The Research of Data Mining Classification Algorithm that Based on SJEP

Liang Zhao [1], Deng-Feng Chen[1], Sheng-Jun Xu[1] and Jun Lu [2]

[1] *Dept. of Info and Automation, Xi'an Univ. Of Arch. &Tech. Xi'an 710055, China*

[2] *School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China*
Macusezhao@gmail.com

Abstract

Data Mining is dedicated to the analysis and understanding of data, revealing hidden data inside knowledge of technology. Classified as an important research topic in data mining , in recent years , data mining sector presents a new knowledge model, called jumping emerging pattern (JEP: Jumping Emerging Pattern), and there have been some JEP -based classification algorithm . This paper presents a special type of JEP, known as the most effective mode hopping revealed (SJEP: Most Significant Jumping Emerging Patterns). The classification algorithm takes full advantage of the ability to distinguish between JEP strong, while using only the most effective jumping emerging patterns (SJEP) as a basis for classification algorithm enhances endurance noise, reducing the complexity of the algorithm.

Keywords: Data mining, JEP, SJEP, classification algorithm

1. Introduction

With the rapid development of information technology and increasing information-gathering capabilities, resulting in a massive data, these vast amounts of data in the form of static or stored on the physical storage companies, or do not appear to be stored and instantaneous dynamic data. Faced with such a wealth of massive data, the traditional method of data processing and the ability to own cannot meet the actual demand. Faced with increasingly fierce competition in the market, people need to be able to help leaders make decisions to extract knowledge from these contains a wealth of information for decision-making data. Under the strong demand driven, data mining techniques emerged. Data mining is a series of integrated application of advanced technology to extract information and knowledge from people interested in a large database or data warehouse, which is implicit, previously unknown and potentially useful concepts, rules, laws, patterns and so on [1].

Data mining is a multidisciplinary field of research. It combines database technology, artificial intelligence, machine learning, statistical research, knowledge engineering, object-oriented methods, Information retrieval, highly performance computing and data visualization of the latest technology. Data mining is known as one of the backbone of information processing technology in the future, mainly because it is a new concept in changing the way people use data. Twentieth century, database technology has made a decisive outcome and has been widely used. However, database technology, as a basic way of information storage and management is still in online Transaction Processing (OLTP (On - Line Transaction Processing) application as the core, the lack of higher functions such as decision-making, analysis and forecast of support mechanisms. As is known to all, with the expansion of the database capacity, especially in Data Warehouse

(Data Warehouse), and the rising popularity of new Data sources such as Web, on-line analytical Processing OLAP (On Line Analytic Processing), Decision Support, Decision Support) (Classification) Clustering and Classification (Clustering), and other complex applications become inevitable, in the face of this challenge, Data mining and Knowledge Discovery (Knowledge Discovery) technology came into being and shows strong vitality, Data mining and Knowledge Discovery in Data Processing technology has entered a more advanced stage, it can not only to query the past Data, and be able to find out the potential link between the past Data, a higher level of analysis, in order to better make ideal decisions, to predict the future development trend, *etc.* [2].

After ten years of research and practice, data mining technology has absorbed the latest research in many disciplines to form a unique research branch. Needless to say, the research and application of data mining will great challenges. In theory, a number of challenging and forward-looking issues are raised, attracting more and more researchers. The concept of data mining from the 1980s have been proposed, its economic value has been revealed, and is respected by many commercial manufacturers, forming a preliminary market. On the other hand, the current data mining system developed is by no means as some businesses in order to promote their products as magic said, there are still many problems to be studied and explored. The current research on data mining described as a divide stage is more accurate to say the so-called gap phase of data mining technology is widely used before are still many gaps need to climb. In the past few decades, we have the ability to generate and collect data has been rapidly increasing, contributing factors include the bar code is widely used in most commercial products, many business, Computerized scientific and administrative matters, as well as by the text and image scanning system platform to progress in satellite remote sensing data collection tools. In addition, as the popular World Wide Web global information system, has been submerged in the ocean of our data and information in the. Despite the early emergence of simple statistical techniques, but advanced intelligent data analysis tool is not yet mature. Thus, between data generation and data understanding there is a big gap. For example, in relation to the introduction of the current business data mining system, they are some of the common auxiliary development tools, these tools only to those familiar with data mining technology experts or senior technical staff to use only corresponding application development professionals from or transverse acceleration to effect solution .However, from the commercial applications of data mining, and commercial applications due to different application areas will exist great differences, most scholars agree the view of data mining in business success cannot expect universal auxiliary development tools, but should longitudinal data mining concepts and solutions in specific areas of business logic of combining [3].

