

# Constructing aspect-based sentiment lexicons with topic modeling

Tutubalina E., Nikolenko S.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

## Abstract

© Springer International Publishing AG 2017. We study topic models designed to be used for sentiment analysis, i.e., models that extract certain topics (aspects) from a corpus of documents and mine sentiment-related labels related to individual aspects. For both direct applications in sentiment analysis and other uses, it is desirable to have a good lexicon of sentiment words, preferably related to different aspects in the words. We have previously developed a modification for several popular sentiment-related LDA extensions that trains prior hyperparameters  $\beta$  for specific words. We continue this work and show how this approach leads to new aspect-specific lexicons of sentiment words based on a small set of “seed” sentiment words; the lexicons are useful by themselves and lead to improved sentiment classification.

[http://dx.doi.org/10.1007/978-3-319-52920-2\\_20](http://dx.doi.org/10.1007/978-3-319-52920-2_20)

## References

- [1] Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Sentiful: generating a reliable lexicon for sentiment analysis. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009, ACII 2009, pp. 1–6. IEEE (2009)
- [2] Tang, D., Wei, F., Qin, B., Zhou, M., Liu, T.: Building large-scale twitter-specific sentiment lexicon: a representation learning approach. In: Proceedings of COLING, pp. 172–182 (2014)
- [3] Severyn, A., Moschitti, A.: On the automatic learning of sentiment lexicons. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT 2015) (2015)
- [4] Kiritchenko, S., Zhu, X., Mohammad, S.M.: Sentiment analysis of short informal texts. *J. Artif. Intell. Res.* 50, 723–762 (2014)
- [5] Vorontsov, K., Frei, O., Apishev, M., Romov, P., Suvorova, M., Yanina, A.: Nonbayesian additive regularization for multimodal topic modeling of large collections. In: Proceedings of the 2015 Workshop on Topic Models: Post-Processing and Applications, TM 2015, New York, NY, USA, pp. 29–37. ACM (2015)
- [6] Liu, B.: *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, Cambridge (2015)
- [7] Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM (2004)
- [8] Chernyshevich, M.: Ihs r&d belarus: cross-domain extraction of product features using conditional random fields. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 309–313 (2014)
- [9] Ivanov, V., Tutubalina, E., Mingazov, N., Alimova, I.: Extracting aspects, sentiment and categories of aspects in user reviews about restaurants and cars. In: Proceedings of International Conference Dialog, pp. 22–34 (2015)
- [10] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: *Advances in Neural Information Processing Systems*, pp. 3111–3119 (2013)
- [11] Blinov, P.D., Kotelnikov, E.V.: Semantic similarity for aspect-based sentiment analysis. In: Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015, vol. 2, pp. 36–45 (2015)

- [12] Tarasov, D.S.: Deep recurrent neural networks for multiple language aspect-based sentiment analysis of user reviews. In: Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015, vol. 2 (2015)
- [13] Lin, C., He, Y., Everson, R., Ruger, S.: Weakly supervised joint sentiment-topic detection from text. *IEEE Trans. Knowl. Data Eng.* 24, 1134–1145 (2012)
- [14] Jo, Y., Oh, A.H.: Aspect and sentiment unification model for online review analysis. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM 2011, New York, NY, USA, pp. 815–824. ACM (2011)
- [15] Yang, Z., Kotov, A., Mohan, A., Lu, S.: Parametric and non-parametric user-aware sentiment topic models. In: Proceedings of the 38th ACM SIGIR (2015)
- [16] Lu, B., Ott, M., Cardie, C., Tsou, B.: Multi-aspect sentiment analysis with topic models. In: 2011 IEEE 11th International Conference Data Mining Workshops (ICDMW), pp. 81–88 (2011)
- [17] Kim, S., Zhang, J., Chen, Z., Oh, A.H., Liu, S.: A hierarchical aspect-sentiment model for online reviews. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, Bellevue, Washington, USA, 14–18 July 2013 (2013)
- [18] Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, pp. 174–181. ACL (1997)
- [19] Chetviorkin, I., Loukachevitch, N.V.: Extraction of Russian sentiment lexicon for product meta-domain. In: COLING, pp. 593–610. Citeseer (2012)
- [20] Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424. Association for Computational Linguistics (2002)
- [21] Tutubalina, E., Nikolenko, S.: Inferring sentiment-based priors in topic models. In: Lagunas, O.P., Alcántara, O.H., Figueroa, G.A. (eds.) MICAI 2015. LNCS (LNAI), vol. 9414, pp. 92–104. Springer, Heidelberg (2015). doi:10.1007/978-3-319-27101-9\_7
- [22] Al-Rfou, R., Perozzi, B., Skiena, S.: Polyglot: distributed word representations for multilingual NLP. In: Proceedings of the Seventeenth Conference on Computational Natural Language Learning, Sofia, Bulgaria, pp. 183–192. Association for Computational Linguistics (2013)
- [23] Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, pp. 1532–1543. Association for Computational Linguistics (2014)
- [24] Arefyev, N., Panchenko, A., Lukanin, A., Lesota, O., Romanov, P.: Evaluating three corpus-based semantic similarity systems for Russian. In: Proceedings of International Conference on Computational Linguistics Dialogue (2015)
- [25] Panchenko, A., Loukachevitch, N., Ustalov, D., Paperno, D., Meyer, C.M., Konstantinova, N.: Russe: the first workshop on russian semantic similarity. In: Proceedings of the International Conference on Computational Linguistics and Intellectual Technologies (Dialogue), pp. 89–105 (2015)
- [26] Loukachevitch, N., Blinov, P., Kotelnikov, E., Rubtsova Yu, V., Ivanov, V., Tutubalina, E.: Sentirueval: testing object-oriented sentiment analysis systems in Russian. In: Proceedings of International Conference Dialog, pp. 3–9 (2015)
- [27] Hagen, M., Potthast, M., Büchner, M., Stein, B.: Twitter sentiment detection via ensemble classification using averaged confidence scores. In: Hanbury, A., Kazai, G., Rauber, A., Fuhr, N. (eds.) ECIR 2015. LNCS, vol. 9022, pp. 741–754. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16354-3\_81