Preface

Data production and collection is rapidly increasing today. The quantity of data is approximately doubled each year. This increase is caused by using computers in the business world and scientific developments as well as the data obtained from tools such as smart cards, cameras, satellite systems and more. In addition, using the internet as a global information system requires large quantities of data. The daily processing data that accumulate electronically due to shopping processes in markets and is known as market basket analysis (Agraval & Srikatan, 1994), and log data which traces the users interacting with websites and information systems (Bucklin & Sismeiro, 2003) are examples of this data accumulation. Information systems collect and accumulate large quantities of data in various formats such as texts, videos, images and more. The collected data may be entirely regular, semi-regular or entirely text. Data in this format make it difficult for researchers to make correct decisions. This has led to the need for new technical and automatic tools. Newly developed computer science theories and tools help researchers extract information quickly from large quantities of data (Berthold, Borgelt, Hoppner, & Klawonn, 2010). This has led to the development of an interdisciplinary research area known by names such as database information extraction and intelligent data analysis. This area is closely related to the techniques such as statistics, machine learning, artificial intelligence, fuzzy logic, rough sets, grey theory, heuristic algorithms and computer science. These techniques complement each other. Many statistical methods need computers for especially large data sets. However, the power of computers alone cannot be replaced by statistical information.

Intelligent Data Analysis (IDA) is associated with Knowledge Discovery from Databases (KDD) (Frawley, Piatetsky-Shapiro, & Matheus, 1991), which is frequently defined as a process, and consists of the following steps (Fayyad, Piatetsky-Shapiro, & Smyth, 1996): Understanding the problem, cleaning the data and making them available for processing, extracting the information hidden in the data and expressing this information in pattern and rule forms. This last step is usually defined as Data Mining (DM) in KDD.

These intelligent techniques help us to obtain useful information from large quantities of data. A research area called database information extraction or data mining has been developed in computer science to achieve this purpose. These fields discover the patterns that are hidden in large quantities of data and include information. These intelligent methods have emerged recently. Twenty years ago, researchers in artificial intelligence and computer science increasingly began to take interest in the processing, analysis and interpretation of large quantities of data. The main reason for this was that large quantities of data were hard to process with the methods of the time such as statistics and conventional methods.

We used the term, Intelligent Data Analysis (IDA), instead of Knowledge Discovery from Databases (KDD), although these two terms have a lot in common. Intelligent data analysis is an approach that
Preface

performs data analysis with artificial intelligence methods (Mohammadzadeh, Safdari, & Mohammadzadeh, 2014). This tool allows the extraction of useful information hidden in large quantities of data. It also helps decision makers make quick decisions. Intelligent data analysis is very similar to data mining; however, it uses the initial information area to analyze the data using iterative and interactive methods. This initial information area increases the effectiveness of the information extraction process and prevents researchers and users from receiving useless results. Probability methods such as statistical classifications, hidden Markov models, Bayesian classification and rule extracting methods are used in traditional data analysis (Holmes & Peek, 2007). Intelligent data analysis tries to solve real world problems using techniques from various fields such as statistics, artificial intelligence, data mining, computational statistics, machine learning and optimization. Thus, intelligent data analysis is useful for researchers and decision makers in many different fields (Vučenović, Trivić, & Kos, 2015).

Therefore, in this book we describe studies of the application of intelligent data analysis techniques in a variety of fields. This book includes a variety of research subjects and practical techniques of data analysis based on theories and various intelligent methods. These subjects include the basic subjects of exploratory data analysis, classification and feature selection, soft computing, information extraction, decision making, data mining, estimation and clustering.

ORGANIZATION OF THE BOOK

The book is organized into fifteen chapters. A brief description of each of the chapters follows:

Chapter 1 presents the notion of ‘wavelet transform’, which has gained an edge over other existing methods due to its highly efficient transforming capabilities in biomedical signal processing. The authors focused on testing and validating the use of this technique on different signals keeping in view the specific needs for various biomedical applications. The importance of this unified technique is highlighted with statistical results and validation on several benchmark datasets.

Chapter 2 examines the allocation problem of shelf space management. The chapter first sets the Yang’s model which includes linear profit function for the shelf space allocation. Then, The authors developed heuristic approaches both based on particle swarm optimization and artificial bee colony algorithm to solve the problem. They showed the heuristics approaches are superior than Yang’s heuristics for the shelf space allocation model by using performance analysis.

