Sarcasm Detection on Twitter:

bolstering lexical features with contextual clues

Samuel Albanie

M.Sc. Computer Science, Networks & Distributed Systems

University of Dublin, Trinity College, 2013

Supervisor: Alexander O'Connor

One of the consequences of the unabated growth of social media has been a surge in the production of opinionated user-generated content. This data represents an immensely valuable resource if meaningful insights and trends can be accurately extracted. One of the challenges facing *sentiment analysis*, the field of research that addresses this task, is the accurate detection of sarcasm, which flips the polarity of literally-interpreted sentiments. While previous research in this area has focused on the use of *lexical feature*-based models to classify sarcasm, this research investigates the impact of using *contextual features* to improve accuracy.

A thorough review is conducted into the literature and theory relating to sarcasm detection and a corpus is generated using the *Twitter API* with a series of gathering and filtering techniques. A study is then conducted into the ability of humans to detect sarcasm and classifiers are trained to recognise sarcasm with a range of lexical and contextual feature sets. Finally, an extensive evaluation of each set of classifiers demonstrates the value of considering the contextual feature-based approach.