Multi-Source Information Decision Model for Internet of Things

Fang Nan, Rui Xiao, Jing Zhang, Yanwen Bao
Qinggong College, North China University of Science and Technology, Tangshan 063009, China

Abstract
In the multi-source information of the Internet of Things, the enterprise decision-making needs to be judged on the basis of the change of data over a period of time. On this basis, the generalized value fuzzy set is introduced into the multi-source information decision-making problem, and the related definitions and proof of nature of the generalized value heuristic decision based on the attribute dependency and mutual information from the geometric view have been put forward respectively, and the corresponding algorithm is provided, which has enriched and developed the generalized Fuzzy set theory. At the same time, provided the idea for the multi-source information analysis. In view of the multi-source information distributed storage architecture, the concept and proof of the Internet of Things generalized value global decision is provided, and the Internet of Things generalized value business decision model is further proposed. In order to achieve better results from the algorithm in the practical application, the concept of the approximate decision is introduced into the proposed three algorithms, to verify the validity of the proposed algorithm by experiments. The experimental results show that, all the three algorithms which have been proposed can greatly reduce both the number of objects and the number of attributes of the data sets under the condition of maintaining high classification accuracy, so as to provide support for the multi-source information decision-making.

Key words: Multi-source Information, Generalized Value, Approximate Decision, Internet of Things, Global Decision.

1. INTRODUCTION

With the development of the emerging information technologies such as cloud computing, Internet of Things and mobile Internet, human beings have been brought into the era of multi-source information, and the ubiquitous multi-source information has become the focus of attention from all walks of life(Lynch, 2008; Gong, Li, Wang and Cheng, 2011; Li and Cheng, 2012). Research shows that, the large-scale enterprise systems today consist of complete data center constituted by thousands of servers distributed in different locations (Rabl, Sadoghi and Jacobsen, 2012). How to quickly and accurately extract the potential value from the distributed, multi-source information, and turn the multi information into the source of economic value has increasingly become a powerful weapon for enterprises to surpass their competitors.

The multi-source information of distributed storage has presented many distinct characteristics: huge data volume, large data volume, fast flow speed and low value density, which put forward higher demands on the processing capability and efficiency of multi-source information. The analysis and processing of multi-source information is no longer keen on the pursuit of accuracy and to find the causal relationship (Naimi and Westreich, 2014). In the face of massive real-time data, appropriate ignoring of the micro-level on the level of accuracy can obtain better insight at the macro-level. Similarly, in the era of multi-source information, to seek the relationship between things without having to focus on the causal relationship between things can provide a very new and valuable point of view.

In many real multi-source information environments, there are a large number of uncertain factors, and the collected data often contain noise, imprecision or even incomplete data. Fuzzy set theory (Qian, Liang, Yao and Dang, 2010) is another powerful mathematical tool to deal with uncertainty after the probability theory and evidence theory.

As a soft computing method, its validity has been confirmed in various application fields, which is the research hotspot in the field of artificial intelligence theory and its applications (Suyun, Tsang and Degang. 2009). Fuzzy set theory has many similar characteristics as the probability theory and the evidence theory, but compared to the latter three, the fuzzy set does not require any priori knowledge, but only through the data itself to obtain knowledge, while the probability theory and evidence theory requires the probability, membership, probability assignment and other information respectively.

One of the core problems in fuzzy set research is attribution decision. Through attribute decision, the minimum expression of decision table can be obtained. That is to say, to delete the unrelated or unimportant attribute in the knowledge expression system; and it is also the key to knowledge acquisition. But it has been
proved that all the decision-making and solving the minimum decision is NP-hard problem. At present, most of the proposed attribute decision-making algorithms are based on heuristic and are for the centralized single decision table (namely, a complete decision table), and not applicable to the multi-source information analysis and mining of distributed storage. Currently, some scholars have researched and implemented the attribute decision-making algorithm of fuzzy set under the distributed platform. (Huang, Hu and Zhou, 2009) However, these algorithms are only the decision-making algorithm itself in the implementation of distributed platforms, and still dealing with centralized single-decision table, not considering the data set of distributed storage. Distributed storage of multi-source information environment decision-making algorithm research can greatly reduce the workload of multi-source information analysis. With distributed tagged multi-source information, the multi-source information can be classified into several categories, each site can be considered as a decision table, the overall multi-source information can be considered to be composed of more than one decision table, and the conditions of these decision table attributes are different, but the decision attributes are the same. Therefore, the distributed storage of multi-source information for decision-making algorithm research can be transformed into the study of the decision-making method of Internet of Things. Literature(Zhang, Wang and Huang, 2009) conducted related research on the distributed Internet of Things approximation decision-making. Literature (Xu, Miao and Wei, 2009) considered some application scenarios based on the above content, and each site hopes to keep its own original data and sensitive information of the local decision table and not be obtained by other sites, the privacy protection strategy is added, and the privacy decision-making algorithm of the Internet of Things is designed. Therefore, it can be seen that the research on the Internet of Things (Distributed storage of multi-source information) is inseparable from the specific applications.