The explosive growth of data storage business has sparked demand for new techniques and automated tools to help us to convert massive amounts of data into information and knowledge. Thus, data mining - from the large amounts of data in a non -trivial way to discover useful knowledge - has become a natural demand. It is this demand has aroused widespread concern, leading to a vigorous development of data mining research. Analysis of the current status of research and application, there are several data mining research needs to focus on the following issues:

1. Data mining technology and the smooth integration issues is specific business logic. It includes key topics in the field of knowledge of the industry or business knowledge mining constraints and guidance, organic business logic embedded data mining processes.

2. Data mining technology is a particular type of data storage adaptation. Implementation of different data storage mechanisms will affect data mining , targeting , technical effectiveness , etc., rely on a common application model suitable knowledge is found to be effective under all data stored unrealistic "So , for different data storage types of features , targeted research is an issue currently popular for some time in the future but also must face .

3. Architecture and interactive data mining techniques is mining system. Even after years of exploration, the basic architecture of data mining systems and processes had become clear, but by the applications, data mining affect the type and mode of knowledge representation, in particular the implementation mechanism, Technology roadmap as well as the various stages or parts, such as data cleaning, knowledge formation, functional positioning mode assessment still needs refinement and depth. Because data mining is concentrated identify potential advance did not know a lot of knowledge in the source data, and user interactive exploratory mining is inevitable, this interaction may occur at various stages of data mining, from different angles or interact with different particle size. So, a good interactive mining (Interaction Mining) Data mining is also the premise of the system success [4].

4. Data mining language visualization problems with the developed and become an important cornerstone of support for database applications, "However, a word of data mining technology, the birth of late, coupled with its complex OLTP applications compared nature , the development of appropriate data mining manipulation language is still a challenging task. Visualization has become the essential technical requirements of an information processing system for a data mining system for it even more so. In addition to mining and visualization of interactive technology combined with good, it must also result in mining or knowledge model visualization, Mining process visualization and visualization to guide users in areas such as mining exploration and practice.

5. Data mining theory and algorithm. After ten years of research, data mining, machine learning has been as relevant in the inheritance and development of basic disciplines, statistics and other existing achievements have made encouraging progress, explore the many unique theoretical system. However, this does not mean mining exploration theory has ended. On the contrary, it is left to the investigator richer theoretical topic. On the one hand, in these large theoretical framework has many practical applications for mining exploration and theoretical target waiting innovation; hand, with the development of data mining technology itself and related technologies, new birth mining theory is inevitable, and may have a particular role in promoting the application of the new theory is bound to promote the development of a new generation of mining algorithms. The effectiveness of these scalable mining algorithms, such as data mining for some stages, certain data types, and such bulk source data set is more effective to improve the accuracy or efficiency of mining, and can be integrated as a particular CRM application object E commerce.

Therefore, to explore the theory and algorithms for data mining will be a long and arduous task. As can be seen from the above, data mining research and exploration of its rich contents and challenging, "This article is in this context, expand research on data mining.

