Recommendation Systems in Social Media - A Systematic Review

João Figueiredo[a], Tiago Nora[b] and João Gaspar[c]

Departamento de Engenharia Informática, Instituto de Engenharia do Porto

ARTICLE INFO

Keywords:
Recommendation systems, Social media, Systematic review, Algorithms, Collaborative Filtering

ABSTRACT

In the era of burgeoning social media platforms, recommendation systems play a pivotal role in enhancing user experience and engagement. This systematic review aims to explore the landscape of recommendation systems in social media, encompassing methodologies, algorithms, and user-centric perspectives. By synthesizing existing literature, this review delineates the evolution of recommendation systems, identifies key challenges, and highlights emerging trends in this domain. Through a comprehensive analysis, it offers insights into the effectiveness, scalability, and ethical considerations of recommendation algorithms in shaping user interactions and content dissemination across diverse social media platforms.

1. Introduction

In an age when social networks are widely used, they are becoming more and more embedded in our daily lives and are becoming a more natural way to communicate, socialize and search for information on a specific subject. With the evolution of technology, particularly the creation of mobile devices, this growth in the number of people using social networks has been strengthened by the increase in their capabilities and the increase in their functionalities.

Many applications used today have some kind of recommendation system, whether it's to recommend a Christmas present, a vacation trip, a restaurant in a certain city or even the next series or movie.

These technologies are embedded in various platforms in different ways, such as blogs, business networks, corporate social networks, forums, micro blogs, photo sharing portals, product recommendation portals, link saving systems, online games, video sharing platforms and virtual worlds (Singh & Pramanik & Dey & Choudhury , 2021).

2. Methods

A systematic review of the literature was conducted to describe the history behind Enterprise Information Systems and it's evolution and possible future trends. The systematic review was prepared according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement.

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) is a methodology used to conduct and register systematic reviews. Several databases were used to carry out the various searches, such as IEEE, ACM, GOOGLE SCHOLAR, Research Gate and Elsevier. A database aggregator called b-on was used to help with the search. It has information on the various databases mentioned above, i.e., it is possible to access articles from these databases on a single platform, but it is also possible to use this site to facilitate the search, since the platform identifies and eliminates repeated publications that may exist, since several articles may have been published by various publishers.

2.1. Eligibility Criteria

The process involved several sequential steps: identification, screening, eligibility assessment, and inclusion of articles for analysis.

2.1.1. Inclusion and Exclusion Criteria

To enhance the research of related articles, it is necessary to define the criteria that should guide the consideration of the found articles. Let's focus on the inclusion criteria for now. An article is considered relevant if it is related to the theme of this article, as well, if it is found using the related keywords that were defined. Regarding temporal analysis, an article is found relevant if it is published at least, after 2000, as articles published in the 20th century are not considered updated regarding social media. In the language domain, only articles published in English or Portuguese are found useful to this work, as they are the only languages that the authors can understand. About the exclusion criteria, articles published before 2000 are not considered, as well as articles that do not have any relation to the theme or the keywords used. Finally, articles that have methodological processes with failures of some types are also excluded.

So, articles were included if they met the following criteria:

- Pertained to recommendation systems in the context of social media.
- Presented original research or systematic review articles.
- Published in peer-reviewed journals or conference proceedings.

And were excluded if they:

- Were not written in English.
- Did not focus on recommendation systems or social media.

1230194@isep.ipp.pt (J. Figueiredo); 1201050@isep.ipp.pt (T. Nora); 1200884@isep.ipp.pt (J. Gaspar)

- Were not primary research or systematic review articles (e.g., editorials, commentaries).

2.1.2. *Research questions*

It is possible to simplify the goals of this article with the following questions, which the available bibliography will help to answer.

- Q1: How are collaborative filtering and content-based algorithms applied in the context of social networks to customize content recommendations?
- Q2: What is the impact of social data mining on improving the accuracy of recommendations in social networks?
- Q3:How is user engagement influenced by the effectiveness of recommendation systems in social networks?
- Q4:What are the main machine learning techniques used to enhance recommendation systems in social networks?
- Q5:How does the interaction between user and system affect the quality of recommendations in social networks?
- Q6:What are the challenges and considerations related to privacy and security in the implementation of recommendation systems in social networks?

2.2. Information sources and search strategy

We extensively searched academic databases like PubMed, IEEE Xplore, ACM Digital Library, Scopus, and Web of Science using a tailored search strategy combining keywords and controlled vocabulary related to recommendation systems, social media, and systematic reviews. Search terms were adjusted for each database, and we only included articles published in English. We applied filters/tags such as Recommender systems, social networks, and collaborative filtering. Our focus was solely on academic papers

2.3. Screening phase

Following the search, duplicate records were removed, and the remaining articles underwent title and abstract screening by two independent reviewers to identify potentially relevant studies. Full-text articles were then retrieved for further evaluation.

Electronic Database	Search Terms
B-ON, ACM, IEEE, Elsevier, Springer Nature, Research Gate	recommender systems; recommendation systems; social media; social networks; algorithms; recommendations

Table 1: Keywords used in the search query

The search query constructed with the terms included:

- recommender systems or recommendation systems and social media or social networks and algorithms and recommendations.

2.4. Selection process (eligibility phase)

Data extraction was performed independently by two reviewers using a standardized data extraction form. The following information was extracted from each included study:

- Study characteristics (e.g., authors, publication year).
- Methodologies and study design.
- Types of recommendation systems examined.
- Key findings and outcomes.

Any discrepancies between reviewers were resolved through discussion, and a third reviewer was consulted if consensus could not be reached.

2.5. Quality assessment and risk of bias

The quality of included studies was assessed using appropriate tools depending on the study design. For primary research studies, we utilized tools such as the Newcastle-Ottawa Scale (NOS) for observational studies and the Cochrane Risk of Bias Tool for randomized controlled trials. For systematic reviews, the Assessment of Multiple Systematic Reviews (AMSTAR) tool was employed.

