RESEARCH AND DEVELOPMENT OF TRAFFIC PREDICTION ALGORITHMS ON CELL LEVEL IN THE WIDE BAND CELLULAR NETWORKS

Oleg Barladean, Nicolae Secrieru, Alexandru Ghincul, Olga Ghincul
Technical University of Moldova
nsecrieru@gmail.com aghincul@hotmail.com

Abstract. *The paper performs research and development of more efective data traffic prediction algorithms in mobile networks using data clustering on the cell level. The developed algorithms were adapted to a specific problem, namely to data with shifted trend. For artificially generated data, that imit real network data, was applied Gausian mixture algorithm to separate the current trend, after last shifting, needed for a more accuare prediction with big timestamps. There was also estimated the potential impact of prediction error if not applying clustering methods.*

Keywords: *prediction algorithms, clustering, shifted trend, data traffic, mobile networks*

I. Introduction

Global mobile data traffic will increase 26-fold between 2010 and 2015 according to Cisco research; Central and Eastern Europe, where Republic of Moldova is situated in, will have mobile data traffic growth at 102 percent. So the next few years will be critical for operators and service providers to plan future network deployments that will create an adaptable platform upon which will deploy the multitude of mobile enabled devices and applications of the future [1]. That is why it is necessary to have a tool which will provide a forecast for network dynamic optimization.

One of the key problems of forecast in cellular networks is the prediction of traffic for each cell separately. This issue is not simple to solve because the network is suffering continuous changes as a result of optimization, the volume of data traffic served by a cell is varying in correlation to the network dynamics respectively and the sharp shifting of traffic trend appeares.

In order to perform more precisely a forecast it is necessary to include in calculations cumulative data only after shifting.

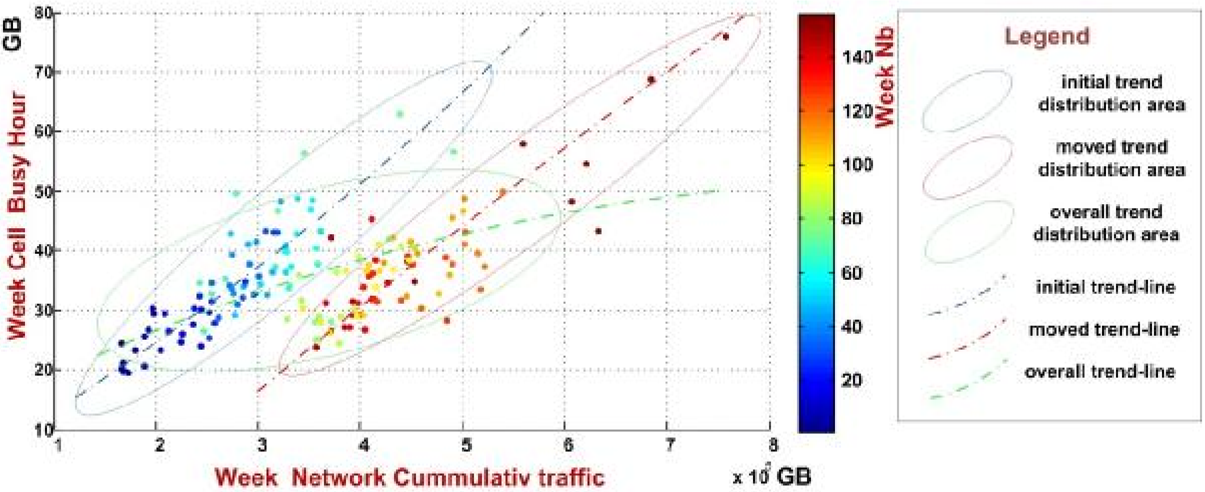

Figure 1. Cell busy hour traffic vs. weekly cummulative traffic scattering for a cell with shifted traffic

Visually it is easy to delimitate this moment (figure 1) but in the case of a big network and the number of cells with sharp shifting is big it appears the necessity to automate this process. In case this delimitation would not be performed one would obtain a wrong forecast that will lead to wrong optimization actions [3].

In order to delimitate the shifting of traffic tred it is proposed to use Gaussian mixture model algorithm of data clustering.

II. The aplication of the Gaussian mixture algorithm for cell traffic prediction

Gaussian mixture models are formed by combining multivariate normal density components. Clusters are assigned by selecting the component that maximizes the posterior probability. Gaussian mixture modeling uses an iterative algorithm that converges to a local optimum. Gaussian mixture modeling may be more appropriate than other clustering when clusters have different sizes and correlation within them.

For convenience data vectors WCUM (weekly cummulative traffic) and CBH (cell busy hour) where merged in anNx2 matrix, where N is the number of values of the vector.

The Gaussian mixture architecture estimates probability density functions (PDF) for each class, and then performs classification based on Bayes' rule [2].

$$P(C_i \mid X) = P(X \mid C_i) \cdot \frac{P(C_i)}{P(X)} \tag{1}$$

where: P(X | C_i) is the PDF of class j, evaluated at X; P(C_j) is the prior probability for class j; P(X) is the overall PDF, evaluated at X.

