Text Tone Determination Using Fuzzy Logic

Igor Olenych1*, Oleh Sinkevych2, Maryana Salamakha3, Marianna Prytula4
1–4Ivan Franko Lviv National University, Lviv, Ukraine

Abstract – The study proposes the text tone detection system based on sentiment dictionaries and fuzzy rules. Computer analysis of texts from different sources has been performed in emotional categories: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. A synonym dictionary has been used to expand the vocabulary. To increase the accuracy and validity of sentiment analysis, the authors of the study have used coefficients that take into account different emotional loads of words of various parts of speech and the action of intensifying or softening adverbs. A quantitative value of the text tone has been obtained as a result of an aggregation of normalized data on all emotional categories by the fuzzy inference methods. It has been found that emotional words have a greater impact on the text tone value in the case of analysis of short messages. The proposed approach makes it possible to contribute to all emotional categories in the final text evaluation.

Keywords – Emotion detection, fuzzy rules, sentiment dictionary, text tone.

I. INTRODUCTION

The development of information technology (IT) has led to a significant increase in the amount of information to be stored, processed, and transmitted by means of computer systems and networks. The text is one of the main forms of information exchange in society. In today’s world, where the exchange of text messages takes place via all possible channels of communication, it is important to quickly analyze the transmitted information. Particular attention is paid to the problem of analyzing the opinions of Internet users. Such an analysis is based on the search and subsequent recognition of words with emotional load in the text. Through emotion detection from the textual information, commercial companies are able to track the needs of users and respond in a timely manner to their feedback on products and services [1]–[3]. Based on the sentiment analysis, the reaction to advertising is monitored in real time, which makes it possible to prevent dangerous situations in society [4]–[6]. The use of IT tools helps not only to understand collective behavior and assess the mood of public response to marketing activities but also to influence social systems and develop the principles of public policy in various fields [7], [8]. Given the significant potential of practical application, the creation and development of text processing technologies is a topical task at all stages of the development of information systems.

Sentiment analysis of the text like many other problems of natural language processing can be considered a classification problem, which usually solves two tasks: classification of subjectivity (because an important feature of thoughts and their tone is subjectivity) and text classification as an expression of a positive, negative or neutral thought known as tone classification. Besides, information systems often focus on the expression of feelings and emotions (anger, happiness, sadness, etc.) or reveal the intentions of the narrator (for example, interested or disinterested). In order to take into account subjective factors and approximate information inherent in the expression of human emotions, it is appropriate to use methods and approaches of fuzzy logic in the process of sentiment analysis and determination of textual information tone. In particular, the combination of emoji classification, hashtag classification, and text polarity in the model of the microblog sentiment analysis by fuzzy logic was proposed in [9]. Fuzzy modeling was also used to increase the interpretability of the film sentiment analysis based on subjective assessments of the plotline and expectations from the film [10], [11]. Fuzzy logic formalizes a person’s ability to approximate reasoning and allows for non-binary values of the truth of fuzzy statements [12], [13].

There are a number of methods and algorithms for implementing systems of emotional text analysis, which represent two main approaches to the automatic classification of textual information. In particular, these are methods based on machine learning technologies [14]–[16], and methods based on dictionaries and rules [17]–[19]. The most popular methods are machine learning methods with or without a teacher, namely, the naive Bayesian classifier, vector support machines and deep learning algorithms that try to simulate the work of the human brain using artificial neural networks to process data. All these methods differ in accuracy and speed. According to another approach, sentiment analysis is carried out by finding and summing words with emotional load in the text by means of pre-constructed tonal dictionaries and rules. Based on the found emotional vocabulary, the text can be evaluated on a scale that contains values for negative and positive words. The advantage of this approach is the ability of sentiment analysis implementation at the level of individual sentences.

In this study, we have created the system of emotion detection from textual information based on tonal dictionaries and rules. The proposed system uses dictionaries that contain

* Corresponding author’s e-mail: igor.olenych@lnu.edu.ua
©2021 Igor Olenych, Oleh Sinkevych, Maryana Salamakha, Marianna Prytula.
This is an open access article licensed under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), in the manner agreed with Sciendo.
several emotional categories. Particular attention is paid to studying ways of more correct emotion detection and finding the quantitative value of the text tone using methods of fuzzy modeling.