2. Related Works

Any technology generation always has its technical background, present data mining technology and universal acceptance is due to the development of computer and related technologies to provide the technical foundation of research and application. Database technology from the beginning of the 1980s has been widely spread and applied research and product promotion process in a relational database; people have been exploring the organization and quick access to large data -related technologies. High-performance relational database engine and related distributed query, Concurrency control techniques are used, the database has been improved solve the application ability, quick access to the data, Extraction issues such as integration with accumulated experience. Database technology in the 1980s after a brilliant, people come to realize, is the slave database queries, is found in the master database. "Data only for staff services, not for policymakers service ", which is leading many units lamented after the enthusiastic

database construction issued. Therefore, demand driven, a lot of research on deductive database scholars from the database turned on Inductive database. From an evolutionary perspective, the generation process of data mining technology actually reflects the evolution of database technology [5]. Data warehouse as a new means of data storage and processing generally accepted by the database vendor, and related auxiliary modeling and management tools to market quickly integrate multiple data sources to become an effective technical support environment. Also the popularity of the Internet has provided people with a wealth of data sources. It is said that in the United States , television reaches 50 million spent about 15 years, while working Internet access reached 50 million in just four years, and has been the development of Internet technology itself is not only simple information browse to the core of Web computing information processing technology can handle a variety of information sources Internet environment, so people already use a variety of methods have the ability to store huge amounts of data , the only way of data mining techniques in order to have it handy .The development of a variety of data storage management and access to technology for the research and application of data mining technology to provide a rich soil. Development of computer chip technology enables computer processing and storage capacity increasing, Moore tells us about the key indicators of computer hardware to speed doubling every 18 months, growing, and now it seems there is still ever-accelerating trend followed by a hard drive, the price CUP other key components of the collection makes it a significant decline in the ability and desire to store and process data continues to increase. After decades of development, especially in computer architecture parallel processing technology to mature and widely used, support for large data processing applications become the basis for the development of improved and advanced performance computer architecture enables data mining research and applied technology possible.

After ten years of development, based on both theoretical and technological achievements was Artificial intelligence, *etc.*, have been successfully applied to the analysis of business processes and, to some extent these applications and the development of data mining technology proposed played a great role in promoting. Core data mining techniques and algorithms to support these systems are inseparable from the theory and techniques. In one sense, the theory itself to the development and application of data mining provides a valuable theoretical and applied accumulation. Mathematical Statistics Applied Mathematics is one of the most important and most active subjects, it was born before the invention of the computer, so far, been a few hundred years of history, however, the combination of mathematical statistics and databases should be said that a recent study of more than a dozen technology was only seriously, the previous method based on the application of mathematical statistics mostly through a dedicated program to achieve. In people with the knowledge from the data query to find the interpretation of the data from the data requirements after induction, probability theory and mathematical statistics gained a new vitality, so will the DMKD (Data Mining and Knowledge Discovery) this juncture, Now showing a boom. We know that most of the statistical analysis is based on the strict application of mathematical theory and superb skills, which makes the average user is difficult to control it calmly. Data mining technology is an extension of the development of mathematical and statistical analysis applications, if the way people use the database query from passive to become active discovery of knowledge, then the probability theory and mathematical statistics, this ancient discipline can we conclude from data to knowledge data mining techniques provide a theoretical basis [6].

Data mining technology is that people long for the results of database technology research and development, and also the result of the natural evolution of information technology. From machine learning to knowledge engineering, from knowledge engineering to expert systems, people came back to the 1980s under the guidance of new neural network theory, machine learning, and subsequently entered into a database of

knowledge discovery, and then they produce complementary data mining. During this period, the emergence of data warehouse technology and data mining prosperity gradually mature into a strong technical force, has recently been Recognizing also the statistical analysis and data mining organically combined would be the best strategy. Therefore, data mining is an interdisciplinary; its development is a spiral process.

