DOI: 10.32347/2412-9933.2024.57.77-82
UDC 004.62:004.8

Stetsyk Oleksii
Phd student, Department of Information Technology Design and Applied Mathematics,
https://orcid.org/0000-0002-1061-0465
Kyiv National University of Construction and Architecture
Terenchuk Svitlana
PhD, Associate Professor Department of Information Technology Design and Applied Mathematics,
https://orcid.org/0000-0001-6527-4123
Kyiv National University of Construction and Architecture

MODEL DEVELOPMENT OF THE SYSTEM FOR AVOIDING ECHO CHAMBERS IN SOCIAL NETWORKS

Abstract. *This article examines the problems that appear in the process of the development of social networks. The focus of the study is on the echo chamber effect in various social networks. It turned out that the echo chamber effect is the result of the work of a deep neural network, which analyzes the interests and priorities of each social network user, checks other similar posts and priorities of other users, and then forms a "circle" of like-minded people for each user. Thus, the main drawback of the standard content generation algorithm is the selection of only publications and comments that support the user's position. The goal of the study is a software architecture to solve the echo chamber problem in highly loaded social networks in real-time. The main idea was to cluster opinions on controversial topics and include user opinions from different clusters in user content. Data on controversial topics is collected from news, scholarly articles, publications, and comments with hashtags. The collected topics are clustered using the K-means algorithm and the "elbow" method is used to find the optimal number of clusters. The result of the clustering of opinions by topic is provided as input data for the generation of a news feed by a recurrent neural network. The system uses Kafka as a message broker between microservices and AZURE blob storage for storing publications and comments. Both solutions support high scalability.*

Keywords: clusterization; echo chambers; Kafka; K-means; PostgreSQL; recurrent neural network; scalability

Introduction

First social networks which were created for connecting individuals based on real-world relationship, appeared only 2 decades ago, namely: Myspace (2003), Facebook (2004), WhatsApp (2009), Telegram (2013). Now social networks are evolving rapidly nowadays, transcend geographical boundaries, are connecting people all over the world, improving communication and connection.

Current social media allows reach number of features like [1]:

– Multimedia (photos, videos, livestreaming);

– Direct communication channel (video calls, private messaging, group chats);

– Personalized recommendation algorithm (in order to show only specific content for individual);

– Interactive features (quizzes, polls, gamification elements and e-commerce integration).

It's nice to see such a rapid development of social medias, with can generate news feed with some concrete objectives. Also platform can show sponsors ads or prioritize paid posts. Order to generate a content for user, actions like sharing, liking, disliking, saving and commenting posts are tracked. A lot of different information are stored about user. Deep neural network [2; 3] analyses the interests and priorities of each user, checks other similar publications and priorities of other users of the social network, and then forms a "circle" of like-minded people for each user.

This algorithm works because user attention are concentrated to similar content he saw previously. That's how echo chamber appears also. This algorithm really helps to capture attention of the users, which is one of the basic business values of social networks, but if ads or paid post is in priority then main objective of news feed generation is to maximize amount of time that user spends on current social network.

So with these improvement, new issues appeared:

– Problem with privacy issues (social medias collect user big amounts data, which can be used to target ads and even identity theft) [4];

– Problem with spam and propaganda (social medias can be used for political manipulation) [5];

– Harassment (anonymity and lack of accountability can lead to bullying) [6].

– Problem with echo chambers (social media create environments where users only see information that confirms their existing beliefs, and this reinforces polarization among people) [4].

– Problem with mental health (excessive use of social media can cause anxiety, feeling of inadequacy and depression) [7].

Major social networks have features like blocking some users, or some harmful content such as propaganda, spam, not true, but it won't help to fight with echo chambers. Therefore, solving the problem of echo chambers is urgent.

Analysis of Recent Research and Publications

Comparison of echo chambers effect in different social network, like Facebook, Twitter, Reddit, was done in [8] Authors of this article was compared 100 million pieces of text that contained discussions about controversial topics like gun control social help, abortion, vaccination in result it has found that Facebook is more influenced by echo chambers effect then Reddit.

