Preface

A time-series describes a discrete sequence of amplitudes of an entity ordered over time. Typically, a time-series may describe the temporal variations in atmospheric temperature, rainfall, humidity, stock price, and any other measurable quantity that has a variation with time and available as a discrete data points sampled uniformly or non-uniformly over time. Prediction of a time-series refers to determining its value at unknown time point $t + 1$ from the knowledge of the time-series at its current time t and also previous time points $(t - 1)$, $(t - 2)$, …, $(t - n - 1)$. On occasions, the predicted value of the time-series depends also on one or more influencing time-series, besides having dependence on its past values. A time-series prediction thus refers to a regression problem in high-dimensional space, where the predicted value describes a highly nonlinear function of its past values and also other relevant time-series. Unfortunately, the nature of nonlinearity being unknown adds more complexity to the prediction problem.

Several approaches to time-series prediction are available in the literature. One of the early works on time-series prediction refers to the well-known ARMA (autoregressive moving average) model, which is appropriate for prediction of a stationary time-series. However, many of the natural/man-made time-series are non-stationary with varying frequency components over the time frames. One indirect approach to predict a non-stationary time-series is to transform it to an equivalent stationary series by taking the differences of consecutive data points and then to test its stationarity with the help of autocorrelation (AC) and partial autocorrelation (PAC) measures. In case the resulting time-series is not stationary yet, we repeat the above step until stationarity is attained. The prediction of the resulting stationary time-series is performed using the ARMA model, and the predicted series is integrated as many times the difference operator has been applied to the original time-series. The complete process of predicting a non-stationary time-series by adoption of the above three steps, transformation to a stationary time-series, prediction using the ARMA model, and integration of the predicted stationary time-series by the requisite number of times is referred to as ARIMA (autoregressive integrated moving average) model.

The fundamental limitation of the ARIMA model lies in the high-order differentiation of the series, which sometimes results in a white noise with zero mean, and therefore, it is of no use from the prediction point of view. This calls for alternative formulation to handle the problem. The logic of fuzzy sets has an inherent advantage to represent nonlinear mapping, irrespective of non-stationary characteristics of the time-series. In addition, the non-Gaussian behavior of a time-series can be approximated by a locally Gaussian formulation. Thus, fuzzy logic can handle nonlinearity, non-stationarity, and non-Gaussian-ness of the time-series. In addition, non-deterministic transition of the time-series from a given partition of a vertically partitioned time-series to others can also be handled with fuzzy logic by concurrent firing of multiple rules and taking aggregation of the resulting inferences generated thereof. Besides fuzzy logic, pattern clustering and classification by neural and/or other means can also take care of the four issues indicated above. This motivated us to develop new prediction algorithms of time-series using pattern recognition/fuzzy reasoning techniques.

Song and Chissom in 1994 proposed fuzzy time-series, where they assign a membership value to each data point to represent the degree of belongingness of each data point in a given partition. They extracted single-point prediction rules from the time-series, where each rule represents a partition P_i in the antecedent and a partition P_j in the consequent, where partition P_i includes the current time point and P_j the next time point. Once the construction of rules is over, they developed fuzzy implication relations for each rule. These relations are combined into a single relational matrix, which is used later to derive fuzzy inferences from membership functions of measured time-series value at the current time point. A defuzzification algorithm is required to retrieve the predicted sample value from the fuzzy inference.

Extensive works on fuzzy time-series have been undertaken over the last two decades to perform prediction from raw time-series data, primarily to handle uncertainty in diverse ways. A few of these that need special mention includes partition width selection, influence of secondary data for the prediction of main time-series, extension of fuzzy reasoning mechanism (such as many-to-one mapping), different strategies for membership function selection, and the like. Pedrycz et al. introduced a novel technique to determine partition width in the settings of optimization and employed evolutionary algorithm to solve the problem. Chen et al. proposed several interesting strategies to utilize secondary memberships for the prediction of main factored economic time-series. They used more stable Dow Jones and NASDAQ as the secondary factor time-series for the TAIEX time-series as the main factor. Huarang et al. proposed different type-1 fuzzy inferential schemes for prediction of time-series. The details of the above references are given in this book.

In early studies of time-series prediction, researchers took active interest to utilize the nonlinear regression and functional approximation characteristics of artificial neural networks to predict time-series from their exemplary instances. Traces of research considering supervised learning techniques for prediction of time-series value are available in the literature. Early attempts include developing functional mapping for predicting the next time point in the series from the current time point value. Most of the researchers employed gradient descent learning-based back-propagation algorithm neural technique and its variant for time-series prediction. Among other significant works, neural approaches based on the radial basis function and support vector machine need special mention for time-series prediction. In recent times, researchers take active interest on deep learning and extreme learning techniques for time-series prediction.