3. Proposed Scheme

Association rule mining is one of the most active data mining methods. Was first used by Agrawal *et al.*, motivation was originally proposed for market basket analysis issues raised, its purpose is to find a link between commodities trading rules in different databases, these rules characterize the customer buying behavior patterns that can be used guide businesses scientifically arranged stock inventory and shelf design, after many researchers question of association rule mining done a lot of research, their work involves exploring the theory of association rule mining, the original algorithm improvements and new design algorithm. Association rule mining (Parallel Association Rule Mining) and the number of association rules mining (Quantities Association Rule Mining) and other issues, to improve the efficiency of mining rule algorithm , adaptability, availability and application promotion, many scholars have made unremitting efforts [7].

3.1. Jumping Emerging Patterns

Year 2000, Dong, Ramamohanara presents a special knowledge model, that jumping emerging pattern (JEP), which is a very important feature is that it contains a very powerful classification ability , which means that it identifies each data set (or the like) unique characteristics . Therefore, based on research carried out classification on the basis of emerging pattern has very important significance. Here, JEP is defined as an item set, its support in a data set to 0, while in another data set is non-zero - to support growth. JEP use as the basis of the classification is given to JEP a strong distinction between capacity. So, by gathering all the JEP can generate strong classification capabilities.

In this paper, based on the basis of emerging pattern theory, we introduce two new classification algorithms: JEP -based classification algorithm (JEP-Classifier) classification algorithm based SJEP (SEJP-Classifier). First introduced to the former.

JEP-Classifier classification process classification algorithm is as follows:

1 use HORIZON-MINER algorithm to find the horizontal boundary of each class [8];

2 half-truth by repeatedly calling algorithm or algorithms based MBD_LLBORDER border dig out all of the JEP;

3 selection of the most expressive JEP;

4 to calculate the effect of each class of collective support;

5 The test is divided into the collective effect instance class highest score .

We classification algorithm scale JEP-Classifier also conducted experimental analysis , the size of the test data set size classification algorithm for influence , "We chose to focus on the original data from 25% , 50%, 75% and 90% of the instances , as we the new data set. Algorithm JEP-Classifier applied to the four data sets found JEP border. The results of running time dependence isshown in Figure 3.1 (a) and 3.1 (b) below.
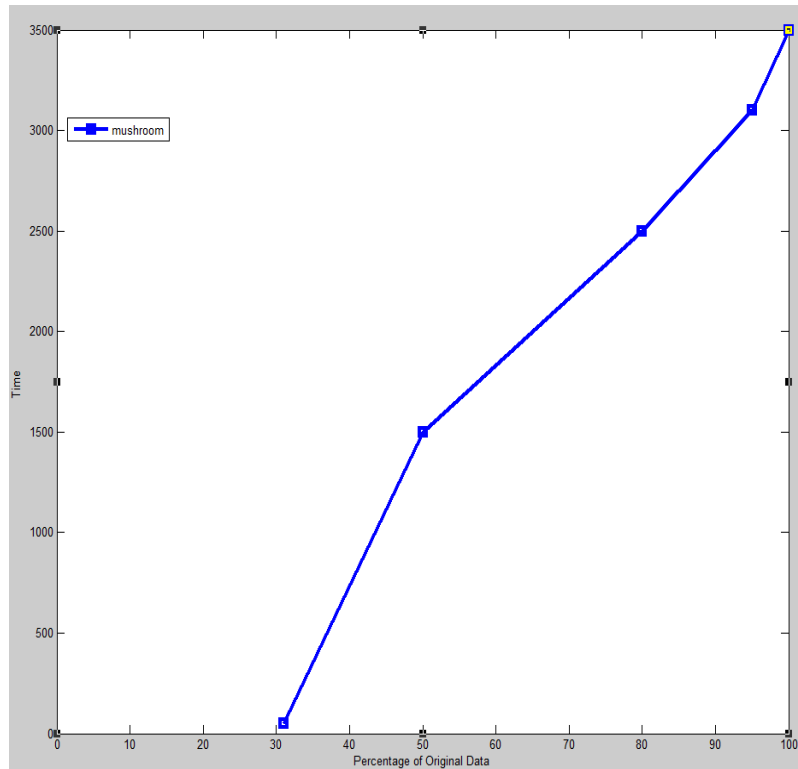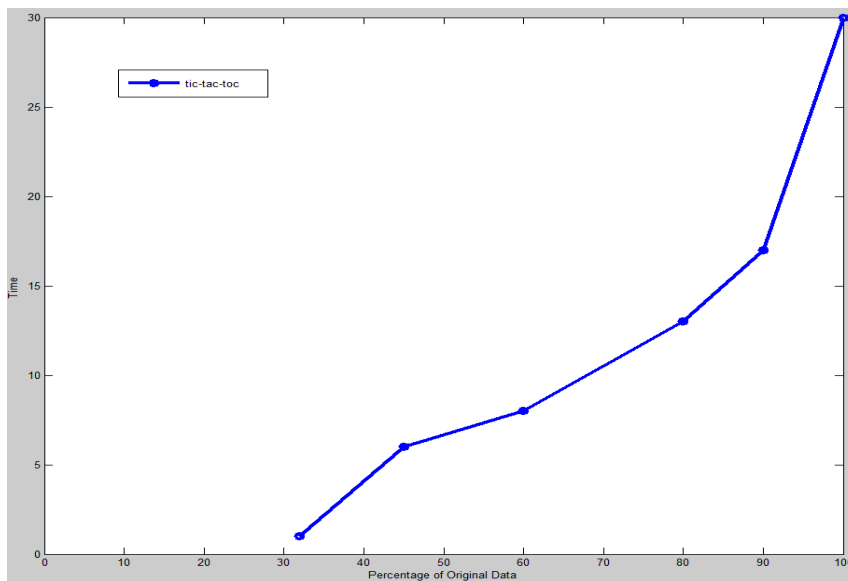
Figure 3.1(a)

