Application of Interactive Computer-Assisted Argument Extraction to Opinionated Social Media Texts

Kostiantyn Kucher
Linnaeus University
Department of Computer Science and Media Technology
Växjö, Sweden
kostiantyn.kucher@lnu.se

Maria Skeppstedt
Linnaeus University
Department of Computer Science and Media Technology
Växjö, Sweden
University of Potsdam
Applied Computational Linguistics
Potsdam, Germany
maria.skeppstedt@lnu.se

Andreas Kerren
Linnaeus University
Department of Computer Science and Media Technology
Växjö, Sweden
andreas.kerren@lnu.se

ABSTRACT

The analysis of various opinions and arguments in textual data can be facilitated by automatic topic modeling methods; however, the exploration and interpretation of the resulting topics and terms may prove to be difficult to the analysts. Opinions, stances, arguments, topics, terms, and text documents are usually connected with many-to-many relationships for such tasks. Exploratory visual analysis with interactive tools can help the analysts to get an overview of the topics and opinions, identify particularly interesting documents, and describe main themes of various arguments. In our previous work, we introduced an interactive tool called Topics2Themes that was used for topic and theme analysis of vaccination-related discussion texts with a limited set of stance categories. In this poster paper, we describe an application of Topics2Themes to a different genre of data, namely, political comments from Reddit, and multiple sentiment and stance categories detected with automatic classifiers.

CCS CONCEPTS

• **Human-centered computing** → **Visual analytics**; • **Computing methodologies** → **Topic modeling**; • **Applied computing** → *Annotation*;

KEYWORDS

visualization, interaction, topic modeling, argument extraction, text visualization, sentiment analysis, sentiment visualization, stance analysis, stance visualization, annotation

ACM Reference Format:
Kostiantyn Kucher, Maria Skeppstedt, and Andreas Kerren. 2018. Application of Interactive Computer-Assisted Argument Extraction to Opinionated Social Media Texts. In *VINCI '18: The 11th International Symposium on Visual Information Communication and Interaction, August 13–15, 2018, Växjö, Sweden.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3231622.3232505

VINCI '18, August 13–15, 2018, Växjö, Sweden

ACM ISBN 978-1-4503-6501-7/18/08.
https://doi.org/10.1145/3231622.3232505

1 INTRODUCTION

The task of qualitative text analysis, in particular, the identification of arguments and themes, requires a lot of effort from the analyst. Computational extraction of main topics in a document or a corpus has been shown to be an effective first step for such analyses [1, 9]. However, the typical output of topic modeling algorithms at the detailed level is also overwhelming. The fields of information visualization and visual analytics provide approaches for representing and interacting with textual data and results of various text analyses (including topic modeling [2] and sentiment & stance analysis [5]) to solve this problem.

In our previous work, we have introduced an interactive visualization tool, called Topics2Themes [7], that is used to assist the task of extraction and annotation of arguments in texts by providing a Jigsaw-like list interface [3, 10]. Topics2Themes was primarily designed to support analyses of vaccination-related texts with a limited number of opinions or stances towards this issue, such as *for*, *undecided*, or *against*. In this poster paper, we describe an application of a customized version of Topics2Themes to a different genre of data (political comments from Reddit) and a different set of supported sentiment and stance analyses (with multiple categories).

2 METHOD

The data processing pipeline of Topics2Themes [6] includes the following steps: 1) optional classification or manual tagging of stances associated with text documents; 2) preprocessing including stop word removal, collocation detection, and clustering of semantically similar words; and 3) topic modeling with either the LDA or NMF algorithm. For the present work, we have customized Topics2Themes to use the classifiers developed as part of the StaViCTA project [8] for the first step. Then, we applied the tool to a data set of about 200 political comments from Reddit created during spring 2018. Each document was automatically labeled with its dominant sentiment category (*positive*, *neutral*, or *negative*) by the VADER sentiment classifier [4] and a set of detected stance categories, such as *certainty* or *contrast*, by our custom stance classifier [8]. Using the frontend of Topics2Themes displayed in Figure 1, we were able to select an interesting topic on Internet neutrality among the output of the NMF algorithm (see Figure 1(c)). By reading the related documents and identifying the recurring themes (see Figure 1(d–e)), we established the main arguments in the ongoing discussion about

