Social Media Mining

In recent years, with the development of networks and virtual communities, social media[1, 2, 3, 4, 5, 6] has become ubiquitous and prosperous for social networking and content sharing based on the foundations of Web 2.0. Everyone can be a media outlet or producer and the barrier of communication disappeared. Research[7] shows that the audience spends 22 percent of their time on social networking sites, people all over the world are creating, exchanging information and socializing online, thus proving how popular social media platforms have become. Furthermore, social media depending on mobiles and web-based technologies to create highly interactive platforms brings people together in more creative and easier ways.

Social media provides opportunities to understand individuals at scale and mine human behavioral patterns by integrating social theories and computational sciences[8, 9, 10]. However, social media data[8] is massive, dynamic, linked, noisy, loosely defined and incomplete which is different from those in traditional data mining area and calls for novel data mining techniques that can handle user-generated content[11] with rich social relations effectively. The study and development of these new techniques is called Social Media Mining, an emerging discipline under the umbrella of data mining.

Social Media Mining(SMM)[10, 12, 13] is the process of representing, analyzing, and extracting actionable patterns from social media data. SMM represents the virtual world of social media in a computable way, and designs models that can help us understand its interactions. SMM covers a series of research areas[14, 15, 16].

1. Social Network Structural Properties and Their Evolution

   - Network Measures[17]
   - Network Models[17]
   - Community Detection[18, 19, 20]

2. Social Groups and Their Interaction Law

   - Sentiment Analysis[21, 22, 23, 24]
   - Behavior Analysis[8, 25, 26]
   - User Migration Patterns[27, 28]
   - Trust in Social Media[29, 30, 31, 32]
   - User Vulnerability Management[33, 34, 35]
   - Link Prediction[36, 37, 38]

3. Information Dissemination in Social Network

   - Information Propagation[39, 40, 41, 42]
   - Influence and Homophily[43]
   - Provenance of Information in Social Media[44, 45, 46, 47]

4. Business and Marketing

   - Crowdsourcing[48]

- Recommendation in Social Media[49, 50, 51]
- Location-based Social Network Mining[52, 28, 53, 54]
- Social Search[55, 56, 57]

5. Others

- Data Cleansing[58, 59]
- Spam Detection[60, 61, 62]
- Social Media in Crisis[63, 64]

However, the emergence of SMM raises new research challenges[65, 66, 8, 11, 67, 4, 68, 69, 70, 20, 71, 72, 10, 16].

1. Too Much Data and Too Less Data[10]

   Social media data is massive, however, it's not just a matter of Scale. We can observe global phenomena the generated from social media data, but invisible at smaller scales, say, don't really know what any one node or link means. So it is important to find the point where the lines of research converge. It's easy to measure macro-level things; hard to pose nuanced questions, especially micro-level, which is the dilemma we are facing.

2. Noise Removal[10]

   Actually, the key point of SMM is to find significant conversations and act on it. However, by some estimates, almost half of Twitter accounts are either ciphers or spams(also, Spam Detection is a hot research area of SMM). Social media noise can be a real productivity issue for anyone involved in SMM area.

3. Privacy Protection[66]

   Many large datasets is based on media (e-mail, IM, voice) which contains lots of user private information. With more detailed data, anonymization has run into trouble. So privacy protection is another factor we can take into account to evaluate a SMM system.

4. Dynamic Networks, Changing All the Time[16]

   As social networks are dynamic, discovering communities and build corresponding social graphs from social data will continue to be a dynamic research challenge. Besides, public opinions or interests will change with time, it's also a hard job to do sentiment analysis.

5. Multiple Provenace[44, 73]

   Social media is decentralized which means that information can be published by anyone on social media networks. This kind of environment provides unique challenges for tracking provenances.

References

[1] WikiPedia. *Social Media*. 16 December 2015. URL: https://en.wikipedia.org/wiki/Social_media.

[2] Eugene Agichtein et al. "Finding high-quality content in social media". In: *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM. 2008, pp. 183–194.

