SOCIAL NETWORK TEXT CLASSIFICATION ALGORITHM AND SOFTWARE TOOL

A.A.Taniberdiyev

"Information Technologies" Department of Gulistan State University,

Doctor of Philosophy in Pedagogical sciences (PhD)

S.F.Saidova

Master's student of Gulistan State University "Information Technologies" department

Abstract: *In this article, today's information in social networks, global information flows and algorithms for their analysis, research in this regard and their effects are studied. Recent effective algorithms are presented and the procedure for their use is shown.*

Key words*: social network, text correspondence, analysis, internet networks, messages, comments, nlp, bert, gpt.*

INTRODUCTION

Today, information on social networks is the main part of the global information flow. Social media text messages are generally written communications, posts, comments, or messages made through social media. This type of correspondence can be in a variety of formats, such as short messages, articles, comments, or replies. This information processing is accelerating on a large scale. The reason is that conducting sufficient analysis in Internet networks with a large amount of information is an important tool in determining the reliability and validity of information.

OBJECTS AND RESEARCH METHODS

The main types of social network text correspondence. Messages, Personal communication (e.g. via Facebook Messenger or Instagram Direct), Group chats (e.g. WhatsApp or Telegram groups), Posts, posts on a social network to communicate their thoughts to a wide audience (Twitter, Facebook, or LinkedIn posts). Comments, Responses or Opinions to Posts or Content. For example, comments left under a video. Tags and discussions (Mentions and Discussions). Mention other users or participate in discussions. Hashtag and contextual correspondence (#Hashtags). Short entries related to a specific topic or keywords that identify topics. Forum and group correspondence. Broader discussions on Facebook groups, Reddit, or other forums. Peculiarities of social network correspondence, the degree of brevity of the text. Posts on platforms like Twitter should be kept short (for example, up to 280 characters). However, there may also be correspondence in the format of a long article (such as LinkedIn or Facebook articles). Informal language and abbreviations. Simple and informal language is often used. Diversity of information. Posts cover a variety of topics: personal life, politics, culture, advertising, and more. Real time communication. In many cases, correspondence occurs in real time

(for example, chats or comments during a broadcast). Multichannel. Text can be mixed with other formats (along with images, videos or voice messages).

RESULTS AND ITS DISCUSSION

Studying social media correspondence has many uses. Determining what users think about the brand. Sentiment analysis. Knowing whether the text is positive, negative or neutral. Clarification of trends. Identifying current topics. Filter spam or malicious content. Block inappropriate or dangerous correspondence and more.

A text classification algorithm is the process of classifying text into predefined categories or classes. These algorithms are implemented using natural language processing (NLP) techniques. Below are the key elements needed to understand and implement the text classification process. In turn, the classification process includes a number of stages. This includes the steps of the classification algorithm. Collection and preparation of information: collection of correspondence from social networks, preparation of texts, cleaning, normalization, tokenization.

Convert text to digital format. The following methods are used to represent the text with numbers: Bag of Words (BoW) - counting the number of words found in the text. TF-IDF (Term Frequency-Inverse Document Frequency) is an assessment of the importance of a word in the text and in the entire data set. Embedding-Text representation in vector format (eg Word2Vec, GloVe, or Transformer-based embeddings).

Special algorithms are required to work with this type of information. These algorithms should be selected depending on the type of information. Algorithm selection and training.

Simple algorithms. These include Naive Bayes (fast and efficient in text classification) and Logistic Regression (linear model suitable for classification).

Powerful machine learning algorithms. Support Vector Machines (SVM) have high accuracy on small data. Random Forest or XGBoost: Used in complex classification problems.

Deep learning algorithms. RNN and LSTM (good at analyzing text with long-range sequences). Transform models (eg BERT, GPT) for deep understanding of text context.

Model evaluation. The effectiveness of the classification model can be evaluated through the following criteria: Accuracy, Precision and coverage.

In data processing, there are transformer-based models, among them BERT, GPT. BERT, GPT has high performance in text classification. BERT analyzes contextual information in depth. For example, it determines the difference between the phrases "I am happy" and "I am not happy". GPT is used for pre-training and comprehension of texts.

CONCLUSION

In the process of selecting an algorithm for classifying social network text messages, it is important to consider data volume, accuracy level requirements, and resource availability. Choosing the right algorithms when analyzing the large amount of data

received first of all increases the speed of data analysis. The quality and speed of data analysis in cloud storage directly depends on the right algorithm.

ADABIYOTLAR:

1. Fundamentals of compilation of electronic tasks for students to test and strengthen their knowledge of database(Article). Toshtemirov, D., Muminov, B., Saidov, J. International Journal of Scientific and Technology Research. Volume 9, Issue 4, April 2020, Pages 3226-3228. https://www.ijstr.org/paper-references.php?ref=IJSTR-0120-29952

2. Tools and Methods of Formation of Professional Competence of Future Teachers of Computer Science and Information Technologies. SaidovJasur Doniyor o'g'li, Baxodirov Muzrob Doniyor o`gli, Normatova Malika Norkulovna, Mavlonov Sherzod Hazratqulovich. Psychology and Education Journal. Vol. 58 No. 2 (2021): Volume 58 No. 2 (2021) https://doi.org/10.17762/pae.v58i2.2282

3. Baxtiyor Baxriddinovich Ergashev, Jasur Doniyor Ogli Saidov, & Sait Xalilovich Islikov (2021). BO'LAJAK INFORMATIKA VA AXBOROT TEXNOLOGIYALARI O"QITUVCHILARI KASBIY KOMPETENTLIGINI SHAKLLANTIRISH VOSITALARI VA METODLARI. Academic research in educational sciences, 2 (2), 1139-1146. doi: 10.24411/2181-1385-2021-00312

4. N.A.Muslimov, K.M. Abdullayeva, O.A.Kuysinov, N.S.Gaipova, N.N.Karimova, M.Kodirov. - Kasb ta'limi o'qituvchilarining kasbiy kompetentligini shakllantirish texnologiyasi. —T.: «Fan va texnologiya», 2013, 128 bet.

5. 3. Y.Y.Chicherina, D.A.Nurkeldiyeva, D.B.Yakubjanova-Mutaxassislik fanlarini o'qitish metodikasi. O'quv qo'llanma Toshkent, 2013