Abstract

Although, the fair amount of works in sentiment analysis (SA) and opinion mining (OM) systems in the last decade and with respect to the performance of these systems, but it still not desired performance, especially for morphologically-Rich Language (MRL) such as Arabic, due to the complexities and challenges exist in the nature of the languages itself. One of these challenges is the detection of idioms or proverbs phrases within the writer text or comment. An idiom or proverb is a form of speech or an expression that is peculiar to itself. Grammatically, it cannot be understood from the individual meanings of its elements and can yield different sentiment when treats as separate words. Consequently, In order to facilitate the task of detection and classification of lexical phrases for automated SA systems, this paper presents AIPSeLEX a novel idioms/ proverbs sentiment lexicon for modern standard Arabic (MSA) and colloquial. AIPSeLEX is manually collected and annotated at sentence level with semantic orientation (positive or negative). The efforts of manually building and annotating the lexicon are reported. Moreover, we build a classifier that extracts idioms and proverbs, phrases from text using n-gram and similarity measure methods. Finally, several experiments were carried out on various data, including Arabic tweets and Arabic microblogs (hotel reservation, product reviews,
and TV program comments) from publicly available Arabic online reviews websites (social media, blogs, forums, e-commerce web sites) to evaluate the coverage and accuracy of AIPSeLEX.

References

- X. Song and W. Ting, "Construction of unsupervised sentiment classifier on idioms resources," Journal of Central South University, vol. 21, pp. 1376-1384, 2014.
- V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in Proceedings of the Joint ACL / EACL Conference, 1997, pp. 174–181.
- P. Turney, "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 ACL ’02, Stroudsburg, PA, USA, 2002, pp. 417-424.
- B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79–86.
- M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2004, pp. 168–177.
- P. Alexander and P. Patrick, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC’10), European Language Resources Association (ELRA), Valletta, Malta, 2010.
- C. Scheible and H. Schütze, "Bootstrapping Sentiment Labels For Unannotated Documents With Polarity PageRank," in Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012), Istanbul-Turki, 2012.
- D. Davidiv, O. Tsur, and A. Rappoport, "Enhanced Sentiment Learning Using Twitter Hash-tags and Smiley-s," in Proceedings of the 23rd International Conference on Computational Linguistics (Coling2010), Beijing, China, 2010, pp. 241–249.
- L. Barbosa and J. Feng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data," in Proceedings of the 23rd International Conference on Computational Linguistics (Coling), 2010.
- M. Abdul-Mageed and M. Diab, "Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire," in Proceedings of the Fifth Law Workshop (LAW V), Association for Computational Linguistics, Portland, Oregon, 2011, pp. 110–118.
- M. Abdul-Mageed, M. Diab, and M. Korayem, "Subjectivity and sentiment analysis of modern standard Arabic," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011.
- A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in Collaboration Technologies and Systems (CTS) International Conference, Denver, CO, USA, 2012, pp. 546-550.
- A. Mourad and K. Darwish, "Subjectivity and Sentiment Analysis of Modern Standard Arabic and Arabic Microblogs," in Proceedings of the 4th Workshop on
Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA), Atlanta, Georgia, 2013, pp. 55–64.
- E. Refaee and V. Rieser, "An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis," in Proceedings of The 9th edition of the Language Resources and Evaluation Conference (LREC 2014), Reykjavik, Iceland, 2014.
- A. S. Zibin and A. R. M. Altakhaineh, "Informativity of Arabic Proverbs in Context: An Insight into Palestinian Discourse," International Journal of Linguistics (IJL), vol. 6, p. 67, 2014.
- Z. Kovecses, Metaphor: A Practical Introduction, First ed.: Oxford University Press, 2002.
- H. S. Ibrahim, S. M. Abdou, and M. Gheith, "Automatic expandable large-scale sentiment lexicon of Modern Standard Arabic and Colloquial," in 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING), Cairo - Egypt, 2015.
- O. F. Zaidan and C. Callison-Burch, "Arabic dialect identification," Computational Linguistics, vol. 40, pp. 171-202, March 2014 2012.
- M. Elmahdy, G. Rainer, M. Wolfgang, and A. Slim, "Survey on common Arabic language forms from a speech recognition point of view," in proceeding of International conference on Acoustics (NAG-DAGA), Rotterdam, Netherlands, 2009, pp. 63-66.
- B. Liu, Sentiment Analysis and Opinion Mining Morgan & Claypool Publishers, 2012.
- A. Basha, "??????? ????????: ?????? ?????? ?? ?????? ?? ?????? ?? ???? ?????? [Colloquial sayings: an annotated and arranged by the first letter of ideals with the Scout TOPTICAL]. Egypt: Al-Ahram Foundation - Al-Ahram Center for Translation and Publishing, 1986.
- A. Saalan, "?????? ??????? ??????? ??????? [Encyclopedia of Egyptian popular sayings]. First ed. Egypt: Dar-alafkalarabia press, 2003.
- F. Husain, "?????? ???????, ????? ???????, ??????? ???????, ????????? ??????? [Colloquial sayings, Egyptian folklore]. Egypt: General Egyptian Book Organization GEBO, 1984.
- G. Taher. (2006). "?????? ??????? ??????? - ????? ???? [Encyclopedia of public sayings - a scientific study]. Available: http://books.google.com.eg/books?id=2CR_EKTjxRgC

- PROz. (2014). PROz website for Arabic Idioms/Maxims/Sayings (Jan 2014). Available: http://www.proz.com/glossary-translations/
- J. C. Carletta, "Assessing agreement on classification tasks: the KAPPA statistic," Computational Linguistics, vol. 22, pp. 249-254, 1996.
- G. Salton, A. Wong, and C. S. Yang, "A vector space model for information retrieval," Communications of the ACM, vol. 11, pp. 613–620, November 1975.
- V. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," Soviet Physics Doklady, 10(8):707-710, 1966. Original in Russian in Doklady Akademii Nauk SSSR, vol. 163, pp. 845-848, 1965.
- X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in Proceeding WSDM &apos;08 Proceedings of the 2008 International Conference on Web Search and Data Mining, New York, NY, USA, 2008, pp. 231-240.
- A. Mudinas, D. Zhang, and M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis," in Proceeding WISDOM &apos;12
Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining Article, New York, NY, USA, 2012.

Index Terms

Computer Science

Information Sciences

Keywords

Sentiment lexicon  Idioms lexicon  AIPSeLEX.