Available online at www.sciencedirect.com

SciVerse ScienceDirect

AASRI Procedia 3 (2012) 209 – 216

AASRI Procedia

www.elsevier.com/locate/procedia

2012 AASRI Conference on Modeling, Identification and Control

A Mining Model for the Worst Plot in the Mobile Communication Networks

Qi Zhang[a], Ping Fu[b],Jinming Peng[a], Daqiao Huang[a,]*

[a] *Zhejiang Communications Services Co.,Ltd Network Technology Company, 310008 Hangzhou, P.R.China*
[b]*College of Computer Science and Information Engineering, Zhejiang Gongshang University, 310018 Hangzhou, P.R.China*

Abstract

With the mobile communication network's coverage and capacity having reached a high level, providing the high-quality network services for users becomes the primary task of the network operation management. The worst cells having the worst quality are closely related with the network performance. In this paper, a mining model for digging out the worst cells in wireless networks comes up. It uses AHP to determine the index weights and introduces these weights into clustering algorithm. An example has verified its feasibility. Therefore, this model has certain significance on the development and optimization of wireless networks.

Keywords: Mining Model, AHP, Network Service, Index weight, clustering;

1. Introduction

Mobile communication is one of the most advanced ways for communication in the world. With its network coverage and capacity having reached a high level, quality, cost and service are more concerned about [1]. How to make the communication network to run at their best and provide users with high-quality network services? They are the primary tasks of the network operation and management. The worst cells have

*Corresponding author. Tel.:13735530027.
E-mail address: fpmyself@126.com.

doi:10.1016/j.aasri.2012.11.035

become the key indicators to measure the wireless quality in the mobile communication networks. At present, one of the worst cells optimization problems is how to dig out the worst district efficiently. In the past, people queried the worst cells just to rely on their definition [2], and the results only were the worst cells lists, which couldn't show the extent of their "worst". In this paper, a mining model based on the clustering algorithm is established to query the worst cells. It can not only dig out the worst cells efficiently, but also reflect their "worst" and priorities of time so as to take optimization measures accurately. So we can say that it plays an essential role in the optimization of wireless networks.

2. Paper Preparation

2.1. Select characteristic indicators

Select data indicators clearly and accurately, understanding that the purposes of data mining are important steps in data mining [3]. In this model, choose seven variables which have great influence on the cells performance as the characteristics description indicators, they are TCH Call Dropping Rate, Each Channel Traffic, TCH call Dropping Number, TCH Connection Rate, TCH jam Number, TCH Jam Rate and Interference Number.[5]

2.2. Data Collection and Data Preprocessing

In this model, the data preprocessing techniques such as vacancy value processing, data cleaning and data transformation are mainly used. There are two kinds of processing methods to handle the vacancy value, one is to use the default value for filling, and the other one is to delete the record. In this paper, use the second method. Data cleaning mainly reduces data quantity which participates in modeling which called variable clip, the processing time and data capacity decrease relatively. [6][10] For example, in the same cells, TCH traffic and each channel traffic has a strong positive correlation in statistical sense, so use one as the indicator in model. Use digital code when deal with each channel traffic. According to the definition in the China Unicom GSM nets mobile communication performance statistics system, when each channel traffic is more than 0.6 erl, call them super busy cells; when each channel traffic is between 0.1 erl and 0.6 erl, call them normal service cells; when each channel traffic is less than 0.1 erl, call them super idle cells; convert them respectively into 1, 2, 3. In addition, the dimension of indicators is not identical, which needs to make these indicator data standardizing.

2.3. Index weight Determination

The above indexes for the target layers have different importance, in another words, they have different weights. In this paper, use the method AHP (Analytic Hierarchy Process) to determine index weight. The method compare the importance between one index and another one, then establish the corresponding judgment matrix, at lastly, solve the eigenvalue of the matrix [9].

