Chapter 12

DATA MINING IN SOCIAL MEDIA

Geoffrey Barbier
Arizona State University
gbarbier@asu.edu

Huan Liu
Arizona State University
huan.liu@asu.edu

Abstract The rise of online social media is providing a wealth of social network data. Data mining techniques provide researchers and practitioners the tools needed to analyze large, complex, and frequently changing social media data. This chapter introduces the basics of data mining, reviews social media, discusses how to mine social media data, and highlights some illustrative examples with an emphasis on social networking sites and blogs.

Keywords: data mining, social media, data representation, social computing, social networks, social networking sites, blogs, blogosphere, event maps

1. Introduction

Data mining, as a young field, has been spearheading research and development of methods and algorithms handling huge amounts of data in solving real-world problems. Much like traditional miners extract precious metals from earth and ore, data miners seek to extract meaningful information from a data set that is not readily apparent and not always easily obtainable. With the ubiquitous use of social media via the internet, an unprecedented amount of data is available and of interest to many fields of study including sociology, business, psychology, entertainment, politics, news, and other cultural aspects of societies. Applying data mining to social media can yield interesting perspectives on human behavior and human interaction. Data mining can be used

C. C. Aggarwal (ed.), *Social Network Data Analytics,*
DOI 10.1007/978-1-4419-8462-3_12,

in conjunction with social media to better understand the opinions people have about a subject, identify groups of people amongst the masses of a population, study group changes over time, find influential people, or even recommend a product or activity to an individual.

The elections during 2008 marked an unprecedented use of social media in a United States presidential campaign. Social media sites including YouTube[1] and Facebook[2] played a significant role in raising funds and getting candidates' messages to voters [51]. Researchers at the Massachusetts Institute of Technology, Center for Collective Intelligence, mined blog data to show correlations between the amount of social media used by candidates and the winner of the 2008 presidential campaign [24]. This powerful example underscores the potential for data mining social media data to predict outcomes at a national level. Data mining social media can also yield personal and corporate benefits. In another example, researchers developed a Group Recommendation System (GRS) for Facebook users using hierarchical clustering and decision tree data mining methods [7]. The GRS matches users, based on their Facebook profiles, with Facebook groups the users are likely to join by applying data mining methods to Facebook groups and their members.

Applying data mining techniques to social media data has gained increasing attention with the significant rise of online social media in recent years. Social media data have three characteristics that pose challenges for researchers: the data are *large*, *noisy*, and *dynamic*. In order to overcome these challenges, data mining techniques are used by researchers to reveal insights into social media data that would not be possible otherwise. This chapter introduces the basics of data mining, reviews social media, discusses how to mine social media data, and highlights some illustrative examples, paving the way for addressing research issues and exploring novel data mining applications.

2. Data Mining in a Nutshell

One definition of data mining is identifying novel and actionable patterns in data. Data mining is also known as Knowledge Discovery from Data (KDD) [28] or Knowledge Discovery in Databases, also abbreviated as KDD [47]. Data mining is related to machine learning, information retrieval, statistics, databases, and even data visualization [43]. One formal definition for data mining is found in Princeton University's WordNet[3] where data mining is defined as:

> "data processing using sophisticated data search capabilities and statistical algorithms to discover patterns and correlations in large preexisting databases; a way to discover new meaning in data"

[1] http://www.youtube.com/
[2] http://www.facebook.com/
[3] http://wordnet.princeton.edu/