

UDK 519.62

Maksym.V. Shopynskyi¹, Nataliia.V. Golian², Iryna.V. Afanasieva³¹Department of Software Engineering, NURE, Ukraine, maksym.shopynskyi@nure.ua²Department of Software Engineering, NURE, Ukraine, nataliia.golian@nure.ua³Department of Software Engineering, NURE, Ukraine, iryna.afanasieva@nure.ua

PRINCIPLES OF SEARCHING AND SORTING OPTIMIZATION IN SOCIAL NETWORKS USING A MULTI-FACTOR ASSESSMENT SYSTEM

The analysis of social networks, which focuses on the relationship between social entities today is an area of active research. It is a set of tools for research, in particular, in combination with artificial intelligence methods such as machine learning, deep learning. The paper examined the current quality of the assessment of information in social networks, analyzed the methods of searching and sorting information in various social networks, as well as the process of providing recommendations to users. Social media data is an inexhaustible source of research and business opportunities. In general, social media data is information gathered from social networks that shows how users interact with content. Methods of improving search results for personalizing recommendations in social networks are given. These indicators and statistics provide an effective understanding of the strategy of behavior in social networks. The advantages and disadvantages of a multifactor assessment system are considered. The possible ways of integrating the combined system of evaluating information elements by the user to optimize search queries and filtering big data are identified.

SOCIAL NETWORK, SEARCH, FILTERING, MULTI-FACTOR ASSESSMENT SYSTEM, RATING SCALE, BIG DATA

Introduction

Social networks are the main direction of dissemination of information on the Internet in recent years. Most of them use a news feed for distribution of records among users.

Social media data is the source that comes from developing or analyzing social networks. After extracting data, analytics is used to sort out raw information. The more data can be collected, the more informed decisions can be made, which will lead to a better result.

Due to the need of providing relevant information, the actual problem is to optimize the searching and filtering the data provided to the user while viewing the feed.

The purpose of this research – analyze the basic algorithms and methods of searching, filtering and sorting information in social networks, possible ways of their optimization and implementation of these methods to the existing and future networks.

To achieve a goal, the following tasks were set:

- to analyze the basic methods of information search in social networks;
- to formulate possible ways of optimization the sorting and filtration of records in the news feed;
- to explore possible options for implementing the obtained methods to the existing social networks.

1. Analysis of the research problem

Before determining the social profitability of investments, it is necessary to first identify key performance indicators (KPI). These are the various business indicators used to measure and measure success. KPI of social networks are indicators that will help suggest what gives results and what does not. In other words, it is necessary to have data, which key performance indicators of social networks should be monitored and analyzed. Only

then can one understand by metrics whether a social strategy is being implemented.

For more than 20 years of history of the social networks' existence, the principles for creating a news feed were largely based on user activity [1]: the posts of friends/subscriptions, groups, and other news sources. In the search process, to sort the found information, initially there was used only the number of preferences (likes) received by a post. During the process of development of neural networks and artificial intelligence, their achievements began to be used to form a "smart" news feed and sort provided information, both in general terms and for each user separately. The relevance of each record is now calculated not only by the total number of preferences, but also by the number of views, related to the subject of the news feed [2].

One of the most important moments when collecting data in social networks is the availability of sufficient information for making an reasoned decision. Understanding user behavior and preferences is important for any business purpose: from determining what content people want to see, to controlling the mood of the community as a whole. The data exists regardless of the network in which the need for research arises: the question is in carrying out the necessary analysis.

Data-driven marketing is the process of obtaining information based on the analysis of indicators extracted from large data on consumer interaction in order to make predictions about future behavior. By collecting data in social networks using targeted platforms, it is possible to make the process of making business decisions more balanced.

Despite the progress in the development of social networks, they have one serious disadvantage. Since

most networks use likes (or their counterparts) to get a response from users [3], for the analysis of user preferences, the algorithms for providing information can only be guided by binary factors (like/dislike). Such a model cannot fully provide a clear assessment and feedback of the extent and the context derived from the user who liked a post.

