

Joint Rumour Stance and Veracity

Anders Edelbo Lillie*

ITU Copenhagen
Denmark
aedl@itu.dk

Emil Refsgaard Middelboe*

ITU Copenhagen
Denmark
erem@itu.dk

Leon Derczynski

ITU Copenhagen
Denmark
ld@itu.dk

Abstract

The net is rife with rumours that spread through microblogs and social media. Not all these claims can be verified. However, recent work has shown that the stances crowds take toward claims alone can be sufficiently good indicators of claim veracity, with for example an HMM using conversational stance sequences as the only input. Existing results are monolingual (English) and mono-platform (Twitter). This paper introduces a stance-annotated Reddit dataset for the Danish language, and describes various implementations of stance classification models. Out of these, a Linear Support Vector Machine provides the best results with an accuracy of 0.76 and macro F_1 score of 0.42. Furthermore, experiments show that stance labels can be used to predict veracity across platforms and also across languages, training on conversations held in one language and using the model on conversations held in another. Using HMM as the sequence model, monolingual scores reach a stance-based veracity accuracy of 0.83 and F_1 of 0.68; applying the model across languages predicts veracity of claims with an accuracy of 0.82 and F_1 of 0.67. Finally, when using automatic stance labels (as opposed to gold) as input to the HMM, only a small drop in performance is observed, showing that the end-to-end system is viable for stance-based veracity assessment.

*: These authors contributed to the paper equally.