European Journal of AI, Computing & Informatics

Vol. 1 No.2 2025

Article Open Access

Development and Optimization of Social Network Systems on Machine Learning

Huijie Pan [1,*]

[1] Identity Department, PayPal Inc., San Jose, California, 95131, United States

* Correspondence: Huijie Pan, Identity Department, PayPal Inc., San Jose, California, 95131, United States

Received: 19 May 2025
Revised: 27 June 2025
Accepted: 16 July 2025
Published: 20 July 2025

Abstract: The intelligent and personalized development of social network system cannot be separated from the support of machine learning technology. This paper discusses the application of machine learning in social network system, focusing on the implementation of data flow design, machine learning model integration, personalized recommendation, social graph analysis, emotion recognition and security protection. At the same time, the strategy of optimizing system throughput, computing efficiency, data transmission and resource allocation through machine learning is studied to improve system performance.

Keywords: machine learning; social networking; personalized recommendation; emotion analysis; system optimization

1. Introduction

In the era of rapid expansion of social media, the user base and data scale have expanded rapidly, making it difficult for the original system architecture to carry the requirements of intelligent and personalized services. Machine learning, as a powerful technical means, can effectively process massive data and improve the user experience and system performance of social networks. This paper aims to explore the development and optimization of social network system based on machine learning, analyze its application in personalized recommendation, social graph analysis, emotion recognition and security protection, and further explore its potential in optimizing system throughput, computing efficiency and data transmission.

2. Machine Learning-Driven Social Network System Development

2.1. Data Flow Design and Real-Time Processing Architecture

In the social network system based on machine learning, data flow design occupies the core position. Systems need to quickly process huge amounts of real-time data covering multiple dimensions such as user behavior, social interaction, and content creation. To ensure the real-time and expandability of the system, streaming data processing architecture is usually adopted. Common architectures include data acquisition layer, data processing layer, storage layer and application layer. At the data acquisition layer, message queues such as Kafka are used to receive data streams in real time. The data processing layer performs stream processing and real-time analysis via Apache Flink or Spark Streaming. The storage layer uses distributed storage systems (such as Hadoop or NoSQL databases) to ensure efficient data storage and fast query. Figure 1 below summarizes the data flow design and real-time processing architecture:

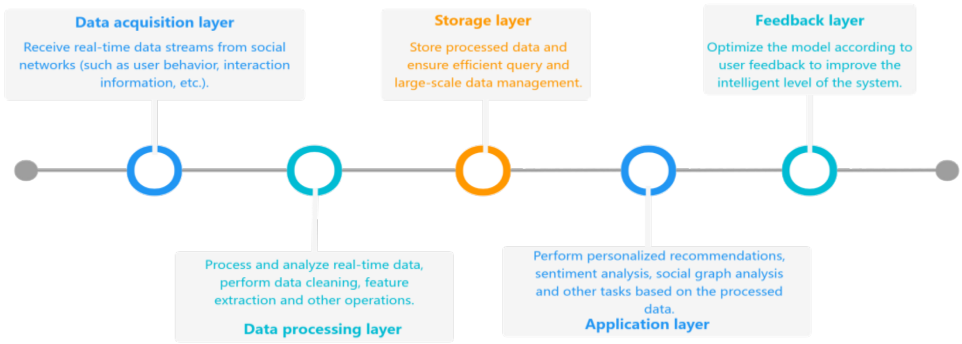

Figure 1. Data flow design and real-time processing architecture.

2.2. Development and Integration of Machine Learning Models

The development of machine learning models includes selecting suitable algorithms, such as deep learning, collaborative filtering, matrix decomposition, random forest, etc., while carrying out the necessary data cleaning and feature extraction work to better enhance the learning efficiency of the model. The model is trained through the historical data set and the parameters are adjusted to obtain more accurate prediction results. After the model is developed, it is connected to the back end of the social network system through API interface to realize real-time personalized recommendation and user behavior prediction. Containerization technologies, such as Docker, can be used to efficiently deploy and scale models, simplifying the go-live process. In order to ensure system stability, the model needs to be regularly monitored and updated, and constantly adjusted based on real-time data to adapt to changing user needs and habits [1].

