MAP-i Thesis Proposal

Relevant Information

Title of the Proposal: Data Mining for Water Distribution Network Managment

Thematic Area: Data Mining

Supervisor: Luis Torgo / Faculty of Sciences, UPorto / LIAAD-INESC Tec, LA

Hosting Research Unit: LIAAD-INESC Tec,LA

Brief Description of the Thesis Proposal

Motivation

Water is considered by many as a key resource for the future of humanity and a potential source for conflicts among nations. In this context, a proper management of this invaluable resource is of key importance and impact on the human society. Water distribution networks play a key role in bringing water to populations, particularly in large metropolitan areas where water consumption is typically higher. The importance of a correct management of this resource leads to tight regulations that impose constraints on the management of these networks. In this context, companies managing these networks typically invest large resources in monitoring these facilities, which leads to the availability of a significant amount of data concerning their functioning in real time. Data mining tools and techniques can play a key role in deriving useful information from these data to help in the tasks of managing these complex networks. The goal of this proposal is to investigate data mining tools that are useful at solving the main problems of this key management activities.

Main Goals / Objectives

The main goal of this proposal is to develop a decision support system for helping in managing water distribution networks. This system should focus on two key issues of this activity: (i) water demand; and (ii) water quality. The software should use available data from a concrete network to provide: (i) monitoring and (ii) prediction operations in the context of water demand and water quality control.

More research oriented goals include the study and development of spatiotemporal data mining techniques to help with the analysis of data arriving from a spatially distributed network of sensors.

Relevant Research Questions

The main research questions to be addressed in this thesis are:

- Monitoring Data
 - Summarization
 - * What are the essential statistical summaries that key to the management of a water distribution network?
 - * Given the large volume of data, how to provide the summaries to managers in a useful manner?
 - Visualization
 - * How to present the huge amount of data and indicators that are available in a synthetic and useful manner to managers?
 - * How to clearly provide visual clues on potential anomalies that arise in real time?

• Forecasting Data

- Water Demand
 - * What type of predictions are useful to managers in this domain? Point predictions? Interval predictions? Other?
 - * Develop accurate predictive models of water demand that answer the previous questions
- Water quality control
 - * What are the major quality constraints in water management? How to pass this information to models?
 - * What are the key events that managers want to avoid given their potential costs?
 - * Develop predictive models that are able to forecast the above key and potentially rare events