

The Transmission of the Combination of Rough sets and Fuzzy Kohonen Clustering Network Technology Based on GPS Data under Large Data Environment

Kun Yang, Ruifen Yu

Jiangxi University of Technology, Nanchang 330098, China

Keywords: Rough set; Neural network; Data fusion; Data fusion transmission

Abstract. the difficulties of wireless sensor data fusion is to switch the sensor cluster state, namely the problem at some point the direction to which the sensor data fusion. First of all, by the acquisition of knowledge in rough set theory, the typical clustering distribution of multiple sensors can be fused as sample space form "decision table data fusion distribution", then according to the method of knowledge reduction in rough sets, attributes and sample data redundancy to delete. And then the clustering analysis and fuzzy Kohonen clustering network, and finally the formation of distribution rules of sensor data fusion. The experimental results show that the model has good classification efficiency, can realizes the clustering distribution of judgment.

Introduction

Wireless sensor network is integration of sensor technology, computer technology and communication technology, the utility and battlefield communications, disaster relief and public gatherings of sudden, temporary occasions. Because the node storage and power constraints, so the power level is one of the main signs to measure the performance of wireless sensor network.

In wireless sensor network, with the widely application of multi-sensor systems, then appeared the processing and use of the multi-sensor information technology emerge as the times require, multi-sensor data fusion, information fusion as a way to reduce network communication data amount, effective means to prolong the survival time of wireless sensor networks, has drawn more and more attention. In the larger scale wireless sensor networks, such as industrial automation and detection, environmental monitoring, transmission of large amount of data, high redundancy, will lead to data transmission collision, reliability problems of power consumption, data transmission, radio frequency transceiver one bit of data consumption energy is hundreds of times to execute an instruction consumes energy the use of data fusion, can significantly reduce the number of RF transceiver. Therefore the node transmission data processing is very important, the use of computing and processing ability of nodes, we can remove the redundant information, improving the performance of the entire network.

Including pattern recognition, decision theory, uncertainty theory, signal processing, estimation theory, optimization technology, artificial intelligence and neural network is the main method in data fusion. While the neural network is a far more common data fusion method. However, based on the existing research shows that the multi sensor information fusion system based on neural network, a

common drawback of these systems is: when the sensor of large quantity of data, the network structure is huge, the neural network training time is longer, the network training burden reduced, leading to performance. Because of the rough set in the treatment of redundant data, a large amount of data and uncertain data are advantages, more and more scholars noticed that can take the rough set and neural network are combined, pre processing of the data by rough set, a simplified neural network training samples, the training speed.

In this paper, the experimental environment is a sensor network consisting of a plurality of sensors, the complexity of the network, sensor data volume than previous experiments to high. The kononen neural network and rough set link. The actual network protocol to modify the Tiny OS stack, the theory of instantiation of Kohonen neural network and rough set, the addition of OS data fusion layer can be directly transplanted to the actual system.

Rough set theory

Rough set theory is a proposed by Poland scholar Z.Pawlakin twentieth Century 80 for theoretical analysis of data. The theory can analyze and process the imprecise, inconsistent and incomplete information, rough set, as a great potential and effective knowledge acquisition tool and the attention of artificial intelligence researchers.

In rough set theory, an information system is a data table. Attribute corresponding object on this list, the corresponding object for research, the whole table are all attributes of the object of values.

The form, an information system as $S = (U, A)$, U and A be a nonempty finite sets, where U is the set of objects is called a domain, A is a set of properties; properties of the elements in the set are defined, such as $a:U, \forall a$. $\forall a$ is a property of arrange; a two yuan between $I(B)$ in a subset of any established on the $U(B)$, this is referred to as the definition of the indiscernibility relation are as follows: (x, Y) in $I(B)$ if and only if $a(x) = a(Y)$ in A , all the A , $a(x)$ value for each x element in the a . Obviously, $I(B)$ is an equivalence relation, an attribute corresponds to an equivalence relation, a table can be viewed as define a family of equivalence relations, namely, knowledge base. A family of equivalence class $I(B)$, is a B classification, will be denoted by $U/I(B)$ or simply expressed as U/B ; an equivalence class $I(B)$, is also the block segmentation U/B containing x will be expressed as $B(x)$ called the B block contains the X . If (x, y) belongs to $I(B)$, X and y are indistinguishable in B . Association of $I(B)$ of the equivalence class is represented group B basic or a subset of the B .

If the attribute in the information system is further divided into condition attributes and decision attributes, the information system is called decision table. The definition of $S = (U, C, D)$, C is known as the condition attribute set, D called the decision attribute set, and $C \cup D = A$. $C(x)$ and $D(x)$ were expressed as X contains conditions block and decision group.

For classification, many of the attributes are redundant. The definition of reduct is free of redundant attributes and ensure correct classification minimum attribute set. Namely the equivalence relation between B belongs to A , when there are attributes of B in B , the $U/B = U/(B - b)$, B can be omitted, otherwise B is omitted in the B . If B is in R can be omitted, so that the $B = B - R$. When the attribute B not be omitted when B is called independent, so that it can obtain a reduction of B , denoted as $red(B)$.

