

INFORMACION TECHNOLOGY. AUTOMATION

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ. АВТОМАТИЗАЦІЯ

UDC 004.9.5

O. Lopakov,

L. Prokopovych, DSc, Assoc. Prof.,

D. Solodkyi

Odessa Polytechnic National University, 1 Shevchenko Ave., Odessa, Ukraine, 65044; e-mail: lopakov.o.s@op.edu.ua

ANALYSIS OF TECHNOLOGIES FOR PERSONALIZED CONTENT RECOMMENDATIONS ON THE INTERNET

O.S. Lopakov, L.V. Prokopovych, D.M. Solodkyi. Аналіз технологій рекомендацій персоналізованого контенту в Інтернеті. Впровадження нових технологій у соціально-інформаційний сектор Інтернету спричинено, у першу чергу, зростанням конкуренції за увагу користувачів. Щоб утримати увагу людини у своїй соціальній мережі необхідно постійно пропонувати лише цікавий, різноманітний контент, і для різних користувачів він має відрізнятися. Для вирішення цих практичних завдань розробляються технології рекомендацій персоналізованого контенту. Актуальність теми цієї статті зумовлена широким застосуванням в Інтернеті технологій рекомендацій персоналізованого контенту та появою публікацій (аналітичних матеріалів, статистики), в яких йдеться про недоліки цих технологій. Мета дослідження – аналіз сучасних технологій рекомендацій персоналізованого контенту з виявленням позитивних та негативних наслідків їх практичного застосування. В роботі проаналізовано алгоритми ранжування контенту, до яких вдаються різні соціальні мережі, відео-хостинги та інші медіа ресурси. Надано теоретичне обґрунтування ризиків, пов'язаних із створенням «інформаційних коридорів». В результаті дослідження встановлено, що ці технології здатні полегшити користувачам процес пошуку та споживання інформації через: рекомендації тільки цікавого контенту; персоналізації реклами, полегшення таргетінгу; зменшення кількості шкідливої інформації або травмуючого контенту. Разом з тим, ці технології призводять й до негативних наслідків: проблеми цензури; інформаційного обмеження, або «інформаційних коридорів»; вплив на концентрацію уваги та спосіб мислення. Наявність цих проблем не дозволяє назвати актуальні технічні досягнення у сфері, що доліджувалася, ультимативними. Втім, це стосується будь-яких технологій, які потребують не лише визнання їх користі, а й ретельного аналізу недоліків задля вчасного їх виправлення та встановлення принципів відповідального використання цих технологій.

Ключові слова: соціальні мережі, реклама, новинні «стрічки», алгоритми ранжування контенту, «інформаційні коридори»

O. Lopakov, L. Prokopovych, D. Solodkyi. Analysis of technologies for personalized content recommendations on the internet.

The introduction of new technologies in the social information sector of the Internet is primarily due to increasing competition for the attention of users. In order to keep a person's attention on your social network, you only need constantly to offer interesting, diverse content, and it should be different for different users. To solve these practical problems, technologies for personalized content recommendations are being developed. The relevance of the topic of this article is due to the widespread use on the Internet of personalized content recommendations and the emergence of publications (analytical materials, statistics), which address the shortcomings of these technologies. The purpose of the study is to analyze modern technologies of personalized content recommendations with the identification of positive and negative consequences of their practical application. The paper analyzes the algorithms of content ranking, which are used by various social networks, video hosting and other media resources. Theoretical substantiation of risks connected with creation of "information corridors" is given. The study found that these technologies could facilitate the process of users to search and consume information through recommendations only interesting content; personalization of advertising, facilitation of targeting; reducing the amount of harmful information or traumatic content. At the same time, these technologies lead to negative consequences: problems of censorship; information restrictions, or "information corridors"; influence on concentration and way of thinking. The presence of these problems does not allow calling the current technical achievements in the field under study, the ultimate. However, this applies to any technology that requires not only recognition of their benefits, but also careful analysis of shortcomings in order to timely correct them and establish the principles of responsible use of these technologies.

Keywords: social networks, advertising, news "feeds", content ranking algorithms, "information corridors"

Introduction

The market of social networks is developing at a very fast pace. Its obsolete representatives are being replaced by new large companies with global and technological projects, and the share of social media users is growing every day. The process of development of this social and information sphere leads to fierce competition for attention and causes problems that we have not encountered so far. De-

DOI: 10.15276/opu.1.65.2022.10

© 2022 The Authors. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

spite significant improvements in social networks as services, and their acquisition of undoubtedly useful mechanisms for filtering harmful content and its personalization, the accompanying shortcomings do not allow to call the modern vector of development uncompromising. Therefore, it is very important to closely monitor, find and eliminate problems that always arise in parallel with the rapid technological development.

