

Report on Twitter topic detection and analysis

Alex Tsilingiris
Christina Pardalidou
Sotiris Karapostolakis

June 27, 2017

CONTENTS

1	Introduction	2
2	Model description	3
3	Method implementation details	4
4	Sample results	5
5	Software implementation details	6

1 INTRODUCTION

Text

2 MODEL DESCRIPTION

Section

Test

3 METHOD IMPLEMENTATION DETAILS

Section

Test

Section

Test

Section

Test

- [1] H. Abdelhaq, C. Sengstock, and M. Gertz. Eventtweet: Online localized event detection from twitter. *Proceedings of the VLDB Endowment*, 6(12):1326–1329, 2013.
- [2] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, pages 61–70. ACM, 2010.
- [3] A. Bakliwal, P. Arora, S. Madhappan, N. Kapre, M. Singh, and V. Varma. Mining sentiments from tweets. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, pages 11–18, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [5] N. N. Bora. Summarizing public opinions in tweets. *International Journal of Computational Linguistics and Applications*, 3(1):41–55, 2012.
- [6] C. Budak, D. Agrawal, and A. El Abbadi. Structural trend analysis for online social networks. *Proceedings of the VLDB Endowment*, 4(10):646–656, 2011.
- [7] R. O. Bueno, A. F. Bruzón, Y. Gutiérrez-Vázquez, and A. Montoyo. Ssa-uo: Unsupervised sentiment analysis in twitter. In *SemEval@NAACL-HLT*, 2013.
- [8] M. Cataldi, L. D. Caro, and C. Schifanella. Personalized emerging topic detection based on a term aging model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):7, 2013.
- [9] D. Chatzakou, V. Koutsonikola, A. Vakali, and K. Kafetsios. Micro-blogging content analysis via emotionally-driven clustering. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 375–380. IEEE, 2013.
- [10] D. Chatzakou, N. Passalis, and A. Vakali. Multispot: Spotting sentiments with semantic aware multilevel cascaded analysis. In *International Conference on Big Data Analytics and Knowledge Discovery*, pages 337–350. Springer, 2015.
- [11] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 759–768. ACM, 2010.
- [12] R. Compton, D. Jurgens, and D. Allen. Geotagging one hundred million twitter accounts with total variation minimization. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 393–401. IEEE, 2014.

- [13] A. Cui, M. Zhang, Y. Liu, and S. Ma. Emotion tokens: Bridging the gap among multilingual twitter sentiment analysis. In *Proceedings of the 7th Asia Conference on Information Retrieval Technology*, AIRS'11, pages 238–249, Berlin, Heidelberg, 2011. Springer-Verlag.
- [14] Y. Dang, Y. Zhang, and H. Chen. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems*, 25(4):46–53, 2010.
- [15] C. A. Davis Jr, G. L. Pappa, D. R. R. de Oliveira, and F. de L Arcanjo. Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, 15(6):735–751, 2011.
- [16] M. Desai and M. A. Mehta. Techniques for sentiment analysis of twitter data: A comprehensive survey. In *Computing, Communication and Automation (ICCCA), 2016 International Conference on*, pages 149–154. IEEE, 2016.
- [17] C. N. Dos Santos and M. Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In *COLING*, pages 69–78, 2014.
- [18] D. Gao, W. Li, and R. Zhang. Sequential summarization: A new application for timely updated twitter trending topics. In *ACL (2)*, pages 567–571. Citeseer, 2013.
- [19] M. Ghiassi, J. Skinner, and D. Zimbra. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst. Appl.*, 40(16):6266–6282, Nov. 2013.
- [20] A. Giachanou and F. Crestani. Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2):28, 2016.
- [21] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12), 2009.
- [22] B. Han, P. Cook, and T. Baldwin. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, 49:451–500, 2014.
- [23] T. Hua, F. Chen, L. Zhao, C.-T. Lu, and N. Ramakrishnan. Sted: Semi-supervised targeted event detection. *KDD'13*, pages 11–14, 2013.
- [24] B. A. Huberman, D. M. Romero, and F. Wu. Social networks that matter: Twitter under the microscope. 2008.
- [25] D. Jurgens. That's what friends are for: Inferring location in online social media platforms based on social relationships. *ICWSM*, 13:273–282, 2013.
- [26] D. Jurgens, T. Finethy, J. McCorriston, Y. T. Xu, and D. Ruths. Geolocation prediction in twitter using social networks: A critical analysis and review of current practice. In *ICWSM*, pages 188–197, 2015.