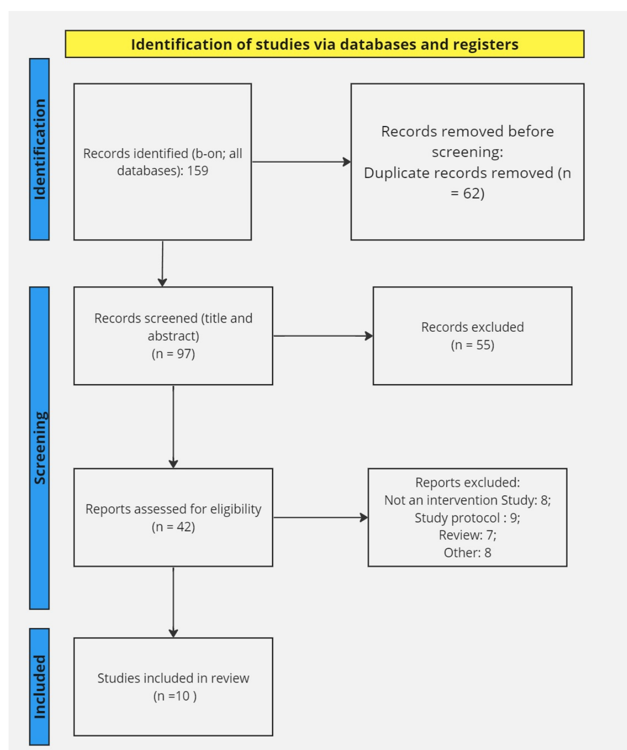

Figure 1: PRISMA flow diagram illustrating the article selection process.

2.6. Data collection process

A narrative synthesis approach was used to summarize the findings from the included studies. Themes and patterns identified across the studies were synthesized to provide insights into the landscape of recommendation systems in social media.

3. Results

To begin with, we need to understand what a recommendation system is, how it works and what its functionalities are. A recommendation system is an intelligent system that, depending on the user's activities, will recommend other actions from a wide range of activities. In the area of recommender systems there are various techniques that can be used to arrive at a recommendation, namely Content-Based Filtering (CBF), Collaborative Filtering (CF), Demographic Filtering (DRS), Hybrid Approaches, Context-Aware Recommendation, Knowledge-Based Recommendation and E-Resource Recommendation.

3.1. Data summary of the researched bibliography

In this section, it is provided a synthesis of the related literature regarding the topic of recommendation systems in social networks. The available literature offers valuable insights into various aspects of recommendation systems, including algorithms, data mining techniques, user engagement, machine learning approaches, user-system interaction, and privacy and security considerations. These related articles cover collaborative filtering as well as content-based methods and explores their application in customizing content recommendations within social network environments and the way in which user-generated data in social networks can be leveraged to enhance the quality of recommendations. These strategies can be used at the same time, resulting in a hybrid approach, which sometimes may be the most adequate one, as these techniques Furthermore, another important aspect discussed in the literature is the influence of recommendation system effectiveness on user engagement in social networks as well as substantial research on the machine learning techniques employed to enhance recommendation systems in social networks. Regarding user-system interaction, this is also a significant factor influencing recommendation quality in social networks as research in this area investigates how user feedback can be incorporated into recommendation algorithms to enhance relevance and effectiveness, Finally, the challenges and considerations related to privacy and security in implementing recommendation systems in social networks are addressed in a select group of articles that examine concerns regarding user data privacy and measures to ensure the security of personal information in social network environments. Collectively, these articles provide a comprehensive overview of the current state of research in recommendation systems in social networks, covering algorithms, challenges, and potential avenues for future exploration.

3.2. Main characteristics of the studies

Fig. 1 shows all the steps that were taken until only 10 articles were included in this systematic review, including the reasons for the exclusions of the remaining articles. Only one reason for the exclusion of each article was identified. However, some articles we consider may simultaneously meet more than one criterion, particularly, review articles and protocols that are also not intervention studies.

Publication dates of the articles spanned over several years, reflecting the evolving landscape of recommendation systems in social media. While some studies date back to the early 2010s, a significant proportion were published in recent years, showcasing the growing interest and advancements in this domain.

Participants across the studies were primarily users of social media platforms, encompassing diverse demographics and user behaviors. The sample sizes varied, with some studies focusing on smaller user groups for in-depth analysis, while others explored larger datasets to uncover broader trends and patterns.

Study designs exhibited diversity, ranging from experimental research methodologies such as randomized controlled trials to qualitative studies and observational analyses. This methodological diversity enabled a comprehensive exploration of the multifaceted aspects of recommendation systems in social media.

Themes explored in the studies included the impact of recommendation algorithms on user engagement, the role of user-generated content in enhancing recommendation accuracy, and the ethical considerations surrounding algorithmic decision-making in social media platforms. These thematic focuses collectively contributed to a nuanced understanding of the dynamics between recommendation systems and user interactions in social media environments.

This section provides a synthesized overview of the main characteristics of the studies included in the systematic review, shedding light on their geographical distribution, participant demographics, study designs, and thematic orientations within the realm of recommendation systems in social media.

3.3. Effect of the Intervention

This section delves into the observed effects and outcomes resulting from the implementation of recommendation systems within social media platforms. Through a comprehensive analysis of the included studies, key insights into the impact of these interventions on user behavior, engagement, and platform performance emerge.

3.3.1. User Engagement and Interaction Patterns

A notable finding across the studies was the significant enhancement of user engagement facilitated by recommendation systems. The implementation of personalized recommendation algorithms led to increased user interactions with content, evident through higher rates of likes, shares, and comments. Moreover, the integration of social data mining techniques contributed to a deeper understanding of user preferences and behaviors, resulting in more relevant and personalized recommendations tailored to individual users.

3.3.2. Platform Performance and Content Consumption

Recommendation systems were found to positively influence platform performance metrics, such as click-through rates and session durations. By effectively guiding users towards relevant content of interest, these systems contributed to prolonged user sessions and increased retention rates. Additionally, studies highlighted the role of recommendation algorithms in diversifying content consumption patterns, thereby mitigating issues of content homogeneity and fostering a more dynamic and engaging user experience within social media ecosystems.