2.3. Sentiment Analysis and Public Opinion Monitoring

Sentiment analysis and public opinion monitoring play a key role in social networks. Sentiment analysis provides insight into the emotional color of user-generated content, including positive, negative, or neutral emotions. This is critical to capturing user emotional responses, fine-tuning content delivery, and enhancing the user experience. Public opinion monitoring detects and predicts social hot events, public mood changes and potential risks through real-time analysis of massive social media data. Machine learning algorithms, such as sentiment dictionaries, sentiment classifiers, and deep learning models, are widely used for such tasks. Combining these algorithms with interactive data in social networks (such as likes, comments, and shares) improves monitoring accuracy and enables comprehensive analysis to better identify and predict trends in network public opinion, providing rapid decision-making references for enterprises and government departments.

3. Application of Machine Learning in Social Networking Systems

3.1. Personalized Recommendation and Content Push

In the social network system, one of the key technologies to achieve user personalized experience is the accurate recommendation and distribution of content. The recommendation algorithm based on machine learning can push the most relevant content to each user by analyzing the user's historical behavior, interest preference, and social relationship. For example, the social platform will predict the user's interest according to the user's past likes, comments, browsing behavior and other data, and push relevant posts, advertisements or social activities to increase the user's participation. Common recommendation algorithms include collaborative filtering, content-based recommendation, matrix decomposition and deep learning. By analyzing similarities among users, collaborative filtering algorithms recommend content liked by similar users. Content-based recommendation recommends content that is relevant to the user's interests based on the characteristics of the content (such as tags, keywords, etc.). Matrix decomposition and deep learning algorithms improve the accuracy of recommendations by analyzing users' potential preferences through more complex models. Table 1 below compares the recommended algorithms:

Table 1. Comparison of recommended algorithms.

Recommendation algorithm	advantage	shortcoming	Application scenario
Collaborative filtering	Easy to implement and can push content according to user interests	Sparse matrix problem, cold start problem	Recommendations based on user behavior
Content-based recommendations	Recommendation accuracy is high, consider the content characteristics	No potential correlations between content can be found	Content related recommendations
Matrix decomposition	It can effectively handle large scale data and capture hidden preferences	The calculation is large and the training time is long	Recommendation of large-scale data
Deep learning	Can handle complex features and improve recommendation accuracy	It requires a lot of data and computing resources	Highly personalized recommendations, complex scenarios

As can be seen from Table 1, by continuously optimizing these recommendation algorithms, social network platforms can continuously improve the accuracy of the recommendation system and user satisfaction.

3.2. Social Graph and Community Discovery in Social Networks

The social graph is a graphical data structure describing user interactions in a social network, representing connections such as friendship and attention between users. By capturing these relationships and their strength, the social graph provides structural information for social networks. Machine learning methods, especially graph neural networks (GNN), are widely used in social graph analysis to help identify users' interests and social groups. Community discovery is a key task in social graph analysis, which aims to discover the close-knit groups of users in a social network through algorithms. Community discovery methods can be divided into graph-based partitioning algorithms and node feature-based clustering algorithms. Graph partitioning algorithms, such as Louvain's algorithm, identify communities by optimizing the connection density between nodes. The clustering algorithm based on node features can divide similar groups by analyzing the attributes of nodes. Community discovery not only helps to reveal the group structure in social networks, but also provides technical support for customized content recommendation, precision marketing and other activities. Social graph and community discovery help to understand user social behavior and group dynamics, and promote the intelligence of social network system [2].

3.3. Public Opinion Analysis and Emotion Recognition

Public opinion analysis and emotion recognition are important tools to understand users' emotions and predict the direction of public opinion in social networks. Emotion recognition uses natural language processing (NLP) techniques to analyze emotional tendencies in user-generated content, such as positive, negative, or neutral. This is crucial for predicting reactions to public events, optimizing content push, adjusting marketing strategies, and more. Sentiment analysis usually includes emotion classification, emotion intensity analysis, etc., which is widely used in social media monitoring, brand management, crisis management and other fields. Public opinion analysis focuses on monitoring and predicting hot topics and public opinion dynamics in social networks. Through real-time analysis of social platform data, public opinion monitoring can reveal users' attitudes towards an event, mood swings and potential risks. Machine learning algorithms, such as naive Bayes, support vector machines, convolutional neural networks, etc., are widely used in sentiment classification and public opinion monitoring to help social network platforms identify crises in a timely manner for intervention and response. Table 2 below compares sentiment analysis and public opinion monitoring methods:

Table 2. Comparison of sentiment analysis and public opinion monitoring methods.

