

Chapter 63

Knowledge Discovery from Online Communities

Luca Cagliero

Politecnico di Torino, Italy

Alessandro Fiori

Politecnico di Torino, Italy

ABSTRACT

During recent years, the outstanding growth of social network communities has caught the attention of the research community. A huge amount of user-generated content is shared among community users and gives researchers the unique opportunity to thoroughly investigate social community behavior. Many studies have been focused on both developing models to investigate user and collective behavior and building applications tailored to the most common community user activities.

This chapter presents an overview of social network features such as user behavior, social models, and user-generated content to highlight the most notable research trends and application systems built over such appealing models and online media data. It first describes the most popular social networks by analyzing the growth trend, the user behaviors, the evolution of social groups and models, and the most relevant types of data continuously generated and updated by the users. Next, the most recent and valuable applications of data mining techniques to social network models and user-generated content are presented. Discussed works address both social model extractions tailored to semantic knowledge inference and automatic understanding of the user-generated content. Finally, prospects of data mining research on social networks are provided as well.

INTRODUCTION

In the last years, social network communities such as Facebook, Flickr, and Twitter have shown a steady growth. Since their introduction, social net-

work sites (SNSs) have attracted millions of users, many of who have integrated these sites into their daily practices. These sites provide the basis for maintaining social relationships, for finding users with similar interests, for improving the content of web documents as well as sharing multimedia

DOI: 10.4018/978-1-4666-2455-9.ch063

content. While their key technological features are fairly consistent, the cultures that emerge around SNSs are varied. Most sites support the maintenance of pre-existing social networks, but others help strangers connect based on shared interests. Some sites cater to diverse audiences, while others attract people based on common social characteristics. Sites also vary in the extent to which they incorporate new information and communication tools, such as mobile connectivity, blogging, and photo/video-sharing.

Many studies have been devoted to developing models for supporting the investigation of user behavior in online communities. For example, understanding the structure of online social networks is very useful for developing systems tailored to real user interests. Moreover, due to the nature of these Web services, the common user social behavior in online communities significantly differs from typical real-life social interactions. Indeed, online social models are worth being thoroughly investigated.

Recent works have investigated social correlations among user behaviors in social network sites and applied their models to emergent real life cases of study, such as conflict of interest detection and social hierarchy detection in household corporations. The well-founded analysis of social correlation in user actions enables the investigation of user behavior at large. Collective behavior modeling is focused on discovering recurrences in groups of users. A wide range of scientific paper on this topic is presented in this chapter (Backstrom et al., 2006; Tang & Liu, 2009; Tang & Liu, 2010). A general-purpose categorization, first proposed in Backstrom et al. (2006), is adapted to well discriminate between analysis focused on group membership and formation and group model evolution. Other approaches tried to figure out discriminative behaviors in specific communities, compare behavioral recurrences in different social networks as well as apply traditional analytical models to mostly used online communities (Mislove, 2008).

Nevertheless, online communities are the new main resource of human knowledge. For example, the effort of thousand of Wikipedia users has produced the biggest and most complete online encyclopedia. All the information provided by the tags on media content is a powerful and huge resource suitable for knowledge inference. Thus, combining network analysis, social networks models, and information extraction algorithms are a challenging and interesting task focused on developing new systems able to improve the human knowledge. Information provided by online community users can be employed for different purposes. For example, the information stored in Wikipedia pages can be extracted to infer ontologies (Suchanek et al., 2008). Moreover, the social tagging is fundamental to classify videos and photos in order to improve search results (Yin et al., 2009).

The main purpose of this chapter is to investigate how the user-generated content (UGC) and the social network models can be employed to build systems based on social network data. The chapter is organized as follows. The first part of the chapter will provide a brief overview of most representative online communities such as Digg, Flickr, Twitter, and Wikipedia. The goal of this section is to introduce the reader to the most notable features of online communities over which outstanding models and applications have been recently developed. Secondly, we will investigate the user behavior in these web communities and the most relevant social models tailored to common community user activities. To this aim, some techniques to extract social network models from web content, such as emails and blogs, will be presented. Furthermore, the semantic social knowledge representation of user-generated content provided by online media data and social tags will be introduced. Since knowledge representation enables the exploitation of well-founded Semantic Web technologies in social network analysis, main efforts in semantic knowledge retrieval from online communities will be outlined as well. The last sec-

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/knowledge-discovery-online-communities/73493

Related Content

Spatio-Temporal OLAP Queries Similarity Measure and Algorithm

Olfa Layouni and Jalel Akaichi (2019). *International Journal of Data Warehousing and Mining* (pp. 22-41).
www.irma-international.org/article/spatio-temporal-olap-queries-similarity-measure-and-algorithm/225805

A Route Recommender System Based on Current and Historical Crowdsourcing

Marlene Goncalves, Patrick Rengifo, Daniela Andreina Rodríguez and Ivette C. Martínez (2017). *Social Media Data Extraction and Content Analysis* (pp. 114-136).
www.irma-international.org/chapter/a-route-recommender-system-based-on-current-and-historical-crowdsourcing/161962

RCUBE: Parallel Multi-Dimensional ROLAP Indexing

Frank Dehne, Todd Eavis and Andrew Rau-Chaplin (2008). *International Journal of Data Warehousing and Mining* (pp. 1-14).
www.irma-international.org/article/rcube-parallel-multi-dimensional-rolap/1810

Ensemble PROBIT Models to Predict Cross Selling of Home Loans for Credit Card Customers

Hualin Wang, Yan Yu and Kaixia Zhang (2008). *International Journal of Data Warehousing and Mining* (pp. 15-21).
www.irma-international.org/article/ensemble-probit-models-predict-cross/1803

Finding Patterns in Class-Labeled Data Using Data Visualization

Gregor Leban, Minca Mramor, Blaž Zupan, Janez Demšar and Ivan Bratko (2008). *Data Mining Patterns: New Methods and Applications* (pp. 106-123).
www.irma-international.org/chapter/finding-patterns-class-labeled-data/7562