Predicting Social Exclusion: A Computational Linguistic Approach to the Detection of Ostracism

Greta Gandolfi
University of Trento, Trento, Italy

Carlo Strapparava
FBK-Irst, Trento, Italy

Abstract

Ostracism is a social phenomenon, shared by most social animals, including humans. Its detection plays a crucial role for the individual, with possible evolutionary consequences for the species.

Considering (1) its relation with communication and therefore language and (2) its social nature, we hypothesised that the combination of linguistic and community-level social features would have a positive impact on the automatic recognition of ostracism in human online communities.

We modelled a linguistic community through Reddit data and we analysed the performance of simple classification algorithms (Nave Bayes and SVM), particularly focusing on the feature selection. Comparing the accuracy scores of the algorithms fed with a) linguistic features, b) extralinguistic features, and c) linguistic + extralinguistic features, we tested our hypothesis, showing how models based on c) generally outperform.

To our knowledge, this is the first attempt to automatise the identification of such a complex phenomenon through NLP techniques.