UDC: 004.93 https://doi.org/10.15407/jai2024.04.228

I. Mysiuk
Ivan Franko National University of Lviv, Ukraine
1, Universytetska st., Lviv, 79000
Iruna.musyk8a@gmail.com
https://orcid.org/0000-0002-3641-4518

EVALUATION OF USER ACTIVITY PARAMETERS IN SOCIAL NETWORKS USING MACHINE LEARNING TOOLS

Abstract. Social networks provide a huge amount of data that can be used to understand user behavior because they cover a wide range of human activities such as posts, comments, likes and follower interactions. Analyzing these data helps to identify patterns and trends that can shape patterns of behavior and assess the parameters of their activity. Key parameters such as the frequency of users' actions, their interactions, the content they share, and the emotional tone of their communications offer valuable insights into user behavior. By formalizing these parameters, large data sets can be turned into meaningful indicators for analyzing behavior, predicting or improving existing models by using machine learning tools, namely: regression methods or by applying their classification. To improve the accuracy of the analysis, it is extremely important to choose appropriate activity indicators and take into account changes over time and different types of interactions. This research suggests analyzing key metrics, including post frequency and user engagement, and visualizing these metrics to gain a deeper understanding of user behavior trends. According to the obtained results, it is possible to calculate the audience engagement factor, the popularity factor of the post in social networks, and the audience retention rate. In addition, an important result when evaluating user parameters for analyzing their behavior is the content influence factor, which shows the contribution of a user's post to other users.

Keywords: user activity, data processing, data collection, data analysis, social networks, machine learning.

Introduction

Social networks contain an extremely large number of parameters and metrics that can be obtained during the data collection stage. In addition, social networks are an important source of data on user behavior, as they reflect a wide range of human activity: posts, comments, likes and followers can be analyzed to identify patterns and trends. Highlighting key parameters will help to form the necessary behavioral model of user activity in social networks and determine their impact. User activity can be considered from different points of view: frequency of actions, interaction with other users, content of publications and emotional coloring of communications. The formalization of such parameters allows structuring a large amount of data and turning them into indicators that can be used to analyze behavior or create forecasts or supplement existing models. For a more accurate analysis, it is important to choose the correct indicators of user activity and take into account their variations over time and by types of interaction. It is proposed to perform an analysis of the main parameters of the activity of social network users, such as the frequency of publications and the number of interactions. In addition, the calculation of relevant indicators and their visualization can help to better understand the behavioral trends of users.

Statement of the problem

The problem of evaluating user activity in social networks with the help of machine learning is the need to create a model that will allow automatic analysis of user behavior and prediction of its changes. In order to be able to work with real data, it is necessary to select and prepare data from social networks using APIs or using web scraping as part of the automation process. After that there is pre-processing of the data. It includes cleaning, normalization of data, removal of abnormal values. Today, the process of selecting and testing machine learning models to achieve the best accuracy is not sufficiently researched. Therefore, it is worth paying attention to the available approaches and parameters of the model that will be taken into account at the stages of their training and testing. After that, the effectiveness of the model is evaluated using metrics, and the most influential factors that determine behavior are identified and recommendations are made to increase interaction with the audience.

Analysis of recent research and publications

The latest research by scientists is mostly based on the use of machine learning models to predict and classify user activity based on social network data tied to geolocation. In addition, the articles highlight different approaches to evaluating and predicting user activity using machine learning techniques, particularly based on their interactions with content and behavioral characteristics on social networks. According to the publication [1], the authors used a dataset from Weibo for analysis and developed four different models, including deep learning and decision trees, to classify and predict user activity. This article [2] explores the possibilities of using neural networks and machine learning algorithms to analyze user data in social networks. Therefore, the main focus is on how these technologies help in recommending content and evaluating the emotional component of interactions in social networks. Research [3] is devoted to forecasting active and passive interaction of users in social networks using context-dependent models based on LSTM networks. The researchers used a large dataset to show how user engagement varies in different contexts. In the study [4], researchers focus on the application of deep learning methods to analyze the behavioral patterns of users in social networks, in particular, interaction with content and the emotional coloring of publications. Article [5] presents models for predicting user interaction based on analysis of likes, comments and reposts, using machine learning algorithms to identify the main factors influencing activity. In [6], the article investigates the use of machine learning algorithms to predict the behavior of Instagram users based on their interaction with content and comments. The publication [7] proposes the application of deep learning for social media analysis and prediction of user engagement on various platforms such as Twitter and Facebook. Research in [8] shows various machine learning techniques for analyzing user sentiment and engagement in social networks such as Facebook using datasets about their activity. Analyzing the publication [9], it is possible to see the factors affecting the interaction of users in social networks, using machine learning methods to identify patterns in behavior. In the article [10], researchers propose a framework based on machine learning for analyzing user activity patterns in social networks and their impact on business strategies and digital marketing.

