

Chapter 2

Multiple Instance Learning

Abstract This chapter provides a general introduction to the main subject matter of this work: multiple instance or multi-instance learning. The two terms are used interchangeably in the literature and they both convey the crucial point of difference with traditional (single-instance) learning. A formal description of multiple instance learning is provided in Sect. 2.1 and we discuss its origins in Sect. 2.2. In Sect. 2.3, we describe different learning tasks within this domain, which may or may not have an equivalent in single-instance learning. Finally, Sect. 2.4 lists a wide variety of applications corresponding to the different multi-instance learning paradigms.

2.1 Formal Description

The traditional data description presented in Chap. 1 corresponds to so-called *single-instance learning*, where each observation or learning object is described by a number of feature values and, possibly, an associated outcome. In our object of study, *multiple-instance learning (MIL)*, the structure of the data is more complex. In this setting, a learning sample or object is called a *bag*. The defining feature of MIL is that a bag is associated with multiple instances or descriptions. Each instance is described by a feature vector, as we saw in single-instance learning, but an associated outcome is never reported. The only information available about an instance, aside from its feature values, is its membership relationship to a bag.

Formally, an instance x corresponds to a point in the instance space \mathbb{X} . It is commonly assumed that $\mathbb{X} \subseteq \mathbb{R}^d$, that is, each instance is described by a vector of d real-valued numbers, its feature values. However, as described in Sect. 1.1, datasets often contain mixed types of features. To model these situations, \mathbb{X} can be generalized to $\mathbb{X} \subseteq \mathcal{A}^d = \mathcal{A}_1 \times \dots \times \mathcal{A}_d$, such that each instance is described by a d -dimensional vector, where each attribute \mathcal{A}_i ($i = 1, \dots, d$) takes on values from a finite or infinite set \mathcal{Y}_i . In this way, we can deal with mixed feature sets in which some of the features are categorical and others are numeric.

A bag X is a collection of n instances, where every instance x_i is drawn from the instance space \mathbb{X} . Each bag is allowed to have a different size, which means that the value n can vary among the bags in the dataset. Multiple copies of the same instance

Table 2.1 Structure of a multi-instance dataset with M bags

Bags	Instances	\mathcal{A}_1	\mathcal{A}_2	...	\mathcal{A}_d	Outcome
X_1	$x_{1,1}$	$x_{1,1,1}$	$x_{1,1,2}$...	$x_{1,1,d}$	y_1
	
	x_{1,n_1}	$x_{1,n_1,1}$	$x_{1,n_1,2}$...	$x_{1,n_1,d}$	
...
X_M	$x_{M,1}$	$x_{M,1,1}$	$x_{M,1,2}$...	$x_{M,1,d}$	y_M
	
	x_{M,n_M}	$x_{M,n_M,1}$	$x_{M,n_M,2}$...	$x_{M,n_M,d}$	

can be included in a bag. For this reason, many authors define a bag as $X \in \mathbb{N}^{\mathbb{X}}$, that is, a multi-set containing elements from \mathbb{X} such that duplicates can occur. Different bags are also allowed to overlap and contain copies of the same instance. This forms an indication of the higher level of complexity of MIL compared to single-instance learning. Throughout this work, we use lowercase letters to represent instances (e.g., x , a , b) and uppercase letters to represent bags (e.g., X , A , B).

As an example, Table 2.1 presents the general structure of a multi-instance dataset. The first column represents the bags, sometimes also referred to as exemplars. Each bag contains a number of instances, represented in the second column. Each instance identifier corresponds to a vector description, of which the attribute values are arranged from columns \mathcal{A}_1 to \mathcal{A}_d . The first instance $x_{1,1}$ in the first bag X_1 is for example represented by the feature vector $\langle x_{1,1,1}, x_{1,1,2}, \dots, x_{1,1,d} \rangle$. The last column represents the outcome associated with the bag. It is important to stress that this outcome is only known for a bag as a whole and not for each individual instance. Depending on the learning task (see Sect. 2.3), the outcome may be a class label (classification) or a real value (regression). In clustering applications, there are no outcome values available. We briefly note that the work of [11] showed that the performance of multi-instance learners on datasets with very similar meta-characteristics, like dimensionality and size, can be very different.

2.2 Origin of MIL

The multi-instance learning paradigm was introduced in the seminal work of [16]. It arose in the context of learning tasks where data observations (bags) can have different alternative descriptions (instances). The authors of [16] focused on an application in biochemistry: the drug activity prediction problem. Here, the task is to predict whether or not a given molecule is a good drug molecule, which is measured by its ability to bind to a given target. Each molecule can be represented as a bag, of which the instances correspond to different conformations (molecular structures) of that particular compound. Figure 2.1 depicts this situation for a butane molecule. In

