# A Study on the Economic Impact of the New Materials Industry on Regional Development Based on Big Data

Wenting Fan

Email: lunafwt@163.com

Ningbo Tongshang Group Co,Ltd;, Ningbo, China

**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 techniques 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 the real-time processing of data, big data tools are needed to process the data and information. This paper builds a new approach to materials industry research and research content based on big data and deep learning technology to lay a very solid foundation and combination for subsequent research in related fields. A new research breakthrough in a new field.

**Keywords:** Big Data; Industrial economy; Regional developments; Environmental Economics

## 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]. The 14th Five-Year Plan proposes to vigorously develop strategic new industries and promote practical breakthroughs in the research and development, production and application of high-end new materials [2]. The new materials are made from aluminium ingots, which are used as the main raw material, by adding a quantitative amount of other metal elements, after a series of processes such as melting and casting, extrusion, cutting and sawing, painting, etc. The alloy products are mainly used in many industries such as construction, transportation, electric power, electromechanical equipment, aerospace and aviation as metal modelling materials. Aluminium profiles are hard, wear-resistant and resistant, lightweight and easy to process, and cheaper compared to copper [3]. In recent years, with the accelerated construction of China's infrastructure and the development of the manufacturing industry, the new aluminium-related materials industry has also achieved rapid development.

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334415

## 2 RECOMMENDER SYSTEM OVERVIEW

Recommendation systems have to solve the problem of recommending items of interest to users in a specific scenario with a huge amount of information. The recommendation system can be defined formally as an abstract function $f(U, I, C)$, which takes data on the characteristics of a given User, a given Context, and a given Item, predicts the user's preference for the candidate item, and sorts the items by preference to obtain the final list of item recommendations [4].

### 2.1 Classical recall layer model

(1) Collaborative filtering algorithm family

Collaborative filtering relies entirely on the behavioural relationships between users and items to make recommendations, the idea of which can be summarised as "things come in groups, people come in groups". The principle of item-based collaborative filtering is that "things are clustered together", i.e. the items that are most similar to the items that the user has acted on are recommended to the user. The preference of user u for an item v, $sim(u, v)$, can be expressed as:

$$sim(u, v) = \sum_{v_i \in V} score(u, v_i) \times sim(v_i, v) \tag{1}$$

where is the set of items for which the user has generated behaviour, $score(u, v_i)$ is the degree of preference of user u for item $v_i$, and $sim(u, v)$ is the degree of similarity between item $v_i$ and item v.

The principle of user-based collaborative filtering is that "people are divided by groups", i.e., the items most similar to the user are recommended to the user, and the preference of user u for an item v, $sim(u, v)$, is expressed as follows:

$$sim(u, v) = \sum_{u_i \in U} sim(u, u_i) \times score(u_i, u) \tag{2}$$

where U is the set of similar users of user u, $sim(u, u_i)$ is the degree of similarity between user u and user $u_i$, and $score(u_i, u)$ has the same meaning as Eq.

The above two equations need to calculate the similarity between users or between items, in the co-occurrence matrix users and items are reflected in the form of row vectors and column vectors respectively (as shown in Figure 1), the similarity between vectors can be calculated using the cosine similarity method of equation 3.

$$sim(i, j) = \cos(i, j) = \frac{i \cdot j}{\|i\| \cdot \|j\|} \tag{3}$$

Collaborative filtering is simple, intuitive， and highly interpretable, but weak in generalisation. Top items, due to the high number of reviews, are easily associated with a large number of items and thus similar, leading to a clear head effect; while long-tail items have few reviews.

The vector is sparse and difficult to associate with other items, resulting in few recommendations. The matrix decomposition algorithm generates a hidden vector for each user and item so that the user and item can represent similarity in a space of hidden vectors of equal dimensionality, essentially decomposing the co-occurrence matrix into the form of a product of hidden vectors. The hidden vectors are denser and therefore enhance the ability to handle sparse matrices.