Assuming the weights of the first grade indexes:

$$A = \{a_m\}, m = 1, 2, ..., 4; a_m \geq 0, \sum_{m}^{5} a_m = 1 \quad (1)$$

The weights of the secondary indexes in their corresponding first grade indexes:

$$A_m = \{a_{mn}\}, m = 1,2,...,4; n = 1,2,...,k; a_{mn} \geq 0, \sum_{n=1}^{k_m} a_{mn} = 1 \tag{2}$$

The weights of the secondary indexes for the target layer:

$$A_{mn} = \{b_j\} = \{a_m \times a_{mn}\}, \quad m = 1,2,...,4; \quad n = 1,2,...,k; j = 1,2,..., p; x_j \geq 0, \sum_{j=1}^{p} b_j = 1 \tag{3}$$

The solving results are provided directly here $A = \{0.3, 0.2, 0.25, 0.25\}$. The detailed content can be seen in table 2.

Table 1. AHP Determine the Indexes Weights

The target layers	The first grade index and the weight a_m	The secondary indexes	Weight a_{mn}	Weight b_j
The main considering cell indicators	Call dropping performance indicator 0.3	Call Dropping Number	0.4	0.12
		Call Dropping Rate	0.6	0.18
	Traffic performance indicator 0.2	TCH Connection Rate	0.5	0.1
		Each Channel Traffic	0.5	0.1
	Connection performance indicator0.25	Jam Number	0.4	0.1
		Jam Rate	0.6	0.15
	Connection performance indicator 0.25	Category 5 External Interference	1	0.25

2.4. Design DM Algorithm

In this model, treat the worst cells as outlier and choose clustering algorithm. Cluster analysis as a quantitative method, gives a more accurate and detailed classification tool on the view of data analysis. K-means is a typical K-means algorithm based on distance, and take the distance as the similarity of the evaluation indicator, which thinks that the closer two objects are, the greater the similarity is. [11] The algorithm considers that clusters are composed of the objects with close distance, and take the Euclidean distance as the similarity measure, so its ultimate goal is to get the compact and independent cluster [4]. The specific approach is shown as follows:

- Choose k objects at random, each object represents the initial mean or center of a cluster, for the rest of the objects, assign them to the most similar cluster according to the distance from them to the mean of each cluster. Assuming the sample set is divided into C classes, by using the vector quantity, the clustering canter of the Class h is represented as $s_j = (s_{1h} s_{2h} ... s_{mh})$, in the formula, $h = 1,2,\cdots,c$; $0 \leq s_{ih} \leq 1$ .The differences between sample j and class h can use the generalized distance to signify.

The generalized distance formula:

$$d_{hj}^0 = \left| \sum_{i=1}^{m} (r_{ij} - s_{ih})^p \right|^{\frac{1}{p}} \tag{4}$$

In other words, this is the foundation of the algorithm similarity measure. Because of the weights' existing, different indexes have different influence to the clustering. It also is an important variable parameter, by introducing the index weight vector into the generalized distance formula, and then we get the generalized weighted distance formula Eq.5.

$$d_{hj} = \left| \sum_{i=1}^{m} w_i (r_{ij} - s_{ih})^p \right|^{\frac{1}{p}} \tag{5}$$

- Calculate the new mean of each cluster.
- Repeat step 2 until criterion function is convergent. Use the squared error criterion. The criterion tries its best to make the k generated clusters compactly and independently.

Nomenclature

p the distance parameter. When p=1, the distance is Hamming distance; when p=2, the distance is Euclidean distance.

m_i the mean of the cluster C_i

3. The Mining Model for the Worst Cell Blending in the Weight of Indictors

3.1. Model Description

This model is mainly divided into four modules, and respectively is the index module, data module, mining module and the optimization module. The index module is divided into index selection and index weight determination. In the process to determine the index weight, use the method AHP, relying on subjective experience. Data module includes related data collection and processing; Mining module is used to combine the index weight into the data mining algorithm and find out the worst cells promptly and accurately; the optimization module for the worst cells mainly includes the correlation analysis of the performance indexes, then takes the corresponding measures according to the different situations.

