Hindawi
Advances in Multimedia
Volume 2022, Article ID 4070827, 6 pages
https://doi.org/10.1155/2022/4070827

Research Article

Research and Implementation of Digital Media Recommendation System Based on Semantic Classification

Xiaoguang Li

Ningbo University of Finance & Economics, Ningbo, Zhejiang 315175, China

Correspondence should be addressed to Xiaoguang Li; 201771451@yangtzeu.edu.cn

Received 12 December 2021; Revised 20 February 2022; Accepted 4 March 2022; Published 27 March 2022

Academic Editor: Qiangyi Li

In order to study the recommendation system of digital media based on semantic classification, the CF-LFMC algorithm based on semantic classification is proposed. Firstly, the traditional algorithm is analyzed. Aiming at some problems existing in the traditional algorithm, a clustering algorithm model based on term meaning and collaborative filtering algorithm is designed by combining the collaborative filtering algorithm and project-based clustering algorithm. Before analyzing sparse data, the cold start and timeliness of the traditional algorithm are improved. Secondly, the performance comparison of three cosine similarity calculation methods of experimental IBCF algorithm, the performance comparison between CF-LFMC algorithm and IBCF algorithm, and the performance comparison between CF-LFMC algorithm and CF-LFMC algorithm without the time function is carried out. The clustering value $N = 10$ in the CF-LFMC algorithm is taken as the experimental result; MAE values of both algorithms decrease with the increase of the nearest neighbor number k. When the number of nearest neighbors is small, MAE values of the two algorithms are close to each other. As the number of nearest neighbors increases, the accuracy of the algorithm does not improve significantly, and the calculation cost of the algorithm will increase with the increase of the number of nearest neighbors, so the number of nearest neighbors between 20 and 30 is more appropriate. CF-LFMC shows better accuracy, and the CF-LFMC algorithm improved by the time function has improved the accuracy, which is better than the traditional algorithm in accuracy.

1. Introduction

Today, we are in an information age. With the rapid popularization of Internet technology, Internet has become an indispensable component of family life [1]. At the same time, with the development of e-commerce, mobile Internet, Internet of things, and other technologies, Internet technology has penetrated into every aspect of life. Every industry generates a huge amount of data every day. We have entered an era of big data [2]. In the era of big data, it is difficult for people to obtain useful information due to the huge amount of information, that is, information overload. At present, the most common tool for people to obtain information is the search engine. However, the search engines of Baidu and Google do not consider the individual characteristics of users; for the same search conditions, the information presented to users is the same, which is difficult to meet the personalized needs. The concept of "personalized information services" provides different types of information services according to the characteristics of users. As an important branch of personalized service research, the recommendation system has attracted the attention of many researchers in recent years. The most important role of a recommendation system is to connect users and information; it obtains user preferences from user behavioral data and makes recommendations based on these preferences by mining potential information favored by users from massive network data [3]. At present, the recommendation system has been widely applied in major e-commerce websites. E-commerce companies use recommendation system to transform users' potential demand into actual purchasing power, so as to improve sales performance. In the context of big data era, the

