Advanced Science and Technology Letters
Vol.134 (CIA 2016), pp.59-64
http://dx.doi.org/10.14257/astl.2016.134.10

Research on Big Data of Smart Home based on Complex Network Theory

Zhao Yan

Beijing Polytechnic, Beijing 100176, China
tozhaoyan@sina.com

Abstract. In the complex network, especially with the typical smart home network, the data is very big, and according to a large number of home smart home based on the complex network analysis of big data behind the information is very valuable, however, it is difficult to analyze the data is big and the meaning of the data represented is abstract. According to the characteristics of complex network, this paper uses a multi-level complex network data analysis based on the inherent properties of complex, which can effectively reduce the complexity of the analysis, and can be easily visualized management, which lays a theoretical foundation for big data cloud computing.

Keywords: Smart Home, Cloud Computing, Complex Network, Big Data, Data Mining

1 Introduction

With the advance of the concept of Internet of things, on the basis of the framework of the Internet, Internet of things give the functions of communication and computing to the objects, so the Internet of things is the internet. For cloud computing, especially under the Internet of things of cloud computing and can analyze the simulation using the Hadoop, Hadoop is a distributed system infrastructure, its origin is an open source web crawler Nutch. Hadoop is the huge amounts of data computation and storage is one of the more mature technologies, especially for offline data processing [1-2]. Especially for large data network intelligent household this complex, just using illuminating data mining algorithms, obviously which cannot meet the needs of the business data mining, business often require fast convergence of the method. Complex network structure of clusters found to intuitive understanding of the function of the network, in-depth analysis of the topology of the network, found hidden in the network characteristics and law and predict the behavior of the network has very important reference value.

ISSN: 2287-1233 ASTL

2 Related works

2.1 Smart home

Smart home is IT technology, control technology and network technology to the traditional home appliance industry is the inevitable result of penetration. With the rapid development of information technology in recent years, high communication liberalization and hierarchical, rapid growth of the business and the people of the working and living environment improvement of safety, comfort and efficiency, the social demand for smart home greatly increase; Moreover in terms of technology, electronic information and communication technology and the development of the computer control technology, also contributed to the birth of smart home. The function of smart home system is mainly composed of the following aspects [3-4]:

1. Electrical appliances intelligent control: the automatic control of electrical appliances, such as according to the indoor temperature automatic began to close the work condition of the air conditioning air conditioning or mediation.
2. Lighting intelligent control: according to the indoor light intensity to automatically adjust the lighting switch or intensity of illumination. Which can keep indoor light brightness is in the appropriate boxes.

2.2 Complex network data mining research status

Complex networks is the fundamental basis of clustering algorithm by using network topology mining which real family structure, and its target is also found that there were a bunch of structure. It is because of this kind of algorithm can be found in the network cluster structure, as a result, this algorithm is of important theoretical significance and application value, it is not only an important technology of the computer science is used to solve the graphic division, is also found in the community of the important methods of sociology, at the same time, also is in the field of biology research fields, such as their main method.

Based on optimization algorithm, the algorithm divide the structure of the network cluster problem to solve with optimization method in the field of mathematics, the idea set an objective function in advance, constantly optimize the function in the process of dividing to determine whether the division of the family structure is the optimal solution, on the basis of this method can be divided into two categories, spectrum method and the local search method. Such as: the average cut algorithm, ratio cut algorithm, FN algorithm and simulated annealing algorithm, etc. Aggregation coefficient C is used to describe the gathering for nodes in the network, the network of collectivization.

