

Ministry of Education of the Republic of Belarus
Educational Institution
Belarusian State University
of Informatics and Radioelectronics

UDC 004.855

Cheng Chengliang

EMOTION RECOGNITION ALGORITHM

ABSTRACT

for a master's degree

Speciality 1-40 80 02 System analysis, information control and processing
(by industries)

Academic Supervisor
German Yuliya Olegovna
Candidate of Technical Sciences,
Associate Professor of the
Department of Electronic Computing
Machines, BSUIR

Minsk 2024

INTRODUCTION

Nowadays individuals cannot possibly know everything in the world, so they often seek advice and recommendations when faced a choice, whether it be regarding a car, technology, book, or film. In the past, people would turn to their friends for traditional advice, but with the rapid development of information and computing technologies and the global Internet, the process of obtaining recommendations has been greatly simplified. However, as the number of Internet users increases, so does the amount of user-generated content.

With the development of computer technologies, the number of people on the Internet has grown at a high rate, with 4.66 billion Internet users and 4.2 billion social media users. Globally, 5.22 billion people use cell phones, equivalent to 66.6% of the world's population. 90% of Internet users access the Internet via smartphones, but 2/3 of them also say they access the Internet using a laptop or desktop computer. As a result, the Internet produces an unimaginable amount of data every day. According to Facebook, it generates 4 petabytes of data every day, including 10 billion messages, as well as 350 million photos and 100 million hours of video viewing. It can be seen that against the background of such a volume of data, it gives artificial intelligence the opportunity to develop at a high speed, and the development of artificial intelligence will deeply change people's way of life.

While it may seem simple to send a request to a search engine and receive information about a desired product or service, the reality is that this task can be complex and tedious. The vast amount of information available on the Internet today often results in the necessary data being lost amidst a sea of irrelevant information.

The demand for automatically analyzing user opinions about goods or services is high. A system is needed to automatically find and analyze opinions on various subjects, helping users make informed decisions. Additionally, online resource owners should be aware of user opinions on any products in their online store. Analysis of user ratings enables quick decision-making regarding the sale of goods, changes in plans for introducing services, and the overall marketing policy of the company.

Sentiment analysis has practical applications in various fields, including sales and marketing. By monitoring social systems, conclusions can be drawn about the popularity of a particular product and current trends among buyers. This information is then used to organize production and marketing of products based on the studied market needs for goods and services. This approach provides a wider range of research opinions compared to surveys on the manufacturer's website and focus groups. It includes analysis of user political positions and

forecasting of election results in politics, extraction of interesting social data such as political and religious views in sociology, definition of de-pression in users in medicine and psychology, market forecasting based on news, blogs, and other social media in finance, and filtering and detection of messages containing information about illegal actions such as terrorist threats and analysis of prohibited content on websites for security protection.

The task of analyzing the sentiment of a text is a relatively recent development, and work on it is ongoing. Although there are tools and platforms available to determine the tone of social media messages, identify their subject and object, analyze opinions about products, and analyze other parameters, there is no single algorithm that can solve this problem with complete accuracy. Research is currently being conducted in this area, and the task of building a system for extracting opinions and analyzing sentiments remains relevant. The purpose of this master's thesis is to develop an algorithm for assessing the emotional tone of reviews.

GENERAL CHARACTERISTICS OF THE WORK

The aim of this research is to develop a sentiment analysis algorithm for sentiment classification of texts such as movie reviews.

This research is relevant because of the increasing use of Artificial Intelligence in text generation and the large amount of text data that needs to be analyzed and processed quickly in a variety of fields ranging from business to scientific research.

The object of the study is the emotional tone of film reviews. The subject of the research is neural networks, decision-making methods, metrics for identifying keywords.

For the purpose of this research, it is assumed that the following tasks will be performed:

- Analyze existing methods and approaches. This phase includes the study of relevant scientific literature and existing technological solutions in the field of natural language processing and word similarity metrics.

- Development of algorithm for sentiment analysis. At this stage, the task is to create a neural network for regression analysis.

- Implementation of algorithm. At this stage, it is important not only to develop the algorithms, but also to ensure that they are implemented correctly in the form of program code.

- Implementation and utilization.

All of the above tasks are aimed at achieving the main goal of this research – to create an effective neural network for sentiment analysis of text.