With the advancement of the construction of the smart grid, the multi-source information pattern of the Internet of Things has been gradually formed. At present, the main forms of obtaining the multi-source information from the power are scattered in different system databases, and the data types are continuous value attributes different from the traditional classification methods, the study of classification of multi-source information is no longer focused on a single entry of data, but adopts the form of data blocks as the research object. This is because the reliance on only one entry of data to determine its classification information has been insignificant, and it is necessary to consider the characteristics of data within a certain period of time to judge the class of the data segment. For example, based on the multi-source information of the Internet to predict the load, a single data does not have the characteristics of load forecasting. It is necessary to compare the data segment to be predicted with the data of a certain time period to determine the load forecast value. Therefore, the classification of multi-source information should start from the data block. In order to quickly and effectively establish a classification model for multi-source information of the Internet of Things, the numerical conditional attribute data block is approximated as the generalized value form, namely, by the maximum and minimum data blocks of the approximate description of the data block (for non-numerical condition attributes can be converted into numerical processing) so as to study the generalized value attribute decision-making strategy, and establish the classification model. Some scholars have studied the generalized conditional attribute decision-making method (Hu, Zhao and Yu, 2008), but these methods are for a centralized data set, and do not take into account the situation of the Internet of Things, therefore, they are not applicable to the distributed storage of multi-source information environment.

In this paper, the multi-source information of distributed storage is regarded as being composed of multiple decision-making tables with the same decision attributes and different condition attributes. On this basis, the multi-source information is partitioned into intervals to study the Internet of Things business decision making fusion method. The significance of this paper is as the following:

1) In view of the characteristics of large data volume and large amount of noise in multi-source information, the fuzzy set method is introduced into multi-source information analysis, and the amount of data involved in multi-source information analysis is reduced by attribute decision making method;

2) As the multi-source information of the Internet of Things mainly has the continuous-valued attribute, and the research on classification of multi-source information should be based on the data block as the object unit, the approximation of the data block is described as the generalized value form. Decision-making table based on the generalized value attribute decision-making. The related concepts and attribute proofs of the generalized value attribute decision-making are given, and the corresponding algorithm is proposed; in order to enhance the practicability of the algorithm, an approximate decision-making concept and method of the generalized value decision table are proposed.

3) In view of the distributed storage of multi-source information, the conception and attributes of the global approximation decision-making under different physical attributes and the same decision attributes are given, and the corresponding decision algorithm is proposed, so that the global decision making with similar classification result is obtained for the multi-source information of distributed storage.

4) The three designed algorithms are tested in the multi-source information real data set, and the results are analyzed and discussed. The experimental results show that: All the three algorithms can select a subset of attributes that maintains the relatively high classification accuracy in the appropriate interval length; with the
increase of the number of attributes, the run time of the generalized decision-making method based on the
dependence degree is slightly longer than the generalized value decision method based on the mutual
information, and the run time of the global decision-making under the Internet of Things is the shortest.

2. GLOBAL APPROXIMATE DECISION OF THE GENERALIZED VALUE λ UNDER THE
INTERNET OF THINGS

The multi-source information is distributed and stored in different locations. Therefore, we discuss the
global decision-making method of generalized value λ under the Internet of Things. This section discusses the
global decision method of the generalized value λ – of the Internet of Things from the perspective of the
information, and the decision method from the perspective of geometry is similar.