Formulation of the problem

Social networks have rapidly and densely entered the daily lives of many Internet users. People use social media to stay in touch and interact with friends, family and different communities. Businesses use social services to market and promote their products, as well as to identify customer problems. Various tools help companies track, measure and analyze the attention they receive from social media, including brand perception and customer understanding. However, along with these useful features, social networks pose a number of problems, including the need to protect users' personal data [1], as well as their digital profiles and accounts [2].

Personalized content recommendation technologies are no exception: they also provide useful opportunities to those who use them (subjects) and even to whom they are aimed (users-objects), but a more detailed analysis reveals the problems of different character.

The purpose of the study is to analyze modern technologies of personalized content recommendations with the identification of positive and negative consequences of their practical application.

Presenting main material

The introduction of new technologies in the social information sector is primarily due to increased competition for user attention. In order to keep a person's attention on your social network, you only need to constantly offer interesting, diverse content, and it should be different for different users. After all, the main source of income for companies that own social media resources is the sale of advertising. Therefore, the more time people spend on the consumption of information on any site / application, the more advertising they will see and, consequently, will bring more profit. Preference algorithms are used not only to hold the attention of consumers, but also to better personalize the advertising itself. It is now possible to set up an advertising campaign in such a way that the advertising message will appear only in the leads of potential customers, thus increasing conversions many times over.

Today, almost all social media uses "feeds" that allow users to scroll through content. Social media companies use special algorithms with elements of artificial intelligence technologies that take into account and analyze user profile data, interests, activity, etc., to determine the content to offer to the client and the order in which this content will be displayed. The feed should include content from subscribers, popular or recommended content, and information from organizations that pay for advertising. The first news aggregators were much simpler and published only information in chronological order, and did not have a strong system of recommendations.

Looking more closely at content ranking systems, such as Facebook, we can determine that the ranking system does not consist of a single algorithm. Instead, several layers of machine learning models are used to predict what the user wants to see. The ranking system first collects candidate posts for each user, including those shared by their friends, groups, or pages since they last logged in. He then gives each post an assessment based on various factors, such as who shared the content and how it corresponds to what the user usually interacts with. Then the lightweight model narrows the pool of shortlist candidates [3]. This allows more powerful neural networks to give each remaining post an estimate that determines the order in which they are placed. Finally, the system adds contextual features such as diversity rules to ensure that the news feed has diverse content. The whole process takes time to open the Facebook application. This speed was achieved with the help of a complex structure of the ranking system, which in real time can process a large amount of information (about 1000 posts per day per user) and give the relevant content of the "tape" [4]. The general view of the algorithm of the Facebook news feed is illustrated in Fig. 1 [5].

Next we will consider the algorithm for creating personalized content on the most famous video hosting, namely YouTube. The general scheme of operation of neural networks of YouTube recommendations is presented in Fig. 2 [6].