Authors assumed of [9] suggested that the reason is because users have higher degree of control of news feed in Reddit platform [10]. That's why in the system under development, we will provide possibility for user to configure level of exposure for different opinion of controversial topics.

In research [11] authors have found that in Twitter users usually are seeing political opinion that supports their own. And users who wants to share different opinions over political spectrum will pay the price of bipartisanship. Also, authors found that when topic is not contentious, then echo chamber doesn't appear. That's why in current article we will concentrate only on topics that are contentious. When user would see post from topic that he/she is interested in, but with different opinion, he will be able to communicate with authors of the post, and this will create a connection between the clusters.

In [12] are checking problem of polarization in Twitter. Authors of this article have found that edges between communities is likely to be deleted. And edges among communities forms in 3-4 times more often than between communities. Because of this, authors concludes that Twitter becomes more polarized. In current article we will provide ways to restore those deleted nodes, from another echo chambers.

There's an interesting research [13] that concentrate their attention of rumors propagation in echo chambers. In research authors have found that top 10 percent of echo chambers rumors contribute to propagation of 24 percent of rumors in whole social media. Authors also have found that 36 percent of retweets for these rumors are done by members of same echo chamber in which rumor has appeared.

Thus, the conducted analysis of recent research and publications showed that echo chambers appear when the number of connections between users of the same cluster is huge, and connections between users of different clusters are rare.

There are several main reasons why echo chambers appear [14]:

1. Algorithm of content generation can prioritize content that confirms existing beliefs and limits exposure to different perspectives and this reinforce intolerance to opposite belief.
2. Chats that can spread misinformation quite quickly.
3. Private groups can support only one point view for some topics.
4. Lack of habits for person to check quality of content, and facts especially comments from bots and deepfakes from bots.

This article will be focused on improvement of algorithm of content generation.

Goal and Tasks of the Publication

Goal of the article is to provide effective software architecture for solving echo chamber issue on high load social networks in real time.

To achieve the goal, it is necessary:

1. Explore the features for working with echo chambers.
2. To improve of news feed generation algorithm by using machine learning and neural networks.
3. To develop of effective scalable architecture, which allows will generated content for user with low latency.

Presenting Main Material

Several features to deal with echo chambers

In this article below proposed several feature to deal with echo chambers.

Detailed design and architecture of the System for Avoiding Echo Chambers in Social Networks (SAECSN) will be proposed for the first two that seems the most important for us, and others would be dealt with in future works.

First feature is to prefer diversified content over relevance. Even if other viewpoints don't have perfect match with user preferences, we should prioritize such responses. One way to implement this is to give such posts bigger priority.

Second feature is to get popular topics with diverse perspectives on topic that current user are interested in. Way to implement this is to aggregate content from diverse sources on this topic and present it to the users. So he or she will have better view about some topic.

Third feature to improve, is to promote healthy discussions. Statistical data is not gathered, but it's rare in social media that under the post after discussion somebody has changed his mind. Toxic speech should be deprioritized, and thoughtful, respectful comments should be prioritized. Comments like "Oh, now I see you point", "I agree with this part, but disagree with this", or "I understand your opinion, but deeply disagree with it" should be prioritized.

Fourth feature is to provide context for each post and conversation. Short summary also could be a way to improve this.

Improvement of news feed generation algorithm by using machine learning and neural networks

To build SAECSN like one above, it is need to gather different type of data like: posts and comments that user liked previously; connection with other users that include following, being followed; number of likes, repost, comments under another user profile; topics that are interesting for current user.

In database we should have tables 1 – 5.

Table 1 – ***Users***

users	
user_id	bigint
email	str
created_on	date
country	str
city	str

Table 2 – ***Posts***

posts	
post_id	bigint
user_id	bigint
hashtag	str
data_link	str
likes	bigint
shares	bigint

Table 3 – ***Comments***

comments	
comment_id	bigint
post_id	bigint
user_id	bigint
likes	bigint
created_on	date
data_link	str

Table 4 – ***Users comment history***

user_comment_history	
user_id	bigint
comment_id	bigint
comment_data_link	str

Table 5 – ***User relationship***

relationship	
following_id	bigint
followed_id	bigint
created_at	date
avg_likes	float
overall_likes	bigint
avg_shares	float
overall_shares	bigint
avg_comments	float
overall_comments	bigint

Overall data above is good for feature engineering for standard feed generation algorithm.