2.1. Related Concepts and Attributes of the Global Decision of the Generalized Value λ under the Internet of
Things

In the distributed environment, the cost of network communication is the key to affect the decision-
making efficiency of the Internet of Things attributes. Therefore, reducing the network traffic effectively is the
key task of solving the global decision-making of the Internet of Things in the distributed environment. Although
the local decision table of each site can be sent to a central site can be simple to achieve attribute decision-
solution, but the network traffic is huge in this practice, particularly in the face of multi-source
information environment (large scale and contain a high number of dimensions) local decision table (a single
site), large volume of data needs to be transmitted.

The conditional information entropy \( H(D|R^i_s) \) of the generalization value \( λ \) is only related to
the compatible class \( R^i_s \) and \( R^i_s \), so it can effectively avoid the transmission of all the local decision tables by
adopting the effective compatible class storage mechanism of \( λ \) – and transmitting only the corresponding
\( λ \) – compatible class strategy. Therefore, for the \( λ \) – compatible class

For \( R^i_s(u_i) \) (1 ≤ i ≤ n) in \( R^i_s \), the following triple is adopted for the storage:

(Site ID, the ID of the generalized value object in \( [R^i_s(u_i)] \). \( R^i_s(u_i) \) incremental sequence).

In which, \( \| \) stands for the number of the generalized value objects.

For the \( λ \) – compatible class \( R^i_s(A \subseteq C_s) \), \( R^i_s(B \subseteq C_s) \) on the different site \( S_s \) and \( S_s' \), the aforementioned
storage methods can be applied to obtain the following lemma:

Lemma 3.1. For \( λ \) – compatible classes \( R^i_s(A \subseteq C_s) \), \( R^i_s(B \subseteq C_s) \) on different site \( S_s \) and \( S_s' \), the
following can be obtained:

\[
R^i_{s,s'} = \{ R^i_s(u_i) \cap R^i_s(u_j) : R^i_s(u_i) \cap R^i_s(u_j) \neq \emptyset, 1 \leq i \leq n, 1 \leq j \leq n \}.
\]

It can be seen that: Through the adoption of the \( λ \) – compatible class transmission method, it can be
obtained that the maximum network traffic of \( R^i_{s,s'} \) is \( n + \max \{ |R^i_s|, |R^i_s'| \} \) (n is the number of objects in a local
decision table); while through the adoption of the transmitting sub-local decision table, the corresponding
network traffic is at least \( \min( |A|, |B| ) \times n \). In the multi-source information environment, the number of selected
attribute subsets is much larger than 1, namely \( \min( |A|, |B| ) > 1 \). Further, by using the following Lemma 3.2 and
Theorem 3.1, through only partially transmission of the \( λ \) – compatible classes of \( R^i_s(A \subseteq C_s) \) or \( R^i_s(B \subseteq C_s) \),
\( R^i_{s,s'} \) can be solved.

Lemma 3.2. For the \( λ \) – compatible class \( R^i_s = \{ X_1, X_2, ..., X_s \} (A \subseteq C_s) \) on different
site \( S_s \) and \( S_s' \), \( R^i_s = \{ Y_1, Y_2, ..., Y_s \} (B \subseteq C_s) \), \( U / D = \{ \psi_1, \psi_2, ..., \psi_l \} \), if \( X_w \subseteq \psi_k \) (1 ≤ w ≤ s, 1 ≤ k ≤ l) ,
then \( X_u \cap Y_j \subseteq \psi_k \).

Theorem 3.1. For the \( λ \) – compatible class \( R^i_s = \{ X_1, X_2, ..., X_s \} (A \subseteq C_s) \) on different
site \( S_s \) and \( S_s' \), \( R^i_s = \{ Y_1, Y_2, ..., Y_s \} (B \subseteq C_s) \), \( U / D = \{ \psi_1, \psi_2, ..., \psi_l \} \), if \( X_w \subseteq \psi_k \) (1 ≤ w ≤ s, 1 ≤ k ≤ l) ,
then \( \forall Y_j \in R^i_s, X_u \cap Y_j \neq \emptyset \), the following can be obtained:
\[ p(X_i \cap Y_j) \sum_{i=1}^{d} p(X_i \cap Y_j \cap \psi_i) \log_2 p(X_i \cap Y_j \cap \psi_i) = 0. \]

In which, \( d \) is the number of the \( \lambda \)-compatible classes that have intersection with \( X_i \), and not null.

