Research and Analysis of Corporate Fintech Development Based on Big Data Innovation

Xuanyi Liu[1,*]

{ lxy_hkzh@163.com }

[1]Hong Kong Chu Hai College, Hong Kong 999077, Chinas

Abstract. In the information explosion mobile Internet era, while people enjoy rich online services, they are also plagued by redundant and inefficient information. By mining information related to users and items, recommendation systems can generate accurate and personalised recommendations for users, which can solve these problems to a certain extent. The development of deep learning technology in recent years has driven the rapid evolution of recommendation algorithms, while at the same time placing greater demands on the feature data of recommendation systems. In order to meet the needs of recommendation algorithms for massive amounts of features and real-time data processing, big data tools are needed to process the data and information. Based on big data and deep learning techniques, this paper constructs a recall ranking algorithm for recommendation systems and designs a complete recommendation system with movie recommendations as the theme. This paper investigates the optimization of new enterprise management platforms based on the background of big data algorithms, which have made a great breakthrough in big data innovation.

Keywords: Financial regulation; Big data; Recommendation algorithms; Evolutionary gaming

1 Introduction

With the rapid development of the Internet and the spread of smartphone devices, as well as the improvement in the quality of people's living standards, everyone can participate in the production, dissemination and consumption of information in the online world [1]. A wide variety of online services have also emerged, flooding people's eyes with information, and with it, an explosion of information. There is no doubt that we have entered an era of information explosion, and it has become a challenge to search for information of interest in the vast amount of information, and the recommendation system is one of the effective tools to solve this contradiction. With the rapid development of financial technology, financial institutions have seized the opportunity of technological innovation and actively developed financial technology by taking advantage of the latter [2]. Commercial banks are an important part of China's financial institutions, with a huge scale of assets, are the main players in China's financial activities, and are also the financial institutions most closely connected with the financial life of the people. This paper is based on the Internet's big data background to optimize the financial platform research, the optimization of the financial platform research for the entire financial big data professional research has a great role in paving and promoting [3].

MSEA 2023, May 26-28, Nanjing, People's Republic of China

DOI 10.4108/eai.26-5-2023.2334465

2 Relevant technical and theoretical foundations

Recommender systems address the problem of recommending items of interest to users in specific scenarios with large amounts of information [4]. The recommendation system can be formally defined as an abstract function $f(U,I,C)$ that takes in feature data such as a specified User, a particular context and a specified Item, predicts the user's preference for the candidate item, and sorts the item by preference procedure to obtain the final list of item recommendations, as shown in Figure 1.

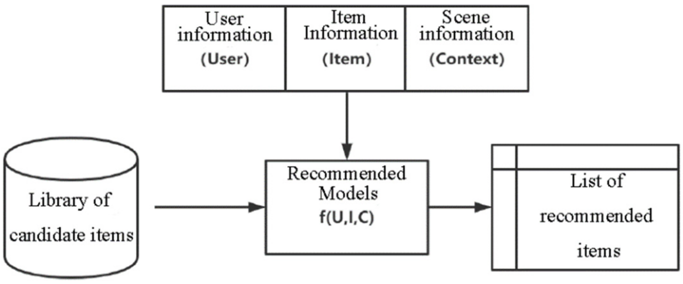

Fig. 1. Logical architecture of the recommendation system

As can be seen, at the heart of the recommendation system is the recommendation model, the abstract function $f(U,I,C)$. The recommendation model determines whether the recommendation system can accurately capture the user's interests and thus provide accurate recommendation results [5].

2.1 Recommendation algorithms

In order to handle recommendations in complex business scenarios and big data scenarios, the business process of industrial grade recommendation systems is generally divided into three phases: recall, sorting and re-ranking. The recall layer is responsible for carefully sorting the recalled items according to their probability of being clicked by the user, often using complex deep learning models to fully explore the user's interest in order to improve accuracy. The three stages in tandem form the pipeline architecture of the recommendation algorithm [6]. As show in Figure 2.