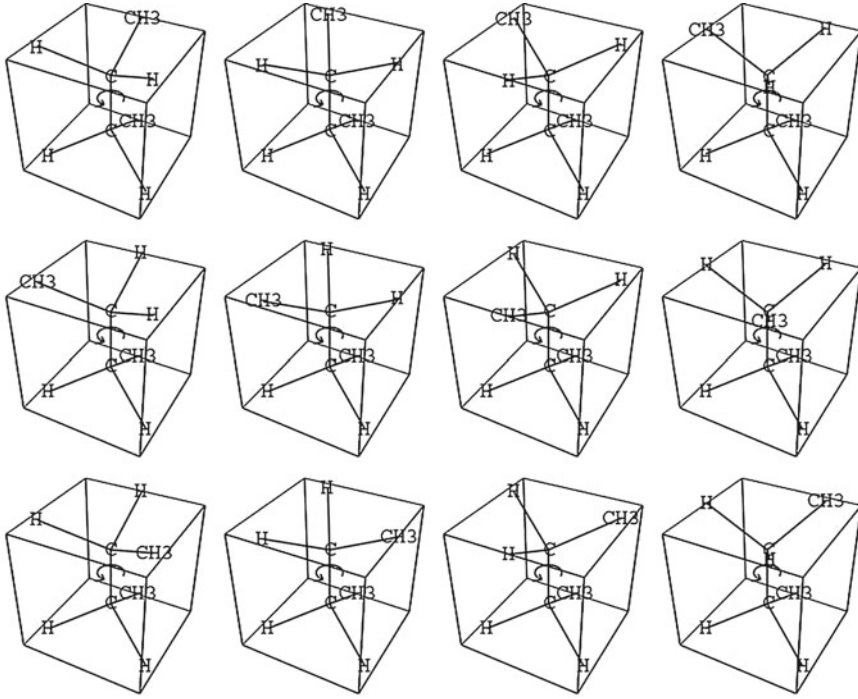


Fig. 2.1 Conformations of a butane molecule

this case, butane would be represented by a bag containing the 12 listed shapes as its instances.

MIL emerged as an extension of supervised learning. bag-instances relationship models the one-to-many relation characteristic of relational databases, since one bag can contain several different instances. More than an extension, MIL can therefore be considered a generalization of single-instance learning and the latter can be understood as a special case of MIL where each bag contains a single instance. Moreover, MIL has proven to be a bridge between two different paradigms: propositional learning on the one hand and relational learning on the other.

2.2.1 Relationship with Propositional Learning

Propositional or attribute-value learning corresponds to the setting described in Sect. 1.1, where the training data is ordered in a single flat table. In single-instance semi-supervised learning (Sect. 1.3.3), only part of the instance outcomes are available and it therefore shows a certain similarity with MIL, where the outcomes are only known for the bags and not their instances. However, there is a fundamental

difference between the two: the relationship between instances and bags in MIL does not exist in semi-supervised learning. In the latter, labeled instances are at the same level as unlabeled instances and there is no specific relationship between them. In MIL on the contrary, a secondary structure is present in the dataset, defining the two different levels of bags and instances. All instances in a bag are somehow interrelated, because of their shared membership to the bag.

2.2.2 *Relationship with Relational Learning*

In relational learning, structured concept definitions are derived from structured training examples [14]. The training data models the different observations as well as the relations between them, for instance by using multiple tables. A clear example is given in [15], where the relational data is represented by two tables, one providing the description of store customers and the other the marital relations between them.

Many learning methods have been developed for propositional learning, but these can only be applied to data organized in a single table and the relations between different observations can not be taken into account. Propositional algorithms can therefore not be directly applied in relational learning problems. Relational data can be transformed into an attribute-value table in a process called propositionalization, but this implies a steep computational cost and its application to real problems is limited as a result of an internal combinatorial explosion [47].

MIL has come to be considered as the missing link between relational and propositional learning, because, as stated above, the bag label models a one-to-many relationship. The contribution of [13] shows that multi-instance problems can also be considered as a special case of inductive logic programming [37]. All inductive logic programming problems (in the form of relational databases) can be transformed by database join operations in a single one-to-many relationship. Such a relation can in turn be naturally represented as a MIL problem [47, 48]. As will be discussed in later chapters, many single-instance learning algorithms have already been adapted to the multi-instance setting. This feature of MIL allows for many relational learning problems to be solved by traditional supervised learning methods.

2.3 MIL Paradigms

As in traditional single-instance learning, discussed in Sect. 1.3, we can distinguish between a number of learning tasks within MIL. In Sect. 2.3.1 we discuss the two supervised learning settings, classification and regression. Section 2.3.2 describes multi-instance clustering. Several other traditional learning tasks, like semi-supervised or multi-label learning, can find a corresponding MIL equivalent (e.g., [44, 82]). However, we must warn the reader that this general similarity between single-instance and multi-instance learning tasks can not be transferred to their solution

methods. Due to the relational nature, MIL solution methods are inherently more complex. This also implies that some MIL tasks have no related single-instance setting. The most prominent example is presented in Sect. 2.3.3.

2.3.1 Multi-instance Classification and Regression

In a *multi-instance classification problem*, the goal is to determine the class label of new bags, based on the class labels in the training set or, more specifically, using a prediction model built on the labeled training bags. The outcome associated with the training bags is categorical.

More formally, in a classification problem, we deal with a training set $D = (\mathbf{X}, \mathbf{L})$, where $\mathbf{X} = \langle X_1, \dots, X_m \rangle$ is a set of bags and $\mathbf{L} = \langle \ell_1, \dots, \ell_m \rangle$ a set of class labels, with $\ell_i \in \mathbb{L}$ ($i = 1, \dots, m$) and \mathbb{L} the finite set of all possible class labels. The bag X_i is assigned the class label ℓ_i . Recall that only the class labels of the bags are known and not those of the instances inside them. Later on in this work, we provide a detailed discussion on the contribution of the individual instances to the bag label. Traditionally, MIL has focused on two-class classification problems, dealing with one positive and one negative class. However, in general the number of classes can be larger, that is, $|\mathbb{L}| \geq 2$. The classification objective is to find a function $\mathcal{H} : \mathbb{N}^{\mathbf{X}} \rightarrow \mathbb{L}$ based on the training set D . This function is the classification model and is used to predict the class labels of new bags as accurately as possible. More details on multi-instance classification will be provided in Chap. 3.

When the outcomes are known for all training bags, but they correspond to real values rather than class labels, we are dealing with a *multi-instance regression problem*. The data description is highly similar to the one for classification data. The main difference is that the bag class labels are replaced by numerical values, that is, \mathbb{L} corresponds to a range of values in \mathbb{R} rather than to a finite set. Multi-instance regression was proposed in [2, 46], independently at the same conference. This task is discussed further in Chap. 6.

