
Forensic Analysis of Time Series

Kumaraswamy Ponnambalam

With the advent of automatic sensors for detection and data collection (for example, SCADA systems), it is now possible to acquire a large number of time series of critical data. In urban water and sewer systems, monitoring stations can collect data on water quantity and quality (for example, dissolved oxygen, electric conductivity, pH and turbidity, among others). The motivation for such data collection usually is to analyze if the systems are working properly. New analytical techniques are needed in order to efficiently analyze such large quantities of data and to answer questions of forensic nature (for example, how well the systems are working and whether any of the components are faulty). An automatic inference system consisting of feature extraction, clustering, and classification steps is developed to answer categorical questions using the large amount of data.

27.1 Proposed Approach

27.1.1 Time Series

Figure 27.1 presents an example of just three time series of water quality collected at a point in a river section. For plotting convenience, all variables are normalized using the maximum observed value for the corresponding variable in the data. A vast amount of such data can be now collected easily. However, it is not clear what method should be chosen such that using these data one can come up with categorical answers to questions such as *Is the treatment system*

upstream working well? or, if not, *What component of the system is not working?* In other time series, the questions can vary. For example, using stock prices or economic variables can we answer the question *Is the stock market is bull or bear?* That is, *Should we buy or sell or hold?* Moreover, in the case of the water quality monitoring system, we may want to rate how well the system is working. Of course, this also depends on the water quality of the inflows to the treatment system and its outflows, as well as the natural variations of the background quality of the receiving rivers. The key contribution of this work is to formulate methodologies that can develop machine inference systems that allow for both automatic and expert inputs and will be able to analyze and infer hundreds of such time series, Liao (2005).

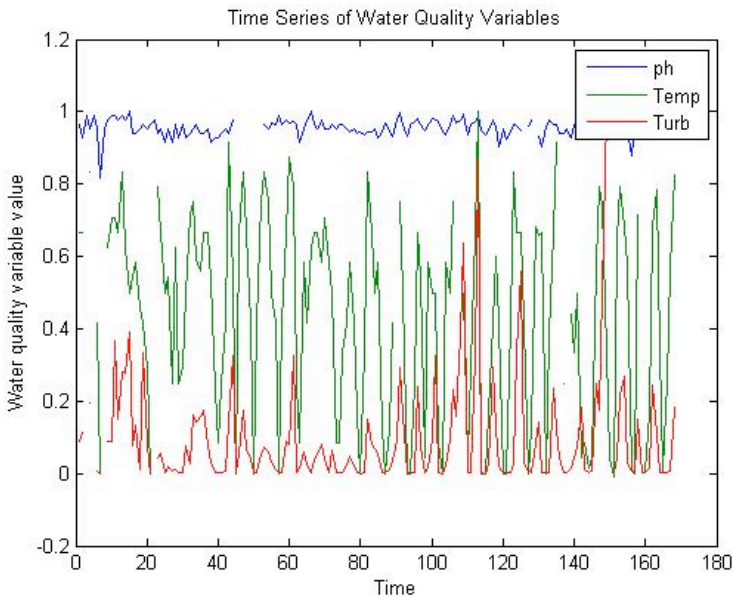


Figure 27.1 Time series of water quality variables.

27.1.2 Inference System

The main steps of the inference system (InferSys) are slicing the long time series into many slices of short length time series, features extraction from each of these slices, clustering sets of samples into groups of similar properties, and classification of future samples. That is, features define various characteristics

of each slice of the time signal, clustering is used to divide the samples into identifiable groups, and classification is used to predict the group of new samples. There are some variations possible in these steps. For example, if the original classification corresponding to a slice is already known or given, then clustering is not needed. Each of these steps of the inference system will be explained in detail next.

