Cell Traffic Prediction Using Joint Spatio-Temporal Information

Enrico Lovisotto, Enrico Vianello, Davide Cazzaro, Michele Polese,
Federico Chiariotti, Daniel Zucchetto, Andrea Zanella, Michele Zorzi
Department of Information Engineering (DEI)
University of Padova, via G. Gradenigo 6/B, Padova, Italy
Email: {lovisott, vianell1, cazzarod, polesemi, chiariot, zucchett, zanella, zorzi}@dei.unipd.it

***Abstract*—In future cellular networks, the ability to predict network parameters such as cell load will be a key enabler of several proposed adaptation and resource allocation techniques. In this study, we consider a joint exploitation of spatio-temporal data to improve the prediction accuracy of standard regression methods. We test several such methods from the literature on a publicly available dataset and document the advantages of the proposed approach.**

I. Introduction

The evolution of cellular networks from 4G to 5G will rely on adaptive techniques in order to manage the increasing complexity of mobile systems [1]. Up to now, cellular networks were designed using worst-case dimensioning, but the increasingly strict capacity, latency and energy efficiency requirements, together with the lower profit margins, make a smarter approach appealing to network operators.

Anticipatory networking [2] is one of the most promising approaches in smart network adaptation: the idea is to exploit knowledge of the dynamics of the system in order to predict future network states and tailor the configuration to the expected profile. There are several possible contexts for the prediction, from a single user's channel gain [3] to large-scale mobility patterns [4]. In this work, we present several prediction techniques whose aim is to estimate future cell load, which is a key parameter in network planning and optimization.

A. Related Work

In the scientific literature, cell load prediction techniques are studied because of the potential gain they can provide to the performance of the network in a wide range of scenarios, such as energy efficient communications and dynamic network planning. In [5] the authors propose to use prediction techniques based on traffic matrices collected for groups of Base Stations (BSs) under the same coordinator in order to optimize the sleeping time of network elements, while in [6] a classification and prediction method is applied to temporal information given by Call Data Records in order to decide when and where it is appropriate to deploy femtocells. The spatio-temporal relation between cells is analyzed in [7], where insights on the predictability of the traffic in a cellular network are given; however, the authors do not attempt to predict future values of the cell load, but use large-scale traffic patterns to examine the correlation. The study in [8] uses traffic variations in cell neighborhoods, using a Markov decision process model, in order to enable energy saving techniques. There are other studies that consider the spatio-temporal context in cellular networks, but their focus is on the prediction of mobility of users [9], [10]. These can be then exploited in association with some knowledge of the network topology, as done in [11].

B. Contribution

The novelty of our work with respect to previous studies is that we consider machine learning techniques that exploit temporal and spatial data jointly: a cell's future load depends not only on its previous values, but also on the loads of neighboring cells. This joint approach can improve the prediction accuracy, especially in the noisiest and most challenging cases. We focus on medium-term prediction with a range of tens of minutes; such a range is still usable for network optimization, but is not as noisy and unpredictable as short-term cell load.

The rest of the paper is organized as follows: first, we present in Sec. II the prediction techniques we employed. Then, we describe the results on real data from the city of Milan in Sec. III. Finally, Sec. IV concludes the paper and lists some possible avenues for future work.

II. Prediction techniques

All the techniques we present in this paper are based on the exploitation of spatio-temporal data, which was first proposed by Ohashi *et al.* [12]. In order to jointly consider the spatial and temporal data, we need to define the concept of *spatio-temporal neighborhood*. If a cell at a given instant is characterized by its position in space and time, given by the vector (x, y, t), we define the distance between two points as

$$d_{i,j} = \sqrt{\left(\frac{x_i - x_j}{d_0}\right)^2 + \left(\frac{y_i - y_j}{d_0}\right)^2 + \alpha\left(\frac{t_i - t_j}{T}\right)^2}, \quad (1)$$

where d_0 is the inter-cell distance and T is the time interval between measurements. Note that the spatio-temporal distance between different instants is non-zero even if the cell is the same, i.e., the spatial distance is 0. The parameter $\alpha \geq 0$ is a weighting factor to combine the spatial and temporal measures.