method	advantage	shortcoming	Application scenario
Naive bayes	Simple and easy to implement, fast training speed	Weak understanding of text context information	Simple emotion classification task
Support Vector Machine (SVM)	High precision, suitable for large-scale data analysis	The calculation cost is large and the training time is long	Large-scale public opinion analysis
Convolutional Neural Networks (CNN)	Can handle complex text features with high accuracy	Large amounts of annotated data are required and high computing resources are required	Deep emotion analysis and complex text processing

As can be seen from Table 2, sentiment analysis methods such as naive Bayes are suitable for simple tasks, SVM is suitable for large-scale data analysis, and CNN provides higher accuracy when processing complex texts. Choosing a method is a trade-off based on task complexity and data volume

3.4. Social Network Security and Anti-cheating Technology

Social networking platforms face security threats such as malicious attacks, disinformation and user privacy breaches. The application of machine learning technology in anti-cheating and abnormal behavior detection is becoming increasingly important in improving platform security. Anti-cheating technology mainly analyzes user behavior data to identify malicious behavior, such as bot attacks, fake account creation, and spam publishing. Anomaly detection methods based on machine learning can automatically identify and flag suspicious behavior, reducing the pressure of manual review. For example, clustering algorithms or neural network models are used to analyze user behavior patterns and find abnormal activities that are significantly different from normal behavior, so as to identify and prevent malicious behavior. In addition, social network platforms can also use graph model-based algorithms to detect potential disinformation dissemination networks by analyzing social relationships and interactions among users. Anti-cheating technology not only helps to improve the security of the platform, but also enhances the trust and experience of users. The combination of privacy protection technologies, such as differential pri vacy, ensures that the user's personal privacy is protected to the maximum extent possible when conducting user behavior analysis and anti-cheat detection [3].

4. Machine Learning-Driven Social Network System Optimization

4.1. Enhance System Throughput and Processing Capability

In machine learning-driven social networking systems, increasing system throughput and processing power is key to improving overall performance. With the rapid growth of users and data volume, the system must have enough processing power to use Streaming data processing architectures (such as Apache Kafka and Spark Streaming) to reduce data latency and enhance system throughput by efficiently processing real-time data streams. By analyzing user behavior and content generation data in real time, the system can quickly identify hot topics and user needs to improve user experience. In terms of algorithm optimization, the compression and parallel computation of machine learning models can reduce the computational complexity and improve the processing power. Distributed computing frameworks (such as Hadoop and Spark) greatly improve computing speed and system throughput by breaking tasks into multiple small tasks and using multiple compute nodes to process them in parallel [4]. The computational efficiency of the optimization algorithm is also the key. For example, in the recommendation system, dimensionality reduction of the interaction matrix between users and items can significantly reduce the computing time and resource consumption, while ensuring the accuracy of the recommendation. In this way, the system is able to maintain an efficient balance between massive data processing and complex calculations. To improve system throughput, the formula can be adopted:

$$T = \frac{N}{\sum_{i=1}^{M}(P_i \cdot C_i)} \tag{1}$$

Among them, T Represents the system throughput, N Is the amount of data, P_i Is the number of processing units, C_i It's the computing power of the unit. M Number of processing units Increasing the number of processing units or increasing the computing power can effectively improve the system throughput.

4.2. Improve Data Transmission and Network Bandwidth Utilization

Data transfer efficiency and bandwidth utilization directly affect system performance, especially in high-traffic and big data environments. The use of compression algorithms, data deduplication, and efficient transport protocols such as HTTP/2 and QUIC can effectively reduce transmission time and bandwidth consumption. The key to improve bandwidth utilization efficiency is to optimize data flow and network resource allocation. Machine learning and network traffic prediction technology can help the system dynamically adjust data traffic, predict network congestion, allocate bandwidth resources rationally, and avoid bottlenecks. For example, deep learning is used to analyze traffic data and predict bandwidth requirements, thereby optimizing data routing and transmission paths and reducing congestion. In addition, a distributed content delivery network (CDN) is used to cache data to nodes closer to the user, reducing transmission distances and improving bandwidth utilization. CDN technology selects the best transmission path based on user location and network conditions to reduce latency and improve user experience. The formula can be selected:

$$B = \frac{D \cdot S}{T \cdot L} \tag{2}$$

Among them, B Is bandwidth utilization, D Is the amount of data transferred, S Is the data compression rate, T It's the transmission time, L It is the network delay. By increasing the amount of data D Or shorten the transmission time T To optimize bandwidth utilization. By optimizing data transfer protocols, adopting intelligent traffic management and distributed caching, social networking systems can effectively improve bandwidth utilization and overall performance.

