ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
September 2013,
Vol. 6, No. (3):
Pgs. 351-355

Analytical Study for Optimal Time Parameter in Cellular Communication Data analysis

WUROOD H. ALBAYATI and TAWFIQ A. AL-ASADI

Al-Mustansiriyah University, Iraq.
Babylon University, Iraq.

(Received: August 16, 2013; Accepted: August 30, 2013)

ABSTRACT

In this research we will try to determine time quantization periods which could be used for cellular data analysis to reduce cost and time in processing such huge data. The research try to answer such questions as, What is the excite time quantization value required to perform network and user analysis? It is possible to eliminate any periods of time without effecting the quality of obtained result? Does it worth to eliminate such data(what is the gain for such elimination)?

Key words: Cellular network, Time quantization, Cellular data complexity.

INTRODUCTION

Cellular communication is a growing filed, that spread fast, with respect to network evolving, result in huge heterogeneous data which are time variant, and dynamically change. Cellular communication company stores user and network status data for months before deleting, thus required a huge amount of storage capabilities devices. Accordingly if it is possible to reduce amount of data to be stored , in other word, could we partially store it (save only part of data) ? If so what part should be delete and what should remain? in our analytical study we will clarify such concepts.

In this research real cellular data were used to analyzed the effect of different time quantization periods, the data were taken from MDC project for Nokia mobile company [NDC] , Cellular database consist of three main type of data

- User data
- Network data
- Environment data

Every minutes cellular network generates all those type of data which transfer among user devices, base station and/or switching center, imagine the amount of data generated for one user during 24 hour/day (3600 second)?

The main idea of this research is about time, in cellular communication time is an important factor, thus many researcher used it in different way, Nathan et al cluster user according to the time of silent for communication tower [Nathan Eagle, et al], while Orlik[Orlik P.V] used time for modeling user mobility and studying teletraffic performance , Tao et al ,analyze the challenges of the interference environment in hybrid network (D2D and cellular connections) they use time hopping based radio resource allocation schema for improve the robustness of hybrid network [Tao Chen *et al*] . In this research we concentrate on time too, but in different way. By observing real cellular communication data , from nokia research group of mobile data challenges project[Juha K. Laurila] , we notice as researchers it is so difficult to analyze those huge data, to reduce cost and computation time , we should found a tradeoff between data size and quality of information obtained, in other word, how to reduce size of data without effecting the quality.

In this research we show how could we facilitate dealing with such data by eliminating some, and select time quantization periods that reflect as many information as possible. First step in analyzing time variant data (for the purpose of prediction or for statistical overview), is the elimination (or discarding) of time periods with no or little user action, if this is done then the amount of data will be reduced, the storage required to deal with will reduced too. Secondly, we could summarize user and network action over little period of time (quantization period) thus result in huge reduction in data but the problem is, how long quantization period should be? Since, user and network behavior change with time, the selection of time depend on the type of application (if you want to study only the network and/ or user behavior.

The paper arrange as follows: in section 2 we analyzed cellular data, practical result and evaluation presented in section 3, finally conclusion is given in section 4.

Analyzing the data

Communication data is a huge databases , BI use such data to generate business application for marketing, such data should be collected then analyzed taking into account that each second reflects, many information about network and customer behavior ,ata attribute used in here consist of : Cell ID, User ID code, Call time, duration, Date, Status, GPS coordinates.

In our schema we want to use both user and network accordingly one second is good to study network analysis but it generate more data , it is not so useful for studying user behavior , also one minute is great for network and user but still generate large data set , 5 minute is good for both ,15 minutes acceptable too, but 2,3,4,5,6 hour is not accurate for analyze network (and /or user behaviour). Thus we should make a trade off between the application domain and amount of space required, in studying user behavior , user could transfer from cell to cell every minute , while using five minutes is acceptable to observe network parameter.

