JURNAL ILMIAH SEMESTA TEKNIKA
Vol. 19, No. 2, 106-112, November 2016

Analysis of Cellular Traffic Homogeneity in A Specific Area using Statistical Test

INDAR SURAHMAT

ABSTRACT

Capacity and coverage are the most important issues on providing services by cellular operator companies. The companies measure the condition of their network and services continuously including data collection, monitoring, evaluation and future prediction. These are part of activities to maintain grade of service. In this paper, homogeneity analysis in cellular traffic data is analysed. This step is a part of evaluation of current situation which it will be used for planning purposes. It is done by processing data provided by network statistic monitoring tool. This paper limits on analysis in a particular area. The area is served by some cell surrounding the area. The result shows a good homogeneity amongst the cell covered a certain area during low traffic (1 am to 1 pm). Otherwise, in high traffic, the serving cells are relatively different in terms of traffic variation and homogeneity test using Fligner-Killeen is more suitable

Keywords: traffic, cellular, homogeneity, variance

INTRODUCTION

In the rapid development of telecommunication technology, some classical problems related to process of delivering services by cellular companies are still interesting to be investigated. The problems, which are familiar amongst the others related to network performance, are capacity and coverage. Both problems are concerned by some telecommunication operators especially cellular systems. Some approaches and methods are delivered to get the best output to overcome these problems.

Some papers introduced methods on how to make traffic predictions based on latest condition traffic. For example, Tran and Dziong in [1] proposed an estimation model based on current traffic statistics. The another work done by An et.al. in [2], reported a prediction model based on self-similar traffic. They investigated current traffic to make a model of near future trends. These methods were designed based on single stream of data which meant an approximation of traffic use only a series of single cell or site data and averaging of data on a cluster of network. The paper worked on cluster is presented by Ma et.al. in [3] which discussed four parameters of radio base stations.

This paper try to examine the possibility of using a different approach which traffic conditions in a certain area are treated by using measurement of serving cells. The term of serving cells in here are cells or sites which cover the same area. This approach is proposed based on a fact that the measurement of serving cells shows an almost similar pattern of traffic trend. Moreover, this paper deliver a homogeneity test, as a first step, to ensure that the traffic pattern of those serving cells is homogeneous. Ma in [3] also discussed homogeneity in a network cluster and define it as temporal homogeneity. This approach use approximation on homogeneity in hourly basis. This research use a difference approximation on showing a cluster homogeneity. Homogeneity tests chosen in this research are Levene's test, Bartlett's test and Fligner Killen Test. The purpose of the test is to measure the homogeneity of variance of traffic in a certain cluster.

The organization of the paper are described as following. Section II explains the approach of observation and the fact or the current measurement taken from network statistic of

cellular operator. Section III constructs three methods as theory of homogeneity tests. Section IV describes analysis of homogeneity test implementation in the data reported in section II. Finally, the last section is conclusion and future work.

PROPOSED MODEL AND MEASUREMENT

To examine homogeneity, we proposed an observation model and represented in a specific location. Then, we decided a list of surrounding cells which served the location. The decision is made based on information of antenna azimuth of single site configuration. Generally, the sites contain three sectors with antennas beamed to a specific azimuth. In the model, the approach only uses the data related to the sectors which serve to the observation area. Figure 1 shows the illustration in a map of the location as the observation area.

As seen in Figure 1, there are four cells served the observation area in the middle of cluster. The BTSs, which corresponds to the area, are BTS A, BTS B, BTS C, and BTS D. The traffic measured in those four cells in a specific week period can be seen in Figure 2.