4.3. Improve Computing Efficiency and Response Speed

Improving computing efficiency and response speed is the key to improving the performance of social networking systems. By optimizing the algorithm and accelerating the calculation process, the efficiency of the system can be effectively improved. By means of model compression, algorithm simplification and parallel computation, the computational complexity is reduced and the processing speed is increased. For example, distributed computing frameworks (such as Apache Spark and Hadoop) can be used to split tasks and process them in parallel, greatly improving computing efficiency. Parallel computing not only speeds up the data processing process, but also improves the response speed of the system, ensuring that user requests can be responded in a short time. At the same time, real-time data processing architectures, such as stream processing systems, also play an important role [5]. Through streaming systems, social networking platforms can quickly process real-time user data, enabling low-latency content recommendations and instant feedback, which greatly improves the user experience. To improve computational efficiency and response speed, the formula can be used:

$$R = \frac{C \cdot P \cdot N}{T \cdot L} \tag{3}$$

Among them, R Represents the speed of response, C It's computing power, L It's the load. P Is the number of processing units, N Is the computing power per processing unit, T It's task processing time. By increasing the number of processing units P 、Improve computing power C And computing power per processing unit N At the same time reduce the system load L And task processing time, can significantly improve the response speed of the system.

4.4. Optimizing Resource Allocation and Load Balancing

Optimizing resource allocation and load balancing is the key to improving system performance and stability. The rational allocation of resources should be based on real-time system load monitoring and forward-looking prediction. With the help of machine learning, the system is able to analyze past data in depth to predict load fluctuations over time and adjust resource allocation flexibly accordingly. Using deep learning, the system has the ability to accurately process massive user data, timely detect and respond to high-load areas, and ensure smooth operation of the system under high-load conditions. Load balancing intelligently schedules and allocates computing resources to ensure that no node in the system is overloaded. Common load balancing policies include weight-based scheduling, polling scheduling, and minimum number of connections scheduling. These policies can allocate requests appropriately according to the current load of the server, avoiding the situation that some nodes are overloaded while others are idle. By adopting the formula:

$$T = \frac{R}{\sum_{i=1}^{N}(P_i \cdot W_i)} \tag{4}$$

Among them, L Represents the load on each node, R Is the total resource amount, P_i Is the processing capacity of the i -th processing uni, W_i Is the weight of the i -th processing unit, N Is the number of processing units. By allocating the weights of processing units reasonably W_i and increase the computing power of the processing unit P_i To optimize resource allocation and reduce the load on each node, thereby improving the overall load balancing effect.

5. Conclusion

With the rapid progress of social networks, the user experience and platform performance have been continuously enhanced. The application of machine learning technology in data processing, personalized recommendation, resource allocation, etc., has made the social network system achieve significant optimization. By increasing computing efficiency, improving data transmission, optimizing bandwidth utilization, and enabling intelligent load balancing, the system can better respond to large-scale user demands. In the future, with the progress of technology, social network systems will achieve further leaps in processing efficiency, response speed and intelligent services, so as to bring users a smoother, reliable and intelligent interactive experience.

References

1. C. K. Ch'ng, "Hybrid machine learning approach for predicting E-wallet adoption among higher education students in Malaysia," *J. Inf. Commun. Technol*., vol. 23, no. 2, pp. 177–210, 2024, doi: 10.32890/jict2024.23.2.2.
2. A. A. de Souza, et al., "249.1: Interpretable prediction of kidney allograft rejection using machine learning: A comparison between linear and non-linear models," *Transplant*., vol. 108, no. 9S, 2024, doi: 10.1097/01.tp.0001064836.44766.95.
3. M. B. Pasha, et al., "Tribo-informatics approach to predict wear and friction coefficient of Mg/Si3N4 composites using machine learning techniques," *Tribol. Int*., vol. 196, p. 109696, 2024, doi: 10.1016/j.triboint.2024.109696.
4. Z. Huang, et al., "Identification of KRAS mutation-associated gut microbiota in colorectal cancer and construction of predictive machine learning model," *Microbiol. Spectr*., vol. 12, no. 5, p. e02720-23, 2024, doi: 10.1128/spectrum.02720-23.
5. K. K. Sinha and S. Suvvari, "PRMNBR: Personalized Recommendation Model for Next Basket Recommendation Using User's Long-Term Preference, Short-Term Preference, and Repetition Behaviour," *Revue d'Intell. Artif.*, vol. 38, no. 4, 2024, doi: 10.18280/ria.380417.