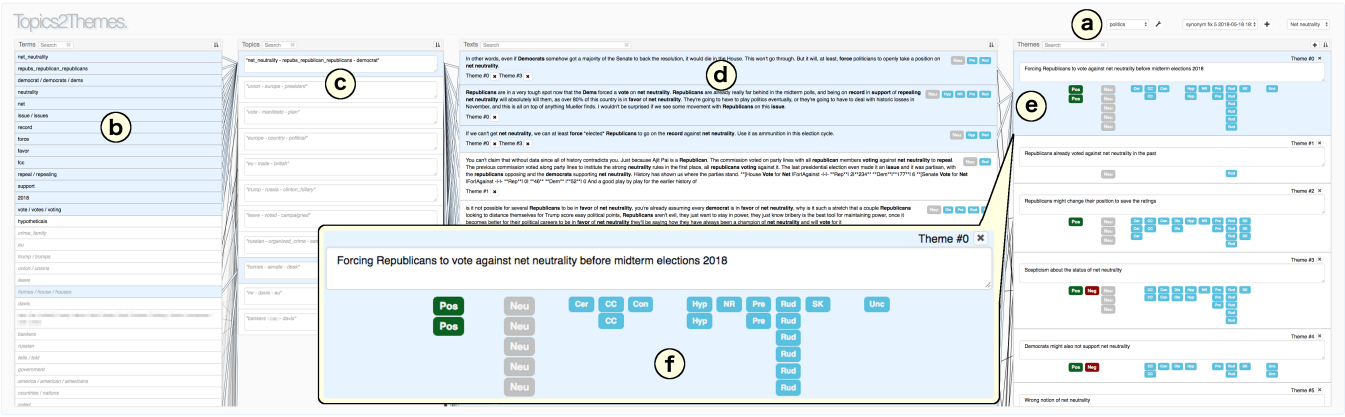

Figure 1: The screenshot of Topics2Themes customized for multiple sentiment & stance categories discovered in political comments from Reddit: (a) data loading controls; (b) list of topic terms; (c) list of topics (the first one is selected); (d) list of documents (displaying detected sentiments & stances and associated themes); (e) list of user-defined themes (displaying sentiments & stances for all associated documents); and (f) zoomed-in cutout of the first theme in (e). Here, the user hovered over the first theme displayed in (e), which affected highlighting of elements and links in other lists (in blue color).

Table 1: Classification categories

Sentiment:	Positive, Neutral, Negative
Stance:	Agreement, Certainty, Concession & Contrariness, Contrast, Disagreement, Hypotheticals, Need & Requirement, Prediction, Rudeness, Source of Knowledge, Tact, Uncertainty

the upcoming U.S. Senate vote on Internet neutrality[1]. The resulting user-labeled themes displayed in Figure 1(e–f) also provide a barchart-like overview of various sentiment and stance categories discovered in the associated documents (see Table 1), thus providing us with an opportunity to compare the opinions related to the themes.

3 CONCLUSIONS AND FUTURE WORK

In this poster paper, we briefly demonstrated the potential applications of our interactive tool Topics2Themes to political texts from social media. Topics2Themes allows the users to visually explore the output of topic modeling and stance classification algorithms, conduct close reading of the original texts, and annotate arguments for various viewpoints by defining recurring themes. Our future work includes collaboration with domain experts, evaluation of our proposed approach, and integration of Topics2Themes into larger visual stance analysis workflows.

ACKNOWLEDGMENTS

This work was partially funded by the project "Navigating in streams of opinions: Extracting and visualising arguments in opinionated texts" from the Swedish Research Council (Vetenskapsrådet) [grant number 2016-06681].

REFERENCES

[1] Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence? *Journal of the Association for Information Science and Technology* 68, 6 (June 2017), 1397–1410. https://doi.org/10.1002/asi.23786

[2] Wenwen Dou and Shixia Liu. 2016. Topic- and Time-Oriented Visual Text Analysis. *IEEE Computer Graphics and Applications* 36, 4 (July 2016), 8–13. https://doi.org/10.1109/MCG.2016.73

[3] Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Haesun Park, and John Stasko. 2013. Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw. *IEEE Transactions on Visualization and Computer Graphics* 19, 10 (Oct. 2013), 1646–1663. https://doi.org/10.1109/TVCG.2012.324

[4] C.J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM '14)*. https://aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109

[5] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. 2018. The State of the Art in Sentiment Visualization. *Computer Graphics Forum* 37, 1 (Feb. 2018), 71–96. https://doi.org/10.1111/cgf.13217

[6] Maria Skeppstedt, Andreas Kerren, and Manfred Stede. 2018. Vaccine Hesitancy in Discussion Forums: Computer-Assisted Argument Mining with Topic Models. *Studies in Health Technology and Informatics* 247 (2018), 366–370. https://doi.org/10.3233/978-1-61499-852-5-366 . Proceedings of the 29th Medical Informatics Europe Conference (MIE '18).

[7] Maria Skeppstedt, Kostiantyn Kucher, Manfred Stede, and Andreas Kerren. 2018. Topics2Themes: Computer–Assisted Argument Extraction by Visual Analysis of Important Topics. In *Proceedings of the 3rd Workshop on Visualization as Added Value in the Development, Use and Evaluation of Language Resources at LREC '18 (VisLR III)*. European Language Resources Association (ELRA), Paris, France. http://lrec-conf.org/workshops/lrec2018/W16/summaries/2_W16.html

[8] Maria Skeppstedt, Vasiliki Simaki, Carita Paradis, and Andreas Kerren. 2017. Detection of Stance and Sentiment Modifiers in Political Blogs. In *Proceedings of the International Conference on Speech and Computer (SPECOM '17)*. Springer, 302–311. https://doi.org/10.1007/978-3-319-66429-3_29

[9] Parinaz Sobhani, Diana Inkpen, and Stan Matwin. 2015. From Argumentation Mining to Stance Classification. In *Proceedings of the 2nd Workshop on Argumentation Mining*. Association for Computational Linguistics, Denver, CO, 67–77. http://aclweb.org/anthology/W15-0509

[10] John Stasko, Carsten Görg, and Zhicheng Liu. 2008. Jigsaw: Supporting Investigative Analysis through Interactive Visualization. *Information Visualization* 7, 2 (2008), 118–132. https://doi.org/10.1057/palgrave.ivs.9500180

[1] The vote eventually took place on May 16, 2018: https://edition.cnn.com/2018/05/16/politics/net-neutrality-vote-senate-democrats/