Pattern Recognition Letters 59 (2015) 11–17

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

On classification with bags, groups and sets ☆

Veronika Cheplygina [a,*], David M.J. Tax [a], Marco Loog [a,b]

[a] *Pattern Recognition Laboratory, Delft University of Technology, The Netherlands*
[b] *Image Group, University of Copenhagen, Denmark*

A R T I C L E I N F O

Article history:
Received 11 November 2014
Available online 28 March 2015

Keywords:
Multiple instance learning
Set classification
Group-based classification
Label dependencies
Weakly labeled data

A B S T R A C T

Many classification problems can be difficult to formulate directly in terms of the traditional supervised setting, where both training and test samples are individual feature vectors. There are cases in which samples are better described by sets of feature vectors, that labels are only available for sets rather than individual samples, or, if individual labels are available, that these are not independent. To better deal with such problems, several extensions of supervised learning have been proposed, where either training and/or test objects are sets of feature vectors. However, having been proposed rather independently of each other, their mutual similarities and differences have hitherto not been mapped out. In this work, we provide an overview of such learning scenarios, propose a taxonomy to illustrate the relationships between them, and discuss directions for further research in these areas.

1. Introduction

In recent years, the field of pattern recognition has seen many problems that are difficult to formulate as regular supervised classification problems where (feature vector, label) pairs are available to train a classifier that, in turn, can predict labels for previously unseen feature vectors. A subset of these problems contains learning scenarios where (part of) the objects are represented by sets or *bags* of feature vectors or *instances*. Such learning scenarios include multiple instance learning [11], set classification [42], group-based classification [48] and many others. In this paper we review these learning scenarios.

There are several reasons why a bag representation might be chosen in a pattern recognition problem. The first reason is that a single feature vector is often too restrictive to describe an object. For example, in drug activity prediction, we are interested in classifying molecules as having the desired effect (active) or not. However, a molecule is not just a list of its elements: most molecules can fold into different shapes or conformations, which can influence the activity of that molecule. Furthermore, the number of stable shapes is different per molecule. A more logical choice is therefore to represent a molecule as a set of its conformations.

The second reason is that labels on the level of feature vectors are difficult, costly and/or time-consuming to obtain, but labels on a coarser level can be obtained more easily. For computer aided diagnosis applications, it can be very expensive for a radiologist to label individual pixels or voxels in an image as healthy or diseased, while it is more feasible to tag a full image or some large image regions with a single label. Such coarsely labeled scans or regions can then be used for train a classifier and predict labels at the bag level, i.e., complete patient scans, or at the finer grained region or instance level, e.g., by labeling individual pixels or voxels.

Another reason to consider the labeling of bags of instances, instead of single feature vectors, is that there can be structure in the labels of the instances. For example, in face verification, where a video of a person is available, considering all the video frames jointly can provide more confident predictions than labeling each of the frames individually and combining the decisions. Similarly, neighboring objects in images, videos, sounds, time series and so forth are typically very correlated, and thus should not be classified independently.

These examples have different goals and assumptions, and therefore may require different representations in the training and the test phase. All possibilities shown in Fig. 1 occur: both training and test objects can be single instances (SI) or bags, i.e. multiple instances (MI). Traditional supervised learning is in the SI–SI scenario, where both training and test objects are instances. Predicting molecule activity is in the MI–MI scenario, where both training and test objects are bags. Image classification problems can be found in the MI–MI scenario (training on images, testing on images) as well as the MI–SI scenario (training on images, testing on pixels or patches). The face verification problem is best represented by the SI–MI scenario (training on a single face, testing on a set of faces).

☆ This paper has been recommended for acceptance by M.A. Girolami.

* Corresponding author. Tel.: +31 152787243.
E-mail address: v.cheplygina@tudelft.nl (V. Cheplygina).

http://dx.doi.org/10.1016/j.patrec.2015.03.008

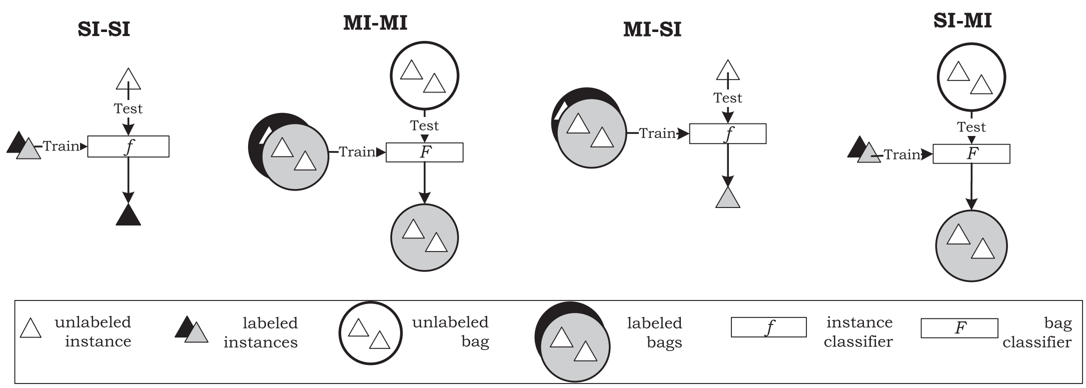

Fig. 1. Supervised learning (SI–SI) and extensions. In the MI–MI scenario (Section 3.3), both training and test objects are bags. In the MI–SI scenario (Section 3.4), the training objects are bags and test objects are instances, while in the MI–SI scenario (Section 3.5), the training objects are instances and the test objects are bags.

