

Wisdom of the Local Crowd: Detecting Local Events Using Social Media Data

Søren B. Ranneries¹, Mads E. Kalør², Sofie Aa. Nielsen³, Lukas N. Dalgaard⁴,
Lasse D. Christensen⁵ and Nattiya Kanhabua⁶

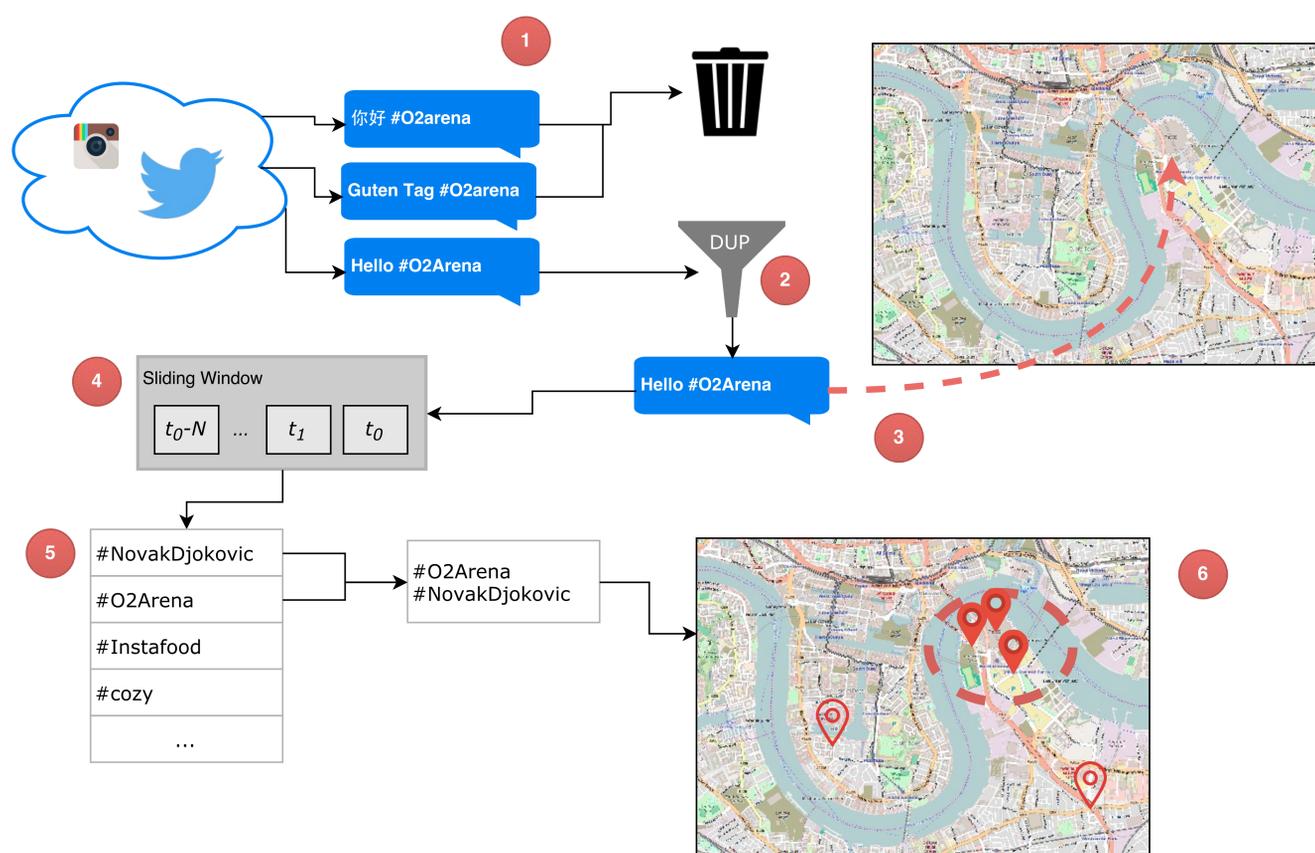
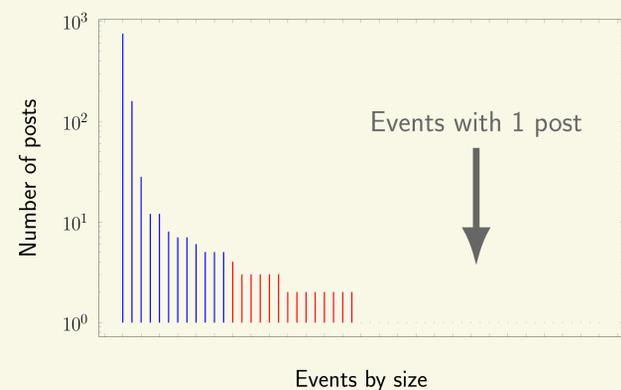
Department of Computer Science, Aalborg University, Denmark

Motivation

- ▶ People share their experiences during events such as concerts and football games on social media.
- ▶ Can we use this data to detect small-scale local events?
- ▶ How does Twitter and Instagram compare regarding the amount of information about events?

Experimental Setup

- ▶ Manually labeled posts
- ▶ London, November 17, 2015, 20:00 to 21:00
- ▶ 9131 posts in total, 7096 from Twitter and 2035 from Instagram
- ▶ 2778 posts with precise location information: 743 from Twitter, and 2035 from Instagram
- ▶ 53 events (10 with at least 5 posts)



Method

The local event detection method is illustrated as a pipeline above with the following stages:

1. Non-English posts are removed to simplify the natural language processing later in the pipeline. We do that with a naive Bayes classifier.
2. A locality-sensitive hashing based near-duplication filter removes cross-posted posts and bulk-posted spam.
3. Location extraction is performed on text from posts without an explicit location based on Wikipedia locations.
4. Posts are inserted into a time-based sliding window. When the contents of the sliding window has changed by some ratio, a snapshot of the posts is sent to the next stage.
5. Posts are grouped in sets of (uncommon) hashtags. Two sets are merged if they share a significant number of posts until no two sets share many posts.
6. A location-based clustering is performed within each hashtag cluster in case multiple events about the same topic are happening at different locations. Outliers are considered to be noise and removed.

Results

- ▶ All data
 - Precision: 0.20
 - Recall: 0.60
 - F_1 : 0.30
- ▶ Twitter only
 - Precision: 0.17
 - Recall: 0.50
 - F_1 : 0.25
- ▶ Instagram only
 - Precision: 0.24
 - Recall: 0.84
 - F_1 : 0.37
- ▶ Precise locations, one event off by 200 meters.

Conclusion

- ▶ Instagram more useful for event detection than Twitter
- ▶ Precision does not outperform contemporary approaches [1, 2, 3]
- ▶ Shorter window than other approaches: More likely ongoing events
- ▶ Precision may be improved with event classifier

References

- [1] R. Lee and K. Sumiya, "Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection," in *SIGSPATIAL*. ACM, 2010.
- [2] M. Walther and M. Kaiser, "Geo-spatial event detection in the twitter stream," in *ECIR*, 2013.
- [3] K. Zickuhr, "Location-based services," <http://www.pewinternet.org/2013/09/12/location-based-services>, Pew Research Center, September 2013, accessed September 18 2015.