2.3.2 Multi-instance Clustering

As discussed in Sect. 1.3.2, clustering is situated in the unsupervised learning domain. The set of outcomes \mathbf{L} associated to the training bags \mathbf{X} in D is not known or not available. The goal is to group these unlabeled bags based on a given similarity measure. A multi-instance clustering method determines a set of groups $\mathcal{G} = \{G_1, \dots, G_k\}$ and a function $\mathcal{H} : \mathbb{N}^{\mathbf{X}} \rightarrow \mathcal{G}$ which assigns bags to groups such that it minimizes the similarity differences between bags of the same group and maximizes the similarity differences between bags of different groups. The choice of an appropriate

similarity measure is crucial in multi-instance clustering. As noted in [74], not all instances within a bag contribute equally to the bag prediction, which implies that the bags should ideally not be considered as collections of independent instances in the definition of the similarity metric. Multi-instance clustering is discussed in more detail in Chap. 7.

2.3.3 Instance Annotation

An important task in some MIL applications, which has no counterpart in single-instance learning, is the instance-level classification. In this setting, apart from predicting a class label for a new bag, the assignment of class labels to its instances is a key objective as well. Depending on the application, there are two possible cases.

In the first situation, given the training set $D = (\mathbf{X}, \mathbf{L})$, the objective is to locate the instance or instances that are key to determining the class of the bag. In general, key instances are considered those that are more likely to have the same (hidden) label as their bag. A function $h : \mathbb{X} \rightarrow \mathbb{L}$ is constructed, such that the corresponding aggregation function $H(h(x_1), \dots, h(x_n)) \rightarrow \mathbb{L}$ can predict class labels of a new bag $X = \{x_1, \dots, x_n\}$ with maximum possible accuracy. This learning strategy is employed by a large group of multi-instance classification algorithms, described in Chap. 4. Some applications require the identification of key instances not only to classify bags, but also because these instances are themselves relevant to the application (e.g., [30]). An example application where the identification of true positive instances is very informative, is that of the stock selection problem [33]. In that setting, true positive instances correspond to stocks that fundamentally perform well, which is an important subgroup to discern from the other stocks.

In the second case, the training set is represented as $D = (\mathbf{X}, \mathbf{L})$, where $\mathbf{X} = \langle X_1, \dots, X_m \rangle$ are bags and $\mathbf{L} = \langle \mathcal{L}_1, \dots, \mathcal{L}_m \rangle$ are sets of instance labels associated to the bags. In this situation, the set $\mathcal{L}_i = \{\lambda_1, \dots, \lambda_{k_i}\}$ of explicit instance labels is assigned to the bag X_i . These labels are drawn from a set $\Lambda = \{\lambda_1, \dots, \lambda_s\}$, which can be different from \mathbb{L} . Unlike the traditional MIL approach, some instance labels are known for each bag. The objective is to find a function that, given a new bag, allows us to find instance labels that best describe it. This setting is very popular in applications such as image annotation (e.g., [7]), where the annotation of image segments (instances) can result in a global label for the complete image (bag). Since one observation (bag) is associated with a set of (instance) labels, this approach shows some similarity with multi-label classification (Sect. 1.3.1). However, multi-label and multi-instance learning remain different paradigms. The former represents each observation by multiple instances and a single global class label, while in the latter an observation corresponds to one instance associated with several labels.

2.4 Applications of MIL

In MIL, a more complex structure of data observations can be represented. The multi-instance setting is required to model several real-world applications that we list in this section. There is an inherent level of representation ambiguity in this type of problems and we can distinguish between several sources. MIL data naturally arises in the following situations:

- **Alternative representations:** different views, appearances or descriptions of the same object are available. A classical example in this case is that of drug activity prediction, the application for which MIL was originally developed in [16] (see also Sect. 2.2).
- **Compound objects:** a compound object consists of several parts. In the example of image recognition, an image corresponds to a bag and each image segment forms an instance. An example is found in Fig. 2.2. The image segments can correspond to different breakfast components like the slice of toast, the sausage, the beans, and so on. Together, they form a full English breakfast.
- **Evolving objects:** in these applications, an evolving object is sampled at different time intervals. This is also referred to as a time-series problem. The bag represents the object, while the time point samples are its instances. An example is the study around the use of MIL in bankruptcy prediction presented in [27].



Fig. 2.2 A full English breakfast

The main research focus within the MIL community has been on multi-instance classification problems. A variety of application domains are listed in Sects. 2.4.1–2.4.6. In Sect. 2.4.7, we consider applications of multi-instance regression, while multi-instance clustering applications are discussed in Sect. 2.4.8.

2.4.1 *Bioinformatics*

We have already discussed the application of *drug activity prediction* in Sect. 2.2. Each bag corresponds to a molecule and its instances are the different molecular shapes, as shown in Fig. 2.1. The objective in the original MIL proposal [16] is the prediction of musky and non-musky molecules. Other drug activity problems concern the mutagenicity prediction of compound molecules [52] and activity prediction of molecules as anticancer agents [6]. Studies like [21, 33, 72, 80] address the drug activity prediction problem with their proposed multi-instance classifiers as well.

Another bioinformatics application of MIL is the *protein identification* task, like the recognition of Thioredoxin-fold proteins, as explored in, e.g., [45, 55, 59]. Binding proteins of the Calmodulin protein are identified in a multi-instance classification process in [36], while the application in [40] is the prediction of binding peptides for the highly polymorphic MHC class II molecules. In [29], multi-instance multi-label classification is used to automate the annotation of gene expression patterns. This method was evaluated on *Drosophila melanogaster* (fruit fly).

