Research on Data Mining Model of Intelligent Transportation Based on Granular Computing

Xiao-Lan Xie and Xiao-Feng Gu[*]

[1]College of Information Science and Engineering, Guilin University Of
Technology, Guilin ,Guangxi Zhuang Autonomous Region, China, 541004
[2]Guilin University of Technology, GuangXi key Laboratory of Embedded
Technology and Intelligent Information Processing, China, 541004
[1]College of mechanical and control engineering, Guilin University Of Technology,
Guilin, Guangxi Zhuang Autonomous Region, China, 541004
gxf199295@163.com

Abstract

Through the analysis of the relationship between current intelligence transportation and large data, after the theory of granular computing concepts involved in preliminary studies, For the current data mining of intelligent transportation ,we presents a Data Mining Model Of Intelligent Transportation Based on Granular Computing. Utilizing granular computing in data mining theory advantages, through constructing a new data mining model to solve the tlarge-scale, complexity, uncertainty and ambiguity problems of massive data from intelligent transportation.

Keyword: Intelligent Transportation; Granular Computing; Data Mining ;Data granulated

1. Introduction

With the process of modernization's development, intelligent transportation has accumulated a mass of complex data. As a result, intelligent transportation raises new requirements on data mining. By this way we aim to increase the efficiency of transportation management and transportation service, through processing the data and comprehensive application. The service of level of transportation system is expected to improve, and the incidence of transportation accident is also hoped to reduce. So far, China intelligent transportation system has stepped into the stage of practical exploitation and application from the exploration stage. In term of application area, the intelligent transportation in China is mainly applied to informationization of transportation, urban road transportation control service, and urban public transportation.

The basic thought of granular computing, which is the world view and methodology as the new emerging object in data mining, is to use the granular in the process of solving the problem. The granule is the group, class or clustering of element. The description of granular structure and the granular computing can be described from two dimensions. The former one focuses on the formation, expression and explanation of the granule. The latter one is to apply the granular to the process of solving problems. Granular computing has already been one of the main methods of soft computing. More details about the theory of granular computing will be introduced in the following.

---

[*] Corresponding Author

2. The Basic Concept and Theory of Granular Computing

2.1. The Basic Theory of Granular Computing

2.1.1. Information Granule, Granular Layer, and Granular Structure:

(1) Granule

In different granular computing models, the granule is the simple block of the most basic elements, which delivers clearer meanings under different forms of the characteristics and performance of data and information.

Definition 2.1 Granule: Provided there is the relationship R of domain of discourse between U and U that $U \rightarrow P(U) \Rightarrow U = U_{i \in \tau} G_i$, one G_i is a granule, of which R is some kind of relevant relation.

Definition 2.2 Granularity: Provided there is the granulation $U_{i \in \tau} G_i$ of domain of U and U, the granularity of G_i is $d(G_i) = card(G_i) = |G_i| = \int_{G_i} dx$.

The granularity is the measuring standard for the "size" of the granule. A quantity of granules in different granularities form the granular layer. Generally speaking, the smaller the granular layer is, the clearer the description to the problem will be.

(2) Granular Layer

In most cases, granular layer is constituted by information granule measured by the granulating standard. In the same layer, all granules share thesame nature. In different layers, granules in higher layers are coarser than the counterparts in the lower one. Thus, the lower layer would be preferred to analyze problems more specifically, by which the addition of parts of nature would deliver more micromesh divisions.

(3) Granular Structure

The granular structure is the relation structure consisted by the interrelation among different granular layers, under the different granulating criterions. It is the combination of either a single multi-layer structure, or multiple layered structures. As a result, granular structure is able to analyze the system from multiple perspectives, more visualized and deepgoing.

2.1.2. Information Granularity and Granular Computing:

(1) Information Granularity

Information granularity is a structural process. Simply speaking, it is the process that a granular layer comes out from the setting granulating criterion; is the creation of the basic unit of granular computing, including granule, granular view, granular net and hierarchical structure. Usually we decompose the coarse granule to get the fine granule by the top-down granulating method, or combine the fine granule to get the coarse granule through the down-top method.

