Chapter 2
Multilabel Classification

Abstract This book is concerned with the classification of multilabeled data and other tasks related to that subject. The goal of this chapter is to formally introduce the problem, as well as to give a broad overview of its main application fields and how it have been tackled by experts. A general introduction to the matter is provided in Sect. 2.1, followed by a formal definition of the multilabel classification problem in Sect. 2.2. Some of the main application fields of multilabel classification are portrayed in Sect. 2.3. Lastly, the approaches followed to face this duty are introduced in Sect. 2.4.

2.1 Introduction

Multilabel classification is a predictive data mining task with multiple real-world applications, including the automatic labeling of many resources such as texts, images, music, and video. The learning from multilabel data can be accomplished through different approaches, such as data transformation, method adaptation, and the use of ensembles of classifiers.

This chapter begins by formally defining the multilabel classification problem, introducing the mathematical notation and terminology that will be used throughout this book. Then, the different areas in which multilabel classification is applied nowadays will be outlined, and the repositories this kind of data can be obtained from are introduced.

The learning from multilabel data is being currently faced through disparate approaches, including data transformation and adaptation of traditional classification methods. The use of ensembles of classifiers is also quite popular in this field. In addition, some specific aspects, such as the use of label dependency information, the problems of high dimensionality, and label imbalance, must be considered. All these topics will be further described, along with an enumeration of the main multilabel software tools currently available.

2.2   Problem Formal Definition

The main difference between traditional[1] and multilabel classification is in the output expected from trained models. Where a traditional classifier will return only one value, a multilabel one has to produce a vector of output values. Multilabel classification can be formally defined as follows.

2.2.1   Definitions

Definition 2.1 Let $\mathcal{X}$ denote the input space, with data samples $X \in A_1 \times A_2 \times ... \times A_f$, being f the number of input attributes and $A_1, A_2, ..., A_f$ arbitrary sets. Therefore, each instance X will be obtained as the cartesian product of these sets.

Definition 2.2 Let $\mathcal{L}$ be the set of all possible labels. $\mathcal{P}(\mathcal{L})$ denotes the powerset of $\mathcal{L}$, containing all the possible combinations of labels $l \in \mathcal{L}$ including the empty set and $\mathcal{L}$ itself. $k = |\mathcal{L}|$ is the total number of labels in $\mathcal{L}$.

Definition 2.3 Let $\mathcal{Y}$ be the output space, with all the possible vectors Y, $Y \in \mathcal{P}(\mathcal{L})$. The length of Y always will be k.

Definition 2.4 Let $\mathcal{D}$ denote a multilabel dataset, containing a finite subset of $A_1 \times A_2 \times ... \times A_f \times \mathcal{P}(\mathcal{L})$. Each element $(X, Y) \in \mathcal{D} | X \in A_1 \times A_2 \times ... \times A_f, Y \in \mathcal{P}(\mathcal{L})$ will be an instance or data sample. $n = |\mathcal{D}|$ will be the number of elements in $\mathcal{D}$.

Definition 2.5 Let $\mathcal{F}$ be a multilabel classifier, defined as $\mathcal{F} : \mathcal{X} \to \mathcal{Y}$. The input to $\mathcal{F}$ will be any instance $X \in \mathcal{X}$, and the output will be the prediction $Z \in \mathcal{Y}$. Therefore, the prediction of the vector of labels associated with any instance can be obtained as $Z = \mathcal{F}(X)$.

2.2.2   Symbols

From these definitions, the following list of symbols, which will be used in this chapter and the next ones, is derived:

$\mathcal{D}$ Any multilabel dataset (MLD).
n The number of data instances in $\mathcal{D}$.
$\mathcal{L}$ The full set of labels appearing in $\mathcal{D}$.
l Any of the labels in $\mathcal{L}$.
k Total number of elements in $\mathcal{L}$.
X The set of input attributes of any instance.

---

[1]In general, we will refer to binary and multiclass classification, which are the most well-known classification kinds, as *traditional* classification.

f The number of attributes comprising X.
$\mathcal{X}$ The full input space in $\mathcal{D}$, consisting of all X instances.
Y The set of output labels (*labelset*) of any instance.
$\mathcal{Y}$ The full output space in $\mathcal{D}$, comprised by all Y instances.
Z The labelset predicted by the classifier.

While referring to specific individual instances, the usual notation $\mathcal{D}_i$ will be used. Analogously, Y_i will make reference to the true labelset of instance i, Z_i the labelset predicted by the classifier to the ith instance, and so on.

