Hamed Bonab

Research Statement

140 Governors Dr., Amherst, MA 01003

(631) 215 4176

✉ *bonab@cs.umass.edu*

Evolving Stream Processing

Designing adaptive classifiers for evolving stream data is a challenging task due to the size and changing nature of data streams. Among existing classifiers, ensemble-based approach is one of the best solutions to this problem. However, there are a few data-fusion algorithms for combining the scores of classifiers in an ensemble classifier. This becomes more important in evolving environments because in different situations it is possible that a subset of classifiers in the ensemble outperforms others. We introduce a novel geometric framework for weighting and combining classifiers in an ensemble. In this framework, we propose a dynamic weighting approach for classifiers called "Geometrically Optimum and Online Weighted Ensemble Classifier (GOOWE)" based on the Euclidean distance between scores and ideal points. In order to measure the robustness of proposed framework, we used real-world datasets and synthetic data generators using MOA Libraries. All the work is implemented in Java programming language using MOA and Weka libraries[1]. We also adapted our geometrical environment to multi-label environments where instead of single class, an instance can be considered with many different labels[2]. This is extremely important for text classification and many other applications. Our theoretical framework, which is applicable to both batch and online environments, shows that the number of classifiers is highly dependent on the number of class labels, where the aggregation method is weighted majority voting [3, 4].

Citation Worthiness for Scientific Text

On this research direction, our aim is to focus more on detecting sentences where a citation is necessary. There are many citation recommendation systems where for any user specified sentence they are able to suggest some published references and let the user choose in between. However, there was no study on finding actual sentences that require a citation to notify to the user. We created citation worthiness dataset and trained neural network based and SVM based classifiers for discriminating such sentences [5].

Neural Cross-lingual Information Retrieval

There are very limited number of studies focusing on standard CLIR tasks in which neural models are used. We aim to tackle this problem by handling the translation and relevance sub-problems using an end-to-end neural system. We are also focused on understanding low-resource languages to better understand the constraints in terms of resources for high and low resources, given that most of the neural models are trained for English language. This is an ongoing project which I work actively on this area.

Publications

H. R. Bonab and F. Can, "Goowe: Geometrically optimum and online-weighted ensemble classifier for evolving data streams," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 12, no. 2, p. 25, 2018.

A. Büyükçakir, H. Bonab, and F. Can, "A novel online stacked ensemble for multi-label stream classification," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018, pp. 1063–1072.

H. R. Bonab and F. Can, "A theoretical framework on the ideal number of classifiers for online ensembles in data streams," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 2053–2056.

H. Bonab and F. Can, "Less is more: a comprehensive framework for the number of components of ensemble classifiers," *IEEE Transactions on neural networks and learning systems*, 2019.

H. Bonab, H. Zamani, E. Learned-Miller, and J. Allan, "Citation worthiness of sentences in scientific reports," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 2018, pp. 1061–1064.