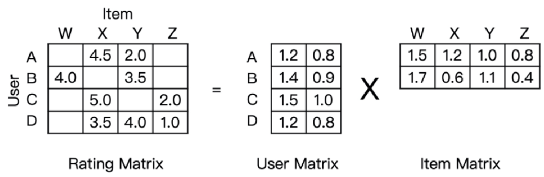


**Figure 1:** Schematic of matrix decomposition

Alternating least squares (Altering Least Squares, ALS) is used in Spark MLlib to solve for matrix decomposition hidden vectors, and we give an introduction to the underlying principles.

Let user u's prediction score for an item v be $\hat{r}_{uv} = p_u * q_v^T, \Delta r = r_{uv} - \hat{r}_{uv}$ indicating the error between the true value and the predicted value, the smaller the$\|\Delta r\|$, the more accurate the prediction. Transforming the above problem into an optimization problem of finding the minimum of $\|\Delta r\|$ by adding the regularization term, the objective function obtained is shown below:

$$\underset{p^*, q^*}{min} \sum_{(u,v)\in A} (r_{uv} - p_u * q_v^T)^2 + \lambda( \|p_u\|^2 + \|q_v\|^2) \tag{4}$$

ALS finds the minimum value by alternating optimization as follows:

(1) Generate random initialization values for $p_u, q_v$.

(2) Fix $p_u$ and solve for the minimum value of $q_v$ by gradient descent.

(3) Fix the current $q_v$ and solve for the minimum value of $p_u$ by gradient descent.

(4) Repeat step (2) and step (3) until the objective function converges or the maximum number of iterations is reached.

## 2.2 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 and makes the input features fully crossed through various modules to increase accuracy as much as possible. We have summarised the differences between the two in the following ways [5].

**Table 1:** Recall layer versus sorting layer.

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

# 3 BIG DATA PROCESSING TECHNOLOGY

## 3.1 Big Data Processing Architecture

The development of Big Data technology has gone through a process of development from batch processing to stream processing and then to batch-stream integration. During its development, two architectures, Lambda and Kappa, have emerged [6].

## 3.2 Speak

Spark, as a distributed computing platform, adopts a standard master-slave architecture. The Driver is responsible for task scheduling in the whole cluster, and the Executor executes the actual Task tasks and returns the finally result to the Driver after the execution is completed [7]. During the physical execution of the Executor, the data in a Task will enter different Partitions for parallel computation according to the Partitioner, whose architecture diagram is shown in Figure 3.

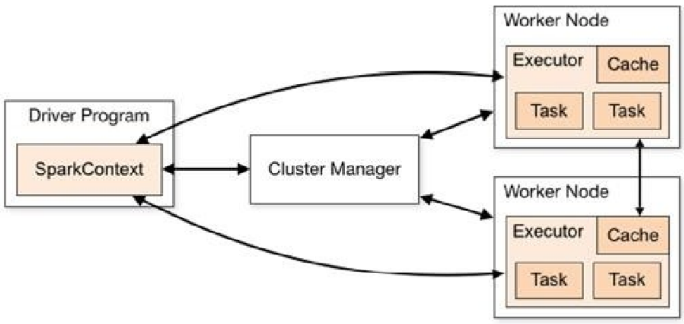


**Figure 2:** Spark Architecture

The input/output and intermediate data in Spark are represented as Resilient Distributed Datasets (RDD), which can perform various data operations such as map(), reduce(), join(), etc. The RDD encapsulates the computational logic and data partitioning information of the data. Spark's logical processing flow is a Directed Acyclic Graph (DAG). During physical execution, Spark will divide a specific Task task into multiple Stage phases based on the DAG and the sequential dependencies of the different RDDs.

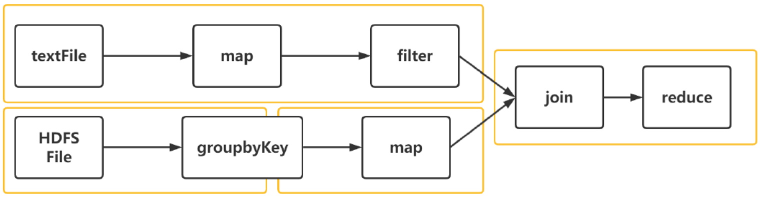


**Figure 3:** DAG example with Stage division

The Spark ecosystem is rich, supporting various deployment models such as Standalone, YARN, Mesos and K8s, as well as various persistence layer access operations such as HDFS and HBase. Spark components are built with Spark Core as the core for big data processing, machine learning, graph computing and other ecologies.

