International Journal of Approximate Reasoning 68 (2016) 194–210

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A scalable pairwise class interaction framework for multidimensional classification ☆

Jacinto Arias [a,*], Jose A. Gamez [a], Thomas D. Nielsen [b], Jose M. Puerta [a]

[a] *Department of Computing Systems, University of Castilla-La Mancha, Albacete, Spain*
[b] *Department of Computer Science, Aalborg University, Denmark*

ARTICLE INFO

Article history:
Received 16 December 2014
Received in revised form 14 July 2015
Accepted 21 July 2015
Available online 29 July 2015

Keywords:
Multidimensional classification
Probabilistic classifiers
Markov random fields

ABSTRACT

We present a general framework for multidimensional classification that captures the pairwise interactions between class variables. The pairwise class interactions are encoded using a collection of base classifiers (Phase 1), for which the class predictions are combined in a Markov random field that is subsequently used for multidimensional inference (Phase 2); thus, the framework can be positioned between multilabel Bayesian classifiers and label transformation-based approaches. Our proposal leads to a general framework supporting a wide range of base classifiers in the first phase as well as different inference methods in the second phase. We describe the basic framework and its main properties, as well as strategies for ensuring the scalability of the framework. We include a detailed experimental evaluation based on a range of publicly available databases. Here we analyze the overall performance of the framework and we test the behavior of the different scalability strategies proposed. A comparison with other state-of-the-art multidimensional classifiers show that the proposed framework either outperforms or is competitive with the tested straw-men methods.

1. Introduction

Supervised classification is the problem of assigning a value to a distinguished variable, the *class* C, for a given instance defined over a set of predictive attributes. In *multi-label* classification, several class variables are simultaneously considered and the task consists of assigning a configuration of values to all the class variables. In the multi-label setting, classes (or *labels*) are binary. *Multi-dimensional* classification is a generalization of multi-label classification that allows class variables to have more than two values [2]. Recent literature, however, also uses the term multi-label classification when dealing with n-ary class variables, so we will use both names in an interchangeable way.

A wide range of applications for multi-dimensional classification has been found [3, Section 1]: bio-informatics, document/music/movie categorization, semantic scene classification, multi-fault diagnosis, etc. One approach to solve a multi-dimensional problem is to first *transform* the problem into a set of single-class classification problems, and then combine the outputs to obtain a joint configuration of the class values. Transformation-based methods range from binary relevance, where no interaction among the class variables is modeled, to brute-force label power set methods, where all the class variables are aggregated into a single compound class. In-between these two extremes, we find new and/or adapted algorithms

☆ This paper is an extension of the conference paper [1].

* Corresponding author.
E-mail addresses: jacinto.arias@uclm.es (J. Arias), jose.gamez@uclm.es (J.A. Gamez), tdn@cs.aau.dk (T.D. Nielsen), jose.puerta@uclm.es (J.M. Puerta).

http://dx.doi.org/10.1016/j.ijar.2015.07.007

that have been developed to deal with the multi-dimensional problem, managing in a natural way the interactions between the variables. From this second family, probabilistic methods and, in particular, those based on Bayesian networks (BNs) [4] have demonstrated a convincing performance [5]. In this paper we focus on the probabilistic approach to multi-dimensional classification.

We propose a two-stage framework for multi-dimensional classification. The framework can be positioned between the transformation-based classifiers and the family of multi-dimensional probabilistic graphical models (PGMs)-based classifiers.[1] In the first stage we learn a single-class classifier for each pair of class variables in the domain, hence this stage of the framework follows a transformation-based approach. The framework does not prescribe a particular type of classifier, but only requires that the outcome of the classifier should be a weighted distribution over the (compound) class values. Standard probabilistic classifiers meet this criterion. In the second stage a Markov random field (MRF) is constructed based on the results from the first stage. The MRF thus models the dependencies between the class variables, and thereby connects the framework to the class of multi-dimensional PGM-based classifiers. Subsequent classification is achieved by performing inference in the induced MRF.

