Use of Fuzzy Ontology for Opinion Mining

Natalya S. Kozhevnikova[1], Yekaterina Yu. Danilova[2]
Assistant of the department
Perm State University
Perm, Russia
[1]kozhevnikovans@gmail.com, [2], ket-eref@yandex.ru

***Abstract*— One of the key parameters in marketing policy selection is users opinions. Number of user feedback in social media grows with increase of the popularity of these services. Users share their experience with products and services in a real time and mentions monitoring as well as timely response became a standard of modern marketing. Meanwhile, the task of opinion mining in social media has several problems, such as finding the most influential opinions, evaluating is the review positive or negative, extraction and summarization of the key features in opinion. In our project, we propose hybrid approach based on fuzzy ontologies and lexico-syntactic patterns, while additional consideration is given to social media metrics that used for opinion ranking and reliability estimation.**

Keywords*— *fuzzy ontology; text analysis; natural language; opinion mining

I. INTRODUCTION

This template, modified in MS Word 2007 and saved as a "Word 97-2003 Document" for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

II. OVERVIEW

There are several commercial products aimed to aggregate data from social networks. For example, Klear [1], Brandwatch [2], Traackr [3] are competing products that provide aggregation services for brands, agencies and enterprises. However, they are mainly focused on improvement of the social networks marketing. Text analytics is presented only by topics clustering without detailing. This products use social networks metrics analyses to detect opinion influencers. However, the detailed text analysis of the selected user reviews and capturing of the main points, which caused the greatest resonance is still made by human. Mentionmapp [4] and Tweetreach [5] are other similar products and they provide a service only for Twitter. These services also concentrate on the relationship of tweets via hashtags, themes and retweets, but have no deeper analysis. It is important to note that no presented tools support the Russian language, so Russian-language content analysis is unavailable.

The scientific community is also engaged in the task of social networks information retrieval. The opinion retrieval problem is a part of broader class of information retrieval tasks. Therefore, the approaches used for these problem solving are often the same, e.g. linguistic text analysis and machine learning methods [6]. However, researchers tackle additional challenges during user review retrieval and summarization. For example, selection of the key features reflected in the opinion, sentiment polarity and its degree assessment, the problem of the relatively short texts analysis, etc. [7]. In our review we have found that in the most works opinions are assessed either unambiguously positive or negative. Moreover, in founded approaches predefined templates or specifiers are used, which does not allow adjusting the algorithm to the problem area specifics. Thus, one of the related works is the study of Hu M. and Liu B. [8], which proposed a three-stage approach to the opinion retrieval and summarization, is based on the phrase frequency analysis. Another interesting method is the Opizer-E [9], which extracts and summarizes the main features in opinion according to sentence position and proximity to the predefined qualifiers. In [10] the authors improved this method. However, the lack of the review context and its relatively short size affects the quality of the analysis for non-standard reviews like products comparisons. Agichtein et al. [11] note that social ranking is an important tool of text quality assessment. However, we noticed that work in a field of opinion mining from the social media and its summarization do not consider this data. Therefore, we proposed a hybrid approach based on fuzzy ontologies and linguistic analysis. This allows to group facts using common knowledge on the one hand, and to take into the account social feedback on the other hand. For this purpose, confidence levels are calculated from the social network metrics. At the same time, we are focusing on the analysis of Russian-language content in this project.

III. A PROPOSED APPROACH

In our approach, we propose to extract concepts from the texts in social networks by context-free grammars and to build the ontology using these concepts. Retrieval of different text parts (key words and phrases): abbreviations, proper nouns, quotations is implemented by Tomita-parser[12] with the use

978-1-5386-5824-6/18/$31.00 ©2018 IEEE

of lexico-syntactic patterns. The parser consists of three standard linguistic processors: a tokenizer (word splitting), a segmenter (splitting into sentences), and a morphological analyzer (mystem). The main components of the parser are a gazetteer, a set of context-free grammars and a set of the facts types that could be generated during the interpretation. The facts are extracted using the Tomita-parser facts descriptions. User can extend the system with new patterns if needed.