Figure 3.1(b)

The classification algorithm based JEP JEP as a basis for classification, and is represented by the boundary JEP, with borders based algorithm to mine JEP, then according to the collective effect of this JEP classification. JEP-Classifier classification algorithm which uses only the most expressive JEP classified, reducing the complexity of the algorithm, enhanced noise endurance. Experimental results prove that with some published classification algorithms (such as C4.5, CAB, and some classification algorithm based on association) compared to the class classification algorithm has higher prediction accuracy. Moreover, the class classification algorithm can be applied to large databases.

3.2. Revealed the Most Effective Mode Hopping

Through the study found, based on the learning phase JEP classification algorithms usually lasting several hours digging out one of the reasons is that a lot of unnecessary JEP. To solve this problem, JEP-Classifiers classification algorithm uses only the most expressive JEP as a basis for classification. For example, the UCI German database, which elected the 32,510 most expressive JEP for classification, but you can see the number of the JEP is still great, so classification algorithm efficiency is still affected, and no major algorithmic complexity amplitude decreases. More importantly, those extra JEP the entire classification does not help, or even increase the noise.For example , the definition of JEP , if there is an entry in a data set only one class or a few times, and in the other class do not appear in the data , then it is also a JEP. Obviously, this classification is JEP useless. Therefore, we propose a special JEP, called SJEP, and support in order to meet the definition of a data class is 0, the data in another class is greater than a given value of the minimum support closing the smallest items sets. As can be seen from its definition, SJEP is a subset of JEP, it removes those items set contains redundant information and noise, their number is much less than the number of JEP 's .

JEP existing mining algorithms often use the border -based algorithm to represent and dig all the JEP. However, this algorithm has two shortcomings : First, require a lot of boundaries to represent all of the JEP; Second, boundary representation does not contain support JEP , and in the subsequent process needs further extract all these JEP the support, the process is very slow (especially for the database that contains many items) . Because of these two problems, making the algorithm is usually based on very low efficiency boundary.For like the UCI Mushroom, Sonar and German databases, algorithms based on the border usually use about two hours to complete mining tasks. Therefore, the paper considers an efficient SJEP mining algorithm (ie, algorithms Mine-SJEPs).