recommendation system has its actual application scenarios in all walks of life [4]. More and more people realize the importance of recommendation system and have carried out extensive research. In addition to e-commerce websites, some community websites such as Douban and Sina Weibo have also achieved great success in the application of recommendation system [5]. In front of big data, who can find users' interests quickly and effectively will occupy important business opportunities. For this research problem, Fresa A. et al. proposed an improved algorithm based on SVD, which mainly improved SVD algorithm by using the gradient descent method. The purpose of using the arcane meaning model for collaborative filtering is to reveal hidden features that can explain the observed scores [6]. Ayachi et al. proposed the idea of collaborative filtering algorithm, which was then applied to news recommendation, which is a recommendation system in practical application [7]. Liu and Yun improved the accuracy of finding the nearest neighbor and reduced the sparsity of the matrix by studying the nearest neighbor method in the domain [8]. The research forms of the personalized recommendation system are still mainly focused on theoretical and experimental verification, and there are still many deficiencies in the actual application of recommendation system. For example, most of the data used in current experiments are explicit rating data given by users, but in practical applications, user behavior data is usually implicit, so the research on implicit data is worth paying attention to. However, there is not much involvement in the cold start of the recommender system and the extensibility of the model. On the basis of current research, the CF-LFMC algorithm proposed based on semantic classification is firstly analyzed, aiming at some problems existing in traditional algorithms; combined with project-based collaborative filtering algorithm and clustering algorithm, a collaborative filtering algorithm based on the argot meaning model and clustering algorithm is designed to improve the traditional algorithm on the issues of data sparsity, cold start, and timeliness previously analyzed; secondly, the performance of three cosine similarity calculation methods of experimental IBCF algorithm is compared: the performance comparison between CF-LFMC algorithm and IBCF algorithm and the performance comparison between CF-LFMC algorithm and CF-LFMC algorithm without time function; CF-LFMC shows better accuracy, and the CF-LFMC algorithm improved by the time function has improved its accuracy, which is better than the traditional algorithm in accuracy [9].

2. Methods

As an important method of image analysis, image classification is widely used in image search and image annotation. In some professional fields, image classification has achieved high accuracy, such as face recognition and handwritten number recognition. However, the classification of images is still a challenging problem due to various changes such as illumination, rotation, and scaling, as well as the complexity of image content itself [10]. Semantic-based image classification methods start from the image data itself and use specific feature extraction methods to extract semantic features in the image; on this basis, the semantic hierarchical structure of the image is established step by step; finally, the semantic features of the high-level image are used to classify the images. Semantic features of images are hierarchical, in which low-level image features are less abstract and highly correlated with the content of image data itself; the features of high-level images have high abstractness and low correlation with the content of image data. Therefore, a hierarchical learning model can be established to learn the image data in an unsupervised way to obtain the characteristics of the data itself; as levels increase, the learned features become higher-order feature descriptions of the input data.

2.1. Dimension Reduction of Arcane Meaning Classification Model. The reason for data sparsity is that the item vector is too long and the user has few scores on the item, resulting in sparse matrix data. Therefore, reducing the dimension of the item-user rating matrix to shorten the length of the item vector can effectively reduce the data sparsity. The arcane meaning model can be used to reduce the dimension of the matrix. The R-matrix is a user scoring matrix, from which LFM algorithm extracts hidden categories, which is mathematically expressed as matrix P and matrix Q multiplied. P matrix is the user-hidden classification matrix, where p_{ij} represents user U's interest in classification C. Matrix Q is the hidden category-item matrix, where Q_{IJ} represents the weight of item I in classification C. The higher the weight is, the more representative this term is of this category:

$$R(u,i) = P_U^T Q_I = \sum_{k=1}^{K} P_{U,k} Q_{k,i}. \tag{1}$$

There is a training set in the algorithm, and for each user U, the training set includes the items that user U prefers and the items that user U is not interested in; the matrix P and matrix Q in the formula are calculated by learning the training set; the specific way is to use the root mean square error RMSE as the evaluation index to minimize the prediction error. The loss function is defined as

$$\begin{aligned} C &= \sum_{(U,I)\in K} \left(R_{UI} - \widehat{R}_{UI}\right)^2 \\ &= \sum_{(U,I)\in K} \left(R_{UI} - \sum_{k=1}^{K} P_{U,K} Q_{K,I}\right)^2 + \lambda \left\|P_U\right\|^2 + \lambda \left\|Q_I\right\|^2, \end{aligned} \tag{2}$$

where $\lambda\|P_U\|^2 + \lambda\|Q_I\|^2$ is used to prevent overfitting of regularization terms and λ needs to be obtained by repeated experiments according to specific application scenarios. The loss function is optimized using stochastic gradient descent algorithm.

