Mining User Interests from Social Media

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

Guangyuan Piao
NOKIA Bell Labs
Dublin, Ireland
guangyuan.piao@nokia-bell-labs.com

Stefano Faralli
University of Rome Unitelma Sapienza
Rome, Italy
stefano.faralli@unitelmasapienza.it

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

ABSTRACT

Social media users readily share their preferences, life events, sentiment and opinions, and implicitly signal their thoughts, feelings, and psychological behavior. This makes social media a viable source of information to accurately and effectively mine users' interests with the hopes of enabling more effective user engagement, better quality delivery of appropriate services and higher user satisfaction. In this tutorial, we cover five important aspects related to the effective mining of user interests: (1) the foundations of social user interest modeling, such as information sources, various types of representation models and temporal features, (2) techniques that have been adopted or proposed for mining user interests, (3) different evaluation methodologies and benchmark datasets, (4) different applications that have been taking advantage of user interest mining from social media platforms, and (5) existing challenges, open research questions and exciting opportunities for further work.

CCS CONCEPTS

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

ACM Reference Format:
Fattane Zarrinkalam, Guangyuan Piao, Stefano Faralli, and Ebrahim Bagheri. 2020. Mining User Interests from Social Media. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3340531.3412167

1 CONTEXT AND MOTIVATION

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 [3, 10].

With the emergence and growing popularity of social media 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 [6]. This has made social media an exciting and unique source of information about users' interests [16]. The development of techniques that can automatically detect such topics and model users' interests towards them from online social media 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 [13], news recommendation [1] and retweet prediction [9], among others.

2 TARGET AUDIENCE AND PREREQUISITES

The target audience for this tutorial are those who have familiarity with social media mining and basics of data mining techniques. Where appropriate the tutorial does not make any assumptions about the audience's knowledge on more advanced techniques such as link prediction, matrix factorization, deep matching, entity linking and knowledge-graph based reasoning, among others. As such, sufficient details are provided as appropriate so that the content are accessible and understandable to those who have a fundamental understanding of data mining principles. The tutorial only assumes familiarity with topics included in an undergraduate data mining course.

3 TUTORIAL OUTLINE

This tutorial presents a comprehensive survey of user interest mining from social media. In particular, this tutorial 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 from various social media such as information sources, representation units to represent each topic of interest

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-6859-9/20/10.
https://doi.org/10.1145/3340531.3412167

and user interest profile, temporal aspects and cross-system user interest modeling. This section also highlights on research questions to which user interest mining from social media would provide an answer for. Finally, we review topics that are covered in the tutorial followed by a disclaimer, i.e., what the tutorial is *not* about.

Techniques and Methods in User Interest Mining from Social Media. Depending on the desirable type of user interest profiles, i.e., explicit or 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. Specifically, in this session, we elaborate on how previous studies have used different techniques such as collaborative filtering [2, 11], topic modeling [11, 20], link prediction [5, 20], graph-based methods [7, 19], Semantic Web technologies [8, 12, 21] and association rule mining [18] to construct a given type of interest profile for users (e.g. implicit interest profile).

Evaluation Methodologies, Benchmark Datasets and Applications of Interest Mining from Social Media. 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. Intrinsic evaluation helps to assess the quality of the constructed user interest profiles based on user studies [4, 12, 14] 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 [15, 19, 20]. Then, we describe the existing benchmark datasets and evaluation metrics [17]. Next, we introduce different applications that have been taking advantage of user interest modeling from social media platforms to improve their services.

Future Directions and Open Challenges. In this session, we present exciting open research questions in the state of the art for mining users' interests from social media. Accurate information extraction from social media 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. In other words, the semantics of the context of the communicated information within a post is often implicit. Moreover, as a large majority of social network users are free-riders and cold start users, the interests of such users is challenging and they cannot be directly identified from their explicit contributions to the online social network. This tutorial presents the open issues that are important but have not been well addressed in recent studies which can inspire future directions in this research field. 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 *User Modeling, Adaptation and Personalization - 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings.* 1–12. https://doi.org/10.1007/978-3-642-22362-4_1

[2] 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 *Sixth ACM International Conference on Web Search and Data Mining, WSDM 2013, Rome, Italy, February 4-8, 2013.* 385–394. https://doi.org/10.1145/2433396.2433445

[3] Alex Beutel, Leman Akoglu, and Christos Faloutsos. 2015. Graph-Based User Behavior Modeling: From Prediction to Fraud Detection. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015.* 2309–2310. https://doi.org/10.1145/2783258.2789985