The purpose of the study

The main research method is the possibility of developing a model that, based on the collected data on user actions (posts, likes, comments, reposts), will be able to estimate the level of their activity and predict its changes. In addition, it is an opportunity to prepare input data for user interaction research. Also, to evaluate the model, it is worth highlighting the main indicators of activity, such as: the frequency of publications, the number of interactions, the time of activity, the growth rate of the audience, etc. The built machine learning model will be used to classify users by activity level and predict their future behavior. The developed behavior model, taking into account the actions of users in the social network, in particular posts, likes, comments and reposts, as well as information about the time of publications, the type of content and the interaction of other users, will help to identify patterns between user activity and calculate the audience engagement factor, the post popularity factor in social networks, the audience retention rate and the content impact factor, which shows the contribution of a user's post to other users. And using machine learning models, it is possible to divide users into groups according to their level of activity (low, medium, high) and apply prediction of activity or interactions based on behavioral characteristics.

Presentation of the main material

During the collection of information, a very important factor is the display of the effectiveness and feasibility of the data that has been obtained. At the stage of collecting all the data, there may be a problem with saving them, because it takes up a lot of computer memory. Another problem may be during the processing of the collected data. Information processing time will increase rapidly with big data processing. That is why the formalization of the main parameters will help to avoid large

loads on computing power and will increase the usefulness of selected metrics for researching user activity in social networks.

Parameters of user activity can be formalized through many indicators that allow to identify general trends of behavior. The main parameters used for this are: Frequency of posts (F_p), the frequency of interactions (F_i). The engagement factor (E) is the ratio of the number of interactions to the number of publications:

$$E = \frac{F_i}{F_p} \ .$$

Frequency of interactions (F_i) parameter reflects how actively the user participates in interactions with the content of other users. The quantitative parameter of the frequency of iterations is necessary to influence the popularity of the publication and monitor the engagement of the audience:

$$F_i = \frac{C_i+L_i+S_i}{T} \ ,$$

where C_i is the number of comments made by the user, L_i is the number of likes placed by the user, S_i is the number of reposts, T is the duration of the analysis period.

The engagement factor has an important role, because the algorithms of social networks are configured in such a way that with an increase in the engagement factor, the influence of the posted content is greater, so the post will reach a larger audience of users.

Popularity factor (P) reflects the user's level of influence on others. The popularity factor directly depends on the engagement factor and vice versa. To gain popularity in social networks, it is necessary to increase the activity of users with a certain post. It is calculated through the number of interactions that the user's posts receive from others (likes, comments, reposts) and his audience:

$$P = \frac{L+C+S}{N_f} \ ,$$

where N_f is the number of subscribers.

Audience retention rate (R) metric shows how many users continue to interact with content over a period of time. The audience retention rate is useful when considering the popularity parameter. This parameter will help predict the popularity of the post over time. It is calculated according to the formula:

$$R = \frac{N_{a2}}{N_{a1}} * 100\% \ ,$$

where N_{a1} is the number of interactions for the first period (for example, for a month), N_{a2} is the number of interactions in the second period.

Content Impact Factor (C_p) shows how successful a user's posts are in attracting new followers and generating activity:

$$C_p = \frac{N_f(from\ start)+N_f(to\ end)}{N_p}$$

where N_p the number of posts for this period.