Because Inspired FP- growth algorithm, we will focus on the training data is also compressed all the information stored in the tree expand the tree structure, called SJEP- tree structure. SJEP- FP- tree structures somewhat similar to the tree structure, with the following two differences:

(1) The count field is, SJEP- tree in, you need to modify the count field based on category instance;

(2) On the pointer field is, SJEP- tree has two pointers that horizon link and vertical link,

They allow us to traverse from the bottom up to SJEP- tree, and can be easily accessed SJEP a tree, those nodes with the same item name.

Based SJEP classification algorithm based on the excavation, so SJEP a direct impact on the efficiency of mining efficiency SJEP classification algorithm. This section first discusses mining optimization algorithm, and then gives a classification algorithm based on SJEP.

3.2.1. Mining Optimization Algorithm

Here are a few rules associated with more representative method of mining optimization algorithm, these methods provide for us to improve the existing algorithm can learn a lot of thought [10].

1. To reduce the number of transactions

A basic principle is that when a transaction does not include the length of k is not necessarily contain large items of length k +1 , so that we can be removed these matters, so that over the next scan can be used fewer number of transaction sets .

2. Based on the method of division

1995 , Sava sere *etc.* based on the design of a division (Partition) frequent itemsets generation algorithm , the algorithm first database into several disjoint blocks from the logic, consider every single one block in each data block discuss problems found in the frequent item sets. This method should only block processed into main memory , reducing the memory requirements, but also for parallel processing algorithm provides the possibility to reduce the amount due after the data block , but also makes the processing efficiency is improved.

3. Based on hash values

In 1995 , Park and other proposed algorithm produces frequent item sets based on hash (Hash) technology, they found through experiments, to find the main computing frequent item sets is to produce 2 - the frequent itemsets L2, therefore, Park and other use the nature of the hashing technique is introduced to improve the yield 2 - frequent item set approach. In this way the scanned items on the type of Hash bucket, each project is only possible up to a certain bucket, this can be tested for the items in each bucket subset Selection reducing costs generated Hou, this approach can also be extended to any of the k- frequent item sets generated on .

4. To minimize the number of passes of the database

1997, Brin put forward a reduction in the number of passes of the database algorithms, specific consideration is in the calculation of k- item set, once we consider a (k+1) - item set might be a frequent item sets on parallel calculate, the (k+1) - the total number of scans to support the needs of the algorithm itemsets is usually less than the maximum number of frequent item sets. Experiments show that it uses less number of passes than Apriori, while using less than the candidate set based sampling method, in addition, FP-Tree and Close algorithm we described earlier also fall into this category.

5. Parallel mining

Using data distribution technology to a subset of data mining, but also among a subset of the possible parallel, another parallel mining is also related to the development of parallel processors for parallel mining models and algorithms.

3.2.2. SJEP-Classifier Classification Algorithm: SJEP-Classifier classification algorithm using a new knowledge model (*i.e.* SJEP) as the basis for classification. Here, an item is an attribute, SJEP is an item set, so SJEP-Classifier classification algorithm is based on multi-attribute test. And because SJEP 's support rate is infinite, so it has a strong ability to distinguish ; because it is where the support of a class given support threshold must be met , so it's at least cover a number of training examples dataset ensure SJEP containing less noise.

SJEP distinguish only a single instance of a class of small parts, if applied to all a SJEP On the instance, the accuracy of the classification algorithm is very low. To construct an accurate classification algorithm requires the ability to distinguish between aggregate multiple SJEP, i.e. for each instance, it is first calculated the score in each class, and standardize these scores, the points with the highest standards class is the class of the instance.