[3]   Nicole B Ellison et al. "Social network sites: Definition, history, and scholarship". In: *Journal of Computer-Mediated Communication* 13.1 (2007), pp. 210–230.

[4]   Daniel Zeng et al. "Social media analytics and intelligence". In: *Intelligent Systems, IEEE* 25.6 (2010), pp. 13–16.

[5]   WikiPedia. *List of social networking websites.* 16 December 2015. URL: `https://en.wikipedia.org/wiki/List_of_social_networking_websites`.

[6]   W Glynn Mangold and David J Faulds. "Social media: The new hybrid element of the promotion mix". In: *Business horizons* 52.4 (2009), pp. 357–365.

[7]   Nielsen Wire. "Social networks/blogs now account for one in every four and a half minutes online". In: *Nielson Wire* (2010).

[8]   Fei-Yue Wang et al. "Social computing: From social informatics to social intelligence". In: *Intelligent Systems, IEEE* 22.2 (2007), pp. 79–83.

[9]   Jiliang Tang, Yi Chang, and Huan Liu. "Mining social media with social theories: A survey". In: *ACM SIGKDD Explorations Newsletter* 15.2 (2014), pp. 20–29.

[10] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. *Social media mining: an introduction.* Cambridge University Press, 2014.

[11] Andreas M Kaplan and Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media". In: *Business horizons* 53.1 (2010), pp. 59–68.

[12] Pritam Gundecha and Huan Liu. "Mining social media: a brief introduction". In: *Tutorials in Operations Research* 1.4 (2012).

[13] Reza Zafarani Mohammad Ali Abbasi Huan Liu. *Social Media Mining: Fundamental Issues and Challenges.* December 10, 2013. URL: `http://ecs.syr.edu/faculty/reza/tutorials/ICDM13/TutorialICDM13SMM.pdf`.

[14] WikiPedia. *Social Media Mining.* 7 September 2015. URL: `https://en.wikipedia.org/wiki/Social_media_mining`.

[15] Huan Liu. *Social Computing: Challenges in Research and Applications.* 2013.03. URL: `http://www.iscintelligence.com/archivos_subidos/liu,_huan._social_computing_challenges_in_research_and_applications.pdf`.

[16] Han Yi Fang Bingxing Jia Yan. "Social Network AnalysisKey Research Problems, Related Work, and Future Prospects". In: (Feb, 2015).

[17] Derek Hansen, Ben Shneiderman, and Marc A Smith. *Analyzing social media networks with NodeXL: Insights from a connected world.* Morgan Kaufmann, 2010.

[18] Santo Fortunato. "Community detection in graphs". In: *Physics Reports* 486.3 (2010), pp. 75–174.

[19] Lei Tang and Huan Liu. "Community detection and mining in social media". In: *Synthesis Lectures on Data Mining and Knowledge Discovery* 2.1 (2010), pp. 1–137.

[20] Symeon Papadopoulos et al. "Community detection in social media". In: *Data Mining and Knowledge Discovery* 24.3 (2012), pp. 515–554.

[21] Bo Pang and Lillian Lee. "Opinion mining and sentiment analysis". In: *Foundations and trends in information retrieval* 2.1-2 (2008), pp. 1–135.

[22] Sitaram Asur, Bernardo Huberman, et al. "Predicting the future with social media". In: *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on.* Vol. 1. IEEE. 2010, pp. 492–499.

[23] Pawel Sobkowicz, Michael Kaschesky, and Guillaume Bouchard. "Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web". In: *Government Information Quarterly* 29.4 (2012), pp. 470–479.

[24] Xia Hu et al. "Exploiting social relations for sentiment analysis in microblogging". In: *Proceedings of the sixth ACM international conference on Web search and data mining.* ACM. 2013, pp. 537–546.

[25] Eileen Fischer and A Rebecca Reuber. "Social interaction via new social media:(How) can interactions on Twitter affect effectual thinking and behavior?" In: *Journal of business venturing* 26.1 (2011), pp. 1–18.

[26] Mohammad-Ali Abbasi et al. "Real-world behavior analysis through a social media lens". In: *Social Computing, Behavioral-Cultural Modeling and Prediction.* Springer, 2012, pp. 18–26.

