Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 486–492

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Internet of Smart Things(IST2021)
November 1-4, 2021, Leuven, Belgium

Analysis of Intelligent Energy Saving Strategy of 4G/5G Network Based on FP-Tree

Ruili Qi[a,b,*], Xuetao Guo[c]

[a] *School of Computer Science & Engineering,Shijiazhuang University, Shijiazhuang,050035, China*
[b] *Hebei Province Internet of Things Intelligent Perception and Application Technology Innovation Center, Shijiazhuang ,050035, China*
[c] *China Telecom Shijiazhuang Branch, Shijiazhuang 056300, China*

Abstract

With the large-scale deployment of 5G network of communication operators, there are more and more 5G devices, and the power consumption of mobile network surges. This paper uses the improved FP-Tree (Frequent Pattern Tree) to mine association rules for the service load of 4G/5G base stations of operators. It reduced the number of times of traversing the data set when generating candidate frequent and complex itemsets, and fully considers the problem of low utilization of equipment resources caused by ”Peak Valley Effect”, so as to improve the accuracy and operation efficiency of the model. It greatly improves the energy-saving potential of the equipment. The evaluation of the operator’s 4G/5G network KPI, KQI and other indicators shows that the network performance is stable and the user perception is basically unchanged.

Keywords: 4G/5G network; Improved FP-Tree; Smart energy saving; User perception

1. Introduction

The scale of domestic mobile network has increased year by year, and 5G base stations have covered major cities, counties and towns. However, the power consumption of equipment caused by 5G NR network high frequency communication is 3-5 times that of 4G. The station construction density is also greater than 4G network. The electricity cost of communication operators is soaring, so it is imperative to study the energy saving and consumption reduction of 5G base station.

For the energy-saving effect of communication base stations, scholars have carried out in-depth research work and achieved good results. The energy-saving methods of base stations can be divided into three methods: the first type is hardware upgrade. The second is simple linear model or expert experience, which depends on expert experience to

* Corresponding author. Tel.: 15369155136.
E-mail address: 309222063@qq.com

1877-0509
10.1016/j.procs.2021.12.274

judge and operate. The third category is machine learning algorithms, such as gradient lifting regression tree, neural network and so on. Practice shows that machine learning algorithm has better energy-saving effect than traditional expert experience. Zhang Qing[1] proposed to apply the statistical analysis and data modeling methods in mining analysis to the base station cell information and wireless service data to mine the traffic tide phenomenon. Jiang Guoqing[2] and others proposed cluster analysis of active home base stations to sleep unnecessary base stations in the network to realize network energy saving. Reference [3] proposed a method of applying K-shape algorithm based on load curve morphological clustering to user daily load data analysis. Literature [4] proposed using Artificial Bee Colony Algorithm to analyze the predicted traffic information in each time period of the network, so as to find the optimal switching state of the load.

At present, there are few literatures on cluster analysis involving base station side data, and the definition of energy-saving scenario is relatively conservative and solidified. Therefore, this paper uses the improved FP-Tree to mine the association rules of the service load of the operator's 4G/5G base station, by reducing the number of traversing the data set when generating the candidate frequency complex item set, and fully considers the low utilization of equipment resources caused by the "Peak Valley Effect", so as to improve the accuracy and operation efficiency of the model and greatly enhance the energy-saving potential of the equipment.

2. Association rules and FP-Tree

2.1. Association rules

Association rule mining is an important part of data mining technology [5, 6, 7]. It is mainly used to find the implicit relationship between a large number of data [8, 9], so as to describe the laws and patterns of some attributes of a thing at the same time. The classical Aprior algorithm [10, 11] adopts the layer by layer search method. When generating frequent itemsets, it needs to scan the original database for many times, and self connection will produce a large number of redundant candidate itemsets, resulting in a long time for model search itemsets and reducing the mining efficiency of the model.

This paper uses the improved frequent pattern tree FP-Tree [12] to cluster and mine the service load of the operator's 4G/5G base station by reducing the number of traversing the data set when generating the candidate frequency complex itemset, so as to improve the energy-saving potential of the equipment.