3.3.3. Impact on User Satisfaction and Loyalty

The implementation of recommendation systems was associated with higher levels of user satisfaction and loyalty towards social media platforms. Personalized recommendations were perceived positively by users, leading to greater perceived value and utility of the platform. Furthermore, the provision of tailored content recommendations fostered a sense of affinity and trust between users and the platform, thereby strengthening user loyalty and retention over time.

3.3.4. Challenges and Considerations

Despite the observed benefits, several challenges and considerations emerged regarding the implementation of recommendation systems in social media. Privacy and security concerns surrounding the collection and utilization of user data were highlighted, necessitating the development of transparent and ethical data handling practices. Additionally, issues of algorithmic bias and filter bubbles were identified, underscoring the importance of algorithmic transparency and accountability in ensuring fair and diverse content recommendations.

3.3.5. Future Directions

Moving forward, future research efforts should focus on addressing the aforementioned challenges while exploring novel approaches to enhance the effectiveness and fairness of recommendation systems in social media. By integrating user-centric design principles and leveraging emerging technologies such as artificial intelligence and machine learning, the potential for recommendation systems to further enrich user experiences and foster vibrant online communities remains promising.

4. Discussion

4.1. Relevant Metrics

In order to check if the recommendations provided are accurate in the context of social media, it can be used some metrics to evaluate its "success". As such, "TP" represents the True Positives, "TN" the True Negatives, "FP" the False Positives and the "FN" the False Negatives in the following formulas:

1. **Accuracy**: - Accuracy measures the proportion of correct predictions among all predictions made by the model and it is calculated as the sum of TP and TN divided by the total number of predictions. - Accuracy provides an overall assessment of the model's performance, however it can be misleading when classes aren't balanced. - Accuracy = TP + TN / TP + TN + FP + FN

2. **Recall** (also known as Sensitivity or True Positive Rate): - Recall measures the ability of the model to correctly identify positive instances from all actual positive instances and it is calculated as the number of TP divided by the sum of TP and FN. - Recall is considered particularly useful when the cost of missing FN is high. - Recall = TP / TP + FN

3. **Precision**: - Precision measures the proportion of TP predictions among all positive predictions made by the model and it is calculated as the number of TP divided by the sum of TP and FP. - Precision is important when the cost of FP is high, as it reflects the model's ability to avoid making incorrect positive predictions. - Precision = TP / TP + FP

4. **F1 Score**: - The F1 score is the harmonic mean of precision and recall. It combines both metrics into a single value and it is calculated as 2 times the product of precision and recall, divided by the sum of these values - The F1 score provides a balanced assessment of a model's performance, considering both FP and FN. It is useful when there isn't an even class distribution or when FP and FN have different implications. - F1 = 2 * Precision * Recall / Precision + Recall

In the universe of recommendation systems in social media, where the aim is to deliver personalized content to users based on their social activities and preferences, there are specific metrics play a crucial role in assessing the system's effectiveness: while traditional metrics like accuracy, precision, and recall remain relevant, additional considerations come into play due to the dynamic and social nature of these platforms. With this factor in mind, it can be said that metrics that reflect user engagement, such as likes, shares, comments, and reposts, provide insights into how well the recommendations resonate with users and encourage interaction within their social circles (Fleder Hosanagar, 2009). However, these aren't the only ones that can provide more valuable information to enhance these recommendation systems as metrics that measure the diversity and novelty of recommendations are key to ensure that users are exposed to a wide range of content that aligns with their interests while also introducing them to new and relevant information from their social network (Castells, 2005). Not only that, but sentiment analysis can be employed to estimate user satisfaction and feelings towards the recommended content given, offering valuable feedback on the quality and relevance of the recommendations within the social media context (Liu et al., 2017). With these indicators, these social media recommendation systems are one step closer to provide users with the right recommendations, extending time on social media, which can be beneficial to these companies, as more time spent on social media, means more probability of watching paid advertisements and so on.

4.2. Recommendations Approaches

4.2.1. Content-Based Filtering

A recommendation can be made using the CBF technique, which uses the user's profile and item descriptions to recommend other items (Pazzani and Billsus, 2007). This technique uses the user's preferences to suggest similar items based on the item's characteristics, i.e. it recommends elements based on the similarities of the characteristics to the items that the user in question prefers. An example that could be given is the following: a user rates an item of a certain brand positively and the system, using this information, can and probably will recommend other items of the same brand.

4.2.2. Collaborative Filtering

Another technique that can be used is CF. The recommendation is based on the user's preferences and interactions. The interactions that users have with the product are analyzed in terms of their similarity and, based on the preferences of similar users or items, the recommendation is made. In other words, this technique requires that information exists and that these interactions have taken place. This problem is called cold start, which will be mentioned later in the article. An example that can be given is the following: if a certain user and another user have a similar purchase history, if the first user buys a certain product, the system will recommend the same product to the second user (Singh & Pramanik & Dey & Choudhury, 2021).

4.2.3. Demographic Filtering

User demographics with clustering techniques are used to categorize users into groups based on similar demographic information, in this context the DRS technique is used. An example that can be presented is the following, a certain user who lives in Portugal bought a certain product, another user who lives in Portugal bought the same product, given this information the two users were categorized in the same group, if another user who lives in Portugal appears they will be recommended the same product that the other two users bought (Singh & Pramanik & Dey & Choudhury, 2021).

4.2.4. Hybrid Approaches

In a hybrid approach, it uses multiple techniques to achieve a more assertive result and overcome some of the limitations inherent in individual techniques by taking advantage of the strengths of the different techniques and thus mitigating their weaknesses (Burke, 2007). The most common set to use is CBF and CF (Singh & Pramanik & Dey & Choudhury, 2021).

4.2.5. Context-Aware Recommendation

An approach that can also be used is CARS, where dynamic user and product information is taken into account to make personalized recommendations, where information such as time, location, mood and social interactions are considered to make recommendations tailored to the user. One of the challenges of this technique is gathering and collecting accurate data at the right time due to the dynamic and fast-changing environment of contextualization data. The use of such a system is important when it comes to personalized marketing campaigns, mobile applications and online services (Singh & Pramanik & Dey & Choudhury, 2021).

