Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2013, 5(12):553-559

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Information fusion of ITS based on granular computing

Seng Dewen, Wu Hao, Fang Xujian and Xu Haitao

School of Software Engineering, Hangzhou Dianzi University, Hangzhou, China

---

ABSTRACT

The construction of Intelligent Transportation Systems (ITS) occupies a crucial position in the current wave of smart city. Effective and efficiency ITS needs two important conditions: plenty of traffic data and effective means of data analysis. Multi-source, heterogeneous, vague, uncertain traffic data fusion and sharing is the focus and difficulty of current research and application of ITS. The granular computing demonstrates a unique advantage in the information analysis and processing of massive, vague, uncertain and incomplete data. In this paper, we study the traffic information granular computing theory and build traffic information fusion model, framework and implementation program based on granular computing. We raise uncertainty reduction algorithms for traffic flow prediction and congestion recognition algorithms based on granular computing theory, which will provide new ideas and methods in the complex decision making under uncertainty problems of the transportation systems.

Key words: Information fusion; granular computing; intelligent transportation

---

INTRODUCTION

The essence of smart city is more thorough perceptions, more extensive connections, more focused and more in-depth calculation [1]. Traffic is the lifeblood of the city development. Intelligent transportation plays an important role in the smart city [2]. The traditional urban traffic management model is difficult to solve population growth, soaring trips, rapid increase of private cars and the increasing congestion in the city. Increase roads alone cannot ease the sharp contradiction between the traffic demand and the supply of transport facilities. Many cities' transportation systems have been on the verge of collapse and have to take extreme measures. Even more cities are difficult to run in severe congestion. The transportation system as show in Fig. 1 is a complex giant system with nonlinear, randomness, variability and uncertainty [3]. Separate from the vehicle or separate from the road considerations are difficult to fundamentally solve the problems [4]. In this context, transport infrastructure, transport devices and traffic participants are combined into a system and take full advantage of the high-tech to solve the traffic problems, that is, the intelligent transportation systems (ITS). With the successful applications of information technology in various fields, the applications of ITS have become internationally recognized the most effective way to solve the traffic problems [5-10].

Data processing has a very important role in every information systems [11]. The traffic data is the research base of all traffic-related problems. The core and key to implement ITS is plenty of data as well as effective data analysis and processing. With a large number of various types of data acquisition equipments put to use and the rapid development of communication technology, traffic data has been greatly improved from quantity to quality, which provides the basic support for a better grasp of the traffic status and more effective ways to solve traffic problems, but at the same time, this is also a challenge to traffic data processing. Face with the massive diverse traffic data, the data fusion and data mining to avoid overlapping and wastage, and discover the internal relations so as to get the real traffic state is very important. Quick smart data analysis to reduce the reaction time and to meet the demand for real-time traffic control is the proper meaning of modern traffic data processing technology and the urgent need.

Granular computing is a new hot spot of the current research in the field of intelligent information processing [10-15]. It takes the idea of "granularity" as a guide to get the research objects' definition, metrics and reasoning. It analyzes and deals with the problem from different angles and different levels in order to achieve an optimal approximate solution. Granular computing demonstrated a unique advantage in the analysis and processing of massive, vague, uncertain and incomplete information. Granular computing theory has been successfully applied to the field of pattern recognition, image processing, intelligent search, but in traffic information processing, there is still lack of a complete theoretical system and practical examples. In our research, we use granular computing to deal with mass uncertain, multi-source and heterogeneous traffic data.

