Expert Systems With Applications 62 (2016) 1–16

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification

Aytuğ Onan[a,b,*], Serdar Korukoğlu[b], Hasan Bulut[b]

[a] Celal Bayar University, Department of Computer Engineering, 45140, Muradiye, Manisa, Turkey
[b] Ege University, Department of Computer Engineering, 35100, Bornova, Izmir, Turkey.

ARTICLE INFO

Article history:
Received 7 December 2015
Revised 10 May 2016
Accepted 3 June 2016
Available online 7 June 2016

Keywords:
Sentiment analysis
Ensemble learning
Weighted majority voting
Multiobjective optimization

ABSTRACT

Typically performed by supervised machine learning algorithms, sentiment analysis is highly useful for extracting subjective information from text documents online. Most approaches that use ensemble learning paradigms toward sentiment analysis involve feature engineering in order to enhance the predictive performance. In response, we sought to develop a paradigm of a multiobjective, optimization-based weighted voting scheme to assign appropriate weight values to classifiers and each output class based on the predictive performance of classification algorithms, all to enhance the predictive performance of sentiment classification. The proposed ensemble method is based on static classifier selection involving majority voting error and forward search, as well as a multiobjective differential evolution algorithm. Based on the static classifier selection scheme, our proposed ensemble method incorporates Bayesian logistic regression, naïve Bayes, linear discriminant analysis, logistic regression, and support vector machines as base learners, whose performance in terms of precision and recall values determines weight adjustment. Our experimental analysis of classification tasks, including sentiment analysis, software defect prediction, credit risk modeling, spam filtering, and semantic mapping, suggests that the proposed classification scheme can predict better than conventional ensemble learning methods such as AdaBoost, bagging, random subspace, and majority voting. Of all datasets examined, the laptop dataset showed the best classification accuracy (98.86%).

1. Introduction

An important subfield of machine learning, ensemble learning aims to construct classification models with better predictive performance by combining the predictions of multiple learning algorithms. In constructing effective ensemble models, identifying base learning algorithms that can perform the classification task is crucial, which ideally involves classifiers with a range of outputs and structures (Ditterich, 2000). Also critical to the predictive performance of ensemble methods is identifying an appropriate combination scheme for base learning algorithms (Moreno–Seco, Iñesta, León, & Micó, 2006).

Assigning appropriate weight values to classifiers can be modelled as an optimization problem, whose optimal solutions can be provided by the well-established means of metaheuristic algorithms. Briefly, metaheuristic methods can be broadly categorized into two groups: one based on single solutions, the other based on population (Talbi, 2009). Whereas single-solution-based metaheuristics encompass simulated annealing, variable neighborhood search, guided local search, and Tabu Search, population-based metaheuristics include ant colony optimization, particle swarm optimization, differential evolution, and genetic algorithms. Among fields of application for metaheuristic algorithms, data mining and machine learning are only two examples (Gogna & Tayal, 2013), for metaheuristic algorithms have been successfully applied in optimizing configurations in ensemble learning (Chen & Wong, 2010) and in identifying the optimal subset of classifiers from a large pool of candidates (Santos, 2012).

Sentiment analysis (i.e., opinion mining) is the process of identifying the contextual polarity of text documents by using tools and techniques from natural language processing, machine learning, and statistics. Among sources of information for sentiment analysis, the Internet comprises an ever-increasing amount of data (Bhaita & Khalid, 2008) in several forms of multimedia (e.g., text and video) that can provide valuable information to governments, business organizations, and individual decision makers. Typical areas of application for sentiment analysis, for example, include identifying public sentiment for the purpose of policymaking and the market analysis of products and services based on customer reviews (Zhang, Zeng, Li, Wang, & Zuo, 2009). In that sense, for

* Corresponding author: Fax: +90 232 3399405.
E-mail addresses: aytug.onan@cbu.edu.tr, aytugonan@hotmail.com (A. Onan), serdar.korukoglu@ege.edu.tr (S. Korukoğlu), hasan.bulut@ege.edu.tr (H. Bulut).

http://dx.doi.org/10.1016/j.eswa.2016.06.005
0957-4174/

decision-making support systems and individual decision makers alike, structured, insightful knowledge obtainable by identifying subjective information from online content can be extremely valuable (Fersini, Messina, & Pozzi, 2014). More specifically, sentiment analysis can be divided into three primary levels: of documents, of sentences, and of entities or aspects (Liu, 2012). Document-level sentiment analysis aims to identify the sentimental orientation of the entire text document, whereas sentence-level analysis aims to identify whether a particular sentence expresses a positive, negative, or neutral opinion. By contrast, more sophisticated analysis at the entity or aspect-level can fully reveal attitudes or opinions toward different features or aspects of a particular entity (Medhat, Hassan, & Korashy, 2014).