For some examples contain SJEP is concerned, a single SJEP can easily distinguish their category. For example, mushroom data sets (obtained from the machine learning database UCL), itemset {ODOR=foul} is a SJEP, which support the edible class 0% toxic class support 55%. If the two test cases included in this SIEP, then we can say with certainty that the instances considered toxic, and not part of the edible kind. However, the examples include this SJEP usually little (and sometimes only 3% of the total number of instances), if we put these SJEP applied to all instances, the predictive accuracy will be

low. Suppose q classes containing C_1, C_2,......,C_q training data set D, SJEP-classifier construction process classification algorithm is as follows.

SJEp_Classifier (D)

{

/*Training phase

All excavated from class $Y_{j=1, j \neq i}^{q} C_j$ to class $C_i (1 \leq i \leq q)$ SJEP;

Out of each class is calculated for all training examples;

Calculated for each group of classes;

/*testing phase

For (each test instance S) {

   S score is calculated at each test case in the class;

S standard test cases calculated in various classes of points ;

S classified into class instance with the highest standards of in ;

  }

  }

4. Experimental Results

Table 1 shows a comparison of several classification algorithm accuracy. Wherein , in the first column is the name of the data set used , the number of instances of the second column is included in the data set , the attribute number and the category number. The third to sixth column in turn gives C4.5, CBA, JEp-Classifier (abbreviated as JEP-C) and SJEP-Classifier (abbreviated as SJEP-C).Accuracy on each dataset. Table 2 compares the various classification algorithms used JEP, CAR and SJEP quantities.

Table 1. Compares the Accuracy

Data set name	Instance, property , class	C4.5	CBA	JEP-C	SJEP-C
Australian	680,15,1	86.1	84.3	85.6	88.5
Mushroom	8123,20,3	-	-	100	99.8
Tic-tac-toe	962,8,1	98.5	99	98	100

Table 2 The number JEP, CAR and SJEP classification algorithm used in comparison

Data set name	JEP	CAR	SJEP
Australian	9905	4655	1350
Mushroom	2980	-	490
Tic-tac-toe	2936	1379	175

As can be seen from Table 1, SJEP-Classifier JEP-Classifier can be obtained almost the same and even higher accuracy, while in most cases higher than the average accuracy of the CAB and C4.5. As can be seen from Table 2, compared to the number of CAR and other JEP classification algorithm used, SJEP-Classifier always use less SJEP.

Because SJEP is a special JEP, so SJEP-Classifier and JEP-Classifier classification algorithms have something in common, but they are, Prediction accuracy of the classification process and the basic idea, learning stage there are still some not same point:

(1) The basic idea: JEP-Classifier classification algorithm using JEP as the basis for classification; SEJP-Classifier uses only a special type of JEP (*i.e.* SJEP) as the basis for classification.

(2) Stage : JEP-Classifier classification algorithm for each class separately using multiple calls to the database as the target boundary based algorithm to mine JEP; SJEP-Classifier classification algorithm will first focus on the training data are compressed all the information stored in an extended the tree structure , and then traverse the tree one time digging through all SJEP.

(3) classification process : JEP-Classifier classification algorithm derived by aggregating JEP score first , and then the class with the highest total score as the final classification category ; SJEP-Classifier classification algorithm first derived by aggregating SJEP score , then out of standardization, finally, with the highest standards of the class as a final classification category.