The metrics aimed to evaluate the performance of a classifier (they will be described in the next chapter) compute the differences between Y_i and Z_i, that is, the real labelset associated with each instance and the one predicted by the classifier. The goal while training a multilabel classifier would be reducing the error ratio brought by these metrics.

2.2.3 Terminology

In addition to the previous notation, the following terms will be frequently used throughout the text:

- **MLD/MLDs**: Any multilabel dataset or group of multilabel datasets.
- **MLC**: Any multilabel classifier or multilabel classification algorithm.
- **Instance/sample**: A row in an MLD, including its input attributes and associated labelset.
- **Attributes/features**: Generally used to refer to the set of input attributes in the MLD, without including the output labelset.
- **Label**: Any of the output attributes associated with an instance.
- **Labelset**: A vector of labels $\{0, 1\}^k$ whose length will be always k.
- **True labelset**: The labelset a sample in the MLD is annotated with.
- **Predicted labelset**: The labelset an MLC is giving as output for a new sample.

2.3 Applications of Multilabel Classification

Once the main concepts linked to multilabel classification have been introduced, the next question arising possibly is what can it be used for. As has been explained in the previous section, a multilabel classifier aims to predict a set of relevant labels for a new data instance.

In this section, several application areas that can take advantage of this functionality are portrayed. In the following chapter, specific use cases belonging to each one of these areas will be detailed.

2.3.1 Text Categorization

Multilabel classification has its roots as a solution for organizing text documents into several not mutually exclusive categories. This is why there are so many publications regarding this topic [15, 21, 40–42]. Most of them will be further described in the next chapter.

Textual documents can be found anywhere, from big companies which store all kind of agreements and reports to individuals filing their invoices and electronic mail messages. All published books and magazines, our historic medical records, as well as articles in electronic media, blog posts, question–answering forum messages, etc., are text documents also. Most of them can be classified into more than one category, thus the usefulness of multilabel classification to accomplish this kind of work.

The usual process to transform a set of text documents into an MLD relies on text mining techniques. The source documents are parsed, uninformative words are removed, and vectors with each word occurrence among the documents are computed. At the end of this process, each document is described by a row in the MLD, and the columns correspond to words and their frequencies or some other kind of representation such as TF/IDF (*Term Frequency/Inverse Document Frequency*, [47]).

2.3.2 Labeling of Multimedia Resources

Although documents containing only text were the most frequent ones in the past, nowadays images, videos, and music are commonly used resources due to the huge growth experienced by storage and communication technologies. Attending to 2015 YouTube statistics, an estimated 432 000 new videos are uploaded every day. The number of new music and sound clips published each day is also impressive, and new images and photographs posted everywhere, from blogs to Web sites such as Tumblr, reach the millions per day barrier.

Multilabel classification has been used to label all these types of resources [5, 7, 28, 48, 56, 58], identifying the objects which appear in sets of images, the emotions produced by music clips, or the concepts which can be derived from video snips. This way, the huge number of new resources can be correctly classified without needing human intervention, something that would be rather expensive.

Even though an image can be represented as a vector of color values, taking each pixel as a column, usually they are preprocessed in order to obtain the most relevant features. For doing so, segmentation techniques aimed at extracting boundaries are applied, the image can be convoluted to transform the pixel space into an energy space, etc. Similar procedures are put in practice with other multimedia resources.

2.3.3 Genetics/Biology

Proteomics and genomics are research fields which have experienced a huge growth in late years. As a result, immense databases with protein sequences have been produced, but only a small fraction of them have been studied to determine their function. Analyzing protein properties and gene expression is a very costly task, but DM techniques can accelerate the process and make it cheaper.

The application of multilabel classification in this area [25, 29] aimed at predicting the biologic functions of genes. Each gene can express more than one function at once, hence the interest in using MLC techniques to face this problem. The features used as predictors are usually the protein's *motifs*, which are traits about its internal structure. Structural motifs indicate how the elements in a secondary structural layer are connected. Certain patterns, such as short chains of amino acids, can be identified and used as predictive features.

