* Pages – 8/9 seems to be the sweet spot
* Total word count – 5000 – 6500
* Introduction – including abstract, 1 – 1.5 pages (600 – 1000 words)
  + C
* Related work – 0.5 – 1 pages (400 – 800 words)
  + X
* Methodology (1.5 – 3 pages)
  + Problem definition
    - X
  + Adaptation technique
    - X
* References – 20 - 25

# Detecting and Adapting to Concept Drift in Continually Evolving Stochastic Processes

Upeka Premaratne

University of Moratuwa, Sri Lanka

upeka@ent.mrt.ac.lk

Sunanda Gamage

rcsunanda@gmail.com

**Abstract**

xxx

## 1. Introduction

The vast majority of machine learning research focuses on building models (eg: regression models, classifiers) that are static; i.e. the model is trained one time on a training dataset and then applied to test data. This works fine when the underlying stochastic process that is modeled remains stationary. However, in many real world applications, the process is non-stationary, and it continuously changes with time.

For example, consider the video feed from a traffic camera, which can undergo changes from morning to night, and from summer to winter. The video feed process in this example undergoes gradual continuous change. The drift can also be abrupt and large, such as an impactful market event that affects stock prices.

The underlying stochastic process in these examples is non-stationary, which means that the joint probability distribution of the process attributes, , is time varying. This is known as concept drift. Applying a model that is trained one time on some initial training dataset to such a process results in performance degradation. Therefore, methods to characterize the concept drift, and adapt machine learning models to drift by incremental learning are an important research direction in machine learning.

So far, these are the methods. But they are not much for continuous, or

In this work, we improve on the drift detection techniques used in [A] and [B], and present an efficient ensemble technique of incremental learning to model a continuously changing stochastic process. The improved drift detection technique tracks a difference metric (what metric?) between probability distributions estimated from two sample windows before and after a time point. The estimated distribution is typically the joint probability distribution of attributes for the unsurprised case, or joint attribute and class distribution for the supervised case. When the continuous sum of this metric since the last adaptation rises above a threshold , it is decided that an adaptation should be made.

At this point, a new model is learned from a suitable number of most recent samples, and it is added to an ensemble of previously learned models. The final time varying model is a weighted some of the sub models in the ensemble. The weights can be set in a way such that older sub models have less relevance (a form of forgetting). They can also be set to reflect the similarity between the probability distributions of the current sample window, and the windows corresponding to each sub model. In the latter case, the distributions relevant to each sub model also need to be stored.

Our drift detection and adaption techniques can act as a wrapper method independent of the machine learning model being employed. The drift detection technique can be made to detect different types of drift [A] by choosing the relevant distribution for computing the difference metric. The methods work well for both the supervised and the unsupervised cases, and the adaptation does not cause catastrophic forgetting. The forgetting factor can in fact be controlled by setting appropriate weights on the sub models in the ensemble. experiment results of evaluating the methods on artificial and real-life datasets show good performance in detecting drift and adapting to it.

The rest of this paper is organized as follows. Section 2 presents related work in detecting concept drift and adaptation methods. In Section 3, the drift problem is formally defined, and solutions for drift detection and adaptation are described. In Section 4, we presents the result of evaluating the method on several artificial and real-life datasets. Conclusions are given in Section 5.

## 2. Related Work

xxx

## 3. Methodology

### 3.1 Problem definition

Xxx

### 3.2 Detection technique

### 3.3 Adaptation technique

xxx

## 4. Experimental Evaluation

xxx

### 4.1 Simulations and results

xxx

### 4.2 Adaptation on network traffic characteristics (or other real life dataset)

xxx

## 5. Conclusion

xxx

## References

[A] Characterizing Concept Drift - Geoffrey I. Webb

[B] Learning with Drift Detection – J. Gama