Mehdi Kaytoue
INSA Lyon, Université de Lyon
LIRIS-UMR CNRS 5205
Lyon, France.
mehdi.kaytoue@liris.cnrs.fr

March 31, 2017

Pr. Peter A. Flach, Machine Learning Journal Editor-in-chief
Pr. Michelangelo Ceci and Pr. Toon Calders, DS'16 Guest Editors

We are pleased to submit our article entitled *Diverse Subgroup Set Discovery in Multi-Label Data* as a candidate for publication in the journal *Machine Learning*, for the *Discovery Science 2016 Special Issue*.

Subgroup Discovery (SD) is a leading data mining technique that allows one to elicit descriptions that unexpectedly occur with a class target. In presence of multi-label data, there are many applications for which one is interested in subgroups that differ from the whole dataset only on subsets of labels, and the interesting subsets of labels are not known beforehand: Typically, descriptive rules that conclude on small label sets. SD, and its extension for more complicated target concepts, Exceptional Model Mining (EMM), fail to produce such rules: They consider either the whole set of labels, each label independently, or a unique and fixed label subset chosen a priori as surveyed by Duivesteijn et al, Data Mining and Knowledge Discovery, 2016 [17].

In this article, we propose to enhance EMM for considering all target subspaces (label subsets): It requires to revisit its formalization, the subgroup search space, and to propose new quality measures expressing how exceptional is a subgroup. Our quality measures consider label distributions dynamically during a heuristic search (i.e., beam search and Monte Carlo Tree Search) that outputs a diversified set of high quality subgroups, whose redundancy is controlled not only on the covered objects, but also on the considered label sets. This is shown through an extensive set of experiments. Finally, we discuss an application in neurosciences which argue the actionability of the discovered subgroups.

Our submission extends a work presented at the DS 2016 Conference. We completely revisited the writing of the article, re-formalized the problem as an EMM instance where diversity has to be taken into account also on targets, introduced and compared several quality measures (with better results), implemented and successfully compared the MCTS discovery technique and proposed a deeper set of experiments.

Bosc, G., Golebiowski, J., Bensafi, M., Robardet, C., Plantevit, M., Boulicaut, J.-F., Kaytoue, M.: Local subgroup discovery for eliciting and understanding new structure-odor relationships. In: Discovery Science, eds. T. Calders, M. Ceci and D. Malerba, pp. 19--34, Lecture Notes in Computer Science 9956, Springer (2016).

We thank you for taking the time to review our request, and hope that you will be convinced to accept starting the reviewing process.

Yours sincerely,
The authors