2.3.4 Other Application Fields

In addition to the ones referenced in the previous sections, multilabel classification has been utilized in many other applications, both with public and with private datasets, sometimes ad hoc generated for a specific need. Most of them could be included in the first two categories, since eventually the goal is to categorize texts, sounds, images, and videos. The following are among the most interesting ones:

- The analysis of nonverbal expressions in speech is the focus in [49], aiming at detecting how people feel. The goal is to predict several not mutually exclusive affective states at once. The predictive attributes are a set of vocal features such as intonation and speech rate.
- In [19], the authors propose a system to automatically classify patent records. The addressed problem is easing the search of previous documents according to inventive principles. In the end, this proposal can be seen as another text categorization solution.
- The method presented in [50] aims to improve the process of searching for relevant information in Twitter. Five different labels are defined to classify twits, including news, opinions, and events. Several of them can be assigned simultaneously to the same twit.
- Analyzing complex motion in events is the goal of the system proposed in [17]. It combines trajectory and multilabel hypergraphs of moving targets in video sequences, detecting the relationship between the low-level features and high-level concepts which are to be predicted.

In general, multilabel classification could be suitable for any scenario in which some kind of information, no matter its type as long as it can be processed by an MLC algorithm, is assigned to two or more categories simultaneously.

2.3.5 MLDs Repositories

Usually, when new datasets are generated, the authors of each original work publicly provide the MLDs they have produced, so that other researchers can use them in their own studies. Notwithstanding, most MLDs can be obtained directly from some of the following data repositories:

- **MULAN**: The well-known multilabel learning package [55] has an associated data repository [54]. As of 2016, more than 20 MLDs are available in it. These MLDs are in ARFF[2] (*Attribute-Relation File Format*), and each one has an associated XML file describing the labels and their relationships. The XML file is needed, since the position of the attributes acting as labels is not assumed to be at the end.
- **KEEL**: KEEL [2] is a open source software providing lots of preprocessing and DM algorithms. It has an associated data repository [1] with almost 20 MLDs. Along with the full datasets, 10-fcv and 5-fcv partitions are also supplied. The KEEL file format, based on ARFF, includes the name of output attributes in the header; therefore, a separate XML file is not needed.
- **MEKA**: MEKA is a multilabel software developed on top of WEKA [37]. As the previous ones, it has also associated a data repository [45] with over 20 MLDs. The file format is also ARFF based, and the information needed to know which attributes are the labels, specifically the number of labels, is included in the header. Some multidimensional datasets are included.
- **RUMDR**: *R Ultimate Multilabel Dataset Repository* [11] provides a compilation of all MLDs publicly available, as well as an R package which eases the partitioning and exporting to several formats, including MULAN, KEEL, MEKA, LibSVM, and CSV.

A common group of MLDs is available in any of the aforementioned repositories, but in the MULAN and MEKA repositories, some specific ones, not available in the other sites, can be found. The partitions used in the experiments conducted in following chapters can be downloaded from the GitHub repository associated with the book [12].

2.4 Learning from Multilabel Data

The process to obtain a multilabel classification model is similar to that used for traditional classification, usually following supervised learning techniques. Most algorithms rely on an initial training phase. It depends on previously labeled samples to adjust the parameters of the model. Once trained, the model can be used to predict the labelset for new instances.

---

[2]ARFF is the file format used by WEKA [37].

When it comes to learning from multilabel data, two main approaches have been applied: data transformation and method adaptation. The former is based on transformation techniques which, applied to the original MLDs, are able to produce one or more binary or multiclass datasets. These can be processed with traditional classifiers. The latter aims for adapting existent classification algorithms, so they can deal with multilabel data, producing several outputs instead of only one. A third alternative, which naturally emerges from the first one, is the use of ensembles of classifiers.

A topic closely related to MLC is label ranking [57], whose goal is to map each instance in a dataset to a set of labels, establishing a total order relationship among them. That is, the output of these algorithms is a ranking of labels, assigning to each one a certain weight. A multilabel classifier can use this label ranking to decide which labels are relevant to the instance, applying a cut threshold that has to be computed in some way. In [8, 30, 38], different proposals on how to do multilabel classification founded on label rankings are presented.

This section introduces the three main approaches to MLC: data transformation, method adaptation, and ensembles of classifiers. In addition, it also advances some of the aspects most often alluded to in multilabel learning, as is taking advantage of label correlation information, the high-dimensionality problem, and the learning from imbalanced data task.

2.4.1   The Data Transformation Approach

MLC is a harder task than traditional classification, since the classifier has to predict several outputs at once. One of the first approaches to arise for solving this problem was to transform it, producing one or more simpler problems. The transformation idea is all about to get an MLD and generate datasets that can be processed by binary or multiclass classifiers. Commonly, the output produced by those classifiers has to be backtransformed in order to obtain the multilabel prediction.