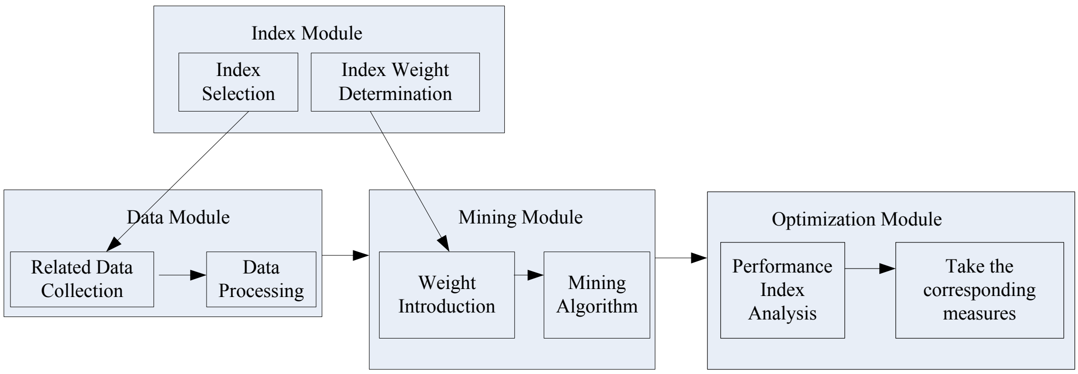


Fig. 1. The Mining Model for the Worst Cell Blending in the Weight of Indictors

3.2. Model Realization

3.2.1. DM Algorithm Realization Blending in The Indicator Weight

Input: k: the number of clusters; in this model, set k=4;

D: data set contains n objects.

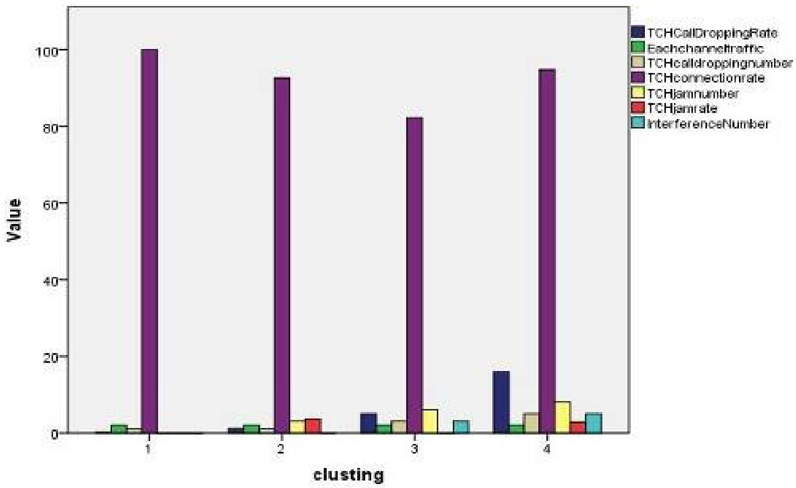
Output: a collection of k clusters. In this model, contains 16 objects.

Methods:

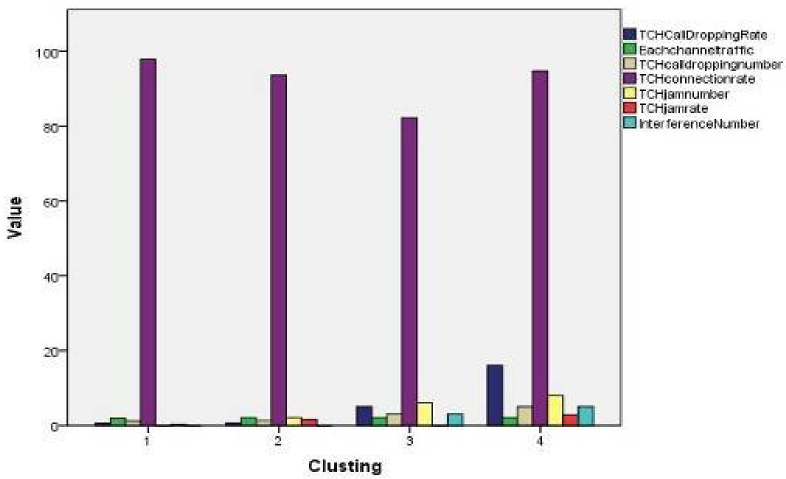
- Select k objects at random from D as the initial cluster center;
- Repeat;
- According to the generalized weighted distance between these objects and the mean of each cluster, assign each object (again) to the most similar cluster;
- Update the cluster mean to calculate the mean of the objects in each cluster until no longer changes.