To calculate the impact of a published post, it is necessary to understand whether the publication gained popularity artificially with the help of money or naturally based on the gradual increase in the volume of published content. If the number of publications and coverage increases gradually, this type of popularity is called organic. In the case of a sharp jump in the collected parameters, it can be argued that the posts have an inorganic influence.

Therefore, by formalizing the main parameters, it is possible to increase the performance of the developed system and improve the initial results for building an analytical system of user behavior in social networks. The engagement factor and the popularity factor allow you to get results for further forecasting of the activity of the post, and the quantification of iterations and the influence factor will show the relevance of the selected content in social networks.

After processing the data, the collected parameters for a specific subject area can be easily visualized and used for further prediction or additional processing. The collected parameters will be sufficient as input parameters to the decision-making system and for processing by machine learning algorithms.

Table 1 describes the main machine learning methods used to evaluate user activity in social networks, as well as tools for their implementation. Logistic regression helps predict user interactions, such as a like or

comment. Deep learning is useful for modeling complex user behavior patterns. Decision trees and Random Forest allow you to classify data, in particular, by the popularity of posts. The support vector method (SVM) is used to classify users based on activity parameters. Gradient boosting is used to accurately predict activity trends based on multiple factors.

Table 1. Machine learning methods and tools for user activity modeling tasks

Machine Learning Method	Description of method usage	Machine Learning Tools
Logistic Regression	Predicting user interaction (like, comment, share)	Scikit-learn, TensorFlow, Keras
Deep Learning (Neural Networks)	Modeling complex interactions and behavior pattern analysis	TensorFlow, PyTorch, Keras
Decision Trees (Random Forest)	Classifying data based on post popularity	Scikit-learn, XGBoost, CatBoost
Support Vector Machine (SVM)	Classifying users by activity parameters	Scikit-learn, LIBSVM
Gradient Boosting	Predicting user behavior based on multiple factors	XGBoost, LightGBM, CatBoost

In general, the calculation of user activity in social networks can be divided into low, medium and high levels. These statistics can be analyzed using data collected from social networks in related work [11].

Table 2. Calculation of low, medium and high level of activity of users in social networks

Parameters	Low level activity	Medium level activity	High level activity
Audience Engagement Factor, F_i	< 10 interactions (likes/comments)	10 –100 interactions	> 100 interactions
Post Popularity Factor, P	< 100 likes, < 10 shares	100 –500 likes, 10-50 shares	> 500 likes, > 50 shares
Audience Retention Rate, R	< 20% repeated interactions	20 –50% repeated interactions	> 50% repeated interactions
Content Influence Factor, C_p	< 10 discussions	10 –50 discussions	> 50 discussions

Table 2 describes the activity of users in social networks through four parameters, taking into account the audience engagement factor, the popularity of the post in social networks, the retention rate and the influence of the content, showing three levels of activity - low, medium and high. When analyzing the audience engagement factor, you can see that the higher the activity, the more interactions, such as likes, comments, reposts. At a low level, users hardly interact with the content, while at a high level, there is regular commenting and discussion.

Evaluating the popularity factor of a post can be seen as the number of interactions with the post using likes, reposts, etc. A low level indicates weak audience interest, a medium level indicates stable popularity, and a high level indicates significant interest and popularity of the post.

The next parameter, the audience retention rate, is an indicator of how effectively the content retains users. A low level indicates that users are not returning to the content, while a high level indicates significant interest and regular return to engagement.

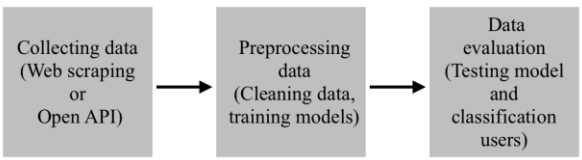

Fig.1. General process of evaluating users

In Figure 1 represent the general process of evaluating the behavior of users in social networks. It includes the selection of data from various sources, be it the process of web scraping or the use of APIs [12]. Data processing or pre-processing describes the step in which the data is cleaned from duplicates or redundant characters and the model is trained with the test data. Data evaluation is the final stage of the scheme and is responsible for testing models and classifying users.

