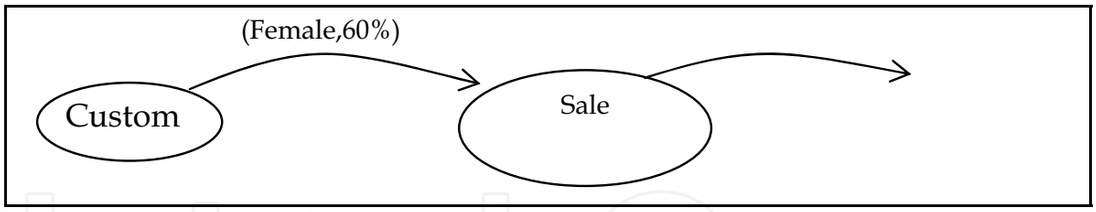

Fig. 12. An edge as a member of the tail part

Revenue	Female	…
?	60%	

Fig. 13. A sample pattern

The pattern produced by the graph has many advantages; for instance it is not local or in intricacy manner and the produced pattern is universal. The efficiency amount of a value on the other value is eliminated. Then the output of both tree and graph is an array. They have no difference in their structure and have equivalent aspect. Each can be proposed at its special position and situation as an important production method. If the attributes and the efficiency are on each other, decision tree would be the best choice. If the attributes independency rate is high in comparison with each other, then decision graph would be the best choice.

10. Pattern Evaluation and Risk Management

10.1 Getting Concepts from Patterns

Pattern evaluation gives new concepts from patterns. While producing a pattern, the probabilities in its inside helps us to be able to recognize the attributes and subjects related to our target. The reason is that the supposed target in impressible attribute and the instances that have influence on it exist in its tail part and their rates are illustrated as numbers. Now, these numbers can be changed into other format or shape and normalize to get proper explanations of them. Pattern production is done in proper periods.

The produced patterns are revised to re-observe the system by that numbers. This is done to explain the data changes and change paths. The policies, protocols, solutions and approach are changed in a way to reach to that target. For this job a general target is introduced and also some sub targets if required. Therefore the final decision pattern is produced like the figure bellow.

Target	Sub-target 1	Sub-target 2	…	Sub-target m

Fig. 14. Targets in the pattern

Therefore a new independent pattern is introduced for each target as the figure bellow.

Target 1	Attribute 1	Attribute 2	…

Fig. 15. Independent pattern for each target

By this way all sub-targets introduce special and effective attributes for themselves. After introducing the final pattern and the sub-pattern, it is the time to be able to lead and guide them in a way to reach to the target. Target is reaching to an ideal number or probability. Then for each pattern, either target or non-target a matrix is supposed.

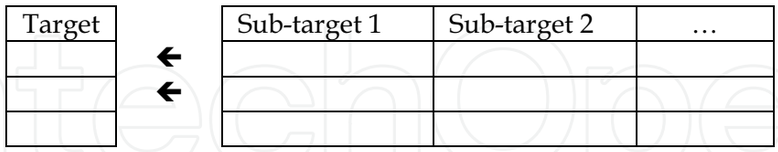

Fig. 16. Target matrix

Now, in various periods and usually after some changes in the system, continually start to produce pattern. This operation should be done and deduced ceaselessly. As an example, a change in the product color may improve the sub-target and consequently have good reflex on the target. In that case, the number of new target will be greater than the number of previous target. In other case the system may change warrantee rate instead of color and see that this change weakens or strengthens a sub-target.

