

ABClass: A multiple instance learning approach for sequence data

Manel Zoghliami^{1,2} Sabeur Aridhi³ Mondher Maddouri⁴ Engelbert Mephu Nguifo ¹

¹University of Clermont Auvergne, LIMOS, PB 10125, 63173 Clermont Ferrand, France

² University of Tunis El Manar, Faculty of sciences of Tunis, LIPAH, 1060 Tunis, Tunisia

³ University of Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

⁴ University of Jeddah, College Of Business, PB 80327, 21589 Jeddah, KSA

manel.zoghliami@gmail.com

In Multiple Instance Learning (MIL) problem for sequence data, the learning data consist of a set of bags where each bag contains a set of instances/sequences. In some real world applications such as bioinformatics comparing a random couple of sequences makes no sense. In fact, each instance of each bag may have structural and/or temporal relation with other instances in other bags. Thus, the classification task should take into account the relation between semantically related instances across bags.

In this work, we present ABClass, a novel MIL approach for sequence data classification. Each sequence is represented by one vector of attributes extracted from the set of related instances. For each sequence of the unknown bag, a discriminative classifier is applied in order to compute a partial classification result. Then, an aggregation method is applied in order to generate the final result. We applied ABClass to solve the problem of bacterial Ionizing Radiation Resistance (IRR) prediction. The experimental results were satisfactory.

Availability: ABClass tool and the used dataset can be found in the following link :

<http://homepages.loria.fr/SAridhi/software/MIL/> . ABClass runs on a Windows or a Linux platform (tested on Ubuntu distribution) that contains a java JRE.