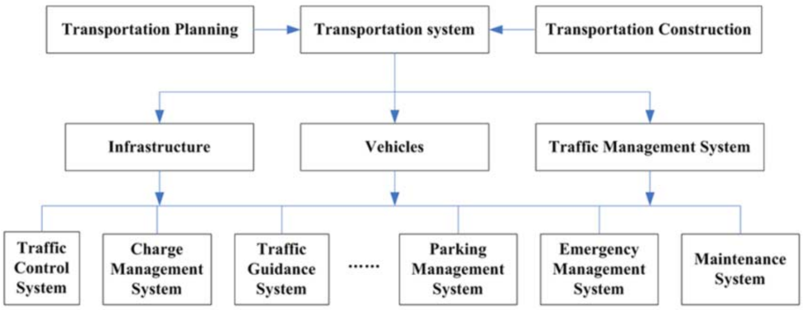

Fig. 1: Complex transportation system

TRAFFIC INFORMATION FUSION TECHNOLOGY

With the advent of the information age, there is more and more number of traffic information. The sources of traffic information are more and more widely and related to increasingly complex levels. The widespread multi-source heterogeneous data in traffic areas make many traffic engineering problems become typical data fusion problems. Specifically, the data fusion in the transport system is to reasonably control and use the information from various sensors, measurement data, statistical data, empirical data, which are real-time or non-real-time, speed change or gradient, fuzzy or identified, complementary or conflicting. Combining the redundant or complementary information in the space or time based on certain criteria, the data fusion is used to explain or describe the consistency of the object being observed. The goal of data fusion is to export more effective information through the optimal combination of data. This is the result of the best synergies. Its ultimate objective is to take the advantage of the multiple sources collaborative work to improve the effectiveness of the entire system. The emergence of data fusion and intelligent decision-making technologies provides intelligent information processing methods for transportation systems.

Traffic information fusion methods include probability methods, logic inference methods, and learning methods. These traffic information fusion methods starts mainly from the perspective of traffic flow data uncertainty, conflict, ambiguity fusion. These models or algorithms are assumed the information needed automatically satisfied, and then data processing such as data conversion and data fusion reasoning are based on this assume. It does not consider the problems of dynamic traffic information sources. Most cities' traffic information system facilities are still not perfect. Loop coil is the primary means of the automatic acquisition of the traffic status information. Road capacity study, road construction and use are not up to the standard. These lead to that the road capacity is difficult to measure accurately. Therefore, the direct application of these models or algorithms has a certain degree of difficulties and problems.

Compared to other data processing technology, information fusion technology has the following advantages [16-20]. It can take appropriate data analysis for data from multiple data sources, and get more accuracy results than single data source analysis. Also, it will take full advantage of the data collected by each of the single data source, and improve the utilization of the data and the utilization of existing hardware. It can make up the information for the lack of sensors or sensor node failure, and improve the robustness of the systems. It is consistent with the systems' characteristics of "high cohesion, low coupling", so the stability and scalability of the entire system will be improved. For example, in the emergency management information database, information fusion model includes transport facilities, passenger traffic information, road traffic status and geographic information data. We need to follow certain standards of these dynamic data to format and digitalize these data, and form a unified feature representation to achieve feature level fusion. The fusion data will be stored in the database for data processing and analyzing in next stage. Traffic information fusion is the focus of the current ITS research and application. At present, the main problems are as shown in Fig. 2.

Traffic information is difficult to share. Traffic information collection, management, processing, and release system have not unified interaction interface, so it is difficult to integrate various systems. Due to the independence of the systems, each system has its own set of databases and data formats, resulting that the data cannot be shared between the systems. So there are numerous data "islands" in traffic system.

Dynamic traffic data analysis is lack of deep-level processing methods. For example, with the development of urban transport infrastructure, beacon and capture-identification methods to obtain a wide range of traffic information have become common acquisition methods for traffic flow data. The current lack of deep analysis of traffic flow data results in that a part of the traffic flow parameters cannot be used. And current research applications using static Origin-Destination analysis method, the results are used to describe traffic state between the origin and the destination point within a relatively long time, and assumed that the traffic conditions in this period is uniform. The dynamic characteristics such as generation and dissipation of congestion, time-varying traffic sections and the path flows, travel time and congestion spatial distribution changes are unable to obtain.

Traffic information has no effective fusion methods. The original traffic information acquisition systems are independently. They get data in different ways. When the traffic information appears contradictory, the systems cannot give accurate and credible traffic information. The traffic flows are often affected by many factors. In the process of traffic information fusion, uncertainty, conflict and ambiguity should all be considered. The current traffic flow information fusion algorithms do not take into account all these factors. This causes the fusion results and the actual value a certain error and cannot be applied to the actual environment.