Unlike the unimodal Gaussian architecture, which assumes P(X | Cj) to be in the form of a Gaussian, the Gaussian mixture model estimates P(X | Cj) as a weighted average of multiple Gaussians [2].

$$P(X \mid C_j) = \sum_{k=1}^{Nc} w_k G_k \tag{2}$$

where w_k is the weight of the k-th Gaussian G_k and the weights sum to one. One such PDF model is produced for each class.

Each Gaussian component is defined as:

$$G_k = \frac{1}{(2p)^{n/2} |V_k|^{1/2}} \cdot e^{[-1/2(X-M_k)^T V_k^{-1}(X-M_k)]} \tag{3}$$

where M_k is the mean of the Gaussian; V_k is the covariance matrix of the Gaussian.

Free parameters of the Gaussian mixture model consist of the means and covariance matrices of the Gaussian components and the weights indicating the contribution of each Gaussian to the approximation of P(X | C_j) [3].

$$P(C_j \mid X) = P(X \mid C_j) \cdot \frac{P(C_j)}{P(X)} \tag{4}$$

where

$$P(X \mid C_j) = \sum_{k=1}^{Nc} w_k G_k \tag{5}$$

$$G_k \equiv p(X \mid G_i) = \frac{1}{(2p)^{d/2} |V_i|^{1/2}} \cdot e^{[-1/2(X-m_i)^T Vi^{-1}(X-m_i)]} \tag{6}$$

It is used EM (estimate-maximize) algorithm for variables μ_i, V_i, w_k to approximate them.

These parameters are tuned using a complex iterative procedure called the estimate-maximize (EM) algorithm, that aims at maximizing the likelihood of the training set generated by the estimated PDF.

Considering that the information about the data traffic of mobile companies is strictly confidential it was needed to generate initial data. There was generated the traffic of over 200 cells for a period of 2 years with the resolution of one hour. There were established three groups of traffic distribution per cell that have the role to imitate the traffic distribution according to activity zones of the population. There was also generated data that emit the specific trend for different scenarios.

The generated data was used as object of research for the prediction method.

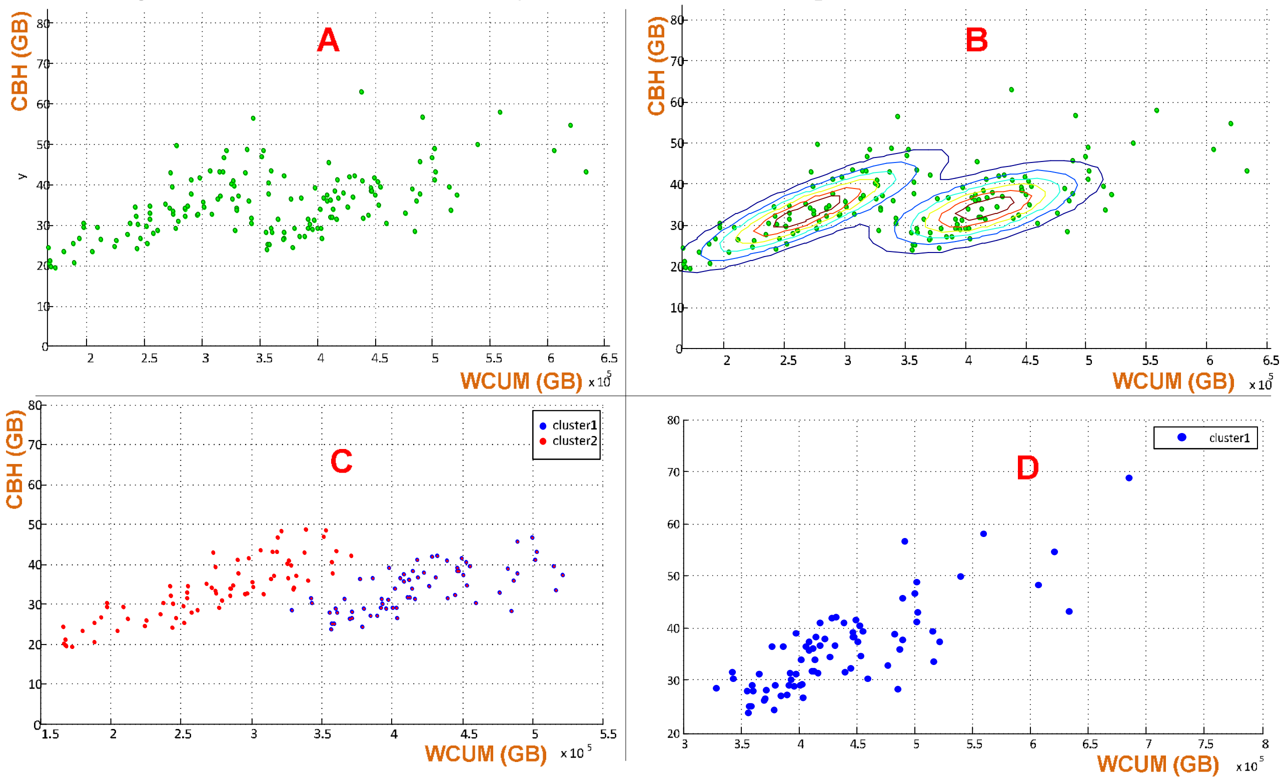

Figure 2. The results of algorithm application at different levels. A- scater ofWCUM and CBH, B – the obtaining of magnitudinal centers of the scatter, C – cluster division, D - extraction of the current cluster.