Chapter 3 attempts to determine the factors affecting housing demand in Turkey. The authors of this chapter established a genetic algorithm-based multivariate grey model in order to reveal their aims. They showed that several factors including M2 money supply, consumer price index and urbanization rate have an impact on housing demand.

Chapter 4 gives an image mining approach. It is an evidence that a huge amount of data is continuously accumulated in databases through internet and various types of smart devices. It is also clear that description-based querying is almost impossible on such a big unstructured data. The author proposed an image mining which provides example-based querying on image databases.

Chapter 5 tries to predict the electricity load. The day-ahead forecasting model is especially important in day-ahead market activities and plant scheduling operations. This chapter’s authors developed several forecasting model in order to predict the hourly loads. Their experimental results indicated that support vector regression and seasonal grey model outperforms other approaches in terms of forecast accuracy for day-ahead load forecasting.
Chapter 6 presents a two-stage segmentation model based on soft computing using the purchasing behaviours of customers in a data mining framework. Segmentation was performed via neuro-fuzzy two stage-clustering approach for a secondary data set, which included more than 300,000 unique customer records, from a UK retail company. The findings indicated that the model provided stronger insights and has greater managerial implications in comparison to the traditional two-stage method with respect to six segmentation effectiveness indicators.

Chapter 7 focuses on available data analysis and data mining techniques to find the optimal location of the Multicriteria Single Facility Location Problem (MSFLP) at diverse business settings. The authors proposed a new approach to the facility location problem with the combination of the PCA-rank and ranking SVMs while presenting a general classification of the MSFLP and its framework.

Chapter 8 addresses the issue of facility location problem with geographical clustering of the customers while considering their demands. The author proposed a new neighborhood structure for the standard particle swarm optimization algorithm to solve the problem. They compared their approach with k-means, fuzzy c-means, fuzzy c-means and center of gravity hybrid method, revised weighted fuzzy c-means algorithm and the standard particle swarm optimization algorithms on several large data sets from the literature. The results indicate that the proposed approach achieves less total transportation cost with less computational cost requirement in facility location problems.

Chapter 9 aims to provide an overview of business intelligence in healthcare context by specifically focusing on the applications of intelligent systems. The author reviewed the current applications into three main categories and presented some important findings of that research in a systematic manner. The discussions made in this chapter can also facilitate the researchers in that area to generate a research agenda for future work in applied health science, particularly within the context of health management and policy and health analytics.

Chapter 10 is about the possible use of chaotic random numbers in the metaheuristic and artificial intelligence algorithms that requires random numbers. The authors only provide the necessary information about the algorithms instead of providing full detailed explanation of the subjects assuming the readers already have theoretical basic information.

Chapter 11 reviews the issue of supplier selection which is an important problem in business. The authors argued that it is possible to solve it by improving the traditional grey relation analysis and VIKOR methods in fuzzy environment. They took a problem from literature and compared their solution with fuzzy and AHP methods.

Chapter 12 analyzes several numbers of techniques in literature, which efficiently provide solutions for document clustering. The author proposed metaheuristic algorithms for document clustering problem because of the limitations of the existing traditional clustering algorithms. They also gave a brief review over various research papers that present the area of document or text clustering approaches with different metaheuristic algorithms.

Chapter 13 addresses the issue of churn with particular reference to Internet Service Providers (ISPs). The authors implemented a questionnaire to investigate customer loyalty status, behavior, and information of the churn possibility in Turkey. They used a real data collected from a survey and obtained a promising and applicable results. It is believed that it will help ISP companies to determine the required advertising campaigns for the customers.

Chapter 14 analyzes liquidity flows into a number of financial instruments such as stocks, commodities, currencies, futures. The authors tried to identify which financial instruments were quantitatively more predictable. They used the data for their model intraday frequencies covering the period between
1993 and 2013. They analysed 72 different artificial neural networks to observe distinction on prediction performances of different financial instrument.

Chapter 15 proposed a novel Wavelet Feature Selection algorithm for ranking and selecting the features from the wavelet decompositions. The authors used support vector machine to rank the features and backward feature elimination to remove the features. The EEG datasets are used to test the algorithm. The results confirmed that the algorithm is able to improve the efficiency of wavelet features in terms of accuracy and feature space.

Numan Çelebi
Sakarya University, Turkey

REFERENCES


