Learning Drifting Data Using Selective Sampling

March 23, 2014

In some on-line learning settings, the optimal function, or classifie,r from a family of functions or classifiers, changes over time, either by gradually drifting or by an abrupt shift (or switch). This occurs due to changes in data distribution or other conditions in real-life problems with which on-line learning algorithms deal, such as information filtering, sentiment analysis, market analysis and more.

In selective sampling approach to on-line learning settings, not all labels or values of input instances are queried, while algorithm performance remains unharmed. Selective sampling suggests a concept of confidence - on instances algorithm is confident about, prediction should be accurate and true value can remain unknown, while maintaining similar future results.

In this work we propose to use the concept of selective sampling to overcome drifting data in on-line learning. We focus on the case of a sudden, abrupt shift (or switch). We use demo classifiers, constructed from a recent time window to evaluate algorithm performance. If our estimator differs significantly from demo classifier on instances we have high confidence about, we assume it is due a switch, a conduct further steps accordingly by restating algorithm to avoid now irrelevant and in-correct former data. We provide assurance that no false detections occur. Furthermore we show that if a switch is remained undetected then we can assure it causes only small harm. We provide analysis for on-line learning classification and regression settings, and show results on synthetic data.