

Preface

Multiple instance learning (MIL) is a recent learning framework that has become very popular lately. In this framework, objects are represented as sets of feature vectors (or bags, in MIL terminology). This kind of representation is well suited for certain problems, such as the prediction of structure–activity relationships, image classification, document categorization or the prediction of protein binding sites. In fact, MIL provides a much more natural representation than the one used in classical machine learning, where a single feature vector is used per object.

The first papers on MIL appeared in the early nineties. Their main interest is solving classification problems where input data are represented as multiple instances. This field, known as multiple instance classification (MIC), is the most popular subparadigm of MIL, but not the only one. In recent years, papers have also appeared on multiple instance regression (multi-instance learning with a continuous output) and multi-instance clustering. This book aims to present a general and comprehensible overview of the MIL paradigm, providing a formal definition of the framework and covering the subparadigms it comprises, the most relevant algorithms and the most representative applications.

The book is divided into three main parts. The first part (Chaps. 1–3) introduces the most important concepts of the discipline that will be necessary to understand the remainder of the book. Chapter 1 contains some introductory concepts on knowledge discovery in databases, data preprocessing, and data mining, whereas Chap. 2 introduces the multiple instance learning paradigm from a descriptive perspective. The first part finishes with a chapter focused on MIC. Besides including a formal definition of the problem and a taxonomy for MIC algorithms, it also carries out a study of two important issues, namely distance metrics and alternative learning hypotheses.

The second part of the book (Chaps. 4–7) provides an exhaustive review of the different MIL algorithms. Chapters 4 and 5 describe the most important classification algorithms, following the taxonomy introduced in Chap. 3. Chapter 6 introduces multiple instance regression, the other main task in supervised learning. Chapter 7 covers two unsupervised learning tasks in the MIL framework: clustering and association rule mining.

The last part of the book (Chaps. 8–10) deals with other recent areas. Data reduction for MIL is addressed in Chap. 8. Chapter 9 discusses the problem of learning with imbalanced multi-instance data. Finally, Chap. 10 introduces multiple instance multiple label learning, a new learning framework that combines MIL with multi-label learning.

The target audience for this book is anyone interested in a good understanding of this important paradigm of machine learning, as well as a deep description of the current state of the art in the discipline. Practitioners in the industry and enterprise should find new insights and possibilities in the breadth of the topics covered. Researchers and data scientists in universities, research centers, and companies could appreciate this comprehensive review and uncover new ideas for productive research efforts.

Granada, Spain
Córdoba, Spain
Santa Clara, Cuba
Ghent, Belgium
Córdoba, Spain
Santa Clara, Cuba
Ghent, Belgium

Francisco Herrera
Sebastián Ventura
Rafael Bello
Chris Cornelis
Amelia Zafra
Dánel Sánchez-Tarragó
Sarah Vluymans

Multiple Instance Learning

Foundations and Algorithms

Herrera, F.; Ventura, S.; Bello, R.; Cornelis, C.; Zafra, A.;

Sánchez-Tarragó, D.; Vluymans, S.

2016, XI, 233 p. 46 illus., 40 illus. in color., Hardcover

ISBN: 978-3-319-47758-9