The spatio-temporal neighborhood of a point m can then be defined as the set of the discrete points in the dataset whose distance from m is smaller than some radius β:

$$N_m^\beta = \{p : d_{m,p} < \beta\}. \tag{2}$$

The points belonging to the spatio-temporal neighborhood are contained in an ellipsoid in space-time, and, given the same β, a smaller α includes in the neighborhood points which are further away in time. The cell load values z_p of the points within the neighborhood can be used in the prediction. In addition to the pure values, we also use as input a series of indicators that capture some of the most relevant dynamics of the cell load, as in [12].

We implemented three indicators, which are listed below:

- The *weighted mean* is an average of the cell load values in the neighborhood, weighted by their spatio-temporal distance, and is given by:

$$\omega(N_m^\beta) = \frac{1}{|N_m^\beta|} \sum_{p \in N_m^\beta} \frac{z_p}{d_{m,p}} \tag{3}$$

- The *spread* is the standard deviation of the cell load values in the spatio-temporal neighborhood:

$$\sigma(N_m^\beta) = \sqrt{\frac{1}{|N_m^\beta|} \sum_{p \in N_m^\beta} (z_p - \bar{z})^2}, \tag{4}$$

  where $\bar{z}$ is the arithmetic mean of the cell load of all the points in the neighborhood.
- The *weighted tendency* is given by the ratio between the weighted means with two radii $\beta_1 < \beta_2$ (following [12], we choose $\beta_2 = \beta = 2\beta_1$):

$$\tau(N_m^{\beta_1,\beta_2}) = \frac{\omega(N_m^{\beta_1})}{\omega(N_m^{\beta_2})}. \tag{5}$$

  This indicator summarizes the trend of the cell load as it approaches the target location. For example, if $\tau(N_m^{\beta_1,\beta_2}) > 1$, then the load on the closest points in time and space is larger than that of farther points.

While in [12] the indicators are added to a purely temporal prediction, in our work we also use the cell load values of all the points in the spatio-temporal neighborhood as predictors.

A. Prediction algorithms

We tested the performance of several well-known prediction algorithms using the input we described above. The algorithms we used represent the state of the art for prediction with time series [13], [14], and are briefly described below:

- The simplest method we tested was the basic *multiple linear regression* [15], using least squares as a loss function.
- Given the highly variable nature of the data, we implemented some regularization techniques in order to avoid the risk of overfitting; we used three methods of *regularized linear regression*.

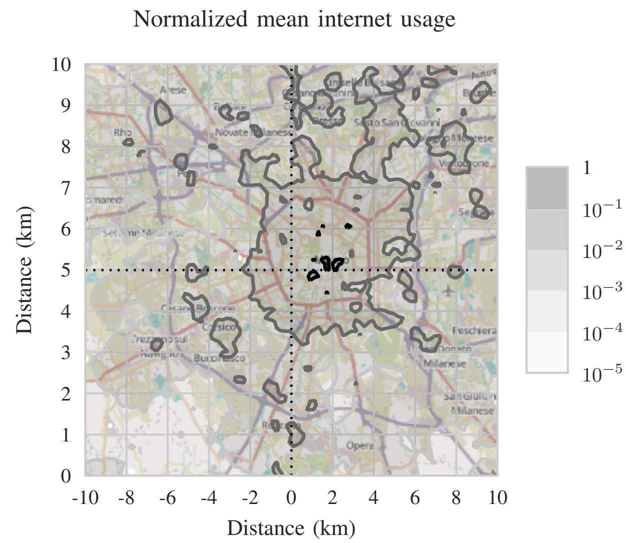

Fig. 1. Normalized average internet usage map.