3.2.2. Results Shown

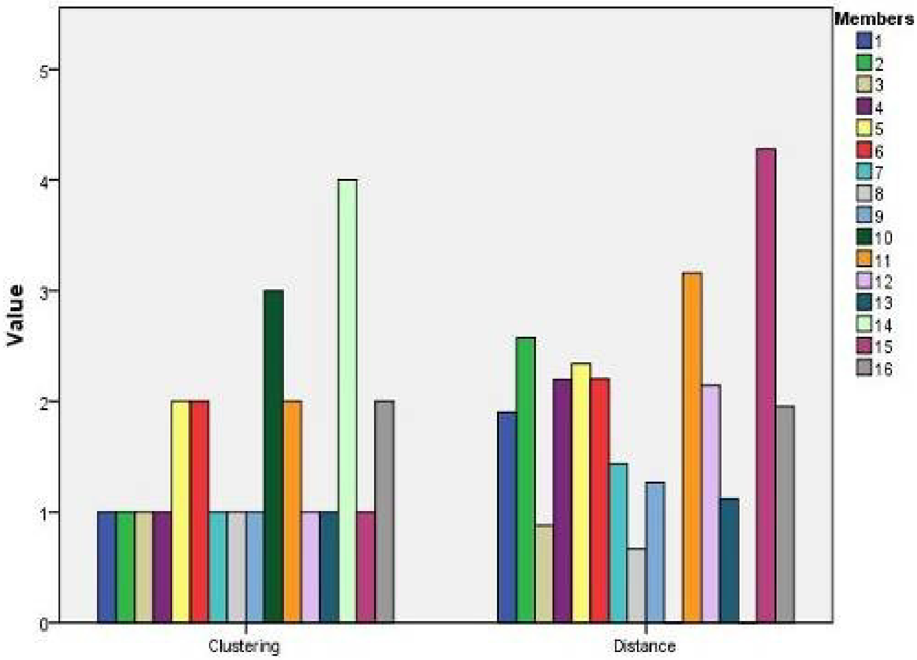
In this model, take the data of GSM cells; see them as objects of study. Get some results by using the model and show them in fig.2.



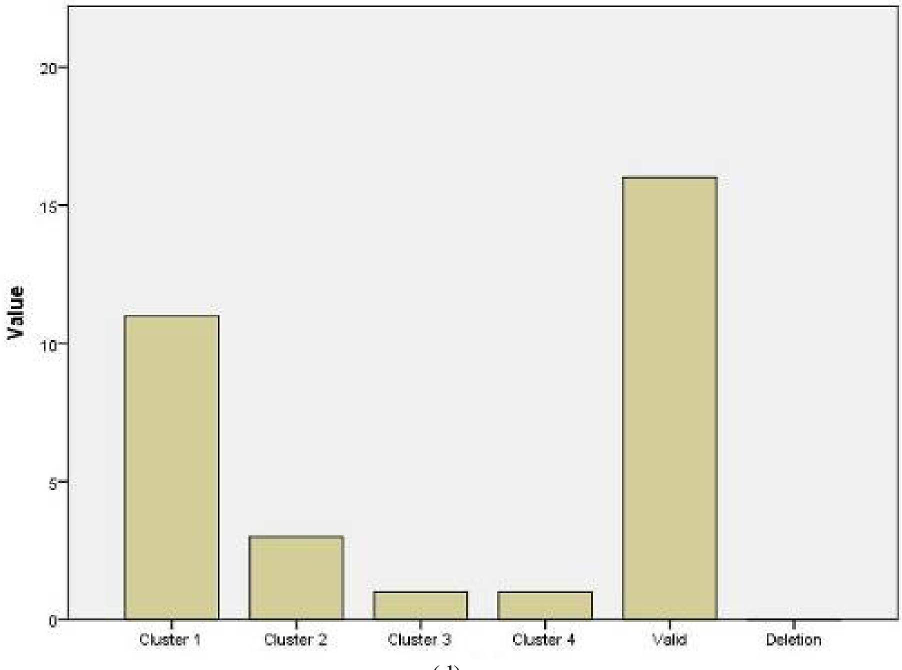
(a)



(b)



(c)



(d)

Fig. 2. (a) Initial Cluster Centers; (b) Final Cluster Centers; (c) Clustering Analysis Tree Diagram; (d) Cases' number in Each cluster.

