

Classifying Blog Texts Based on the Psycholinguistic Features of the Texts

Authors : Hyung Jun Ahn

Abstract : With the growing importance of social media, it is imperative to analyze it to understand the users. Users share useful information and their experience through social media, where much of what is shared is in the form of texts. This study focused on blogs and aimed to test whether the psycho-linguistic characteristics of blog texts vary with the subject or the type of experience of the texts. For this goal, blog texts about four different types of experience, Go, skiing, reading, and musical were collected through the search API of the Tistory blog service. The analysis of the texts showed that various psycholinguistic characteristics of the texts are different across the four categories of the texts. Moreover, the machine learning experiment using the characteristics for automatic text classification showed significant performance. Specifically, the ensemble method, based on functional tree and bagging appeared to be most effective in classification.

Keywords : blog, social media, text analysis, psycholinguistics

Conference Title : ICAIA 2017 : International Conference on Artificial Intelligence and Applications

Conference Location : Singapore, Singapore

Conference Dates : January 08-09, 2017