II. DESCRIPTION OF EMOTION DETECTION METHOD

In order to correctly identify the emotional tone of the text, the technique using a fuzzy inference system is proposed. The developed method is schematically shown in Fig. 1.

The automatic definition of text tone takes place at several stages. First, the tokenization procedure is applied to the downloaded text. Textual information is divided into simple units (tokens), i.e., meaningful groups of symbols that correspond to certain patterns (e.g., words). Typically, tokenization is used as an initial and fundamental step in many data analysis tasks. As a result, a word array is obtained, which after excluding punctuation and stop-words is used for sentiment analysis.

The next stage of the algorithm is to sequentially search for each word from the resulting array in the sentiment dictionaries [20], [21]. Sentiment lexicons used in the study contain a number of emotional categories: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. In these dictionaries, words are related to some category as a whole (affiliation is denoted by one), or their degree of affiliation by a fractional number from zero to one is specified. If the affiliation degree of the chosen word to some category is different in various sentiment dictionaries, then it is assigned the value obtained by the max or min aggregation of emotional evaluations in varied lexicons. In order to expand the vocabulary, WordNet dictionary of synonyms is used. If there is no word in the sentiment lexicons, the emotional weight of its nearest synonym is used for further calculations.

Besides, the proposed method provides the ability to take into account the emotional load of different speech parts to increase the accuracy and validity of sentiment analysis of textual information. Since different parts of speech have different effects on the overall assessment of the text tone, a corrective factor is applied to the obtained values of the affiliation degree of words to some emotion category. In particular, the coefficient is equal to one for adjectives that convey the main emotional load. Verbs and nouns are assigned corrective coefficient values in the range of 0.6–0.8 and 0.4–0.6, respectively. The applied coefficient decreases the weight of words that have less emotional load and, accordingly, have less effect on the results of sentiment analysis of the whole text.

Another corrective factor is related to words that increase or decrease the emotional load of the following words in a sentence. For example, such adverbs as very, strongly, highly, extremely, exceptionally, incredibly, amazingly, absolutely, utterly, awfully, frantically, unusually, etc. cause an increase in the emotional weight of the following words. In this case, the value of the coefficient is equal to two. Otherwise, slightly, partially, scantily, quite, ridiculously, and other similar adverbs cause a decrease in the emotional weight of related words. Then it is appropriate to assign a coefficient of 0.5. The optimal values of the correction factors can be selected in the process of setting up the sentiment analysis system.

The final stages of the algorithm include procedures for summarizing the emotional weight of words in each of the categories and their normalization as well as the determination of text tone. In order to determine not only qualitative but also a quantitative assessment of the emotional tone of the text, a fuzzy inference system is used.

III. DESIGN OF FUZZY INFERENCE SYSTEM

It should be noted that in the simplest case of qualitative analysis, the sentiment evaluation of the text is determined by the emotional category, which contains the largest word number or is characterized by the largest total value of the emotional weight of words in this category [22]. However, this approach does not take into account the contribution of other emotional categories, which can sometimes be quite significant although not decisive. Therefore, the algorithm of fuzzy inference is implemented to aggregate the obtained data by different emotion categories and determine the quantitative value of the text tone. According to the proposed method, an expert system of the fuzzy inference is developed with one input and one output. The terms of the “emotion” input linguistic variable are associated with eight emotion categories: “anger”, “disgust”, “fear”, “sadness”, “anticipation”, “surprise”, “trust”, and “joy”. As the normalized values of the affiliation degree of the text to eight emotional categories are used, the fuzzification procedure of the input variable is missing. Membership functions of fuzzy sets that characterize the terms of the “text tone” output linguistic variable are set by piecewise-linear functions, as shown in Fig. 2.

The mechanism for obtaining a fuzzy inference is usually based on expert opinions, which are presented in the form of fuzzy production rules $R^{(n)}$:
Very characterized by tone values, the use of correction factors makes it possible to respond to terms of the output variable. The procedure of accumulation by max
production rules, the truthfulness degree of conditions of which, respectively. The aggregation of the sub-conditions truthfulness in each of the fuzzy rules is carried out by the fuzzy conjunction operation.

The verity of the conclusions of each fuzzy rule is determined using the min-activation of conclusion procedure. Only active production rules, the truthfulness degree of conditions of which is different from zero, are taken into account.

