



Membership-margin based feature selection for mixed type and high-dimensional data: Theory and applications



Lyamine Hedjazi ^{a,b,*}, Joseph Aguilar-Martin ^b, Marie-Véronique Le Lann ^{b,c},
Tatiana Kempowsky-Hamon ^b

^a *Institute of Cardiometabolism and Nutrition, Department of Omics Sciences, 91 boulevard de l'hôpital, 75013 Paris, France*

^b *CNRS, LAAS, 7, Avenue du Colonel Roche, F-31031 Toulouse, France*

^c *Univ de Toulouse, LAAS, F-31031 Toulouse, France*

ARTICLE INFO

Article history:

Received 1 July 2014

Received in revised form 13 May 2015

Accepted 11 June 2015

Available online 20 June 2015

Keywords:

Feature selection

Fuzzy classifier

Mixed-type data

Machine learning

Margin

ABSTRACT

The present paper describes a new feature weighting method based on a membership margin. Distinctive properties of the proposed method include its capability to process problems characterized by mixed-type data (quantitative, qualitative and interval) as well as a huge number of features. The key idea is to map simultaneously all the features of different types into a common space; the membership space. Once all features are represented in a homogeneous space, a feature weighting task can be performed in unified way. This weighting approach is integrated here within a fuzzy classifier through a fuzzy rule weighted concept in order to improve its performance. Each antecedent fuzzy set in the fuzzy if–then rule is weighted to characterize the importance of each proposition and therefore its corresponding feature. Weight estimation process is based on membership margin maximization to estimate a fuzzy weight of each feature in the membership space. Experiments on low and high dimensional real-world datasets demonstrate that the proposed approach can improve significantly the performance of the fuzzy rule-based as well as other state of the art classifiers and can even outperform classical feature weighting approaches. In particular, we show that this approach can yield meaningful results on two real-world applications for cancer prognosis and industrial process diagnosis.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

The data generated by recent advanced high-throughput technologies present multiple challenges, including high-dimensional and mixed-type data. The first challenge related to high-dimensionality appears as a topical problem that justifies the success of “Machine Learning” approach; in particular for statistically ill-posed problems involving a huge number of features (e.g. several thousands to ten thousands for DNA microarray data [23,64]). In such applications, most of these features are irrelevant with a relatively small sample size (tens or at most hundreds). These situations arise particularly in the bioinformatics field involving the analysis of gene expression and proteomic profiles for different purposes such as disease diagnosis, prognosis and treatment response prediction [72,78]. The second challenge concerns the problem of processing simultaneously mixed-type and heterogeneous data (qualitative, quantitative, interval, etc.) which are present

* Corresponding author at: Institute of Cardiometabolism and Nutrition, Department of Omics Sciences, 91 boulevard de l'hôpital, 75013 Paris, France.
E-mail address: l.hedjazi@ican-institute.org (L. Hedjazi).

almost in all daily produced datasets (for instance most of the UCI repository datasets are of mixed type). However, with the recent trends towards an integrative bioinformatics that aims to integrate different data sources, the occurrence of both challenges jointly is possible in many applications (e.g. integration of clinical data with gene expression or proteomic profiles [20,76]). Accordingly, developing efficient feature selection methods to overcome both challenges simultaneously is gaining increasing attention within machine learning community [51].