2.4.2 *Image Classification and Retrieval*

Another widely studied MIL application area is that of *image classification*, where the goal is to, given an image, decide on what it represents or to which of a given set of categories it belongs. As an example, consider the early work of [34] that revolves around the classification of natural scene images, e.g., images of waterfalls. In the data representation, an image corresponds to a bag. The instances within this bag are subimages, encoded as templates describing color and spatial characteristics of that specific region. The subimages can be obtained by a partitioning process or, possibly more appropriately, an image segmentation procedure. In a perfect segmentation, the resulting regions correspond to individual objects. The classification objective is to predict what the complete image represents. If we consider Fig. 2.2, a multi-instance classifier should derive that it is processing an image of a full English breakfast based on the different objects on the plate. This type of region-based image categorization was also evaluated in [3, 9, 10, 24, 42], although not all of these referenced works developed multi-instance classification methods specific for image data. They often consider more general algorithms and evaluate them on a variety of applications. Multi-instance image datasets have indeed become popular benchmarks to evaluate new proposals on. One specific type of image classification, facial recognition, where

a bag of instances can represent images taken of the same person from different angles, was studied in, e.g., [8, 19].

More complex models for the mapping of images to multi-instance data were studied in later works. The method of [43] models the interrelations of instances (regions) in a bag (image) to improve the categorization process, while [25] considers image annotation by means of a joint multi-instance mapping and feature selection process. The recent proposal of [20] develops a multi-instance semi-supervised classification method based on sparse representation and evaluates it on image data.

A task related to image categorization is that of *image retrieval*. The aim in this case is to obtain images from a dataset that are semantically relevant to the user, based on his specified query or presented examples of images of interest. Multi-instance approaches to this challenge represent, as above, an image as a bag, containing many of its subimages as instances. Examples can be found in, e.g., [7, 66, 71, 75–77].

2.4.3 *Web Mining and Text Classification*

Another application domain of MIL lies in web mining. The web index recommendation problem was introduced as a multi-instance problem in [81]. In this application, a bag corresponds to a web index page and its instances refer to other websites to which the page links. The recommendation task is to suggest relevant web pages to users based on their browser history. Such knowledge is useful for the construction of intelligent web browsers. This problem domain was also the central focus of [67, 69], in which genetic programming algorithms were developed to solve it. In [51], a multi-instance classifier based on the Rocchio classifier [49] was developed for this application.

A related task is that of document classification. In [3], the proposed multi-instance classification method is evaluated on a document categorization problem. In this case, a bag corresponds to a document and the instances are particular passages within that document. In the experiments of [45], the dataset obtained in the biomedical study of [5] is used. A bag corresponds to a biomedical article about a particular protein and the instances are the paragraphs of the text. A positive bag is one that can be labeled with a Gene Ontology code, while a negative bag cannot. The classification goal is to discern between positive and negative bags.

2.4.4 *Object Detection and Tracking*

This domain requires methods that discern an object of interest in image or video data. Examples are the application of the proposed multi-instance boosting method to horse detection and pedestrian detection in [1]. In [32], the detection of landmines based on radar images is studied in a multi-instance classification context. The study of [61] considers the related aspect of saliency detection, which is the detection of

the object in the image that draws the visual attention, as humans focus more on some parts of pictures than on others. It is not known in advance what the object is, only that it draws the attention of the observer.

In an object tracking application, a specific object is followed during the course of a video sequence. Online methods have been proposed in, e.g., [4, 73]. In the recent contributions of [31, 83], online multi-instance boosting algorithms for visual object tracking problems are developed.

2.4.5 Medical Diagnosis and Imaging

Several studies on multi-instance data focus on applications within the medical domain. In [22], a multi-instance classification framework is developed for computer-aided medical diagnosis, like the detection of tumors. It is shown that the use of this framework significantly improves the diagnostic accuracy in the evaluated applications. The study of [53] concerns the automatic detection of myocardial infarction based on electrocardiography (ECG) recordings. For each patient, a 24-h ECG is taken, which traces his or her heart activity for a full day. Such a recording is too large to be interpreted by a cardiologist. Automated prediction tools are required to detect any heart abnormalities in the data. In the input data for the multi-instance classifier, a bag corresponds to a full ECG, while each instance represents a recorded heartbeat.

The proposal of [41] studies the early detection of illnesses, like frailty and dementia, in senior citizens. This is done in a noninterfering way, namely by using sensor data, collected from a number of sensors monitoring elderly people in nursing homes. A bag consists of 24 hourly sensors measurements (instances) taken in one day for a single patient. The label of a bag is determined based on the report made by a nurse for the patient on that particular day. It indicates whether the patient exhibited health problems (positive) or not (negative).

A fourth study [60] develops a multi-instance classification algorithm for the detection of colonic polyps, abnormal growths in the colon. It revolves around video classification. When a possible polyp is present in the colon, images of it are collected from several viewpoints and combined into a video. Each candidate polyp consequently corresponds to a bag. The different viewpoints or video frames are the instances. The prediction aim is to decide whether the videoed candidate is an actual polyp or not.

2.4.6 Other Classification Applications

In this final section on applications of multi-instance classification, we collect a number of miscellaneous applications that do not fall within any of the categories listed in the previous sections.

Multi-instance classification has been applied to *prediction of student performance* [68]. This problem allows interesting relationships to be obtained that can suggest activities and resources to students and educators that favor and improve both learning and the effective learning process. From the MIL perspective, each student is regarded as a bag which represents the work carried out and is composed of one or several instances where each instance represents the different types of work that the student has done. This representation has shown better results than traditional single-instance representation [68]. The work of [70] proposes a genetic programming model to solve this problem more efficiently.