Figure 3.11 A presents the scatter of WCUM / CBH. It is clearly observed that there are two subsets of points that confirmed that the traffic was shifted. In 2 B are presented the areas of points around magnitudinal centers of the subsets. One can observe that these round areas delimit the two subsets quite accurately. For this case it was obtained that "mu" (magnitudinal unit) have the following coordinates: mu1=[429515.22 , 35.90]; mu2=[274073.97 , 34.39].

These values also match with those presented in 2 B. Is it can be observed the clusters are nominalized from left to right. Figure 2 C shows in what way the values where attributed to those two clusters. Cluster 2 presented with red points included 51% from the total number, but cluster 2 presented with blue pointsincluded 49% respectively. In figure 2 D is represented the actual cluster extracted from the total values set. In a practical way it presents the main goal of this paper.

Estimation of the impact of traffic forecast of a cell without using data clustering

In order to estimate the impact of traffic forecast in the case when data are not deided in clusters it is ptoposed to perform a forecast on Busy Hour (BH) for the same cell for 2 cases:

forecast1 – using th cluster of actual data, that is cell extract (Figure 3D), forecast2 – using all the data; Both models would give a theoretical prediction of the maximal awaited traffic „**fBH"** (future Bussy Hour), in relation to the total traffic of the network – WCUM=8*10^5GB.

In Figure 3 A is presented the graph that contains the theoretical curve of CBH in function of WCUM as well as the confidence interval for cluster 1 and in figure 3 B is represented the graph that contains the theoretical curve of CBH in the function of WCUM as well as the interval of confidence for all the data.

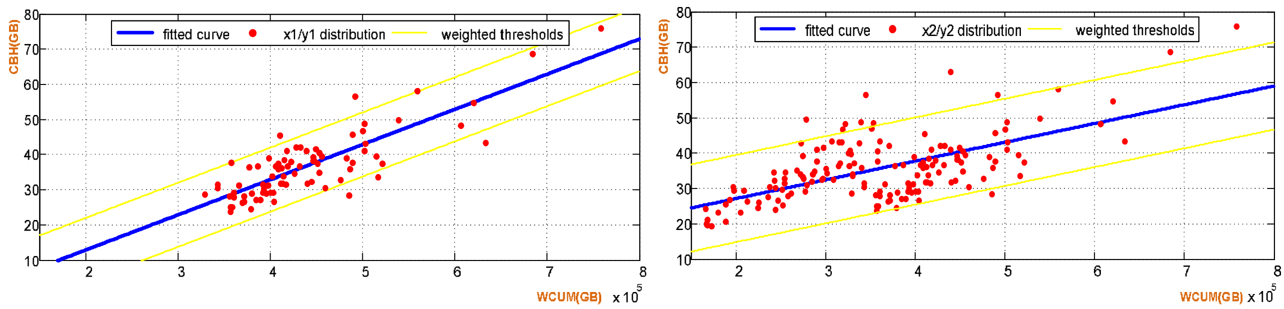

Figure 3. A - grafical representation of the forecast using the extracted cluster, B - grafical representation of the forecast without data clustering

If analyzing the obtained results it could be observed that the model that corresponds to data from cluster 1, figure 3 A is more accurate, has a confidence interval of about ±9 GB in comparison with the fitted curve, and the weighted average traffic of CBH forWCUM=8·10^5GB constitutes ~ 71 GB. In figure 3 B is presented the result of the forecast for the whole set of data, which has the confidence interval of ± 13 GB in comparison with the fitted curve, but the average trafficof CBH waited forWCUM=8·10^5GB constitutes approximately ~ 60 GB.

III. Conclusion

The Gaussian mixture algorithm was applied in order to perform data clustering for cells with shifted trend. As a result data was accurately divided into two subsets and the cluster with actual data was extracted. For the given cluster there was obtained the analytical model and there was performed a forecast for average traffic of busy hour in relation to the weekly cumulative traffic. In the case when for a real network the forecasting was performed without preliminate clustering, the result could be erroneous and as a result the cell could be configured with insuficient capacity in order to serve the awaited traffic and the KPIs of the network would degrade. For the presented example the difference in data prediction was estimated to 11 Gb.

IV. References

1. Cisco Visual Networking Index, Global Mobile Data Traffic Forecast Update, 2011
2. Andrew Moore, K-means and Gaussian Clustering - Tutorial Slides, http://www-2.cs.cmu.edu/~awm/tutorials/kmeans.html, accessed 10.02.12
3. Alaj R. Mishra, Advanced cellular network planning and optimization, John Wiley & Sons, Ltd, USA - 2007- 521 p
4. StefaniaSesia, I. Roufik, LTE The UMTS Long Term Evolution From theory to practice, John Wiley &Sons, Ltd, USA - 2009 - 626 p.