Recurrent neural network (RNN) is chosen for news feed generation, as it's useful for predicting sequence and would be optimal to forecast user behavior (Fig. 1).

In order to be able to show posts for user from different spectrum of view on some topic, additional data should be gathered. Firstly, posts should be clusterized by contentious topic and secondly, data in contentions topic should be clusterized over different opinions on this topic.

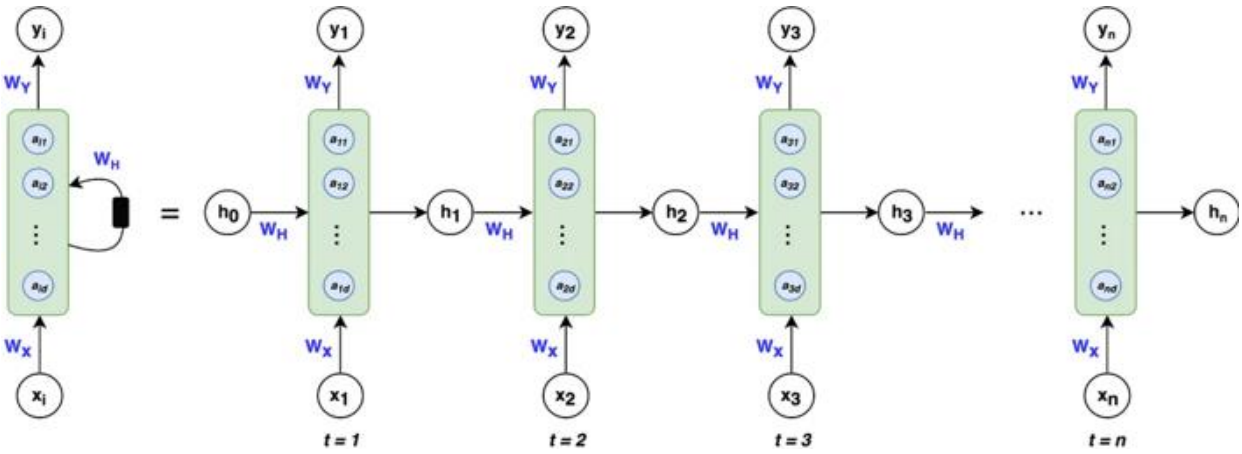

Figure 1 – RNN calculation algorithm [15]

First of all list of contentious topics in politics (gun controls, Russia's war against Ukraine, Roe and Wade laws about abortion, Europe immigration topics), sport (who is better Messi or Ronaldo, Djokovic, Federer or Nadal), economics (market vs planned), ecology, fashion trends should be chosen manually. After that texts for these topics should be collected from forums, articles, papers, posts, and comments with hashtags.

Secondly clusterization by different opinions on every concrete subtopic are needed. For example, texts about gun control should be gathered in two clusters one for those who support this and others who are against it.

K-means algorithm [16] is used here for clusterize data over spectrum of views on different topic. Elbow method [17] for K-means algorithm is used here to determine optimal number of clusters (opinions).

Result of K-means clusterization will be used as an input for RNN for news feed generation. In this way will makes sure that user will not receive news that are identical with his current opinion over some topic. This also will help to exposure current users to others, who holds different opinions and possibility to have a dialogue with them.

System architecture design

Fig. 2 architecture of SAECSN, that is developing is presented.

Supposing system for avoiding echo chambers in social networks will have hundreds of millions of users, then it architecture should be highly scalable.

First part in gathering information would be to receive it through message broker like Kafka [18], or RabbitMQ [10]. That will allow not to have excessive load on consequent services.

After it goes to feature microservice. If request is create new post, system will store post metadata and post itself in different locations. Metadata will be stored in relations database PostgreSQL.