3 The proposed scheme

3.1 Data mining based on intelligent household

In smart home network, which carries on the segmentation, general way of segmentation for the triangle subdivision, to determine the area of distribution network, which found that the distribution of the smart home network has the character of asymmetric area, sections of its objects interact with each other, the relation between according to the theory of random graphs, each node cluster on the influence of other nodes cluster each are not identical, and according to the size of the radius of influence is quantified. The object of study for the random network, passing information between nodes is fixed, the stochastic model can be used in the model as the basis of research. Definition of topology matrix encoding, the encoding for

$$C(B) = b_{12}b_{13}b_{23}\cdots b_{(n-1)n} \quad (1)$$

In a directed graph can be defined as

$$C(B) = d_{12}d_{13}d_{23}\cdots d_{(n-1)n} \quad (2)$$

The backlog of candidate solutions through the comparison of Pareto partial order relation, there will be a selective collection into an external. The local search procedure is described below [5-7]:

Step 1, randomly select a node s_i in the locality of submanifold, computing the similarity of two nodes, according to the similarity to determine the direction, the similarity is defined as

$$d(s_i, s_j) = 1 - \frac{\sum_{k=1}^{n} s_{ik} \oplus s_{jk}}{n} \quad (3)$$

Where, $\oplus$ as the exclusive or operation, if the two values are less than the threshold, the results of an exclusive or operation is 0, or 1.

Step 2, if $d(s_i, s_j) > \lambda$,processing step3. Otherwise, start with the step1.

Step 3, making discretization of differential, the differential algorithm as shown in (4)

$$s_{ij}^{n} = s_{ij} + G(s_{ij} - s_{kj}) \quad (4)$$

Where, $G(\square)$ is a step function, its decision to the speed of convergence.

Step 4, comparing with s_{ij}^{n} and each child node of links, if $s_{ij}^{n} < s_i$, s_i is substituted for s_{ij}^{n}, otherwise, repeat the first step, until replacing the constraints.

3.2 The overall algorithm process

According to the raw data to calculate the similarity, the similarity matrix put the similarity matrix is greater than the value closed with said that is less than the value of local closed with 0 so it generated 0, 1 matrix as a network of adjacency matrix to do so under the premise of guarantee the network performance greatly simplifies the calculation. The source of chromosome number belongs to this kind of class in the node corresponding to the purpose of the class number with such number of chromosomes, thus getting the new offspring chromosomes [8-10].

Step 1. Initialization related parameters, randomly generated randomly generated a q of individual species, in this article is the child node cluster, which does not limit the length of the external collection;

Step 2. The population of the inferior individuals into the external collection.

Step 3. Choosing u from randomness in current population individuals, perform crossover and Gaussian mutation operator to generate a new individual;

Step 4. To choose a new individual to perform operation " selecting one individual into ES, if it is not feasible to another by comparing individuals, perform the local search process"

Step 5.Repeating step 3 and 4 until generating n individuals, replacing the original population;

Step 6. If termination conditions are not met, then turned to step 2, otherwise, the algorithm performs over.

4 Simulation results and analysis

This paper uses the Matlab language to realize, in the Intel 64 bit processors and 8 G memory on the machine running, the Windows 7 operating system. Researchers have proposed a number of different evaluation criteria to evaluate complex network data mining standards, such as

Table 1. The testing software environment

Data network	nodes	edge	Snapshot number
Job site topology	4873	343303	20
home	251	14057	20
Stadium	2359	642959	28
Business zone	58186	160224	10

The simulation of the road is generally shown in the following figure, according to the virtual task and scheduling to achieve the model, the core

algorithm for the development of the preparation of its scheduling interval according to the different simulation environment, need to set up a separate. Figure 1 shows the running time of the job site network under different snapshot numbers T. It can be found that the running time is not proportional to the number of snapshots, in contrast, it is inversely proportional to the number of snapshots.

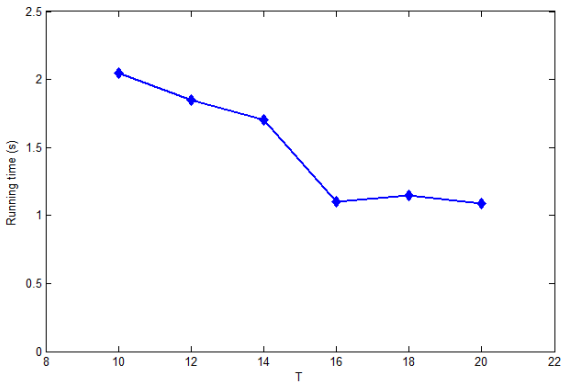

Fig. 1. Running time

As can be seen from Figure 1, the time cost of the algorithm of single machine mode can be increased sharply with the increase of sample data, and the time cost is increasing slowly, which is more obvious with the increase of data quantity. It can be seen that, with the increase of the number of nodes, the running time is not increasing, but it tends to be steady.