[27] Shamanth Kumar, Reza Zafarani, and Huan Liu. "Understanding User Migration Patterns in Social Media." In: *AAAI.* Citeseer. 2011.

[28] Eunjoon Cho, Seth A Myers, and Jure Leskovec. "Friendship and mobility: user movement in location-based social networks". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM. 2011, pp. 1082–1090.

[29] Jiliang Tang, Huiji Gao, and Huan Liu. "mTrust: discerning multi-faceted trust in a connected world". In: *Proceedings of the fifth ACM international conference on Web search and data mining.* ACM. 2012, pp. 93–102.

[30] Jiliang Tang et al. "eTrust: Understanding trust evolution in an online world". In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM. 2012, pp. 253–261.

[31] Jiliang Tang et al. "Exploiting homophily effect for trust prediction". In: *Proceedings of the sixth ACM international conference on Web search and data mining.* ACM. 2013, pp. 53–62.

[32] Jiliang Tang and Huan Liu. "Trust in Social Media". In: *Synthesis Lectures on Information Security, Privacy, & Trust* 10.1 (2015), pp. 1–129.

[33] Pritam Gundecha, Geoffrey Barbier, and Huan Liu. "Exploiting vulnerability to secure user privacy on a social networking site". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM. 2011, pp. 511–519.

[34] Munmun De Choudhury et al. "Predicting Depression via Social Media." In: *ICWSM.* 2013.

[35] Pritam Gundecha et al. "User vulnerability and its reduction on a social networking site". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 9.2 (2014), p. 12.

[36] David Liben-Nowell and Jon Kleinberg. "The link-prediction problem for social networks". In: *Journal of the American society for information science and technology* 58.7 (2007), pp. 1019–1031.

[37] Linyuan Lü and Tao Zhou. "Link prediction in complex networks: A survey". In: *Physica A: Statistical Mechanics and its Applications* 390.6 (2011), pp. 1150–1170.

[38] Dashun Wang et al. "Human mobility, social ties, and link prediction". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM. 2011, pp. 1100–1108.

[39] Meeyoung Cha, Alan Mislove, and Krishna P Gummadi. "A measurement-driven analysis of information propagation in the flickr social network". In: *Proceedings of the 18th international conference on World wide web.* ACM. 2009, pp. 721–730.

[40] Daniel M Romero et al. "Influence and passivity in social media". In: *Machine learning and knowledge discovery in databases*. Springer, 2011, pp. 18–33.

[41] Eytan Bakshy et al. "The role of social networks in information diffusion". In: *Proceedings of the 21st international conference on World Wide Web*. ACM. 2012, pp. 519–528.

[42] Greg Ver Steeg and Aram Galstyan. "Information transfer in social media". In: *Proceedings of the 21st international conference on World Wide Web*. ACM. 2012, pp. 509–518.

[43] Munmun De Choudhury et al. "" Birds of a Feather": Does User Homophily Impact Information Diffusion in Social Media?" In: *arXiv preprint arXiv:1006.1702* (2010).

[44] Geoffrey Barbier and Huan Liu. "Information provenance in social media". In: *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, 2011, pp. 276–283.

[45] Geoffrey Barbier et al. "Provenance data in social media". In: *Synthesis Lectures on Data Mining and Knowledge Discovery* 4.1 (2013), pp. 1–84.

[46] P Gundecha, Z Feng, and H Liu. "Seeking provenance of information in social media". In: *The 22nd ACM International Conference on Information and Knowledge Management*. 2013.

[47] Tom De Nies et al. "Towards Multi-level Provenance Reconstruction of Information Diffusion on Social Media". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 2015, pp. 1823–1826.

[48] Anhai Doan, Raghu Ramakrishnan, and Alon Y Halevy. "Crowdsourcing systems on the world-wide web". In: *Communications of the ACM* 54.4 (2011), pp. 86–96.

[49] Ido Guy et al. "Personalized recommendation of social software items based on social relations". In: *Proceedings of the third ACM conference on Recommender systems*. ACM. 2009, pp. 53–60.