By finding the partial derivatives of parameters P and Q, the fastest downward direction can be determined as

$$\frac{ac}{aP_{UK}} = -2\left(R_{UI} - \sum_{k=1}^{K} P_{U,K}Q_{K,I}\right)Q_{K,I} + 2\lambda P_{U,K},$$
$$\frac{ac}{aQ_{KI}} = -2\left(R_{UI} - \sum_{k=1}^{K} P_{U,K}Q_{K,I}\right)Q_{U,K} + 2\lambda Q_{K,I}. \quad (3)$$

Then, according to the stochastic gradient descent algorithm, the iterative calculation formula (4) is obtained:

$$P_{U,K} = P_{U,K} + a\left(\left(R_{UI} - \sum_{k=1}^{K} P_{U,K}Q_{K,I}\right)Q_{K,I} - \lambda P_{U,K}\right),$$
$$P_{K,I} = Q_{K,I} + a\left(\left(R_{UI} - \sum_{k=1}^{K} P_{U,K}Q_{K,I}\right)Q_{U,K} - \lambda Q_{K,I}\right). \quad (4)$$

When the loss function reaches the minimum value, the iteration ends and the user-hidden classification matrix P and hidden classification-item matrix Q are obtained. Rate represents the user's rating on the project, error represents the error, F represents the number of hidden categories, and alpha represents the learning rate. The greater the alpha value, the faster the iterative decline.

2.2. K-Means Clustering Algorithm: Clustering Users. The user-item scoring matrix R can be decomposed by the argot meaning model in Section 2.1 to obtain the user-classification matrix P and classified-item matrix Q. After dimension reduction, user-classification matrix P contains the weight of users in each implicit classification, k-means clustering algorithm is applied to matrix P to classify users into a certain category.

2.2.1. Generate the Item Vector after Dimension Reduction. Users have been classified into several categories by means of the semantic model and clustering algorithm. In order to reduce data sparsity, user category score can be used to replace the user score, so as to shorten the length of item vector and reduce the data sparsity of the item-user matrix.

2.2.2. Algorithm Application Example. The CF-LFMC algorithm uses the argot meaning model and clustering algorithm to classify the users in the item-user scoring matrix, which plays a role in dimension reduction. Therefore, the CF-LFMC algorithm focuses on the group characteristics of users. In the field of e-commerce, the group characteristics of users have certain regularity; most of the items purchased by students are popular and inexpensive, while the consumption level of high-income business people is generally higher. These two methods can classify the users while reducing the matrix dimension, so they are very suitable for application in e-commerce system. Take the following recommended scenario as an example to explain the basic idea of the CF-LFMC algorithm. This is the user record of an e-commerce system with eight users: *A*, *B*, *C*, *D*, *E*, *F*, *G*, and *H* and four items: YONEX badminton racket, Li Ning badminton racket, YONEX badminton, and Li Ning badminton. Among users, *A* and *B* are professional badminton players, *C*, *D*, and *E* are badminton lovers, and *F*, *G*, *H*, *I*, *J*, and *K* are ordinary players who play badminton for entertainment. YONEX is a high-end brand with high price but better quality and hand feel. Li Ning brand is a public brand, cost-effective, and loved by the public. 1 represents the purchase and 0 represents the nonpurchase. The purchase of the above users is expressed by the item-user matrix, as shown in Table 1.