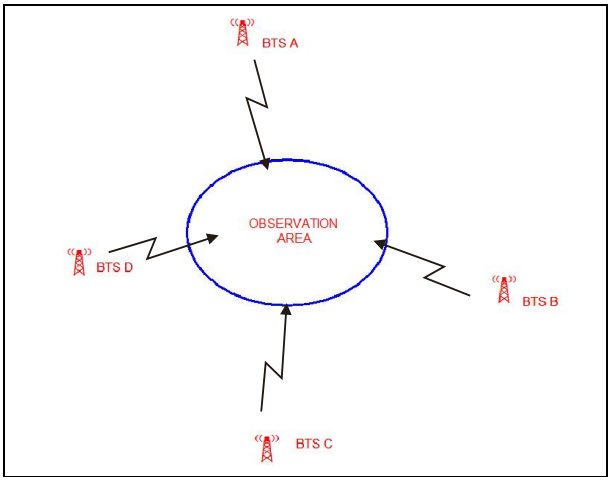

FIGURE 1. The observation model is represented by four serving cells

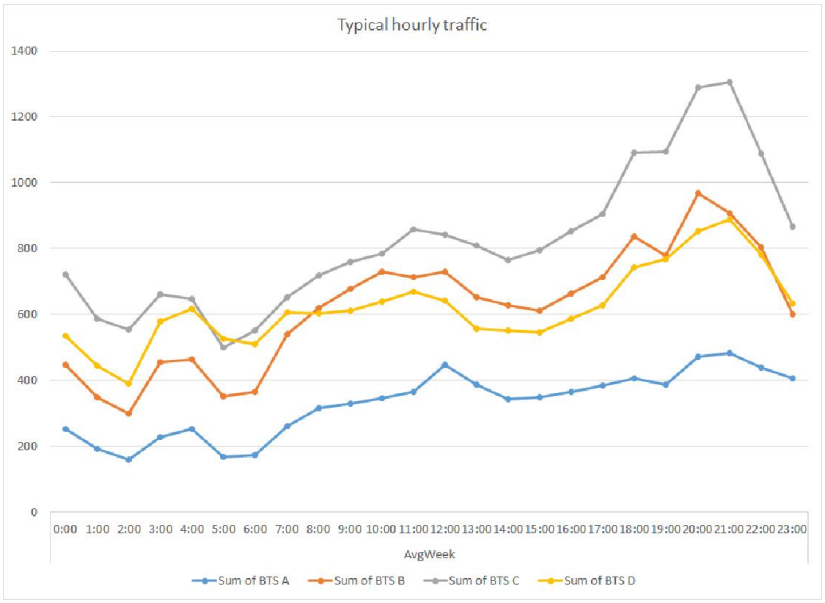

FIGURE 2. Weekly Traffic of Four BTSs in 24 Hours

HOMOGENEITY OF VARIANCE TEST CONCEPT

In statistical analysis, there are three methods commonly used for evaluating homogeneity of variance of data and all of them are explored in this paper. The first similarity between the three-method is in the first step. They use null and alternative hypotheses as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 \quad (1)$$

H_1 : other (not all variance has the equal value) (2)

where Ho is null hypotheses which represents that data being compared are equal or have no significant difference in variance, otherwise, H1 represents there is significant difference. σ^2 is variance of a single group and k is index of group (1, 2, ..., k).

The second similarity is on choosing the value of significance level. The most common significance level (α) used in statistical consideration is 0.05 or it means the error of hypotheses is below 5% based on normal distributions. If the result of p-value is greater than α, then the hypotheses is not significant or simply said that Ho cannot be rejected.

1. Levene's Test

First of all, Levene's test is method to measure homogeneity using one of three criteria as the following:

a. Use of mean, suitable for symmetric and moderate tailed distribution
b. Use of median, suitable for data under skewed distribution
c. Use of 10% trimmed mean, suitable for data with a heavy-tailed distribution

For simplification, this paper only investigates Levene's test using mean approximation. The step to find homogeneity is begun by calculating residual e_{ij},

$$e_{ij} = x_{ij} - \bar{x}_j \quad (3)$$

Then, a complete one-way analysis of variance (one-way ANOVA) is done to get the result. With modification of residual mean e_{ij}, total sum of square required by one-ANOVA can be defined :

$$TS = \sum_j \sum_i (e_{ij} - \bar{e})^2 \quad (4)$$

While, sum of square within groups is formulated as

$$WS = \sum_j \sum_i (e_{ij} - \bar{e}_j)^2 \quad (5)$$

And sum of square in terms of between groups is

$$BS = \sum_j n_j (\bar{e} - \bar{e}_j)^2 \quad (6)$$

Where n_j is the number of members of jth group. After these, degree of freedom criteria are defined as

$$dFT = n - 1,\ dFB = k - 1, \text{ and } dFW = n - k \quad (7)$$

Where dFT, dFB and dFW are degree of freedom of total, between groups, and within groups respectively. The parameter n is total elements in all groups and k is the number of groups.