- [27] V. Kharde, P. Sonawane, et al. Sentiment analysis of twitter data: A survey of techniques. *arXiv preprint arXiv:1601.06971*, 2016.
- [28] V. N. Khuc, C. Shivade, R. Ramnath, and J. Ramanathan. Towards building large-scale distributed systems for twitter sentiment analysis. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC '12*, pages 459–464, New York, NY, USA, 2012. ACM.
- [29] L. Kong, Z. Liu, and Y. Huang. Spot: Locating social media users based on social network context. *Proceedings of the VLDB Endowment*, 7(13):1681–1684, 2014.
- [30] C. Li, J. Weng, Q. He, Y. Yao, A. Datta, A. Sun, and B.-S. Lee. Twiner: named entity recognition in targeted twitter stream. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 721–730. ACM, 2012.
- [31] R. Li, S. Wang, and K. C.-C. Chang. Multiple location profiling for users and relationships from social network and content. *Proceedings of the VLDB Endowment*, 5(11):1603–1614, 2012.
- [32] R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang. Towards social user profiling: unified and discriminative influence model for inferring home locations. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1023–1031. ACM, 2012.
- [33] J. Mahmud, J. Nichols, and C. Drews. Where is this tweet from? inferring home locations of twitter users. *ICWSM*, 12:511–514, 2012.
- [34] J. McGee, J. Caverlee, and Z. Cheng. Location prediction in social media based on tie strength. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 459–468. ACM, 2013.
- [35] K. R. McKelvey and F. Menczer. Truthy: Enabling the study of online social networks. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 23–26. ACM, 2013.
- [36] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [37] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson. Semeval-2013 task 2: Sentiment analysis in twitter. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 312–320, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.

- [38] G. Paltoglou and M. Thelwall. Twitter, myspace, digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4):66, 2012.
- [39] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics, 2004.
- [40] S. Petrović, M. Osborne, and V. Lavrenko. Streaming first story detection with application to twitter. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 181–189. Association for Computational Linguistics, 2010.
- [41] D. Rout, K. Bontcheva, D. Preoțiuc-Pietro, and T. Cohn. Where’s@ wally?: a classification approach to geolocating users based on their social ties. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 11–20. ACM, 2013.
- [42] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *GIS*, 2009.
- [43] H. Sayyadi and L. Rschid. A graph analytical approach for fast topic detection.
- [44] A. Severyn and A. Moschitti. Unitn: Training deep convolutional neural network for twitter sentiment classification. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Association for Computational Linguistics, Denver, Colorado, pages 464–469, 2015.
- [45] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the conference on empirical methods in natural language processing*, pages 151–161. Association for Computational Linguistics, 2011.
- [46] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First Workshop on Unsupervised Learning in NLP, EMNLP ’11*, pages 53–63, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [47] C. Suen, S. Huang, C. Eksombatchai, R. Sasic, and J. Leskovec. Nifty: a system for large scale information flow tracking and clustering. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1237–1248. ACM, 2013.
- [48] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL (1)*, pages 1555–1565, 2014.

- [49] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang. Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 1031–1040, New York, NY, USA, 2011. ACM.
- [50] J. Weng and B.-S. Lee. Event detection in twitter. *ICWSM*, 11:401–408, 2011.
- [51] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu. Combining lexicon-based and learning-based methods for twitter sentiment analysis. 2011.