10.2 Matrix Analysis

In this part the pattern is analyzed through a matrix. Suppose p is the number of patterns and k is the number of distinct attributes in the pattern. We have:

$$N_i = \sum_{j=1}^{k} M_{ij} \qquad \forall M_{ij} = 0$$

$$S_i = \sum_{j=1}^{p} M_{ji} \qquad \forall M_{ji} = 0$$

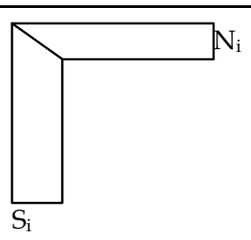

Fig. 17. A sample matrix

The greatest Ni determines that the pattern has the least covering with other patterns and the greatest Si determines that the attribute number i at the column i is the attribute that has the least repetition in the pattern. Therefore it can be concluded that some of the greatest Ni and Si can be deleted. It means that the Mpk matrix can be reduced and recreated by the existing attributes. This job is done over and over till the growth interval of the Ni and Si becomes limited. When the maximum value of Ni is 3 then it means that there is 3 unrelated attributes. When the maximum value of Si is 3 then it means that there is an attribute which cannot be found in 3 patterns and the patterns are independent from it.

Resistance in a matrix is the confirmation in a pattern and come to a resistant conclusion in patterns. After reach in a resistant state in the matrix, it can be supposed that the numbers loose their value and importance and the average of numbers in the columns can be used as the coefficient of that attribute in the major pattern. The pattern will be produced through the matrix and all the attributes will be used in that process with the average of values in the pattern. The numerical analysis on the pattern will represent the resistance of the final pattern and its fluctuation will be at the least amount because of using the average value.

10.3 Targets and Sub-targets

Having targets and sub-targets helps to notice the effect of changes in policies, rules, approaches and etc easily. If the process of these changes and the curve of general changes of target attribute be taken under consideration, after a short time the system will simply access to a knowledge that helps to notice the effect of different instances in its pattern.
These changes should be lead in a way that generally and in one trade off among the parameters, the final estimate and resultant will be the considered target of system. Such guidance is called risk management. If the system operates in a way that it never comes down from special values, the system targets are gained. So risk management should be accompanied with a diagram for the targets. The figure bellow gives example.

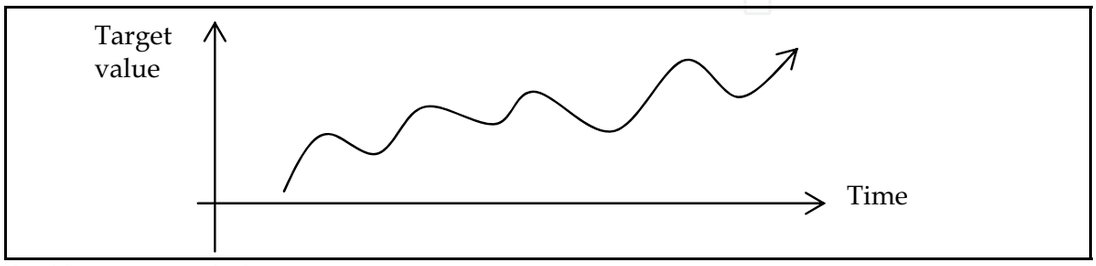

Fig. 18. Target variation

It is possible that the improvement in a sub-target does not always improve the target and some times lowering of a sub-target may cause an ideal improvement in other sub-targets which finally ends in the improvement of the target. Figure bellow demonstrates a sample.

