

Design and Implementation of Agricultural Production and Market Information Matching Recommendation System in the Cloud Environment

Tao Chen¹

Key Laboratory of Agri-information Service Technology, Ministry of Agriculture, Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing, 100081, China
E-mail: lotus.ct@hotmail.com

Shihong Liu²

Key Laboratory of Agri-information Service Technology, Ministry of Agriculture, Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing, 100081, China
E-mail: lotusct@163.com

At present, in China, small farmers decentralized production doesn't keep pace with the agricultural products market development requirements. This paper provides a new way by using the cloud computing technology to design and implement the agricultural production and marketing information matching recommendation system in the cloud computing environment. The platform collects agricultural market information from internet through NUTCH (an open source search engine), builds a market space data warehouse and data sources, studies and proposes the user's interested model, which is more suitable for the agriculture-related objects and agricultural production and marketing docking combination push algorithm. Studying and setting up three subsystems of the platform: the market price data acquisition subsystem in respect of the agricultural products based on WEB, the agricultural products deal-making subsystem and the recommendation subsystem of marketing network information. The system's implementation results show that it runs normally, gains good economic benefits and achieves the expected design purpose.

*CENet2015
12-13 September 2015
Shanghai, China*

¹Speaker

²Corresponding Author

1. Introduction

With the relentless march of the information technology and internet technology, the agricultural information in China has obtained considerable development and also confronted with ushered challenges in the era of big data. The internet development enables farmers to search their own interested web information, most of information are useless thought[1]; thus it becomes difficult for farmers to get the information that they indeed demand. It thus results in the problems of information explosion and information overload [2]; at meanwhile, it becomes more difficult to access the demanded agricultural market information in an efficient, timely and accurate manner, which sharpens the mismatch of agricultural production and market liquidity. The information-match recommendation system can well solve these problems, but with the generalization of information explosion, the system also faces such difficulties as storage scalability, analysis, calculation efficiency and the like. The cloud computing, as a new business model, has attracted the public attention[3]. Hadoop, acting as a cloud computing platform, is easy for the development and could parallel the process mass data. It is characterized by scalability, low cost, high processing, efficient computing and sound reliability; therefore, a new matching system is imperative to provide farmers with high-precision and quick agricultural market information. This system combines the existing information-match recommend system and the Hadoop platform to establish the cloud computing analysis model based on accurate network-wide user interest, collect network-wide market information and build large data centers. The establishment of the agricultural market information supporting platform, which can gather and analyze amounts of data, may solve the contradiction between the "small production" and the "big market"; and this action will have far-reaching significance.

2. Overall Framework

In the agricultural production and marketing information matching recommendation system, there should be a clear direction of data flow and information flow to define the division and logic association among various business levels; therefore, the agricultural production and marketing information matching recommendation system is divided into four levels, specifically, the basic data layer is mainly for collecting and storing the market data of grain, vegetables, fruits and livestock; the data processing layer uses the model algorithm library to process, cluster and classify the data as uploaded and then saves the processed data in the way of HDFS-distributed storage so as to build the basic data warehouse; the application layer, with the integration of platform technology, use the data of data warehouse to feature extract, constructs the vector space model for matching the recommended match algorithms library with the user interest model, and then inputs the results into the system; and the display layer, through mobile phone, computer, TV and other media, transmits the forecasting results to farmers, government agencies, research institutions and consumers to realize "the combination of production and sales " and thus serve the society.

Researches on the agricultural production and market information matching recommendation system under the cloud computing environment involve the Nutch open source search engine, the Hadoop distributed cluster, HTML content extraction and processing, user interest model building, information recommendation, matching algorithms and the like. The technical roadmap is as shown below:

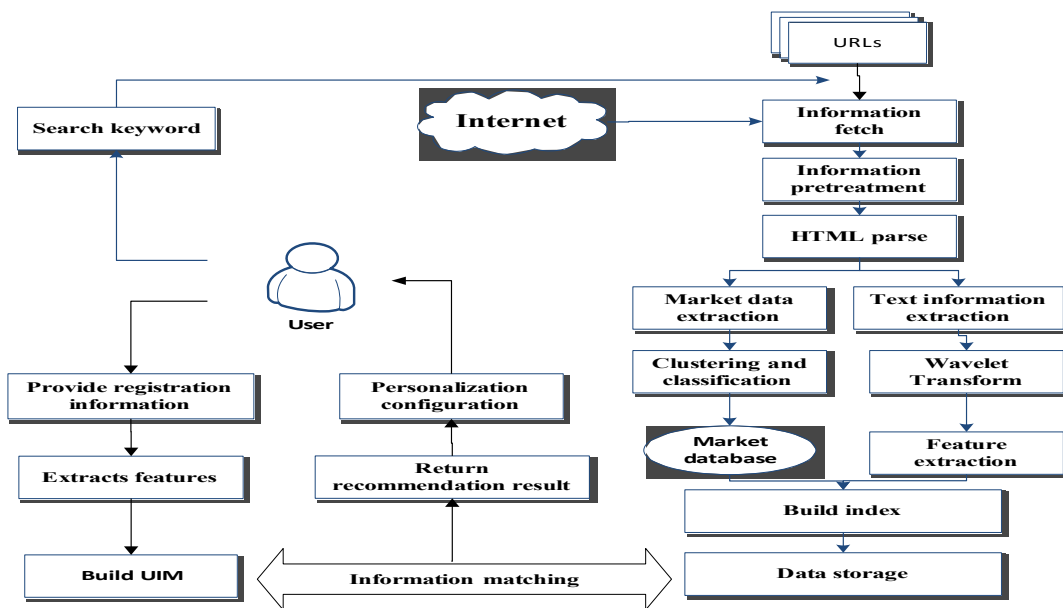


Figure 1: Technology Roadmap

3 System Design

3.1 Design of the Data Acquisition Subsystem in respect of Agricultural Products Market Price Based on WEB

The construction process of agricultural market database includes: the selection of web data sources, analysis of the web structure, selection of the information extraction methods, writing the extract code and storage of the extraction results. Thus the subsystem can be divided into the module of market prices crawling, the module of web content analysis and data storage in respect of the agricultural products' market price. The web data sources are picked from a lot of authority agricultural sites based on the actual application requirements; the web information analysis comes from the feature extraction of climb back WEB data; in terms of the data extraction: select the appropriate algorithm on the structural characteristics of Web pages to extract the interesting content; store data upon processing in a data set of data warehouse. The construction process diagram is shown as follows:

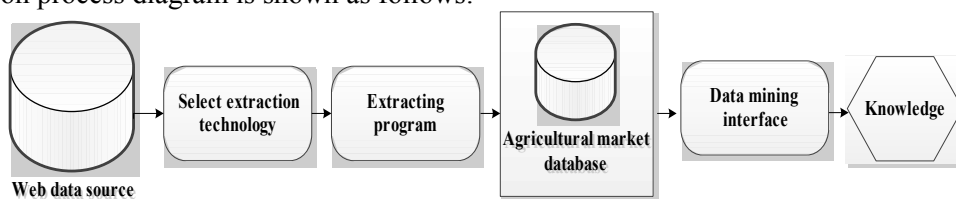


Figure 2: Flowchart of Agricultural Market Database Construction

Currently, most agricultural sites use HTML syntax structure. The system is to crawl the web pages having similar structural features, such as data storage in the table form. The specific market data is often nested in <table> tag or <td> tag. Upon analysis in detail, we can know that those agriculture website pages have the following characteristics: several data tables are nested in the same page; the agricultural market data is usually nested in <td> </td> in <table> tag, and it takes up most of the page layout.

POS (CENet 2015) 045

3.2 Design of the Recommendation Subsystem in respect of the Network Market Information

The subsystem includes the information acquisition and the pre-processing, establishment of the User Interest Model, design and implementation of the recommendation algorithm.

(1) Information Acquisition and Preprocessing

We need to obtain the information sources in this subsystem covering the product and sales information as well as relevant information from the Internet. The data processing of the recommendation pretreatment layer prepares for the recommendation layer. There are two plans to access the information source. On the one hand, use open source search engine; and on the other hand, use the cloud engine. Such two plans have their advantages and disadvantages; and we adopt the first one in the subsystem.