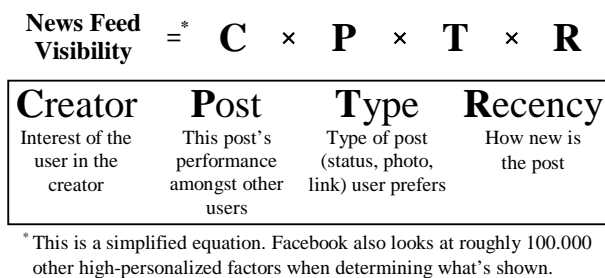


Fig. 1. Simplified scheme of the Facebook news feed algorithm

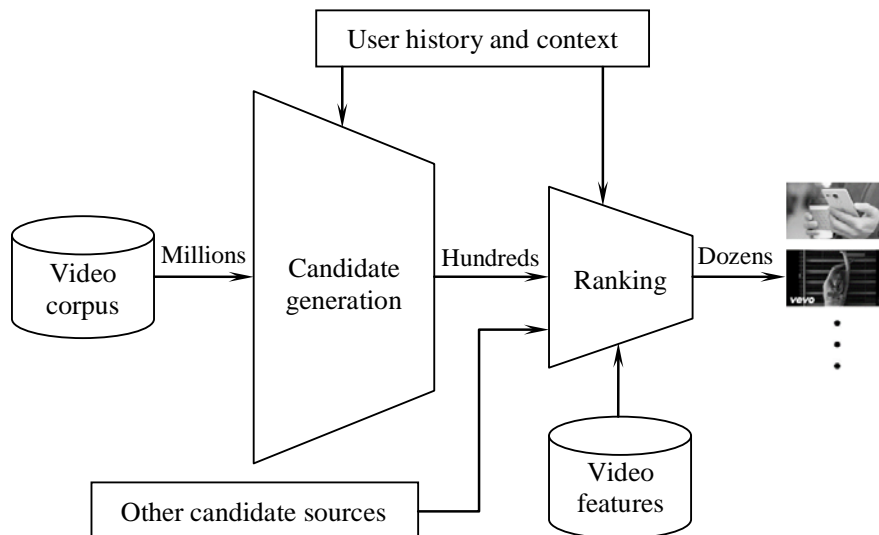


Fig. 2. Scheme of neural networks of YouTube recommendations

When creating a personalized feed, the service focuses on 80 billion information objects (“signals”), which include standard criteria (clicks, viewing time, content links, likes and dislikes), and survey results. Surveys help determine “valuable viewing time” – time that users actually find useful. Based on the answers received, machine learning predicts possible estimates. To test the accuracy of these predictions, YouTube excludes some ratings already provided by users from the training task. This allows you to check whether the data obtained by the system correspond to the real answers. These algorithms are more suitable for music and entertainment content. In the case of news and information videos, additional tools are connected. In particular, fact-checking. The platform separates verified content from “borderline” content using classifiers. The quality of information is assessed by experts from around the world who study according to detailed public rules of video service.

In 2018, the director of YouTube products said that 70% of the time spent watching on YouTube is spent watching videos recommended by the algorithm.

The market for artificial intelligence in social networks is developing so rapidly that experts estimate that by 2023 it will grow to more than 2.3 billion USD [6].

The main advantages of the technological trend to create personalized content are:

- recommendations only for interesting content;
- personalization of advertising, facilitated targeting;
- reducing the amount of harmful information or traumatic content.

However, there are a number of significant shortcomings.

First of all, we should consider the problem, which was formulated in 2010 by Eli Pariser called “Filter bubble” (Engl. Filter bubble, sometimes an information bubble). A Filter bubble is intelligent isolation

that can occur when web services (including mobile applications) use algorithms to selectively accept information that a user would like to see and then provide information to the user according to this assumption. Websites make these assumptions based on user-related information, such as past preferences, browsing history, search history, and location. For this reason, websites are likely to provide only information that will follow the user's past activities. Thus, the filter bubble can lead to significantly less contact with conflicting points of view, resulting in the user becoming intellectually isolated [7].

For example, user A believes that the Android operating system is the best among mobile systems. User B, on the other hand, believes that iOS is the best. Then each of them will watch, read and share only those posts/videos that support their opinion. Social networking algorithms will quickly process the input and start recommending only the content and community of people that the account owner likes, thus not allowing to assess the situation objectively.

In our opinion, this process is similar not so much to the formation of “information bubbles” as to the creation of “information corridors” through which the user is guided by algorithms. It is clear that such “corridors” can be created not only to offer users certain goods and services, but also to impose certain ideas, cutting off from the rest of the information space.

In other words, following the conditional “white rabbit”, a person may not even suspect that there are also black rabbits, as well as butterflies, hedgehogs and other interesting creatures (Fig. 3). This practice significantly reduces a person's ability to think critically and, as a consequence, leads to isolation and radicalization of certain social groups.

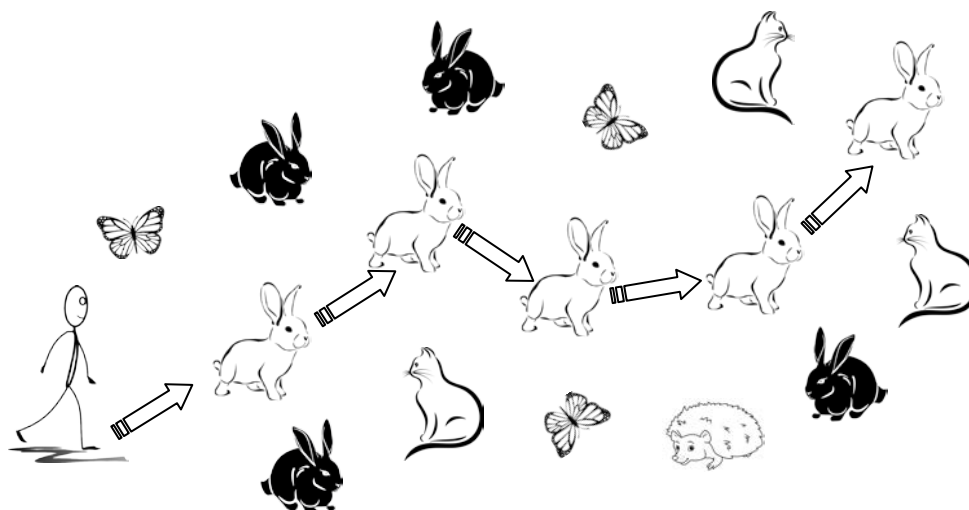


Fig. 3. The effect of the “information corridor”: walking behind the conditional “white rabbit”, a person may not even suspect that there are also black rabbits, as well as butterflies, hedgehogs and other interesting creatures