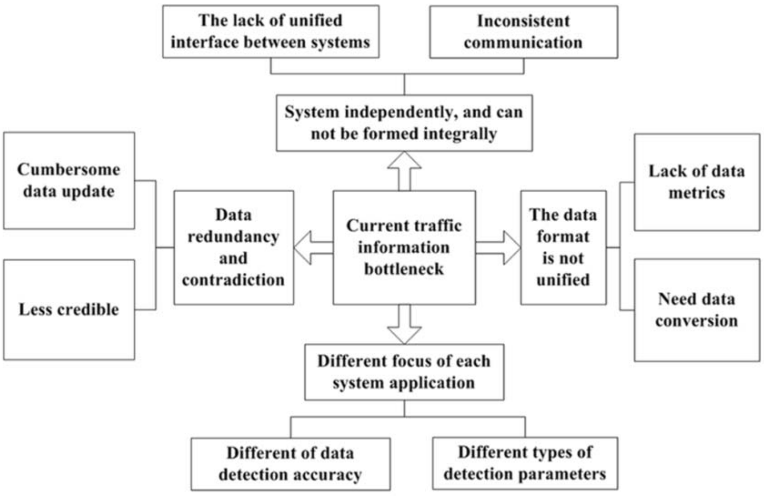

Fig. 2: Main data problems of current ITS application

GRANULAR COMPUTING

Granular computing [10] concept was first proposed by T.Y. Lin in solving the multi-level problem in 1997. Since then, granular computing has aroused the attention of many scholars, and become the hot spot of the field of artificial intelligence research. It has become an important tool in the research of vague, imprecise, incomplete and massive information processing. An accepted feature of human intelligence is that people can observe and analyze the same problem from very different granularity. People not only can solve problems in the world of different granularity, but also can skip quickly from a granularity world to another. The ability to handle different granularity world is a powerful performance of human problem solving. This feature of human intelligence is the basic idea of granular computing. The essence of granular computing is to reduce the difficulty of the problem by selecting a suitable granularity so as to find a better solution.

Granular computing is a simulation of human global analysis capability. The computer's processing speed is far greater than the speed of the human brain, but the computer is not as intelligent as human beings. This is mainly because humans have a very strong global analysis capability to turn complex problems into a relatively simple model from a variety of different size or level. Granular computing is a computational paradigm of information processing, covering all the granularity related theory, methods, techniques and tools. Most of the researches on granular computing are theoretical research, and less in the applications. The analysis, discussion and application of granular computing model to specific problems are an urgent need for the study.

ROUGH SET THEORY AND TRAFFIC FLOW PREDICTION

Rough set theory is a mathematical tool for fuzzy, uncertain, incomplete knowledge [20]. Its main idea is to export decisions or classification rules by knowledge reduction while maintaining the classification ability. Rough set theory early has been successfully used in decision analysis, process control, pattern recognition and data mining. In recent years, with the rise of artificial intelligence theory, rough set as an effective tool for data processing in machine learning, knowledge acquisition, knowledge discovery field has been extensive research and application, such as expert systems, speech recognition, fault diagnosis, intelligent dispatch system analysis of the cause of the incident, and traffic flow characteristics analysis. The basic idea of the rough set is to use equivalence relation to classify the domain objects and generate a partition for problem domain. Based on the principle that elements of same equivalence classes cannot be distinguished, it builds information granularity of different equivalence classes, to achieve the purpose of information reduction [21]. Rough set theory has significant advantages in data analysis and data processing, mainly reflected in the following aspects.

Let U be a set of objects called the universe and R be an equivalence relation on U, then for any X belongs to U, we can associate two subsets with X:

$$R_*X = \{x \in U | [x]_R \subseteq X\}$$
$$R^*X = \{x \in U | [x]_R \cap X \neq \phi\} \quad (1)$$

where $[x]_R$ is the equivalence class containing x. R_*X and R^*X are called the R-lower and R-upper approximation of X respectively. The pair (R_*X, R^*X) is called the rough approximation of X and any such pair is called a rough set. The pair (U, R) is thus called a Pawlak approximation space (PAS).

To relax the constraints on R is one of the direct generalizations of the above-mentioned idea. To allow R to be an arbitrary binary relation, different useful generalizations of the Pawlak rough set model can be obtained. To distinguish the PAS and the generalized one, the latter is referred as relational approximation space (RAS). When (U, R) is an RAS, the lower and upper approximations of a set X will be modified as

$$R_*X = \{x \in U | R(x) \subseteq X\}$$
$$R^*X = \{x \in U | R(x) \cap X \neq \phi\} \quad (2)$$

where $R(x) = \{y \in U | (x, y) \subseteq R\}$.

Even further generalizations of RAS are possible, such as the neighbourhood systems. A neighbourhood system (NS) is a pair (U, N), where $N: U \longrightarrow 2^{2^U}$ satisfies $\phi \notin N(x)$ for all $x \in U$, and for all $x \in U$ and $X \subseteq Y \subseteq U$, if $X \in N(x)$, then $Y \in N(x)$.

