

Preface

The huge growth of information stored everywhere, from mobile phones to datacenter servers, as well as the large user base of many Internet services, such as social networks and online services for publishing music, pictures, and videos, demands automated systems for categorizing and labeling all this information. A common characteristic of texts published in news sites and blogs, videos, images, and pieces of music is that all of them can be assigned to multiple categories at once. Hence, the need to have algorithms able to adequately classify the data assigning it the proper labels.

Multilabel classification is a data mining area that encompasses several tasks specific for this type of data, including custom metrics aimed to characterize multilabel datasets and also to evaluate results, specialized preprocessing methods able to solve the peculiarities of multilabeled data, and also specific classification algorithms qualified for learning from this type of data, among others. Most of these techniques are pretty new and many of them are still in development.

Multilabel classification is a topic which has generated a notable interest in late years. Beside its multiple applications to classify different types of online information, it is also useful in many other areas, such as genomic and biology. Consequently, the demand for multilabel techniques is constantly growing. This book will guide the reader to the discovery of all aspects of multilabel classification.

Based on the experience of the authors after several years focused on multilabel learning techniques, this book reviews the specificities of this kind of classification, including all the custom metrics and techniques designed to deal with it, and provides a comprehensive reference for anyone interested in the field.

After portraying the context that multilabel classification belongs to, in the introduction, a formal definition of this problem along with a broad view on how it has been faced and the fields it has been applied to are provided in the second chapter. The third one is devoted to introducing most of the publicly available multilabel use cases, as well as the metrics defined to characterize and evaluate them. Chapters 4–6 review multilabel classification methods grouping them into three groups, depending on the approach followed to tackle the task, data

transformation, method adaptation, or the use of ensembles. Two of the most relevant obstacles in working with multilabel data, high dimensionality and class imbalance, are discussed in Chaps. 7 and 8. Chapter 9 introduces several software tools and frameworks aimed to ease the work with multilabel data, including obtaining this kind of datasets, performing exploratory analysis and conducting experiments.

Although multilabel learning is still in an early development stage with respect to other data mining techniques, the amount of proposed algorithms, most of them classification methods, is impressive. In the foreseeable future, it predictably will further expand to additional application fields, and the volume of new techniques grows almost every day.

The intended audience of this book are developers and engineers aiming to apply multilabel techniques to solve different kinds of real-world problems, as well as researchers and students needing a comprehensive review on multilabel literature, methods, and tools. In addition to the text itself, the authors supply the readers with a software repository containing data, code, and links, along with two R packages as tools to work with multilabel data.

We wish to thank all our collaborators of the research groups “Soft Computing and Intelligent Information Systems” and “Intelligent Systems and Data Mining.” We are also thankful to our families for their helpful support.

Granada, Spain
Granada, Spain
Jaén, Spain
Jaén, Spain
May 2016

Francisco Herrera
Francisco Charte
Antonio J. Rivera
María J. del Jesus

Multilabel Classification

Problem Analysis, Metrics and Techniques

Herrera, F.; Charte, F.; Rivera, A.J.; del Jesus, M.J.

2016, XVI, 194 p. 72 illus., Hardcover

ISBN: 978-3-319-41110-1