However, the problem of personalization of content concerns not only those who consume content, but also those who create it. The concept of “shadow ban” was formed. Many people mistakenly believe that there are special algorithms for “punishing” users who do not use the tools of the social network in good faith, for example, cheating such indicators as “Likes”, comments, etc. None of the companies that own social networks have confirmed that they have a special punishment algorithm. In fact, the phenomenon of cutting the coverage of posts and entire accounts exists and arises due to the complex process of content filtering and personalization. Representatives of Instagram said that recently the social network uses a set of algorithms to form a news feed and even issue hashtags.

Instagram algorithms work on the following principles:

- algorithms try not to bore users with frequent display of posts from the same profile;
- rankings and coverage are influenced by reactions to the content (the better the audience responds, the better for the creator);
- news feeds are personalized – everyone sees them as content, to which they will react with high probability [8].

That is, social networking algorithms are not intended to intentionally reduce someone's reach. They simply offer users what they will be interested in (according to those who provide this service). This complex process of making recommendations makes it easy to hide "unwanted" content, both intentionally and directly by instructing algorithms not to take it into account and not to recommend it to users. Thus, the material complained about, for example, by the government of a country or, the material that the neural network considers harmful will be "pessimistic" (less recommended or completely removed from the search). In general, the tool for pessimizing and filtering content is useful, performs its functions well to remove spam and really harmful information, but may be wrong or be configured to hide the necessary content, creating a threat of censorship and artificial perception of reality.

Attention should also be paid to, so to speak, the "side effect" of the widespread use of recommendation algorithms.

Content personalization algorithms are used not only in social networks, but also in music services, news aggregators, streaming services for movies and serial, etc. However, they are most popular on social media: according to the analytical agency "Datareportal" 93.4% of people who use the Internet – have accounts on social networks and are in them for almost 2.5 hours a day [9]. But the desire of companies to keep users' attention on their platform for as long as possible, by constantly recommending interesting and easy-to-consume content, has led to a sharp decline in people's concentration. A 2000 study by Microsoft found that the average duration of a person's concentration was 12 seconds. In 2015, the same study found that this figure decreased to 8 seconds. This significantly complicates the process of working with the analysis of large data sets and leads to more frequent attention deficit disorder [10].

Conclusions

In today's world, social networks are no longer just a means of communication, but also a complex socio-cultural phenomenon that has penetrated almost all spheres of human life. As a result of their global spread, social networks have developed and changed rapidly, but in tandem with advanced technologies of artificial intelligence, algorithmization and machine learning, they have become important mechanisms influencing the culture, psychology and even physiology of people. Unfortunately, these impacts are not always beneficial, and along with progress, they create new problems for people and society.

Such ambiguous technologies include algorithms for recommending personalized content on the Internet. Research has shown that these technologies can make it easier for users to find and consume information through:

- recommendations of only interesting content;
- personalization of advertising, facilitation of targeting;
- reducing the amount of harmful information or traumatic content.

However, these technologies lead to negative consequences:

- censorship problems;
- information restrictions, or "information corridors";
- influence on concentration and way of thinking.

The presence of these problems does not allow to call the current technical achievements in the field under study, the ultimate. However, we believe that any development of technology is important for the formation of a better reality. We should not only turn a blind eye to the problems associated with technological progress and focus only on the invention and development of new technologies, but also work to eliminate the identified problems, which are always the result of rapid development.

Література

1. Прокопович Л.В., Лопаків О.С., Солодкий Д.М. Шляхи підвищення захисту персональних даних користувачів соціальних мереж. *The scientific heritage*. 2021. № 65. С. 3237. DOI: 10.24412/9215-0365-2021-65-32-37.

