Fuzzy feature ranking method based on PageRank

Pablo Carmona[1], Juan Luis Castro[2], and José Ignacio Suárez[1]

[1] Industrial Engineering School, UEx, Badajoz, Spain
{pablo,jmarcelo}@unex.es
[2] Department of Computer Science and Artificial Intelligence, UGr, Granada, Spain
castro@decsai.ugr.es

Abstract. This work proposes a new fuzzy feature ranking method based on PageRank in the context of supervised classification. In the directed graph in which the algorithm relies on, the nodes represent fuzzy expressions and the weights of the arcs assess the degree of information gained when adding the head fuzzy expression to the tail one.

Keywords: fuzzy logic, classification, feature ranking, PageRank

1 Introduction

In many contexts where a huge amount of information need to be analyzed, the overwhelming processing requirements can be mitigated by evaluating the usefulness of each pieze of information for the goal we are pursuing and selecting the best ones. In supervised classification problems, where the data is expressed through a set of labeled examples (each example represented by a set values for a fixed set of features and its class label), those techniques fits with the concept of feature selection [1]. That way, the best features are selected from the whole set based in their usefulness to correctly classify new examples.

We propose a new ranking method for feature selection inspired by PageRank [2], a well-known ranking algorithm initially used by Google web to rank the results from a query and widely extended to other contexts [3, 4]. This algorithm relies in a directed graph model that is traversed by a random walker. In our proposal, the nodes represent fuzzy expressions and each directed arc has a weight that assess the degree of information gained by adding the expression in the head node to the expression in the tail node. This weight is based on concept of the entropy of the fuzzy expressions in the light of the labeled examples.

2 The Feature Ranking Method

We consider a set of d classes $\mathcal{C} = \{c_1, \dots, c_d\}$, n features $\mathcal{F} = \{f_1, \dots, f_n\}$ and m examples $\mathcal{E} = \{e_1, \dots, e_m\}$, each example with the form $e_i = ([x_i^1, \dots, x_i^n], c^i)$, where x_i^j is its value for the feature f_j and c^i is its class label. Each feature f_i is associated with a fuzzy variable X_i whose fuzzy domain is denoted as

$\widetilde{\mathcal{X}}_i = \{LX_i^1, \ldots, LX_i^{p_i}\}$, being p_i the number of fuzzy values associated with the variable and being LX_i^j the linguistic label of its jth fuzzy value.

The method starts by adding one node for each possible fuzzy expression $S_i^j : X_i$ is LX_i^j, obtaining a graph with p_i nodes for each feature f_i. Besides, in order to grasp the individual potential of each fuzzy expression, a fictitious node is also added to the graph that represents the void expression. Then, a weight is assigned to the arc from node A to node B that measures the usefulness of adding the expression $S_B = S_{i_B}^{j_B}$ to the expression $S_A = S_{i_A}^{j_A}$, defined as

$$w_{AB} = U(S_A) - U(S_A, S_B), \tag{1}$$

where U is a measure of uncertainty. Concretely, we propose

$$U(S_A) = \rho_A \cdot H(S_A) + \rho_{\neg A} \cdot H(S_{\neg A}) \tag{2}$$

$$\begin{aligned} U(S_A, S_B) = \rho_{A\wedge B} \cdot H(S_{A\wedge B}) + \rho_{A\wedge\neg B} \cdot H(S_{A\wedge\neg B}) + \cdots \\ \cdots + \rho_{\neg A\wedge B} \cdot H(S_{\neg A\wedge B}) + \rho_{\neg A\wedge\neg B} \cdot H(S_{\neg A\wedge\neg B}), \end{aligned} \tag{3}$$

where ρ_A measures the ratio of examples satisfying S_A, and $H(S_i^j)$ is the entropy of S_i^j, whose associated distribution of probability is

$$p(c_k) = \frac{\sum_{e_i\in\mathcal{E}_k} \mu_{LX_i^j}(x_i^j)}{\sum_{e_i\in\mathcal{E}} \mu_{LX_i^j}(x_i^j)}, \tag{4}$$

being $\mathcal{E}_k$ the subset of examples within c_k. For $H(S_{i_A}^{j_A} \wedge S_{i_B}^{j_B})$ the expression $\mu_{LX_i^j}(x_i^j)$ in (4) is replaced by $\mu_{LX_{i_A}^{j_A}}(x_{i_A}^{j_A}) \wedge \mu_{LX_{i_B}^{j_B}}(x_{i_A}^{j_A})$, being $\wedge$ a t-norm.

Since PageRank does not allow negative weights, some mechanism must be introduced to satisfy this requirement in (1). The proposal is to remove arcs with negative weights, since traversing them is worthless.

Finally, after the PageRank algorithm is applied, the value of each node gives a measure of the information provided by the fuzzy expression it represents in the light of the dataset.

Acknowledgments. Work co-funded by the IB16048 proyect (Junta Extremadura).

References

1. Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H.: Feature selection. ACM Computing Surveys **50**(6) (dec 2017) 1–45
2. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab (November 1999) Previous number = SIDL-WP-1999-0120.
3. Mihalcea, R., Tarau, P.: TextRank: Bringing order into texts. In: Proceedings of EMNLP-04and the 2004 Conference on Empirical Methods in Natural Language Processing. (July 2004)
4. Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., Zadeh, R.: Wtf: the who to follow service at twitter. In: Proceedings of the 22nd international conference on World Wide Web - WWW'13, ACM Press (2013)