All the patterns are split by the domain in order to adjust the system for the specifics of the text processing. Implementation of this function is based on Tomita-parser gazetteers. The extracted key phrases are stored in the database. Texts are separated by the topics using the classifier from the Zamgi project [13]. Its algorithm brings text to one of the 13 predefined classes. The keywords and phrases are extracted from the text and then used as concepts of the ontology. This ontology of the review is mapped on an external subject ontology – the open Wikidata project [14]. The mapping algorithm uses the Wikidata API, which provides querying concept by name that is extracted from the keywords and phrases. Use of the Wikidata ontology helps to take into account knowledge that were implied by user, but not explicitly mentioned in a review. The external ontology mapping facilitates the mapping of review ontologies to each other; thereby it is easier to find similar views. In addition, mapping result are used to extend Tomita-parser gazetteers with possible synonyms, increasing the quality of further extractions of keywords and phrases.

Since different user reviews can be contradictory, as different people's opinions, their ontologies should reflect the knowledge fuzziness. There are several approaches to fuzzy ontology creation and their main difference is in which knowledge aspect is fuzzified [15]. To define fuzzy ontology, we will use OWL 2 language, similarly to the Bobillo and Straccia approach [16]. Extracted relation's confidence levels will be set to theirs data properties. The fuzzy ontology building methodology is described in detail in [17]. We rely on social ranking to calculate confidence levels. Thus the "engagement rate" metric

$$\frac{likes + saves + comments}{followers}$$

reflects the level of user's opinion popularity and, consequently, trust level to the facts in the text which corresponds with the message rate level. At the same time, comment analysis, helps to correct the confidence level if it denies the facts from the starting publication. Confidence levels correction will be made as proposed in [18]:

$$f = f + \frac{f_{new} - f}{Q + 1}, \quad (1)$$

where Q is the number of previous changes. Weighted ontology with summarized facts is obtained as a result of the user query processing. Further work with that ontology could be: visualization of the extracted information, uploading of the retrieved facts to the database and traditional DataMining methods usage, working with fuzzy ontology like with an expert system and testing business hypotheses, etc.

IV. Case study

Consider an example of proposed approach to representation of knowledge extracted from social media. We will analyze an opinion about some property: "We like this building. Beautiful architecture and interior. It was nice to communicate with developer representative. He was polite and informative. Yesterday we visited construction site, road building is in progress". Suppose this post has 135 likes and 1 repost. It was published in a group with 500 followers "Building 'Willow' is a best place to live in".

Firstly, key words and collocations are extracted. Thus, concept "building" will be mapped to Wikidata and similar concepts: "liquid-crystal" and "building" (in Russian language) would be found. Text topic would be clustered as "real estate". Concept "building" in Wikidata has attribute category with value "real property", which has main category as a "real estate". Thus using topics clustering and metaontology mapping, we will find concepts that have multiple meaning.

Links and mentions in social media provide a context for text analysis. Thus, group name of the post "Building 'Willow' is a best place to live in" helps to concretize discussed building, because "building "willow" is an instance of "building". After the key word analysis based on Wikidata knowledge, gazetteers are formed. The next step is context free grammar analysis of the text. In or example the result would contain the following facts: like building, nice architecture, nice interior, nice to communicate, representative was polite, representative was informative, road building in progress.

The next step is to calculate confidence levels using collected metrics. Confidence level for example post is 0.272. During the repost analysis, some additional comment could be found. For example, "Disagree. This building is definitely ugly" with one like. Founded negation forces recalculation of confidence levels of «nice architecture" fact. Calculation of likes and comments count, as well as number of friends give confidence level to this negative post. Let it be 0.002. Then the post confidence level would become 0.27. Using formula 1, we obtain 0.271. Next, consider that there is another post with "nice architecture" fact having confidence level equal to 0.45. Thus, these concepts would be mapped, found related and the confidence level of that fact would be changed to 0.331. As a result, ontology will contain two opposite facts, but with different confidence levels. This approach allows consider even minimal contradictions in analysis, which helps improve marketing decisions. However, fuzzy logic usage helps to keep ontology consistent.