The universe U is open if $U \in N(x)$ for all $x \in U$, or equivalently, $N(x) \neq \phi$ for all $x \in U$. The lower and upper approximations of a set X in an NS (U, N) are based on the definition of interior and closure in topology [22].

$$N_*X = \{x \in U | \exists Y \in N(x), Y \subseteq X\}$$
$$N^*X = \{x \in U | \forall Y \in N(x), Y \cap X \neq \phi\} \quad (3)$$

Given an RAS (U, R), an NS (U, N) can be defined by $N(x) = \{S \subseteq U | R(x) \subseteq S\}$, so the latter is indeed a generalization of the former.

For a PAS or RAS (U, R), and $X \subseteq U$, the accuracy of the approximation of X can be defined by

$$\rho(X) = \frac{|R_*X|}{|R^*X|} \quad (4)$$

Furthermore, the rough membership function associated with X is defined by $\mu x: U \longrightarrow [0,1]$

$$\mu x(u) = \frac{|X \cap R(u)|}{|R(u)|} \quad (5)$$

This provides a numeric characterization of rough sets. A variable precision rough set model is proposed based on the definition of rough membership function [22]. For $0 \leq \alpha < \beta \leq 1$, the α and β approximation of X is defined by

$$R_{\alpha *}X = \{u \in U | \mu x(u) \geq 1 - \alpha\}$$
$$R^*_{\beta}X = \{u \in U | \mu x(u) > 1 - \beta\} \quad (6)$$

The rough membership function and the accuracy of approximation are well-defined for finite universe U; however, it uses the cardinality which may be not finite in the infinite case. To cope with this situation, the RAS is extended to probabilistic approximation space (PRAS). A PRAS is just a triple (U, R, Pr), where (U, R) is still an RAS and Pr is a probability distribution on U. Then the definition of accuracy and rough membership function can be replaced by the following equations:

$$\rho(X) = \frac{Pr\ (R_* X)}{Pr\ (R^* X)} \quad (7)$$

$$\mu x(u) = \frac{Pr\ (X \cap R(u))}{Pr\ (R(u))} \quad (8)$$

For convenience, $\mu x(u) = 1$ if $PPr\big(R(u)\big) = 0$. When U is finite and Pr is a uniform distribution, the definitions just reduced to the original ones [22].

Rough set theory analyze only data itself, no need any prior knowledge or additional information and subjective evaluation given in advance. This is also the most important difference between rough set theory and other data processing mathematical tools for fuzziness and uncertainty. Conventional data processing method is a hard computing method. It use the accurate, fixed and unchanging algorithm to express and solve the problems, while the rough set theory is a soft computing method allowing the use of imprecision, uncertainty and partially true data in order to become more easy to handle, and get strong robustness and low cost solutions. The rough set theory has a powerful data analysis capability to express and deal with incomplete information. It can obtain the minimum expression of knowledge by simplifying the data with the key information still being retained. It is able to identify and assess the dependencies between the data to reveal the simple patterns of the concepts. The rough set theory provides a series of mathematical algorithms to various analysis and data processing operations. The algorithms are simple and easy to analyze, and the conclusions obtained are very understandable and easy to verify. At the same time, the rough set theory has its own shortcomings. When using the rough set theory to process data, the continuous quantitative and qualitative description of the property value need to be discrete. The determination of the discrete intervals requires manual operation, which has a direct impact on the data processing.

The traffic flow forecasting is the real-time prediction of traffic flow at time t for the next decision time. It is an important research content of dynamic traffic control and guidance. The accuracy and reliability of the real-time prediction results are the basis of dynamic traffic management.

In this study, we use the advantages of granular computing in dealing with massive, heterogeneous traffic data to build a traffic flow forecasting model. First, we use the rough set theory to construct the input parameters of the prediction models to distinguish the information granularity, exclude weak correlation parameters with traffic flow prediction, and determine a set of input parameters for the prediction models. Secondly, we construct mapping relationship information granularity of traffic input parameters set and prediction traffic flow sets, extract the relation rules, and thus use these rules to construct flow prediction information granularity, that is, the neural network prediction model for short term traffic flow forecasting, and then we study the space-time extension of the prediction models.

Compared with the other predictive models, our model use rough set to pre-process the traffic data and exclude the invalid and redundant data. Therefore, it greatly reduces the amount of computation. The prediction results are real-time and high accuracy. At the same time, due to the neural network model for prediction is built by the association rules, we get a better solution for the "black box" of the traditional neural network, which makes the model become good transparency, easy to explain, good ability of generalization, and strong anti-interference ability. The comparison of artificial neural networks (ANN) and rough set theory prediction values and actual values of traffic flow are as show in Fig. 3, which shows that the prediction values of rough set theory is closer to the true values.