In this paper, we present an ensemble classification scheme for text sentiment classification combining base learning algorithms by a novel weighted voting scheme, for which identifying algorithms that can be used as base learners is crucial. Based on an extensive empirical analysis of several classifiers, we identified an optimal set of base learning algorithms, including Bayesian logistic regression, Naïve Bayes, linear discriminant analysis, logistic regression, and support vector machines.

To combine the outputs of base learning algorithms with an optimal weighting scheme, we evaluated several soft computing-based schemes of weighted voting, including ones based on a genetic algorithm, multiobjective particle swarm optimization, multiobjective simulated annealing, and multiobjective differential evolution. To the best of our knowledge, our paper is the first to extensively analyze the performance of conventional and soft computing-based weighted voting schemes for sentiment analysis. Although we used the two novel soft computing-based weighting schemes for text sentiment classification, they are readily adaptable to other domains provided the careful selection of base learning algorithms. Comparative experiments on benchmark classification datasets from several other domains—software defect prediction, credit risk modeling, spam filtering, and semantic mapping—demonstrated the validity and effectiveness of the proposed classification scheme.

The contributions of the paper can be summarized as follows:

- We have proposed a novel, efficient ensemble classification scheme based on an optimization technique using a multiobjective differential evolution algorithm to efficiently combine the classification algorithms. Our experimental comparisons of conventional weighted voting schemes (e.g., simple weighted voting, rescaled weighted voting, best–worst weighted voting, and quadratic best–worst weighted voting) and well-known metaheuristic weighted voting schemes (e.g., genetic algorithm-based weighted voting, multiobjective simulated annealing-based weighted voting) in sentiment analysis indicate that the proposed scheme outperforms the methods compared.
- To the best of our knowledge, our study is the first to use weighted voting schemes in ensemble classification for sentiment analysis and to use a multiobjective differential evolution algorithm to adjust weights in ensemble learning. Whereas our use of multiobjective differential evolution is a novel contribution to ensemble learning, our use of metaheuristic weighted voting is a novel contribution to sentiment analysis.
- Although the proposed classification scheme was developed based on an empirical analysis of sentiment analysis, the proposed scheme was also applied to several other classification tasks, including software defect prediction, credit risk modeling, spam filtering, and semantic mapping—all domains for which the classification scheme shows promising results.
- Though majority voting schemes are widely used in classification tasks, including sentiment analysis, conventional and metaheuristic weighted voting schemes for sentiment analysis receive less attention. To overcome that oversight, in this original comprehensive analysis of voting schemes for text sentiment analysis, we present an empirical comparison of five unweighted voting schemes (i.e., average of probabilities, product of probabilities, maximum of probabilities, minimum of probabilities and majority voting), four weighted voting schemes (i.e., simple weighted voting, rescaled weighted voting, best–worst weighted voting, and quadratic best–worst weighted voting), and four metaheuristic-based weighted voting schemes (i.e., genetic algorithm, differential evolution, simulated annealing, and particle swarm optimization), all for sentiment analysis.
- Along with an analysis of voting schemes, our paper presents a comprehensive empirical analysis of well-known ensemble methods (e.g., AdaBoost, bootstrap aggregation or bagging, dagging, random subspace, and stacking) for sentiment analysis.
- Since identifying base learners for classifier ensembles is crucial to developing an efficient ensemble scheme, we experimented with using a statistic selection model based on majority voting error and forward search to identify optimal classifiers. Based on an extensive empirical analysis, we present an optimal set of classifiers for text sentiment analysis using five base learning algorithms: Bayesian logistic regression, Naïve Bayes, linear discriminant analysis, logistic regression, and support vector machines.

In what follows, Section 2 briefly reviews the state of the art in ensemble learning and sentiment analysis, after which Section 3 presents the theoretical foundations of our study and Section 4 the proposed weighted voting scheme. Section 5 presents the results of our experiments with sentiment analysis datasets, whereas Section 6 presents experimental results for datasets from several other domains. Lastly, Section 7 concludes the paper with a review of the study's contributions and limitations, as well as with suggestions for future research.

2. Related literature

This section briefly reviews related work on ensemble learning in sentiment analysis and weighted voting schemes in ensemble learning.

2.1. Ensemble learning in sentiment analysis

Thanks to their predictive performance, machine learning methods have been widely used in sentiment analysis. By extension, as recent studies addressing sentiment analysis based on machine learning have indicated, ensemble learning can enhance the predictive performance of sentiment classification (Prabowo & Thelwall, 2009; Wang, Sun, Ma, Xu, & Gu, 2014; Xia, Zong, & Li, 2011). An important subfield of machine learning and natural language processing, sentiment classification is often subject to research focused on linguistic characteristics or machine learning-based features, if not both, in order to obtain more robust classification models for sentiment analysis. For instance, Xia et al. (2011) examined the performance of several linguistic features of text documents, including part of speech, word relation, and term frequency–inverse document frequency (TF–IDF), in conjunction with ensemble learning methods for sentiment classification. In that scheme, several classification algorithms such as Naïve Bayes, maximum entropy, and support vector machines were combined by fixed-rule output combination, meta-classifier, and weighted