demonstrates the capability of reducing human labels by leveraging keyword information.

4.5 Evaluation on Keywords Generation

Tables 3, 4, and 5 list example keywords of one time window over the three events. Note that in the first time window for all three events, the top-ranked keywords for *AL* and *SSEM* are ranked by the importance score fs (see Section 3.2.2 for the definition of fs), which is the single time window ranking. From the second time window on, the top ranked keywords are generated by considering keyword graph changes. Results from all three tables are generated by considering keywords' dynamic changes over time. Some topic-model based methods are with fewer keywords because they generate the same keywords in different topics.

From the results, some important keywords appear all the time along the development of the event, such as *earthquake, chile* for the *EQ* event, *flood, flooding* for the *Flood* event, and *NBA, dunk* for the *NBA* event. This kind of keywords could be produced by all baseline methods and ranked higher than other words. However, generating such kind of keywords all the time does not help introduce new aspects of events. To make the newly added keywords effective on collecting event-related data, our keyword generation method tries to rank such kind of words lower. Emerging words are given higher ranks. Results from the three tables demonstrate the desired results, *e.g., terremoto, @reuters* for *EQ*, *disaster, #infloodwetrust* for *Flood*, and *#dunkcontest, #legend* for *NBA* are extracted and rank higher. That is the proposed keyword generation method could produce keywords representing the emerging aspects of an event and also the important user accounts, *e.g., @reuters, @bleacherreport*, which is also a desirable feature for depicting a developing event. However, the other methods could not produce such useful keywords and would generate some irrelevant words.

The fs and *ReRank* in Tables 3, 4 and 5 give a comparison example of word ranking by the importance score fs and re-ranking results by considering keywords' dynamic changes. Observe that, before considering word changes over time, top ranked words by importance score fs include those core keywords of each event or words that have already been added before, *e.g., #prayforchile, NBA*. After considering the keyword graph changes over time, more relevant keywords are introduced, *e.g., #tsunami, iquique, #dunkcontest.*

5 CONCLUSION

We propose a semi-supervised expectation and maximization mechanism to collect high-quality tweets related to an event of interest from the Twitter stream. We make the first attempt to leverage word properties to help identify event-related tweets, which could reduce human annotation effort and maintain high performance in terms of F_1 measure. By considering the word graph changing over time, we can generate keywords with few overlap with historical ones. Existing works do not consider words' dynamic changes. Since keywords that could depict the emerging aspects of events are preferred, one can also refer to the top-ranked keywords to learn about the event development. Our proposed method outperforms state-of-the-art methods on both event-related tweets identification and keywords generation.

ACKNOWLEDGMENTS

The first author is in the SAP Industrial Ph.D Program, partially funded by the Economic Development Board and the National Research Foundation of Singapore.

REFERENCES

[1] Hila Becker, Dan Iter, Mor Naaman, and Luis Gravano. 2012. Identifying content for planned events across social media sites. In *WSDM*. 533–542.
[2] Hila Becker, Mor Naaman, and Luis Gravano. 2011. Beyond Trending Topics: Real-World Event Identification on Twitter. In *ICWSM*.
[3] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.
[4] Praveen Bommannavar, Jimmy J. Lin, and Anand Rajaraman. 2016. Estimating topical volume in social media streams. In *SAC*. 1096–1101.
[5] Chris Buckley, Gerard Salton, James Allan, and Amit Singhal. 1995. Automatic query expansion using SMART: TREC 3. *NIST special publication sp* (1995), 69–69.
[6] Moses Charikar, Chandra Chekuri, Tomás Feder, and Rajeev Motwani. 2004. Incremental Clustering and Dynamic Information Retrieval. *SIAM J. Comput.* 33, 6 (2004), 1417–1440.
[7] Yan Chen, Hadi Amiri, Zhoujun Li, and Tat-Seng Chua. 2013. Emerging topic detection for organizations from microblogs. In *SIGIR*. 43–52.
[8] Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malú Castellanos, and Riddhiman Ghosh. 2013. Discovering coherent topics using general knowledge. In *CIKM*. 209–218.
[9] Xueqi Cheng, Jiafeng Guo, Shenghua Liu, Yanfeng Wang, and Xiaohui Yan. 2013. Learning Topics in Short Texts by Non-negative Matrix Factorization on Term Correlation Matrix. In *SDM*. 749–757.
[10] Steffen Eger and Alexander Mehler. 2016. On the Linearity of Semantic Change: Investigating Meaning Variation via Dynamic Graph Models. In *ACL(2)*.
[11] Feifan Fan, Yansong Feng, Lili Yao, and Dongyan Zhao. 2016. Adaptive Evolutionary Filtering in Real-Time Twitter Stream. In *CIKM*. 1079–1088.
[12] Kazi Saidul Hasan and Vincent Ng. 2014. Automatic Keyphrase Extraction: A Survey of the State of the Art.. In *ACL (1)*. 1262–1273.
[13] Shiva Prasad Kasiviswanathan, Prem Melville, Arindam Banerjee, and Vikas Sindhwani. 2011. Emerging topic detection using dictionary learning. In *CIKM*. 745–754.
[14] Henry Oliver Lancaster and Eugene Seneta. 2005. *Chi-Square Distribution*. Wiley Online Library.
[15] David D. Lewis and Jason Catlett. 1994. Heterogeneous Uncertainty Sampling for Supervised Learning. In *ICML*. Morgan Kaufmann, 148–156.
[16] Cheng Li, Yue Wang, Paul Resnick, and Qiaozhu Mei. 2014. ReQ-ReC: high recall retrieval with query pooling and interactive classification. In *SIGIR*. 163–172.
[17] Zhiyuan Liu, Wenyi Huang, Yabin Zheng, and Maosong Sun. 2010. Automatic Keyphrase Extraction via Topic Decomposition. In *EMNLP*. 366–376.
[18] Walid Magdy and Tamer Elsayed. 2014. Adaptive Method for Following Dynamic Topics on Twitter. In *ICWSM*.
[19] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into texts. In *EMNLP*, Vol. 4. 404–411.
[20] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to Expect When the Unexpected Happens: Social Media Communications Across Crises. In *CSCW*. 994–1009.
[21] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank citation ranking: bringing order to the web. (1999).
[22] Mehdi Sadri, Sharad Mehrotra, and Yaming Yu. 2016. Online Adaptive Topic Focused Tweet Acquisition. In *CIKM*. 2353–2358.
[23] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *WWW*. 851–860.
[24] Simon Tong and Daphne Koller. 2002. Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research* 2 (2002), 45–66.
[25] Abraham Wald. 1945. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics* 16, 2 (1945), 117–186.
[26] Shuai Wang, Zhiyuan Chen, Geli Fei, Bing Liu, and Sherry Emery. 2016. Targeted Topic Modeling for Focused Analysis. In *SIGKDD*. 1235–1244.
[27] Shuai Wang, Zhiyuan Chen, Bing Liu, and Sherry Emery. 2016. Identifying Search Keywords for Finding Relevant Social Media Posts. In *AAAI*. 3052–3058.
[28] Ming Yang, Melody Y. Kiang, and Wei Shang. 2015. Filtering big data from social media - Building an early warning system for adverse drug reactions. *Journal of Biomedical Informatics* 54 (2015), 230–240.
[29] Yisong Yue and Thorsten Joachims. 2009. Interactively optimizing information retrieval systems as a dueling bandits problem. In *ICML*. 1201–1208.
[30] Deyu Zhou, Liangyu Chen, and Yulan He. 2015. An Unsupervised Framework of Exploring Events on Twitter: Filtering, Extraction and Categorization. In *AAAI*. 2468–2475.