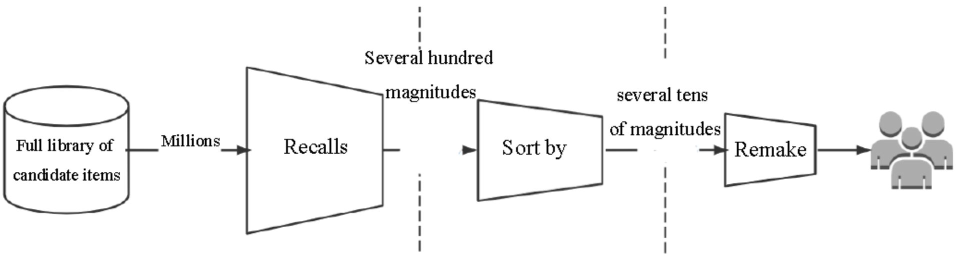

Fig. 2. Recommendation algorithm pipeline architecture

2.2 Classical recall layer model

The input features for the Wide section are the installed application and the application to be recommended. The two types of features are combined and the combination function is known as the Cross Product Transform and is defined as shown below:

$$\phi_K(X) = \prod_{i=1}^{d} x_i^{c_{ki}}, c_{ki} \in \{0,1\} \quad (1)$$

where x_i is the value of the ith feature and c_{ki} is a Boolean value. c_{ki} takes 1 when the ith feature belongs to the kth combined feature and 0 otherwise.

The structure of the Deep part is typical of the embedding layer + MLP, and the combined features of the Wide part are fed into the final LogLoss layer together with the Deep part for the final fit.

Wide & Deep continues to be influential today, and the idea of its combined model has inspired a large number of subsequent hybrid models that use Wide & Deep as their base structure.

2.3 DIN

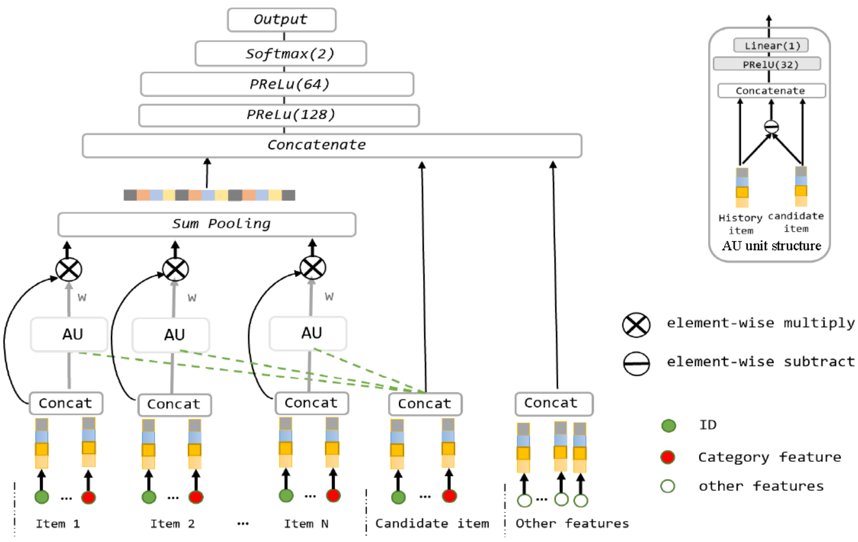

Fig. 3. DIN network structure

As show in Figure 3.DIN (Deep Interest Network) is a recommendation model used by Alibaba for e-commerce advertising, and the most important innovation is the introduction of the attention mechanism into the recommendation model. The DIN model is based on the assumption that different items in a user's historical behaviour sequence contribute differently to predicting the click-through rate of the current candidate ad. Such an assumption comes from the engineers' precise grasp of the business and deep insight into user behaviour [7].