Relational database is chosen, as in future joins with other tables is needed and non-relational databases don't allow to do joins effectively [19].

If in future problems with scalability appears, sharding by user_id can be used in order to solve this problem.

If problem with queries latency appear, database indexes based on B-tree or Log-Structured Merge (LSM) tree can be used, depending on which queries we have more, read or write queries. B-trees are popular solutions for read heavy queries and LSM-trees are popular solutions for write heavy queries. Post data itself like images, audio, or just text is stored in Azure blob storage. This is a popular solution to store raw data and it's highly scalable [20].

From feature microservice post also will be passed to topic clusterization neural network service to clusterize posts by topics and to separate this by opposite and different opinions. Feature vector will be created for each post as an output from topic clusterization neural network. Later in the flow it will be given as an input to ranking neural network service.

Ranking neural network service will give priority to all posts by the users, including posts that show different from current user opinion and news feed generation service will return them upon request.

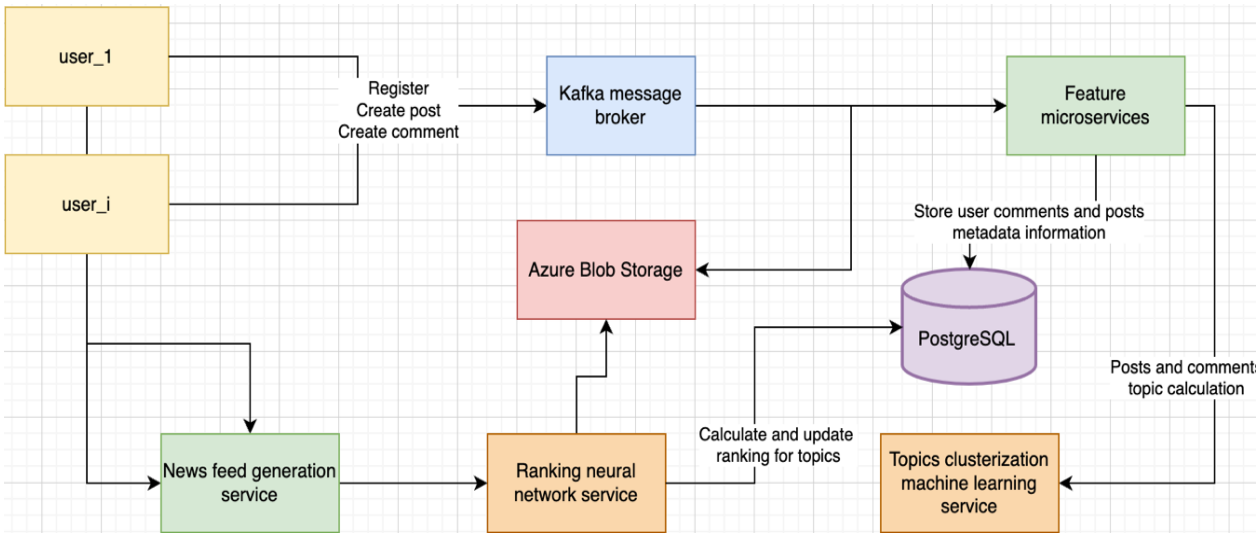

Figure 2 – Architecture of the SAECSN

Conclusion

1. The article provides an effective model for dealing with the echo chamber effect in social media, by improving the algorithm of news feed generation.

Main idea here was to have clusterization using K-means algorithm over opinions in contentious topics and including these results afterward during content generation for user.

2. Combination of recurrent neural network and K-means machine learning algorithm to deal effectively with problem of echo chambers in social networks is proposed.

3. Algorithm of news feed content generation was improved in such a way, that exposes user to wide range of opinion over contentious topic, that he/she is interested in. In this way user will be exposed also to users from different echo chambers of social network, will be able to communicate with such users, and connections between users from different clusters will be formed.

4. In the future it's planned to design several other improvements for content generation algorithm. Several directions in which it can be designed: adding context for topic post, adding summarization of views over some contentious topic, fact checking for data provided in post or comment.