The activated conclusions of production rules are accumulated by max-disjunction of obtained fuzzy sets that respond to terms of the output variable. The procedure of defuzzification of the output variable is carried out by the Centre of Gravity method [12]. As a result, the quantitative value of the emotional tone of the text in the range from −1 to 1 is obtained.

IV. RESULTS AND DISCUSSION

Testing the proposed sentiment analysis system has been carried out using texts from different sources and various emotional tones. In particular, art literature fragments with a negative and positive emotional tone and scientific articles in a neutral tone have been used. Messages, feedback and other social media content have also been analyzed. For example, the proposed sentiment analysis system can be adapted to detect emotions in texts from different sources and various emotional tones.

If the truthfulness of the rules $R^{(1)}$–$R^{(8)}$ conditions prevail, then the calculated value of the emotional tone is in the $[-1, 0]$ range, as shown in Fig. 3a, b. Fig. 3c, d illustrates the case when the truthfulness of the conditions of the rules $R^{(10)}$–$R^{(15)}$ prevails and the tone value is within $[0, 1]$. Moreover, texts with more expressive emotions are characterized by tone values approaching the limits of the $[-1, 1]$ range. In cases where the truthfulness of the rule $R^{(9)}$ condition prevails or the verity of the negative and positive conclusions of the fuzzy rules is almost the same, the value of the text tone is close to zero (see Fig. 3e, f). Usually, such results are obtained when analyzing scientific information, which is completely devoid of emotional load.

The study of the influence of the applied correction factors on the sentiment analysis results has been carried out on the basis of textual information of various formats. Both large fragments of texts and short text messages have been used. As can be seen in Table I, the influence of emotional words increases in the case of sentiment analysis of individual sentences or short messages. It is expressed in larger modulo quantitative values of text tone. This is likely due to the greater ratio of emotional words to their total number in short texts, which causes less fuzzification of emotion evaluation.

The various emotional weights of adjectives, verbs and nouns have a different effect on the whole text tone value by slightly increasing or decreasing it. Presumably, the resulting value depends on which part of speech outweighs and forms the dominant emotional category. The coefficient associated with intensifying or softening adverbs also does not significantly change the assessment of all text because it adjusts the emotional weight of only the category that contains the related words. Nevertheless, the use of correction factors makes it possible to increase the validity of sentiment analysis of the texts.

Based on the proposed approach, the sentiment analysis system can be adapted to detect emotions in texts from different sources by a simple correction factor setup procedure. The system sensitivity to different emotion categories can be additionally adjusted using the weight coefficients applied to each of the fuzzy production rules.

An important evaluation metric of the text tone definition method is its accuracy. The accuracy of the proposed method has been determined by means of the F-score. A dataset of preprocessed Twitter tweets has been used for sentiment analysis [23]. As a result of the analysis of tweets, we have obtained the method accuracy values of 0.8799 and 0.8865 in cases without and with the use of correcting coefficients, respectively. Analysis of the same dataset by VADER and TextBlob methods.
shows an accuracy of 0.82 and 0.87, respectively. Thus, the simple method based on dictionaries and rules is effective.

Fig. 3. The sentiment analysis results of texts (a, c, e) and short messages (b, d, f) with negative (a, b), positive (c, d) and neutral (e, f) emotional tones.

| Text          | Correction factor for | Emotional tone |
|---------------|-----------------------|----------------|
|               | adjectives | verbs | nouns | intensifying adverbs | softening adverbs |                     |
| Negative text | 1          | 1     | 1     | 1                    | 1                  | -0.18143           |
|               | 1          | 0.7   | 0.5   | 1                    | 1                  | -0.16901           |
|               | 1          | 0.7   | 0.5   | 2                    | 0.5                | -0.16905           |
| Positive text | 1          | 1     | 1     | 1                    | 1                  | 0.23417            |
|               | 1          | 0.7   | 0.5   | 1                    | 1                  | 0.24475            |
|               | 1          | 0.7   | 0.5   | 2                    | 0.5                | 0.26092            |
| Negative message | 1          | 1     | 1     | 1                    | 1                  | -0.39407           |
|               | 1          | 0.7   | 0.5   | 1                    | 1                  | -0.42410           |
| Positive message | 1          | 1     | 1     | 1                    | 1                  | 0.39054            |
|               | 1          | 0.7   | 0.5   | 1                    | 1                  | 0.38185            |
|               | 1          | 0.7   | 0.5   | 2                    | 0.5                | 0.40079            |
V. CONCLUSION

The text tone automatic determination system based on sentiment dictionaries and rules has been created. First, the texts have been classified into the emotional categories of anger, anticipation, disgust, fear, joy, sadness, surprise and trust by means of tonal and synonym dictionaries. Then using fuzzy rules, conclusions about the text belonging to different emotional categories have been aggregated and the quantitative value of the text tone has been obtained.