V. Conclusion

In this article, we considered the existing approaches to opinion mining from social networks. The shortcomings of existing approaches were noted, such as the complexity of non-standard opinion analysis, when some knowledge is not explicitly presented in the text. In addition, binary assessment of opinion and insufficient consideration of the social metrics associated with the opinion are weak sides of founded projects. In order to solve these problems we proposed an approach based on fuzzy ontologies that uses social media metrics to estimate the confidence levels.

It is important to note that the proposed approach could be extended to the other data mining tasks, in addition to the opinion mining. Furthermore, the ontological approach enables user to customize the system and obtain more accurate results, by expanding the ontology of Wikidata, the addition of Wikipedia. Thus, it allows on one hand to take advantage of crowdsourcing of semantics, and on the other, avoid the "cold start", a common problem of such products.

Further research would be devoted to the quality evaluation of the obtained confidence levels, as well as the development of the fuzzy ontology analysis tools.

REFERENCES

[1] Klear: Influencer Marketing Software (2018). Available at: https://klear.com/ (accessed 04 April 2018) (in Russian)

[2] Brandwatch: Home (2018) Available at: https://www.brandwatch.com/ (accessed 04 April 2018)

[3] Traackr: Influencer Marketing that matters (2018) Available at: http://www.traackr.com/ (accessed 04 April 2018)

[4] Mentionmapp (2018) Available at: http://mentionmapp.com/ (accessed 04 April 2018)

[5] TweetReach: How Far Did Your Tweets Travel? (2018) Available at: https://tweetreach.com/ (accessed 04 April 2018)

[6] Paltoglou G., Giachanou A. Opinion retrieval: Searching for opinions in social media. Professional Search in the Modern World. Springer. 2014, pp. 193-214.

[7] Pang B., Lee L. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval. 2008, vol. 2, pp. 1-135.

[8] Hu M., Liu B. Mining and Summarizing Customer Reviews. Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '04). 2004, pp. 168-177.

[9] R.E.L. Condori, Pardo T.A.S. Opinion summarization methods: Comparing and extending extractive and abstractive approaches. Expert Systems with Applications. 2017, vol. 78, pp. 124-134.

[10] Anchieta R., Sousa R.F., Moura R., Pardo T. Improving Opinion Summarization by Assessing Sentence Importance in On-line Reviews. Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology. 2017, pp.32-36.

[11] Agichtein E., Castillo C., Donato D., Gionis A., Mishne G. Finding high-quality content in social media. Proc. of the Intl. Conf. on Web search and web data mining WSDM '08. 2008, pp. 183-194.

[12] Tomita-parser - Yandex technologies. (2018) Available at: https://tech.yandex.ru/tomita (accessed 04.04.2018)

[13] Zamgi: auto classifier of the Russian texts (2018) Available at: https://github.com/zamgi/lingvo--classify (accessed 04.04.2018)

[14] Wikidata (2018) Available at: https://www.wikidata.org/wiki/Wikidata:Main_Page (accessed 04.04.2018)

[15] El-Sappagh S., Elmog M. A fuzzy ontology modeling for case base knowledge in diabetes mellitus domain. Engineering Science and Technology, an International Journal. 2017, vol. 20, pp. 1025–1040.

[16] Bobillo F., Straccia U. Fuzzy ontology representation using OWL 2. Int. J. Approx. Reason. Oct. 2011, vol. 52, no. 7, pp. 1073-1094.

[17] Alexopoulos P., Wallace M., Kafentzis K., Askounis D. IKARUS-Onto: A Methodology to Develop Fuzzy Ontologies from Crisp Ones. Knowledge Information Systems. Sep. 2012, vol. 32, no. 3, pp. 667-695.

[18] Calegari S., Ciucci D. Integrating Fuzzy Logic In Ontologies. Proceedings of the 8th International Conference on Enterprise Information Systems. 2006, pp 66–73.