Electricity Load Forecast using Data Streams Techniques

Pedro Pereira Rodrigues and João Gama

LIACC-NIAAD, University of Porto
R. de Ceuta 118-6
4050-190 Porto,Portugal
{prodrigues,jgama}@liacc.up.pt

Abstract. Sensors distributed all around electrical-power distribution networks produce streams of data at high-speed. From a data mining perspective, this sensor network problem is characterized by a large number of variables (sensors), producing a continuous flow of data, in a dynamic non-stationary environment. In this work we analize the most relevant data mining problems and issues: online learning and change detection. We propose an architecture based on an online clustering algorithm where each cluster (group of sensors with high correlation) contains a neural-network based predictive model. The goal is to maintain in real-time a clustering model and a predictive model able to incorporate new information at the speed data arrives, detecting changes and adapting the decision models to the most recent information. We present preliminary results illustrating the advantages of the proposed architecture.

Keywords: Electricity demand forecast, online clustering, incremental neural networks.

1 Motivation

Electricity distribution companies usually set their management operators on SCADA/DMS products (Supervisory Control and Data Acquisition / Distribution Management Systems). One of their important tasks is to forecast the electrical load (electricity demand) for a given sub-network of consumers. Load forecast is a relevant auxiliary tool for operational management of an electricity distribution network, since it enables the identification of critical points in load evolution, allowing necessary corrections within available time. In SCADA/DMS systems, the load forecast functionality has to estimate, on a hourly basis, and for a near future, certain types of measures which are representative of system's load: active power, reactive power and current intensity. In the context of load forecast, near future is usually defined in the range of next hours to the limit of seven days, for what is called *short-term* load forecast.

Traditionally, real knowledge extraction problems faced a barrier on the relative scarcity of data. Nowadays, not rarely the amount of available data is so huge that traditional systems, based on memory and several reading of same information, cannot operate efficiently. Moreover, on current real applications, data are being produced in a continuous flow, at high speed, producing examples over time [4]. In this context, faster answers are usually required, keeping an anytime model of the data, enabling better decisions.

Given its practical application and strong financial implications, electricity load forecast has been targeted by innumerous works, mainly relying on the non-linearity and generalizing capacities of neural networks, which combine a cyclic factor and an auto-regressive one to achieve good results [5]. Nevertheless, static iteration-based training, usually applied to estimate the best weights for network connections, is not adequate for the high speed production of data usually encountered. Moreover, a predictive system may be developed to serve a set of thousands of load sensors, but the load demand values tend to follow a restrict number of profiles, considerably smaller than the total of registered sensors. This way, clustering of sensors greatly allows the reduction of necessary predictive models. However, most work in data stream clustering has been concentrated on example clustering and less on variable clustering [8].

© Revue MODULAD, 2007