Expanding MLkNN Using Extended Rough Set Theory

Gabriela Pérez[1]
Marilyn Bello[1,2]
Gonzalo Nápoles[2],
María Matilde García[1]
Rafael Bello[1]
Koen Vanhoof[2]

[1]Department of Computer Science,Central University "Marta Abreu" of Las Villas, Cuba. mbgarcia@uclv.cu
[2]Faculty of Business Economics, Hasselt University, Hasselt, Belgium

Abstract

Multi-label classification refers to the problem of associating an object with multiple labels. This problem has been successfully addressed from the perspective of problem transformation and adaptation of algorithms. MultiLabel k-Nearest Neighbour (MLkNN) is a lazy learner that has reported excellent results, still there is room for improvements. In this paper we propose a modification to the MLkNN algorithm for the solution to problems of multilabel classification based on the Extended Rough Set Theory. More explicitly, the key modifications are focused in obtaining the relevance of the attributes when computing the distance between two instances, which are obtained using a heuristic search method and a target function based on the quality of the similarity. Experimental results using synthetic datasets have shown promising prediction rates. It is worth mentioning the ability of our proposal to deal with inconsistent scenarios, a main shortcoming present in most state-of-the-art multi-label classification algorithms.

Keywords: *Multi-label classification k-Nearest Neighbour Extended Rough Set Theory Measure Quality of Similarity*