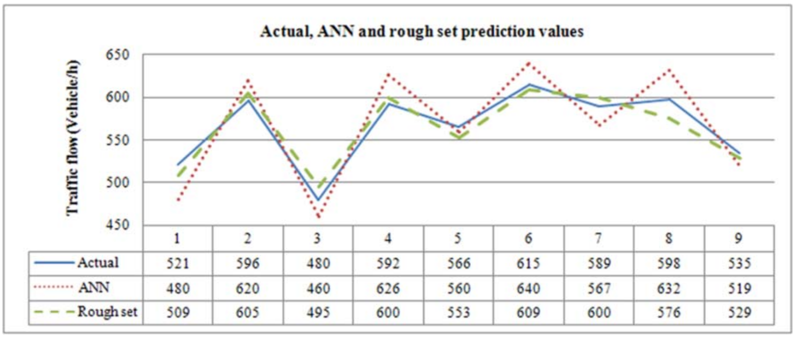

Fig. 3: Prediction and actual values of traffic flow

QUOTIENT SPACE THEORY AND TRAFFIC CONGESTION DETECTION

When referring to the classification, we always explicitly or implicitly determine the classification standard. Different standards will get different classification results. For a given a classification result, we can enhance or relax their corresponding classification standard conditions thereby to obtain more detailed than the original classification results or more rough division. For two classification standards given, we can do "and" computing, or "or" computing, then we can get a new classification standard, and get the new division. In other words, human beings can do appropriate adjustments to the classification according to their needs, so the problems will be solved in a complete different level. Zhang B. and Zhang L. proposed the quotient space theory of problem solving [23]. Quotient space theory for solving problems at different levels of granularity has aroused great concern of international scholars and peers, and has become one of the three important models of granular computing.

The causes of traffic congestion are very complicated. The impact factors mainly include three categories: traffic flow characteristics, such as car speed, traffic flow and share; road characteristics, such as road grade, lanes, capacity; traffic control measures, such as setting a one-way street, limiting speed, traffic signal control. To achieve the traffic congestion recognition, the above factors are indispensable. Some of the existing models cannot complete the massive, fuzzy data analysis, and some even completed but will take long computation time, so they are unable to meet the real-time traffic control requirements.

In this paper, we take the advantages of the quotient space theory in dealing with massive, fuzzy data to establish the congestion impact factor index system, and build congestion recognition model based on quotient space theory. The model is based on the raw data. The use of the quotient space theory normalized data, reduce the vector dimension, and improve the efficiency of model operation. Meanwhile, the appropriate level of granularity is selected through the quotient space. The fuzzy set theory will be used for clustering analysis and determining the congested roads in the road network. The basic principle of traffic congestion recognition based on granular computing theory is to use fuzzy quotient space theory for the hierarchical cluster analysis of sample data. First, find the optimal clustering hierarchical by quotient space theory attributes, reduce the impact factors for the evaluation system, and reduce the amount of computation, thereby improve the efficiency for the clustering in next step. Secondly, use the fuzzy similar matrix to build dynamic clustering pedigree chart. Finally, use the quotient space theory in different granularity, select a reasonable threshold, and get the final clustering results. The comparison of the prediction accuracy of ANN and quotient space theory of traffic congestion is as show in Fig. 4, which shows that the prediction accuracy of quotient theory is higher than the ANN.