(4) The predictive accuracy: SJEP-Classifier classification algorithm achieved and JEP-Classifier classification algorithm is almost the same or even higher prediction accuracy.
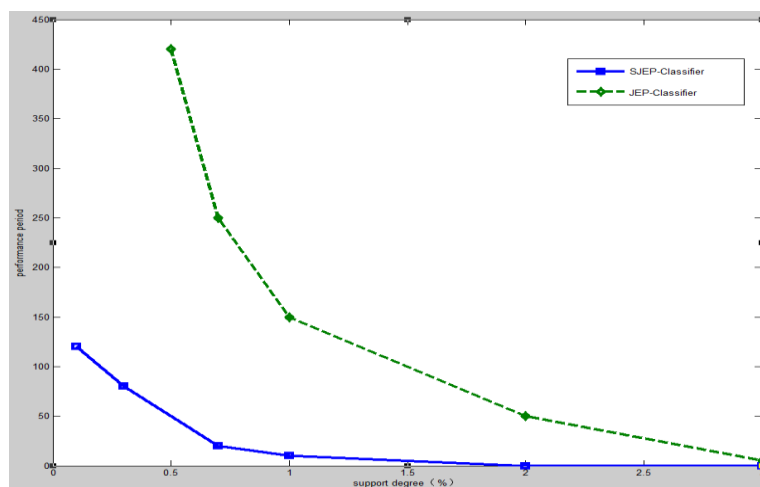

Figure 4.1. Algorithm SJEP-Classifier Compared with JEP-Classifier

5. Conclusions

This paper proposes a new classification algorithm based SJEP (SJEP-Classifier). The classification algorithm takes full advantage of the ability to distinguish between JEP strong, while using only the most effective jumping emerging patterns (SJEP) as a basis for classification algorithm enhances endurance noise, reducing the complexity of the algorithm. Experiments show that, SJEP-Classifier than CAB and C4.5 classification algorithm has higher classification accuracy than the JEP-Classifier with higher efficiency.

Acknowledgements

References

[1]   J. Han and M. Kmaber, "Data Mining ConeePtsnad Techniques", Morgna Kaumfann Publishers, **(2008)**.
[2]   R. Agrawal, *et.al,* A Mining assertion rules between sets of items in large databases Rore.ACMSIGMODConfonManagementofData.2 003:207216
[3]   S. Chuadhuri *et.al*. "An overview of data ware housing and OLAP technology ACMSIGMOD Record", **(2007)**, vol.26, pp. 6574.

[4] Hnajetal, "Generalization-based data mining in object-oriented data bases using an Object-cube model", Data knowledge Engineering, **(2008)**, vol. 25, pp. 55-97.

[5] A. Capiluppi, J. F. Ramil, "Studying the Evolution of Open SourceSystems at Different Levels of Granularity", Two Case Studies, Proceeding of the 7th International Workshop of Principles of Software Evolution, **(2004)**, pp. 113 - 118.

[6] J. E. Istiyanto, A. W. R. Emanuel, "Success Factors of Open SourceSoftware Projects using Datamining Technique", Proceeding ofInformation Technology and Communication International Seminar (ITIS), June **(2009)**.

[7] G. Dong and J. Li, "Efficient mining of emerging Patterns: Discovering trends and differences", In proc. Of KDD 99, **(2005)**, pp. 15-18.

[8] J. R. Bayard, "Efficiently mining 1ong patterns from databases", In Proceedings of the 1998ACM-SIGMOD International Conference on Management of Data, **(1998)**, pp. 85 93.

[9] J. F. Rosen and C. Stone, "Classification and Regression Trees", Madwort International Group, **(2000)**.

[10] R. D. Hart, "Pattern Classification and Scene Analysis", New York, John Wiley Sons, **(2003)**.

Authors
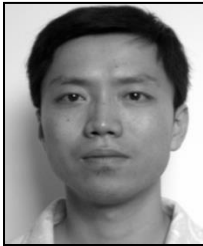
Liang Zhao, he was born 1980. He completed his Ph.D. degree in Xi'an University of Architecture and Technology. He research focus on computer control, measure and detection technique.

Deng-Feng, he completed his Ph.D. degree in Northwestern Polytechnic University. He mainly focus research on computer detection technique, control engineering.

Sheng-junXu, he received Ph. D. degree from Xi'an JiaoTong University. His research interests include image processing, computer vision, and pattern recognition.

Jun Lu, she received M.S degree from XiDian University. Her research interests in computer control, information system.