Some of the simplest methods originally proposed in the literature are the ones described below:

- **Selecting a single label**: It is the transformation named *Model-s* in [5], *s* standing for *single label*. When a sample is associated with a set of labels, one of them is chosen as single class. This selection can be random or be based on some heuristic method. The result produced by this transformation is a multiclass dataset having the same number of instances than the original one, but each sample has only one class.
- **Ignoring multilabel instances**: Dismissing all the samples associated with more than one label, this transformation obtains a new dataset of multiclass nature, with only one label per instance. It is introduced as *Model-i* in [5], *i* standing for *ignore*. The resulting dataset usually will have less instances than the original one, unless none of the samples had two or more labels. Since there is only one label per instance, any multiclass classifier can be used.

- **Unfolding samples with multiple labels**: Introduced in [53] as PT5 (*Problem Transformation 5*), this method decomposes each instance into as many instances as labels it contains, cloning the input attributes and assigning to each sample one of the labels. A weight can be optionally assigned to each label, depending on its distribution on the MLD. The output of this transformation is also a multiclass dataset, in this case containing more samples than the original one.
- **Using the labelset as class identifier**: Instead of discarding labels or samples, the method presented in [5] as *Model-n*, the *n* standing for *new class*, proposes using each different combination of labels (each labelset) as identifier of a new class. The resulting dataset has the same number of instances, but only one class. Therefore, it can be processed with any multiclass classifier. Nowadays, this transformation method is best known as LP (*Label PowerSet*).
- **Applying binarization techniques**: Binarization techniques were already used to deal with multiclass classification using binary classifiers; thus, they were a clear choice also for multilabel classification. The most usual approach, used in [35], consists in training k classifiers, each for one label, taking the instances in which the labels appear as positive and all the others as negative. Another proposal, called *cross-training* in [5], also trains k classifiers but prefers to use samples with multiple labels always as positive. The former way is currently known as BR (*Binary Relevance*), and it is possibly the most common transformation method.

For each specific transformation, a method to build the predicted labelset has to be defined. For instance, if the BR transformation is applied to the original MLD, the individual predictions obtained from the binary classifiers have to be joined making up the corresponding labelset.

In late years, the LP and BR transformations have been the foundation for multiple MLC solutions, including many of which are based on ensembles of binary and multiclass classifiers. Additional details related to these techniques will be provided in Chap. 4, along with some experimental results.

2.4.2 The Method Adaptation Approach

Automatic classification [26] has been a traditional DM task for years. Throughout this time, several families of algorithms [18] have been gradually developed, tested, and fine-tuned. These algorithms are an essential foundation to develop new, more specific ones, including those aimed at working with MLDs.

Some classification models were initially designed to solve binary problems and then extended to also consider multiclass cases [4, 31]. An example of this is SVM [9]. By contrast, other approaches are able to deal with several classes with great simplicity. A kNN classifier [20], for instance, can predict the most frequent class among

the neighbors of the new instance, whether there are two or more classes. Analogously, trees, classification rules, neural networks, and statistical models, among others, have been used to tackle both binary and multiclass classifications.

The lessons learned from adapting binary classifiers to the multiclass scenario, such as binarization, voting methods, and divide-and-conquer procedures, have been also useful while adapting proven algorithms to tackle multilabel classification. The main difficulty is to build a model capable of predicting several outputs at once. Again, some approaches are easily adaptable, whereas others require more effort. A kNN-based algorithm can take the multiple outputs present on its neighbors to elaborate its own prediction, but a SVM has to find the maximum margin boundaries between multiple labels.

> Proposals of new multilabel classifiers based on the adaptation approach have proliferated lately [34]. Many of them will be introduced in Chap. 5, and the most representative ones will be detailed and experimentally tested.

2.4.3 Ensembles of Classifiers

Classification ensemble is a widespread technique aimed to improve the performance obtained by individual classifiers. An ensemble is compounded by a set of classifiers, whose output is usually combined by weighted or unweighted averaging [24, 46]. The point is that a group of weak classifiers, with different biases, is able to perform better than a strong classifier, following the famous Wolpert's No Free Lunch Theorem [59]. One key aspect in ensemble construction is diversity. The more diverse the individual classifiers are, the more likely they have different biases. Diversity can be achieved through training homogeneous models with different data, for instance with bagging techniques [6], or alternatively by using heterogeneous classification models.