  - *Ridge regression* [16] is a shrinkage method that adds a square penalty to the least squares loss, weighted by a regularization parameter λ_R.
  - *Lasso regression* [17] is a shrinkage method very similar to ridge regression, but uses a linear penalty instead of a square penalty.
  - *Elastic net regression* [18] is a linear combination of the lasso and ridge regularization techniques, and is particularly useful when the number of predictors is larger than the number of observations and in the presence of highly correlated predictors.
- Support Vector Machines (SVMs) are mostly known as a classification tool, but they can be adapted to output real numbers, giving us the *Support Vector Regression (SVR)* technique [19]. We used SVR with a linear kernel, which has a regularization parameter C.
- *Random Forest (RF)* [20] is an ensemble estimator that consists of a number of regression trees, whose output is the average output of all the trees. For optimal performance, the trees' decisions should be uncorrelated, and dataset bagging and random training techniques are employed to obtain this property.
- *Neural Networks (NNs)* [21] are well-known learning tools which use back-propagation to learn an objective function. In our work, we use the stochastic gradient descent method of back-propagation, using the $\tanh$ activation function.

III. Results

All the prediction methods we described above were trained and tested using the *Telecom Italia Big Data Challenge 2014* dataset,[1] which contains the records of the internet usage for a

[1] `https://dandelion.eu/datamine/open-big-data/`

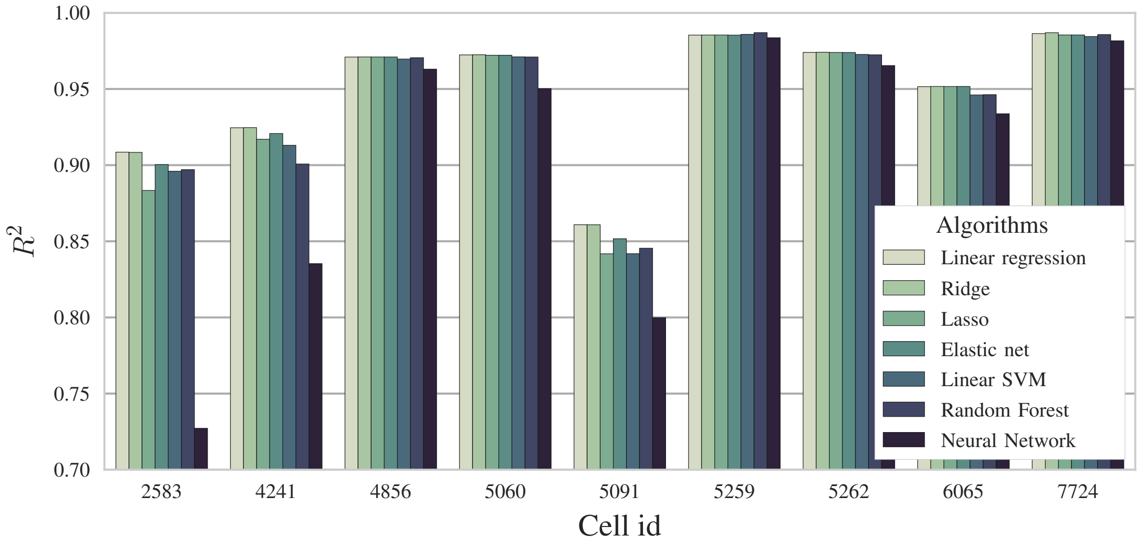

Fig. 2. Performance of the tested regression methods.

grid of square cells with 200 m sides (which makes $d_0 = 200$ m in Eq. (1)) in the city of Milan, Italy, for the last two months of 2013. The data had a sampling period of 10 minutes (i.e., $T = 600$ s in Eq. (1)). The normalized mean internet usage is overlaid on a map of the city in Fig. 1.

For computational reasons, we only predicted the load of a small but representative subset of cells, namely, the cells with id 2583, 4241, 4856, 5060, 5091, 5259, 5262, 6065 and 7724. These cells were selected because they are placed in different areas of the city and they show different traffic patterns. In particular, cells 2583 and 4241 have an average traffic that is close to the average traffic for the whole city, cells 5060, 5091 and 7724 show very high peak usage, and cells 4856, 5259, 5262 and 6065 have a very high average traffic.

The metric we chose for the results was the coefficient of determination R^2 [22], which is a commonly used metric in the regression literature, and gives an indication of how well the regression model describes the observed data.