Practical result

During our usage of cellular communication data we found that: those huge data could be analyze with reasonable resource taking into account a perfect quantization periods as well as reduced storage and computation complexity by eliminating data that reflects little information about user and network

In analyzing user behavior, an active user could be in more than one cell at a minute, from our database we notice that an active user use (in middle case) three cell in a minute during usage of cellular network . thus to study the action of such user taking an hour or more is unacceptable for identifying user pattern or behavior, now, let as now calculate processing time (complexity) required to deal with such data , the selection of the time periods (quantization period) is accomplish with summarization process (that include sum all network data and user action during that period) the time required to do so measured in second , table (1) shows the result of analyzing the information of one user (User Id=05) clarifying time periods, processing time and the effect of quantization periods on file size, (assuming to do that in ordinary computer with 3 core to simulate natural situation in communication company where

not all employing super computers or advance technology), we show (in table 1and 2) the result of applying different periods of time to two user ((user ID= 5(in table(1)) and user ID= 63in table(2)), the tables display the results when taking time periods of one minute(t_q=0.01),five minute (t_q=0.05),quarter ((t_q=0.15),half hour ((t_q=0.30), hour(t_q=1),hour and half(t_q=1.30),two hour(t_q=2),two hour and half(t_q=2.30),three hour (t_q=3) and three hour and five(t_q=3.30).

For each period ,a summarization process take place, which summarize ,user action ,cell ,location and network data, calculating the time for such summarization(to give an indication of the time required when do some calculation on such data) and the resulted file size, notice that only important data of the user and network where included on the original file which mean the original file we used on that paper is a filtered version of the original data taking from (MDC) that include user identification , location ,action(call miss call ,message),network cell , network time, and few other network parameters.

From table 1,2 it is obviously that increase time periods result in decreasing computation time as well as file size, taking t_q=0.15 or t_q=30 reflect

Table 1: Shows time quantization periods and processing time

Quantized time	Time required to process data measure in (second)	No of cell	File_Size MB (original size=3.8)
Tq=0.01	1813.475974	2	3.5
Tq=0.05	350.088589	5	0.9
Tq=0.15	118.237314	13	0.4
Tq=0.30	58.152568	21	0.3
Tq=1	29.653158	25	0.3
Tq=1.30	20.159814	25	0.2
Tq=2	15.298435	25	0.2
Tq=2.30	12.530809	31	0.2
T_q=3	10.591135	31	0.2
T_q=3.30	9.271826	33	0.2

Table 2: Shows time quantization periods and processing time

Quantized time	Time to process (second)	Max. No.Cell	File_SizeMB (4.94 original size)
T_q=0.01	41497.472292	2	4.6
T_q=0.05	1297.031279	5	1.9
T_q=0.15	450.130855	13	1.5
T_q=0.30	226.430932	25	1.3
T_q=1	120.404673	29	1.5
T_q=1.30	82.209793	39	1.3
T_q=2	65.803950	47	1.3
T_q=2.30	54.928002	39	1.3
T_q=3	46.912186	55	1.3
T_q=3.30	41.634105	47	1.3

acceptable computation time and size, from other hand , when studying network behavior, the effect of time on network representing in cell, where at each minute user could hop from cell to other , figure (1) represent maximum number of cell could our users used at each time (the same time period of table(1 & 2), from that figure it is easily to notice that the increasing in maximum number of cell increase slowly at $t_{q=}$0.30 while it increase rapidly before that ,Thus we could exploit this , to judge on time periods required for sampling(or quantized) time in cellular communication,

Now, as we stated earlier communication company stores cellular data(call detail recorder and network data) for months before discard it, storing such data is costly for those company, what if we could store only important data ,in other word

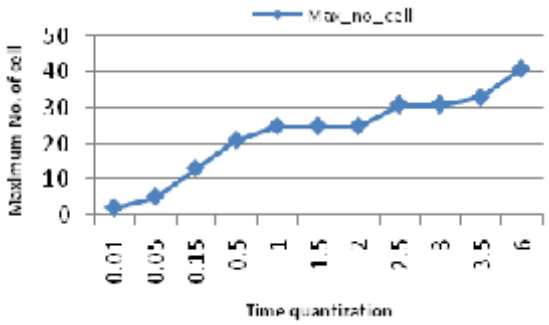

Fig. 1: Shows the effect of time quantization on cellular cell (for 20 user)