The proposed framework is flexible: (1) different types of classifiers can be applied in the first stage; (2) preprocessing can be done separately for each single-class classifier, thus allowing one to take advantage of state-of-the-art algorithms for supervised discretization and feature selection; and (3) different types of MRF-based inference algorithms can be used for the subsequent classification, and the choice of method can therefore depend on the complexity of the model (exact or approximate inference) and the score to be maximized (calculation of marginal probabilities or a most probable explanation). Furthermore, the general method scale with the computational resources available as the base classifiers in the first stage can be learned independently and chosen to also support scalability at the individual classifier level [7]. Nevertheless, naively dealing with all pairs of class variables imposes a strong limitation on the number of class variables the algorithm can handle. We therefore also outline strategies for scaling up the algorithm to datasets having a large number of class variables. Experiments carried out over a collection of benchmark datasets confirm the feasibility of the approach and show that the proposed method significantly outperforms or is comparable to the straw-men methods included in the comparison. This paper has a companion website, where the source-code and additional experimental results can be found: http://simd.albacete.org/supplements/FMC.html.

We would like to remark that the proposal does not fit into the so-called *pair-wise multi-label approach*, which interprets the labels of the instances as preferences and whose goal is to obtain a *ranking* among the labels [8] and not a joint configuration of class values. Furthermore, although our approach trains several classifiers in the first phase, it does not strictly fall into the class of *ensemble* methods either, as each base-classifier only provides a partial answer to the multi-dimensional problem.

This paper extends a previous preliminary study [1]. Specifically, we explore additional strategies to ensure scalability of the framework and we provide a significantly expanded experimental analysis and comparison with other state-of-the-art straw-men methods covering both accuracy and run-time performance.

The rest of the paper is structured as follows: In Section 2 we introduce the notation and the required background in the field of multidimensional classification; in Section 3 we describe our framework and discuss its main properties; in Section 4 we introduce and discuss several improvements to the framework with regards to scalability; in Section 5 we evaluate our approach by carrying out several experiments with real world data, and finally, in Section 6, we summarize the obtained results and introduce ideas for future work.

2. Background

2.1. Notation and problem definition

We assume that the available dataset consists of a collection of instances $\mathbf{D} = \{(\mathbf{a}^{(1)}, \mathbf{c}^{(1)}), \ldots, (\mathbf{a}^{(t)}, \mathbf{c}^{(t)})\}$, where the first part of an instance, $\mathbf{a}^{(i)} = (a_1^{(i)}, \ldots, a_n^{(i)})$, is a configuration of values defined over a set $\mathbf{A} = \{A_1, \ldots, A_n\}$ of predictive attributes, while the second part, $\mathbf{c}^{(i)} = (c_1^{(i)}, \ldots, c_m^{(i)})$, is a configuration of values defined over a set $\mathbf{C} = \{C_1, \ldots, C_m\}$ of classes.[2] Due to the restrictions of the models analyzed in this paper, we will assume that all the variables are discrete (nominal), i.e., the state spaces $\mathrm{dom}(()A_i)$ and $\mathrm{dom}(()C_j)$ are finite sets of mutually exclusive and exhaustive states, $\forall i, 1 \leq i \leq n$ and $\forall j, 1 \leq j \leq m$. In Section 3.1 we will discuss the preprocessing capabilities of our proposal in relation to continuous data.

Our goal is to induce a multidimensional classifier f that maps configurations of the predictive variables to configurations of the class variables:

[1] We refer to PGM based classifiers in order to accommodate a wider range of graphical models [6] in addition to the more common approaches based on Bayesian networks.

[2] We will omit the superscript when no confusion is possible, just writing $(\mathbf{a}, \mathbf{c})$ and $((a_1, \ldots, a_n), (c_1, \ldots, c_m))$ instead of $(\mathbf{a}^{(i)}, \mathbf{c}^{(i)})$ and $((a_1^{(i)}, \ldots, a_n^{(i)}), (c_1^{(i)}, \ldots, c_m^{(i)}))$, respectively.