Feature selection is defined as the problem of choosing a small subset of features that ideally is necessary and sufficient to describe the target concept [46]. Many research efforts have been directed in the last two decades towards developing efficient feature selection methods [6,28,29,41,53,54,82,84]. Feature selection methods are traditionally categorized as filter, wrapper, hybrid or embedded methods, with respect to the criterion used to search for relevant features [27,48]. In filter methods an independent evaluation function based typically on a measure of information content is used to select a set of features that maximizes this function, regardless of their effects on model performance. Filter methods are fast but lack in robustness against feature interactions and redundancy [21]. Wrapper methods use the performance of a learning method to assess the relative usefulness of the selected feature subset (e.g. by cross validation) [27,48]. Wrapper methods employ search algorithms to determine an optimal subset of features. The most generally used search approaches are backward elimination and forward selection [27]. Stochastic algorithms, developed for solving large scale combinatorial problems, such as ant colony optimization, genetic algorithms, and simulated annealing are also used [29,59,61,80,84,90]. Although these algorithms efficiently capture feature redundancy and interaction they are computationally expensive. Recently, some authors take advantage of both filter and wrapper methods and propose hybrid algorithms [33,77]. The idea is to apply first a filter method to select a feature pool and then the wrapper method is applied to determine the optimal subset of features from the selected pool. This can reduce significantly the dimensionality but the global optimality of the selected subset of features is not assured. Embedded methods incorporate feature selection task into the learning process. Just like in wrapper approaches, embedded approaches require therefore a learning algorithm. Embedded approaches have the advantage of integrating the interaction with the learning method, while at the same time being less computationally intensive than wrapper methods. The most popular learning algorithm used in embedded methods is the support vector machines (SVM) [28,85]. Nevertheless, the accuracy of an SVM strongly depends on the choice of the parameters and the kernel function [5]. Even more, they are prone to over-fitting and poor generalization. Therefore to enhance the performance of wrapper methods it is necessary to use fast learning algorithms which performs well when dealing with noisy and imprecise data.

Fuzzy logic was introduced in 1965 by Zadeh [89] and has known a wide use in system control especially for nonlinear systems [52,74]. Its use has been then successfully extended to several applications such as business, medicine, bioinformatics, and computational biology [62,63]. This is mainly due to its interesting fundamental concept which enables to handle and manipulate imprecise and noisy data. In a machine learning framework, the relationship between the outcome and the original feature is usually nonlinear and complicated. However, if the original feature is appropriately fuzzified, the relationship might become nearly linear and a simple classifier can be used [55]. An approach is defined as “fuzzy” if we consider that each pattern belongs with a certain degree of membership to each class, unlike the “hard” approach where each pattern is considered to belong exclusively to only one class [40]. Fuzzy approaches have recently shown their effectiveness in different learning tasks (classification, clustering, feature selection) but their performances decrease significantly whenever high dimensional and/or heterogeneous problems have to be faced [1,4,12,36,67,68]. Although many attempts to develop new fuzzy feature selection have been proposed to cope with this problem, most of these methods end up with a high sophistication [50,56,57,65,69,90]. Either they depend on a specified method (e.g Fuzzy C-Means[56,57,90]) which is designed originally for clustering, or they use an arbitrary choice to determine the linguistic terms of the fuzzified features, which is not always possible and accurate enough whenever a big number of features is faced. For example, a heuristic method based on a min–max learning rule and extension matrix is proposed in [55] for searching the optimal feature subset, but only a random feature fuzzification has been performed. The obtained fuzzy feature space increases rapidly the dimensionality, making the feature selection procedure more complicated and time consuming. Recently, a feature selection approach based on a fuzzy multi-objective ant colony optimization has been proposed [80]. To evaluate the features importance, this approach takes use of an exception ratio which measures the degree of overlaps in the class region. In the approach proposed in [24] a feature selection mechanism is integrated into a genetic learning algorithm in order to reduce its computational cost. For the same aim, a fuzzy entropy is introduced in [53] for selecting relevant features. Schmidt and co-authors in [69] proposed a feature selection approach based on the concept of *Choquet intergral* to improve the performance of fuzzy rule-based classifiers. Although these approaches are able to perform well on intermediate and low dimensional problems, it is still difficult to apply them to accomplish a gene selection task involving several thousands of irrelevant features. Moreover, most of these approaches are designed for handling quantitative features and, if some of features are of mixed type the problem becomes more challenging. We propose in this paper a new approach of feature weighting based on a membership margin to improve the performance of fuzzy rule-based classifiers on high dimensional and heterogeneous problems. Each antecedent fuzzy set in the fuzzy if–then rule is weighted to characterize the importance of each proposition and therefore its corresponding feature. The estimation process of feature's weight is based on membership margin maximization to compute a fuzzy weight of each feature in the membership space. It has been shown that the margin concept can efficiently decrease the computing complexity and standard optimization techniques can be adopted to avoid combinatorial search.

Download English Version:

<https://daneshyari.com/en/article/6857673>

Download Persian Version:

<https://daneshyari.com/article/6857673>

[Daneshyari.com](https://daneshyari.com)