### 3.3 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 [8].

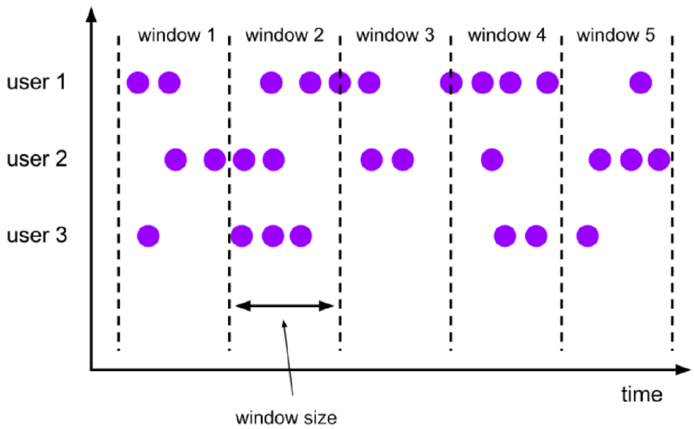


**Figure 4:** Flink scrolling window example

Flink cuts an infinite stream of data into finite sets for processing by means of windows, of which there are two types: temporal windows and counting windows. Time windows generate windows according to time and counting windows generate windows according to the number of data bars specified, where time windows can be divided into scrolling, sliding and session windows according to the principle of implementation [9].

# 4 ANALYSIS OF THE EXTERNAL ENVIRONMENT OF S NEW MATERIALS

## 4.1 Analysis of the economic environment

Since 2008, the country began to accelerate the upgrading of road traffic, the national power grid and other infrastructure construction, with mechanical and electrical equipment, made in China to the world, the market for aluminum wire and cable, aluminum alloy rods, aluminum enameled wire and other materials increasing demand. Due to their lightness, corrosion resistance and relatively low price, new materials involving aluminium are gradually being used in a wide range of fields [10].

## 4.2 Technical environment analysis

China Nonferrous Metals Industry Association has held the "2017 China International Aluminium Week" with the theme of "industry chain, environmental protection and application", to discuss the development direction of aluminium-related industries, market dynamics, technical cooperation and the construction of upstream and downstream industry exchange platforms. In order to promote the high-quality development of aluminium-related industries, the State Ministry of Industry and Information Technology and relevant departments have carried out a series of work from various aspects and achieved certain results. The aluminium-related industry is an important industrial category in the construction of China's modern economic system. The aluminium-related industry actively responds to market changes, earnestly implements the requirements of the new development concept, pays more attention to development quality and development efficiency, carries out technological innovation and transformation and upgrading of the aluminium-related industry, continuously expands the practical application of new aluminium-related materials, and promotes the development of the aluminium-related industry to achieve multiple changes in quality, efficiency and power [11].

## 4.3 The threat of alternative products

From the perspective of the development of new materials involving aluminium, new materials involving aluminium have strong electrical conductivity, good ductility, strong corrosion resistance, lighter materials and other characteristics, and their products themselves appear as a substitute for products involving copper. Copper materials due to the high price of raw materials, not high temperature resistance, heavy quality, and other reasons, in the product development, production, application and other aspects of a greater degree of restriction, involving the new aluminum new materials once put into the market has been widely welcomed by all parties. Although the raw material as bauxite also has the characteristics of scarcity, but compared with copper ore, its reserves are large, low price, light material, more application prospects. In other alternative research and development production, due to the current market for new materials related to aluminium is widely optimistic, new materials related to aluminium can meet the vast majority of the market demand, new materials related to aluminium can replace a small variety of products, especially in wire and cable, aerospace, automotive lightweight, electromechanical manufacturing, etc., has not yet appeared new alternatives, which is also in recent years new materials related to aluminium enterprises generally develop faster, profit This is also an important market environment for the faster development and more substantial profits of new aluminium-related material enterprises in recent years. As show in figure 5.

