European Journal of Operational Research 230 (2013) 581–595

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

A memetic approach to construct transductive discrete support vector machines

Hubertus Brandner, Stefan Lessmann *, Stefan Voß

Institute of Information Systems, Department for Business and Economics, University of Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany

ARTICLE INFO

Article history:
Received 2 May 2012
Accepted 6 May 2013
Available online 23 May 2013

Keywords:
Data mining
Transductive learning
Support vector machines
Memetic algorithms
Combinatorial optimization

ABSTRACT

Transductive learning involves the construction and application of prediction models to classify a fixed set of decision objects into discrete groups. It is a special case of classification analysis with important applications in web-mining, corporate planning and other areas. This paper proposes a novel transductive classifier that is based on the philosophy of discrete support vector machines. We formalize the task to estimate the class labels of decision objects as a mixed integer program. A memetic algorithm is developed to solve the mathematical program and to construct a transductive support vector machine classifier, respectively. Empirical experiments on synthetic and real-world data evidence the effectiveness of the new approach and demonstrate that it identifies high quality solutions in short time. Furthermore, the results suggest that the class predictions following from the memetic algorithm are significantly more accurate than the predictions of a CPLEX-based reference classifier. Comparisons to other transductive and inductive classifiers provide further support for our approach and suggest that it performs competitive with respect to several benchmarks.

1. Introduction

Classification analysis is an important approach to support decision making in various disciplines including medical diagnosis, information retrieval, risk management and marketing. A classification model categorizes objects into disjoint groups. The group assignment is based on a set of attributes that characterize the objects. Depending on the application, the objects can, e.g., represent patients who are to be categorized into medical risk groups on the basis of symptoms, clinical tests, or their health behavior (e.g. [1–3]). Similarly, financial institutions discriminate between high and low risk loan applicants to support money lending decisions (e.g. [4,5]), and service companies divide customers into loyal clients and likely churners to target retention programs to the right customers (e.g., [6–8]). Independent of the application, classification analysis always aims at constructing a model that predicts group memberships with high accuracy.

The prevailing approach toward classification is to employ a sample of objects with known group memberships. The relationship between the attribute values of these objects and the corresponding class labels is then inferred in an inductive manner (e.g. [9]). Several techniques pursuing this principle have been proposed in Statistics, Machine Learning, and Operations Research. Statistical classifiers often rely on probability theory and estimate the conditional probability (i.e., the *a posteriori* probability) of an object belonging to a class given the object's attribute values (e.g. [10]). Many machine learning methods adopt a data-driven paradigm. For example, tree-based classifiers recursively partition a data set through a sequence of tests on attribute values (e.g. [11]). Eventually, this produces a clear separation of objects of disjoint classes. Operations Research methods typically ground on linear and mixed integer programming (e.g. [12–16]).

In this work, we consider the transductive learning (TL) setting [17]. Standard (inductive) classification aims at creating a global prediction model that facilitates classifying arbitrary decision objects. TL differs from this approach in that it advocates a direct estimation of group memberships for a fixed set of objects called the working set. The fundamental assumption of TL is thus that the decision maker knows all objects that are to be classified in advance. These a priori known decision objects are called the working set. A transductive model can be characterized as a local model that is applicable to working set objects only. The main advantage of TL compared to the more general classification setting is that the additional constraint of a fixed working set simplifies the learning task [17]. This, in turn, will often facilitate more accurate class predictions for working set objects (e.g. [18,19]). With respect to the applicability of TL, it has been shown that several important corporate planning tasks do not require a global model and could potentially benefit from TL [20]. Consequently, developing and testing transductive classification models is an important task to support decision making in organizations.

* Corresponding author. Tel.: +49 (0)40 42838 4706; fax: +49 (0)40 42838 5535.
E-mail addresses: hubertus.brandner@studium.uni-hamburg.de (H. Brandner), stefan.lessmann@uni-hamburg.de (S. Lessmann), stefan.voss@uni-hamburg.de (S. Voß).

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.05.010

From the perspective of inferential statistics, TL is simpler than inductive classification because it explicitly considers the working set objects when building the local classification model (e.g., [21]). In other words, the objects for which classification accuracy matters are taken into account during classifier construction. Though, this approach brings about new algorithmic challenges. First, it is not obvious how to best exploit the predictive information contained in the working set. Second, creating a transductive classifier involves working with both labeled and unlabeled data. This is because the class labels of working set objects are (by definition) unknown. Accommodating labeled and unlabeled data in a learning algorithm is a nontrivial task in its own right. In this work, we propose solutions to these challenges and develop a novel transductive classifier.

Our approach is based on two foundations. First, it relies on the principle of maximal margin separation, which has been put forward in the context of support vector machine (SVM) classifiers (e.g., [22]). The maximal margin principle is also a common approach toward TL (e.g. [18,23–25]). According to the overall risk minimization (ORM) theory, maximizing the margin of separation of a linear classifier helps to minimize a bound of the classifier's error on working set objects (e.g. [21,26]).

