



Working Group on Statistical Learning Seminar

Title: Analysis of Sentiment in Twitter

Speaker: Michael Salter-Townshend (UCD)

Date: Fri 25th February 2011 at 1:00PM

Location: Statistics Seminar Room- L550 Library building

Abstract: This is early work on the analysis of sentiment in Twitter data. We have a large volume (millions of tweets) from twitter.com and the goal is to analyse the sentiment contained in the tweets. We have also manually labelled a small subset of commonly occurring words and smileys as relating to either positive or negative sentiment.

We model the number of positive words in a tweet as Binomially distributed with mean equal to the number of words in the tweet times the tweet sentiment. We employ an EM algorithm where the E step calculates the expected sentiment label for the unlabelled words appearing in the tweets and the M step

finds maximum-likelihood estimates for the sentiment in each tweet.

We present preliminary results and would like to discuss challenges, choices and extensions to the work.