Next steps, mean square of total, within groups, and between groups are calculated with

$$MST = \frac{TS}{dFT},\ MSB = \frac{BS}{dFB},\ MSW = \frac{WS}{dFW} \quad (8)$$

F-value is result from operation $F=MSB/MSW$. Finally, p-value is probability that a single observation, calculated from cumulative density function of F distribution with α, dFB and dFW as parameters, will happen at range 0 to α. The complete function is defined as follow:

$$p - value = F(\alpha \mid dFB, dFW)$$

$$= \int_0^{\alpha} \frac{\Gamma\left[\frac{dFB + dFW}{2}\right]}{\Gamma\left(\frac{dFB}{2}\right)\Gamma\left(\frac{dFW}{2}\right)} \left(\frac{dFB}{dFW}\right)^{\frac{dFB}{2}} \frac{t^{\frac{dFB-2}{2}}}{\left[1+\left(\frac{dFB}{dFW}\right)t\right]^{\frac{dFB+dFW}{2}}} dt \quad (9)$$

2. Bartlett's Test

Another parametric test to measure homogeneity is Bartlett's test. To analyse homogeneity using Bartlett's Test, the user need to define some parameters. First parameter, it is required to define chi-square

approximation to get the value of B; therefore, the value of B is determined as:

$$B = \frac{dFW \ln(WS) - \sum_j jdFj \ln s_j^2}{1 + \frac{1}{3(k-1)}\left(\sum_j \frac{1}{dFj} - \frac{1}{dFW}\right)} \quad (10)$$

Where dFj is degree of freedom of jth group and sj2 is variance . The p-value than calculated using chi-square test with parameter B and degree of freedom (dFW)

$$p - value = F(B \mid dFW) = \int_0^B \frac{t^{(dFW-2)/2} e^{-t/2}}{2^{dFW/2}\Gamma(dFW/2)} dt \quad (11)$$

3. *Fligner-Killeen Test*

A Fligner-Killeen Test is another homogeneity of variance test to measure two or more data have a significant of difference or not. In this test, the data, which are median-centering processed similar with Levene's Test, are ranked and normalized. Next, parameter of Fligner-Killeen is calculated as

$$FK = \frac{\sum_{j=1}^{k} n_j (\bar{c}_j - \bar{c})^2}{s^2} \quad (12)$$

Where is the mean of normalized value and the p-value is defined by using right tailed chi-square distribution.

Fligner -Killeen test is categorized as non-parametric statistical test which is suitable for data with non-normal distribution.

TABLE 1. Weekly Traffic of Four BTSs in 24 Hours

Week	TIME	BTS A	BTS B	BTS C	BTS D
AvgWeek	0:00	251.7	446.4	719.8	535
AvgWeek	1:00	191.8	348.6	586	443.7
AvgWeek	2:00	158.9	298.1	552.4	390
AvgWeek	3:00	226	454.1	661.3	577.1
AvgWeek	4:00	251.1	464.1	648.1	616.5
AvgWeek	5:00	166.6	349.6	499.1	525.1
AvgWeek	6:00	172.8	365.2	551.5	509
AvgWeek	7:00	259.8	540.9	653.8	604.8
AvgWeek	8:00	314.2	620.3	717.1	602.2
AvgWeek	9:00	329.6	677.8	759.9	610.2
AvgWeek	10:00	344.7	730.6	783.4	637.4
AvgWeek	11:00	364.8	712.2	857.8	668.2
AvgWeek	12:00	447.5	728.2	841	640.3
AvgWeek	13:00	386.7	652	807.7	555.3
AvgWeek	14:00	341.7	627	766.3	550
AvgWeek	15:00	347.9	612.2	796.1	546.2
AvgWeek	16:00	364	664.8	851.4	586
AvgWeek	17:00	384.7	711.7	905.2	627.4
AvgWeek	18:00	405.8	836.7	1092.2	742.8
AvgWeek	19:00	386.3	779.6	1095.2	767.7
AvgWeek	20:00	470.7	968	1289.5	853.2
AvgWeek	21:00	483.1	906.8	1305.1	887.2
AvgWeek	22:00	438.5	803.1	1087.7	780.6
AvgWeek	23:00	405.5	599.7	865.3	633.5