The study of [35] proposes a method for automatic *subgoal discovery* in reinforcement learning [54]. The trajectory of an agent in a reinforcement learning process is encoded as a bag. The observations made along this trajectory are the instances. The bag label states whether the trajectory is successful or not, where the definition of success depends on the problem description.

Multi-instance classification has been applied to several *computer-related tasks* as well, for instance in the work of [50] that focused on computer security applications. Impending failure of computer hard drives is predicted in [38]. A bag corresponds to a single drive and its instances are observations of this drive taken at different time points. In [26], the quality of object-oriented software is estimated. A class hierarchy is transformed into a bag, containing the constituent classes as instances.

The proposed classification method of [33] was evaluated on a *stock selection problem*. In this work, each bag represents a month of trading. A positive bag contains the 100 stocks (instances) with the highest returns in that month, while a negative bag consists of the five stocks with the lowest returns.

The final classification application that we list, is graph mining, the process of extracting knowledge from graph structured data. *Multi-graph learning* is a further generalization of MIL, where every bag consists of several graphs. In MIL, all instances in the bags are drawn from the same feature space, but this is no longer the case in multi-graph learning. This area was the focus of the recent works [64, 65].

2.4.7 Regression Applications

Although to a lesser extent than for classification problems, we also encounter real-world applications of multi-instance regression. We collect these examples in this section.

The application referenced in one of the original proposals of multi-instance regression [46] is related to the *drug activity prediction problem*. Instead of treating this as a yes-or-no question, as done in the classification scenario, real-valued activity levels are estimated for the molecules. The second initial proposal on multi-instance regression [2] also interpreted drug activity prediction as a regression problem, where the binding strength of a molecule is the prediction objective. The theoretical study on multi-instance regression in [17] refers to the real-valued drug activity prediction problem as an important application as well. In [12], the authors develop a method

to predict the binding affinity of molecules based on their three-dimensional structure. They evaluate their method on thermolysin inhibitors, dopamine agonists, and thrombin inhibitors. In later work, [56] considers the prediction of protein-ligand affinities and [18] the prediction of the binding affinity of MHC class II molecules.

The study of [23] uses a real-valued outcome in the interval $[0, 1]$ to express the satisfaction degree of a bag to the concept. One of the evaluated applications is *landmark recognition* for robot vision. In a navigation assignment, robots are required to recognize whether or not they find themselves near one of a given set of landmarks.

Multi-instance regression has also been used in *remote sensing applications*. The contribution of [57] focuses on an agricultural process, namely the modeling of crop yield based on remote sensing data. A bag corresponds to one county in the United States. The instances in the bag are image pixels covering different parts of that county. The same application was evaluated in [58], where the authors developed a multi-instance regression method for structured data. In [62], a climate research application related to aerosols is considered. The prediction value is the so-called aerosol optical depth, which is a number related to the induced attenuation of radiation. This value characterizes aerosols and is central in the construction of climate models. Aerosols are globally monitored by satellites that provide data in the form of multi-spectral images. In this application, a bag corresponds to a set of neighboring pixels (instances) in such an image. The bag is labeled with an aerosol optical depth value. The two remote sensing applications, aerosol optical depth prediction and crop yield modeling, were also studied in [63].

Finally, we also list the multi-instance regression study of [39]. The authors develop a robust system for *age estimation* of a person based on an image of his or her face.

2.4.8 Clustering Applications

In this section, we review the applications for multi-instance clustering that have been presented in the literature. Recall that the goal of this learning paradigm is to arrange the bags in a number of well-separated groups of similar observations.

The proposal of [74] references an application in *biochemistry*. The execution of experiments to determine the functionality of specific molecules can be costly. Multi-instance clustering can be used in the often necessary step to derive the functionality of a molecule by identifying similar molecules with known characteristics. The method of [28] was evaluated on two types of clustering problems. The first one consists of enzyme data, where a bag corresponds to an enzyme and its instances to amino acid sequences. The second problem is the clustering of the molecules in the drug activity prediction datasets taken from [16].

In [78, 79] a multi-instance clustering method based on the maximum margin principle was proposed. It was evaluated on two separate applications. In *image clustering*, the method is used to detect common hidden concepts or patterns in

images. As was done in the image classification applications listed in Sect. 2.4.2, the images correspond to bags and the instances are image segments. The second application is *text clustering*. In this case, a bag represents a document and is made up from (possibly overlapping) passages taken from this document.