2.2. FP-Tree

The idea of FP-Tree is to construct a tree, mapping the data in the data set to the tree, and then find all frequent itemsets according to the FP-Tree.It is a prefix tree used to store the occurrence times of items. Each item is stored in the tree structure in the form of path. FP-Growth algorithm [13, 14, 15] means that by scanning the transaction dataset twice, the frequent items contained in each transaction are stored in FP-Tree in descending order of their support. In the process of discovering frequent patterns in the future, there is no need to scan the transaction dataset, but only search in FP-Tree. Frequent patterns can be generated directly by recursively calling FP-Growth method, so candidate patterns do not need to be generated in the whole discovery process, and the execution efficiency is obviously better than Apriori algorithm. Algorithm 1 describes the construction process of FP-Tree.

Algorithm 1	Constructing FP-Tree
Input:	Transaction dataset D, minimum support SM
Output:	FP-Tree
Step 1:	Scan D to obtain frequent itemset F and its support, and arrange F in descending order of support. The result is frequent itemtable L;
Step 2:	Create the root node of FP-Tree T = null;
Step 3:	For each transaction T in D, repeat step 4 and step 5;
Step 4:	Sort all frequent items in T according to the order in L;
Step 5:	Each item I in T. If there are nodes with the same path in the FP Tree, its count will be increased by 1. Otherwise, add the project node, set its count to 1, and link it to its parent node T.

2.3. Improved FP-Growth algorithm

This paper uses the idea of ”Vertical Segmentation-Resegmentation-Sub Database Mining-Result Integration” to improve FP-Growth algorithm. Through the analysis of cell traffic load data, this paper mines the daily traffic data law of a specific cell, finds the daily peak phenomenon and low energy-saving period, and takes energy-saving measures periodically.

Firstly, the mobile network service load data is divided into different sub databases according to the cell ID, and then divided into cell daily sub databases according to the day. Then, the improved FP-Growth algorithm is used to mine the frequent itemsets of each subset, and then the frequent itemsets of each subset are integrated to provide services for service load data analysis.

Algorithm 2 describes the execution process of the improved FP-Growth algorithm.

Algorithm 2	Improved FP-Growth Algorithm
Input:	Service load database D,minimum support SM
Output:	Frequent mode set
Step 1:	Divide the data in D into cell sub database Cell-Database according to cell ID;
Step 2:	Divide each Cell-Database into days again to obtain the daily service load database Cell-Daily-Database of each cell;
Step 3:	For each Cell-Daily-Database, construct FP-Tree according to algorithm 1, and obtain the frequent itemset of Cell-Daily-Database from FP-Tree;
Step 4:	Repeat step 2 and step 3 for each Cell-Database obtained by step 1 to get the frequent itemset of Cell-Database;
Step 5:	Integrate the frequent itemsets of step 4 to obtain the global candidate frequent itemsets.

3. Application of improved FP-Growth algorithm in business load data analysis

3.1. Data cleaning

The data in this paper comes from the historical network load data of a telecom operator for 12 consecutive weeks in recent 3 months in 2021, with a total of 120 cells and 215016 records. Each record includes 21 related indicators of user traffic data, such as *Start Time* (i.e. collection start time, hereinafter abbreviated as *ST* for convenience), *End Time* (i.e. collection end time, hereinafter abbreviated as *ET*), *Query Granularity* (i.e. collection duration, hereinafter abbreviated as *QG*), *Physical Cell* (Cell Identification, hereinafter abbreviated as *CI*), *Load*(i.e. cell load data value,hereinafter abbreviated as *LD*),*Average Occupancy of Downlink Physical Resource Block*(hereinafter abbreviated as *AD*) and *Average Occupancy of Uplink Physical Resource Block*(hereinafter abbreviated as *AU*). Some original data are shown in Table 1.

Table 1. Partial raw data.