2. Matthew Hudson. What is social media? Definition and examples of social media. 2020. URL: <https://www.thebalancesmb.com/what-is-social-media-2890301>. (дата звернення 10.01.2021).
3. Thomas Macaulay. Here's how AI determines what you see on the Facebook News Feed. 2021. URL: <https://thenextweb.com/news/heres-how-ai-determines-what-you-see-on-facebook-news>. (дата звернення 10.12.2021).
4. Akos Lada, Meihong Wang, Tak Yan. How machine learning powers Facebook's News Feed ranking algorithm. 2021. URL: <https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/>. (дата звернення 11.12.2021).
5. Josh Constine. How Facebook News Feed Works. 2016. URL: <https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/>. (дата звернення 12.12.2019).
6. Paul Covington, Jay Adams, Emre Sargin. Deep Neural Networks for YouTube Recommendations. 2016. URL: <https://research.google/pubs/pub45530.pdf>. (дата звернення 12.01.2022).
7. Markets and Markets. AI in Social Media Market by Technology (Deep Learning & Machine Learning, and NLP), Application (Sales & Marketing, Customer Experience Management, and Predictive Risk Assessment), Component, Enterprise Size, End-User, and Region - Global Forecast to 2023. 2018. URL: <https://www.marketsandmarkets.com/Market-Reports/ai-in-social-media-market-92119289.html>. (дата звернення 12.01.2022).
8. Daily Dish. The Filter Bubble. 2010. URL: <https://www.theatlantic.com/daily-dish/archive/2010/10/the-filter-bubble/181427/>. (дата звернення 12.01.2022).
9. Adam Mosseri. Shedding More Light on How Instagram Works. 2021. URL: <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>. (дата звернення 10.01.2022).
10. Simon Kemp. Digital 2022: global overview report. 2022. URL: <https://datareportal.com/reports/digital-2022-global-overview-report>. (дата звернення 12.02.2022).
11. Kevin Mcspadden. You Now Have a Shorter Attention Span Than a Goldfish. 2015. URL: <https://time.com/3858309/attention-spans-goldfish/>. (дата звернення 01.12.2019).

References

1. Prokopovich, L.V., Lopakov, O.S., & Sweet, D.M. (2021). Ways to increase the protection of personal data of social network users. *The scientific heritage*, 65, 3237. DOI: 10.24412/9215-0365-2021-65-32-37.
2. Matthew Hudson. (2020). What is social media? Definition and examples of social media. Retrieved from <https://www.thebalancesmb.com/what-is-social-media-2890301>. (Last access 10.01.2021).
3. Thomas Macaulay. (2021). Here's how AI determines what you see on the Facebook News Feed. Retrieved from <https://thenextweb.com/news/heres-how-ai-determines-what-you-see-on-facebook-news>. (Last access 10.12.2021).
4. Akos Lada, Meihong Wang, & Tak Yan. (2021). How machine learning powers Facebook's News Feed ranking algorithm. Retrieved from <https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/>. (Last access 11.12.2021).
5. Josh Constine. (2016). How Facebook News Feed Works. Retrieved from <https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/>. (Last access 12.12.2019).
6. Paul Covington, Jay Adams, & Emre Sargin. (2016). Deep Neural Networks for YouTube Recommendations. Retrieved from <https://research.google/pubs/pub45530.pdf>. (Last access 12.01.2022).
7. Markets and Markets. AI in Social Media Market by Technology (Deep Learning & Machine Learning, and NLP), Application (Sales & Marketing, Customer Experience Management, and Predictive Risk Assessment), Component, Enterprise Size, End-User, and Region - Global Forecast to 2023. (2018). Retrieved from <https://www.marketsandmarkets.com/Market-Reports/ai-in-social-media-market-92119289.html>. (Last access 12.01.2022).
8. Daily Dish. (2016). The Filter Bubble. 2010. Retrieved from <https://www.theatlantic.com/daily-dish/archive/2010/10/the-filter-bubble/181427/>. (Last access 12.01.2022).
9. Adam Mosseri. (2021). Shedding More Light on How Instagram Works. Retrieved from <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>. (Last access 10.01.2022).

-
10. Simon Kemp. (2022). Digital 2022: global overview report. Retrieved from <https://datareportal.com/reports/digital-2022-global-overview-report>. (Last access 12.02.2022).
 11. Kevin Mcspadden. (2015). You Now Have a Shorter Attention Span Than a Goldfish. Retrieved from <https://time.com/3858309/attention-spans-goldfish/>. (Last access 01.12.2019).

Лопаків Олексій Сергійович; Oleksii Lopakov, ORCID: <https://orcid.org/0000-0001-6307-8946>
Прокопович Лада Валеріївна; Lada Prokopovych, ORCID: <https://orcid.org/0000-0001-8636-9172>
Солодкий Денис Миколайович; Denys Solodkyi

Received February 13, 2022

Accepted March 26, 2022