References

1. Ali, K., Saenko, K.: Confidence-rated multiple instance boosting for object detection. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), pp. 2433–2440. IEEE, Los Alamitos (2014)
2. Amar, R.A., Dooly, D.R., Goldman, S.A., Zhang, Q.: Multiple-instance learning of real-valued data. In: Brodley, C.E., Danyluk, A. (eds.) Proceedings of the 18th International Conference on Machine Learning (ICML 2001), pp. 3–10. Morgan Kaufmann Publishers, San Francisco (2001)
3. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information, vol. 15, pp. 561–568. MIT press, Cambridge (2002)
4. Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. *IEEE Trans. Pattern Anal.* **33**(8), 1619–1632 (2011)
5. Blaschke, C., Leon, E., Krallinger, M., Valencia, A.: Evaluation of BioCreAtIvE assessment of task 2. *BMC Bioinform.* **6**(1), 1 (2005)
6. Braddock, P., Hu, D., Fan, T., Stratford, I., Harris, A., Bicknell, R.: A structure-activity analysis of antagonism of the growth factor and angiogenic activity of basic fibroblast growth factor by suramin and related polyanions. *Br. J. Cancer* **69**(5), 890 (1994)
7. Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.: Supervised learning of semantic classes for image annotation and retrieval. *IEEE Trans. Pattern Anal.* **29**(3), 394–410 (2007)
8. Chang, K., Bowyer, K., Flynn, P.: An evaluation of multimodal 2d+3d face biometrics. *IEEE Trans. Pattern Anal.* **27**(4), 619–624 (2005)
9. Chen, Y., Wang, J.: Image categorization by learning and reasoning with regions. *J. Mach. Learn. Res.* **5**, 913–939 (2004)
10. Chen, Y., Bi, J., Wang, J.Z.: MILES: Multiple-instance learning via embedded instance selection. *IEEE Trans. Pattern Anal.* **28**(12), 1931–1947 (2006)
11. Cheplygina, V., Tax, D.: Characterizing multiple instance datasets. In: Feragen, A., Pelilo, M., Loog, M. (eds.) Similarity-Based Pattern Recognition, pp. 15–27. Springer, Switzerland (2015)
12. Davis, J., Costa, V.S., Ray, S., Page, D.: An integrated approach to feature invention and model construction for drug activity prediction. In: Ghahramani, Z. (ed.) Proceedings of the 24th international conference on Machine learning (ICML 2007), pp. 217–224. ACM, New York (2007)
13. De Raedt, L.: Attribute-value learning versus inductive logic programming: the missing links. In: Page, D. (ed.) Inductive Logic Programming. Lecture Notes in Computer Science, vol. 1446, pp. 1–8. Springer, Berlin (1998)
14. De Raedt, L.: Logical and Relational Learning. Springer Science & Business Media, Berlin (2008)
15. Deroski, S.: Relational data mining. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 887–911. Springer, New York (2009)
16. Dieterich, T.G., Lathrop, R.H., Lozano-Perez, T.: Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.* **89**(1–2), 31–71 (1997)
17. Dooly, D.R., Goldman, S.A., Kwek, S.S.: Real-valued multiple-instance learning with queries. *J. Comput. Syst. Sci.* **72**(1), 1–15 (2006)
18. El-Manzalawy, Y., Dobbs, D., Honavar, V.: Predicting MHC-II binding affinity using multiple instance regression. *IEEE ACM Trans. Comput. Biol.* **8**(4), 1067–1079 (2011)

19. Faltemier, T., Bowyer, K., Flynn, P.: Using a multi-instance enrollment representation to improve 3D face recognition. *Comput. Vis. Image Underst.* **112**(2), 114–125 (2008)
20. Feng, S., Xiong, W., Li, B., Lang, C., Huang, X.: Hierarchical sparse representation based multi-instance semi-supervised learning with application to image categorization. *Signal Process.* **94**, 595–607 (2014)
21. Fu, G., Nan, X., Liu, H., Patel, R.Y., Daga, P.R., Chen, Y., Wilkins, D.E., Doerksen, R.J.: Implementation of multiple-instance learning in drug activity prediction. *BMC Bioinform.* **13**(15), 1 (2012)
22. Fung, G., Dundar, M., Krishnapuram, B., Rao, R.B.: Multiple instance learning for computer aided diagnosis. *Adv. Neural Inf.* **19**, 425 (2007)
23. Goldman, S.A., Scott, S.D.: Multiple-instance learning of real-valued geometric patterns. *Ann. Math. Artif. Intel.* **39**(3), 259–290 (2003)
24. Han, Y., Qi, X.: A complementary svms-based image annotation system. In: *Proceedings of the 2005 IEEE International Conference on Image Processing (ICIP 2005)*, vol. 1, pp. 1185–1188. IEEE, Los Alamitos (2005)
25. Hong, R., Wang, M., Gao, Y., Tao, D., Li, X., Wu, X.: Image annotation by multiple-instance learning with discriminative feature mapping and selection. *IEEE Trans. Cybern.* **44**(5), 669–680 (2014)
26. Huang, P., Zhu, J.: Multi-instance learning for software quality estimation in object-oriented systems: a case study. *J. Zhejiang Univ.-Sci. C* **11**(2), 130–138 (2010)
27. Kotsiantis, S., Kanellopoulos, D., Tampakas, V.: Financial application of multi-instance learning: two greek case studies. *J. Conver. Inf. Technol.* **5**(8), 42–53 (2010)
28. Kriegel, H.P., Pryakhin, A., Schubert, M.: An EM-approach for clustering multi-instance objects. In: Ng, W., Kitsuregawa, M., Li, J., Chang, K. (eds.) *Lecture Notes in Artificial Intelligence*, pp. 139–148. Springer, Berlin (2006)
29. Li, Y.X., Ji, S., Kumar, S., Ye, J., Zhou, Z.H.: Drosophila gene expression pattern annotation through multi-instance multi-label learning. *IEEE ACM Trans. Comput. Biol.* **9**(1), 98–112 (2012)
30. Liu, G., Wu, J., Zhou, Z.: Key instance detection in multi-instance learning. In: Hoi, S., Bun-tine, W. (eds.) *JMLR: Workshop and Conference Proceedings: Asian Conference on Machine Learning*, pp. 253–268 (2012)
31. Liu, J., Lu, Y., Zhou, T.: Instance significance guided multiple instance boosting for robust visual tracking (2015). arXiv preprint. [arXiv:1501.04378](https://arxiv.org/abs/1501.04378)
32. Manandhar, A., Morton, K.D., Collins, L.M., Torriano, P.A.: Multiple instance learning for landmine detection using ground penetrating radar. In: Harmon, R., Holloway, J., Broach, J. (eds.) *Proceedings of SPIE, Detection and Sensing of Mines, Explosive Objects and Obscured Targets*, pp. 721–835. SPIE, Bellingham (2012)
33. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. In: Jordan, M., Kearns, M., Solla, S. (eds.) *Advances in Neural Information*, vol. 10, pp. 570–576. MIT press, Cambridge (1998)
34. Maron, O., Ratan, A.L.: Multiple-instance learning for natural scene classification. In: Shavlik, J. (ed.) *Proceedings of the 15th International Conference on Machine Learning (ICML 1998)*, vol. 98, pp. 341–349. Morgan Kaufmann Publishers, San Francisco (1998)
35. McGovern, A., Barto, A.G.: Automatic discovery of subgoals in reinforcement learning using diverse density. In: Brodley, C., Danyluk, A. (eds.) *Proceedings of the 18th International Conference on Machine Learning (ICML 2001)*, pp. 361–368. Morgan Kaufmann Publishers, San Francisco (2001)
36. Minhas, A., ul Amir, F., Ben-Hur, A.: Multiple instance learning of calmodulin binding sites. *Bioinformatics* **28**(18), i416–i422 (2012)
37. Muggleton, S., De Raedt, L.: Inductive logic programming: theory and methods. *J. Logic Program.* **19**, 629–679 (1994)
38. Murray, J., Hughes, G., Kreutz, K.: Machine learning methods for predicting failures in hard drives: a multiple-instance application. *J. Mach. Learn. Res.* **6**, 783–816 (2005)

