Chapter 12

TEXT ANALYTICS IN SOCIAL MEDIA

Xia Hu

Computer Science and Engineering
Arizona State University

xiahu@asu.edu

Huan Liu

Computer Science and Engineering
Arizona State University

huanliu@asu.edu

Abstract   The rapid growth of online social media in the form of collaboratively-created content presents new opportunities and challenges to both producers and consumers of information. With the large amount of data produced by various social media services, text analytics provides an effective way to meet usres' diverse information needs. In this chapter, we first introduce the background of traditional text analytics and the distinct aspects of textual data in social media. We next discuss the research progress of applying text analytics in social media from different perspectives, and show how to improve existing approaches to text representation in social media, using real-world examples.

1.   Introduction

Social media such as blogs, microblogs, discussion forums and multimedia sharing sites are increasingly used for users to communicate breaking news, participate in events, and connect to each other anytime, from anywhere. The social media sites play a very important role in current

web applications, which accounts for 50% of top 10 sites according to statistics from Alexa[1], as shown in Table 12.1. Besides that, the Twitter messages are even archived in the US Library of Congress[2]. These social media provides rich information of human interaction and collective behavior, thus attracting much attention from disciplines including sociology, business, psychology, politics, computer science, economics, and other cultural aspects of societies.

Table 12.1.   Internet Traffic Report by Alexa on March 3rd, 2011

Rank	*Website*	*Rank*	*Website*
1	Google	6	**Blogger**
2	**Facebook**	7	Baidu
3	**Youtube**	8	**Wikipedia**
4	Yahoo!	9	**Twitter**
5	Windows Live	10	QQ.com

We present a definition of Social Media from a social media source, Wikipedia[3], as follows:

> "Social media are media for social interaction, using highly accessible and scalable communication techniques. It is the use of web-based and mobile technologies to turn communication into interactive dialogue."

Moturu [43] defines social media as:

> "Social Media is the use of electronic and Internet tools for the purpose of sharing and discussing information and experiences with other human beings in more efficient ways."

Traditional media such as newspaper, television and radio follow a unidirectional delivery paradigm, from business to consumer. The information is produced from media sources or advertisers and transmitted to media consumers. Different from this traditional way, web 2.0 technologies are more like consumer to consumer services. They allow users to interact and collaborate with each other in a social media dialogue of user-generated content in a virtual community. We categorize the most popular social media web sites into groups, shown in Table 12.2.

---

[1] www.alexa.com
[2] http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/
[3] http://en.wikipedia.org/wiki/Social_media/