Design of algorithm of data fusion

The wireless network may use the cluster based routing mechanisms, such as the LEACH algorithm is divided into different clusters, the whole network is divided into three types of nodes, base station, the cluster head nodes and the nodes in cluster. This paper mainly according to the cluster configuration, the design of two layer data fusion mechanism, the first layer of nodes and the cluster head nodes fusion, intra cluster nodes according to the threshold to determine whether you need to send data, the cluster head nodes based on the received data, data consistency test, eliminating abnormal data, second layer using the fusion BP algorithm for cluster head nodes and a base station alarm, get the results required.

In wireless sensor networks, the nodes in the cluster as a basic sensor node, used for data collection of A nodes within a cluster, assuming current collected data value is D_i , then the last data for D_{i-1} , the cluster nodes threshold is P , then the nodes in the cluster can be set according to the threshold to decide whether to send the current value of the data, i.e., when $|D_i - D_{i-1}| > P$, a node sends data in a cluster, the cluster nodes to change the memory value, replace the numerical last collected data, otherwise not transmitting, memory value remains unchanged last. In this way, the data has not changed greatly, the amount of change depends on the threshold of P size, the nodes in the cluster can reduce the number of data is sent more, can prevent to the cluster head node sends redundant and highly similar data. Save cluster member node energy consumption.

Assume that the cluster head nodes with 10 nodes in clusters, and each cluster nodes have different numbers, are numbered from 1 to 10, the cluster head nodes is not received by the nodes in the cluster data, is stored on a cluster node collected data, when the cluster head node connection received the nodes in cluster data, replace the corresponding node data value, data storage, the collected otherwise, pertaining to data nodes of the cluster inner value and great changes have taken place, directly calling the corresponding cluster node last collected data value represents the data. In this way the cluster head nodes can know the corresponding 10 intra cluster nodes of the current collected by the temperature data values, for example, the 10 data collected in accordance with the hypothesis of value from 4, numerical N large arrangement after $T_1, T_2, T_3, \dots, T_{10}$, too, because the node may receive external dry interference, and its hardware failure effect of, the cluster head nodes first to check the consistency of data is used in the detection technique of data distribution map method to weed out the mistake error data, assuming that the median is indicated by T_M , four quintiles are indicated by F_u and four quintile with F_L said and the four quartile dispersion expressed using dF . According to the above collected data can be obtained, $T_M = (T_5 + T_6) / 2$, range four quantile F_L for $[T_1, T_M]$, interval four quantile F_U for $[T_M, T_{10}]$, the four quantile dispersion with $dF = F_U - F_L$. We have identified and median distance is greater than the adF data for divorce data, should be removed. Void data discrimination range of $[T_i - T_M] > adF$. A is a constant of type, can set the size according to the actual need. In this way, the cluster head nodes can eliminate the abnormal data out of simple and effective.

Conclusions

Rough set and two layers of data fusion algorithm effectively combines, in the larger advantage in multi sensor data fusion. In the first layer of nodes in the cluster and the cluster head node data processing, reliable and effective to reduce the number of nodes in the

cluster the data forwarding. The second layer uses BP algorithm for data fusion, can get the user needed results directly, without the need to transfer large amounts of original data collection. Reduce the amount of information in a wireless network to network advantage is obviously, and can reduce the number of node information retransmission, save energy limited networks, improve information accuracy, prolong the network survival time. But in practical application, if the sensor OS to add protocol layer can greatly reduce the amount of communication data sensor. Data fusion, sensor fusion problem in single direction was solved. Although there are network fusion structure of the sample is difficult to summarize, sensor data processing requirements are higher shortcoming, but with the sensor hardware technology continues to improve, I believe that in the near future, these problems will be solved satisfactorily.

Acknowledgements

This work was supported by Nature of Jiangxi University of Technology [No. ZR13ZD03], and Jiangxi science and technology project [No. 20143BBM26128] and also give us help.

References

- [1] FAN S L, CHANG L, GUO B J. Common model control-CMC base in fuzzy neural network-FNN. *Automation and Instrumentation*. 2009
- [2] LIU J SH, JIN J, CHEN H X. An algorithm of multi-sensor information fusion based on the combination of rough set and neural network. *Computer Applications and Software*. 2009
- [3] LI X Q, ZHUANG D H, ZHANG Q. A new fault diagnosis model of electric power grid based on rough radical basis function neural networks. 2009
- [4] Ostroff S, *Temporal Logic for Real-Time Systems*, Research Studies Press LTD. John Wiley & Sons Inc, 1989, 22~28.
- [5] Berndt J D, Clifford J, *Finding Patterns in Time Series: A Dynamic Programming Approach*, *Advances in Knowledge Discovery and Data Mining*, MIT Press, 1996, 229~248.
- [6] Bazan Jan G, Skowron A, Synak Piotr, *Market Data Analysis: A Rough Set Approach*, Technical Report: 6 / 94 University of Warsaw, 1994.
- [7] Golan R, Edwards D. , *Temporal Rules Discovery using Datalogic / R+ with Stock Market Data*, *Rough Sets, Fuzzy Sets and Knowledge Discovery*, Workshops In Computing Series, Springer-Verlag, 1993, 74~81.
- [8] Golan R, Ziarko W. , *A Methodology for Stock Market Analysis Using Rough Set Theory*, *Proceedings of IEEE / IAFE Conference on Computational Intelligence for Financial Engineering*, New York City, 1995, 32~40.
- [9] Anders T B, *Mining Time Series Using Rough Sets—A Case Study*, *Proceeding of first European Symposium, PKDD' 97*, Norway: Trondheim, 1997, 351~358.