The attention weights for the different items in the user's behaviour are calculated in DIN by the attention unit AU (top right in Figure 3), which is structured as a small neural network. The inputs are the embedding vector of the user's historical behavioural goods and the embedding vector of candidate advertisements, and the attention weights are obtained by joining the two input vectors and the result of their outer product over the Relu fully connected layer, and the weights are calculated as shown in equation 2:

$$V_u = f(V_a) = \sum_{i=1}^{N} w_i * V_i = \sum_{i=1}^{N} MLP(V_i, V_a, V_i\theta V_a) * V_i \quad (2)$$

Where V_u, V_a is the embedding vector of users and candidate ads respectively, and V_i is the vector of items in the user's history of behaviour. The DIN model brought new ideas to recommendation algorithms, and many subsequent recommendation algorithms introduced attention mechanisms, making the attention component a standard in the recommendation field.

2.4 Recall layer sorting layer comparison

As can be seen from the previous classical models, the recall layer has a simple and efficient recall strategy with a relatively simple model structure and feature combination approach, allowing the recommender system to build recall sets quickly; the ranking layer makes use of all features as much as possible, with various modules to make the input features fully crossed to improve accuracy as much as possible [8]. We have summarised the differences between the two in the following table 1:

Table 1. Comparison of the recall layer with the sorting layer

	Candidate set size	Model Complexity	Number of features	Speed	Precision
Recall layer	Millions	Simple	Less	Quick	Poor
Sorting layer	Hundred magnitude	Complex	More	Slower	Good

3 Big Data Processing Technology

3.1 Spark

The input/output and intermediate data in Spark are represented as Resilient Distributed Datasets (RDD), on which various data operations such as map(), reduce() and join() can be executed, and the computational logic and data partitioning information of the data are encapsulated in the RDD [9]. Spark's logical processing flow is a Directed Acyclic Graph (DAG). During physical execution, Spark divides a specific Task into multiple Stage phases based on the DAG and the sequential dependencies of the different RDDs. As show in Figure 4.

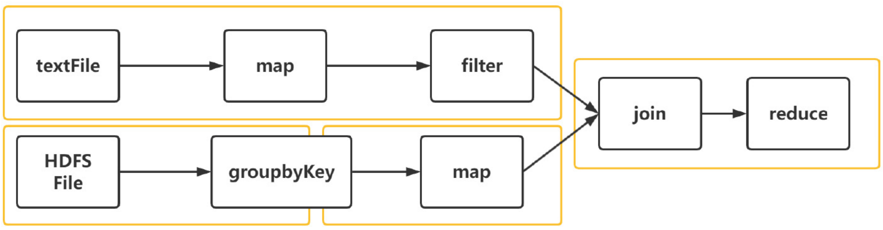

Fig. 4. DAG example with Stage division

The Spark ecosystem is rich, supporting multiple deployment models such as Standalone, YARN, Mesos and K8s, as well as multiple persistence layer access operations such as HDFS and HBase [10].

3.2 Flink

The Flink runtime architecture consists of two main processes, JobManager and TaskManager. The JobManagerr is responsible for task scheduling, failure handling, and coordinating checkpoints for Flink programs; the TaskManager is responsible for executing specific Flink jobs. The smallest unit of resource scheduling in Flink is the slot, which represents the number of tasks that can be executed in parallel in the TaskManager. As show in Figure 5.

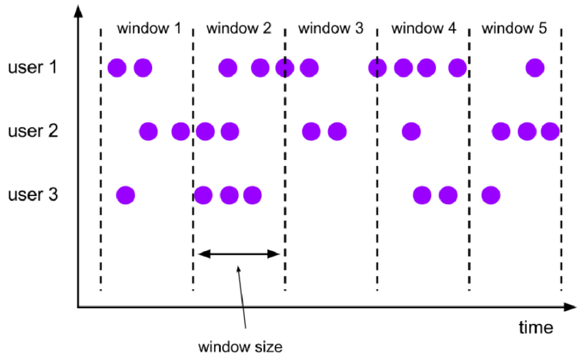

Fig. 5. Example of a Flink scrolling window

Flink uses windows to cut an infinite stream of data into finite sets for processing, with two categories: time windows and count windows. Time windows generate windows according to time, while count windows generate windows according to a specified number of data items.