4.2.6. Knowledge-Based Recommendation

Knowledge-based recommendation uses explicit knowledge about items and users, as well as their preferences, to make personalized recommendations. This system uses knowledge bases, association rules and information about a specific domain to make informed recommendations (Middleton, 2009), but for this it needs experts in the domain to build and maintain the knowledge bases, but on the other hand, it has the advantage of being able to deal with the cold start problem and provide explanations for the recommendations it has made (Singh & Pramanik & Dey & Choudhury, 2021).

4.2.7. Problems in recommendation systems

Despite the capabilities of recommender systems, they present certain problems, including Black Box, Long Tail, cold start, sparsity, limited content analysis, excessive specialization, scalability problems, synonyms and abbreviations. Starting with the Black Box problem, this is related to the lack of transparency and the failure to specify how the recommendation algorithm works (Ramaswamy, 2015; Lee, 2014; Asmus, 2014). With regard to the Long Tail problem, this problem stems from elements that for some reason are not classified or are poorly classified (Singh & Pramanik & Dey & Choudhury, 2021). With regard to the Cold Start problem, this is a very common problem when a new product is added or when a new user is created, in which there is no history of previous users and it is therefore not possible to make an accurate recommendation (Lakshmi and Lakshmi, 2014; Su and Khoshgoftaar, 2009). With regard to sparsity, as recommendation systems are built with a large amount of data, this can lead to sparsity in the matrices that relate a certain item to a certain user. Although there is content analysis, this analysis can be inadequate and not aligned with user preferences. In parallel, there is a lack of understanding of user preferences and behavior, which obviously has a negative effect on the recommendation system (Lakshmi and Lakshmi, 2014; Su and Khoshgoftaar, 2009). Another problem is related to the specific focus on user preferences that a recommendation system can have, which can cause other broader interests to be ignored and therefore the discovery and consequent exploration of new elements can be limited (Singh & Pramanik & Dey & Choudhury, 2021). With the constant increase in the volume of data, scalability can become a problem, and the efficiency of processing this volume of data must be taken into account (Thorat et al., 2015).Finally, synonyms and abbreviations can lead to problems, since different users express themselves differently, which means that there is no connection with the items (Singh & Pramanik & Dey & Choudhury , 2021).

4.3. Usage of recommendation systems in social Media

Social media platforms have revolutionized how people interact with each other, share information about their daily lives or unique experiences, discover new people near them as well as internationally, whether through groups where people share the same interests and content or events, and how they consume content on these same platforms more frequently on a daily basis, spending more time on each use and less time between uses. With the growing penetration and rapid development of new features that increasingly tie people to social media platforms, users now spend a significant part of their time interacting with various social platforms. From chatting and watching videos to shopping and discovering new products, social media has become a central and permanent part of everyday life for millions of users around the world, which means that the role of recommendation systems cannot be underestimated and must be taken seriously in today's context. Recommendation algorithms play a crucial and very important role in helping and in a way coercing platform users to discover new, relevant and interesting content, thereby increasing customer satisfaction and making them more engaged, which can often result in increased returns for the company, and it can be said that it helps to establish contacts with like-minded people and to make informed decisions, especially in the field of e-commerce. As users navigate the vast sea of information available on social media platforms, personalized recommendations serve as valuable guides that assist the user with their experience on the platform and in their decision-making, improving the overall user experience and engagement.

The following subsections will discuss the use of recommendation systems on specific social media platforms, in particular the techniques used and the use to which the recommendation system is put, in other words, the purpose for which it exists. It is important to mention that this analysis is carried out on the following social networks:

- Facebook
- Instagram
- Twitter
- YouTube

4.3.1. Facebook

One of the oldest and most popular online social networks is Facebook, where users can communicate with their friends, join groups, create groups, play games and make friends all over the world. Although Facebook was once the most used and trending social network, its popularity has declined over the years with the emergence of new social networks and the lack of new features that differentiate it from other platforms, users are increasingly abandoning this platform and starting to use other platforms more often that offer them other experiences. One of the most popular online social networks is Facebook, where users can communicate with their friends, join groups, create groups, play games and make friends all over the world. Although its popularity has been declining over the years with the emergence of new social networks and the lack of new features that differentiate it from other platforms, users are increasingly abandoning this platform and starting to use other platforms more often that provide them with other experiences.

The exponential growth in the amount of data is making it increasingly complicated to implement automatic learning algorithms, which is why the technique used in recommendation systems to help users discover items that are relevant to them is the collaborative filter. In the case of Facebook, this technique can be used in groups, on the page where other users' posts can be found, games and other elements of the platform to recommend entities through the rating given by other users with similar likes.

According to Ilic and Kabiljo, the average Facebook data set for CF has 100 billion ratings, more than 1 billion users and millions of items. That's why it's necessary to design a distributed architecture that can manage the scale of the datasets, and that's why they came to the conclusion that they should use Apache Giraph, which is a powerful platform for iterative distributed processing and graphs, and the work we put into making it scale according to the platform's needs (Ilic, Kabiljo, 2015).

On Facebook, the most common approach to CF is through matrix factorization where the problem is considered taking into account a set of users and a set of items, and a very sparse matrix representing the known ratings from user to item. In order to predict the missing values, each user and each item are represented in a vector of latent features, given that the scalar products of these vectors correspond closely to the known ratings from user to item. A latent feature vector represents the characteristics or preferences of users, or items that are not directly observable, which are inferred from observed data, such as interactions between users and items (Ilic, Kabiljo, 2015).

Other methods can be used, such as stochastic gradient descent (SGD) and alternating least square (ALS) optimization. In the case of SGD, the aim is to minimize the loss function during training. This method iteratively updates the model parameters based on small random subsets of the training data. ALS is an evolution of matrix factorization. This algorithm alternates between fixing the user vectors and updating the article vectors, and fixing the article vectors and updating the user vectors, until the convergence criteria are met (Ilic, Kabiljo, 2015).