NO.	ST	ET	QG(*minutes*)	CI	LD(%)	AD(%)	AU(%)
1	2021/7/19 00:00	2021/7/19 01:00	60	48	4.50	4.20	4.13
2	2021/7/20 01:00	2021/7/20 02:00	60	1	1.80	1.75	1.31
3	2021/7/21 03:00	2021/7/21 04:00	60	3	2.10	1.88	1.83
4	2021/7/22 01:00	2021/7/22 02:00	60	4	2.60	1.83	2.46
5	2021/7/23 11:00	2021/7/23 12:00	60	48	88.20	87.91	66.00
6	2021/7/24 16:00	2021/7/24 17:00	60	49	92.00	90.74	55.34

Firstly, the original load data is cleaned to eliminate inconsistencies and noise interference data. This paper cleans up the original data by removing noise data, deleting outliers and solving inconsistencies, so as to eliminate abnormal data, correct errors and remove duplicate data. In the original data table, delete the records with more missing attribute values. For the records with less missing numerical attribute values, fill them with the average value of this column. There are 193512 records in the cleaned data set.

3.2. Data conversion

In order to facilitate data representation and processing, it is necessary to convert the original data. The numerical data in the service load data table is discretized. Some discretization rules are as follows: according to the network performance key performance index(KPI) and user perception key quality index(KQI), *LD,AD* and *AU* are divided into four intervals: [0,20], [21,40], [41,60] and [61,100]. Simplified representation of text data. The collection Start Time is expressed as ST0, ST1, … and ST23 from 0 to 23, and the End Time is expressed as ET0, ET1, ... and ET23 from 0 to 23. For example, if the *ST* and *ET* of a record are ST0 and ET1 respectively, it means that the user's load data is collected for 60 minutes from 0 to 1.The partially converted data are shown in Table 2.

Table 2. Partially discretized data table.

NO.	ST	ET	QG(*minutes*)	CI	LD(%)	AD(%)	AU(%)
1	ST0	ET1	60	48	LD[0,20]	AD[0,20]	AU[0,20]
2	ST1	ET2	60	1	LD[0,20]	AD[0,20]	AU[0,20]
3	ST3	ET4	60	3	LD[0,20]	AD[0,20]	AU[0,20]
4	ST1	ET2	60	4	LD[0,20]	AD[0,20]	AU[0,20]
5	ST11	ET12	60	48	LD[61,100]	AD[61,100]	AU[61,100]
6	ST16	ET17	60	49	LD[61,00]	AD[61,100]	AU[41,60]

3.3. Rule analysis

For the processed data set, algorithm 2 is used for experiment. The experimental environment is CPU: Intel (R) core (TM) i7-11370h and the operating system is windows 10. It is mined in the environment of MATLAB 2014.The minimum support threshold is 0.50 and the minimum confidence threshold is 0.60. Some association rules are shown in Table 3. The rules in the table show that the load data of the base station changes obviously with time. Rule 1 shows that between 1:00 a.m. and 6:00 a.m., when the average occupancy rate of downlink physical resource blocks is not higher than 20%, it is more likely that the user service load is not higher than 20%. Rule 5 shows that between 10:00 a.m. and 12:00 a.m., the user's service load is more likely to be greater than 61%. Subsequently, corresponding energy-saving measures will be taken for the cell base station according to the result data.

Table 3. Partial association rules.

NO.	Association Rules	Degree of Support	Degree of Confidence
1	(ST1,ET6,AD[0,20])->LD[0,20]	0.60	0.70
2	(ST7,ET9, AD[21,40],AU[41,60])->LD[21,40]	0.55	0.75
3	(ST21,ET23, AD[21,40],AU[21,40])->LD[21,40]	0.58	0.75
4	(ST13,ET15, AD[20,40],AU[41,60])->LD[41,60]	0.56	0.72
5	(ST10,ET12, AD[61,100],AU[41,60])->LD[61,100]	0.59	0.80
6	(ST16,ET18, AD[61,100],AU[61,100])->LD[61,100]	0.60	0.75

4. Results and analysis

4.1. Identification of service load characteristics

4.1.1. Recognition of week feature

Matlab is used to model the business load using the improved FP-Growth algorithm to identify the energy-saving characteristics of the same site. For the daily load data of each frequency band cell at the same station, set the support and confidence thresholds, and adopt the improved FP-Growth algorithm to obtain the analysis results.

