

# Opinion Mining: Exploiting the Sentiment of the Crowd

Diana Maynard and Adam Funk

University of Sheffield, 211 Portobello, Sheffield, S1 4DP, United Kingdom  
email: [diana,a.funk]@dcs.shef.ac.uk  
phone: +44 114 222 1936  
fax: +44 114 222 1810

## 1 Overview

The proposed tutorial will introduce the concepts of opinion mining and sentiment analysis from unstructured text, looking at why it is useful and what tools and techniques are available. We will describe a variety of general rule-based and machine learning techniques, provide some background information on the key underlying NLP processes required, and focus specifically on some of the major problems and solutions, such as detection of sarcasm, use of informal language, spam opinion detection, trustworthiness of opinion holders, and so on. We will also discuss problems associated with opinion detection in social media such as blogs, forum posts, twitter, etc. The techniques will be demonstrated with key open-source tools and applications.

## 2 Motivation

Web 2.0 nowadays provides a great means to share knowledge, including opinions which may be useful to all kinds of people: for example product reviews, public reactions to political events, scientific blogs, fashion and music trends and so on. This information is useful not only to other consumers but to archivists, companies looking for feedback on both their own products and those of their rivals, governments and social historians, amongst others. It is not straightforward, however, to access these opinions - first, because information from individual users is not sufficient, but needs to be aggregated; second, because understanding what is actually meant is not always easy due to the vagaries of natural language; and third, because not all information is useful or indeed, trustworthy. Current off-the-shelf tools for opinion mining, while numerous, tend to be quite limited, often offering little more than a keyword-based lookup for opinionated words, or a bag-of-words-based classification approach. Effective opinion mining is still very much a hot research topic rather than a solved problem.

This tutorial will address these needs by introducing and demonstrating techniques for extracting the relevant information from unstructured text, and in particular from social media, so that participants will have a better understanding of both the problems and solutions in opinion mining, and will be equipped with the necessary building blocks of knowledge to build their own tools and tackle complex issues. The tutorial will cover state-of-the-art research as well as established methods and tools for important subtasks. Since all of the NLP tools to be presented are open source, the tutorial will provide the attendees with skills which are easy to apply and do not require special software or licenses.

### 3 Outline of the tutorial

The tutorial will be divided into 3 sections, as follows:

**Introduction to opinion mining** The first part of the tutorial will explain the motivation for such tools, introduce the main linguistic subcomponents needed for an application, and outline the major difficulties. It will look at existing opinion mining tools and discuss their strengths and weaknesses.

**Linguistic challenges** This section will explore further the concept of opinion mining and describe ways in which NLP techniques can be used. We will introduce and discuss some of the major linguistic challenges for an opinion mining system, such as negatives, conditional statements, incorrect English, slang/swear words, irony/sarcasm and so on.

**Applications** Finally we will illustrate the techniques introduced with examples of real research applications. These will cover both rule-based and machine-learning approaches to opinion mining, in the domains of product reviews, political opinions, and discussions of social events. Furthermore, we will discuss (with real examples) how these techniques can be adapted for different languages and for texts in multiple languages.

### 4 Estimated audience

The tutorial will have an expected audience of 30-50 people.

### 5 Audience profile

The **target audience** will consist of researchers from any background looking to extend their work to opinion mining applications. No previous knowledge of opinion mining or machine learning is necessary, but a basic knowledge of natural language processing techniques is useful.

### 6 Duration

half a day

### 7 Technical requirements

Internet connection

## 8 Tutorial speakers

**Speaker 1 (Main Contact): Dr Diana Maynard (University of Sheffield)** is a Research Fellow in the Natural Language Processing Group at the University of Sheffield. She holds a PhD in Automatic Term Recognition from Manchester Metropolitan University (UK) and has almost 20 years of experience in the field. Her main interests are in Information Extraction, opinion mining, text mining, semantic web technology and robust and adaptable tools for language engineering. Over the past 12 years she has led the development of Sheffield's open-source multilingual text mining tools, and has led research teams on a number of national and international projects including the EU NoE KnowledgeWeb and the EU projects NeOn, Musing and Arcomem. She has published over 50 scientific papers in conferences, journals and books, has reviewed numerous conference and journal papers, organised several workshops and is chair of the Semantic Web Challenge at ISWC, one of the most prestigious annual Semantic Web events. She is responsible for managing training and consultancy with GATE, has taught a text mining course at the 2011 International Summer School in Language and Speech Technologies, given keynote speeches, invited talks, tutorials, lectures and courses on a number of NLP topics at international NLP and Semantic Web conferences.

**Speaker 2: Dr Adam Funk** is a Research Associate in the Natural Language Processing Group at the University of Sheffield, where he has worked for five years. He holds a PhD in Computational Linguistics from the University of Manchester, and has worked in computational linguistics and natural language processing for 10 years. He has published scientific papers and reviewed conference and journal papers, and regularly teaches at GATE courses and tutorials. He has worked on the implementation of linguistic annotation standards and the development of NLP applications for information extraction, opinion mining and the semantic web, particularly in national and international research projects. **Website:** <http://www.dcs.shef.ac.uk/~adam>