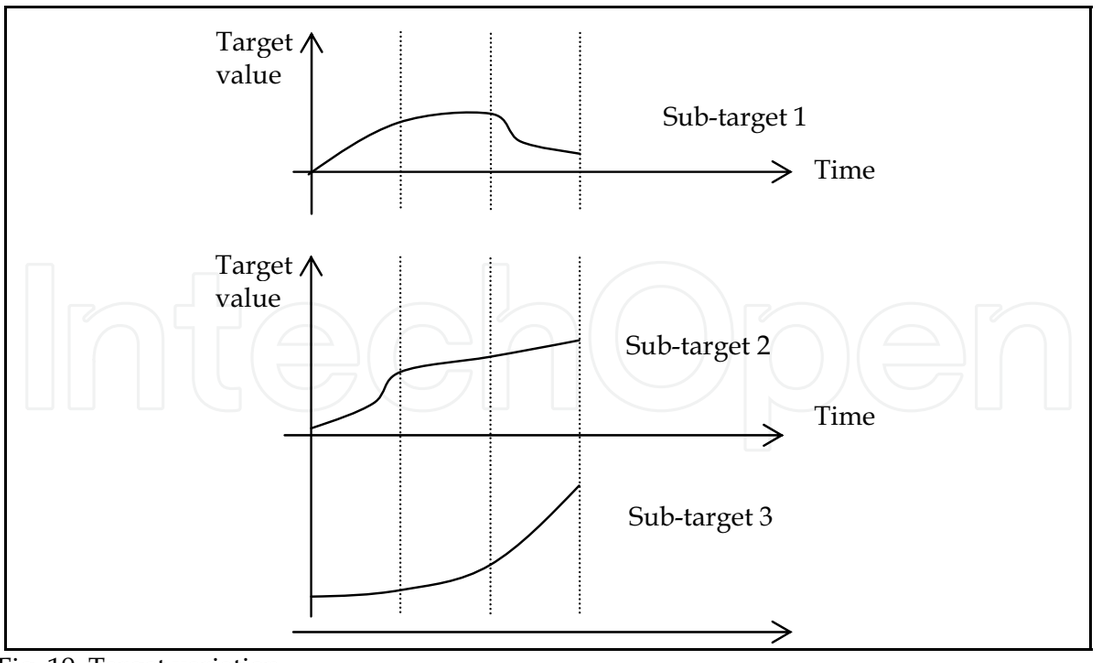

Fig. 19. Target variation

By noticing the variation curve of the value of sub-target 1, it is seen that a lowering in a

sub-target may affect other sub-targets in the positive direction so much and finally improve the target. Now if these variations and their effect sources become recognized, certainly the system management will be done simply and ideally. By such management the system can improve the parameter and factor which likes to grow even it causes lowering or becoming negative in other parameters.

11. Conclusion

There are so many problems for great systems by the huge amount of data in data-bases. Data-mining by using patterns from data-bases can solve the problem very much, as describe in the first section. So many kinds of patterns are suggested, designed and used in systems that we have introduced some in section 2. The new pattern designing suggested in this paper, is done using probability and decision trees. It is valid, novel, useful and very understandable because of giving mathematical information. The relationship between the relations (tables) is done according to system analysis and the designing and gathering of data is based on that analysis. Various analysis give various designing. As described in section 3 it can be used in risk management operations and guide great systems to increase their profitable attributes and reduce the harmful ones. In section 4 the way to use these patterns in risk management is explained and some important points are mentioned.

11.1 Future Research

The method we presented in this paper sets the stage ready for two interesting topics of research:

Knowledge production along with the risk management by probability based patterns in data-mining.

A New Approach:

A method can be suggested, in which using patterns in risk management can lead us to change the direction of job from data-mining to the knowledge bases in expert systems. In this case the control of mankind managers on the system will be given to the programmed machines and the rate of errors surely come down.

12. References

Aflori, C.; Leon, F. (2004), Efficient distributed data mining using intelligent agents, *supported in part by the National University Research Council under Grant AT no 66 /* 2004

Alvarez, J.L.; Mata J. & Riquelme, J.C. (1994), Data mining for the management of software development process, *International Journal of Software Engineering and Knowledge Engineering*, World Scientific Publishing Company, p.3.

Bloedorn, E. (2000), Mining Aviation Safety Data: A Hybrid Approach, *The MITRE Corporation*

Fayyad, U.; Piatetsky-Shapiro, G. & Smyth, P. (1996), From Data Mining to Knowledge Discovery in Databases, Copyright © 1996, *American Association for Artificial Intelligence*, pp. 37-49

Hand, D. J.; Mannila, H. & Smyth, P. (2001), Principles of Data Mining (Adaptive Computation and Machine Learning), *The MIT Press* (August 1, 2001), Ch 6: models and patterns

Karasova, V. (2005), Spatial data mining as a tool for improving geographical models, *Master's Thesis, Helsinki University of Technology*, pp.6-7