Take for example one of the most popular social networks – Twitter. Twitter has two basic parameters for evaluating each post (tweet):

- the number of likes (bookmarks). It has two usage options: “liked” (the user rated a tweet) and “not defined” (the user did not rate a tweet);
- the number of retweets (allows to share tweets on your page). Also, it has only two states: the user either shared a tweet or not;
- the owner of the tweet. It allows to focus on the number of author’s subscriptions to be displayed in the search top.

Twitter has two methods for displaying found tweets: top results for the latest time and tweets sorted by date of publication. The formation of the top is most influenced by the status of the tweet’s author (number of subscribers, total activity). The number of retweets and likes [4] affect tweet’s rating less significantly.

The first problem of forming the top of the tweets when searching on Twitter is the universality of the results. The search considers only the user’s location (country and city) to create the feed of tweets. There is no more personalized search, since the feedback to users is quite low.

The second problem follows from this: Twitter does not have a unique method for analyzing user actions, since the likes are acting as favorites and cannot be used for real evaluation, and the number of retweets is not informative option for content creation algorithms.

So, the general problem of social networks is paying little attention to active user actions [5]. To create a news feed, they involve highly effective algorithms for machine learning, which should predict what might be liked by a user without having enough information. Due to this, search results in social networks are rather approximate, this may lead to an information collapse against the backdrop of exponential growth in the amount of data in the Internet, depriving many users of the benefit of one of the most valuable information sources.

3. Institution of segmented grading scale

For more detailed feedback from the reaction of users to an information post, it is necessary to expand the simplified scheme of a post evaluation in a social network.

Instead of using likes, the notion of rating ranking is introduced. This rating should have a differentiated

scale, depending on the required degree of detail. Major scales may range from 1 to 5 (like a five-point rating system used by some educational institutions) and a scale from 1 to 10 (due to the use of the decimal number system). In addition, since the emoji have become widely popular recently, they can also be used as an emotional assessment (from annoying emotion to satisfaction one, from a sad one to a fun one, etc.).

As the system of records evaluation, in the form of likes, has an advantage in the form of greater ease of use [6], to implement a segmented scale, it will be necessary to re-engineer the user’s post evaluation interface to maintain satisfactory ease of use.

The segmented scale requires one action more than the likes system (in the worst situations, where the scale is highly differentiated, there may be needed more actions, but such options are not considered here). In addition, for web-based systems and high-resolution smart devices, additional actions may not be even needed – the whole scale can be compactly positioned on the user’s screen. Considering that such a system will not greatly influence the usability of the evaluation, it can be assumed that such a method would not cause dissatisfaction with users.

The segmented scoring scale is now used in many online services: movie evaluation systems, mobile app distributors (Google Play), online stores (AliExpress, Rozetka), and more. It should be noted that the rating scale has already been applied to social networks. An example is a network, where the scale from 1 to 5 is used to evaluate photos. But due to incorrect design and lack of a logical context on the rating scale, this rating system was not able to get the right value.

A differentiated scale affects several aspects of social networking optimization. Thus, the user is given a wider choice to evaluate an item: instead of one option, he can choose from several (depending on the segmentation of the scale). In addition, users can express a negative reaction using low scores on the scale (like the score on Google Play, where the dissatisfied users give an app 1-2-star, etc.).

From the point of view of the internal structure of the social network, the assessment scale will help to optimize the statistics for an information element [7]. Instead of prioritizing the quantitative factor of evaluation (total number of likes), the qualitative factor (number of positive assessments and average score) is now on the foreground.

Also, in addition to the information item ratings, the segmented scale can optimize the recommended entries that are provided to the user. Now, recommendations can consider not only the records that were reviewed and approved by the user, but also their overall rating and preferences of the user (evaluation of related records). This will create an additional impact factor on

search results and recommendations that will make them more differentiated and independent for each network user.

