# Social User Interest Mining: Methods and Applications

Fattane Zarrinkalam
Ryerson University
Toronto, ON, Canada
fzarrinkalam@ryerson.ca

Hossein Fani
University of New Brunswick
Fredericton, NB, Canada
hfani@unb.ca

Ebrahim Bagheri
Ryerson University
Toronto, ON, Canada
bagheri@ryerson.ca

## ABSTRACT

The abundance of user generated content on social networks provides the opportunity to build models that are able to accurately and effectively extract, mine and predict users' interests with the hopes of enabling more effective user engagement, better quality delivery of appropriate services and higher user satisfaction. While traditional methods for building user profiles relied on AI-based preference elicitation techniques that could have been considered to be intrusive and undesirable by the users, more recent advances are focused on a non-intrusive yet accurate way of determining users' interests and preferences. In this tutorial, we cover five important aspects related to the effective mining of user interests: (1) we introduce the information sources that are used for extracting user interests, (2) various types of user interest profiles that have been proposed in the literature, (3) techniques that have been adopted or proposed for mining user interests, (4) the scalability and resource requirements of the state of the art methods, and finally (5) the evaluation methodologies that are adopted in the literature for validating the appropriateness of the mined user interest profiles. We also introduce existing challenges, open research question and exciting opportunities for further work.

## CCS CONCEPTS

• **Information systems** → **Social networks**; **Information extraction**; • **Human-centered computing** → **User models**; **Social networks**.

## 1 INTRODUCTION

Mining user interests from user behavioral data is critical for applications such as online advertising. Based on user interests, service providers such as advertisers, can significantly reduce service delivery costs by offering the most relevant products (e.g., ads) to their customers. The challenge of accurately and efficiently identifying user interests has been the subject of increasing attention in the past several years. Early approaches were based on explicit input

from individuals about their own interests. To avoid the extra burden of manually filling in and maintaining interest profiles, most methods in the past two decades have focused on the development of techniques that can automatically and unobtrusively determine users' interests based on user behavioral data from data sources such as browsing history, page visits, the links they click on, the searches they perform and the topics they interact with [15].

With the emergence and growing popularity of social networks such as blogging systems, wikis, social bookmarking, and microblogging services, many users are extensively engaged in at least some of these applications to express their feelings and views about a wide variety of social events/topics as they happen in real time by commenting, tagging, joining, sharing, liking, and publishing posts [22]. This has made social networks an exciting and unique source of information about users' interests. The development of techniques that can automatically model users' interests from online social networks would be highly important and have the potential to improve the quality of applications that work on a user modeling basis, such as filtering Twitter streams [21], news recommendation [1] and retweet prediction [13], among others.

In this tutorial, we comprehensively introduce different strategies proposed in the literature, including our own work [4, 10, 11, 27, 29–33], for mining user interests from social networks with respect to the following five perspectives:

(1) *Information Sources*: The type of information sources used for extracting user interests from within social networks such as textual content (comments, #tags), social network structure, and images [4, 28]. Additionally, we review external background knowledge sources such as semantic web resources and knowledge graphs that have been incorporated by some researchers to enhance the accuracy of user profiles [6, 30].

(2) *Profile Types:* Most of works in user interest mining from social networks extract users' explicit interests that are directly observable from user content [2, 23, 24, 31]. However, given the increasingly noticeable free-rider, some other techniques focus on passive users and extract their implicit interests by considering the interaction patterns between users and topics [25, 28, 29]. There is another line of work that is dedicated to predict users' future interests instead of modeling current interests of users [19, 30].

(3) *Underlying Techniques*: Previous methods have employed different techniques to build user profiles including neural embeddings [10, 16, 19], collaborative filtering [3, 5, 8, 18], topic modeling [17, 18, 32], link prediction [7, 29, 32], regression [4, 14], graph-based methods [9, 31] and Semantic Web technologies [12, 20, 30]. We review the techniques that have been used for identifying user interests and their different architectural variations.

(4) *Scalability and Resource Requirements:* Scalability is fundamental to user interest mining to accommodate torrents of social

content. To this end, we provide a comprehensive overview of the speed-accuracy (efficiency-accuracy) trade-off when building user interest profiles for existing techniques of the literature [26].

