Knowledge-Based Systems 136 (2017) 46–57

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

An ensemble approach to multi-view multi-instance learning

Alberto Cano

Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA

ARTICLE INFO

Article history:
Received 25 January 2017
Revised 26 August 2017
Accepted 28 August 2017
Available online 30 August 2017

Keywords:
Multi-view
Multi-instance
Classification
Ensemble

ABSTRACT

Multi-view learning combines data from multiple heterogeneous sources and employs their complementary information to build more accurate models. Multi-instance learning represents examples as labeled bags containing sets of instances. Data fusion of different multi-instance views cannot be simply concatenated into a single set of features due to their different cardinality and feature space. This paper proposes an ensemble approach that combines view learners and pursues consensus among the weighted class predictions to take advantage of the complementary information from multiple views. Importantly, the ensemble must deal with the different feature spaces coming from each of the views, while data for the bags may be partially represented in the views. The experimental study evaluates and compares the performance of the proposal with 20 traditional, ensemble-based, and multi-view algorithms on a set of 15 multi-instance datasets. Experimental results indicate the better performance of ensemble methods than single-classifiers, but especially the best results of the multi-view multi-instance approaches. Results are validated through multiple non-parametric statistical analysis.

1. Introduction

Multi-view learning [1] is a relatively new paradigm that exploits data represented by multiple distinct feature sets called views which have been obtained from different data sources. The objective is to learn functions to model the views and jointly optimize and exploit the redundancy and complementarity of data in the views [2]. The multi-view foundations are based on the consensus and complementary principles of data. The consensus principle maximizes the agreement of distinct views, i.e. the relationships between different subsets of features. The complementary principle advocates for the distribution of the information among the views, i.e., each view contains partial data that may not provide interesting information when analyzed separately, but when merged together provide meaningful and comprehensive knowledge to the learner. Multi-view learning has received much attention in machine learning, especially for multi-label and image classification [3–5]. Learning classification models from the fusion of multiple views increases the strength of the classification predictions as compared to independent views [6–8]. However, it is not straightforward to accomplish such task and many times multi-view data cannot be easily integrated due to the heterogeneous data representation from the multiple sources.

Multi-instance learning [9–11] is a generalization of traditional supervised learning in which each example, called *bag*, comprises a variable number of non-repeated *instances* described by feature vectors. The bag is associated with a single class label, although the labels of the particular instances are unknown. Dietterich et al. [9] defined the multiple instance problem, in which a bag is classified as positive if it contains at least one positive instance. More recently, other generalized multi-instance models have been formalized [12–14]. The intersection between multi-view and multi-instance learning is natural and represents a flexible representation paradigm for supervised learning. Multi-instance learning can be adapted to train from data allocated in multiple views, reflecting the distribution of the information into heterogeneous feature sets with different sets of disjoint attributes. The fusion of the views provides more complete information about the bags that should lead to better accuracy as compared with the prediction using the partial data from the isolated views.

Despite the flexibility of the multi-instance data representation, it is difficult to combine multiple multi-instance views. While bags contained in multiple views share the bag identifier, the instances are represented using different feature vectors (attributes provided by each view). In traditional single-instance classification, it is straightforward to combine data from two views having common examples by simply joining the feature sets. However, this process cannot be directly performed in multi-instance learning since there is no matching between the particular instances but between the whole bags by means of their bag identifiers. Current multi-view multi-instance approaches perform problem transformations of the data into multiple single meta-instance views [15]. Nonetheless,

E-mail address: acano@vcu.edu

http://dx.doi.org/10.1016/j.knosys.2017.08.022
0950-7051/

this *flattening* transformation destroys the multi-instance representation of the data and refuses to resolve the true multi-view multi-instance nature of the problem.

This paper presents an ensemble approach to directly learn from multi-view multi-instance (MVMI) data. Ensembles combine the predictions of multiple classifiers in order to reduce the variance and bias of the predictions, and specifically they have shown to improve accuracy in single-instance multi-view [16,17]. Our main contributions are the following:

1. Propose ensemble learning to fuse the information from multi-view multi-instance data relations with heterogeneous feature sets without conducting any multi-instance problem transformation.
2. Identify the best performing base classifier for each of the views. Views are represented by different feature vectors providing diverse information. Therefore, rather than using the same base classifier on all of the views homogeneously, we propose to evaluate different families of base classifiers on each of the views, and then selecting the model which performed best for each view. Moreover, considering the prediction of a diverse family of classifiers has shown to improve performance [18].
3. Weight the prediction of the base classifiers based on their local accuracy on the views. This is motivated because not all the views provide useful high-quality information to the classifier, but on the contrary, some views may provide irrelevant, contradictory, or noisy information to the ensemble. Therefore, the weighting will disregard predictions coming from low-quality classifiers, in contrast to the default majority voting that would decrease the performance of the ensemble.

