Election Outcome and Political Party Prediction using SVM and k-NN Classifiers

Behzad Savabi (338222)
behzad.savabi@aalto.fi

Abstract

In this project, we aimed to predict the outcome of an election for a set of candidates based on the answers to a set of predefined questions as well as the properties of comments about each candidate by using an ensemble of support vector machine classifiers. Moreover, for each candidate, we predicted the political party, which he/she belongs to by using a k-nearest neighborhood classifier. The experimental results showed that these classifiers with a set of properly selected features are able to predict the election outcomes and political parties of the candidates with a convenient accuracy.

1 Introduction

During this course, we have been familiarized with using classification algorithm on real datasets. Challenging to choose the best method helped us understand the importance of comparing results and trying to test different methods. Undoubtedly, supervised classification is one of the most important tasks carried out by Intelligent Systems. Therefore, plenty of methods have been developed based on Articial Intelligence (Logic-based techniques, Perceptron-based techniques) and Statistics (Bayesian Networks, Instance-based techniques). The goal of supervised learning is to build a concise model of the distribution of the class labels in terms of predictor features. The resulting classifier is then used to assign class labels to the new instances where the values of the predictor features are known, but the value of the class label is unknown.
In this project, two tasks have been carried out; first, predicting the winner of the election; second, predicting the political party of every candidate. Classifications are based on the answers to a set of predefined questions and the properties of comments about each candidate. We predicted the political party of every candidate based on their background & political views; in machine learning and statistics this kind of problem is called classification. Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. The individual observations are analyzed into a set of quantifiable properties, known as various

explanatory variables, features, etc. These properties may variously be categorical, ordinal, integer-valued or real-valued. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. For this task, we utilized the k-nearest neighbors (k-NN) classifier [1] to predict political parties (classes). k-NN classifier algorithm is a non-parametric method for classification and regression that predicts objects' "values" or class memberships based on the k closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. However, k-NN classifier is a very intuitive method that classifies unlabeled examples based on their similarity to examples in the training set.
On the other hand, in order to predict the outcome of an election based on candidates answers we have utilized support vector machines (SVM) [1]. In machine learning, SVM is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, this method used for regression analysis and classification. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories; an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

2 Methods

In the following, we describe our methods for predicting the outcome of the election as well as the party of each candidate.

2.1 Feature Selection

Since the dimension of the input feature vectors may be large, we must consider a method to reduce the dimension of the input, efficiently. However, we should note that in some cases, not all the features provided for the task are informative and sometimes they may even cause misclassification. Therefore, we considered a forward feature selection strategy [2] for each experiment to select an appropriate subset of features by first, defining a cost function which given a set of training and test instances as well as their corresponding labels, calculates $(1-\text{F-score})$ for the prediction using a base classifier and then, adding the feature which decreases the cost function over the validation set, the most. (stopping criterion)In our experiments, the feature set $\{22, 45, 46, 98, 149\}$ was selected by the algorithm for the election output prediction experiment where 23 features were chosen for the party prediction. This also eliminates the need

Method	F-score	ϕ-score	ACC	FP-Rate	SENS	Pos. Pred.
PCA	0.1890	0.0956	0.6084	0.3872	0.5581	0.1137
Proposed	0.4828	0.4362	0.9144	0.0476	0.4884	0.4773

Table 1: Experimental results for election outcome prediction

for any dimensionality reduction technique to reduce the dimension of the input features.

3 Classification

3.1 Election Outcome Prediction

We performed a 10-fold cross validation on the training data. Since the number of positive instances is much smaller than the negative ones, we considered an ensemble of support vector machine (SVM) classifiers where each classifier was trained using all the positive instances along with a set of randomly selected negative instances from the negative examples. We trained 101 SVM classifiers with radial basis function (RBF) kernel function and selected the sigma value, which maximized the F-score on the validation set. We used a majority-voting scheme where we considered the result of the classification as positive if the results of more than half of the classifiers were positive for that instance.

3.2 Political Party Prediction

We considered a k-NN classifier for predicting the political party of each candidate. We selected the optimal number of neighbors for each party by performing a 10-fold cross validation and maximizing the F-score over the validation set.

4 Experimental Results

The data consists of a set of 199 features, which are answers to a set of predefined questions as well as properties of the comment e.g. the length of the textual answer for 1300 different candidates from 17 different political parties for a Finnish election event. The data has been preprocessed and normalized so each feature has zero mean and unit variance over the set of all candidates. We selected an appropriate subset of features for each experiment, as described in Section 2. We performed a 10-fold cross validation on the training data as mentioned in Section 3. We found the optimal value of sigma for the SVM classifiers to be 1 and the optimal number of neighbors equal to 1 for the k-NN classifier. We also performed an experiment where we used principal component analysis (PCA) [3] to reduce the dimension of the original data to the corresponding values used in our previous experiments, namely 5 and 10 for election outcome and political party prediction, respectively. The results are shown in Table 1

Method	F-score	ϕ-score	ACC	FP-Rate	SENS	Pos. Pred.
PCA	0.1828	0.1441	0.9141	0.0472	0.1879	0.1998
Proposed	0.2501	0.2086	0.9224	0.0426	0.2500	0.2539

Table 2: Experimental results for political party prediction (average over all parties)

and Table 2. As it can be seen, the feature selection approach outperforms the dimensionality reduction in all the cases. This indicates that using all the features in the dataset for a particular task is not always appropriate and it can even be misleading.

5 Conclusion

We predicted the outcome of an election for a set of candidates as well their corresponding political parties using ensemble of SVM and k-NN classifiers, respectively. The results show that these classifiers provide an appropriate performance for the task. Additionally, we showed that for these tasks, a feature selection scheme outperforms the dimensionality reduction approach. This can be justified by the fact that in some scenarios, a feature or a subset of features are not informative and may lead to a reduced accuracy.

References

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[2] Ethem Alpaydin. *Introduction to Machine Learning, second edition.* The MIT Press, 2010.

[3] J.E. Jackson. *A User's Guide to Principal Components*. Wiley, 1991.