INTERNATIONAL JOURNAL OF ELECTRONICS; MECHANICAL and MECHATRONICS ENGINEERING Vol.2 Num.2 pp.(162-165)

CLUSTERING GSM INTERNET USAGE WITH SELF ORGANIZING MAPS

Ulas CELENK[1] Osman Nuri UCAN[2]

[1]*Department of Electric Electronic Engineering*
Istanbul University, Avcilar, Istanbul , Turkey
ulascelenk@hotmail.com

[2]*Department of Electric Electronic Engineering*
Istanbul Aydin University, Kucukcekmece, Istanbul, Turkey
uosman@aydin.edu.tr

Abstract:*In the recent years, the mobile internet usage has been common in daily life by the development of GSM technologies. The internet provides a very easy and flexible connection everywhere, when the users need to use internet. It supports messaging, media, finance and lots of services throughout its connection. Data can be transferred by constant quota values of internet users. For all new quota requests, charging system makes queries. There is some kind of performance problems occurs when the applications of quota is adjusted due to low level users. If the quotas are planned to provide internet services regarding intensive level internet users, then the low level users become aggrieved. The aim of this study is to set up dynamic quota values according to users behaviors and thus ensure high performance income using most used artificial neural network in grouping studies as SOM (Self-Organizing Maps)*

Key Words: *GSM Charging, Chapman, SOM Neural Networks, Cluster Analysis*

1. INTRODUCTION

Voice and data calls in mobile Networks start with access of incoming calls from the mobile devices to Base Transceiver Station (BTS). Then this incoming request is transferred to Base Station Controller (BSC) and its transmission to GSM network is provided via forwarding this call to Mobile Switching Center (MSC). If the call is not a voice call, an IP address is provided to subscriber from Gateway Serving GPRS Support Node unit, which is at the main network. After then subscriber is routed to Radius(Remote Authentication Dial-in User Service) and subscriber is checked if its connection is right

the subscriber continues to use the internet, all

or not. After authorization requirement is taken, internet Access is started together with a protocol that processes as a counter with a start of a session named as Pocket Data Protocol. Related counter access user's accounts over Home Location Register (HLR) or Visited Location Register (VLR), after all, the system blocks a particular quota. Meanwhile a call record (CDR-Call Detail Record) is generated. As the subscriber surfing and creating data traffic on the internet, the counter starts to deduct the used part form the quota. If the quota blocked on the counter is consumed, the counter communicates with HLR and confirms that the quoted data is fully consumed; a charge set CDR is terminated. In case where

define constant quota value for the groups with the processes that described above are repeated again and a new quota is blocked. By this way, in setting charge systems; CDR's are created, the steps explained above are repeated in the process of GSM subscriber's internet. Access is summarized in Fig. 1,

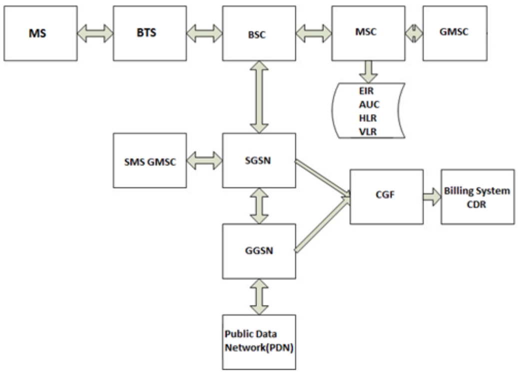

Figure 1 Structure of Mobile Networks

A charging process for created CDR, corresponding to a quota as it is explained in this section, takes place in billing the GSM subscriber's internet Access. Processing the billing according to each created CDR having a standard quota for each user, causes a processing load because of putting the all user to same filtration processing. Thus lots of processing loads and ten thousands CDRs, in one minute are created for users with different user habits. Nowadays, in order to increase performance of the billing systems and determine reserved quotas to the subscribers dynamically, alternative billing systems is an important working field for the software programmers that develop billing, charging of mobile operator systems. Charging and billing are one of the main topics among the most resource consuming systems at the GSM operators which charge over subscribes' tariff information via monitoring voice and data traffic of the subscribers. In classical charging systems, a standard quota is defined for all the subscribers' CDRs and charging process are done over this quota. But realization of performance increment on CDRs which are created as tens of thousands in one second will provide effective and productive working chance and this will an important cost advantage. In our model, instead of defining quota brackets separately for each user, we similar properties. Prediction over millions of users will cause high time consuming burden to the system. Alternatively, here we specify thousands of users that are grouped by aggregation of the users with the same habits and then we predict statistical properties of these groups which provides convenience. The method that is considered for the grouping is SOM (Self-Organizing Maps) type artificial neural networks.