References

1. Adisa, D. (2023, October 30). Everything you need to know about social media algorithms. Sprout Social. URL: https://sproutsocial.com/insights/social-media-algorithms/

2. Lada, A., Wang, M., & Yan, T. (2021, January 26). How machine learning powers Facebook's News Feed ranking algorithm. Core Infra, ML Applications. URL: https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/

3. Koumchatzky, N., & Andryeyev, A. (2017, May 9). Using deep learning at scale in Twitter's timelines. Twitter Engineering Blog. URL: https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.

4. Hetler, A. (2023, January 30). 6 common social media privacy issues. TechTarget. URL: https://www.techtarget.com/whatis/feature/6-common-social-media-privacy-issues.

5. Stetsyk, O., Pasieka, P., & Yeremenko, B. (2023). Designing an Effective System Architecture for Detecting Propaganda and Spam in Social Media News Feed. In 2023 IEEE 4th KhPI Week on Advanced Technology (KhPIWeek). IEEE. DOI: https://doi.org/10.1109/khpiweek61412.2023.10312932.

6. Peebles, E. (2014). Cyberbullying: Hiding behind the screen. Paediatrics & Child Health, 19(10), 527–528. DOI: https://doi.org/10.1093/pch/19.10.527.

7. Pantic, I. (2014). Online social networking and mental health. Cyberpsychology, Behavior, and Social Networking, 17(10). DOI: https://doi.org/10.1089/cyber.2014.0070.

8. Alatawi, F.H., Cheng, L., Tahir, A., Karami, M., Jiang, B., Black, T., & Liu, H. (2021). A Survey on Echo Chambers on Social Media: Description, Detection and Mitigation DOI: https://doi.org/10.48550/arXiv.2112.05084.

9. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. Proceedings of the National Academy of Sciences, 118(9), Article e2023301118. DOI: https://doi.org/10.1073/pnas.2023301118.

10. Moehring, A. (2022). News Feeds and User Engagement: Evidence from the Reddit News Tab. Massachusetts Institute of Technology (MIT) - Sloan School of Management. pp. 4–6.

11. Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Political Discourse on Social Media. In the 2018 World Wide Web Conference. ACM Press. DOI: https://doi.org/10.1145/3178876.3186139.

12. Du, S., & Gregory, S. (2016). The Echo Chamber Effect in Twitter: does community polarization increase? In Studies in Computational Intelligence (pp. 373–378). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-50901-3_30.

13. Choi, D., Chun, S., Oh, H., Han, J., & Kwon, T. (2020). Rumor Propagation is Amplified by Echo Chambers in Social Media. Scientific Reports, 10 (1). DOI: https://doi.org/10.1038/s41598-019-57272-3.

14. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., & Starnini, M. (2020). Modeling Echo Chambers and Polarization Dynamics in Social Networks. Physical Review Letters, 124(4), 048301. DOI: https://doi.org/10.1103/PhysRevLett.124.048301.

15. Khuong, B. (2019, June 24). The basics of recurrent neural networks (RNNs). Towards AI. Retrieved July 25, 2023, URL: https://towardsai.net/p/machine-learning/the-basics-of-recurrent-neural-networks-rnns.

16. Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. Electronics, 9(8), 1295. DOI: https://doi.org/10.3390/electronics9081295.

17. Umargono, E., Suseno, J. E., & Gunawan, S. K. V. (2020). K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. In Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019). DOI: https://doi.org/10.2991/assehr.k.201010.019.

18. John, V., & Liu, X. (2017, April 3). A Survey of Distributed Message Broker Queues. arXiv:1704.00411. https://doi.org/10.48550/arXiv.1704.00411.

19. Stetsyk, O. & Terenchuk, S. (2021). Comparative analysis of NoSQL databases architecture. Management of Development of Complex Systems, 47, 78–82, dx.doi.org\10.32347/2412-9933.2021.47.78-82.

