Scrutinizing Changes in the Water Demand Behavior

Manuel Herrera, Rafael Pérez-García, Joaquín Izquierdo and Idel Montalvo

Abstract. Time series novelty or anomaly detection refers to automatic identification of novel or abnormal events embedded in normal time series points. In the case of water demand, these anomalies may be originated by external influences (such as climate factors, for example) or by internal causes (bad telemetry lectures, pipe bursts, etc.). This paper will focus on the development of markers of different possible types of anomalies in water demand time series. The goal is to obtain early warning methods to identify, prevent, and mitigate likely damages in the water supply network, and to improve the current prediction model through adaptive processes. Besides, these methods may be used to explain the effects of different dysfunctions of the water network elements and to identify zones especially sensitive to leakage and other problematic areas, with the aim to include them in reliability plans. In this paper, we use a classical Support Vector Machine (SVM) algorithm to discriminate between nominal and anomalous data. SVM algorithms for classification project low-dimensional training data into a higher dimensional feature space, where data separation is easier. Next, we adapt a causal learning algorithm, based on the reproduction of kernel Hilbert spaces (RKHS), to look for possible causes of the detected anomalies. This last algorithm and the SVM’s projection are achieved by using kernel functions, which are necessarily symmetric and positive definite functions.

1 Introduction

The anomaly detection of water demand time series aims to correct likely data errors in measures from telemetry systems. These systems are used by most water
companies in big cities for control and operation purposes. This will allow more accurate estimations that can be used to immediately detect severe anomalies, such as service disruption. Simultaneously, it can identify more rapidly light anomalies which can develop insidious and progressively [7]. If no errors are found in data, the novelties significance is the occurrence of some physical change in the water supply network or in the demand behavior caused by external influences, such as climate factors. Changes in time series behavior may exhibit permanent or transitional effects. The causes of these are diverse, but could be divided into external and internal causes. Examples of the first class are weather or calendar factors. For the internal causes one can have wrong telemetry readings, water leakage or failure of one or more valves. There is a need to divide the problem into two phases: anomaly detection and action taking. In this way, one can obtain early warning methods to identify and mitigate likely damages in the water supply network or to improve the current prediction model through some adaptive process.

To distinguish between normal and abnormal deviations, novelties will be sought in three specific cases: when data loggers identify a disruption of service, when the discrepancies between the last observations and their prediction are significant, and when the last observations lack of the expected random characteristics. Here, we consider working with sliding time windows to include all possible cases. The sliding window method is based on a window size $W$; only the latest $W$ observations are used for detection. As an observation arrives, the oldest observation in the sliding window expires. An alert processing method based on Support Vector Machines will be proposed to extract trends and to highlight punctual discrepancies between observed and predicted data [14]. To look for possible causes of the detected anomalies, we propose using the recently developed causal learning algorithm based on the reproduction of kernel Hilbert spaces [20]. By using this methodology the statistical dependences can always be detected by correlations after the data are mapped into an appropriate feature space. The algorithm is an improvement of the inductive causation (IC) algorithm [18], which generalizes in several ways. The control of the consequences of novelties in the earlier stages can avoid, among other things, economic and water losses, which are of great importance from the point of view of water as a scarce resource. This paper will focus on the development of markers of the different possible types of anomalies in water demand time series, explaining their causes, and proposing a feasible integration mechanism in the prediction system. Figure 1 summarizes the process.

The paper is organized as follows. Section 2 shows a brief literature review about detection of anomalies in water distribution systems. In section 3 we present a methodology to detect the possible types of novelties within time series data, to develop a way of classifying them and to discuss their causes. Section 4 gives some application results and summarizes the conclusions.