
Preface

Knowledge Discovery demonstrates intelligent computing at its best, and is the most desirable and interesting end-product of Information Technology. To be able to discover and to extract knowledge from data is a task that many researchers and practitioners are endeavoring to accomplish. There is a lot of hidden knowledge waiting to be discovered – this is the challenge created by today’s abundance of data.

Knowledge Discovery in Databases (KDD) is the process of identifying valid, novel, useful, and understandable patterns from large datasets. Data Mining (DM) is the mathematical core of the KDD process, involving the inferring algorithms that explore the data, develop mathematical models and discover significant patterns (implicit or explicit) -which are the essence of useful knowledge. This detailed guide book covers in a succinct and orderly manner the methods one needs to master in order to pursue this complex and fascinating area.

Given the fast growing interest in the field, it is not surprising that a variety of methods are now available to researchers and practitioners. This handbook aims to organize all major concepts, theories, methodologies, trends, challenges and applications of Data Mining into a coherent and unified repository. This handbook provides researchers, scholars, students and professionals with a comprehensive, yet concise source of reference to Data Mining (and additional selected references for further studies).

The handbook consists of eight parts, each part consists of several chapters. The first seven parts present a complete description of different methods used throughout the KDD process. Each part describes the classic methods, as well as the extensions and novel methods developed recently. Along with the algorithmic description of each method, the reader is provided with an explanation of the circumstances in which this method is applicable, and the consequences and trade-offs incurred by using that method. The last part surveys software and tools available today.

The first part describes preprocessing methods, such as cleansing, dimension reduction, and discretization. The second part covers supervised methods, such as regression, decision trees, Bayesian networks, rule induction and support vector machines. The third part discusses unsupervised methods, such as clustering, association rules, link analysis and visualization. The fourth part covers soft computing

methods and their application to Data Mining. This part includes chapters about fuzzy logic, neural networks, and evolutionary algorithms.

Parts five and six present supporting and advanced methods in Data Mining, such as statistical methods for Data Mining, logics for Data Mining, DM query languages, text mining, web mining, causal discovery, ensemble methods, and a great deal more. Part seven provides an in-depth description of Data Mining applications in various interdisciplinary industries, such as finance, marketing, medicine, biology, engineering, telecommunications, software, and security.

The motivation: Over the past few years we have presented and written several scientific papers and research books in this fascinating field. We have also developed successful methods for very large complex applications in industry, which are in operation in several enterprises. Thus, we have first hand experience in the needs of the KDD/DM community in research and practice. This handbook evolved from these experiences.

The first edition of the handbook, which was published five years ago, was extremely well received by the data mining research and development communities. The field of data mining has evolved in several aspects since the first edition. Advances occurred in areas, such as Multimedia Data Mining, Data Stream Mining, Spatio-temporal Data Mining, Sequences Analysis, Swarm Intelligence, Multi-label classification and privacy in data mining. In addition new applications and software tools become available. We received many requests to include the new advances in the field in a second edition of the handbook. About half of the book is new in this edition. This second edition aims to refresh the previous material in the fundamental areas, and to present new findings in the field. The new advances occurred mainly in three dimensions: new methods, new applications and new data types, which can be handled by new and modified advanced data mining methods.

We would like to thank all authors for their valuable contributions. We would like to express our special thanks to Susan Lagerstrom-Fife of Springer for working closely with us during the production of this book.

Tel-Aviv, Israel
Beer-Sheva, Israel

Oded Maimon
Lior Rokach

April 2010