Comparison of two opinion mining methods to classify opinions in online political discussions

Hugo Hammer[1]
Department of Computer Science
Oslo and Akershus University College of Applied Sciences

Per-Erik Solberg
Språkbanken
The National Library of Norway

Lilja Øvrelid
Department of Informatics
University of Oslo

Abstract

Bla bla

keywords:

1 Introduction

Over the past years there has been an alarming growth in hate against minorities like Muslims, Jews, Gypsies and gays in Europe, driven by right wing populism parties and extremist organizations (Fekete, 2013; Wilson and Hainsworth, 2013). A similar increase in hate speech has been observed on the Internet (Goodwin et al., 2013; Bartlett et al., 2013), and experts are concerned that individuals influenced by this web content may resort to violence as a result (Strømmen, 2012; Sunde, 2013). Hateful speech is not only observed on extremst sites, but also as comments on Twitter, YouTube or online newspaper articles.

Social media and online discussions contain a wealth of information which can make us able to understand the extent of hate speech on the Internet. However, it turns out that academia is lacking research on social media and

[1]hugo.hammer@hioa.no

online radicalization (Taylor, 2013). Opinion mining is the discipline of automatically extracting opinions from a text material and may be one important tool in the understanding online radicalization. Opinion mining have mostly been analyzed toward extracting opinions in comments and reviews of commercial products, but have also been used to analyze political tweets and discussions, see e.g. Tumasjan et al. (2010); Chen et al. (2010). Opinion mining towards political discussions is known to be hard since citations, irony and sarcasm is very common (Liu, 2012).

Opinion classification is perhaps the most studied topic within opinion mining. It aims to classify a set of text as either positive or negative and sometimes also neutral. There are mainly two approaches, one based on machine learning and one based on based on using a list of known sentiments words (lexical approach). One simple lexical approach is to count the number of words with positive and negative sentiment in the document as suggested by Hu and Liu (2004). One may classify the opinion of larger documents like movie or product reviews or smaller documents like tweets, comments or sentences. See Liu (2012), chapters three to five and references therein for several methods to do opinion classification.

In this paper we focus on classifying the opinion of sentences and use the lexical-based approach. Suppose we want to classify the opinion toward a keyword, say 'immigration', in a sentence containing the keyword. One intuitive approach is to also find the the words with sentiment in the sentence and classify the sentiment of the sentence based on the polarity of these sentiment words. We expect that sentiment words that is close or related to the keyword in some sense is more important than other sentiment words in the sentence as suggested by Ding et al. (2008). An other approach is to use parsing to develop dependence paths between keywords and sentiment words, see e.g. Jiang et al. (2011). The aim of this paper is to compare the method suggested in Ding et al. (2008) with a suggested method based on dependence paths to classify opinions in political discussions.

The paper is organized as follows.

2 Opinion mining methods

2.1 Word distance method

2.2 Dependence paths method

2.3 Measuring differences in opinion mining methods

3 Results

4 Closing remarks

References

Bartlett, J., Birdwell, J., and Littler, M. (2013). The rise of populism in Europe can be traced through online behaviour... Demos, `http://www.demos.co.uk/files/Demos_OSIPOP_Book-web_03.pdf?1320601634`. [Online; accessed 21-January-2014].

Chen, B., Zhu, L., Kifer, D., and Lee, D. (2010). What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model. In *AAAI*.

Ding, X., Liu, B., and Yu, P. S. (2008). A Holistic Lexicon-based Approach to Opinion Mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, pages 231–240, New York, NY, USA. ACM.

Fekete, L. (2013). Pedlars of hate: The violent impact of the European far Right. Institute of Race Relations, `http://www.irr.org.uk/wp-content/uploads/2012/06/PedlarsofHate.pdf`. [Online; accessed 21-January-2014].

Goodwin, M., Ramalingam, V., and Briggs, R. (2013). The New Radical Right: Violent and Non-Violent Movements in Europe. Institute for Strategic Dialogue, `http://www.strategicdialogue.org/ISD%20Far%20Right%20Feb2012.pdf`. [Online; accessed 21-January-2014].

Hu, M. and Liu, B. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Jiang, L., Yu, M., Zhou, M., Liu, X., and Zhao, T. (2011). Target-dependent Twitter Sentiment Classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 151–160, Stroudsburg, PA, USA. Association for Computational Linguistics.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Strømmen, Ø. (2012). *The Dark Net. On Right-Wing Extremism, Counter-Jihadism and Terror in Europe*. Cappelen Damm.

Sunde, I. M. (2013). Preventing radicalization and violent extremism on the Internet (Norwegian). The Norwegian Police University College 2013:1.

Taylor, H. (2013). Social Media for Social Change. Using the Internet to Tackle Intolerance. Institute for Strategic Dialogue, `http://tsforum.event123.no/UD/rehc2013/pop.cfm?FuseAction=Doc&pAction=View&pDocumentId=46414`. [Online; accessed 21-January-2014].

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the fourth international aaai conference on weblogs and social media*, pages 178–185.

Wilson, R. and Hainsworth, P. (2013). Far-right Parties and discourse in Europe: A challenge for our times. European network against racism, `http://cms.horus.be/files/99935/MediaArchive/publications/20060_Publication_Far_right_EN_LR.pdf`. [Online; accessed 21-January-2014].