Ensembles of binary classifiers have been used to face multiclass classification [31], either by way of OVA or by way of OVO decompositions. Furthermore, ensembles are usually applied to fight with problems such as imbalance [32]. Therefore, it is not surprising that ensemble techniques are also applied in the multilabel field. In fact, it is the most usual approach to do multilabel classification.

> As will be shown in Chap. 6, a plethora of ensembles, made up of binary and multiclass classifiers, have been defined in the literature as a way to fulfill the multilabel classification needs. Some of the best-known proposals are ensembles of classifier chains [44] and ensembles of pruned sets [43]. Both will be further detailed along with other related ideas.

2.4.4 Label Correlation Information

Many of the published multilabel classification algorithms rely on a process to sim-
plify the original problem, producing several easier to confront ones. Through this
process, usually a complete independence between the labels is assumed. However,
most researchers highlight [3, 22, 52, 60] the importance to take into account label
dependency information. Allegedly, these correlations would help in designing better
classification models.

   The BR transformation method, as well as many algorithms based on it, builds an
individual classifier for each label. The classifiers are completely independent, so the
prediction made by one of them does not influence how the others make their work.
This is a first approach to label dependency, which consists in assuming that they
are fully independent. However, there are use cases where the presence of a certain
label could determine whether another one is also more likely to be present or not.
In scene labeling, for instance, the probability of the label `beach` would be higher
if the label `sea` is also relevant.

   A second mechanism is to implicitly incorporate label dependency information
into the classification process. The LP transformation, and some other algorithms
based on taking full or partial labelsets as class identifiers, follows this way. Since sets
of labels are treated like a unit, the dependency among them is implicitly embedded
into the model through the training process. A similar approach, but relying on binary
classifiers instead of multiclass ones, is the one based on chains of classifiers [44].
This technique introduces the label predicted by one classifier into the data given as
input to the next one, as will be detailed in Chap. 6.

   Explicit procedures for taking advantage of label correlation information have
been also developed. The authors of the CML (*Collectible Multilabel*) algorithm
[33], for instance, propose the use of conditional random fields to model correlations
between label pairs. The authors of [36] define conditional dependency networks
to capture correlations among labels. Similar statistical models are being used to
explicitly represent this information.

2.4.5 High Dimensionality

High dimensionality is a problem which affects multilabel classification even more
than it does in traditional classification. Usually, three dimensionality spaces are
considered:

- **Instance space**: There are MLDs with millions of instances, but this is something
  increasingly common in all contexts, including traditional classification, due to
  the rise of big data.
- **Input feature space**: Although there are binary and multiclass datasets with many
  input features, MLDs stand out in this area as they usually have dozens of thousands

of features. The dimensionality of this space is what mostly makes difficult the learning process.

- **Output attribute space**: Before multilabel classification was considered, the classifiers only had to deal with one output attribute. Whether it was binary or multivalued, there was only one output, always. By contrast, all MLDs have several outputs and many of them have thousands, one for each label. Due to this fact, the dimensionality treatment in MLDs is a more complex topic than in traditional classification.

That learning from a high-dimensional input space (i.e., a large number of features) imposes serious difficulties is something well known. Even there is an ad hoc expression to name this problem, *the curse of dimensionality*. Therefore, it is an obstacle deeply studied and analyzed in DM, with dozens of different proposals to face it to some extent [51].

There are non-supervised methods, such as Principal Component Analysis (PCA) [39] and Latent Semantic Indexing (LSI) [27], able to reduce dimensionality by analyzing only the input features. These can be applied to MLDs, since no knowledge about the output variables is needed. The supervised approaches, characterized for using the output attributes to infer which input features provide more useful information, have to be adapted prior to be used in the multilabel context.

> Dimensionality in the label space is the least studied problem until now, with only a few proposals. Most of them will be detailed in Chap. 7, specifically devoted to this topic.

2.4.6 Label Imbalance

The learning from imbalanced data is another of the casuistics intrinsically linked to multilabel classification. Like high dimensionality in the input feature space, imbalance is also a well-studied problem in traditional classification. There are many proven techniques to face imbalance in binary classification, and many of them have been adapted to work with multiclass datasets. A few of these techniques will be described in Chap. 8, including how some of them have been adjusted to deal with MLDs.