Second, our algorithm builds upon the Discrete Support Vector Machine (DSVM) of Orsenigo and Vercellis [27]. DSVMs improve upon standard SVMs in the sense that they implement the principles of statistical learning more accurately [28]. To achieve this, Orsenigo and Vercellis propose to capture classification errors through a discrete step function, which is exactly the notion of errors used in the risk bounds of statistical learning theory. In a large number of simulations, Orsenigo and Vercellis as well as others show that the discrete error measurement produces highly accurate classifiers that outperform standard SVMs and other challenging benchmarks under several experimental conditions [27–33]. We hypothesize that the appropriateness of a discrete error measurement extends to the TL setting. A first contribution of this paper is thus the development of a transductive DSVM (tDSVM) classifier.

Building a transductive classification model is challenging from an algorithmic point of view. In general, classifier construction involves optimizing some measure of model fit over the objects of the training set. Inductive classification algorithms often rely on continuous optimization methods (e.g. [34]). Contrary, the mathematical program underlying classifier construction is typically a mixed integer program (MIP) within TL (e.g. [18,23]). This is also true for Orsenigo and Vercellis's DSVM classifier and our tDSVM in particular. A second contribution of this paper is associated with the development of a memetic algorithm to solve the MIP underlying tDSVMs. Our approach, which we call $\text{tDSVM}_{\text{mem}}$, incorporates population-based and local search operators. We design these operators so as to account for characteristics of the MIP underlying our tDSVM classifier. Additional characteristics of $\text{tDSVM}_{\text{mem}}$ include a self-adaptive tuning of endogenous strategy parameters and an inheritance of solution characteristics.

We test the effectiveness of $\text{tDSVM}_{\text{mem}}$ through several empirical experiments on synthetic data and real-world data from the *UCI Machine Learning Repository* [35]. The results show that $\text{tDSVM}_{\text{mem}}$ performs significantly better than CPLEX. More specifically, whenever the two solvers find the same solution, this solution is also the optimal solution of the corresponding problem instance. Whenever finding an optimal solution is computationally infeasible, $\text{tDSVM}_{\text{mem}}$ gives significantly better objective values than a truncated CPLEX benchmark (i.e., better than the best objective value obtained with CPLEX for a given time limit of reasonable length). We also find that $\text{tDSVM}_{\text{mem}}$ produces classification models that predict significantly more accurately than the CPLEX-based reference classifier. These results confirm the appropriateness of our approach and suggest that $\text{tDSVM}_{\text{mem}}$ is well suited to construct tDSVM classifiers. Regarding the tDSVM classifier itself, we conduct several experiments to assess its predictive performance in comparison to other inductive and transductive methods. The results confirm the effectiveness of a discrete error measurement in TL settings. Furthermore, we find that tDSVM performs often but not always better than other inductive classifiers. This suggests that TL and tDSVM in particular are not necessarily preferable to inductive classifiers, even if class predictions are sought for a known group of working set objects only. Through a set of follow-up experiments, we gain some insight what factors influence the suitability of TL. For example, we observe that the ratio between labeled and unlabeled examples in a data set is an important determinant of TL success. Overall, the analysis allows us to provide some practical recommendations under which circumstances TL is preferable to an inductive approach.

A general implication of our study is that it emphasizes the efficacy of relatively simple heuristic procedures for combinatorial optimization under the condition that the focal problem is well understood, appropriately formalized in a mathematical model, and that the search operators within the heuristic are well adapted to this formulation. Our tDSVM formulation is well grounded in theory and thus captures the learning task in a suitable way. On this basis, a carefully selected set of standard search mechanisms suffices to devise an effective solver and obtain promising results. On the one hand, this evidences the power and generality of the heuristic search framework. On the other hand, it puts the popular approach to extend this framework and invent novel metaheuristics somewhat into perspective. Efforts related with the development of novel metaheuristics are best geared toward novel problems, whereas the techniques known today are well suitable to approach a wide range of standard combinatorial problems. We provide empirical evidence in favor of this view for the problem of building tDSVM classifiers, which can be considered a further contribution of our study.

The paper is organized as follows: Section 2 introduces the original DSVM classifier and explains our modifications to extend it to the TL setting. We then develop our memetic algorithm in Section 3. Section 4 introduces the design of our empirical study. The corresponding results are presented and discussed from an optimization and predictive modeling point of view in Section 5. We conclude the paper with a summary of the main findings and an outlook to future research in Section 6.

2. Classification with transductive discrete support vector machines

The objective of a classification model is to group objects $x_j^\star \in \mathcal{R}^n$ into fixed, disjoint classes $y_j^\star$. In other words, a classifier defines a mapping from objects to classes $f\colon x \mapsto y$. An object is characterized by a set of n attributes. The fundamental assumption of classification analysis is that attribute values determine class memberships. However, the specific (functional) relationship between attribute values and class memberships is unknown. A classification method strives to reconstruct this relationship from a training sample L that consists of objects with known class labels $L = \{x_i, y_i\}_{i=1}^{l}$. The model resulting from this step facilitates predicting the class memberships of novel objects $U = \left\{x_j^\star\right\}_{j=1}^{u}$. Without loss of generality [36], we concentrate on binary classification problems in this paper and assume that $y_j^\star \in \{\pm 1\}\ \forall\, j$.

2.1. Discrete support vector machines

SVMs are a popular approach toward classification. They are inspired by statistical learning theory and the principle of structural-risk-minimization (SRM) in particular [17]. Roughly