Proof: As \( X_i \cap Y \neq \emptyset \), it can be known from Lemma 3.2 that \( X_i \cap Y \subseteq \psi_i \), then \( p(X_i \cap Y \cap \psi_i) = 1 \) is established, therefore, Theorem 3.1 is established.

Theorem 3.2. For the \( \lambda \)-compatible class \( R^i = \{X_1, X_2, \ldots, X_s\} \{A \subseteq C_r\} \) on different site \( S_\epsilon \) and \( S_{\lambda} \), \( R^i = \{Y_1, Y_2, \ldots, Y_s\} \{B \subseteq C_r\} \), \( U/D = \{\psi_1, \psi_2, \ldots, \psi_t\} \), if assume \( Y_i (1 \leq i \leq d) \subseteq \psi_{\lambda(v)} \{1 \leq y(v) \leq l\} \). \( X_i \{1 \leq w \leq q\} \subseteq \psi_{\lambda(w)} \{1 \leq x(w) \leq l\} \) (d has the meaning of the above), then:

\[ H(D|R^i_{\lambda/\lambda}) = \sum_{i=1}^{d} \sum_{j=1}^{s} p(X_i \cap Y_j) \sum_{k=1}^{l} p(X_i \cap Y_j \cap \psi_k) \log_2 p(X_i \cap Y_j \cap \psi_k). \]

Theorem 3.2 can be proven by Theorem 3.1 and Definition 2.9.

As can be known from Theorem 3.2: For \( R^i_{\lambda} = \{X_1, X_2, \ldots, X_s\} \{A \subseteq C_r\} \), \( R^i_{\lambda} = \{Y_1, Y_2, \ldots, Y_s\} \{B \subseteq C_r\} \), \( U/D = \{\psi_1, \psi_2, \ldots, \psi_t\} \), \( X_i \{1 \leq w \leq q\} \subseteq \psi_{\lambda(w)} \{1 \leq x(w) \leq l\} \). To solve \( H(D|R^i_{\lambda/\lambda}) \), it is only required to transmit the \( \lambda \)-compatibility classes from site \( g \) to \( e \), therefore, the required maximum network traffic is \( \sum_{i=1}^{d} |X_i| + s - q \).

While \( s - q < |R^i_{\lambda}| \), \( \sum_{i=1}^{d} |X_i| < \sum_{i=1}^{d} |X_i| \), \( R^i_{\lambda} = \{Y_1, Y_2, \ldots, Y_s\} \{B \subseteq C_r\} \), \( U/D = \{\psi_1, \psi_2, \ldots, \psi_t\} \), \( X_i \{1 \leq w \leq q\} \subseteq \psi_{\lambda(w)} \{1 \leq x(w) \leq l\} \). And usually under the multi-source information environment, \( s - q \) can be negligible relative to \( \sum_{i=1}^{d} |X_i| \). For the convenience, record \( Z(R^i) = \{X_i \in R^i_{\lambda} : X_i \cap \psi_k \neq \emptyset \wedge X_i \cap \psi_r \neq \emptyset, 1 \leq k \neq r \leq l\} \) to represent the compatible classes that are not contained in a decision class; record the amount of network traffic required to transfer \( \lambda \)-compatible class \( R^i_{\lambda} \) from one site to another as \( NZ(R^i_{\lambda}) \).

It can be known from the decision-making algorithm based on mutual information that: During the running of the algorithm, the network transmission cost will decrease rapidly with the expansion of important attributes.

2.2. Global Approximation Decision Algorithm for the Generalized Value \( \lambda \) under the Internet of Things

According to the generalized value attribute approximation method based on mutual information and the compatible class transfer strategy in Section 2.4, the global approximation algorithm of the generalized value under the Internet of Things can be designed, please refer to algorithm 3.

Algorithm 3. \( \lambda \)-global approximate reduction in interval-valued multi-decision tables (referred to as GARIv for short).

Input: \( DT = \langle U, C \cup D, V, f, \lambda \rangle \).