The differences among label distributions in MLDs arise firstly by their own nature. In an MLD with thousands of labels, for instance categorizing text documents, it is usual that some of them are much more frequent than others. More specialized documents would have rare labels, which will be in minority against the most common ones.

Secondly, the transformations applied to multilabel data tend to increase the imbalance between labels. This is the case of BR. Taking as positive only the instances in which a certain label appears and as negative all other samples, the original distri-

bution changes affecting the representation of the considered label. The situation is even worse if this label is already a minority label.

Imbalance in multilabel classification is one of the specificities more recently tackled in the literature. Specific metrics [13] and several methods to balance label distributions [14, 16, 23] have been proposed. Most of them will be presented in Chap. 8.

2.5   Multilabel Data Tools

To conclude this chapter, in which the topics of what multilabel classification is, what it is used for, and how it has been faced have been dealt, the main tools currently used to work with this kind of information are briefly introduced. They will be explained in more detail in Chap. 9.

Although multilabel data can be analyzed and multilabel classification conducted with any ad hoc software, specifically designed to accomplish a certain task, there are also available some generic tools. Many researchers and practitioners have relied on them in late years. The most noteworthy are the following:

- **MULAN**: Presented in 2010 [55], it is the most mature multilabel software and probably the most widely used. MULAN is an open source library written in Java. It provides a programming interface (API) to help in the development of multilabel classification applications, including also evaluation of results and label ranking generation. Several of the most common transformations and MLC algorithms are already included, ready to use. Unlike WEKA [37], the tool it is based on, MULAN does not offer a graphic user interface (GUI) to ease the accomplishment of these tasks.
- **MEKA**: As MULAN, MEKA [45] is also founded on WEKA, but unlike MULAN it supplies a GUI from which the user can load MLDs, perform some exploration tasks, and apply different MLC algorithms. Since the GUI is very similar to that of WEKA, WEKA users will immediately feel comfortable with this tool. It is also open source and has been developed in Java.
- **mldr**: R is one of the most used tools to explore and analyze data nowadays, but it lacked the ability to present the specificities of MLDs. The mldr [10] package adds this functionality to R, providing functions to load, modify, and write MLDs, as well as general exploratory analysis methods. The package also includes a Web-based GUI from which most tasks can be easily performed.
- **mldr.datasets**: Built upon the infrastructure provided by the preceding one, this R package [11] eases the process of obtaining multilabel datasets. In addition, it provides methods aimed to retrieve basic information about them, including citation data, as well as to facilitate the partitioning process.

- **scikit-multilearn**: Based on the well-known scikit-learn Python module, scikit-multilearn[3] is an extension still in early development. The current version is 0.0.1 and it provides BR, LP, and RAkEL implementations, with many other methods whose development is in progress.

As mentioned above, the details about how to use these tools to explore MLDs and conduct experiments with them are the main topic of Chap. 9.

References

1. Alcala-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: KEEL multi-label dataset repository. http://sci2s.ugr.es/keel/multilabel.php
2. Alcala-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: KEEL data-mining software tool: data set repository and integration of algorithms and experimental analysis framework. J. Mult-valued Log. Soft Comput. **17**(2–3), 255–287 (2011)
3. Alvares-Cherman, E., Metz, J., Monard, M.C.: Incorporating label dependency into the binary relevance framework for multi-label classification. Expert Syst. Appl. **39**(2), 1647–1655 (2012)
4. Aly, M.: Survey on multiclass classification methods. In: Technical Report, pp. 1–9. California Institute of Technology (2005)
5. Boutell, M., Luo, J., Shen, X., Brown, C.: Learning multi-label scene classification. Pattern Recogn. **37**(9), 1757–1771 (2004)
6. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
7. Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J.K., Hadley, A.S., Betts, M.G.: Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach. J. Acoust. Soc. Am. **131**(6), 4640–4650 (2012)
8. Brinker, K., Hüllermeier, E.: Case-based multilabel ranking. In: Proceedings of 20th International Joint Conference on Artificial Intelligence, IJCAI'07, pp. 702–707. Morgan Kaufmann (2007)
9. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)
10. Charte, F., Charte, D.: Working with multilabel datasets in R: the mldr package. R J. **7**(2), 149–162 (2015)
11. Charte, F., Charte, D., Rivera, A.J., del Jesus, M.J., Herrera, F.: R Ultimate multilabel dataset repository. In: Proceedings of 11th International Conference on Hybrid Artificial Intelligent Systems, HAIS'16, vol. 9648, pp. 487–499. Springer (2016)
12. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: Multilabel classification. In: Problem Analysis, Metrics and Techniques Book Repository. https://github.com/fcharte/SM-MLC
13. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: Addressing imbalance in multilabel classification: measures and random resampling algorithms. Neurocomputing **163**, 3–16 (2015)
14. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: MLSMOTE: approaching imbalanced multilabel learning through synthetic instance generation. Knowl.-Based Syst. **89**, 385–397 (2015)
15. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: QUINTA: a question tagging assistant to improve the answering ratio in electronic forums. In: Proceedings of IEEE International Conference on Computer as a Tool, EUROCON'15, pp. 1–6. IEEE (2015)
16. Chen, K., Lu, B., Kwok, J.: Efficient classification of multi-label and imbalanced data using min-max modular classifiers. In: Proceedings of IEEE International Joint Conference on Neural Networks, IJCNN'06, pp. 1770–1775 (2006)

