Weekly Progress Report

Detecting Conversations Going Viral

Matt Chapman

Week 14, 03/04 - 07/04

1 Dataset Selection

Selecting the correct dataset to carry out my analyses is one of the most important parts of the work I will carry out in this project. To this end, I will explain the data set that I will be working with, as well as my rationale for selecting it.

I have chosen to work exclusively with Dutch language Twitter data for the following reasons:

- The vast majority of Buzzcapture clients are Dutch companies, so it is more likely that I will be able to find relevant data by restricting gathered conversations by language.
- Twitter is *the* platform for breaking news and discussions thereof. It is likely that changes in volume that the likes of which I am looking for, will happen on Twitter before they happen on other platforms such as Facebook, Pinterest or Facebook.
- Conversation volume is much higher on Twitter than on Facebook or other competing services. As such spikes in conversation volume are likely to be higher and more easily annotated by hand for both testing and demonstrating the effectiveness of a given algorithm.

2 Additional Research

In addition to making some decisions regarding the experimental dataset, I have also carried out some additional research into the field. The following is a short annotated bibliography of the additional sources found:

[3] **Anita M Pelecanos, Peter a Ryan, and Michelle L Gatton. "Outbreak detection algorithms for seasonal disease data: a case study using Ross River virus disease." In: *BMC medical informatics and decision making* 10.1 (2010), p. 74. ISSN: 1472-6947. DOI: 10.1186/1472-6947-10-74. URL: http://www.biomedcentral.com/1472-6947/10/74**

[2] **Martin Kulldorff et al. "A space-time permutation scan statistic for disease outbreak detection". In: *PLoS Medicine* 2.3 (2005), pp. 0216–0224. ISSN: 15491277. DOI: 10.1371/journal.pmed.0020059**

[1] **Jeremy Ginsberg et al. "Detecting influenza epidemics using search engine query data." In: *Nature* 457.7232 (2009), pp. 1012–4. ISSN: 1476-4687. DOI: 10.1038/nature07634. URL: http://www.ncbi.nlm.nih.gov/pubmed/19020500**

References

[1] Jeremy Ginsberg et al. "Detecting influenza epidemics using search engine query data." In: *Nature* 457.7232 (2009), pp. 1012–4. ISSN: 1476-4687. DOI: 10.1038/nature07634. URL: http://www.ncbi.nlm.nih.gov/pubmed/19020500.

[2] Martin Kulldorff et al. "A space-time permutation scan statistic for disease outbreak detection". In: *PLoS Medicine* 2.3 (2005), pp. 0216–0224. ISSN: 15491277. DOI: 10.1371/journal.pmed.0020059.

[3] Anita M Pelecanos, Peter a Ryan, and Michelle L Gatton. "Outbreak detection algorithms for seasonal disease data: a case study using Ross River virus disease." In: *BMC medical informatics and decision making* 10.1 (2010), p. 74. ISSN: 1472-6947. DOI: 10.1186/1472-6947-10-74. URL: http://www.biomedcentral.com/1472-6947/10/74.