[50] Ido Guy et al. "Social media recommendation based on people and tags". In: *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM. 2010, pp. 194–201.

[51] Jiliang Tang, Jie Tang, and Huan Liu. "Recommendation in social media: recent advances and new frontiers". In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2014, pp. 1977–1977.

[52] Yu Zheng et al. "GeoLife2. 0: a location-based social networking service". In: *Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on*. IEEE. 2009, pp. 357–358.

[53] Huiji Gao, Jiliang Tang, and Huan Liu. "Exploring Social-Historical Ties on Location-Based Social Networks." In: *ICWSM*. 2012.

[54] Jie Bao, Yu Zheng, and Mohamed F Mokbel. "Location-based and preference-aware recommendation using sparse geo-social networking data". In: *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. ACM. 2012, pp. 199–208.

[55] Shenghua Bao et al. "Optimizing web search using social annotations". In: *Proceedings of the 16th international conference on World Wide Web*. ACM. 2007, pp. 501–510.

[56] David Carmel et al. "Personalized social search based on the user's social network". In: *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM. 2009, pp. 1227–1236.

[57] W Bruce Croft, Donald Metzler, and Trevor Strohman. *Search engines: Information retrieval in practice.* Addison-Wesley Reading, 2010.

[58] Mauricio A Hernández and Salvatore J Stolfo. "Real-world data is dirty: Data cleansing and the merge/purge problem". In: *Data mining and knowledge discovery* 2.1 (1998), pp. 9–37.

[59] Heiko Müller and Johann-Christph Freytag. *Problems, methods, and challenges in comprehensive data cleansing.* Professoren des Inst. Für Informatik, 2005.

[60] Benjamin Markines, Ciro Cattuto, and Filippo Menczer. "Social spam detection". In: *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web.* ACM. 2009, pp. 41–48.

[61] Jonghyuk Song, Sangho Lee, and Jong Kim. "Spam filtering in twitter using sender-receiver relationship". In: *Recent Advances in Intrusion Detection.* Springer. 2011, pp. 301–317.

[62] Xia Hu et al. "Social spammer detection in microblogging". In: *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence.* AAAI Press. 2013, pp. 2633–2639.

[63] Leysia Palen. "Online social media in crisis events". In: *Educause Quarterly* 31.3 (2008), pp. 76–78.

[64] Shari R Veil, Tara Buehner, and Michael J Palenchar. "A Work-In-Process Literature Review: Incorporating Social Media in Risk and Crisis Communication". In: *Journal of contingencies and crisis management* 19.2 (2011), pp. 110–122.

[65] Manoj Parameswaran and Andrew B Whinston. "Research issues in social computing". In: *Journal of the Association for Information Systems* 8.6 (2007), p. 22.

[66] Jon M Kleinberg. "Challenges in mining social network data: processes, privacy, and paradoxes". In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM. 2007, pp. 4–5.

[67] Robert V Kozinets. *Netnography.* Wiley Online Library, 2010.

[68] Jim Hendler and Tim Berners-Lee. "From the Semantic Web to social machines: A research challenge for AI on the World Wide Web". In: *Artificial Intelligence* 174.2 (2010), pp. 156–161.

[69] Lev Manovich. "Trending: the promises and the challenges of big social data". In: *Debates in the digital humanities* (2011), pp. 460–475.

[70] Judd Antin and Elizabeth F Churchill. "Badges in social media: A social psychological perspective". In: *CHI 2011 Gamification Workshop Proceedings (Vancouver, BC, Canada, 2011).* 2011.

[71] Jan H Kietzmann et al. "Social media? Get serious! Understanding the functional building blocks of social media". In: *Business horizons* 54.3 (2011), pp. 241–251.

[72] Jie Yin et al. "Using social media to enhance emergency situation awareness". In: *IEEE Intelligent Systems* 6 (2012), pp. 52–59.

[73] Ariel M Greenberg, William G Kennedy, and Nathan D Bos. *Social Computing, Behavioral-Cultural Modeling and Prediction.* Springer, 2012.