Provisions for defense:

1. The combined approach that uses the TF-IDF metric to find keywords, as well as the Saaty method to rank review scores. This made it possible to increase the accuracy of describing the output results in the range from 0 to 1.

2. The vector data compression method, which made it possible to reduce the vector length by 3 times. Thanks to this, we can submit fewer combinations and reduce training time. The quality of teaching suffers slightly.

The topic of the dissertation work corresponds to the list of priority areas determined by the Decree of the President of the Republic of Belarus dated 07.05.2020 (No. 156) “On priority areas of scientific and technical activity in the Republic of Belarus for 2021-2025” (direction of development of the information society, electronic state and digital economy).

The main provisions and results of the research were discussed at various conferences.

SUMMARY OF THE WORK

The focus of the work is on the research and development of sentiment analysis techniques for optimizing the size of text data sets in the sentiment analysis process. The basic goal consists in the use of optimized text data sets to provide good analysis results and their potential usage for good application.

The focus of this research is to formulate the problem of accurate query and its impact on the quality of the generated product description. In particular, the importance of the correctness of writing words in the query is emphasized.

Algorithm for preprocessing data sets has been developed. This takes into account various parameters that affect the quality of the information description of the text dataset. Special attention was paid to the definition of keywords in text data sets using the TF-IDF method. The keyword priorities were defined by means of Saati method, and the compression of data methods such as Huffman Coding was considered.

Neural networks based on developed algorithms for preprocessing data sets and regression analysis have been realized to analyze and verify the emotions of texts. As a result, it successfully analyzed the emotions of different texts.

In the end of the work, experiments were carried out to describe the accuracy of the suggested sentiment analysis model.

CONCLUSION

Throughout the course of the study, all goals and objectives were successfully accomplished, leading to valuable insights into the capabilities of cognitive computing platforms such as Google Cloud Natural Language and IBM Watson. The strengths and weaknesses of these platforms were thoroughly explored, providing a comprehensive understanding of their potential in the field of text analysis.

The thesis introduces a novel approach to text analysis, encompassing several key steps. First, the TF-IDF technique was employed to identify keywords within the corpus of reviews. This step enabled the extraction of essential information and facilitated subsequent analysis. Additionally, the Saaty technology was utilized to determine the priorities of these keywords, allowing for a more refined evaluation of their significance in the context of sentiment analysis. Finally, a regression neural network was constructed, employing a reduced vector length, which optimized the efficiency and performance of the model.

To facilitate the experimental process and assess the accuracy of the proposed model, bespoke programming software was developed. This software served as an essential tool in organizing and conducting the experiment, ensuring robust and reliable results.

The research objectives were effectively achieved, and all assigned tasks were successfully completed. Notably, a significant contribution was made through the introduction of a novel method to reduce the width of the training data, resulting in a substantial reduction in size by a factor of three. This achievement demonstrates the innovation and ingenuity of the research, highlighting its relevance and importance in the field of text data processing, sentiment recognition, and analysis.

The findings of this study have significant implications for numerous applications, ranging from market research and customer sentiment analysis to political forecasting and content filtering for security purposes. The developed methodology and the insights gained contribute to the advancement of sentiment analysis techniques, ultimately enhancing decision-making processes and fostering a deeper understanding of user preferences and attitudes.

Overall, this research provides a solid foundation for further advancements in the field of sentiment analysis and text data processing. The successful accomplishment of the study objectives, coupled with the development of a novel method and the implementation of dedicated software, underscores the significance of this research and its potential impact on various industries and domains.

LIST OF PUBLISHED WORKS

1-A. Cheng, C. Facial expression recognition method on static and dynamic image / C. Cheng, Yu. O. German // Information Technologies and Systems 2022 (ITS 2022): Proc. of the International Conference, 23 November 2022 / BSUIR, Minsk – 2022. –P. 137–138.

2-A. Cheng, C. Text sentiment recognition for economic and financial application / C. Cheng // III International Scientific and Practical Conference "Current vectors of Belarusian-Chinese trade and economic cooperation", 16 December 2022 / BSEU, Minsk – 2022. – P. 204–212.

3-A. Cheng, C. Sentiment extraction and recognition algorithm / C. Cheng // 59th conference of postgraduate students, undergraduates and students of the educational institution "Belarusian State University of Informatics and Radioelectronics", 17-21 April 2023 / BSUIR, Minsk – 2023. – P. 54.