Output: \( \lambda \)-Global Approximate Decision \( red \).

Step1. Let \( red = \emptyset \).

Step2. Conduct parallel computing \( \lambda \)-compatible class on all sites \( R^i_{\lambda_{\alpha}} \{a \in C_r\} \); Conduct parallel computing at each site for \( Z(R^i_{\lambda_{\alpha}}) \{a \in C_r\} \), find each site \( S_i \), that makes \( H(D|R^i_{\lambda_{\alpha}}) \) achieve the minimum attribute \( a_i \); and obtain in site \( S_j \) the attribute \( a_j \) that enables \( H(D|R^i_{\lambda_{\alpha}}) \leq H(D|R^i_{\lambda_{\beta}}) \{a \in C_r\} \) (namely, \( a_j \) is one of the \( a_i \) selected by each site with the minimum conditional entropy), \( red = red \cup \{a_j\}, H(D|R^i_{\lambda_{\beta}}) = H(D|R^i_{\lambda_{\alpha}}) \);
Step3. If \((C_i - \text{red}) \neq \emptyset, i \neq j\), transmit the \(Z(R_{\text{int}}^i)\) obtained in site \(S_j\) to each site \(S_j\), conduct parallel computing in each site for \(Z(R_{\text{int}(a_j)}^i)\) \((a_i \in (C_i - \text{red}))\) to find each site \(S_j\), so that \(H(D|R_{\text{int}(a_j)}^i)\) achieves the minimum attribute \(a_j\) (namely, \(a_j\) is one of the \(a_i\) selected by each site with the minimum conditional entropy), recorded as:

\[
SIG(a_j, \text{red}, D) = H(D|R_{\text{int}}^i) - H(D|R_{\text{int}(a_j)}^i);
\]

Step4. If \(SIG(a_j, \text{red}, D) > \varepsilon\) and red \(\neq \bigcup_{i=1}^m C_i\) (m is the number of sites), red = red \(\bigcup\{a_j\}\), go to Step3; otherwise, output the \(\lambda\) – global approximation decision \text{red}.

In the practical applications, the \(\lambda\) – compatible classes with relatively small volume can be transmitted to the site with a relatively large volume of \(\lambda\) – compatible classes to further optimize the algorithm.

3. EXPERIMENT AND ANALYSIS

However, most of the existing methods of business decision-making do not pay attention to the characteristics of the Internet of Things data, in the absence of steady-state data to determine the case of direct business decisions. As the division of work condition is not clear enough, it leads to the result that the comparability of the mining results with the actual operation data is not very good. In this experiment, according to the production characteristics of the power plant, the steady state decision of the production data is carried out, and the classification model is established, and the validity of the proposed algorithm is evaluated by the accuracy of the classification result and the time of the classification model.

3.1. Experimental Data

In this experiment, the multi-source data of a power company is selected, with the data of the first half of 2012 as the experiment objects, and a total of 107,184 records have been generated. The integrated data contains 427 attributes, excluding the system automatically generates the keyword ID number and data retention time, and a total of 425 conditional attributes (all in numerical type) are obtained. The steady-state and unsteady-state annotations of the steady-state condition parameters are obtained from the steady-state condition parameters, and a large scale decision table is formed. In order to test the effectiveness of algorithm 3 (GARIv), the original system data is not processed, and the same decision attributes are added to each database data.

In order to evaluate the performance of the algorithm, we design a variety of partitioning methods for the data interval, such as every 10 minutes, 20 minutes, ..., 90 minutes as an interval. If in the partition process, a range corresponds to different decision class, the data of the same decision class is classified into a small interval, and the next interval starts from the different decision classes.

All experiments are run on an Intel Xeon (R) Processor (Four Core, 2.5GHz, 16GRAM) workstation and are programmed in JAVA. In order to test the GARIv algorithm, two virtual machines are set up on the workstation as two sites. To ensure the fairness of the comparison of the experiments, tenfold cross is adopted to validate the accuracy of the estimation classification.

