A review of “Detecting Concept Drift with Support Vector Machines (2000)”

Introduction

The paper is prepared by Ralf Klinkenberg & Thorsten Joachims in 2000. The objective of the paper is to recognize and handle concept changes with support vector machines (SVM). The method maintains a window on the training data. The key idea is to automatically adjust the window size so that the estimated generalization error is minimized. The window adjustment approach described in this paper uses support vector machines (Vapnik, 1998) as their core learning algorithm. The new approach is both theoretically well-founded as well as effective and efficient in practice. Since it does not require complicated parameterization, it is simpler to use and more robust than comparable heuristics. The paper includes an experiment which shows the efficiency of this method on real world concept drift scenario of text data. The experiments use a subset of 2600 documents of the data set of TREC consisting of English text. The texts are randomly split into 20 batches of equal size containing 130 documents each. Each text is assigned some categories from 1 to 6. To perform the actual experiment, we will take four data management approaches: Full memory, No memory, Window of fixed size and window of adaptive size. It is done three times with three different scenarios of concept drift.

Pros and Cons

There are many pros of the method used in the paper. Some of them are briefly discussed below:

* **Does not require complex parameter tuning:** The method used here uses simple SVM to select an appropriate window size that does not involve complicated parameterization. They key idea is to select the window size so that the estimated generalization error on new examples is minimized. Helmbold, & Long in 1991, Helmbold and Long 1994, Widmer & Kubat in, 1996, Lanquillon in 1997 and Klinkenberg & Renz in 1998 proposed many model to tackle concept drift but those were intuitive and work well in their particular application domain, they usually require tuning their parameters.
* **Consume less time and lesser calculation:** Many other method on handling concept drift working on SVM uses complex and time taking calculations to predict the errors by iterating the leave-one-out process over all the data. While the leave-one-out estimate is usually very accurate, it is very expensive to compute. With a training sample of size n, one must run the learner n times. Our method use ksaai alpha estimators to overcome this problem using an upper bound on the number of leave-one-out errors instead of calculating them brute.
* **Works under the constraint:** The model works on the assumption that the next batch of the data will be similar to the immediate previous batch. If this is not the case, the model will fail deliver accuracy and will give false prediction. However the real world data is also similar to the assumption.

Technical Error

The estimator is pessimistically biased, overestimating the true error rate on average. The model just uses the average error rate and not the true error rate. However experiments show that the bias is acceptably small for text classification problems and that the variance of the estimator is essentially as low as that of a holdout estimate using twice as much data. So we can say that the model will work fine for text classification task, but the accuracy of the model is under question for any other task where the bias is large enough for acceptance.

Also there are only three scenarios taken during experimentation, all of them following the assumption we have made, if there were more scenarios which does not follow the assumption we would have knowledge about the accuracy of this model under unfavourable conditions.

Suggestion

The approach is better for task where concept drift takes place less frequently. We can work on a model which can sense a pattern of concept drift and make a generalised model for this. Here in the experiment we are taking a subset of 2600 documents of the data set consisting of English text. The texts are randomly split into 20 batches of equal size containing 130 documents each. Each text is assigned some categories from 1 to 6.

Here we took only four data management approaches: Full memory, No memory, Window of fixed size and window of adaptive size. We can test different other adaptive methods also discussed in many previous papers. It is done three times with three different scenarios of concept drift, we can test in under extreme condition to check if it is a perfect generally accepted model.

Further Research

There can further research to if there is any pattern of concept drift and if we can learn to predict concept drift. As this model has peaked the task of predicting concept drift. Instead of handling concept drift, now we can work on predicting concept drift.

Prajjaval Verma

B19ME054