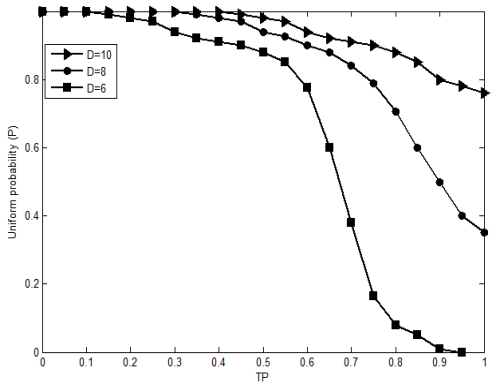

Fig. 2. Unified probability affects by TP change

The TP value of the variable clustering non-scale network increases, the average maximum number of points will be reduced, the clustering coefficient has an obstacle effect. The main reason is that the higher clustering coefficient

implies that the individual neighbor relationship is more closely related to the formation of local individual.

5 Conclusion

In this paper, according to the characteristics of the smart home network, and the characteristics of the data correlation, based on the nature of the complex network of self-organized criticality, the data is data mining;Data mining method for the improved genetic algorithm, a simulation experiment can be found that the convergence of the proposed algorithm has robustness for network topology.

References

1. Adl, R. K., Askari, M., Barker, K., Safavi-Naini, R., Privacy consensusin anonymization systems via game theory. in Proc. 26th Annu. DataAppl. Security Privacy, 2012, pp. 74–89.
2. Karimi Adl, R., Barker, K., Denzinger, J.: A negotiation game: Establishing stable privacy policies for aggregate reasoning,'' Dept.Comput. Sci., Univ. Calgary, Calgary, AB, Canada, Tech. Rep.,Oct. 2012. [Online]. Available: The paper is available at http://dspace.ucalgary.ca/jspui/bitstream/1880/49282/1/2012-1023-06.pdf
3. Kargupta, H., Das, K., Liu, K., Multi-party, privacy-preserving distributed data mining using a game theoretic framework. in Proc.11th Eur. Conf. Principles Pract. Knowl. Discovery Databases (PKDD), 2007, pp. 523–531.
4. Miyaji, A., Rahman, M. S.: Privacy-preserving data mining: A game-theoretic approach. in Proc. 25th Data Appl. Security Privacy,2011, pp. 186–200.
5. Ge, X., Yan, L., Zhu, J., Shi, W.: Privacy-preserving distributed association rule mining based on the secret sharing technique. In Proc. 2nd Int. Conf. Softw. Eng. Data Mining (SEDM), Jun. 2010, pp. 345–350.
6. Nanavati, N. R., Jinwala, D. C.: A novel privacy preserving game theoretic repeated rational secret sharing scheme for distributed data mining,'' vol. 91, 2013. [Online].Available:http://www.researchgate.net/publication/256765823_A_NOVEL_PRIVACY_PRESERVING_ GAME_THEORETIC_REPEATED_RATIONAL_SECRET_SHARING_SCHEME_FOR_DISTRIBUTED_DATA_MINING
7. Halkidi, M., Koutsopoulos, I.: A game theoretic framework for data privacy preservation in recommender systems. in Proc. Mach. Learn. Knowl. Discovery Databases, 2011, pp. 629–644.
8. Ioannidis, S., Loiseau, P.: Linear regression as a non-cooperative game. in Proc. Web Internet Econ., 2013, pp. 277–290.
9. Chakravarthy, S. L., Kumari, V. V., Sarojini, C.: A coalitional game theoretic mechanism for privacy preserving publishing based on k-anonymity,'' Proc. Technol., vol. 6, pp. 889–896, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212017312006536
10. Nix, R., Kantarciouglu, M.: Incentive compatible privacy-preserving distributed classification. IEEE Trans. Dependable Secure Comput., vol. 9, no. 4, pp. 451–462, Jul. 2012.