A. Parameter optimization

All the parameters of the prediction algorithms were optimized by exhaustive search with 10-fold cross-validation, after dividing the dataset into training, validation and testing sets. The chosen values of the parameters are listed in Table I.

The values of the spatio-temporal weighting factor α and of the neighborhood radius β were optimized for each cell and are listed in Table II, for a number of neighbors from 27 to 46.

Parameter	Value	Description
λ_R	[1.637e-6, 0.074]*	Ridge regularization parameter
λ_L	[1e-06, 4.665e-6]*	Lasso regularization parameter
$\lambda_{R,E}$	[0, 1.105e-5]*	Ridge regularization (elastic net)
$\lambda_{L,E}$	[0, 4.665e-6]*	Lasso regularization (elastic net)
C	[0.22, 34.081]*	SVR linear kernel penalization term
N_t	200	Number of RF trees
γ	10^{-3}	NN learning rate
N_{iter}	10^4	Maximum NN iterations
ε	10^{-10}	NN convergence tolerance

*These parameters were optimized for each cell.

TABLE I
PARAMETERS USED IN THE SIMULATION.

Cell id	α	β	Number of neighbors
2583	0.25	2	27
4241, 4856	2.25	3	25
5060	0.09	2	46
5091	0.19	2	28
5259, 5262, 6065	0.12	2	37
7724	0.19	2	28

TABLE II
OPTIMAL NEIGHBORHOOD DEFINITION FOR EACH CELL.

B. Prediction results

Fig. 2 shows the prediction accuracy on the test set for each regression method. The figure clearly shows that the NN is not an accurate method, probably due to an insufficient training set size, whereas the other algorithms often have a similar performance. The reason is that the cell load can be easily predicted in most cells, and therefore the differences among different algorithms are minimal. On the other hand, in cells with poor prediction accuracy different methods show some performance difference. This reveals that, when the behavior of the load in a cell is less predictable, the prediction performance can be improved using different algorithms and additional context information. Indeed, the simple linear regression and ridge regression have a slightly better performance in cells 2583, 4241 and 5091, which are all located in peripheral areas of the city, close to major traffic roads or hubs (Via Gianbellino for cell 2583, the A1 highway for cell 4241, and Linate airport for cell 5091). In locations like these, with high

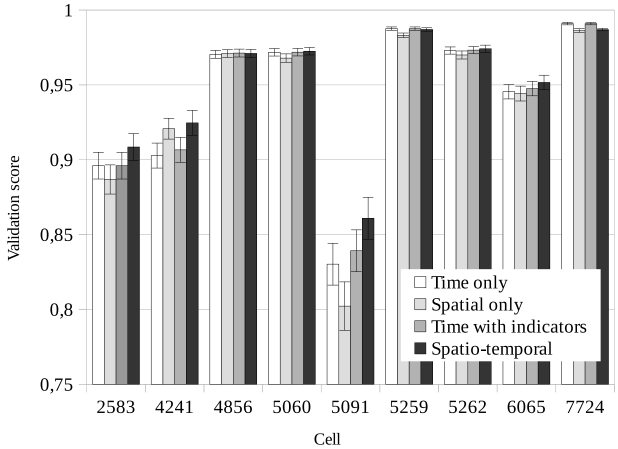

Fig. 3. Performance of the prediction algorithms for different neighborhood definitions.

mobility and bursty traffic, the benefit of combining spatial and temporal information is intuitive, and the performance improvement can be seen in Fig. 3. While only temporal or spatial data is sufficient in the highly predictable cells, the same 3 cells mentioned above show a marked improvement in the R^2 score when spatio-temporal data are considered jointly in the prediction. It is also worth noting that the use of temporal indicators does not result in a significant improvement by itself, but only when combined with the spatio-temporal neighborhood data.

The most accurate prediction methods are also the simplest: both training and parameter optimization for the linear, ridge, lasso and elastic net algorithms were significantly faster than for RF, SVR and NN. This offsets the increased complexity due to the bigger size of the neighborhood due to the inclusion of the spatial dimension in its definition.

