

Foreword

The importance of knowledge discovery and data mining is evident by the great plethora of books and papers dedicated to this subject. Such methods are finding applications in almost any area of human endeavor. This includes applications in engineering, science, business, medicine, humanities, just to name a few. At the same time, however, there is a great confusion about the development and application of such methods. The main reason for this situation is that many, if not most, of the books examine issues on data mining in a narrow manner. Very few books study issues from the mathematical/algorithmic and also the applications point of view simultaneously. Even fewer books present a comprehensive view of all the critical issues involved with the development and application of such methods to many real-life domains. The present book, edited by two world-renowned scholars, Drs. Giovanni Felici and Carlo Vercellis, is a bright example of the most valuable books in this fast emerging field. The emphasis of this book on the mathematical aspects of knowledge discovery and data-mining methods makes the presentations scientifically sound and easy to understand in depth.

The 19 chapters of this book have been written by a number of distinguished scholars, from all over the world, who discuss the most critical subjects in this area. The book starts by discussing an important first step for any application of such methods, that is, how to discretize the data. This step is essential as many methods use binary data, while real-life applications may be associated with nonbinary data. If the analyst is not careful at this step, then it is possible to end up with too many nonrelevant variables that generate computational problems associated with highly dimensional data. A related step is that of cleaning the data before they are used to extract the pertinent models. As before, if this step is not done properly, the validity of the final results may be in jeopardy. Another interesting topic discussed in this book is the development of sophisticated visualization techniques, which are presented in relation with many diverse domains, ranging from astronomy to genetics. The successful application of the proposed visualization techniques to these two highly demanding application areas witnesses the high potential of these methods for a wide spectrum of applications. The high volumes of log data, produced by recording the way people surf the Web, provides an exciting opportunity, amid with interesting algorithmic challenges, for knowledge discovery and data-mining methods. This fascinating topic is also discussed here. Another very interesting subject discussed is how to mine data that come from virtual environments that involve some kind of spatial navigation. Such studies involve the analysis of sequences of routes of actions. A highly promising direction of research seems to be based on the development of methods that attempt to combine characteristics of various approaches. Such methods are known as hybrids, and an interesting development of a new hybrid approach and its applications are presented as well.

No book in this area would be complete without the discussion of logic-based methods that offer some unique algorithmic and application advantages. The relevant discussions are done by some of the most knowledgeable world-renowned scholars on this subject; logic methods are also applied to the analysis of financial data. The potential of using grids for distributed approaches and also parallelism is

explained too. Another prominent topic is the use of clustering approaches, which are discussed by developing some specialized evolutionary approaches. Clustering is also used in one of the most promising application areas for the future, the analysis of time series of data. Applications can be found in many domains as one realizes that systems or phenomena of interest usually generate data over time. In such settings, data from one point of time are somehow related to the data of the next point of time. Again, this very fascinating problem is discussed in great depth, and in an easy-to-understand manner by a distinguished expert in this fast-growing field. An extensive treatment of the classification technique, known as support vector machines, is provided by three chapters of the book; altogether with a complete treatment of the main theory of this method and of the related kernel function theory, the new extension of discrete support vector machines is described and applied to bioinformatics. This type of data is also the topic of other applications described in the book. The picture is completed with the description of data-mining approaches for service quality measurement, of fuzzy set and rough set theory applied in different contexts, and of other industrial applications of data mining.

It is quite clear that this book is very valuable to all practitioners and researchers working on different fields but unified by the need to analyze their voluminous and complex data. Therefore, it is strongly recommended to anyone who has an interest in data mining. Furthermore, it is hoped that others will follow the example of this book and present more studies that combine algorithmic developments and applications in the way this edited book by Drs. Felici and Vercellis does so successfully.

Evangelos Triantaphyllou, PhD
Professor
Department of Computer Science
Louisiana State University
Baton Rouge, LA 70803 USA