The previously mentioned article describes the standard approach and its problems, as well as the solution that plans to solve the problems presented. The normal approach is characterized by the use of the distributed matrix factorization explained above. This approach presents some problems, namely high network traffic, distorted distribution of item grade and inaccurate implementation of the SGD. In the case of high network traffic, it is the main bottleneck of all distributed matrix factorization algorithms, since the transmission of feature vectors through the edges of the graph during each iteration results in a substantial volume of data transfer. As for the skewed distribution of item degree, as some items in the dataset have an unbalanced distribution of item degrees, this leads to some of the items being very popular. This imbalance will create memory constraints and processing bottlenecks. Finally, the normal approach departs from the conventional stochastic gradient descent (SGD) approach and instead of using the most recent feature vectors, the vertices work with feature vectors from the initial iteration which can lead to a drift that can prevent convergence, especially in scenarios involving simultaneous updates, potentially slowing down the overall optimization process. The solution that plans to solve the previous problems presented is characterized as a hybrid approach. This approach requires to extend Giraph framework with worker-to-worker messaging where users are still presented as the vertices of the graph, but items are partitioned in workers disjoint parts, with each of these parts stored in global data of one of the workers. Then all the workers are put in a circle, and rotate the items in clockwise direction after each superstep, by sending worker-to-worker messages containing items from each worker to the next worker in the line (Ilic, Kabiljo, 2015).

4.3.2. Instagram

Instagram, one of the main social media platforms and a subsidiary of Facebook, was launched in 2010 and has grown rapidly due to its innovative features and simple interface. On this platform, users can upload photos and videos, apply filters such as editing effects, add captions and tag locations or other users who have an account on the social network. One of the innovative features that Instagram has brought is the ability to support features such as Stories, where users can share temporary posts with other users of the platform that disappear after 24 hours. Another feature was IGTV, a space dedicated to videos of a longer duration than stories as previously presented, which were later replaced by Reels or videos in the Feed.

Instagram's recommendation system utilizes advanced algorithms and machine learning to personalize content based on user preferences, interactions, and engagement history. By analyzing various signals such as post interactions, search history, and accounts followed, the system dynamically adapts recommendations to align with evolving user interests and trends. This ensures a more engaging user experience, with individuals consistently presented with content that resonates with their preferences and keeps them actively engaged on the platform.According to a blog post from Meta, authored by Vladislav Vorotilov and Ilnur Shugaepov, we'll delve into the Explore section to understand how its recommendation system works.

Instagram's Explore section utilizes advanced machine learning techniques, including Two Towers neural networks, to recommend personalized content to users. The recommendation system operates through multiple stages:

- **Retrieval:** This stage selects hundreds of relevant items from a vast pool of media using various sources, including heuristics and machine learning algorithms. The Two Towers neural network generates embeddings for users and items, allowing efficient retrieval of similar items based on user preferences and interactions history.
- **Ranking:** Content candidates are ranked in two stages. The first stage employs a lightweight model trained to predict the output of the second stage, which uses a heavier multi-task multi-label neural network model. The ranking considers factors such as the probability of user engagement events like clicks or likes.
- **Final Reranking:** Additional filtering and adjustments are applied to the ranked items to ensure integrity, diversity, and alignment with user preferences. Parameters tuning, either through Bayesian optimization or offline tuning, is crucial for optimizing the system's performance.

The system's complexity continues to evolve, leading to ongoing improvements and exploration of new ranking models and retrieval sources to enhance the user experience on Instagram's Explore section.

4.3.3. Twitter

Born in 2006, Twitter is a fast-paced social networking platform known for its concise format. Users share short messages called "tweets". At its inception, these messages were limited to 280 characters and were limited to text only. Over the years, this limit has been increased and the ability to add photos and long-form videos has been added. Twitter has become a vital tool for the dissemination of real-time news, discussions about current events and social interactions between users around the world. With Elon Musk's purchase of Twitter, there has been a modernization and renovation of the platform with the aim of making it a platform where people can express themselves freely.

The recommendation systems that Twitter uses are made up of various models and features that extract information from Tweets, users and engagement data.

The recommendation pipeline is made up of three main stages that consume these features:

1. Fetch the best Tweets from different recommendation sources in a process called candidate sourcing.
2. Rank each Tweet using a machine learning model.
3. Apply heuristics and filters, such as filtering out Tweets from users you've blocked, NSFW content, and Tweets you've already seen.

Twitter has several candidate sources, which are used to obtain recent Tweets that are most relevant to the user. Whenever there is a new request, the best 1500 tweets are selected from millions of sets of publications. These results are obtained by combining candidates that the user follows with those that they don't. On average, the distribution is even, but of course this distribution can be different from user to user (Twitter Team, 2023).

When it comes to candidates from users the user follows, these are the main source and are intended to provide the user with the most relevant content and the most recent posts from users the user follows. Based on the relevance of each post, they are ranked using the logistic regression model, after which the posts at the top of the ranking are sent to the next stage. One of the components that is also important to mention is the use of a model called Real Graph which predicts the likelihood of engagement between two users, whereby the higher the Real Graph score between the user and the author of the Tweet, the more of their tweets will be included (Twitter Team, 2023).

In the case of candidates outside the user's network, that is, those users they don't follow, this is a more difficult process since finding relevant Tweets outside a user's network is a complex challenge and twitter has two approaches for this:

- Social Graph
- Embedding Spaces

The first approach consists of estimating relevant content by analyzing the interactions of people who follow each other or have similar interests, and to do this GraphJet was developed, a graph processing engine that maintains a real-time interaction graph between users and Tweets. This system was developed in order to obtain answers to the following questions: What Tweets did the people I follow recently engage with? and Who likes similar Tweets to me, and what else have they recently liked? On the other hand, Embedding Spaces aims to answer a more general question about the similarity of content: Which Tweets and Users are similar to my interests? and with this it has become the largest source of Tweets outside the user's contact, given its wider reach. This approach uses SimClusters, which discovers communities anchored by a group of influential users using a customized matrix factorization algorithm (Twitter Team, 2023).

On the classification side, there are 1500 candidates that may be revealing to the user, but which need to be sorted according to their relevance. To achieve this organization of content, the Twitter team uses a neural network with around 48 million parameters that is continuously trained on Tweet interactions to optimize positive engagement (e.g. likes, retweets and replies). The presented classification mechanism takes into account thousands of characteristics and produces ten labels to assign a score to each Tweet, where each label represents the probability of engagement (Twitter Team, 2023).