The success of a classifier in one application, such as molecule activity prediction, often motivates other researchers to use the same method in a different application, such as image classification. However, it is not necessarily the case that the assumptions of the first application still hold. For example, the assumptions on the relationships of bag and instance labels can be different for molecules and for images, which can lead to poor performances. On the other hand, it can also happen that the same type of problem occurs in two different applications, and that researchers in the respective fields approach the problem in different ways, without benefiting from each other's findings. We therefore believe that understanding the relationships between such learning scenarios is of importance to researchers in different fields.

With this work, our goal is to provide an overview of learning scenarios in which bags of instances play a role at any of the stages in the learning or classification process and to provide insight in their interconnections. We have gathered papers that proposed novel learning scenarios, often combining synonyms of the word "set" with words such as "classification" or "learning". Our work is intended as a survey of learning problems, not of classifiers for a particular scenario, although we refer to existing surveys of this type whenever possible. Furthermore, we focus on a single-label, binary classification scenario. Our focus is complementary to the multi-label and/or multi-class setting and the problem formulations covered in this work can be extended to multi-label and multi-class. Examples can be found in [51,62].

This paper begins with an overview of applications which motivate the bag representation in Section 2, and the assumptions (such as on the relationship of instance and bag labels) associated with these applications. We then explain the categories of learning scenarios and the methodologies used to learn in such scenarios in Section 3. The paper concludes with a discussion in Section 4.

2. Applications and assumptions

2.1. Molecule activity prediction

In molecule activity prediction, the goal is to predict whether a previously unseen molecule has the desired activity, for example, whether a protein binds to another protein and thus influences a biological process. Often molecules have different conformations, or 3D shapes they can fold into, which influence their binding properties. Naturally, different molecules have different numbers of conformations. Therefore, one possibility is to represent molecules by the set of their conformations. For existing molecules, however, the information of which conformations are active, and which are not, is not available. A possible assumption in this case is that if at least one of the conformations is active, that the molecule can be regarded is active. This assumption is used in [11,14] and entails that the instances have labels, and if at least one instance is positive, the bag is positive as well.

Another possibility is to represent a molecule by a 3D cloud of atoms. Atom clouds with similar shapes are expected to display similar activity. Therefore, by aligning the clouds and comparing them directly, the function of previously unseen molecules can be predicted. This assumption is used in [18]. Here the instances (atoms) do not have labels, as it is not logical for an atom to be active or inactive, but certain combinations of instances do lead to different bag labels. In other words, most, or all instances contribute to the bag label. Both applications can be found in the MI–MI scenario (Section 3.3).

2.2. Image classification

In one group of image classification applications, bags are images, and the instances are parts of the images, such as pixels, blobs or segments. Examples include natural scene classification [8,36], object recognition [2,45] or medical imaging [9,22,23,48]. Often the assumption is that not all parts of the image contribute to the image label. For example, in an image of a tiger, other surroundings can be present, or in a lung scan of a patient with a lung disease, healthy lung tissue can be present as well [9]. Each instance therefore has a label (positive, i.e. containing a tiger, or not) and a popular assumption is that if at least one instance is positive, then the bag is also positive. The goal is to label novel images (i.e. bags).

On the other hand, the standard assumption might not always be sufficient. For example, if the instances are pixels, it might not be suitable to define pixels as belonging to the tiger concept. Perhaps a fraction of positive instances is more suitable. Or, for the beach concept, both instances containing sand and instances containing water might be needed, therefore asking for a conjunction of concepts. Relaxed assumptions to deal with such problems are described in [56]. Such problems fall into the MI–MI scenario (Section 3.3). Another assumption is that all instances in the bag share the same label. This assumption is used in [48], when classifying groups of cells as healthy or anomalous, with the added information that all cells in a group share the same label. Although training can be done using labeled cells, in the test phase, it might be advantageous to classify the cells jointly, rather than using a two-step approach where cells are classified first, and their decisions are combined. This problem is therefore in the SI–MI scenario (Section 3.5).

In general, the definition of instances influences what is reasonable for the application at hand. Typically, the more knowledge is involved in generating the instances, the more assumptions could be applicable. Consider an application with photographs, where each photograph is labeled with the people in that photo. If we use as a face detector to generate candidate instances [17], it is reasonable to