IV. Conclusions and future work

In this paper, we applied several regression methods taken from the literature, combined with joint spatio-temporal information with indicators, to predict the future cell load on a 10 minute scale. We used real data from the Telecom Italia network in Milan to perform the training and evaluation of the different methods.

Our work proves the usefulness of joint spatio-temporal information in the most difficult prediction scenarios, confirming the importance of context information for network optimization.

Future work on the prediction methods might consider the introduction of new indicators which could capture network-specific dynamics, along with a more in-depth study of the effect of the neighborhood size on the prediction accuracy.

Another possible avenue for future research is a more systematic study of the dataset, applying the methods described in this work to all the cells in the dataset and correlating other geographical features with the prediction accuracy, as well as using them as additional indicators.

References

[1] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, Nov. 2014.

[2] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "Anticipatory networking in future generation mobile networks: a survey," *submitted to IEEE Communications Survey and Tutorials*, June 2016. [Online]. Available: https://arxiv.org/abs/1606.00191

[3] F. Chiariotti, D. Del Testa, M. Polese, A. Zanella, G. M. Di Nunzio, and M. Zorzi, "Learning methods for long-term channel gain prediction in wireless networks," in *International Conference on Computing, Networking and Communications (ICNC2017)*. IEEE, Jan. 2017.

[4] Y. Jiang, D. C. Dhanapala, and A. P. Jayasumana, "Tracking and prediction of mobility without physical distance measurements in sensor networks," in *International Conference on Communications (ICC)*. IEEE, June 2013, pp. 1845–1850.

[5] R. Li, Z. Zhao, X. Zhou, and H. Zhang, "Energy savings scheme in radio access networks via compressive sensing-based traffic load prediction," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 4, pp. 468–478, Nov. 2012.

[6] S. E. Hammami, H. Afifi, M. Marot, and V. Gauthier, "Network planning tool based on network classification and load prediction," in *2016 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, Apr. 2016, pp. 1–6.

[7] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *IEEE INFOCOM 2011 - The 30th Annual IEEE International Conference on Computer Communications*. IEEE, Apr. 2011, pp. 882–890.

[8] R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, "TACT: a transfer actor-critic learning framework for energy saving in cellular radio access networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 2000–2011, Apr. 2014.

[9] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, "NextPlace: a spatio-temporal prediction framework for pervasive systems," in *International Conference on Pervasive Computing*. Springer, May 2011, pp. 152–169.

[10] H. Gao, J. Tang, and H. Liu, "Mobile location prediction in spatio-temporal context," in *Nokia Mobile Data Challenge Workshop*, June 2012.

[11] W.-S. Soh and H. S. Kim, "QoS provisioning in cellular networks based on mobility prediction techniques," *IEEE Communications Magazine*, vol. 41, no. 1, pp. 86–92, Jan. 2003.

[12] O. Ohashi and L. Torgo, "Wind speed forecasting using spatio-temporal indicators," in *20th European Conference on Artificial Intelligence (ECAI'12)*. IOS Press, Aug. 2012, pp. 975–980.

[13] J. Ma and J. C. Cheng, "Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology," *Applied Energy*, vol. 183, pp. 182–192, Dec. 2016.

[14] F. Herrema, V. Treve, R. Curran, and H. Visser, "Evaluation of feasible machine learning techniques for predicting the time to fly and aircraft speed profile on final approach," in *International Conference for Research in Air Transportation*, June 2016.

[15] D. F. Andrews, "A robust method for multiple linear regression," *Technometrics*, vol. 16, no. 4, pp. 523–531, Nov. 1974.

[16] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970.

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, Jan. 1996.

[18] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B*, vol. 67, pp. 301–320, Apr. 2005.

[19] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Information Processing-Letters and Reviews*, vol. 11, no. 10, pp. 203–224, Oct. 2007.

[20] U. Grömping, "Variable importance assessment in regression: linear regression versus random forest," *The American Statistician*, vol. 63, no. 4, pp. 308–319, Sep. 2008.

[21] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568–576, Nov. 1991.

[22] N. R. Draper and H. Smith, *Applied regression analysis*. John Wiley & Sons, May 1998.