(2) Granular Computing

Granular computing is usually used to solve the practical problem based on the different granules, granular layers and granular structures. The mutual transformation of the same problem in the same layer or among different layers helps us to raise the efficiency to solve these complex problems. Based on the deep analysis and study of granular computing theory, we discover that most discipline principles are interrelated. Meanwhile, granular computing and discipline are mutual independent. Once the thought

of granular computing is applied to solving the practical problem, granular computing would be likely to be applied to any area.

3. A Data Mining Model of Intelligent Transportation Based on Granular Computing

Due to the uncertainty, ambiguity, large-scale and complex of the massive transportation data, for mining analysis of integration, pretreatment and subsequent across multiple data centers of mass urban transport information data, and unpredictable traffic system behavior, the task of urban traffic data mining reaches the bottleneck, using existing data mining algorithms is very difficult. So we propose the data mining model of intelligent transportation based on granular computing here.
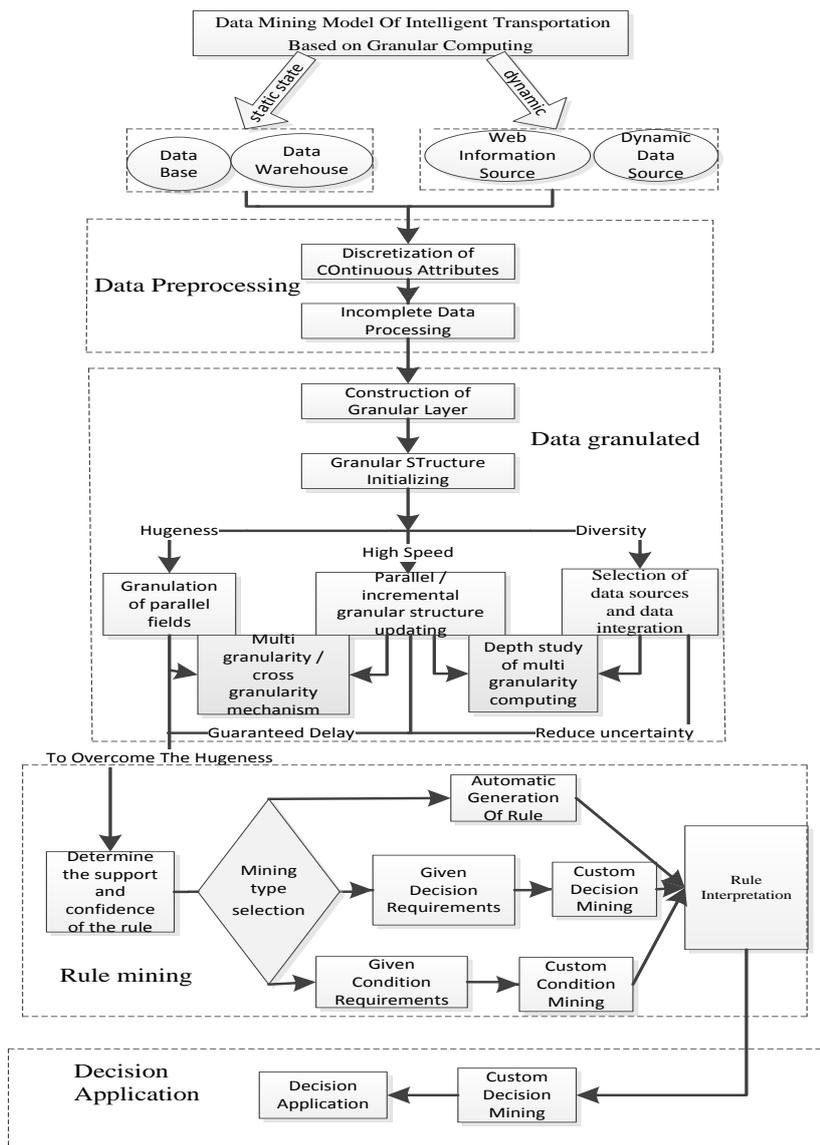
Figure 1. Data Mining Model of Intelligent Transportation based on Granular Computing

(1) Data Preprocessing

When we conduct intelligent transportation data excavation, after screening data preliminarily from a data source, it must first is data preprocessing.