Kargar, M.; Isazadeh, A. & Fartash, F. & Saderi, T. & Habibizad Navin, A. (2008), Data-mining by the probability-based patterns, *Proceedings of the 30th International Conference on Information Technology Interfaces (ITI'08),* pp.353-360, Croatia, June 2008, IEEE, Cavtat

Kargar, M.; Mirmiran, R. & Fartash, F. & Saderi, T. (2008), Risk-Management by Numerical Pattern Analysis in Data-Mining, *Proceedings of world academy of science, engineering and technology* Vol. 31, pp.497-501, Austria, July 2008, WASET, Vienna

Kargar, M.; Mirmiran, R. & Fartash, F. & Saderi, T. (2008), Risk-Management by Probability-Based Patterns in Data-Mining, *Proceedings of the 3rd International Symposium on Information Technology 2008 (ITSim'08)*, Malaysia, August 2008, IEEE, Kuala Lumpur

Kuonen, D. (2004), A Statistical Perspective of Data Mining, *published by CRM Today*, December 2004, in CRM Zine (Vol. 48)

Larose, D. T. (2005), *Discovering Knowledge in Data: An Introduction to Data Mining*, Copyright C 2005 John Wiley & Sons, Inc. Ch 1, pp.2-4

Laxman, S.; Sastry, P S. (2006), A survey of temporal data mining, *Sadhana Vol. 31, Part 2*, April 2006, © Printed in India, p.173

McGrail, A.J.; Gulski, E. & Groot, E.R.S. (2002), Data mining techniques to access the condition of high voltage electrical plant, *School of Electrical Engineering, University of New South Wales*, SYDNEY, NSW 2052, AUSTRALIA, On behalf of WG 15.11 of Study Committee 15, 2002

Menzies, T.; Hu, Y. (2003), Data mining for very busy people, *Published by the IEEE Computer Society*, © 2003 IEEE, P.19

Nycz, M.; Smok, B. (2003), Intelligent support for decision-making: a conceptual model, *Informing Science InSITE - "Where Parallels Intersect"*, June 2003, P. 916-917

Ordieres Meré, J. n B.; Limas, M. C. n. (2005), Data mining in industrial processes, *Actas del III Taller Nacional de Miner a de Datos y Aprendizaje, TAMIDA2005*, P.60

Pei, J.; Upadhyaya, Sh. J.& Farooq, F. & Govindaraju, V. (2004), Data Mining for Intrusion Detection: Techniques, Applications and Systems, *Proceedings of the 20th International Conference on Data Engineering (ICDE'04)* © 2004 IEEE

Piatetsky-Shapiro, G.; Djeraba, Ch. & Getoor, L. (2006), What are the grand challenges for data mining, *KDD-2006 Panel Report*, SIGKDD Explorations, Volume 8, Issue 2

Thau Loo, B.; Condie, T. & Garofalakis, M. & Gay, D. E. & Hellerstein, J. M. (2006), Declarative networking: language, execution and optimization, *SIGMOD 2006*, Chicago, Illinois, USA, Copyright 2006 ACM

Wul, X. (2004), Data Mining: An AI Perspective, *IEEE Computational Intelligence Bulletin*, December 2004, Vol.4 No.2

IntechOpen

New Trends in Technologies

Edited by Blandna ramov

ISBN 978-953-7619-62-6

Hard cover, 242 pages

Publisher InTech

Published online 01, January, 2010

Published in print edition January, 2010

This book provides an overview of subjects in various fields of life. Authors solve current topics that present high methodical level. This book consists of 13 chapters and collects original and innovative research studies.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Masoud Kargar, Farzaneh Fartash and Taha Saderi (2010). A Novel Theory in Risk-Management by Numerical Pattern Analysis in Data-Mining, New Trends in Technologies, Blandna ramov (Ed.), ISBN: 978-953-7619-62-6, InTech, Available from: http://www.intechopen.com/books/new-trends-in-technologies/a-novel-theory-in-risk-management-by-numerical-pattern-analysis-in-data-mining

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen