



#MSM2013 Workshop Chairs' Welcome

It is our great pleasure to welcome you to the *Making Sense of Microposts 2013 - #MSM2013* workshop. We started the #MSM adventure 3 years ago with the modest goal of exploring the impacts of the novel way people interact with the Web by using low-effort publishing tools (such as tweets, likes, check-ins). Today, we are proud to say that #MSM has become a premier venue for communicating research results and shaping the research agenda of our emerging field of study.

Throughout the past #MSM events we have learned about the importance of some practical issues that Micropost researchers experience, especially those related to information extraction. As a response to the growing academic and industrial interest in these questions, this year we have organised an Information Extraction challenge as a part of the Workshop. The challenge gained a lot of interest in the community and attracted 21 submissions. Six submissions, selected for their innovative nature, were given the opportunity for a live presentation, and another 8 were given space in the proceedings.

The main conference track attracted 13 submissions from North and South America, Asia, and Europe. The program committee accepted 6 papers (4 long and 2 short) that cover a variety of topics, including Machine Learning and Statistical Analysis of Microposts, Trend and Event detection in Microposts and Using Microposts for Prediction and Intent Detection.

A modern perspective and issues specific to Microposts treatment are the focus of papers in the first group dedicated to Machine Learning. The growth in Micropost publication requires approaches that enable effective categorisation and retrieval. A core component of such processes is the use of suitable labels for Microposts that describe their inherent content and topic. In the short paper 'Using Topic Models for Twitter Hashtag Recommendation' by Godin et al. a two-stage approach is presented for hashtag recommendation to enable Micropost categorisation. The two-stage approach begins by first identifying English-language posts using a naive Bayes classifier trained using Expectation-Maximization, achieving an accuracy level of 97.4%; before second, using a Latent Dirichlet Allocation model to induce topic models for each post. This latter stage provides a collection of high probability terms that capture the implicit topic of a given post and can therefore be used as hashtags.

Another important issue is the noisy nature of Microposts that, due to their informal and cryptic language style, often hinders existing named entity recognition approaches that function over well-formed language structures. Additionally, the scale of Micropost publication requires methods that can process content quickly and extract entities found therein. The paper 'FS-NER: A Lightweight Filter-Stream Approach to Named Entity Recognition on Twitter Data' by Marinho de Oliveira et al. tackles this problem through a filter-based approach. Individual filters are constructed by inducing discrete probability distributions from various observed features of Microposts (terms, context indicators, dictionary lookups). The work's assumption is that individual filters are weak in isolation, however when combined together they allow entities to be detected in Microposts. Marinho de Oliveira et al. demonstrate the feasibility of this approach by comparing their filter-ensemble approach with an approach based on Conditional Random Fields and demonstrate the superior performance and faster execution time of the filter-based approach.

Due to their abundance, Microposts are often seen as a way to observe and react to tendencies and real-world events. Two papers explore this important topic. The paper ‘Nerding Out on Twitter: Fun, Patriotism and #Curiosity’ focuses on a real-world event – the landing of NASA Mars rover, Curiosity, in observing the dynamics of tweets, sentiment and most common message types exchanged on Twitter in relation to this event. This paper takes a novel perspective of using advanced information retrieval and natural language processing techniques to provide answers to social science questions. It demonstrates the importance of a multidisciplinary perspective in treating and making sense of Microposts.

In addition to analysing Tweets known to be related to a particular event; a major topic on the research agenda of the Microposts research community is the automatic detection of events and classification of tweets by events they refer to. The paper ‘ET: Events from Tweets’ presents a general vision for automatic event detection and explores in detail the detection of bigrams that, by their dynamics and frequency, can be considered as indicative of an event taking place. This particular approach is an important piece of the automatic event detection puzzle that is beginning to take shape in response to the growing need for such approaches.

Finally, the topic of using Microposts for prediction tasks and for detecting user intentions emerged in two papers. The paper ‘Meaning as Collective Use: Predicting Semantic Hashtag Categories on Twitter’ by Posch et al. explores the pragmatics of hashtag usage to identify whether such information could be used for the semantic categorisation of Microposts. The authors model several pragmatic features that capture properties of hashtag streams (i.e. the social network dynamics of users who mention the same hashtag in their posts) and examine how these features differ between semantic categories of tweets (e.g. politics, technology, games). The authors find statistically significant differences between categories based on hashtag dynamics, and are able to use this information to classify Microposts into their semantic categories accordingly.

Increasingly, market actors are looking for ways to leverage Twitter for economic purposes, to solicit items and to advertise items for sale. Hollerit et al., in ‘Towards Linking Buyers and Sellers: Detecting Commercial Intent on Twitter’, discuss the use of Twitter posts to advertise intent to buy and sell. The approach feeds a set of (pre-selected) keywords into a classification model in order to detect tweets that indicate implicit and explicit commercial intent. The authors report precision of 57% and recall 77%, and conclude that an additional benefit of their work is the provision of an alternative search method for Twitter feeds.

Putting together #MSM2013 was a team effort. We first thank the authors for providing the content of the programme. We are grateful to the programme committee, who worked very hard in rigorously reviewing papers and providing highly valuable feedback for authors. We also wish to thank Bernardo Pereira Nunes, our local chair, who helped #MSM attract relevant submissions and audience in Brazil, thus allowing our community to spread to new spaces. We are especially grateful to Amparo Elizabeth Cano Basave for chairing the Information Extraction challenge, which was a very difficult task involving great responsibility. Finally, we wish to thank eBay for sponsoring the challenge award and for pushing us forward by their continuous interest in #MSM research topics.

We hope that you will enjoy this programme and that you will join us in future #MSM adventures, determined to search for novel ways of treating, understanding, publishing and making use of Microposts.

Matthew Rowe

University of Lancaster, UK

Milan Stankovic

Université Paris-Sorbonne, France

Aba-Sah Dadzie

University of Sheffield, UK