For the most data mining algorithms, in order to continue to use, the continuous attribute data must be discreted. Discretization of continuous attributes is an important step in data mining study. The discrete method of granule calculation can be seen as split and merge ,Based on the interval merged guidelines given, under the premise of the minimized information loss, through adding to breakpoints to split adjacent intervals or removing breakpoints to merged adjacent intervals.

We will see the all average values of two adjacent and not same property values form sample set as a candidate breakpoints, two adjacent candidate breakpoints consisting of an interval granule. To define evaluation criteria of merged range, at the same time, merging to achieve the most appropriate neighboring interval of the evaluation criteria value. thus the condition attribute range into a number of the appropriate interval granule, then using the selected interval granule to discrete for conditional attributes.

For example, according to the value $x_1, x_2 \ldots x_n$ of conditions property named a, we can get the initial range granules that the amount is n when the average value is obtained by sorting from small to large of each attribute, The first interval granule referred to as

$$\left[x_1, \frac{x_1 + x_2}{2} \right]$$

The second interval granule referred to as

$$\left[\frac{x_1 + x_2}{2}, \frac{x_2 + x_3}{2} \right]$$

By parity of reasoning, we can get

$$\left[\frac{x_{n-1} + x_n}{2}, x_n \right]$$

When the initial interval granule is the same as the total number of samples, that is, all the attribute values are different, each interval contains a property value. But there are more of the same attribute values in actual data, in order to get less initial interval granule, we can use some algorithms to find all non redundant breakpoints, at the same time, the two adjacent break points are formed into a sequence of intervals, thus we can obtain a smaller number of candidates for a set of breakpoints.

After discretization of data attributes, there may be some part of the object attribute value is not unique or unknown, incomplete data processing is to be performed at this time. The definition of incomplete information system is given here:

Definition 2.1 In information system $S = (U, A, V, f)$, if there is at least one property $a \in A$ and an object $x \in U$ make the value is default, it is called incomplete information system.

In incomplete data processing, you can merge all attribute values with the same object and increase the number of a column for the same class of objects to make some small objects into a large object. Each object is a collection of small objects, and constitute an information granule, this process reduces the number of objects in the original decision system, the length of the decision table system is compressed, it accord with the thought of granular computing.

(2) Data Granulated

After incomplete data processing, according to the data of the nature is not clear, or similar set of functional aggregation obtained an object, at this time, the data size is suitable for particle size. The same granular layer can be divided into the same granular information, at the same time to initialize the granular structure. The free switching of granularity takes into account the decomposition and combination of multiple granularity and the rapid construction of the corresponding solutions. Construction of the corresponding granular layer and the structure of each particle layer on some specific problems are needed to consider multiple granularity layers of information, and through using the "cross granularity" mechanism to solve the problem. From the whole process, it can be found that the original data whether has the appropriate size to provide guidance for confirming whether need to adjust and how to adjust collection of the generation data. Selection of data sources and data integration is to confirm which data to solve the problem may be helpful, and which is not related to the topic.

We according to the way that is the same attribute values to granulate, there is a information system $S = (U, AT = C \cup \{d\})$, it is recorded as $GS = (U, GC \cup GD)$ after information granulated, for a decision table with value N, information system after granulated is recorded as $GS = (U, GC \cup GD, N)$.

(3) Rule Mining

In this model, mining rules are called classification association rules. The attributes of information system reduced a lot after attribute reduction. A series of decision rules are obtained by the AND operation of each information granule after reduction and decision attribute granule. We can get the rules of reduction after putting the decision rule granule whose support degree and certainty factor (CF) reach the requirements into the rule base. Users give the support degree and CF threshold of rules and dig out the rules that meet the requirements of support and CF, we call it the rules automatically. It can also define the decision requirements or conditions requirements according to user requirements, and dig out the simplest rules that satisfy these requirements, we call them custom decision mining and custom condition mining. Finally explain the mining rules with the original meaning of the attribute.

