2016 15th IEEE International Conference on Machine Learning and Applications

# Improving HSDPA Traffic Forecasting Using Ensemble of Neural Networks

Isah A. Lawal*, Salihu A. Abdulkarim*, Muhammad K. Hassan† and Jibrin M. Sadiq‡
*Dept. of Computer Science, Federal University Dutse, Nigeria
Email: i.alawal@fud.edu.ng, sakwami@fud.edu.ng
†Pennsylvania State University, USA
Email: mkh171@psu.edu
‡Nokia Networks Nigeria, Abuja, Nigeria
Email: jibrin.sadiq@nokia.com

***Abstract*—Accurate forecasting of data traffic demand is very crucial for profitable operation of cellular data networks because it helps in facilitating the optimization and planning of the network resources. Many machine learning regression models including Support Vector Regression and Abductive Networks have been applied to this problem, but this paper study the concept of ensemble method for improving the forecasting accuracy. Specifically, a cooperative ensemble training strategy using two optimization algorithms is proposed to train a Neural Network model. The trained model is characterized with good forecasting performance due to the exchange of experience and knowledge among the two optimization algorithm during the training process. A dataset consisting of 44160 recordings of hourly High Speed Data Packet Access (HSDPA) data traffic, which was collected over a period of 30 days from sixty different sites of a UMTS based cellular operator was used to evaluate the performance of the propose method. Experimental results show the superiority of the Neural Network model trained with the proposed ensemble training strategy over other state-of-the-art methods.**

***Index Terms*—Neural Networks; Ensemble Methods; Forecasting;**

## I. Introduction

Forecasts of data traffic demand are useful for optimizing cellular data networks and for planning/scheduling of short-term network facility maintenance [1]. Accurate forecasting is necessary for profitable operation of cellular data networks, as high forecasting error leads to greater losses for the network operators. For example, poor forecasts may lead to allocation of inappropriate size of the network. Statistical regression techniques are the widely used models for forecasting data traffic demand in cellular data networks [2], [3]. Due to inherent noise and non-stationary in the historical data use in forecasting traffic demand, statistical regression models however, fall short of desired forecasting accuracy in many cases [4]. This is because they have pre-defined model structures and assume linearity of the historical data use in forecasting. Data-based machine learning techniques such as Neural Networks (NN) [5] and Support Vector Regression (SVR) [1] on the other hand, provide a better forecasting performance. These models are able to automatically discover patterns in the historical data traffic records without bias or influence by prior assumptions.

Ensemble of machine learning models have been used for improving accuracy in many application [6], but its potential have not been widely explore in cellular network data traffic demand forecasting [7]. With committee ensemble strategy, an odd number of machine learning models can be trained to solve the same problem independently. During prediction the models operate simultaneously on the input data and their outputs are combined to produce the final ensemble output. A major drawback of this approach is how to determine the best method among "majority voting", "simple averaging" and "weighted averaging" for combining the output of the individual committee models. On another hand, the ensemble strategy can be cooperative where two or more machine learning models exchange their experience and knowledge during the training process to produce the final ensemble model for prediction. This study is concerned with the latter ensemble strategy. We explore the cooperative use of global and local optimization algorithms to trained a NN model for short-term (6-hour to 12-hour ahead) forecasting of High Speed Data Packet Access (HSDPA) data traffic demand. Therefore, following a review of related earlier work in Section 2, Section 3 introduces the ensemble NN modeling approach. Section 4 describes the HSDPA traffic dataset used. Section 5 summarizes the experimental work using the proposed method and the results obtained. Finally, section 6 concludes the paper.

## II. Related work

The task of forecasting data traffic demand for cellular data networks have been studied in the literature as a time series regression problem, whereby historical data traffic are used to build the forecasting models [3]. Svoboda et al [2] present a study which employs two different modelling methods namely; Auto Regressive Moving Average (ARMA) and Dynamic Harmonic Regression (DHR) for forecasting packet switched traffic demand from 3G cellular networks. Yu et al [8] on the other hand, suggested the combination of ARMA model with fractional auto-regressive integrated moving average method for forecasting the throughput of 3G cellular network. This aimed at improving the forecasting accuracy of the regression method when the input data exhibit multifractal characteristics. However, the aforementioned methods have two major