It is proposed to use coefficients that take into account the different emotional load of various parts of speech and the influence of intensifying or softening adverbs. The influence of the applied coefficients on the results of sentiment analysis of texts from different sources and of various emotional tones has been studied. A slight change in the quantitative value of the text tone due to the introduction of such coefficients has been found. In addition, the greater impact of emotional words on the text tone of short messages has been found. The use of additional tools provides the analysis of more words in the text and their emotional weight, which increases the validity of emotion detection and accuracy of the method.

REFERENCES

[1] C. Gallagher, E. Furey, and K. Curran, “The application of sentiment analysis and text analytics to customer experience reviews to understand what customers are really saying,” International Journal of Data Warehousing and Mining, vol. 15, no. 4, pp. 21–47, 2019. https://doi.org/10.4018/IJDWM.201910102

[2] F. Poeze, C. Ebster, and C. Strauss, “Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts,” Procedia Computer Science, vol. 130, pp. 660–666, 2018. https://doi.org/10.1016/j.procs.2018.04.117

[3] Z. Drus and H. Khalid, “Sentiment analysis in social media and its application: Systematic literature review,” Procedia Computer Science, vol. 161, pp. 707–714, 2019. https://doi.org/10.1016/j.procs.2019.11.174

[4] R. P. Schumaker, Y. Zhang, C. N. Huang, and H. Chen, “Evaluating sentiment in financial news articles,” Decision Support Systems, vol. 53, no. 3, pp. 458–464, June 2012. https://doi.org/10.1016/j.dss.2012.03.001

[5] J. Hao and H. Dai, “Social media content and sentiment analysis on consumer security breaches,” Journal of Financial Crime, vol. 23, no. 4, pp. 855–869, 2016. https://doi.org/10.1108/JFC-01-2016-0001

[6] S. Mansour, “Social media analysis of user’s responses to terrorism using sentiment analysis and text mining,” Procedia Computer Science, vol. 140, pp. 95–103, 2018. https://doi.org/10.1016/j.procs.2018.10.297

[7] R. Feldman, “Techniques and applications for sentiment analysis,” Communications of the ACM, vol. 56, no. 4, pp. 82–89, Apr. 2013. https://doi.org/10.1145/2346256.2346274

[8] A. D’Andrea, F. Ferri, P. Grifoni, and T. Guzzo, “Approaches, tools and applications for sentiment analysis implementation,” International Journal of Computer Applications, vol. 125, no. 3, pp. 26–33, Sep. 2015. https://doi.org/10.5120/IJCA.2015.905866

[9] K. Howells and A. Ertugan, “Applying fuzzy logic for sentiment analysis of social media network data in marketing,” Procedia Computer Science, vol. 120, pp. 664–670, 2017. https://doi.org/10.1016/j.procs.2017.11.293

[10] P. Jain, A. Sovastava, V. Singh, and B. Hazela, “Aspect based sentiment analysis by fuzzy logic,” International Journal of Current Engineering and Technology, vol. 9, no. 2, pp. 243–248, Apr. 2019. [Online]. Available: https://inpressco.com/wp-content/uploads/2019/04/Paper8243-248.pdf

[11] H. Liu and M. Caoa, “Fuzzy rule based systems for interpretable sentiment analysis,” in 26th International Conference on Advanced Computational Intelligence, Feb. 2017, pp. 129–136. https://doi.org/10.1109/ICACI.2017.7974497

[12] Y. Bai and D. Wang, “Fundamentals of fuzzy logic control – fuzzy sets, fuzzy rules and defuzzifications,” in Advanced Fuzzy Logic Technologies in Industrial Applications. Advances in Industrial Control, Y. Bai, H. Zhuang, and D. Wang, Eds. London: Springer, 2006, pp. 17–36. https://doi.org/10.1007/978-1-84628-469-4_2

[13] I. Olenych, Y. Olenych, A. Kostruba, and Y. Pryima, “Fractal analysis of porous structures using a fuzzy logic system,” in 2019 Xth International Scientific and Practical Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, Nov. 2019, pp. 97–101. https://doi.org/10.1109/ELIT.2019.8892344

[14] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, Jul. 2008. https://doi.org/10.1561/1500000011

[15] A. Khurshid, Affective Computing and Sentiment Analysis: Metaphor, Ontology, Affect and Terminology. Berlin: Springer Science & Business Media, 2011.