The algorithm for ranking of the record displayed in the search results should change according to the entered scale. For a social network with estimates in the form of likes, in the simplest form it may look like this:

$$R = K * \text{likes} \quad (1)$$

R – overall post’s rating; K – the user influence factor (calculated for each network separately, depends on the user’s environment and their activity); likes – the number of likes received by a post.

The easiest version of a ranking with a segmented scoring scale is as follows:

$$R = K * \frac{\text{sum}_{rate}}{\text{totalRate}} \quad (2)$$

sum_{rate} – the total sum of received rating; totalRate – the maximum amount of ratings for a post (the number of ratings multiplied by the maximum rating value in numerical form).

Given that most modern systems use a special policy for minimum and maximum ratings to protect against deliberate increase of the rating, the formula (2) allows you to enter the coefficients of the significance of each estimate from a segmented scale. Then the formula will look like this:

$$R = K * \frac{\sum(\text{rate}_i * \text{rate}K_i)}{\text{totalRate}} \quad (3)$$

rate_i – the i -th element of the assessment scale in numerical form; $\text{rate}K_i$ – the coefficient of significance of the i -th element of the scale.

Therefore, a segmented scale for rating of information elements leads to more feedback from the user, which can help to customize the data filtering for each user more independently.

3. Assessment of the information post by factors

One of the major drawbacks in the evaluation of information elements with the help of the likes is the lack of a concrete reason for the assessment or its absence [8]. This system partially answers the question “Did the user like this item?”, Which has options for “Yes” and “Not at all”. Such an assessment is not very informative for the search algorithms of the social network, therefore for its additional analysis, the subject of the element and its affinity with other elements are considered.

The answer to the question “What exactly the user liked the item?” should expand the range of possible choices for the user. One of the possible solutions to this problem is the introduction of factors for the evaluation of information elements.

Evaluation factor is the unit of evaluation of the information element in the social network. It is an analog

of the like, but it has a certain context or category that is used to specify the user’s motives.

Compared to the segmented scale, which is responsible for the quantitative assessment of the elements, the task of the factor system is a qualitative analysis of data and obtaining reasons for the preference of an element by the user.

The factors can be categorized according to their versatility, way of display, and the purpose.

In terms of universality, the factors are divided into:

- general – can be applied to any information elements, there are universal evaluation options (for example, “Utility”, “Relevance”, “Truth”, etc.);
- specific – inherent to elements of a certain type or certain subjects (for example, factors such as “Perspective”, “Processing”, “Exposition”, etc.) can be used for evaluation of the photographer works;
- custom – factors created by the user for an individual assessment of the information element.

By the way of displaying, factors can be divided into:

- linguistic (verbal) – transmit values using words and phrases;
- figurative (graphical) – transmit values using images (pictures, emoji, etc.).

By the purpose, the factors are:

- statistical – used to consider the popularity and calculation of the overall rating of the information element;
- logical – used as a voting system, have several alternative choices for shaping further actions and making decisions based on user ratings;
- emotional – used as a psychological element evaluation; have a similarity to a segmented scale of evaluation (when using a scale of images), but they focus not on the quantitative but on the qualitative characteristics of the element.

The introduction of the factor system of evaluation takes place by increasing the number of factors of evaluation from one (likes) to several. At the same time, the main disadvantage of such a system is an increase in the complexity of the element’s evaluation by the user. In terms of the latter, the complexity increases in proportion to the number of factors. By Miller’s law [9] (about the number of objects an average person can keep in working memory), the maximum number of factors for evaluating one element should be $7 (\pm 2)$ variants.

Despite the mathematically increasing complexity of the factor system evaluating, many users already have experience working with such a system. Facebook, Github, and Stack provide the evaluation feature with emoji, which is one of the options for factor estimation, so in practice, the introduction of the multifactorial system will not cause inconvenience when evaluating the information elements of the network.

From the point of view of the internal structure of the social network, the introduction of the multi-factor evaluation system can greatly affect the algorithms for data search and filtration [10]. Like filtration systems in online stores, factors can be considered when getting search results, while reducing the number of possible variations and increasing their quality.