The experimental study compares the performance of 20 multi-instance classifiers on 15 datasets with regards of four performance measures. The experimental results are validated through the analysis of non-parametric statistical tests [19], namely the Bonferroni–Dunn, Holm, and Wilcoxon tests that evaluate whether there are statistically significant differences in multiple and pairwise comparisons of algorithms. Results indicate that ensemble and multi-view methods improve the performance of traditional multi-instance classifiers.

This paper is structured as follows. Section 2 reviews related works. Section 3 describes the proposed multi-view multi-instance approach. Section 4 describes the experimental study, discusses, and analyzes the results. Finally, Section 5 presents the conclusions.

2. Related works

This section defines the basis and reviews related works on multi-instance learning, multi-view learning, and ensemble classification.

2.1. Multi-instance learning

In multi-instance classification the examples are called bags, and represent a set of instances. The class is associated with the whole bag although the instances are not explicitly associated with any particular class. Therefore, multi-instance learning inducts a prediction function $f(bag) \rightarrow C$ where the bag is a set of k instances $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k\}$ associated with a class label C. Instances are feature vectors $\bar{x} = [x_1, x_2, \ldots, x_f]$ with f attribute–values.

The standard hypothesis by Dietterich et al. [9] assumes that if a bag is positive, then at least one of its instances must be positive. However, if the bag is negative, then none of its instances could be positive. Therefore, a bag is positive if and only if at least one of its instances is positive. This can be modelled by introducing a second function $g(bag, j)$ that takes a single variant instance j. The classification function $f(bag)$ can be defined as follows:

$$f(bag) = \begin{Bmatrix} 1 & if\ \exists\ j\ \mid\ g(bag, j) = 1 \\ 0 & otherwise \end{Bmatrix} \quad (1)$$

There are two recent multi-instance reviews in literature by Foulds and Frank [12] and Amores [20], which cover multi-instance learning assumptions and the taxonomy of algorithms. Methods are categorized based on whether they focus on instance-level information (instance-space paradigm) or bag-level information, and in the latter whether they extract the relevant information implicitly (bag-space paradigm) or explicitly (embedded-space paradigm). Langone and Suykens [21] combined instance and bag level information for aggregated feature learning though the collective assumption. Qiao et al. [22] presented diversified dictionaries to address the problem of bridging instance-level representations to bag-level labels, which exploits bag-level label information for training class-specific dictionaries. Weidmann et al. [23] presented a two-level methodology which transformed a multi-instance dataset into single meta-instance dataset. They create multiple propositional single-instance datasets using decision trees and clustering, which are learned using a correlation-based multi-view learner. However, it comes at the cost of losing the nature of the multi-instance representation. On the contrary, the advantage of our contribution is the lack of problem transformation, and therefore it operates on multi-instance bags directly.

2.2. Multi-view learning

Multi-view learning algorithms can be categorized into three groups [24]: co-training, multiple kernel learning, and subspace learning. Co-training learns from different views alternately to maximize the mutual agreement among two distinct views of the data [25], i.e., the consensus principle. It relies on three assumptions: sufficiency, compatibility, and conditional independence. Multiple kernel learning performs linear or non-linear combinations of the views. Subspace learning focuses on the latent subspace shared by views [26,27], which has lower dimensionality. Many related works on multi-view multi-instance learning are mainly oriented to image annotation or semi-supervised learning [28–32].

MV-TLC is a two-level method by Wang et al. [15] which transforms a multi-view multi-instance dataset into a multi-view single meta-instance dataset. The first level creates multiple single-instance views using decision trees and clustering, which are learned using correlation-based learners. The second level constructs the final classification model using regular propositional classifiers combined in a multi-view algorithm. This is closest related work to our proposal, but they conduct flattening of the multi-instances into single-instances via problem transformation. On the contrary, our contribution is capable of handling multi-view multi-instance data directly without conducting any problem transformation.

Consistency between the views in multi-instance learning was the primary objective defined by Zhang et al. [33] for improving the performance. Furthermore, cross-view feature selection was presented by Wu et al. [34] to identify the most representative features across the views in multi-instance. However, this approach may not be the most effective in the case of heterogeneous base learners, for which it is better to let them select the best performing features for each view, as considered in our proposal.

MI2L was proposed by Li et al. [35], which employs a sparse graph model to represent context relations in a bag into a unified framework. They also implement a multi-view graph dictionary to improve the discrimination performance, similar to Qiao et al. [22].