3.2. Experimental Comparison

Firstly, conduct experiment on algorithm 1 to algorithm 3 in the selected data set, record the length of each algorithm in different lengths of time to select the number of different attributes of time. Figure 1 (a) to Figure 1 (d) shows that when \(\lambda = 0.7\), under the interval length of 10 minutes, 20 minutes, ..., 90 minutes, the run time diagram for the selection of select different attribute number. Figure 2 (a) to Figure 2 (c) shows when \(\lambda = 0.5, 0.7, 0.9\), under different interval length, the run time required to select 3 attributes.
It can be seen from Figure 1 that:

With the increase of the interval length, the data objects are reduced in multiples and the run time of the three algorithms is also greatly reduced.

However, as the length of the interval increases, the degree of coincidence between the intervals will increase, it is prone to cause the increase in the number of $\lambda$-compatible class elements, when an attribute is added, the compatibility class cross computing amount increases. Therefore, the run time of the algorithm does not show the linear change. Even at certain times, especially when the interval length becomes longer, the run time is not reduced with the reduction of the objects, but increased. This may be due to the fact that although the length of the interval increases, the runtime still increases.

When the number of attributes is small, the run time of RIvD algorithm and RIvMI algorithm is relatively similar; but with the increase in the number of attributes, RIvD algorithm run time more than RIvMI. This may be because with the increase in the number of attributes, the number of compatible classes is increased by the cross operation, while in the calculation of the positive domain, it is required to re-judge whether each compatible class belongs to the positive domain, which will cause the increase in the computing time of RIvMI algorithm. For RIvMI algorithm, although the addition of an attribute, the number of compatibility class will increase accordingly, however, when the new conditional entropy is calculated, the new conditional entropy can be calculated by the original condition entropy. Therefore, as the number of attributes increases, the RIvMI algorithm generally takes less time than the RIvD algorithm.
It can also be seen from Figure 1 that, the running time of GARIv algorithm is the shortest. In theory, the run time required by GARIv algorithm should be 1/2 of Rlvl algorithm (a total of 2 sites in parallel computing), but as in the calculation of compatibility classes, it involves the transmission and cross computing of partial compatibility classes, and the virtual machine configuration is lower than the entire workstation, therefore, the GARIv algorithm runs faster than the theoretical time. But generally speaking, GARIv algorithm runs slower than the RlvlMI. With the increase of the number of sites, the actual run time of GARIv algorithm shall be significantly lower than RlvlMI.

It can be seen from Figure 2 that, the run time of the three algorithms affected by the value of $\lambda$ does not have regularity. The more refined the compatible class is, the less number of compatible class elements are. For GARIv algorithm, the required transmission compatible classes are more, therefore, under the same conditions, the run time is slightly longer.

Compare the average classification rate on the attribute subsets selected by the three algorithms when $\varepsilon = 0.01$, the results are shown in Figure 3. Figure 3 (a) is the algorithm average classification accuracy rate when $\lambda = 0.5$; Figure 3 (b) is the algorithm average classification accuracy rate when $\lambda = 0.7$; Figure 3 (b) is the algorithm average classification accuracy rate when $\lambda = 0.9$. As can be seen from the figures, the average classification accuracy rate of the three algorithms basically can achieve above 80% in the appropriate interval length. When $\lambda = 0.7$, the interval average is 20 minutes, the average classification accuracy rate of all the algorithms is the highest. As RlvlD algorithm and RlvlMI algorithm has different measurement methods on the significance, resulting in the different selection of the subset of attributes. As can be seen from Figure 3, RlvlMI algorithm selected subset of attributes has higher classification accuracy than that of RlvlD. This may be due to inconsistency of the data after the data is partitioned into intervals, while the method based on dependency (positive domain) is not suitable to deal with the inconsistency problem, therefore, the classification accuracy is slightly lower than RlvlMI. GARIv algorithm is equivalent to the decision algorithm with the conduction of RlvlMI algorithm in the vertical partitioned data set in parallel, but the attribute subsets selected by the two algorithms are not the same. This is because as RlvlMI algorithm selects the attributes, if the significance of some attributes is the same, it will select the most left attribute to the decision set. While GARIv algorithm processed dataset can be regarded as the vertical partition of the RlvlMI algorithm dataset, so the attributes are sorted differently, and the attributes subsets selected by the two algorithms are not the same.