DATA ANALYSIS

The following are the process and the result of homogeneity test using three model described in the previous section. The tests are conducted to the data explored in section II. Table 1 shows the data extracted from network measurement which corresponds to figure 2. At the first stage, we construct the hypotheses; null hypothesis (*Ho*) is the data have same or no significance variance amongst the BTS's traffic and the other condition is alternative hypothesis (H¬1). This test use common value of significance level ($\alpha = 0.05$) as threshold to determine the significance of the data.

The tests are done on hourly basis. For example, table 2 contains raw values of four-cell traffic at 9 am and 3 pm. Next, for those hours, the calculation of homogeneity tests is conducted. At 9 am, all of the tests results p-value greater than $\alpha = 0.05$ which means that the null hypotheses cannot be rejected; there is no significant difference in the variance of four cells traffic. The planner can consider the data for planning or predicting future traffic by using the data at this hour.

For comparison, at 3 pm (table 4), not all the test results p-value higher than α. The p-value using Levene's test is only 0.02. If the network planners consider this test, they should be careful to forecast the following week by using the data. An exclusion of outliers, the numbers of traffic which are unusual resulting high variations of traffic, should be done. Another alternative, the result of Fligner-Killeen test is suitable for this case since the result of the test shows p-value higher than significance threshold.

The line graph in figure 3 shows the result of all methods of homogeneity test mentioned. The differences amongst the test results are the characteristics of each test which is not discussed in this paper. The most striking feature is both Levene's and Bartlett's show p-values under the threshold of hypotheses at 2pm to 9pm. The traffics has significant differences in variance amongst four cells at those hours. While, in the other hours, except at 1am and 2am, no significant difference amongst them or, in another word, the traffics are homogenous in variance.

Another interesting thing, Fligner-Killen test illustrates a little bit different meaning. The implementation of this method in the same data results the p-value higher than α at all hours except at 2 am. It means the data of four cells are homogenous at all hours over the day except at 2 am. In this case, the non-parametric test for evaluating homogeneity at 2 pm to 9 pm is more considerable. This result also explains that the condition of traffic at 2 pm to 9 pm is non-normal distribution in every single hour.

As aforementioned efidence, the parametric test (Levene and Bartlett) need some adjustment by selecting parts of the data which are far from the mean (average) or far from the median. Deleting or excluding them from the distribution may improve the homogeneity of the data.

TABLE 2. Hourly traffic at 9 am and 3 pm of four cells during one week

DATE	TIME	BTS A	BTS B	BTS C	BTS D	TIME	BTS A	BTS B	BTS C	BTS D
Day1	9:00	45	109	115.2	85.9	15:00	54.2	85.4	119	87.9
Day2	9:00	54	99.5	97.8	90.4	15:00	45.5	99.1	93.1	79.7
Day3	9:00	33.9	104.4	109.8	89.6	15:00	46.4	78.9	97	69.7
Day4	9:00	56.1	102.6	77.6	65.2	15:00	40.8	69.4	87.6	58.4
Day5	9:00	50.8	71	88.4	91.3	15:00	44.8	96.8	140.8	90.7
Day6	9:00	56.5	94.8	155.3	97.4	15:00	44.9	85.3	151.4	81.2
Day7	9:00	33.3	96.5	115.8	90.4	15:00	71.3	97.3	107.2	78.6

TABLE 3. Homogeneity of variance tests of traffic at 9 am

Levene's test					
Source of Variance	Sum of Square	Degree of Freedom	Means of Square	F value	p-value
Between Group	537.5241	3	179.1747	1.73355	0.186876
Within Group	2480.571	24	103.3571		
Total of Group	3018.095	27	111.7813		