(5) *Evaluation Methodology:* Intrinsic vs. extrinsic evaluations are two main evaluation techniques, which have been widely adopted in the literature. Intrinsic evaluation helps to assess the quality of the constructed user interest profiles based on user studies [6, 20] while extrinsic evaluations measure the quality of the user interest profiles by looking at its impact on the effectiveness of other applications such as news recommendation and retweet prediction [31, 32]. We review how each of these evaluation methodologies have been used in the literature.

## 2 CONTENT OVERVIEW

This tutorial presents a comprehensive survey of user interest mining from online social networks and covers the following sections:

**Background and Introduction to Theory of User Interest Mining:** The tutorial begins with a session about basics of user interest mining and various online social networks. This includes preliminaries, motivations, and highlights on research questions to which user interest mining from online social networks would provide an answer for. Then, we introduce different third-party applications that can take advantage of user interest mining from social network to improve the accuracy of their results.

**Techniques and Methods in User Interest Mining from Online Social Networks:** Depending on the desirable type of user interest profiles, i.e., explicit, implicit or future user interest profiles, previous work have adopted different approaches for addressing the problem. Within these three categories, we lay out the details and provide a comparative analysis of existing methods in terms of their representation power, flexibility, resource needs and scalability.

**Evaluation Methodologies, Future Directions and Open Challenges:** In this session, we first elaborate on different resources and two main approaches used in the literature to evaluate user interest profiles, namely intrinsic vs extrinsic evaluation techniques. Next, this session presents exciting open research questions in the state-of-the-art for mining users' interests from online social networks. Accurate information extraction from online social networks poses unique challenges due to the special characteristics of them. Social posts are rather short, noisy and informal and they often do not provide sufficient contextual information for identifying their semantics. This tutorial presents the open issues that are important but have not been well addressed in recent studies. We cover potential resources (e.g., Linked Open Data) and techniques (e.g. Learning-to-Rank, deep learning architectures and causal inference) that can be relevant for mining user interests.

## REFERENCES

[1] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. 2011. Analyzing User Modeling on Twitter for Personalized News Recommendations. In *UMAP*. 1–12.
[2] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. 2011. Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web. In *ESWC*. 375–389.
[3] Amr Ahmed, Bhargav Kanagal, Sandeep Pandey, Vanja Josifovski, Lluis Garcia Pueyo, and Jeffrey Yuan. 2013. Latent factor models with additive and hierarchically-smoothed user preferences. In *WSDM*. 385–394.
[4] Negar Arabzadeh, Hossein Fani, Fattane Zarrinkalam, Ahmed Navivala, and Ebrahim Bagheri. 2018. Causal Dependencies for Future Interest Prediction on Twitter. In *CIKM*. 1511–1514.

