

Preface

Ensemble methodology imitates our second nature to seek several opinions before making a crucial decision. The core principle is to weigh several individual pattern classifiers, and combine them in order to reach a classification that is better than the one obtained by each of them separately.

Researchers from various disciplines such as pattern recognition, statistics, and machine learning have explored the use of ensemble methods since the late seventies. Given the growing interest in the field, it is not surprising that researchers and practitioners have a wide variety of methods at their disposal. *Pattern Classification Using Ensemble Methods* aims to provide a methodic and well structured introduction into this world by presenting a coherent and unified repository of ensemble methods, theories, trends, challenges and applications.

Its informative, factual pages will provide researchers, students and practitioners in industry with a comprehensive, yet concise and convenient reference source to ensemble methods. The book describes in detail the classical methods, as well as extensions and novel approaches that were recently introduced. Along with algorithmic descriptions of each method, the reader is provided with a description of the settings in which this method is applicable and with the consequences and the trade-offs incurred by using the method. This book is dedicated entirely to the field of ensemble methods and covers all aspects of this important and fascinating methodology.

The book consists of seven chapters. Chapter 1 presents the pattern recognition foundations that are required for reading the book. Chapter 2 introduces the basic algorithmic framework for building an ensemble of classifiers. Chapters 3-6 present specific building blocks for designing and implementing ensemble methods. Finally, Chapter 7 discusses how ensembles should be evaluated. Several selection criteria are proposed - all are

presented from a practitioner's standpoint - for choosing the most effective ensemble method.

Throughout the book, special emphasis was put on the extensive use of illustrative examples. Accordingly, in addition to ensemble theory, the reader is also provided with an abundance of artificial as well as real-world applications from a wide range of fields. The data referred to in this book, as well as most of the Java implementations of the presented algorithms, can be obtained via the Web.

One of the key goals of this book is to provide researchers in the fields of pattern recognition, information systems, computer science, statistics and management with a vital source of ensemble techniques. In addition, the book will prove to be highly beneficial to those engaged in research in social sciences, psychology, medicine, genetics, and other fields that confront complex data-processing problems.

The material in this book constitutes the core of undergraduate and graduates courses at Ben-Gurion University. The book can also serve as an excellent reference book for graduate as well as advanced undergraduate courses in pattern recognition, machine learning and data mining. Descriptions of real-world data-mining projects that utilize ensemble methods may be of particular interest to the practitioners among the readers. The book is rigorous and requires comprehension of problems and solutions via their mathematical descriptions. Nevertheless, only basic background knowledge of basic probability theory and computer science (algorithms) is assumed in most of the book.

Due to the broad spectrum of ensemble methods, it is impossible to cover all techniques and algorithms in a single book. The interested reader can refer to the excellent book "pattern classifiers: methods and algorithms" by Ludmila Kuncheva (John Wiley & Sons, 2004). Other key sources include journals and conferences' proceedings. The Information Fusion Journal and the Journal of Advances in Information Fusion are largely dedicated to the field of ensemble methodology. Nevertheless, many pattern recognition, machine learning and data mining journals include research papers on ensemble techniques. Moreover, major conferences such as the International Workshop on Multiple Classifier Systems (MCS) and the International Conference on Information Fusion (FUSION) are especially recommended as sources for additional information.

Many colleagues generously gave comments on drafts or counsel otherwise. Dr. Alon Schclar deserves special mention for his particularly detailed and insightful comments. I am indebted to Prof. Oded Maimon for lend-

ing his insight to this book. Thanks also to Prof. Horst Bunke and Prof. Patrick Shen-Pei Wang for including my book in their important series in machine perception and artificial intelligence. The author would also like to thank Mr. Steven Patt, Editor, and staff members of World Scientific Publishing for their kind cooperation throughout the writing process of this book.

Last, but certainly not least, I owe my special gratitude to my family and friends for their patience, time, support, and encouragement.

Lior Rokach

Ben-Gurion University of the Negev

Beer-Sheva, Israel

September 2009