27.1.3 Slicing

A long time series is sliced into a large number of smaller length time series using an appropriate windowing method. Depending upon the length of the data and the frequency of data collection, it is possible to estimate an appropriate windowing system. A common window used is the rectangular window which is easy to use as it simply keeps the original data in the given window and zeros the data elsewhere. Alternatively, one can use other windows such as the Hamming or Blackman window, each of which can be specialized for selected applications. The window can be used also as a sliding window to create a larger number of samples when some overlapping features are important. The shorter the sliced data, the larger the number of samples we can analyze, and vice versa. The tradeoff is that a shorter window may provide too little information, and too large a window will smear the information contained. In this chapter, we use the rectangular window in a non-overlapping manner using natural frequencies (e.g., weekly) and the results later show the sensitivity of the results to window sizes. In addition to windowing, the data may need to be scaled and in some cases missing values need to be estimated. A common scaling technique is to scale all the corresponding values to lie between either 0 and 1 or -1 and 1 using the corresponding maximum or minimum value of the time series. The scaling is chosen such that relative differences between categories are not lost.

27.1.4 Feature Selection

An important step in the design of any time series inference system is the selection of a good set of features that are capable of characterizing the time series in the feature space. Features are extracted from the sliced time series by applying suitable transformations. Domains from which the features are extracted are listed in Table 1. Each of these features may provide a set of vectors of multiple dimensions. Therefore, features define the time series in a compact manner and the selection of these at the moment is highly problem dependent and depends

on the understanding of the general nature of the time series where an expert may help.

Table 27.1 A sample of feature vectors.

Domain	Features
Statistical	Mean, Standard Deviation, Skew, Kurtosis, Correlation
Derivatives	First, Second, Third
Fourier Analysis	Fourier coefficients
Wavelet Analysis	Wavelet detail coefficients
ARIMA model	Model coefficients
Cepstral Analysis	MFC coefficients

There are hundreds of possible features. In addition, the principal component analysis (PCA) can be used to produce a smaller number of vectors that combine the effects of all the features that have been obtained. At the end of this step of analysis, a large table is formulated where the samples (each slice of time series after windowing) are in rows, and features are in columns. If necessary, these columns can be scaled or normalized. Note that while a slice of time series may have thousands of data values, they have all been used to calculate the features, which are much fewer in number. In addition, features are expected to abstract the various essential characters of the time series while ignoring small errors.

27.1.5 Clustering Analysis

After the end of the feature selection step, each sample needs to be categorized as belonging to one of the few important system states. For example, in the case of stock prices, either bull or bear. An expert may be able to provide such a classification but even experts may make mistakes. Alternatively, clustering analysis helps in providing an objective method for such categorization into a cluster number and is called an unsupervised method as. Figure 27.2 provides a sample result of clustering analysis.

The x - and y - axes are the values of either two features or the two PCA vectors and these are clustered into three clusters (the choice of three was chosen by the analyst). There are many clustering methods as described in Karray and Silva (2004) and Jain (2009). We use the k-means method to demonstrate. This method minimizes the sum, over all clusters, of the within-cluster sums of point-to-centroid cluster distances. Samples that are closer to each of the cluster centroids are assigned the number of that cluster by this method. At the end of

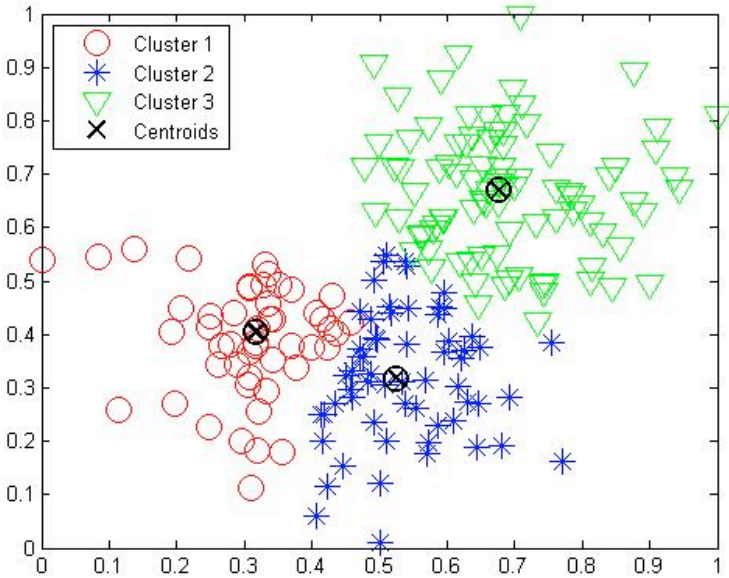


Figure 27.2 Clustering analysis using the k-means method.

this step, each of these clusters may be linked to some system-wide meaning such as “Treatment plant N, component L not working.”