(1) Obvious weekly characteristics: if the support of the frequent itemset is greater than 0.55 and the confidence is greater than 0.70, it indicates that the clustering effect of the daily load curve of the station is better, and the weekly characteristics of the station can be considered to be obvious.

(2) Consistent weekly characteristics: if the support of the frequent itemset is less than 0.65 and the confidence is greater than 0.72, it indicates that the daily load fluctuation of the station is extremely similar, that is, the trend of the base station is consistent throughout the week.

(3) No obvious characteristics: in addition to the above two types of sites, it can be considered that there is no obvious change law.

In particular, energy-saving measures are adopted according to the peak valley coefficient for daily peak valley feature recognition.The first step is to calculate the peak valley effect coefficient of the base station and judge whether there is daily peak valley phenomenon.The second step is to calculate the daily peak period and valley energy-saving period for the base station with daily peak and valley phenomenon, and define a more detailed energy-saving scenario for the base station in combination with the weekly effect and valley energy-saving period.

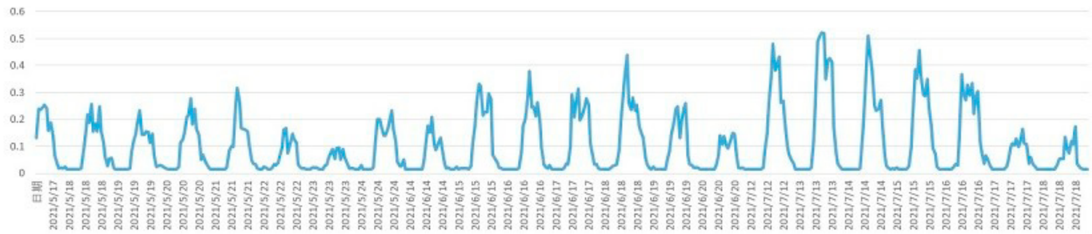

Fig. 1. hourly service load curve of an office building in the urban area.

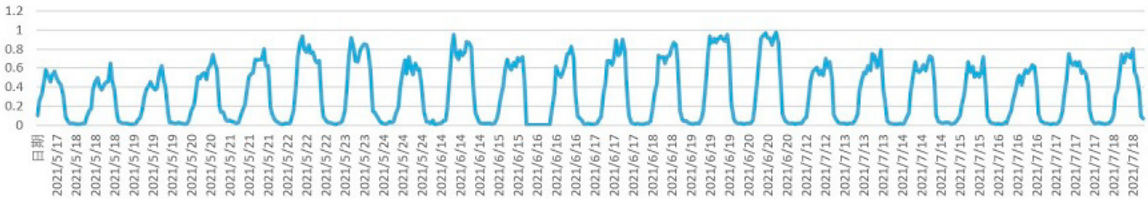

Fig. 2. hourly business load curve of a commercial complex in urban area.

4.1.2. *Comprehensive energy saving measures*

The platform connects with the operator's north interface to automatically collect load data, establish energy-saving prediction model according to the above steps, and generate site intelligent energy-saving strategy. According to the threshold index, the platform adopts symbol level, channel level, carrier level and station energy-saving measures and energy-saving time periods.

(1) Energy-Saving of Symbol. When there is no baseband data transmission, the orthogonal frequency division multiplexing(OFDM) symbol is turned off, and vice versa.

(2) Energy-Saving of Channel. The channel saves energy and controls the startup and shutdown of radio remote unit(RRU) power amplifier. When the cell load is low, close some transmission channels, increase the service load and restore the original channel transmission.

(3) Energy-Saving of Carrier. Control the startup and shutdown of RRU different frequencies. In the scenario of different frequencies and same coverage, the sum of "basic coverage cell" and "energy-saving cell" is less than the threshold, and turn off the "energy-saving cell", so as to achieve the purpose of energy saving.

(4) Energy-Saving of Station. The Internet of things communication technology is used to send commands to the controller outside the base station for automatic energy-saving operation.

4.1.3. *Effect evaluation of comprehensive energy saving*

(1) Week features obvious of base station

Fig.1 shows the change of business load of an office building station in Hebei during 12 weeks from 20210503 to 20210725. This type of site repeats the same fluctuation law every week. The weekly working day load is relatively high, more than 50%, and the load is at a low point on the rest day, less than 20%.