(4) Decision Application

The function we design in this module is very little, because decision application requires people's subjective regulation and decision. Therefore it requires users determine whether it can be used in data decision, rules that not be used for decision must be abandoned; only those rules with high value or revealing the internal rules can be used for decision. The rules after the explanation are derived from data mining or the results of the coincidence of data. As this module involves more subjective regulatory judgments and does not involve the specific content of granular computing, we will not elaborate.

4. Conclusion

Intelligent transportation is an important part of smart cities and the basic necessities of life, and society and it plays a crucial role for the development of city. Granular computing, as the research object of artificial intelligence machine learning, data mining and other popular areas, is rarely applied to the development and construction of Intelligent transportation. So on this basis, this paper proposed the intelligent transportation data mining model based on granular computing. This model preliminary shaping and explorations are given a detailed explanation, and the thought and methods of analysis and solving problem of granular computing are integrated into this model, the unique advantages that using it to reduce uncertainty and complexity problem can solve the uncertainty, fuzziness and complexity in intelligent transportation data mining.

Because this paper only makes a preliminary study on the model, so we don't make too much statement about some algorithms of data mining involved in this paper. We will continue to explore and research for future practical applications.

Acknowledgments

References

[1] Z.-Luo and W.-Jun, "Granular Computing Applied to Data-mining of Tunnel Information[C]", 2009 First International Workshop on Education Technology and Computer Science, **(2009)**, pp. 670-674.

[2] Y. Yao, "Apriori model of granular computing[C]", LNCS, Transactions on rough sets, **(2004)**.

[3] T. Y. Lin, "Data Mining and Machine Oriented Modeling: A Granular Computing Approach [J]", Journal of Applied Intelligence, vol. 13, no. 2, **(2000)**, pp. 113-124.

[4] Y. Y. Yao and J. T. Yao, "Induction of Classification Rules by Granular Computing Fuzzy Systems[C]", 2002, FUZZ, IEEE'02. In: Proc. ofthe 2002 IEEE Intl. Conf. on., no. 1, **(2002)**, pp. 773-778.

[5] Y.Y. Lin, "Granular Computing: Examples, Intuitions and Modeling[C]", In: the Proceedings of 2005 IEEE International Conference on Granular computing", Beijing China, **(2005)** July 25-27, pp. 40-44.

[6] Zhang L. and Zhang B., "A quotient space approximation model of multire solution signal analysis[J]", Journal of Computer Science and Technology, vol. 20, no. 1, **(2005)**, pp. 90-94.

[7] W.-Deng and G.-Y. Wang and Y.-Wu, "Summary of granular computing [J]", Computer Science, vol. 31, no. 10A, **(2004)**, pp. 178-181.

[8] W.-Deng, G.-Y. Wang and Y.-Wu. "An Improvement in Evolutionary Computation for Granular Computing[J]", Computer Science, vol. 31, no. 10, **(2004)**, pp. 209-211.

[9] T. Y. Lin, "Granular computing: fuzzy logic and rough sets[C]", Computing with Words in Information/Intelligent Systems, Physica2Verlag, **(1999)**.

[10] A. Skowrom, J. Stepaniuk and J. F. Peter, "Extracting Patterns Using Information Granules [J]", Bulletin of international rough set society, vol. 5, no. 1/2, **(2001)**, pp. 135-142.

Authors

Xiao-Lan Xie, She got her PhD in Xidian University, Shan Xi, China. She is a Professor in School of information science and engineering, Guilin University of Technology. Areas of research include Cloud computing, Grid computing and Intelligent decision system. She is a committee member and deputy secretary general of Cloud computing expert committee of China communication society. She is also a member of China computer society CCF and IEEE CS.

Xiao-Feng Gu, He is Mr. candidate. His research interests include Cloud computing, data mining, *etc*.

作者姓名	职称/学位	单位	地址	手机	Email
谢晓兰	教授	桂林理工大学	雁山校区	15277336998	237290696@qq.com
谷晓峰	硕士	桂林理工大学	雁山校区	18290060910	gxf199295@163.com