A PageRank-based method to extract fuzzy expressions as features in supervised classification problems*

*Note: The full contents of this paper have been published in the volume *Lecture Notes in Artificial Intelligence 11160* (LNAI 11160)

Pablo Carmona
Industrial Engineering School
University of Extremadura
Badajoz, Spain

Juan Luis Castro
Department of Computer Science and Artificial Intelligence
University of Granada
Granada, Spain

Jesús Lozano
Industrial Engineering School
University of Extremadura
Badajoz, Spain

José Ignacio Suárez
Industrial Engineering School
University of Extremadura
Badajoz, Spain

***Abstract*—This work presents a new ranking method inspired on PageRank to reduce the dimensionality of the feature space in supervised classification problems. More precisely, as it relies on a weighted directed graph, it is ultimately inspired on TextRank, a PageRank based method that adds weights to the edges to express the strength of the connections between nodes. The method is based on dividing each original feature used to describe the training set into a set of fuzzy predicates and then ranking all of them by their ability to differentiate among classes in the light of this training set. The fuzzy predicates with the best scores can be then used as new features, replacing the original ones. The novelty of the proposal relies on being an approach halfway between feature selection and feature extraction approaches, being able to improve the discrimination ability of the original features but preserving the interpretability of the new features in the sense that they are fuzzy expressions. Preliminary results supports the suitability of the proposal.**

***Index Terms*—fuzzy logic, supervised classification, ranking methods, feature selection, feature extraction, PageRank, TextRank**