



Using Sentiment Analysis to Observe How Science is Communicated

David Topping (1) and Sam Illingworth (2)

(1) University of Manchester, Centre for Atmospheric Science, Manchester, United Kingdom (david.topping@manchester.ac.uk), (2) Manchester Metropolitan University, School of Research, Enterprise & Innovation, United Kingdom (s.illingworth@mmu.ac.uk)

‘Citizen Science’ and ‘Big data’ are terms that are currently ubiquitous in the field of science communication. Whilst opinions differ as to what exactly constitutes a ‘citizen’, and how much information is needed in order for a data set to be considered truly ‘big’, what is apparent is that both of these fields have the potential to help revolutionise not just the way that science is communicated, but also the way that it is conducted. However, both the generation of sufficient data, and the efficiency of then analysing the data once it has been analysed need to be taken into account.

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. The process of sentiment analysis can be automated, providing that an adequate training set has been used, and that the nuances that are associated with a particular topic have been accounted for. Given the large amounts of data that are generated by social media posts, and the often-opinionated nature of these posts, they present an ideal source of data to both train with and then scrutinize using sentiment analysis.

In this work we will demonstrate how sentiment analysis can be used to examine a large number of Twitter posts, and how a training set can be established to ensure consistency and accuracy in the automation. Following an explanation of the process, we will demonstrate how automated sentiment analysis can be used to categorise opinions in relation to a large-scale science festival, and will discuss if sentiment analysis can be used to tell us if there is a bias in these communications. We will also investigate if sentiment analysis can be used to replace more traditional, and invasive evaluation strategies, and how this approach can then be adopted to investigate other topics, both within scientific communication and in the wider scientific context.