[16] U. Jain and A. Sandhu, “A review on the emotion detection from text using machine learning techniques,” International Journal of Current Engineering and Technology, vol. 5, no. 4, pp. 2645–2650, 2015. [Online]. Available: https://inpressco.com/a-review-on-the-emotion-detection-from-text-using-machine-learning-techniques/

[17] F. K. Chopra and R. Bhatia, “Sentiment analyzing by dictionary based approach,” International Journal of Computer Applications, vol. 152, no. 5, pp. 32–34, Oct. 2016. https://doi.org/10.5120/jica2016911814

[18] M. Thelwall, K. Buckley, G. Paltoglou, A. Kappas, and D. Cai, “Sentiment strength detection in short informal text,” Journal of the American Society for Information Science and Technology, vol. 61, no. 12, pp. 2544–2558, Dec. 2010. https://doi.org/10.1002/asi.21416

[19] K. Denecke, “Using SentiWordNet for multilingual sentiment analysis,” in 2008 IEEE 24th International Conference on Data Engineering Workshops, Apr. 2008, pp. 507–512. https://doi.org/10.1109/ICDEW.2008.4489830

[20] S. M. Mohammadvand and P. D. Turney, “Crowdsourcing a word-emotion association lexicon,” Computational Intelligence, vol. 29, no. 3, pp. 436–465, Aug. 2013. https://doi.org/10.1111/j.1467-8642.2012.00460.x

[21] S. M. Mohammadvand, “Word affect intensities,” in Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018), 2018, pp. 174–183.

[22] Y. Deng, Y. Zhang, and H. Chen, “A lexicon enhanced method for sentiment classification: An experiment on online product reviews,” IEEE Intelligent Systems, vol. 25, no. 4, pp. 46–53, Jul.-Aug. 2010. https://doi.org/10.1109/MIS.2009.105

[23] S. Yadav, “Pre-processed Twitter tweets,” 2017. [Online]. Available: https://www.kaggle.com/shashank1558/preprocessed-twitter-tweets. Accessed on: Mar. 15, 2021.

Igor Olenych received the degree of PhD Candidate in Physical and Mathematical Sciences from the Institute for Physical Optics, Ukraine in 2010 and a degree of Doctor of Sciences in Physics and Mathematics from Ivan Franko Lviv National University, Ukraine in 2020. He is presently the Head of the Department of Radioelectronic and Computer Systems at Ivan Franko Lviv National University. His current research interests include fuzzy modeling as well as smart solutions and systems.

E-mail: igor.olenych@ina.lviv.ua

ORCID ID: https://orcid.org/0000-0002-6642-0222

Oleh Sinkeych received the Master degree in System Analysis and Management from Ivan Franko Lviv National University, Ukraine in 2010. From 2010 to 2017, he was conducting research at Pedelechov Institute for Applied Problems of Mechanics and Mathematics, National Academy of Sciences of Ukraine. He is presently a PhD student and Assistant Professor of the Department of Radioelectronic and Computer Systems at Ivan Franko Lviv National University. The main research topics are machine learning applications in fog computing, embedded systems and data mining.

E-mail: oleh.sinkeych@ina.lviv.ua

ORCID ID: https://orcid.org/0000-0001-9834-2700
Maryana Salamakha received the degree of PhD Candidate in Philological Sciences from Ivan Franko Lviv National University, Ukraine in 2016. She is presently an Associate Professor at the Department of Foreign Languages for Natural Sciences, Ivan Franko Lviv National University. Her current research interests include the structure and semantics of terms and paradigmatic relations in terminology.
E-mail: maryana.salamakha@lnu.edu.ua
ORCID iD: https://orcid.org/0000-0002-9873-4728

Marianna Prytula received the Master degree in Computer Science from Ivan Franko Lviv National University, Ukraine in 2020. She is presently a PhD student at Ivan Franko Lviv National University. Her research interests include smart solutions and systems.
E-mail: mariannaprytula@gmail.com
ORCID iD: https://orcid.org/0000-0003-3410-3350