2. SOM NEURAL NETWORKS

Instead of classical statistical methods, artificial neural networks can be used in grouping studies. Artificial Neural Networks (ANN) do not need the distribution assumptions for datum. Having a large quantity of elements and variables in a data set does not raise difficulties for neural networks

The most used artificial neural network in grouping studies is SOM (Self-Organizing Maps) [1]. This network is developed by Teuvo Kohonen in 1982. For this reason they are also known as Kohenen SOM Networks. SOM networks can process for both K-Averaging and multiple dimensional scaling methods. In other words, it makes both grouping and mapping in the data label. Thus these networks become very popular in recent years [2].SOM Networks are one layered Networks and they consist input and output neurons. Number of variables in the data set determines the number of input neurons. Each output neurons represents one group. A SOM network is shown in Fig.2. As aparting from other artificial neural networks, the positioning of neurons at the output layer is very important. This positioning can be linear, rectangular and hexagonal or cube shape. Mostly rectangular and hexagonal positioning are preferred. In practice, rectangular positioning is applied as quadratic. This positioning is important in terms of topologic neighborhood. Reference vectors (code-book-vectors) show the connection between input neurons and each output neurons. It is conceivable to think these vectors as columns of coefficients matrixes. This topologic neighborhood is used renewing the reference vectors while SOM neural networks is trained.

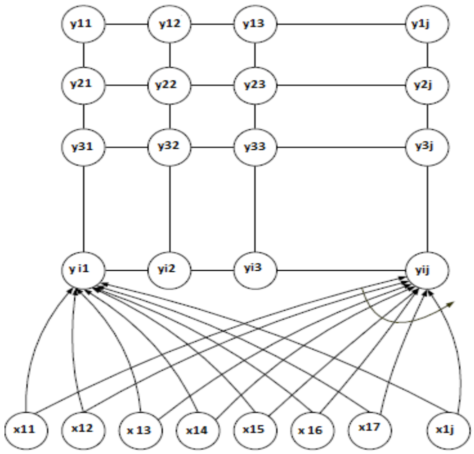

Figure 2 Kohonen SOM Neural Network

3. SOM LEARNING ALGORITHM

Kohonen Networks use SOM algorithm which also gives the name to these Networks. The learning algorithm used in these Networks is uncontrolled. As input vectors in the data set enter the network, the network arranges itself and form reference vectors. This algorithm is given at below [3].

3.1. SOM Algorithm

1) Assign initial value to w_{ij} coefficients. Determine topologic neighborhood parameters. Adjust learning coefficient parameters.

2) While finish condition is wrong, follow steps 3-9

3) For each x input vector follow steps 4-6

4) Calculate Euclidean distance D(i,j)= $\sum_{ij}\left(w_{ij}-x\right)^2$ for each i,j

5) Find value of i,j where D(i,j) is minimum

6) For all the output neurons in the defined neighborhood of I,J

$$w_{ij} \text{ (new)= } w_{ij} \text{ (old)+ a (x- } w_{ij} \text{ (old))}$$

7) Update learning coefficient

8) Decrease topologic neighborhood parameter at specified times.

9) Control the finish condition

The symbols that are used in this algorithm as;

w_{ij} : the reference vector that belongs to output neuron which is at i. row j. Column

x: input vector

D (i,j) : square of Euclidean distance of x vector to the output neuron which at (i,j) coordinates.

i,j. : the coordinates of output neuron which is closest output neuron to the x vector.