Evaluating the influence factor of the content, the extent to which the content shapes the thoughts or behavior of the audience. A low level means that the content has almost no influence on users, a medium level means that the interaction depends on the topic, and a high level means that the content is actively discussed and influences the opinions of the audience.

Thus, it can be argued that the factors are interrelated: with the growth of one of the parameters, as a rule, the others also grow. A high level of retention and engagement indicates effective work with the audience and the creation of relevant, influential content.

Conclusions

Estimating the parameters of user activity behavior allows you to build detailed models of user behavior based on calculated parameters such as: frequency of publications, interaction with content, tone of communication and popularity of posts, etc. Using machine learning techniques to analyze large data sets helps to formalize these metrics and turn them into meaningful indicators. This contributes to the improvement of existing forecasting models and emphasizes the importance of these technologies for deep analysis of audience behavior. Machine learning allows you to automate the processes of predicting interactions with content, evaluating the levels of engagement and impact of publications, which helps to more accurately determine the effectiveness of marketing and communication strategies. The use of various algorithms, such as logistic regression, neural networks and gradient boosting, allows not only to analyze current activity, but also to predict future trends in user behavior.

References

1. Riaz, R., Jiang, S., & Wang, X. (2023). Prediction and classification of user activities using machine learning models from location-based social network data. *Applied Sciences, 13*(6), 3517.
DOI: https://doi.org/10.3390/app13063517

2. Taherdoost, H. (2023). Enhancing social media platforms with machine learning algorithms and neural networks. *Algorithms, 16*(6), 271.
DOI: https://doi.org/10.3390/a16060271

3. Xu, X., et al. (2023). Context-aware prediction of active and passive user engagement on social media platforms. *Journal of Big Data, 10*(5).
DOI: https://doi.org/10.1186/s40537-023-00677-8

4. Wang, Y., Xu, Z., & Sun, X. (2023). Mining user behavior in social media using deep learning. *Journal of Big Data, 10*(4), 67.
DOI: https://doi.org/10.1186/s40537-023-00578-9

5. Yu, H., Zhang, K., & Lee, D. (2023). Predicting user engagement in social media with machine learning. *IEEE Access, 11*, 12345-12359.
DOI: https://doi.org/10.1109/ACCESS.2023.3047890

6. Kim, S., & Lee, J. (2023). Predicting social media user behavior with machine learning models: A case study on Instagram. *Computers in Human Behavior*, 139, 107617.
DOI: https://doi.org/10.1016/j.chb.2023.107617

7. Ali, F., Zhou, X., & Zheng, H. (2023). Enhancing social media analytics using deep learning: A study on user engagement. *Expert Systems with Applications*, 216, 119451.
DOI: https://doi.org/10.1016/j.eswa.2023.119451

8. Nguyen, T., & Pham, M. (2023). Machine learning approaches for sentiment analysis and user engagement in social networks. *Knowledge-Based Systems*, 263, 110047.
DOI: https://doi.org/10.1016/j.knosys.2023.110047

9. Zhang, W., & Li, Y. (2023). Understanding the factors of user engagement in social media through machine learning models. *Information Processing & Management*, 60(2), 103001.
DOI: https://doi.org/10.1016/j.ipm.2023.103001

10. Ghasemaghaei, M., & Hassanein, K. (2023). A machine learning-based framework for understanding user activity patterns in social media. *Journal of Business Research*, 159, 113220.
DOI: https://doi.org/10.1016/j.jbusres.2023.113220

11. Mysiuk I., Mysiuk R., Shuvar R. (2023) Collecting and analyzing news from newspaper posts in facebook using machine learning. Stuc. intelekt. 28(1). PP. 147-154.
DOI: https://doi.org/10.15407/jai2023.01.147

12. I. Mysiuk, R. Shuvar. Automating Web Scraping of User Comments for Sentiment Analysis in Social Networks, 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, P. 77-81.
DOI:http://doi.org/10.1109/ELIT61488.2023.10310867

The article has been sent to the editors 18.10.24.
After processing 29.10.24.
Submitted for printing 30.12.24.