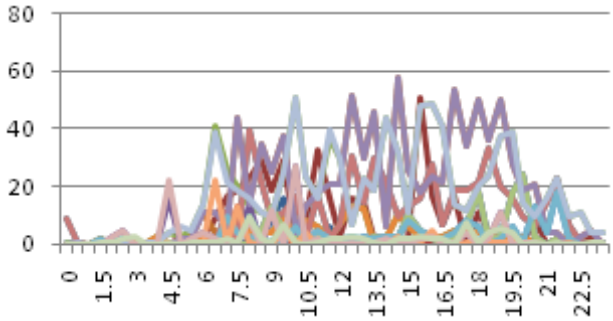

Fig. 2: Relation between user action and time (along 24 hour/day for 100 day)

discard some information (which is not important to take decision),this is possible , analyzing 20 user action for more than 3 month(100 days)indicated that there is a time slot that most user did not have any action (or minor action) and network data are constant during those hour, divide 24 hour a day into half's,(which mean we have 48 half or value) starting at 0 (which mean 24 oclock, if our timming based on 24 hour) ending with 23.5,

Figure (2) shows the effect of time quantization value of 30 second on user where data of 20 user have been analyzed according to their action (network usage), (y-axis represent user action ,x-axis represent time) the observation was taken over 100 day/24 h, It is so easily to say that there is a slot of time seems to have very little action ,that time could be eliminated from the data and never save, cause it does not reflect too many information.

CONCLUSION

There is many unsolved challenges in cellular communication which need researchers effort , in order for researcher to start their search in that area, they definitely need to reduce data size before they could continuing their work , in real cellular communication data ,communication company record every second of cellular network action , those data is about network and user call information , it is huge and confusing data, the main corner stone are from where to start, by deciding the accurate time quantization periods which ensure reduction of data size and reflect as much information as possible we could facilitate the researcher work .

From table (1&2) and figure (1) we see that taking time quantization periods to be T=0.30 (or half an hour) is a good choice for reduce data, which mean instead of dealing with 3600 *24 =86400 second a day (with each second have its data record) we could deal with 1800 *24 =43200 second (or 43200 record) a day (summarizing the information related to that half hour).

Also , eliminating times that represent little information have its effect in reducing cost for communication company where they could save less data instead of saving 24 hour a day it could be seen from figure (2) that time 0 ,1 2 and time 23 have no or very little user action that mean we can eliminate those 4hour and saving about (20*1800=36000) second (or record), thus we save about 7200 record each day of network usage for each user. This is a very useful reduction of communication data without effecting the accuracy of such data.

REFERENCES

1. [Juha K. Laurila et al], Juha K. Laurila, Daniel Gatica-Perez, Imadaad, Jan Blom, Olivier Bornet, Trinh- Minh-Tri Do, Olivier Dousse, JulienEberle, Markus Miettinen, “The Mobile DataChallenge: Big Data for Mobile Computing Research” ,2012, Nokia research centre: http://research.nokia.com/files/public/MDC2012_Overview_LaurilaGaticaPerezEtAl.pdf
2. [NDC] http://research.nokia.com/about_nrc
3. [Nathan et al] Nathan Eagle, et al, “methodologies for continuous cellular Tower Data Analysis “,7th international conference pervasive ,Nara, Japan, may 11-14,2009 ,pp342-353.
4. [Orlik P.V] Orlik, P.V., Rappaport Stephen, S., “A model for teletraffic performance and channel holding time characterization in wireless cellular communication with general session and dwell time distributions “, IEEE journal on Selected Areas in Communications, Volume:16 , Issue: 5, Aug.2002,pp 788 – 803
5. [Tao Chen et al] Tao Chen ; Charbit, G. ; Hakola, S.,” Time Hopping for Device-To-Device Communication in LTE Cellular System”,IEEE Wireless Communications and Networking Conference (WCNC), 2010 ,pp1-6.