Once the sorting phase is complete, filters and checks are applied, which may include :

- Removing Tweets from accounts that you block or mute.
- Avoiding too many consecutive Tweets from a single author.
- Determining whether Tweets that are currently on a device are out of date.

Once the selection of relevant content is complete, the selected content is mixed with other content such as ads (Twitter Team, 2023).

4.3.4. Youtube

YouTube, the leading video-sharing platform, has transformed media consumption and social interaction since its inception in 2005. Central to its success is its recommendation system, which utilizes advanced algorithms to personalize content suggestions based on user preferences. While this system enhances user engagement, it has raised concerns about filter bubbles, misinformation, and algorithmic transparency. Understanding the impact and implications of YouTube's recommendation system is vital for various stakeholders.

> YouTube's recommendation system stands at the forefront of digital content curation, shaping user experiences and content consumption patterns on the platform. Building upon the foundational understanding provided earlier, this section delves deeper into the mechanics of YouTube's recommendation algorithm. By elucidating the intricate processes underlying content recommendation, we aim to offer insights into how YouTube tailors its suggestions to individual users, balances user satisfaction with responsible content promotion, and navigates the complexities of content quality and relevance.
>
> This exploration draws extensively from information obtained from YouTube's official blog post written by Cristos Goodrow.

Breakdown of the Recommendation Algorithm:

1. **Constant Evolution**: YouTube's recommendation system operates as a dynamic entity, continually learning and adapting from an extensive pool of over 80 billion data points, or "signals." This continuous evolution ensures that recommendations remain relevant and responsive to shifting user preferences and behaviors.
2. **Multifaceted Signals**: The recommendation algorithm integrates a diverse array of signals to gauge user satisfaction and content relevance. These signals encompass various user interactions, including clicks, watchtime, survey responses, sharing, likes, and dislikes. Each signal provides valuable insights into user preferences, guiding the algorithm in suggesting content that aligns with individual tastes and interests.
3. **Clicks**: Initially, clicks on videos were considered indicative of user interest. However, YouTube recognized the limitations of this metric, as clicking on a video did not necessarily guarantee viewer satisfaction. To address this, the algorithm evolved to

incorporate watchtime as a more robust indicator of content value.

4. **Watchtime**: Watchtime, or the duration of time users spend watching videos, emerged as a pivotal signal in shaping recommendations. By analyzing which videos users watch and for how long, the algorithm gains deeper insights into content relevance and user engagement. Videos that garner longer watchtimes are prioritized in recommendations, reflecting their perceived value to viewers.
5. **Valued Watchtime**: Beyond mere watchtime, YouTube introduced the concept of "valued watchtime" to measure user satisfaction more accurately. Through user surveys, YouTube solicits feedback on the perceived value of watched content, enabling the algorithm to prioritize videos rated highly by users. This nuanced approach ensures that recommendations align not only with viewing habits but also with user preferences and satisfaction levels.
6. **Sharing, Likes, and Dislikes**: User engagement metrics such as sharing, likes, and dislikes further inform the recommendation algorithm. Videos that are shared or liked are deemed more likely to satisfy viewers, influencing subsequent recommendations. Conversely, dislikes serve as signals of user dissatisfaction, prompting the algorithm to adjust recommendations accordingly.
7. **Responsible Recommendations**: In addition to optimizing user satisfaction, YouTube prioritizes responsible content promotion, particularly in sensitive areas such as news and information. The recommendation algorithm employs classifiers to identify and mitigate low-quality or harmful content, including misinformation and borderline material. Human evaluators assess the credibility and potential impact of videos, guiding the algorithm in promoting authoritative and trustworthy content while demoting misleading or harmful material.

In tandem with the comprehensive breakdown of YouTube's recommendation algorithm, it is imperative to explore the underlying technologies that power its intricate processes. As such, this supplementary article, authored by Paul Covington, Jay Adams, and Emre Sargin, delves into the realm of neural networks and deep learning, illuminating the sophisticated mechanisms driving YouTube's recommendation system. By delving into the technical foundations of neural networks, this piece aims to provide a deeper understanding of how advanced algorithms are leveraged to personalize content suggestions and navigate the complexities of user engagement and content quality on the platform.

During candidate generation, the vast YouTube corpus is filtered down to hundreds of videos that might be relevant to the user. The approach involves:

- **Matrix Factorization Approach:** Previously, a matrix factorization approach was employed, trained under rank loss. Early neural network models mimicked this factorization behavior, serving as a nonlinear generalization.

- **Recommendation as Classification:** Recommendation is framed as extreme multiclass classification, accurately classifying a specific video watch among millions based on user and context. Implicit feedback, such as watches, is utilized for training, enabling recommendations deep into the tail.

- **Efficient Training:** To train efficiently with millions of classes, negative class sampling and importance weighting are utilized. This technique speeds up training significantly compared to traditional softmax, with comparable accuracy.

- **Model Architecture:** Inspired by continuous bag of words language models, embeddings for each video are learned and fed into a feedforward neural network. User watch history is represented as a variable-length sequence of sparse video IDs, mapped to dense vectors via embeddings.

- **Heterogeneous Signals:** The model accommodates arbitrary continuous and categorical features. Search history, demographic features, and other user attributes are integrated into the model.

- **Example Age Feature:** To recommend recently uploaded content effectively, the age of the training example is fed as a feature during training. This helps correct for the implicit bias towards past content in machine learning systems.

- **Label and Context Selection:** Recommendations are generated from all YouTube watches, including those embedded on other sites. The selection of training examples per user is fixed to prevent a small cohort of users from dominating the loss function. Care is taken to prevent overfitting and exploitation of the site's structure by withholding information from the classifier.

- **Experiments with Features and Depth:** Adding features and depth significantly improves precision on holdout data. Experiments show the impact of varying network depth on model performance.

The primary role of ranking in YouTube's recommendation system is to specialize and calibrate candidate predictions for the particular user interface. This involves utilizing impression data to tailor recommendations to individual user preferences. Key points include:

- **Deep Neural Network for Ranking:** YouTube employs a deep neural network with a similar architecture to candidate generation for ranking. Logistic regression is used to assign an independent score to each video impression, which is then sorted to generate the recommendation list for the user.