a :learning coefficient

As it is understood from the algorithm above, firstly, an initial value is given to the reference vectors. Before starting to loop, a high value is assigned to learning coefficient (a) and neighborhood variable (R). A value between 0 and 1 is assigned to a. It is preferred to have this value closer to 1. R variable starts with the value which is larger than height or width of positioning of output layer. One loop for algorithm is submission of all the rows in the data set to SOM network input. One of the row of data set is x vector. The square of Euclidean distance of x vector to each neuron at the output layer is being found. Each neuron at the output layer is represented by a reference vector (w_{ij}). Hence, this distance is the distance between x vector and wij. The smallest value is being found among the calculated distances. Whichever output neuron has this smallest distance is the Winner neuron. In other words, SOM Networks are ‘‘competitor’’ networks. The reference vectors of winner neuron and neighborhood neurons are being calculated again. The linear neighborhood of the winner neuron seemed in Fig.3, and rectangular neighborhood winner neuron seemed in Fig.4. As it seems from these figures, there are more neighborhood neurons around winner neuron in rectangular neighborhood. w_{ij} (new)= w_{ij} (old)+ a(x- w_{ij} (old)) equation is used in this calculation. For this reason if a small values are given to references vectors as initial values, the value of a must be taken as closer to 1. In this way reference vectors have a chance to generate themselves. By this way, a loop will be completed when these process are completed for each row in the data set. Reference vectors keep changing as long as loops continue. A and R values decreases in particular periods of loop. There is no particular rule that determines that the values decreased in how many loops. There are different opinions for this subject. Mostly it is adequate to decrease it with a linear function. Looping finishes when the changing in the reference vectors finish.

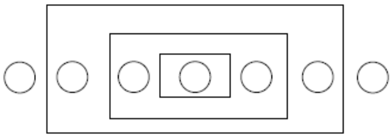

Figure 3 Linear Neighborhood of Winner Neuron (#) (in the order of inside to outside R=0, R=1, and R=2).

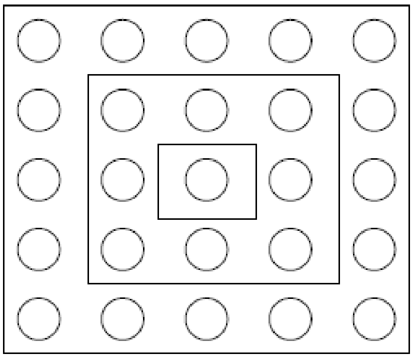

Figure 4 Rectangular Neighborhood of Winner Neuron (#) (in the order of inside to outside R=0, R=1, and R=2)

After completing the training of the network and forming the reference vectors, the elements that are in the data set are grouped together. All the rows in the data set are entered to network in a row. Entry vector is multiplied by reference vectors of output neurons. The element will be belonged to the group which the result is bigger. At the end of this process, elements are both grouped and placed in a two dimensional map. It is possible to see the elements that are close to and remote to each other from this map. If the positioning on the output layer 3 dimensional, the map will be 3 dimensional also. These maps can be colored and shadowed in different patterns according to features of the groups. Therefore, a more visual map can be gathered.

4. CONCLUSION

In this paper ,we propose a new approach to reduce the intensity of GSM systems and not to leave the users grievance. For this aim, we use different pricing algorithms and quota systems for different groups. The users are grouped according to their current balances and internet usage history. SOM-type neural networks are preferred for grouping. In this way, using high-dimensional inputs may be organized low-dimensional output. Usage of the internet varies according to circumstances and time. In this study; the neighborhood relationship between the groups in SOM-type neural networks are crucial data. The distances of the users for their own groups and other groups must be calculated. Future usage of internet is estimated by Markov processes [4-6]. The probability of switching from one group to another group in the future is estimated using Chapman Kolmogorov equations. In this step; the groups which have calculated by SOM-type neural networks and the distance of these groups will be studied. After estimation of the user's group, the most suitable quota value can be given. In this way, optimal adjustment of the quota values and maximum efficiency for signalization will be obtained.

5. REFERENCES

[1] Teuvo KOHONEN, *Self-Organizing Maps* ,Springer Series in Information Sciences, 2001

[2] Hudaverdi BIRCAN,Metin ZONTUL,Ahmet Gurkan YUKSEK,A study of clustering exporting countries of Turkey using SOM Neural Networks, *Ataturk University Journal of Economics and Administrative Sciences*,Vol 20,No.2 ,2006, pp.219-238

[3] Laurene FAUSETT,*Fundamentals of Neural Networks,* NJ:Prentice Hall,1994

[4] A. Hamdy TAHA, *Operation Research,* NJ:Prentice Hall,1992

[5] Howard ANTON, Chris RORRES, *Elementary Linear Algebra Applications versions,* John Wiley & Sons Inc Canada,1994

[6] Nursel RUZGAR,Analysis of a company balance of payments with markov processes *Dokuz Eylul University Journal of Graduate school of social sciences,* Vol 5, No 1, 2003, pp164-179.