If we consider as the basic formula of the rating of an element for a one-factor evaluation system formula (2), then when introducing a multi-factor system, the formula of the element rating in filtering by a factor takes the following form:

$$R_f = K * \frac{factorsCount_f}{sum(factorsCount)} \quad (4)$$

R_f – post rating when filtering by selected factor; $factorsCount_f$ – the number of element ratings for the selected factor; $sum(factorsCount)$ – the number of element ratings by all factors.

To form a general assessment based on factors, coefficients of the influence of a factor on the overall assessment of the element can be introduced. In this case, the formula for the overall rating of the item becomes the following:

$$R = K * \frac{\sum(K_i^f * factorsCount_i)}{totalRatesCount} \quad (5)$$

K_i^f – the coefficient of influence of the i -th factor on the overall rating; $factorsCount_i$ – the number of element ratings on the i -th factor; $totalRatesCount$ – the total number of element ratings by all factors.

Thus, the multi-factor system allows users to evaluate information elements in a qualitative way. In addition, filtering can be applied to certain data by searching for data, which can greatly optimize the search results, including for each user independently (based on only his factor ratings).

4. Ways of factor system integration

To optimize the search and filtering of information in social networks and increase the role of users in the evaluation of elements, the combination of a segmented scale and a multi-factor evaluation system is optimal [11].

In this case, each factor will have a universal scale (numeric or figurative) for element evaluation. Thus, the number of options for the user's assessment will increase $g * f$ times (g – the number of segmented scale options; f – the number of factors) in comparison with the one-factor evaluation system based on the likes.

Given the formulas (3), (5) and integration options the general formula of ranking information element in the social network with a combined system testing (segmented scale with the multifactor system) should be as follows:

$$R = K * \frac{K_i^f * factorsCount_i * avgRate_i}{totalRatesCount} \quad (6)$$

$$avgRate_i = K * \frac{\sum(rate_j * rateK_j)}{totalRate_i} \quad (7)$$

R – the overall post rating; K – the user influence coefficient; K_i^f – the coefficient of the influence of the i -th factor on the overall rating; $factorsCount_i$ – the number of element ratings on the i -th factor; $totalRatesCount$ – the total number of element ratings by all factors; $avgRate_i$ – the average element evaluation for the i -th factor; $rate_j$ – the j -th element of the estimation scale in numerical form; $rateK_j$ – the coefficient of the significance of the j -th element of the scale; $totalRate_i$ – the maximum sum of ratings for a certain element on the i -th factor.

The integration of such a combined assessment system is simpler to consider using the Facebook social network as an example. The network already has a simplified system of factors for evaluating records (in the form of emoji). You can use a 5-point or 10-point rating system with stars (or other images) to enter a segmented scale for each factor. In this case, the overall assessment composition will continue (use of graphic images). The disadvantage of such a system is the increase of the number of user actions to evaluate an item. Depending on the device, the number of actions displayed will increase by one for the users of the web interface of personal computers (when the scale for each factor in the form of a popup card is displayed) and two for mobile users (opening of the scale of the factor and the choice of an option).

The integration of such a system into the Twitter network, which was discussed earlier, needs to be done gradually, as this network, unlike Facebook, has neither a multivariate system nor a segmented scale of evaluation. The first is to integrate the system of factors based on graphic images (like those already known to the user evaluation options). With enough level of development by the users of the factor system, you can enter a rating scale for each factor (like Facebook, as described above). In this case, the role of likes becomes minor, they should be revisited for use only as a saving of the elements to the chosen bookmarks.

Integration of such a system involves possible risks during operation. The most likely of these is the increase in the complexity of the assessment system and the lack of instant evaluation with one action. On the one hand, the complexity of evaluating a multi-factor system using a segmented scale causes the user to spend more time to make a decision, but on the other hand, such a system provides the user with much greater freedom of action and options, which, of course, brings the social network to a new level of progress in terms of searching, filtering and sorting information.