[5] Hongyun Bao, Qiudan Li, Stephen Shaoyi Liao, Shuangyong Song, and Heng Gao. 2013. A new temporal and social PMF-based method to predict users' interests in micro-blogging. *Decision Support Systems* 55, 3 (2013), 698–709.
[6] Ceren Budak, Anitha Kannan, Rakesh Agrawal, and Jan Pedersen. 2014. Inferring User Interests From Microblogs. In *Technical Report, MSR-TR-2014-68*.
[7] Charalampos Chelmis and Viktor K. Prasanna. 2013. Social Link Prediction in Online Social Tagging Systems. *ACM Trans. Inf. Syst.* 31, 4 (2013), 20:1–20:27.
[8] Ernesto Diaz-Aviles, Lucas Drumond, Zeno Gantner, Lars Schmidt-Thieme, and Wolfgang Nejdl. 2012. What is happening right now ... that interests me?: online topic discovery and recommendation in twitter. In *CIKM*. 1592–1596.
[9] Yuxin Ding, Shengli Yan, Yibin Zhang, Wei Dai, and Li Dong. 2016. Predicting the attributes of social network users using a graph-based machine learning method. *Computer Communications* 73 (2016), 3–11.
[10] Hossein Fani, Ebrahim Bagheri, and Weichang Du. 2017. Temporally Like-minded User Community Identification through Neural Embeddings. In *CIKM*. 577–586.
[11] Hossein Fani, Ebrahim Bagheri, Fattane Zarrinkalam, Xin Zhao, and Weichang Du. 2018. Finding Diachronic Like-Minded Users. *Computational Intelligence* 34, 1 (2018), 124–144.
[12] Stefano Faralli, Giovanni Stilo, and Paola Velardi. 2017. Automatic acquisition of a taxonomy of microblogs users' interests. *J. Web Semant.* 45 (2017), 23–40.
[13] Wei Feng and Jianyong Wang. 2013. Retweet or not?: personalized tweet re-ranking. In *WSDM*. 577–586.
[14] Li Gao, Jia Wu, Chuan Zhou, and Yue Hu. 2017. Collaborative Dynamic Sparse Topic Regression with User Profile Evolution for Item Recommendation. In *AAAI*. 1316–1322.
[15] Fabio Gasparetti. 2017. Modeling user interests from web browsing activities. *Data Min. Knowl. Discov.* 31, 2 (2017), 502–547.
[16] Sadid A. Hasan, Yuan Ling, Joey Liu, and Oladimeji Farri. 2015. Exploiting Neural Embeddings for Social Media Data Analysis. In *TREC*.
[17] Liangjie Hong and Brian D. Davison. 2010. Empirical Study of Topic Modeling in Twitter. In *1st Workshop on Social Media Analytics*. 80–88.
[18] Liangjie Hong, Aziz S. Doumith, and Brian D. Davison. 2013. Co-factorization machines: modeling user interests and predicting individual decisions in Twitter. In *WSDM*. 557–566.
[19] Jaeyong Kang, Hongseok Choi, and Hyunju Lee. 2019. Deep recurrent convolutional networks for inferring user interests from social media. *J. Intell. Inf. Syst.* 52, 1 (2019), 191–209.
[20] Pavan Kapanipathi, Prateek Jain, Chitra Venkatramani, and Amit P. Sheth. 2014. User Interests Identification on Twitter Using a Hierarchical Knowledge Base. In *ESWC*. 99–113.
[21] Pavan Kapanipathi, Fabrizio Orlandi, Amit P. Sheth, and Alexandre Passant. 2011. Personalized Filtering of the Twitter Stream. In *Proceedings of the second Workshop on Semantic Personalized Information Management: Retrieval and Recommendation*. 6–13.
[22] Xin Li, Lei Guo, and Yihong Eric Zhao. 2008. Tag-based social interest discovery. In *WWW*. 675–684.
[23] Shangsong Liang, Zhaochun Ren, Yukun Zhao, Jun Ma, Emine Yilmaz, and Maarten de Rijke. 2017. Inferring Dynamic User Interests in Streams of Short Texts for User Clustering. *ACM Trans. Inf. Syst.* 36, 1 (2017), 10:1–10:37.
[24] Matthew Michelson and Sofus A. Macskassy. 2010. Discovering users' topics of interest on twitter: a first look. In *CIKM*. 73–80.
[25] Guangyuan Piao and John G. Breslin. 2017. Inferring User Interests for Passive Users on Twitter by Leveraging Followee Biographies. In *ECIR*. 122–133.
[26] Nemanja Spasojevic, Jinyun Yan, Adithya Rao, and Prantik Bhattacharyya. 2014. LASTA: large scale topic assignment on multiple social networks. In *ACM SIGKDD*. 1809–1818.
[27] Anil Kumar Trikha, Fattane Zarrinkalam, and Ebrahim Bagheri. 2018. Topic-Association Mining for User Interest Detection. In *ECIR*. 665–671.
[28] Jinpeng Wang, Wayne Xin Zhao, Yulan He, and Xiaoming Li. 2014. Infer User Interests via Link Structure Regularization. *ACM TIST* 5, 2 (2014), 23:1–23:22.
[29] Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, and Mohsen Kahani. 2016. Inferring Implicit Topical Interests on Twitter. In *ECIR*. 479–491.
[30] Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, and Mohsen Kahani. 2017. Predicting Users' Future Interests on Twitter. In *ECIR*. 464–476.
[31] Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, Mohsen Kahani, and Weichang Du. 2015. Semantics-Enabled User Interest Detection from Twitter. In *WI-IAT*. 469–476.
[32] Fattane Zarrinkalam, Mohsen Kahani, and Ebrahim Bagheri. 2018. Mining user interests over active topics on social networks. *Inf. Process. Manage.* 54, 2 (2018), 339–357.
[33] Fattane Zarrinkalam, Mohsen Kahani, and Ebrahim Bagheri. 2018. User interest prediction over future unobserved topics on social networks. *Information Retrieval Journal* (2018), 1–36.