Applications of Social Media Text Analysis

Atefeh Farzindar
NLP Technologies Inc.
Université de Montréal
farzindar@nlptechnologies.ca

Diana Inkpen
University of Ottawa
diana.inkpen@uottawa.ca

Abstract

This tutorial reviews the current research on NLP tools and methods for processing the non-traditional information such as social media content. We will focus on innovative NLP tasks (opinion mining and emotion analysis, geo-location detection, event and topic detection, entity linking and disambiguation, summarization and machine translation for social media) that can be used in order to integrate appropriate linguistic information in various fields such as social media monitoring for health care, security and defence, business intelligence and politics.

1 Description

Analyzing social media texts is a complex problem that becomes difficult to address using traditional Natural Language Processing (NLP) methods. Our tutorial focuses on presenting new methods for NLP tasks and applications that work on noisy and informal texts, such as the ones from social media.

Automatic processing of large collections of social media texts is important because they contain a lot of useful information, due to the increasing popularity of all types of social media. Use of social media and messaging apps grew 203 percent year-on-year in 2013, with overall app use rising 115 percent over the same period, as reported by Statista[1], citing data from Flurry Analytics[2]. This growth means that 1.61 billion people are now active in social media around the world and this is expected to advance to 2 billion users in 2016, led by India. The research shows that consumers are now spending daily 5.6 hours on digital media including social media and mobile internet usage[3].

At the heart of this interest is the ability for users to create and share content via a variety of platforms such as blogs, micro-blogs, collaborative wikis, multimedia sharing sites, social networking sites. The unprecedented volume and variety of user-generated content, as well as the user interaction network constitute new opportunities for understanding social behavior and building socially intelligent systems. Therefore it is important to investigate methods for knowledge extraction from social media data. Furthermore, we can use this information to detect and retrieve more related content about events, such as photos and video clips that have caption texts.

2 Content Overview

NLP for social media text is a new research area, and it requires adapting the traditional NLP methods to this kind of texts or developing new methods suitable for information extraction, categorization and clustering, automatic summarization, semantic search engine, and statistical machine translation methods.

NLP for Social Media could be used in various contexts such as detecting criminal/terrorist activities, monitoring healthcare (e.g., monitoring for signs of Ebola infection), predicting public behavior from social media, and increasing public safety. Likewise, business informatics and stock market analysis can benefit from the semantic analysis of social media. Several industries are interested in this topic, such as the entertainment industry, news agencies, marketing and advertising industry, security and defense.

[1] http://www.statista.com/chart/1778/app-use-in-2013/
[2] http://www.flurry.com/solutions/analytics
[3] http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Our tutorial focuses on various applications of social media text analysis and the need for showcasing the new research studies and techniques for Natural Language Processing of these types of data. We present text analysis techniques for social media data and applications using a variety of computational approaches that are able to process large amounts of noisy data.

We emphasize the importance of this dynamic discipline and the great potential of NLP in the coming decade, in the context of changes in mobile technology, wearable technology, cloud computing and social networking. The content of the tutorial will be based on our upcoming book (Farzindar and Inkpen, 2015).

This tutorial is the follow-up of the workshops organized by Dr. Farzindar and Dr. Inkpen in collaboration with ACL, NAACL-HLT and EACL (Semantic Analysis in Social Media SASM 21012 held in Avignon, France with EACL 2012, Language Analysis in Social LASM 2013 held in Atlanta, GA, with NAACL/HLT 2014, and Language Analysis in Social Media LASM 2014 held in Gothenburg, Sweden with EACL 2014).

3 Outline

1. Introduction
   - 1.1. Challenges in social media data
   - 1.2. Semantic analysis of social media
   - 1.3. NLP tools for Social Media Texts

2. Semantic Analysis of Social Media Texts
   - 2.1. Geo-location detection
   - 2.2. Opinion mining and emotion analysis
   - 2.3. Event and topic detection
   - 2.4. Entity linking and disambiguation
   - 2.5. Summarization in social media data
   - 2.6. Machine translation in social media

3. Applications
   - 3.1. Health care applications
   - 3.2. Financial applications
   - 3.3. Predicting voting intentions
   - 3.4. Media monitoring
   - 3.5. Security and defence applications
   - 3.6. NLP-based information visualisation
   - 3.7. Disaster response applications
   - 3.8. NLP-based user modelling
   - 3.9. Applications for entertainment

4. TRANSLI™ a case study of an NLP-based social media monitoring system

4 About the Presenters

Dr. Atefeh Farzindar is the CEO of NLP Technologies Inc. and Adjunct Professor at University of Montreal. She has served as Chair of the technology sector of the Language Industry Association Canada (AILIA) (2009-2013), vice president of The Language Technologies Research Centre (LTRC) of Canada (2012-2014) and a member of the Natural Sciences and Engineering Research Council of Canada (NSERC) Computer Science Liaison Committee (2014-2015). Recently, she authored a book chapter in Social Network Integration in Document Summarization, Innovative Document Summarization Techniques: Revolutionizing Knowledge Understanding, IGI Global publisher January 2014.

Dr. Diana Inkpen is a Professor the School of Electrical Engineering and Computer Science at the University of Ottawa. Her research interests and expertise are in natural language processing, in particular lexical semantics as applied to near synonyms and nuances of meaning, word and text similarity, classification of texts by emotion and mood, information retrieval from spontaneous speech, extraction of semantic frames, and lexical choice in natural language generation. She published more than 25 journal papers, 85 conference papers, and 6 book chapters. She is an associated editor of the Computational Intelligence and Natural Language Engineering journals.

References

Atefeh Farzindar and Diana Inkpen, 2015. Natural Language Processing for Social Media. Synthesis Lectures on Human Language Technologies. Morgan and Claypool Publishers (to appear).

Atefeh Farzindar. 2014. Social Network Integration in Document Summarization, In Innovative Document Summarization Techniques: Revolutionizing Knowledge Understanding: IGI-Global, 139-162.

Atefeh Farzindar and Wael Khreich, 2013. A Survey of Techniques for Event Detection in Twitter. In Computational Intelligence: Wiley, 31(1): 134-164.