27.1.6 Classification

Once we have the samples with their features and corresponding categories, an inference system is developed which, in future, when given a set of features, will recognize its category. This can be done using a number of classification methods or a fusion of these methods. A sample list of classification methods is: linear classifiers (support vector machines); k-nearest neighbour; decision trees; neural networks; Bayesian networks; and hidden Markov models. For simplicity, we describe the k-nearest neighbour (KNN) method which is similar to the k-means clustering method. A sample point is assigned to the class to which the majority of the k-nearest neighbours belong. There could be other rules to assign the class, like consensus. The distance can be measured using various metrics, the simplest being the Euclidean. In contrast to clustering methods (which are unsupervised learning methods), classification methods are

supervised learning because the sample category (or class) is assumed to be given.

Each classifier (or a subset of classifiers) may be strong in identifying certain cases and weak in others. We can fuse the results such that the overall prediction rate is significantly better than the best classifier. The simplest and still the most commonly used fusion method does the classification based on majority voting.

27.2 Results and Discussions

The purpose of this example is to demonstrate the kind of results one can get by using the inference system and its sensitivity to some of the parameters. The data used came from a process monitoring system and only three time series were used with 65% of data used for training the classifier and 35% of data used for testing. These data were picked from random sampling of the original time series with the chosen window size and Table 27.2 presents some results.

Table 27.2 K-means method and window size sensitivity.

Window size in Time Period Total Data = 8341 points	Accuracy of Classification (%) (Mean, Std. Dev)
20	(13, 60)
30	(22, 38)
40	(28, 28)
60	(78, 24)

It is clear from these results that window size has a large importance both in mean and standard deviations of prediction accuracy. In Table 27.3, we present the sensitivity to the size of the training and testing sets for a window size of 40.

Table 27.3 K-means method and the sensitivity to training and testing sets.

(Training, Testing) % Total Data = 8341 points	Accuracy of Classification (%) (Mean, Std. Dev)
(65, 35)	(28, 28)
(75, 25)	(85, 17)

The reason for such high sensitivities is most likely due to the relatively small number of data points in the original time series consisting of only 8 341

points. Sensitivities of prediction results to features, the type of clustering and classification method used may also exist. Therefore, for any given problem design of experiments must be conducted for choosing an appropriate method and the set of parameters that give the best results

27.3 Conclusion

In recent times, much data has become available, a large number of them belonging to the class of time series. Use of this data in a forensic sense to answer the question of what caused changes is a new field of study, in contrast to the classical study of prediction of the time series data. The problem considered here is to identify the different categorizations of these time series in order to make decisions. Moreover, the need for automation and the rapidness with which the answers need to be found also make this problem challenging. Examples of such problems and their approximate prediction accuracies are (these are from our own works in progress):

- Music genre recognition (70–95%);
- Epilepsy seizure prediction (70–95%);
- Stock market evaluation (80%);
- Process fault detection (85%); and
- Water quality monitoring (80%).

Acknowledgment

Much of this work was explored in the last three years with two of my graduate students, Arvind Dorai and Suriyapriya Vallamsundar.

References

- Jain A.K. Data Clustering : 50 Years Beyond K-Means, Technical Report TR-CSE-09-11, Michigan State University, 2008. (Pattern Recognition Letters doi:10.1016/j.patrec.2009.09.011).
- Karray, F. and De Silva, C.2004, Soft Computing and Intelligent Systems Design. Addison Wesley, Pearson Education Ltd., Toronto.
- Liao T. W. 2005. Clustering of time series data—a survey, Pattern Recognition, 38, 1857-1874.