---

[3]http://scikit.ml.

17. Chen, X., Zhan, Y., Ke, J., Chen, X.: Complex video event detection via pairwise fusion of trajectory and multi-label hypergraphs. Multimedia Tools Appl. 1–22 (2015)

18. Cherkassky, V., Mulier, F.: Learning from Data: Concepts. Theory and Methods. Wiley-IEEE Press (2007)

19. Cong, H., Tong, L.H.: Grouping of triz inventive principles to facilitate automatic patent classification. Expert Syst. Appl. **34**(1), 788–795 (2008)

20. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theory **13**(1), 21–27 (1967)

21. Crammer, K., Dredze, M., Ganchev, K., Talukdar, P.P., Carroll, S.: Automatic code assignment to medical text. In: Proceedings of Workshop on Biological, Translational, and Clinical Language Processing, BioNLP'07, pp. 129–136. Association for Computational Linguistics (2007)

22. Dembszynski, K., Waegeman, W., Cheng, W., Hüllermeier, E.: On label dependence in multi-label classification. In: ICML Workshop on Learning from Multi-label data, pp. 5–12 (2010)

23. Dendamrongvit, S., Kubat, M.: Undersampling approach for imbalanced training sets and induction from multi-label text-categorization domains. In: New Frontiers in Applied Data Mining, *LNCS*, vol. 5669, pp. 40–52. Springer (2010)

24. Dietterich, T.: Ensemble methods in machine learning. In: Multiple Classifier Systems. *LNCS*, vol. 1857, pp. 1–15. Springer (2000)

25. Diplaris, S., Tsoumakas, G., Mitkas, P., Vlahavas, I.: Protein classification with multiple algorithms. In: Proc. 10th Panhellenic Conference on Informatics, PCI'05, vol. 3746, pp. 448–456. Springer (2005)

26. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. John Wiley (2000)

27. Dumais, S., Furnas, G., Landauer, T., Deerwester, S., Deerwester, S., et al.: Latent semantic indexing. In: Proceedings of 4th Text Retrieval Conference, TREC-4, pp. 105–115. NIST (1995)

28. Duygulu, P., Barnard, K., de Freitas, J., Forsyth, D.: Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In: Proceedings of 7th European Conference on Computer Vision, ECCV'02, vol. 2353, pp. 97–112. Springer (2002)

29. Elisseeff, A., Weston, J.: A kernel method for multi-labelled classification. In: Advances in Neural Information Processing Systems, vol. 14, pp. 681–687. MIT Press (2001)

30. Fürnkranz, J., Hüllermeier, E., Loza Mencía, E., Brinker, K.: Multilabel classification via calibrated label ranking. Mach. Learn. **73**, 133–153 (2008)

31. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. Pattern Recogn. **44**(8), 1761–1776 (2011)

32. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: A review on ensembles for the class imbalance problem: bagging, boosting, and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(4), 463–484 (2012)

33. Ghamrawi, N., McCallum, A.: Collective multi-label classification. In: Proceedings of 14th ACM International Conference on Information and Knowledge Management, CIKM'05, pp. 195–200. ACM (2005)

34. Gibaja, E., Ventura, S.: A tutorial on multi-label learning. ACM Comput. Surv. **47**(3) (2015)

35. Gonçalves, T., Quaresma, P.: A preliminary approach to the multilabel classification problem of Portuguese juridical documents. In: Proceedings of 11th Portuguese Conference on Artificial Intelligence, EPIA'03, pp. 435–444. Springer (2003)