Depending on the number of factors and the differentiation of the scale, the number of options for evaluating one element for the user increases by 10-50 times, which will take into account the preferences of the users when giving recommendations (including advertisements) and provide the user with an expression of their opinion when evaluating the item much more concretely and clearer than usual “like it”.

Conclusion

The main problem of popular social networks in data retrieval and filtration is the lack of an extensive system of evaluation of information elements that allows the user to evaluate the records more qualitatively and specifically.

The paper considers the main evaluation options used in popular social networks. Using Twitter as the example of the social network, the features were analyzed of the use of the evaluation system in the form of likes and possible options for its improvement. The problems connected with the sorting of data during search and giving personal recommendations to the user were analyzed too.

It has been determined that the introduction of a segmented scale will help users to evaluate information more variably and critically (like the evaluation of movies in online cinemas and goods in online stores). On the other hand, it was found that the multi-factor system greatly improves the qualitative assessment of the element, allowing users not only to express their preferences for some information element, but also to indicate the reason for such an assessment.

Determined two development options of the evaluation system, it has been determined that a combination of a segmented scale with a multi-factor system increases the number of possible evaluation options for the user by tens of times. This allows to optimize search results using variables scale factors and increase the number of possible options for filtering data using factors and their impact coefficients on the overall rating of the information elements.

Using Facebook and Twitter as examples of the social networks there were considered the possible ways of integrating the combined assessment system, possible risks and ways to eliminate them. Thus, the emphasis is placed on the fact that a social network requires a more sophisticated evaluation system, which will help to sort information qualitatively with the increase in the amount of data in the network.

Subsequent studies include a detailed analysis of the integration of such a system of evaluation and creating a new social network based on it. In addition, one of the key tasks is to get feedback from users on the complexity of the combined rating system to determine the main areas of information evaluation development to optimize the sorting and filtering of big data [12], whose number is continuously increasing.

Conflict of interest

The authors declare no Conflict of interest.

References

- [1] *Wasserman S.* Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences) // Cambridge University Press. – 2012. – 857p.
- [2] *Kadushin C.* Understanding Social Networks: Theories, Concepts, and Findings // Oxford University Press. – 2011. – 264p.
- [3] *Easley D, Kleinberg J.* Networks, Crowds, and Markets // Cambridge University Press. – 2010. – 744p.
- [4] *Cha M., Haddadi H., Benevenuto F., & Gummadi P. K.* Measuring User Influence in Twitter: The Million Follower Fallacy // Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. – 2010. – P. 10-17.
- [5] *McCulloh I., Armstrong H., Johnson A.* Social Network Analysis with Applications // Wiley. – 2013. – 320p.
- [6] *Tsvetovat M., Kouznetsov A.* Social Network Analysis for Startups: Finding Connections on the Social Web // O'Reilly Media. – 2011. – 192p.
- [7] *Golbeck J.* Analyzing the Social Web // Morgan Kaufmann. – 2013. – 290p.
- [8] *Stieglitz S., Mirbabaie M., Ross B., Neuberger C.* Social media analytics – Challenges in topic discovery, data collection, and data preparation // International Journal of Information Management. – 2017. – P. 156-168.
- [9] *C. Amit, J. Van Hillegersberg.* Exploring the Impact of Socio-Technical Core-Periphery Structures in Open Source Software Development, // Journal of Information Technology. – 2010. – P. 216-229.
- [10] *T. Kesava, G. Mohan Ram.* A Novel Sorting Algorithm for Data Analysis // International Journal of Scientific Research in Computer Science, Engineering and Information Technology. – 2018. – P. 1454-1456.
- [11] *S. Stieglitz, L. Dang-Xuan, A. Bruns, C. Neuberger.* Social Media Analytics – An Interdisciplinary Approach and Its Implications for Information Systems // Business & Information Systems Engineering. – 2014. P. – 89-96.
- [12] *Yang M., Kiang M., Shang W.* Filtering big data from social media – Building an early warning system for adverse drug reaction // Journal of Biomedical Informatics. – 2015. – P. 230-240.

The article was delivered to your editory stuff on the 22.05.2019