- **Feature Representation:** Features are categorized into traditional taxonomy of categorical and continuous/ordinal features. These features describe properties of the item (impression) or properties of the user/context (query), contributing to the ranking process.

- **Feature Engineering:** Despite advancements in deep learning, feature engineering remains crucial for representing temporal sequences of user actions and their relation to the video impression being scored. Important signals include user's past interactions with the item and similar items, along with features from candidate generation.

- **Embedding Categorical Features:** Sparse categorical features are mapped to dense representations using embeddings. Shared embeddings are utilized to improve generalization, speed up training, and reduce memory requirements.

- **Normalizing Continuous Features:** Proper normalization of continuous features is essential for model convergence. Linear interpolation on feature quantiles is used for normalization, with additional powers of continuous features enhancing model expressiveness.

- **Modeling Expected Watch Time:** The goal is to predict expected watch time using weighted logistic regression. Positive impressions (clicked) are weighted by observed watch time, while negative impressions (unclicked) receive unit weight.

- **Experiments with Hidden Layers:** Different hidden layer configurations are tested to optimize model performance. Increasing the width and depth of hidden layers improves results, but there's a trade-off with server CPU time. The chosen configuration balances performance with computational resources.

5. Conclusion

Based on the analyses presented on the use of recommendation systems on specific social media platforms such as Facebook, Instagram, Twitter, and YouTube, it's clear that these algorithms play a pivotal role in personalizing user experience and promoting engagement on their respective platform. Each one uses advanced Machine Learning methods and specific algorithms to analyze user behavior, interactions and preferences, which is fundamental to provide relevant and personalized recommendations. This article showed that regardless of the platform, recommendation systems analyze user behavior, interactions, and preferences to provide suggestions that suit each and every user and this process involves retrieving relevant content, ranking it based on user relevance, and enhancing recommendations to ensure diversity but most importantly alignment with user interests, as well as dealing with problems such as the cold-start or the black sheep. With that in mind, while recommendation systems highly contribute to user engagement and satisfaction it is worth noting that their implementation should raise important questions regarding some themes such as privacy, content diversity, and algorithmic transparency to avoid legal problems that sometimes can ruin a social media platform image. It is crucial for these systems to be not only developed but deployed responsibly, ensuring that user privacy is respected, diverse content is promoted, and the algorithms' decision-making processes are transparent to users. It is possible to conclude that recommendation systems play a pivotal role in shaping user experiences on social media platforms, but their responsible development and deployment are essential to keep trust as well as providing positive user engagement.

References

- Shaba, T. S., Khurshid, S. A., Khattak, A. M., Khan, I. U., & Sajjad, A. W. M. (2021). Techniques and Applications of Social Network Analysis for Cybersecurity: A Survey. IEEE Access, 9, 166523-166547. DOI: 10.1109/ACCESS.2021.3076833.

- K, Suraj Suresh, Srivastava, Mehul, & Mohana. "YouTube and Movie Recommendation System Using Machine Learning." Electronics and Telecommunication Engineering, RV College of Engineering®, Bengaluru, Karnataka, India, 2023.

- Mehlhose, F. M., Petrifke, M., & Lindemann, C. (2021). Evaluation of Graph-based Algorithms for Guessing User Recommendations of the Social Network Instagram. In 2021 IEEE 15th International Conference on Semantic Computing (ICSC) (pp. 409-410). IEEE.

- Fleder, D., & Hosanagar, K. (2009). Recommender systems and their impact on sales diversity. In Proceedings of the 18th international conference on World Wide Web (pp. 697-706).

- Castells, P. (2005). Informationalism, networks, and the network society: A theoretical blueprint. In Castells, M. (Ed.). The Network Society: A Cross-Cultural Perspective (pp. 3-45). Edward Elgar Publishing.

- Liu, B., Gao, Z., Zhi, Y., Chen, X., & Wei, Y. (2017). Sentiment analysis of user comments for one-class collaborative filtering based recommendation. Neurocomputing, 260, 5-14.

- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2018). A comparison of social media and traditional news media: Which one is more influential? In Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM).

- Zhang, M., Tong, H., Cui, P., Wang, J., & Shao, J. (2019). Attention-aware deep learning for automatic news recommendation. IEEE Transactions on Knowledge and Data Engineering, 32(11), 2181-2194.

- Gao, H., Tang, J., Hu, X., & Liu, H. (2018). Content-Aware Point-of-Interest Recommendation on Location-Based Social Networks. IEEE Transactions on Knowledge and Data Engineering, 30(12), 2354-2367.

- Al-Daihani, S., & Dennis, C. (2019). What Drives Instagram Users to Like and Comment? The Effects of Content, Personal Characteristics, and Marketing Strategy on User Engagement. Journal of Marketing Management, 35(9-10), 855-875.

- Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In Proceedings of the 31st International Conference on Machine Learning (ICML'14).

- Saputra, Y. M., Kusumadewi, S., Munir, R., & Arifin, A. Z. (2020). Instagram Post Recommendation System using Collaborative Filtering and Personalized PageRank Algorithm. In 2020 4th International Conference on Informatics and Computational Sciences (ICICoS) (pp. 1-6). IEEE.

- Mehlhose, C., Petrifke, M., & Lindemann, J. (2021). Deep Learning for Instagram Image Content Recommendation. arXiv preprint arXiv:2106.03161.

- Zhang, Y., Guo, H., Lan, Y., Xu, J., Cheng, X., & Liu, Y. (2018). Time-weighted collaborative filtering for personalized POI recommendations. Information Sciences, 432, 543-561.

- Mittal, H., & Sharma, A. (2019). Personalized Tweet Recommendation System. In Proceedings of the International Conference on Data Engineering and Communication Technology (ICDECT'19) (pp. 1-5). Springer, Singapore.

- Chen, X., Wang, P., & Lin, X. (2017). Content-Aware Collaborative Filtering with Network Regularization. IEEE Transactions on Knowledge and Data Engineering, 29(9), 1963-1976.