[4] C. Budak, A. Kannan, R. Agrawal, , and J. Pedersen. 2014. Inferring User Interests From Microblogs. In *Technical Report, MSR-TR-2014-68.*

[5] 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. https://doi.org/10.1145/2516891

[6] Pin-Yu Chen, Chun-Chen Tu, Pai-Shun Ting, Ya-Yun Lo, Danai Koutra, and Alfred O. Hero III. 2016. Identifying Influential Links for Event Propagation on Twitter: A Network of Networks Approach. *CoRR* abs/1609.05378 (2016). arXiv:1609.05378 http://arxiv.org/abs/1609.05378

[7] 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. https://doi.org/10.1016/j.comcom.2015.07.007

[8] Stefano Faralli, Giovanni Stilo, and Paola Velardi. 2017. Automatic acquisition of a taxonomy of microblogs users' interests. *J. Web Semant.* 45 (2017), 23–40. https://doi.org/10.1016/j.websem.2017.05.004

[9] Wei Feng and Jianyong Wang. 2013. Retweet or not?: personalized tweet re-ranking. In *Sixth ACM International Conference on Web Search and Data Mining, WSDM 2013, Rome, Italy, February 4-8, 2013.* 577–586. https://doi.org/10.1145/2433396.2433470

[10] Fabio Gasparetti. 2017. Modeling user interests from web browsing activities. *Data Min. Knowl. Discov.* 31, 2 (2017), 502–547. https://doi.org/10.1007/s10618-016-0482-x

[11] Liangjie Hong, Aziz S. Doumith, and Brian D. Davison. 2013. Co-factorization machines: modeling user interests and predicting individual decisions in Twitter. In *Sixth ACM International Conference on Web Search and Data Mining, WSDM 2013, Rome, Italy, February 4-8, 2013.* 557–566. https://doi.org/10.1145/2433396.2433467

[12] Pavan Kapanipathi, Prateek Jain, Chitra Venkatramani, and Amit P. Sheth. 2014. User Interests Identification on Twitter Using a Hierarchical Knowledge Base. In *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings.* 99–113. https://doi.org/10.1007/978-3-319-07443-6_8

[13] 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 2011, Bonn, Germany, October 24, 2011.* 6–13.

[14] Fedelucio Narducci, Cataldo Musto, Giovanni Semeraro, Pasquale Lops, and Marco de Gemmis. 2013. Leveraging Encyclopedic Knowledge for Transparent and Serendipitous User Profiles. In *User Modeling, Adaptation, and Personalization - 21th International Conference, UMAP 2013, Rome, Italy, June 10-14, 2013, Proceedings.* 350–352. https://doi.org/10.1007/978-3-642-38844-6_36

[15] Guangyuan Piao and John G. Breslin. 2016. Analyzing Aggregated Semantics-enabled User Modeling on Google+ and Twitter for Personalized Link Recommendations. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, UMAP 2016, Halifax, NS, Canada, July 13 - 17, 2016.* 105–109. https://doi.org/10.1145/2930238.2930278

[16] Guangyuan Piao and John G. Breslin. 2018. Inferring user interests in microblogging social networks: a survey. *User Model. User-Adapt. Interact.* 28, 3 (2018), 277–329. https://doi.org/10.1007/s11257-018-9207-8

[17] Giorgia Di Tommaso, Stefano Faralli, Giovanni Stilo, and Paola Velardi. 2018. Wiki-MID: A Very Large Multi-domain Interests Dataset of Twitter Users with Mappings to Wikipedia. In *17th International Semantic Web Conference.* 36–52.

[18] Anil Kumar Trikha, Fattane Zarrinkalam, and Ebrahim Bagheri. 2018. Topic-Association Mining for User Interest Detection. In *ECIR.* 665–671.

[19] Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, Mohsen Kahani, and Weichang Du. 2015. Semantics-Enabled User Interest Detection from Twitter. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015, Singapore, December 6-9, 2015 - Volume I.* 469–476. https://doi.org/10.1109/WI-IAT.2015.182

[20] 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. https://doi.org/10.1016/j.ipm.2017.12.003

[21] Fattane Zarrinkalam, Mohsen Kahani, and Ebrahim Bagheri. 2019. User interest prediction over future unobserved topics on social networks. *Inf. Retr. Journal* 22, 1-2 (2019), 93–128. https://doi.org/10.1007/s10791-018-9337-y