Invited Talk : Ensemble Learning from Data Streams with Active and Semi-Supervised Approaches

Bartosz Krawczyk

Department of Computer Science
Virginia Commonwealth University, Richmond, VA
bkrawczyk@vcu.edu

Abstract. Developing efficient classifiers which are able to cope with big and streaming data, especially with the presence of the so-called concept drift is currently one of the primary directions among the machine learning community. This presentation will be devoted to the importance of ensemble learning methods for handling drifting and online data. It has been shown that a collective decision can increase classification accuracy due to mutually complementary competencies of each base learner. This premise is true if the set consists of diverse and mutually complementary classifiers. For non-stationary environments, diversity may also be viewed as a changing context which makes them an excellent tool for handling data shifts. The main focus of the lecture will be given to using these mentioned advantages of ensemble learning for data stream mining on a budget. As streaming data is characterized by both massive volume and velocity one cannot assume unlimited access to class labels. Instead methods that allow to reduce the number of label queries should be sought after. Recent trends in combining active and semi-supervised learning with ensemble solutions, such as online Query by Committee or Self-Labeling Committees, will be presented. Additionally, this talk will offer discussion on emerging challenges and future directions in this area.