- Huang, J., Zeng, D. D., Li, X., Wang, R., & Li, X. (2020). Real-time Event-aware Recommendation on Twitter: A Multi-task Learning Approach. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20) (pp. 1409-1418).

- Singh, P. & Pramanik, P. & Dey, A. & Choudhury, P. (2021). Recommender Systems: An Overview, Research Trends, and Future Directions. International Journal of Business and Systems Research. 15. 14–52. 10.1504/IJBSR.2021.10033303.

- Pazzani, M.J. and Billsus, D. (2007) 'Content-based recommendation systems', in The Adaptive Web, pp.325–341, Springer-Verlag, Berlin, Heidelberg

- Burke, R. (2007) 'Hybrid web recommender systems', in The Adaptive Web, pp.377–408, Springer-Verlag, Berlin, Heidelberg.

- Middleton, S.E., Roure, D.D. and Shadbolt, N.R. (2009) 'Ontology-based recommender systems', in Handbook on Ontologies, pp.779–796, Springer Berlin, Heidelberg.

- Ramaswamy, S. (2015) The Challenges in Adopting Merchandise Optimization Solutions [online] `http://234w.tc.tracom.net/SiteCollectionDocuments/White%20Papers/Retail_Whitepaper_The-Challenges-in-Adopting-Merchandise-Optim` `0812-1.pdf` (accessed 31 March 2024).

- Lee, N., Jung, J.J., Selamat, A. and Hwang, D. (2014) 'Black-box testing of practical movie recommendation systems: a comparative study', Comput. Sci. Inf. Syst., Vol. 11, No. 1, pp.241–249.

- Asmus, J., Borchmann, D., Sbalzarini, I.F. and Walther, D. (2014) 'Towards an FCA-based recommender system for black-box optimization', in 3rd International Workshop on 'What Can FCA Do for Artificial Intelligence?'.

- Lakshmi, S.S. and Lakshmi, T.A. (2014) 'Recommendation systems: issues and challenges', International Journal of Computer Science and Information Technologies, Vol. 5, No. 4, pp.5771–5772.

- Su, X. and Khoshgoftaar, T.M. (2009) 'A survey of collaborative filtering techniques', Advances in Artificial Intelligence, Vol. 2009, No. 4, p.4:2.

- Thorat, P.B., Goudar, R.M. and Sunita, B. (2015) 'Article: survey on collaborative filtering, content-based filtering and hybrid recommendation system', International Journal of Computer Applications, Vol. 110, pp.31–36.

- Ilic A., Kabiljo M. (2015) Recommending items to more than a billion people [online] `https://engineering.fb.com/2015/06/02/core-infra/recommending-items-to-more-than-a` (accessed 31 March 2024).

- Twitter Team (2023) Twitter's Recommendation Algorithm [online] `https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm` (accessed 31 March 2024).

- Vorotilov, V., & Shugaepov, I. (2023, August 9). Scaling the Instagram Explore recommendations system. Posted to ML Applications. Retrieved from: `https://engineering.fb.com/2023/08/09/ml-applications/scaling-instagra`

- "Powered by AI: Instagram's Explore Recommender System." (2019, November 25). Retrieved from: `https://ai.meta.com/blog/powered-by-ai-instagrams-explore-recommende`

- Twitter. (2023, March 31). Twitter's Recommendation Algorithm. Retrieved from: https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm

- Pazzani, M.J., Billsus, D. (2007). Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_10

- Singh, Pradeep & Dutta Pramanik, Pijush & Dey, Avick & Choudhury, Prasenjit. (2021). Recommender Systems: An Overview, Research Trends, and Future Directions. International Journal of Business and Systems Research. 15. 14–52. https://doi.org/10.1504/IJBSR.2021.10033303

- Kumar, B.; Sharma, N.; Sharma, B.; Herencsar, N.; Srivastava, G. Hybrid Recommendation Network Model with a Synthesis of Social Matrix Factorization and Link Probability Functions. Sensors 2023, 23, 2495. https://doi.org/10.3390/s23052495

- Berkani, Lamia. (2020). A semantic and social-based collaborative recommendation of friends in social networks. Software: Practice and Experience. 50. https://doi.org/10.1002/spe.2828

- Cai, Chongchao & Xu, Huahu & Wan, Jie & Zhou, Baiqing & Xie, Xiongwei. (2020). An Attention-Based Friend Recommendation Model in Social Network. Computers, Materials & Continua. 65. 2475-2488. https://doi.org/10.32604/cmc.2020.011693

- Ha, I., Oh, KJ. & Jo, GS. Personalized advertisement system using social relationship based user modeling. Multimed Tools Appl 74, 8801–8819 (2015). https://doi.org/10.1007/s11042-013-1691-6

- Margaris, D., Vassilakis, C. & Spiliotopoulos, D. Handling uncertainty in social media textual information for improving venue recommendation formulation quality in social networks. Soc. Netw. Anal. Min. 9, 64 (2019). https://doi.org/10.1007/s13278-019-0610-x

- Shokeen, J., Rana, C. Social recommender systems: techniques, domains, metrics, datasets and future scope. J Intell Inf Syst 54, 633–667 (2020). https://doi.org/10.1007/s10844-019-00578-5

- Tahmasebi, H., Ravanmehr, R. & Mohamadrezaei, R. Social movie recommender system based on deep autoencoder network using Twitter data. Neural Comput & Applic 33, 1607–1623 (2021). https://doi.org/10.1007/s00521-020-05085-1

- Shokeen, J., Rana, C. A study on features of social recommender systems. Artif Intell Rev 53, 965–988 (2020). https://doi.org/10.1007/s10462-019-09684-w

- Chang, L., Chen, W., Huang, J. et al. Exploiting multi-attention network with contextual influence for point-of-interest recommendation. Appl Intell 51, 1904–1917 (2021). https://doi.org/10.1007/s10489-020-01868-0

- Zhu, Tianyu & Liu, Guannan & Chen, Guoqing. (2020). Social Collaborative Mutual Learning for Item Recommendation. ACM Transactions on Knowledge Discovery from Data. 14. 1-19. https://doi.org/10.1145/3387162