World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:9, No:06, 2015

Classification of Political Affiliations by Reduced Number of Features

Authors : Vesile Evrim, Aliyu Awwal

Abstract : By the evolvement in technology, the way of expressing opinions switched the direction to the digital world. The domain of politics as one of the hottest topics of opinion mining research merged together with the behavior analysis for affiliation determination in text which constitutes the subject of this paper. This study aims to classify the text in news/blogs either as Republican or Democrat with the minimum number of features. As an initial set, 68 features which 64 are constituted by Linguistic Inquiry and Word Count (LIWC) features are tested against 14 benchmark classification algorithms. In the later experiments, the dimensions of the feature vector reduced based on the 7 feature selection algorithms. The results show that Decision Tree, Rule Induction and M5 Rule classifiers when used with SVM and IGR feature selection algorithms performed the best up to 82.5% accuracy on a given dataset. Further tests on a single feature and the linguistic based feature sets showed the similar results. The feature “function” as an aggregate feature of the linguistic category, is obtained as the most differentiating feature among the 68 features with 81% accuracy by itself in classifying articles either as Republican or Democrat.

Keywords : feature selection, LIWC, machine learning, politics
Conference Title : ICMLA 2015 : International Conference on Machine Learning and Applications
Conference Location : Copenhagen, Denmark
Conference Dates : June 11-12, 2015

Open Science Index, Computer and Systems Engineering Vol:9, No:06, 2015 publications.waset.org/abstracts/27684/pdf