(2) Week Feature Consistent of Base Station

Fig.2 shows the change of business load of a commercial complex in Hebei in 12 weeks from 20210503 to 20210725, The business load characteristics of the site type are the same on weekdays and rest days. The energy-saving measures taken from 0 a.m. to 7 a.m. have a significant effect.

(3) Irregular of base station

Fig.3 shows the change of business load of a residential house in Hebei in 12 weeks from 20210503 to 20210725, The data shows that the daily business load of the site fluctuates randomly, and there are no obvious rules. It is not suitable to adopt energy-saving strategies, otherwise it will affect the user's perception.

(4) Energy saving benefit evaluation and analysis

According to the improved FP-Growth algorithm, it can quickly and effectively automatically identify the intelligent energy-saving measures of multi band stations, periodically select energy-saving strategies, and automatically start and stop energy-saving operations, as shown in Table 4.

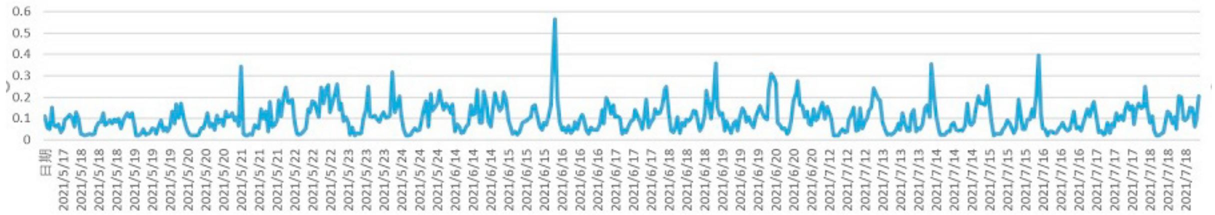

Fig. 3. hourly business load curve of a residential house in urban area.

The analysis results show that the valley effect energy-saving periods of the base station on rest days and working days are 0-7 o'clock, 19-23 o'clock (Monday to Friday), 0-23 o'clock (Saturday and Sunday), 7-9 o'clock and 14-16 o'clock (Monday to Friday). According to the energy-saving methods, two policies or even more policies can be adopted every week to open and close the appropriate energy-saving methods during the energy-saving period, The site energy consumption can be saved by up to 44%. The valley effect energy-saving period of the base station with consistent trend effect is 0-6, 19-23 (Monday to Sunday) and 0-23 (Monday to Sunday). According to the energy-saving method, the strategy of two weeks a week can be implemented, and the maximum energy consumption of the station can be saved by 26%. In addition, irregular base stations are prone to emergencies and random fluctuations. In order to ensure user perception, symbol energy saving can be adopted, rather than deep energy saving strategy.

Table 4. Partial association rules.

Features of scene	Week features obvious	Week feature consistent	Irregular of base station
Association results	Working days (Monday to Friday) and rest days (Saturday and Sunday) Rest day and workday effects	Monday to Sunday	-
Feature Identification	Rest day and workday effects	Consistent trend	No obvious rules
Energy saving period1	0-7am,19-23pm	0-6am,19-23pm	0-23pm
Energy saving period2	0-23	0-23	-
Energy saving period3	7-9am,14-16pm	-	-
Energy Saving Strategy	Two policies a week	One week policy	It is not suitable to implement deep energy-saving strategy
Saving Mode of Energy	(1) Symbol off (2) channel off (3) Carrier off (4) station off (5) Deep dormancy	(1) Symbol off (2) channel off (3) Carrier off (4) deep sleep	(1) Symbol off
Saving Mode of Space	44%	26%	3%

4.2. Sleep and wake-up of base station equipment based on user perception

For different network coverage scenarios, the impact of turning on and off 5G active antenna unit(AAU) energy-saving function on 4G network KPI and KQI. For the cell in Shijiazhuang that turns on AAU energy-saving function for one week, extract the KPI, KQI and complaint indicators at the cell level of the same 4G network. It is found that turning on and off 5G AAU energy-saving function has little impact on 4G network indicators and user perception.