39. Ni, B., Song, Z., Yan, S.: Web image mining towards universal age estimator. In: Proceedings of the 17th ACM international conference on Multimedia, pp. 85–94. ACM, New York (2009)
40. Pfeifer, N., Kohlbacher, O.: Multiple instance learning allows MHC class II epitope predictions across alleles. In: Crandall, K., Lagergren, J. (eds.) Algorithms in Bioinformatics, pp. 210–221. Springer, Berlin (2008)
41. Popescu, M., Mahnot, A.: Early illness recognition using in-home monitoring sensors and multiple instance learning. *Method. Inform. Med.* **51**(4), 359 (2012)
42. Qi, X., Han, Y.: Incorporating multiple svms for automatic image annotation. *Pattern Recogn.* **40**(2), 728–741 (2007)
43. Qi, G.J., Hua, X.S., Rui, Y., Mei, T., Tang, J., Zhang, H.J.: Concurrent multiple instance learning for image categorization. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1–8. IEEE, Los Alamitos (2007)
44. Rahmani, R., Goldman, S.A.: MISSL: Multiple-instance semi-supervised learning. In: Cohen, W., Moore, A. (eds.) Proceedings of the 23rd International Conference on Machine Learning (ICML 2006), pp. 705–712. ACM, New York (2006)
45. Ray, S., Craven, M.: Supervised versus multiple instance learning: an empirical comparison. In: De Raedt, L., Wrobel, S. (eds.) Proceedings of the 22nd International Conference on Machine Learning (ICML 2005), pp. 697–704. ACM, New York (2005)
46. Ray, S., Page, D.: Multiple instance regression. In: Brodley, C., Danyluk, A. (eds.) Proceedings of the 18th International Conference on Machine Learning (ICML 2001), pp. 425–432. Morgan Kaufmann Publishers, San Francisco (2001)
47. Reutemann, P.: Development of a propositionalization toolbox. Master’s thesis, Albert Ludwigs University of Freiburg, Germany (2004)
48. Reutemann, P., Pfahringer, B., Frank, E.: A toolbox for learning from relational data with propositional and multi-instance learners. In: Webb, G., Yu, X. (eds.) Lecture Notes in Artificial Intelligence, pp. 421–434. Springer, Berlin (2005)
49. Rocchio, J.J.: Relevance feedback in information retrieval. In: Salton, G. (ed.) The SMART Retrieval System: Experiments in Automatic Document Processing, pp. 313–323. Prentice-Hall, Englewood Cliffs (1971)
50. Ruffo, G.: Learning single and multiple instance decision trees for computer security applications. Ph.D. thesis, Department of Computer Science, University of Turin, Turin, Italy (2000)
51. Sánchez Tarragó, D., Cornelis, C., Bello, R., Herrera, F.: A multi-instance learning wrapper based on the Rocchio classifier for web index recommendation. *Knowl.-Based Syst.* **59**, 173–181 (2014)
52. Srinivasan, A., Muggleton, S., King, R.D., Sternberg, M.J.: Mutagenesis: ILP experiments in a non-determinate biological domain. In: Wrobel, S. (ed.) Proceedings of the 4th international workshop on inductive logic programming, vol. 237, pp. 217–232. Gesellschaft fr Mathematik und Datenverarbeitung MBH, Bonn (1994)
53. Sun, L., Lu, Y., Yang, K., Li, S.: ECG analysis using multiple instance learning for myocardial infarction detection. *IEEE Trans. Bio-Med. Eng.* **59**(12), 3348–3356 (2012)
54. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT press, Cambridge (1998)
55. Tao, Q., Scott, S., Vinodchandran, N., Osugi, T.T.: Svm-based generalized multiple-instance learning via approximate box counting. In: Greiner, R., Schuurmans, D. (eds.) Proceedings of the 21st International Conference on Machine Learning (ICML 2004), p. 101. ACM, New York (2004)
56. Teramoto, R., Kashima, H.: Prediction of protein-ligand binding affinities using multiple instance learning. *J Mol. Graph. Model.* **29**(3), 492–497 (2010)
57. Wagstaff, K.L., Lane, T.: Saliency assignment for multiple-instance regression. In: Proceedings of the ICML 2007 Workshop on Constrained Optimization and Structured Output Spaces. Citeseer (2007)
58. Wagstaff, K.L., Lane, T., Roper, A.: Multiple-instance regression with structured data. In: Bonchi, F., Berendt, B., Giannotti, F., Gunopulos, D., Turini, F., Zaniolo, C., Ramakrishnan, N., Wu, X. (eds.) Proceedings of the 2008 IEEE International Conference on Data Mining Workshops (ICDMW 08), pp. 291–300. IEEE, Los Alamitos (2008)