20. Microsoft Azure. (n.d.). What is Azure Blob storage? Retrieved February 27, 2024, URL: https://learn.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction

Стаття надійшла до редколегії 10.03.2024

Стецик Олексій Андрійович
Аспірант кафедри інформаційних технологій проєктування та прикладної математики,
https://orcid.org/0000-0002-1061-0465,
Київський національний університет будівництва і архітектури, Київ, Україна
Теренчук Світлана Анатоліївна
Кандидат фізико-математичних наук, доцент кафедри інформаційних технологій проєктування та прикладної математики
https://orcid.org/0000-0001-6527-4123,
Київський національний університет будівництва і архітектури, Київ, Україна

РОЗРОБКА МОДЕЛІ СИСТЕМИ УНИКНЕННЯ ЕФЕКТУ «БУЛЬБАШКИ» В СОЦІАЛЬНИХ МЕРЕЖАХ

Анотація. *У статті розглянуто можливості сучасних соціальних медіа та проблем, що з'являються в процесі розвитку соціальних мереж. Фокус дослідження спрямовано на ефект «бульбашки» в різних соціальних мережах. З'ясовано, що ефект «бульбашки» є результатом роботи глибинної нейронної мережі, яка аналізує інтереси і пріоритет кожного користувача соціальної мережі, перевіряє інші подібні публікації та пріоритети інших користувачів, після чого формує «коло» однодумців для кожного користувача. Таким чином в соціальних мережах з'являються кластери користувачів, які мають однакову думку щодо суперечливої теми. «Соціальні бульбашки» з'являються, коли кількість з'єднань між користувачами одного кластера величезна, а з'єднання між користувачами різних кластерів рідкісні. Отже, основним недоліком стандартного алгоритму генерації контенту є добірка лише публікацій і коментарів, що підтримують позицію користувача. Метою дослідження є архітектура програмного забезпечення для вирішення проблеми «бульбашки» у високонавантажених соціальних мережах у реальному часі. Основна ідея полягала в кластеризації думок у суперечливих темах і включення у вміст контенту користувача думок користувачів з різних кластерів. Дані щодо суперечливих тем збираються з новин, наукових статей, публікацій і коментарів із хеш тегами. Зібрані теми кластеризуються за допомогою алгоритму K-середніх. Для знаходження оптимальної кількості кластерів використовується метод «ліктя». Потім у різних кластерах користувачів розміщуються тексти з різними думками щодо певних тем. Результати кластеризації думок за темою надаються як вхідні дані для генерації стрічки новин рекурентною нейронною мережею. Отже, алгоритм генерації контенту вибрано як основний компонент вирішення проблеми виникнення «соціальної бульбашки». При цьому запропоновано удосконалити алгоритм генерації стрічки новин за допомогою машинного навчання і нейронних мереж. Також в цій статті запропоновано ефективну архітектуру системи для генерації стрічки новин. У системі використовується Kafka як посередник повідомлень між мікросервісами і Ажур блоб-сховище для зберігання публікацій і коментарів. Обидва рішення обґрунтовані високою масштабованістю. Реляційна база даних PostgreSQL використовується для метаданих публікацій, коментарів, користувачів та зв'язків між користувачами. При розробці архітектури системи передбачено виникнення певних проблем і окреслено шляхи їх вирішення. Наприклад, якщо в майбутньому виникнуть проблеми з масштабованістю, то для їх вирішення можна використати шардинг за user_id; якщо будуть виникати проблеми, спричинені затримкою запитів, то можна використовувати індекси бази даних на основі B-дерева або Log-Structured Merge-дерева, залежно від того, чи буде більше запитів на читання, чи на запис.*

Ключові слова: алгоритм K-середніх; соціальна бульбашка; Kafka; кластеризація; масштабованість; PostgreSQL; рекурентна нейронна мережа

Link to publication

APA Stetsyk, O., & Terenchuk. S., (2024). Model development of the system for avoiding echo chambers in social networks. *Management of Development of Complex Systems*, 57, 77–82, dx.doi.org\10.32347/2412-9933.2024.57.77-82

ДСТУ Стецик О. А., Теренчук С. А. Розробка моделі системи уникнення ефекту «бульбашки» в соціальних мережах. *Управління розвитком складних систем*. Київ, 2024. № 57. С. 77 – 82, dx.doi.org\10.32347/2412-9933.2024.57.77-82.