Through the comparisons of the selected subset of attributes in the aforementioned experiments, it can be found that most of them have the five attributes including generator power, main steam pressure (machine side), main steam temperature (machine side), reheat temperature (machine side) and #1 high temperature of the inlet steam, which is consistent with the actual assessment of the indicators involved in steady state.

In order to further investigate the effectiveness of the three algorithms proposed in this paper, the average classification accuracy of the above experiments is compared with the average classification accuracy rate of the traditional classifiers KNN ($k$ nearest neighbor), RBF (Radial basis function neural network) and REPTree (Error reduction pruning tree), and the results are shown in Table 1.
Table 1. Average Accuracies of Classification

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Interval Length (Unit: Minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>RIvD</td>
<td></td>
</tr>
<tr>
<td>λ = 0.5</td>
<td>79.2</td>
</tr>
<tr>
<td>λ = 0.7</td>
<td>79.2</td>
</tr>
<tr>
<td>λ = 0.9</td>
<td>80.1</td>
</tr>
<tr>
<td>RIvMI</td>
<td></td>
</tr>
<tr>
<td>λ = 0.5</td>
<td>84.8</td>
</tr>
<tr>
<td>λ = 0.7</td>
<td>83.6</td>
</tr>
<tr>
<td>λ = 0.9</td>
<td>84.3</td>
</tr>
<tr>
<td>GARIV</td>
<td></td>
</tr>
<tr>
<td>λ = 0.5</td>
<td>78.5</td>
</tr>
<tr>
<td>λ = 0.7</td>
<td>80.7</td>
</tr>
<tr>
<td>λ = 0.9</td>
<td>79.8</td>
</tr>
<tr>
<td>1NN</td>
<td></td>
</tr>
<tr>
<td>2NN</td>
<td></td>
</tr>
<tr>
<td>3NN</td>
<td></td>
</tr>
<tr>
<td>5NN</td>
<td></td>
</tr>
<tr>
<td>10NN</td>
<td></td>
</tr>
<tr>
<td>RBF</td>
<td></td>
</tr>
<tr>
<td>REPtree</td>
<td></td>
</tr>
</tbody>
</table>

As can be seen Table 1: KNN has the best classification effect in the traditional classification method, and the accuracy rate of RBF neural network and REPTree decision tree classification method is low, as the multi-source information of IOT. It is difficult to judge whether it is in steady state or unsteady state for a certain data. KNN classification effect is better, mainly due to the fact that the value of decision-making class is segmented, and the decision class of the same data class is same, so in the calculation of the k nearest neighbor, the k points that are closes from the test data corresponding to the decision-making class may be the same as the adjacent test data; and as the value increases, it is prone to have cross-class phenomenon, resulting in the decrease in the average classification accuracy. The three algorithms proposed in this paper are all based on interval and more consistent with the conditions of the stability of judgments, so the average classification accuracy is higher than the traditional methods. It can also be seen that, the three algorithms proposed in this paper are effective for multi-source information decision-making. The value λ is determined according to the different length of the interval selected by different applications, and the value ε is determined according to the required number of attributes. For multi-source information of distributed storage, GARIV algorithm can be adopted directly to process the data to obtain the global approximation decision.

4. CONCLUSION AND OUTLOOK

In this paper, the generalized value decision algorithm based on the dependency and mutual information is put forward in view of the multi-source information decision problems in the Internet of Things, and the concept and method of business decision making fusion under information theory is proposed, which has achieved good results in the application of the multi-source information judgment in the Internet of Things. As in most of the Internet of Things multi-source information analysis applications, it is required to consider change of a data segment rather than a particular entry of data, the interval data setting not only can greatly reduce the data volume of the multi-source information, reduce the difficulty of multi-source information analysis, but also conforms to the specific application of multi-source information of the Internet of Things. While the attribute decision for the data set, under the condition of not affecting the whole dataset classification, has also reduced the multi-source information from the dimension, lowered the data volume of the multi-source information and reduced the difficulty of the multi-source information analysis. It can be seen from the experimental results that,
all of the three algorithms are effective, which have provided not only new idea for the generalized value decision method, but also solutions for the multi-source information decision-making problem.

REFERENCES


Naimi, A. I., Westreich, D. J. (2014) “Big data: a revolution that will transform how we live, work, and think”, American Journal of Epidemiology, 17(9), 181-183.