36. Guo, Y., Gu, S.: Multi-label classification using conditional dependency networks. In: Proceedings of 22th International Joint Conference on Artificial Intelligence, *IJCAI'11*, vol. 2, pp. 1300–1305 (2011)

37. Holmes, G., Donkin, A., Witten, I.H.: WEKA: a machine learning workbench. In: Proceedings of 2nd Australian and New Zealand Conference on Intelligent Information Systems, ANZIIS'02, pp. 357–361 (2002)

38. Hüllermeier, E., Fürnkranz, J., Cheng, W., Brinker, K.: Label ranking by learning pairwise preferences. Artif. Intell. **172**(16), 1897–1916 (2008)

39. Jolliffe, I.: Introduction. In: Principal Component Analysis, pp. 1–7. Springer (1986)
40. Katakis, I., Tsoumakas, G., Vlahavas, I.: Multilabel text classification for automated tag suggestion. In: Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML PKDD'08, pp. 75–83 (2008)
41. Klimt, B., Yang, Y.: The enron corpus: a new dataset for email classification research. In: Proc eedings of 15th European Conference on Machine Learning, ECML'04, pp. 217–226. Springer (2004)
42. Lewis, D.D., Yang, Y., Rose, T.G., Li, F.: RCV1: a new benchmark collection for text categorization research. J. Mach. Learn. Res. **5**, 361–397 (2004)
43. Read, J., Pfahringer, B., Holmes, G.: Multi-label classification using ensembles of pruned sets. In: Proc eedings of 8th IEEE International Conference on Data Mining, ICDM'08, pp. 995–1000. IEEE (2008)
44. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. Mach. Learn. **85**, 333–359 (2011)
45. Read, J., Reutemann, P.: MEKA multi-label dataset repository. http://sourceforge.net/projects/meka/files/Datasets/
46. Rokach, L.: Pattern classification using ensemble methods. World Scientific (2009)
47. Salton, G., Fox, E.A., Wu, H.: Extended Boolean information retrieval. Commun. ACM **26**(11), 1022–1036 (1983)
48. Snoek, C.G.M., Worring, M., van Gemert, J.C., Geusebroek, J.M., Smeulders, A.W.M.: The challenge problem for automated detection of 101 semantic concepts in multimedia. In: Proceedings of 14th ACM International Conference on Multimedia, MULTIMEDIA'06, pp. 421–430 (2006)
49. Sobol-Shikler, T., Robinson, P.: Classification of complex information: Inference of co-occurring affective states from their expressions in speech. IEEE Trans. Pattern Anal. Mach. Intell. **32**(7), 1284–1297 (2010)
50. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: Proceedings of 33rd international ACM SIGIR conference on Research and development in information retrieval, pp. 841–842. ACM (2010)
51. Sun, L., Ji, S., Ye, J.: Multi-label dimensionality reduction. CRC Press (2013)
52. Tenenboim-Chekina, L., Rokach, L., Shapira, B.: Identification of label dependences for multi-label classification. In: Proceedings of 2nd International Workshop on Learning from Multi-Label Data, MLD'10, pp. 53–60 (2010)
53. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining multi-label data. In: Data Mining and Knowledge Discovery Handbook, pp. 667–685. Springer (2010)
54. Tsoumakas, G., Xioufis, E.S., Vilcek, J., Vlahavas, I.: MULAN multi-label dataset repository. http://mulan.sourceforge.net/datasets.html
55. Tsoumakas, G., Xioufis, E.S., Vilcek, J., Vlahavas, I.: MULAN: a java library for multi-label learning. J. Mach. Learn. Res. **12**, 2411–2414 (2011)
56. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.: Semantic annotation and retrieval of music and sound effects. IEEE Trans. Audio Speech Lang. Process. **16**(2), 467–476 (2008)
57. Vembu, S., Gärtner, T.: Label ranking algorithms: a survey. In: Preference learning, pp. 45–64. Springer (2011)
58. Wieczorkowska, A., Synak, P., Raś, Z.: Multi-label classification of emotions in music. In: Intelligent Information Processing and Web Mining. AISC, vol. 35, Chap. 30, pp. 307–315 (2006)
59. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. IEEE Trans. Evol. Comput. **1**(1), 67–82 (1997)
60. Zhang, M., Zhang, K.: Multi-label learning by exploiting label dependency. In: Proceedings of 16th International Conference on Knowledge Discovery and Data Mining, ACM SIGKDD'10, pp. 999–1008 (2010)