3.3. Results Analysis and Model Verification

Realize the mining model for the worst cells, and get some conclusions such as the follows:

- According to the sorting, 14 sample cells have the worst quality, followed by 10 sample cells, they all need to find the reasons and be processed as far as quickly.
- According to the clustering results, the cells in cluster 1 have good indicators, cannot be temporarily processed; the cells in cluster 4 have the worst indicators, so they are the worst ones. We need to find out the reasons and process them in time. The "worst" extent of the cells in cluster 2 and cluster 3 is between the cluster 1's and the cluster 4's, with different situations, so we need to take various measures to treat them.
- According to the definition of the worst cells in the Chinese Unicom GSM nets mobile communication performance statistics system, 14 and 10 samples of the telephone traffic are set to 2, and satisfy the requirements that every channel traffic value must be during 0.1 erl and 0.6 erl. They also have more than 3% call dropping rate. Verify the correctness of the definition by the model results.[8]All tables should be numbered with Arabic numerals. Headings should be placed above tables, left justified. Leave one line space between the heading and the table. Only horizontal lines should be used within a table, to distinguish the column headings from the body of the table, and immediately above and below the table. Tables must be embedded into the text and not supplied separately. Below is an example which authors may find useful.
- Get rid of the history that just relied on the definition to query the worst cells, and mine out the worst cells efficiently, also reflect the "worst" extent of the worst cells, at last improve the efficiency of dealing with the cells which have the worst quality.

4. Conclusion

The proposed mining model of the worst cells based on clustering algorithm, blending in the index weights, just offers a new idea to dig out the worst cells effectively. To some extent, it has certain significance on the development and optimization of wireless network, but still has a lot of problems. One is that only consider the congestion and interference problems when indicators are selected, but actually there are still many other performance indicators; another one is that K-means algorithm itself has many defects to improve and the method(AHP) has strong subjectivity. Therefore, we still need further studies to optimize and solve the worst cells problems.

Acknowledgements

This paper is supported by Zhejiang Gongshang University Graduate Student Scientific Innovative Project.

References

[1] Ye Qing, Zhou Ming-zheng, Zha Wei-xing. A Wireless Network Optimization Method Based on Reducing the Proportion of the Worst Plot, Modern Computer, 2010.

[2] Wang Yan-ping, Yue Chun-xia. Analysis and implement of worst cell based on data mining. Computer Engineering and Design, 2007.

[3] Shen Liang. The Research of Data Mining in Mobile Network Optimization, ShangHai Jiao Tong University, 2009.

[4] Sun Geng, Feng Yan-hong; GUO Xian-jiu; Zhang Chun-ping. Research on K-means Clustering Algorithm, Journal of Changchun Normal University,2011.

[5] Zhang Yongmin. Application and Realization of Data mining in the Mobile Network Optimization. University of Electronic Science and Technology of China,2007.
[6] Jiawei Han, Micheline Kamber.Data Mining Concepts and Techniques. China Machine Press,2007.
[7] WU Jinhai, LI Yinan, ZHOU Yaobiao. The GSM Network Optimization of Mobile Communication. Journal of Liaoning University(Natural Sciences Edition),2006.
[8] Han Bingjie.GSM principle and its network optimization. China Machine Press,2004.
[9]WANG Wei, ZHANG Yu-zhi. Method of Obtaining Eigenvector for a Fuzzy AHP. Control and Decision,2006(02).
[10] S. Anahory, D. Murray, Data Warehousing in the Real World,Addison Wesley, 1997, pp. 327– 339.
[11]J. Han, M. Kamber, Data Mining Concepts and Techniques, 2000.