3.3 HBase

HBase is a distributed, highly available NoSQL database, built on top of HDFS. HBase is used extensively in the data storage of this paper, so it is briefly described. As show in Figure 6.

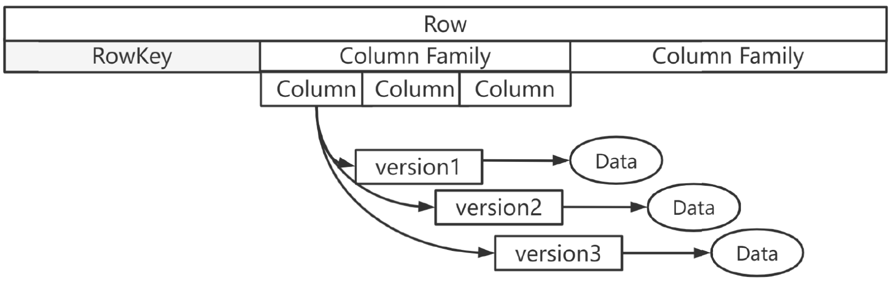

Fig. 6. HBase logical storage structure

A row of data in an HBase table consists of a RowKey and multiple Column Families, which consist of multiple Column Qualifiers, i.e. real columns, in which the real data is stored in each column.

HBase uses a timestamp to identify different versions of the same row of data. If this field is not specified when entering data, the system will use the time of writing as the default value.

Data in HBase is stored in dictionary order according to RowKey, and inserted data can only be indexed according to RowKey. A data cell (Cell) is uniquely identified by the format of the <RowKey, Column Family:Column Qualifier, TimeStamp> triplet.

4 Conclusion

This paper gives a detailed introduction to two very important parts of a recommendation system: the algorithm and the data part. Firstly the pipeline of recommendation algorithms is described, then the classical models of the recall and ranking layers and the characteristics of the two types of models are summarised and compared. The data section provides a detailed explanation of the big data processing architecture, the principles of Spark and Flink, two powerful tools for batch and stream processing, and a brief introduction to the storage structure of HBase, laying the foundation for the subsequent chapters.

References

[1] He Baocheng, Song Mengmeng. How fintech empowers corporate R&D innovation - the moderating role of big data development level[J]. Friends of Accounting,2023(09):9-15.

[2] Wang Xiaobo. The development trend of harmonization between enterprise economy and financial technology[J]. National circulation economy,2023(07):165-168. doi:10.16834/j.cnki.issn1009-5292.2023.07.028.

[3] Wang YZ. The era of "technology flow" is here [N]. Fujian Daily, 2022-05-20(003). doi:10.28232/n.cnki.nfjrb.2022.001686.

[4] Zhu Xiaoyue. Exploration of the impact of financial technology on small and micro enterprises in the era of big data[J]. Mall Modernization, 2022(02):98-100. doi:10.14013/j.cnki.scxdh.2022.02.037.

[5] Yan Lulu. Big data credit to help innovation and reflection of financial services for small and micro enterprises[J]. Marketing World, 2022(01):11-13.

[6] Ding Lianye. Big data finance: innovation and reflection of financial services for small and micro enterprises[J]. Southwest Finance,2021(07):62-73.

[7] Zhang M. The innovation strategy of financial service model for small and micro enterprises under the environment of big data[J]. Journal of Beijing Printing Institute,2021,29(05):39-41.DOI:10.19461/j.cnki.1004-8626.2021.05.012.

[8] Li S. Reflections on the development path of fintech in the era of 5G intelligence [J]. Financial Technology Times,2021,29(04):14-18.

[9] Zhou B, Wang X. Research on the construction of Nanjing SME financial service platform under the background of big data[J]. Contemporary Economy, 2020(02):69-71.

[10] Song Hua. Smart supply chain finance [M]. People's University of China Press:, 201909.528.