(2) Design User Interest Model

The subsystem establishes the user's interest model by means of display and implication. In terms of establishment of the display mode, get the user interest source through multiple channels, proceed features extracted from the user's information source upon the wavelet feature extraction[4] and use the notation to represent the user's interest features in a vector space. In terms of establishment of the implication mode: obtain the user's interest source by collecting his/her information in the course of user accession to information[5], after screening, the wavelet feature extraction, feature extraction and forgetting algorithm to establish the model. This process makes the calculation to become more convenient and more accurate in respect of characterizing the user's interest [6]. The flowchart of establishing the User Interest Model is shown in Fig. 3:

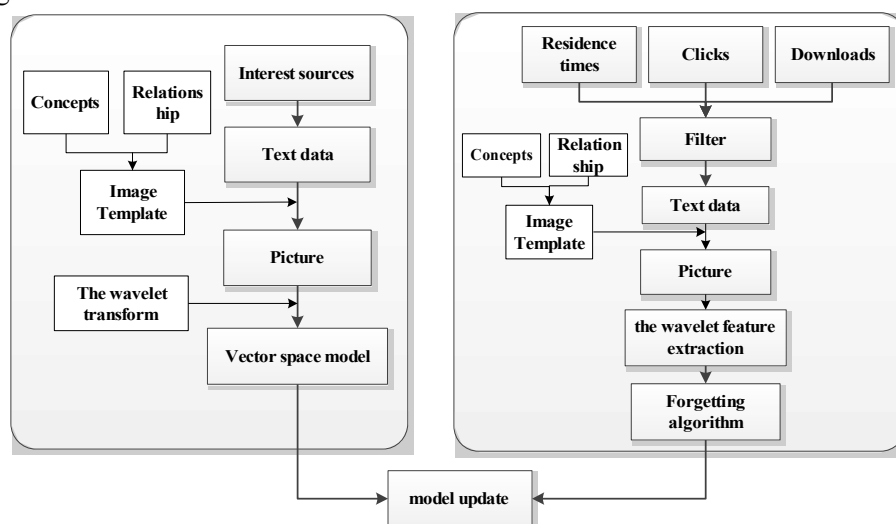


Figure3: Flowchart of Establishing the User Interest Model

The wavelet transform feature extraction can be realized by two steps [4]: Step1: establish image template via specific rules: concept, conceptual relations and the like, then convert text into image information; Step2: select the appropriate wavelet bases, perform wavelet transformation on image, extract the image feature as the features of the original text.

(3) Recommendation Algorithm Implementation Module

The recommendation algorithm of the system is divided into two modules. As to the first module, use the content-based recommendation algorithm through BP neural network to retrieve the feedback data classification according to the custom themes and put pages of the same characteristics together[7]; as to second module, use collaborative filtering recommendation algorithm based on projects by SOM neural algorithm to match the cluster and UIM. The page whose similarity is greater than the set threshold would be pushed to users.

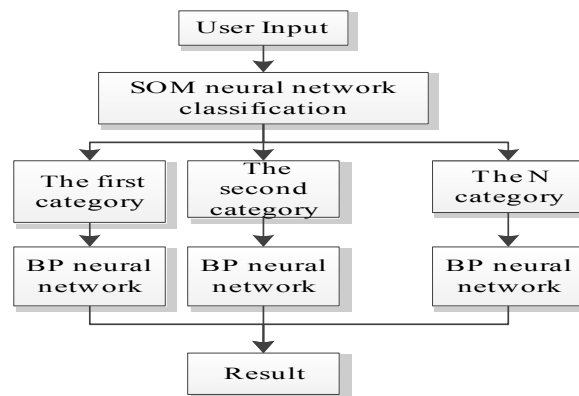


Figure 4: Flowchart of Combined Neural Network

Step1: read input and processing; Step2: preliminary classification by SOM neural network; Step3: select appropriate BP neural network algorithm and parameters, and make adjustments; Step4: calculate the output value of each layer; Step5: calculate and adjust the error; Step6: calculate the output of combined neural network.