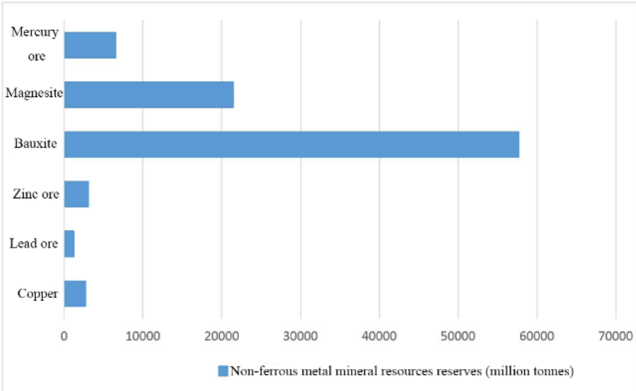


**Figure 5:** Selected non-ferrous metal mineral reserves in China in 2021

### 4.4 External factors evaluation matrix

Based on the key external factors of S New Materials Company, an external factor evaluation matrix was constructed to evaluate the external factors of S New Materials Company. In terms of specific evaluation, 10 experts from the aluminium-related new materials industry and 10 members of the management of S New Materials were invited to study and evaluate the weighting and scoring of each factor. First, each person was assigned a value for each indicator individually, and then the average of the 20 people's ratings was calculated as the basis.

## 5. CONCLUSION

This article gives a detailed introduction to two very important parts of a recommendation system: the algorithm and the data part. The pipeline of recommendation algorithms is first described, followed by a summary comparison of the classical models of the recall and ranking layers and the characteristics of the two types of models. 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 subsequent use. This paper combines the new technology of computer big data with the new materials industry for research and exploration, and has made some significant developments in the new field, laying a certain foundation for the subsequent research.

## REFERENCES

[1] Big data industry builds new advantages of digital economy[J]. Software and integrated circuits,2022(08):62.

[2] Huang Fenmei,Wei Chaoying,Lu Meifang. Data mining and decision making in agricultural financial statistics in the context of big data application--a review of "Big data in agriculture and quantitative analysis of agricultural industry economy"[J]. Journal of Tropical Crops,2021,42(02):637.
[3] Li Yesong. Exploring the design of regional industrial economic big data platform[J]. Shanxi Agricultural Economics,2020(15):34+48. DOI:10.16675/j.cnki.cn14-1065/f.2020.15.019.
[4] Lu Shan, Zhu Biqin. Design of analysis tool for economic development planning of county oil tea industry based on big data analysis[J]. China New Technology and New Products,2022(06):137-139.DOI:10.13612/j.cnki.cntp.2022.06.015.
[5] Pei Yan,Su Le,Wang Yue. Analysis of economic impact of big data industry based on input-output model[J]. Telecommunications Network Technology,2018(05):56-60.
[6] Ren Yao. Research on the impact of big data industry development on economic growth in Guizhou Province [D]. Guizhou University of Finance and Economics, 2022. doi:10.27731/d.cnki.ggzcj.2022.000317.
[7] Wu Xiaolin,Zhang Meng. Research on innovation-driven economic development of information industry--based on the perspective of big data application in broadcasting network[J]. Radio and Television Network,2020,27(10):92-95.DOI:10.16045/j.cnki.catvtec.2020.10.027.
[8] Xu Zhan, Mai Shiyu. The mechanism of the role of big data industry in promoting sustainable regional economic development--a review of "Sustainable Economic Development and Industrial Structure Evolution"[J]. Business and Economic Research, 2022(21):193.
[9] Wang Junchao. Analysis of industrial economic information in the era of big data and its application in macro decision making[D]. China University of Geosciences (Beijing), 2016.
[10] Zhang Qiwen. Economic status and development construction of China's medical big data industry[J]. Tianjin Science and Technology,2020,47(06):79-80+84.DOI:10.14099/j.cnki.tjkj.2020.06.023.
[11] Zhou Benhai. Analysis of the impact of big data on the economy of media industry[J]. Journal of Economic Research,2015(05):32+77.