This book includes six chapters. Chapter 1 provides a thorough introduction to the problem undertaken with justification to the importance of the selected problem, limitations of the existing approaches to handle the problem, and the new approaches to be adopted. Chapters 2–5 are the original contributions of the present research. Chapter 2 examines the scope of uncertainty modeling in time-series using interval type-2 fuzzy sets. Chapter 3 is an extension of Chap. 2 with an aim to reason with both type-1 and type-2 fuzzy sets for the prediction of close price time-series. The importance of using both type-1 and type-2 fuzzy sets is apparent from the availability of number of data points in a given partition of a prepartitioned time-series. When there exist a fewer data points in a single contiguous region of a partition, we represent the partition by a type-1 fuzzy set, else we go for an interval type-2 representation. Chapter 4 introduces a clustering technique for subsequence (comprising a few contiguous data points) prediction in a time-series. This has a great merit in forecasting applications, particularly for economic time-series. Chapter 5 attempts to design a new neural technique to concurrently fire a set of prediction rules realized on different neural networks and to combine the output of the neural networks together for prediction. Chapter 6 is the concluding chapter, covering the self-review and future research directions.

In Chap. 2, a new technique for uncertainty management in time-series modeling is proposed using interval type-2 fuzzy sets. Here, the time-series is partitioned into k number of blocks of uniform width, and each partition is modeled by an interval type-2 fuzzy set. Thus, transition of consecutive data points may be described by type-2 fuzzy rules with antecedent and consequent containing partitions P_i and P_j, respectively, where P_i and P_j denote two consecutive data points of the time-series. We then employ interval type-2 fuzzy reasoning for prediction of the time-series. The most important aspect of the chapter is the inclusion of secondary factor in time-series prediction. The secondary factor is used here to indirectly control the growth/decay rate in the main factored time-series.

Chapter 3 is an extension of Chap. 2 to deal with fuzzy reasoning using both type-1 and interval type-2 fuzzy sets. The motivation of using both types of fuzzy sets has been discussed, and principles used for reasoning with prediction rules containing both types of fuzzy sets are discussed. The improvement in performance of the proposed technique with only type-1 and interval type-2 has been demonstrated, indicating the significance of the technique.

Chapter 4 deals with a novel machine learning approach for subsequence (comprising a few contiguous data points) prediction. This is undertaken in three phases, called segmentation, clustering, and automaton-based prediction. Segmentation is required to represent a given time-series as a sequence of segments (time blocks), where each segment representing a set of contiguous data points of any arbitrary dimension should have a semantically meaningful shape in the context of an economic time-series. In the current book, segmentation is performed by identifying the slopes of the data points in the series for a sequence of a finite number of data points and demarcating them into positive, negative, or near-zero slopes.

A segment is labeled as rising, falling, or having a zero slope, depending on the maximum frequency count of any of the three primitives: positive, negative, or near-zero slope, respectively, for every five or more consecutive data points in the series. After the segments are identified, we group them using a clustering algorithm based on shape similarity of the normalized segments. As the number of clusters is an unknown parameter, we used the well-known DBSCAN algorithm that does not require the said parameter. In addition, we extend the DBSCAN clustering algorithm to hierarchically cluster the data points based on descending order of data density. The cluster centers here represent structures of specific geometric shape describing transitions from one partition to the other. To keep track of the transition of one partition to the others using acquired structures, we develop an automaton and use it in the prediction phase to predict the structure emanating from a given time point. The predicted information includes the following: Given a terminating partition, whether there exists any feasible structure that terminates at the desired partition starting at the given time point? In addition, the probable time to reach the user-defined partition and its probability of occurrence are additional parameters supplied to the user after prediction. Experiments undertaken on three standard time-series confirm that the average accuracy of structure prediction is around 76%.

In Chap. 5, we group prediction rules extracted from a given time-series in a manner, such that all the rules containing the same partition in the antecedent can fire concurrently. This requires realization of the concurrently fireable rules on different neural networks, pretrained with supervised learning algorithms. Any traditional supervised learning algorithms could be used to solve the problem. We, however, used the well-known back-propagation algorithm for the training of the neural networks containing the prediction rules.

This book ends with a brief discussion on self-review of this book and relevant future research directions in Chap. 6.

Kolkata, India

Amit Konar
Diptendu Bhattacharya

http://www.springer.com/978-3-319-54596-7

Time-Series Prediction and Applications
A Machine Intelligence Approach
Konar, A.; Bhattacharya, D.
2017, XVIII, 242 p. 69 illus., 13 illus. in color., Hardcover
ISBN: 978-3-319-54596-7