As to advantages of this algorithm, it can not only provide retrieve information for users, but also retrieve associated content as pushed to the users. When the user submits a request, the system will convert the input into feature vectors, connect input and output by weighting the characteristics, for example, when the user retrieves "corn acquisition", the system will not only return relevant pages in respect of such information as "corn acquired information" but also "corn policy guidance", "corn market forecast", "corn planting technology", "corn seed information" and other topics as related to the information as recommended to the user.

3.3 Design of the Agricultural Products Deal-making Subsystem

The subsystem can provide the purchase and sales of goods as released, query and match functions for the agricultural production users. The agricultural traders can make deal via the system. In order to ensure the system's ability, including process information and the speed of feedback while taking into account the sustainability of agricultural production and trade, the subsystem adopts the framework as shown in Fig. 5:

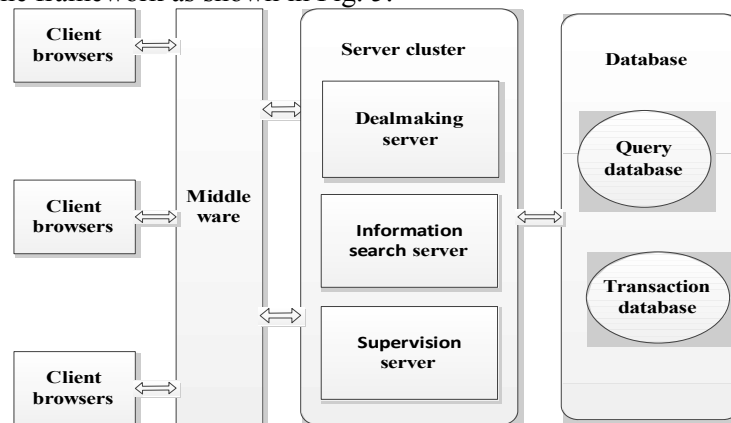


Figure 5: The Subsystem Framework

The subsystem consists of four parts, namely, the business logic layer, the intermediate layer, the expression layer and the data accession layer. This subsystem is to realize the matching between the market information in respect of the agricultural products and their transaction information. The main function modules cover: (1) the user's identity management module for the purpose of the user's registration, authentication and login; (2) the module of releasing the supply and purchase information in respect of agricultural products to facilitate information dissemination, release the supply and demand information on the system for the

sake of transaction inquiries and negotiations among parties concerned; besides, it also provides a data base for the matching module; (3) the agricultural products dealmaker module for the online deal-making of agricultural products. It is primarily to design the program and parameters of deal-making, provide traceability services for agricultural products and solve problems in respect of the traceability management; (4) the system management module which is used by the system administrator to manage the user's privilege, role, menu, transaction data, users and price information; (5) the agricultural products market data analysis module according to the information of web-based agricultural price data subsystem, which provides users with analysis and forecasting of the agricultural market price.

4. Realization of System

4.1 Development Environment

The system can be divided into the information publishing subsystem and the WEB information subsystem, both of which are using B/S structure. The information publishing subsystem adopts C# language on the basis of .NET platform while the WEB information subsystem uses Java language development on Eclipse to build the theme of search engine on the basis of NUTCH & SOLR & MYSQL, and build the distributed system on Hadoop; finally, in the Linux environment to build a Hadoop & Nutch clusters, adopt Tomcat server and SQL server and use MySQL database as a background database.

4.2 System Realization

This study selects a number of sites for test from the authoritative agricultural portal. Upon retrieval, analysis and testing for three months, we chose Anhui Agriculture website (<http://www.ahnw.gov.cn/>), Agriculture website (<http://www.hnnw.net/price/>) and other 80 sites as the main crawling sources of website. After pretreatment, we used Heritrix crawl data with the main interface shown as follows:

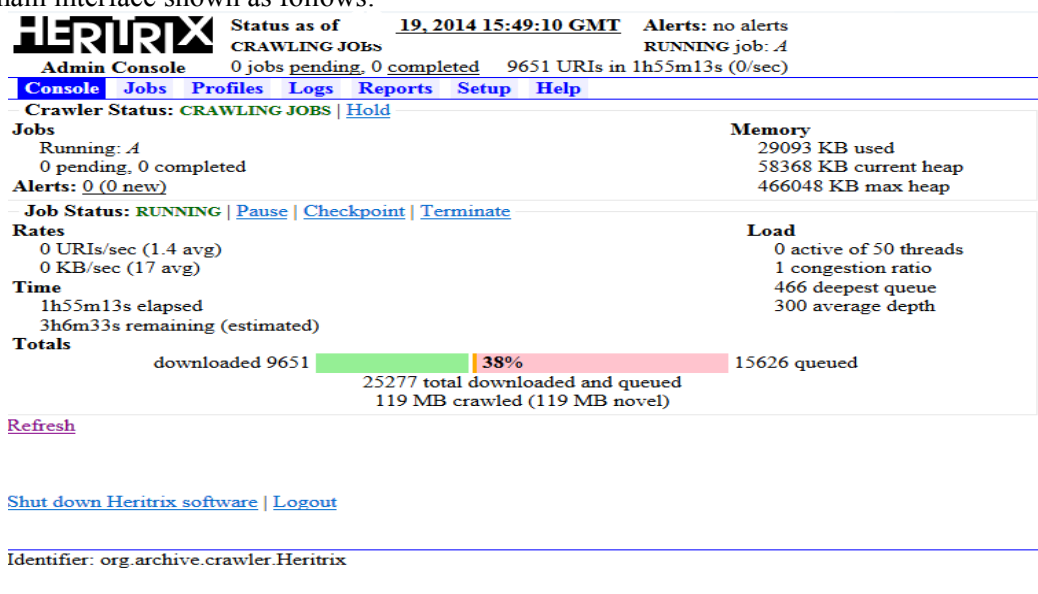


Figure 6: Process and Results of Grasp

5. Conclusion

In order to meet the user's demand of agricultural production and market information so as to improve the transaction efficiency of agricultural products and maximize the interests of agricultural workers. The recommendation system in respect of the agricultural production and the market information matching information is proposed in this paper by using the hybrid

recommendation algorithm to deal with arranging algorithm of agricultural products and using open source search engine technology under the cloud computing environment. Under the background of agricultural production in China, we firstly introduce the theory and cutting-edge technology which can solve the existing problems, such as cloud computing, open source search engine, information recommendation algorithms and dealmakers algorithm as well as key technical analysis; then the system includes three subsystems: the acquisition subsystem in respect of market price data of agricultural products based on WEB, the arranging subsystem in respect of deals of agricultural products and the subsystem in respect of production and sales recommendation across entire network; then we introduce the system functions module in detail and the implementation process of key technologies such as open source search engine technology, information fusion technology, coordination filtering and sorting technology page. As far as the experimental data and application are concerned, the system can ensure that the information of agricultural production and purchase can be fully indexed up to date with matching efficiency and accuracy to be improved at meanwhile.

References

- [1] Wenjing Zhang, Haixia Lu, Yanhong Feng, Peng Yang, Bin Dong. *The Application of Datamining in Agriculture Production*[J]. Journal of Agricultural Mechanization Research,2008(08):244-246.(In Chinese)
- [2] Hua Chen, *Algorithms recommend research on personalized search engine*[J]. Application Research Of Computers,2010,27(1):48-50.(In Chinese)
- [3] Xiaofeng Meng, Xiang Ci. *Big Data Management: Concepts, Techniques and Challenges*[J]. Journal of Computer Research and Development,2013,50(1):146-169.(In Chinese)
- [4] Hong Shen. *A Study on Feature Selection Methods Based on Wavelet Analysis and Semantic Information*[D]. Shanghai Jiao Tong University, Shanghai, 2005.(In Chinese)
- [5] Huimin Fang. *User Modeling of Personalized Web Interface Based on BP Neural Network*[D]. Henan University, Municipality of Kaifeng, 2008.(In Chinese)
- [6] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins. *Recommendation Systems: A Probabilistic Analysis*[J]. Journal of Computer and System Sciences,2001,63(1):42-61.
- [7] Duen-Ren Liu, Pei-Yun Tsai, Po-Huan Chiu, *Personalized Recommendation of Popular Blog Articles for Mobile Applications*[J], Information Sciences, 2011,181(9):1551-1572.