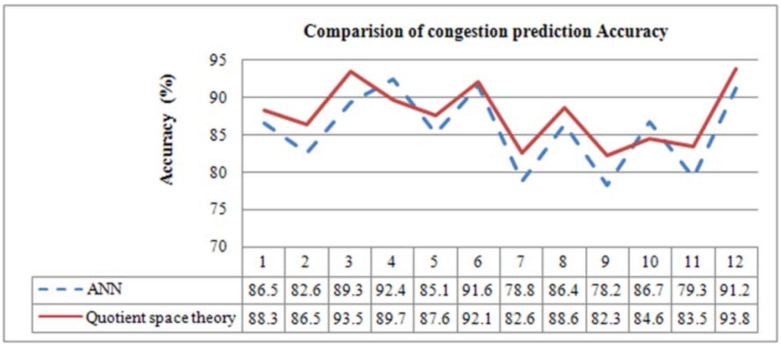

Fig. 4: Congestion prediction accuracy of ANN and quotient space theory

CONCLUSION

Quotient space theory and rough set theory are granular computing models, and are discussed in the framework of set theory. They regard the granularity as a subset of the problem domain. They are not mutually exclusive, but different in main focus. From the model point of view, both are described in the model of the ability of human to deal with things in different granularity. They believe that the concepts can be expressed in subsets. Different concepts of granularity can be used different sizes of subsets. All of these representations can be described by equivalence relations. Form the research object point of view, the quotient space theory and rough set theory all take the collection of objects discussed as the problem domain. When discussing the relationship between the objects, they are different. The prototype of rough set theory is probably derived through relational database abstraction, that is, use different attribute values of the elements to describe the relationship between the elements and represent different concept granularity according to the classification of different attributes. The prototype of the quotient space theory is Hierarchical method that in addition to the properties of the elements, it also introduced the relationship between the elements. In our approach, we use rough set theory to predict the traffic flow and use quotient space theory to predict the traffic congestion. The results show the effectiveness and efficiency of our methods.

Acknowledgments
The authors wish to thank the National Natural Science Foundation of China for contract 61100194, the foundation of education department of Zhejiang province of China for contract Y201120520 and the research foundation of Hangzhou Dianzi University for contract KYS105612008 and YB1205, under which the present work was possible.

REFERENCES

[1] Paskaleva K. *International Journal of Innovation and Regional Development*, v.1, n.4, pp.405-422, 25 January, **2009**.
[2] Arora A.. *Traffic Engineering and Control*, v 53, n 10, p 375-378, November **2012**.
[3] B. Chen and H. H. Cheng. *IEEE Trans. Intelligent Transportation Systems*, vol.11, no.2, pp.485–497, **2010**.
[4] Seng Dewen. *Applied Mechanics and Materials*, v. 189, p. 482-485, **2012**.
[5] SENG Dewen, LI Zhongxue. *Journal of Liaoning Technical University*, Vol.27(1), pp.9-12, **2008**.02.
[6] Dikaiakos M. D.. Minersoft: *ACM Transactions on Internet Technology*, v.12, n.1, June **2012**.
[7] Seng Dewen, Shu Yueqing. *Advances in Intelligent Systems and Computing*, p. 393-400, **2013**.
[8] Di Lecce, V. Amato, A. *IET Intelligent Transport Systems*, v 5, n 3, p 149-158, September **2011**.
[9] Prashant T., Kavita D., Manish S.. *International Journal of Soft Computing and Engineering*, v.2, July **2012**.
[10] LIN T. Y. Granular Computing: From Rough Sets and Neighborhood Systems to Information Granulation and Computing in Words. European Congress on Intelligent Techniques and Soft Computing, Heidelberg, Germany, pp.1602-1606, September 8-12, **1997**.
[11] Seng Dewen. *Applied Mechanics and Materials*, v. 189, p. 478-481, **2012**.
[12] Zadeh L A. *Fuzzy Sets and Systems*, **1997**, 90(2): 111-127.
[13] Seng Dewen. *Applied Mechanics and Materials*, v. 189, p. 457-460, **2012**.
[14] Seng Dewen, Shu Yueqing. *Advances in Intelligent Systems and Computing*, p. 401-408, **2013**.
[15] Yao Y Y. *IEEE Conference on Granular Computing*, pp.85- 90, **2005**.
[16] Zhang Q, Ma Z.K, Wang L. *Pattern Recognition Letters*, v.34, n.2, pp.185-193, January 15, **2013**.
[17] Du P. J, Liu S.C., Xia J. S. *Information Fusion*, v.14, n.1, pp.19-27, January **2013**.
[18] Shang L., Liu G.H., Zhang R. *Acta Astronautica*, v.85, pp.33-40, **2013**.
[19] Wu S.L. *Information Processing and Management*, v.49, n.1, pp.108-122, January **2013**.
[20] Skowron A., Stepaniuk J., Swiniarski R.. *Information Sciences*, v.184, n.1, p.20-43, February 1, **2012**.
[21] Pawlak Z. Rough Set: Theoretical Aspect s of Reasoning about Data. Kluwer Academic Publishers , Boston, **1991**.
[22] C.J. Liau. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, V. 8, No.1, pp.93-118, **2000**.
[23] Zhang Ling, Zhang Bo. The Quotient Space Theory of Problem Solving. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing Lecture Notes in Computer Science, V.2639, pp.11-15, **2003**.