Through the analysis, it is found that the downlink physical resource block utilization has obvious positive correlation with the number of RRC connected users, and has obvious inverse correlation with the downlink user experience. By analyzing the 4G network Deep Packet Inspection(DPI) data, combined with the DPI user type and perception index, and taking the user's video delay and Caton as the guarantee objectives, there is no difference in user experience when ZTE equipment user experience rate greater than or equal to 19mbps and Huawei equipment user experience rate not less than 22mbps.

In addition, the historical DPI data is used to establish the prediction model for 4G users and 5G users in the next 24 hours. Combined with the analysis of historical DPI data, when 4G/5G not less than 26%, wake up 4G AAU or 5G AAU.

5. Conclusion

Based on the site load characteristic index, this paper uses the improved FP-Growth algorithm and combined with the peak and valley characteristics of site load to effectively identify the site energy-saving scene. An example shows the reliability of the effect of association rules and the effectiveness of the energy-saving field identification algorithm, which greatly improves the energy-saving potential of equipment on the premise of ensuring that the KPI and KQI indexes of 4G/5G network of operators remain unchanged. It helps telecom operators to realize intelligent management of 4G/5G site energy consumption, optimize network resources, and improve operators' market competitiveness and sustainable development ability.

References

[1] Zhang Qing. (2019) "AI based 5G base station energy saving scene intelligent recognition." *5G Network Innovation Seminar thesis*.
[2] Jiang Guoqing, Pan Yong, and Hu Feiyue. (2015) "Energy saving method of home base station based on clustering algorithm." *Computer Application and Software* **32** (11): 103–106+175.
[3] Wang Xiaodi, Liu Junyong,and Liu Youbo. (2019) "Typical load curve morphological clustering algorithm using adaptive piecewise aggregation approximation." *Power System Automation* **43** (01): 110–121.
[4] Du Wenfeng, Liu Gangyi, and Gao sining. (2016) "Energy saving dynamic optimization deployment of green wireless networks based on artificial bee colony algorithm ", in Joanne Evans and Lester Hunt (eds) *Small Microcomput System* **37** (12): 2607–2611.
[5] Wang Min, Zhu Xijun. (2021) "Essociation rule analysis based on improved Apriori algorithm." *Computer science and application* **11** (6): 11.
[6] Wei Lihong, Chen Song.(2014) "Research on data mining technology and association rule mining algorithm." *Journal of Jiamusi Institute of education* **11** (1): 498.
[7] Xu huizeng.(2012) "Research on data mining methods of association rules." *Science, technology and engineering* **12** (01): 60–63.
[8] Wang Chaohui, Wang Tingting.(2007) "Research on data transformation and analysis in teaching evaluation information data mining." *Journal of Guizhou University of Technology (NATURAL SCIENCE EDITION)* **4** (36): 40–42.
[9] Guo Jia.(2012) "Application of data mining technology in enterprise customer relationship management (CRM)." *China business theory* **34** (03): 79–82.
[10] Kong Dejian.(2011) "Research on improving the efficiency of Apriori algorithm in association rule mining ." *China Science and technology information* **1** (23): 85.
[11] Dong Zhenhua, Li Xiyan, Zhang zhangbian.(2012) "Research on classical Apriori algorithm based on association rules ." *Science and technology information* **1** (06): 148–149.
[12] Han Jiawei, Pei Jian, Yin Yiwen.(2000) "Mining Fre-quent Patterns without Candidate Generation." *ACM SIGMOD Record* **29** (02): 1–12.
[13] Peng Hao, Dian Songyi.(2020) "Design and implementation of user interest recommendation algorithm based on improved FP-Growth." *Microelectronics and computer* **1** (2): 8–13.
[14] Yan Jie, Qi Wenjuan.(2020) "Research based on Aprior & FP-growth algorithm." *Computer system application* **22** (05): 122–125.
[15] Tian Qing, Liu Yongmei.(2013) "Application of FP growth algorithm in shopping basket analysis." *Journal of computer science and technology: Chinese and English edition* **1** (02): 24–30.