Bartlett's test		
Degree of Freedom	B value	p-value
3	7.256974	0.064143

Fligner-Killeen test		
Degree of Freedom	FK value	p-value
3	4.949075	0.175564

TABLE 4. Homogeneity of variance tests of traffic at 3 pm

Levene's test					
Source of Variance	Sum of Square	Degree of Freedom	Means of Square	F value	p-value
Between Group	752.4059	3	250.80197	3.94523	0.020236
Within Group	1525.704	24	63.571		
Total of Group	2278.11	27	84.3744		

Bartlett's test		
Degree of Freedom	B value	p-value
3	6.837651	0.077256

Fligner-Killeen test		
Degree of Freedom	FK value	p-value
3	5.453431	0.141451

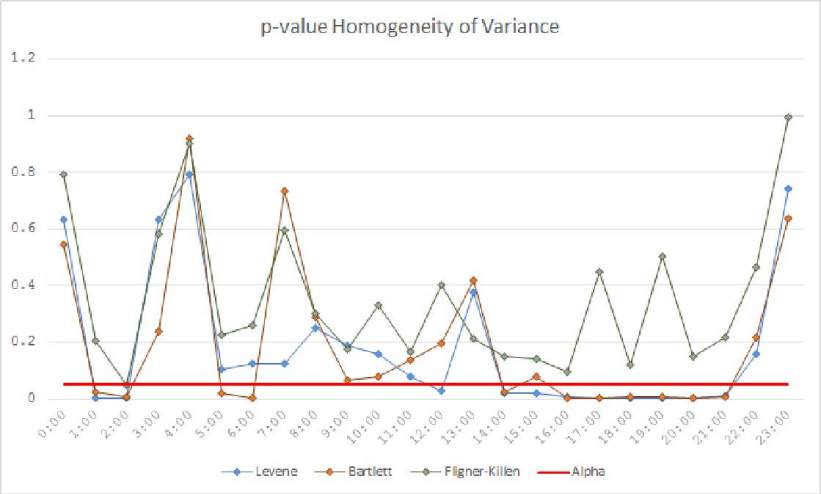

FIGURE 3. Homogeneity test result in hourly basis (p-value)

CONCLUSION AND FUTURE WORK

1. *Conclusion*

Regarding the test, the data extracted from network monitoring for the four-cell observed fulfill a homogenous assumption. The data of the cells can be considered to perform planning tasks. A very good homogeneity in almost all hours over the day promises an approach to evaluate capacity of the network based on serving cells in a certain area.

2. *Future work*

This paper is limited of one cluster in the discussion. Two or more clusters are needed to

investigate the difference between a certain cluster or area and another cluster. The works toward this requirement are being researched. Further, the forecasting considering this homogeneity approach is interesting to be explored more.

ACKNOWLEDGEMENT

These tests are performed to the real data extracted from network measurement of one of typical week of traffic. The identity of the four-cell has been modified to keep confidentiality. We would like to thank Electrical Engineering Department, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta for supporting tools which are used for processing the data.

REFERENCES

C. Zaiontz Real Statistics Using Excel. www.real-statistics.com, 2015.

J Ma, et.al., "Modelling Social Characteristics of Mobile Radio Networks", IEEE International Conference on Communication Workshop (ICCW), DOI: 10.1109/ICCW.2015.7247404, 2015.

Levene, Howard. "Robust tests for equality of variances". In Ingram Olkin; Harold Hotelling; et al. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press. pp. 278-292,1960

Snedecor, George W. and Cochran, William G. (1989), Statistical Methods, Eighth Edition, Iowa State University Press. ISBN 978-0-8138-1561-9

Tran and Z. Dziong, "Traffic Trend Estimation for Profit Oriented Capacity Adaptation in Service Overlay Networks", IEEE Transactions on Network and Service Management, Vol. 8, No. 4, December 2011.

X. An, L Qu and H. Yan, "A Study Based on Self-similar Network Traffic Model", Sixth International Conference on Intelligent Systems Design and Engineering Applications, IEEE DOI 10.1109/ISDEA.2015.28, 2015

https://www.mathworks.com/help/search.html?submitsearch=&qdoc=p-value.

AUTHOR:

Indar Surahmat

Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, South Ringroad, Tamantirto, Kasihan, Bantul, Yogyakarta, 55163.

Email: indar.surahmat@umy.ac.id