Abstract

This paper presents High Probability Minimum Redundancy, HPMR, as a new algorithm for employing the most predictive features to contribute dimensionality reduction. The proposed algorithm is useful for finding new, optimal, and more informative features maintaining acceptable classification accuracy. A problem encountered in many large-scale information applications relevant to expert and intelligent systems such as pattern recognition, bioinformatics, social media content classification where data sets containing massive numbers of features. Implementing categorization on these huge, uneven, useless datasets with the overwhelming number of features is just a waste of time degrading the efficiency of classification algorithms and hence the results are not much accurate. HPMR controls the tradeoff between relevance and redundancy by selecting new feature subset that retains sufficient information to discriminate well among classes.

To emphasize the significance of HPMR, it has been relied upon to develop an intelligent system for Arabic sentiment analysis on social media. Additionally, the performance of such
algorithm is quantitatively compared with other traditional dimensionality reduction techniques in terms of performance accuracy, dataset reduction percentage, training time. Experimental results prove that HPMR cannot only diminish the feature vector but also can significantly enhance the performance of the well-known classifiers.

References


17. Jiguang Liang; Xiaofei Zhou; Li Guo1; Shuo Bai; “Feature Selection for Sentiment Classification Using Matrix Factorization”, pp. 63 – 64. Florence, Italy, ACM. May 2015,
30. Satoshi Niijima , Yasushi Okuno “Laplacian Linear Discriminant Analysis Approach to


Index Terms

Computer Science Information Sciences

Keywords

Feature Selection, HPMR, Chi-squared, Social Media, Arabic, SA, SVM.