59. Wang, C., Scott, S., Zhang, J., Tao, Q., Fomenko, D.E., Gladyshev, V.N.: A study in modeling low-conservation protein superfamilies. CSE Technical reports, p. 35 (2004)
60. Wang, S., McKenna, M.T., Nguyen, T.B., Burns, J.E., Petrick, N., Sahiner, B., Summers, R.M.: Seeing is believing: video classification for computed tomographic colonography using multiple-instance learning. *IEEE Trans. Med. Imaging* **31**(5), 1141–1153 (2012)
61. Wang, Q., Yuan, Y., Yan, P., Li, X.: Saliency detection by multiple-instance learning. *IEEE Trans. Cybern.* **43**(2), 660–672 (2013)
62. Wang, Z., Radosavljevic, V., Han, B., Obradovic, Z., Vucetic, S.: Aerosol optical depth prediction from satellite observations by multiple instance regression. In: Apte, C., Park, H., Wang, K., Zaki, M. (eds.) *Proceedings of the 2008 SIAM International Conference on Data Mining*, pp. 165–176. SIAM, Philadelphia (2008)
63. Wang, Z., Lan, L., Vucetic, S.: Mixture model for multiple instance regression and applications in remote sensing. *IEEE Trans. Geosci. Remote* **50**(6), 2226–2237 (2012)
64. Wu, J., Zhu, X., Zhang, C., Yu, P.S.: Bag constrained structure pattern mining for multi-graph classification. *IEEE Trans. Knowl. Data. Eng.* **26**(10), 2382–2396 (2014)
65. Wu, J., Pan, S., Zhu, X., Cai, Z.: Boosting for multi-graph classification. *IEEE Trans. Cybern.* **45**(3), 416–429 (2015)
66. Yang, C., Lozano-Pérez, T.: Image database retrieval with multiple-instance learning techniques. In: *Proceedings of the 16th International Conference on Data Engineering*, pp. 233–243. IEEE, Los Alamitos (2000)
67. Zafra, A., Romero, C., Ventura, S., Herrera-Viedma, E.: Multi-instance genetic programming for web index recommendation. *Expert Syst. Appl.* **36**(9), 11470–11479 (2009)
68. Zafra, A., Romero, C., Ventura, S.: Multiple instance learning for classifying students in learning management systems. *Expert Syst. Appl.* **38**(12), 15020–15031 (2011)
69. Zafra, A., Gibaja, E.L., Ventura, S.: Multiple instance learning with multiple objective genetic programming for web mining. *Appl. Soft Comput.* **11**(1), 93–102 (2011)
70. Zafra, A., Ventura, S.: Multi-instance genetic programming for predicting student performance in web based educational environments. *Appl. Soft Comput.* **12**(8), 2693–2706 (2012)
71. Zhang, C., Chen, X.: Region-based image clustering and retrieval using multiple instance learning. In: Leow, W., Lew, M., Chua, T., Ma, W., Chaisom, L., Bakker, E. (eds.) *Lecture Notes in Computer Science*, pp. 194–204. Springer, Berlin (2005)
72. Zhang, Q., Goldman, S.A.: EM-DD: an improved multiple-instance learning technique. In: Dieterich, T., Becker, S., Ghahramani, Z (eds.) *Advances in Neural Information*, pp. 1073–1080. MIT press, Cambridge (2001)
73. Zhang, K., Song, H.: Real-time visual tracking via online weighted multiple instance learning. *Pattern Recogn.* **46**(1), 397–411 (2013)
74. Zhang, M.L., Zhou, Z.H.: Multi-instance clustering with applications to multi-instance prediction. *Appl. Intell.* **31**(1), 47–68 (2009)
75. Zhang, Q., Goldman, S.A., Yu, W., Fritts, J.E.: Content-based image retrieval using multiple-instance learning. In: Sammut, C., Hoffman, A. (eds.) *Proceedings of the 19th International Conference on Machine Learning (ICML 2002)*, pp. 682–689. Morgan Kaufmann Publishers, San Francisco (2002)
76. Zhang, C., Chen, S.C., Shyu, M.L.: Multiple object retrieval for image databases using multiple instance learning and relevance feedback. In: *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME 2004)*, vol. 2, pp. 775–778. IEEE, Los Alamitos (2004)
77. Zhang, C., Chen, X., Chen, M., Chen, S.C., Shyu, M.L.: A multiple instance learning approach for content based image retrieval using one-class support vector machine. In: *Proceedings of the 2005 IEEE International Conference on Multimedia and Expo (ICME 2005)*, pp. 1142–1145. IEEE, Los Alamitos (2005)
78. Zhang, D., Wang, F., Si, L., Li, T.: M3IC: maximum margin multiple instance clustering. In: *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*, vol. 9, pp. 1339–1344 (2009)

79. Zhang, D., Wang, F., Si, L., Li, T.: Maximum margin multiple instance clustering with applications to image and text clustering. *IEEE Trans. Neural Netw.* **22**(5), 739–751 (2011)
80. Zhao, Z., Fu, G., Liu, S., Elokely, K.M., Doerksen, R.J., Chen, Y., Wilkins, D.E.: Drug activity prediction using multiple-instance learning via joint instance and feature selection. *BMC Bioinform.* **14**(Suppl 14), S16 (2013)
81. Zhou, Z., Jiang, K., Li, M.: Multi-instance learning based web mining. *Appl. Intell.* **22**(2), 135–147 (2005)
82. Zhou, Z.H., Zhang, M.L., Huang, S.J., Li, Y.F.: Multi-instance multi-label learning. *Artif. Intell.* **176**(1), 2291–2320 (2012)
83. Zhou, T., Lu, Y., Qiu, M.: Online visual tracking using multiple instance learning with instance significance estimation. *Comput. Res. Repos.* (2015)