Because the relationship between rackets and rackets and between balls is generally competitive and the similarity is low, so the calculation is not done. The similarity between racket and ball is calculated by cosine similarity formula, S_{YY} represents the similarity between YONEX racket and YONEX ball, and S_{YL} represents the similarity between YONEX racket and Lining ball, and formula (5) is obtained:

$$S_{YY} = 0.655,$$
$$S_{YL} = 0.535. \quad (5)$$

The above results indicate that customers who buy YONEX rackets are highly likely to be recommended to buy YONEX badminton. However, combining with the actual user attributes, we can find that YONEX badminton rackets are very popular among both professional players and amateurs, and YONEX badminton is generally purchased by professional players and Li Ning badminton is very popular among amateurs and ordinary players. According to the above analysis, there is a strong correlation between user group characteristics and purchase preferences; therefore, the user-classification attribute is introduced and the purchase information of users with the same attribute is combined; for the combined project, the length of the 11-user vector is reduced to the length of the 3-user category vector. S'_{YY} represents the similarity between YONEX racket and YONEX ball and S'_{YL} represents the similarity between YONEX racket and Li Ning ball, and formula (6) is obtained:

$$S'_{YY} = 0.651,$$
$$S'_{YL} = 0.843. \quad (6)$$

So, for customers who buy YONEX badminton rackets, the system will recommend Li Ning badminton first. In practice, this makes sense, as Li Ning is more popular with amateurs and casual players, and most of its customers fall into those two categories. Therefore, this method uses the characteristics of user groups, and the recommendation results will be more general, which is also more in line with the actual situation in the e-commerce system. The above examples are only used to describe the customers who buy badminton rackets and badminton; in the actual e-commerce system, users can be classified into thousands of categories according to the purchase habits, consumption level, and user gender. The basic principle is the same.

3. Results and Analysis

3.1. Experimental Evaluation Criteria. There is a basic assumption in recommender systems that users prefer more

Table 1: Item-user matrix.

	A	B	C	D	E	F	G	H	I
YONEX pat	1	1	1	0	1	1	0	1	1
Li Ning pat	0	0	0	1	0	0	1	0	1
YONEX ball	1	1	1	0	0	0	0	0	0
Li Ning ball	0	0	0	1	1	1	1	1	1

accurate recommender systems. Therefore, the recommendation algorithm with higher accuracy is the key to the recommendation system. The prediction accuracy can be measured by offline experiments. Generally speaking, the evaluation indexes of the recommendation system include user satisfaction, prediction accuracy, coverage, diversity, novelty, surprise, trust, real-time, and robustness. Offline evaluation uses the existing data and models' user behavior to evaluate the performance of the recommendation system, especially the accuracy, which is the most important offline evaluation standard of the recommendation system. This index is calculated through the offline dataset of user behavior, which is then divided into training set and test set in a certain proportion; the user's performance in the test set is predicted by calculating the user's information and data in the training set, and the degree of consistency between the user's performance in the test set and the actual situation is calculated as the evaluation accuracy [11].

Define the user set as U, the item set as I, R as the system score set, and S as the score set for scoring optional. At the same time, r_{ui} is represented as the score of user $u \in U$ for a specific item $i \in I$; at the same time, it is assumed that the number of values of r_{ui} cannot be more than one set of users who have graded item i in the set denoted by U_i. Similarly, I_u represents the collection of items rated by user collection U. The set of users rated by both u and r can be expressed as I_{UV}. U_{ij} is used to represent the set of users who have rated both item i and item j.

The optimal term and the optimal N term are the two most important problems in the recommendation system. The best item is the new item $i \in I/I_u$ that user u is most likely to be interested in. When the score value is present, the optimal term can usually be defined as a regression or classification problem with the goal of using the learning function $f: U \times I \longrightarrow s$ to predict user u's score $f(u, i)$ for item i. Then, we can use this function and use the following formula to predict for which item I the user set U has the highest score:

$$i^* = \arg\max f(u_a, j). \tag{7}$$

3.2. Performance Comparison of Three Cosine Similarity Calculation Methods of IBCF Algorithm. There are three similarity calculation methods for the collaborative filtering algorithm based on items, namely, cosine similarity, modified cosine similarity, and Pearson correlation coefficient; the MAE values of these three methods vary with the number of nearest neighbors k, as shown in Figure 1.

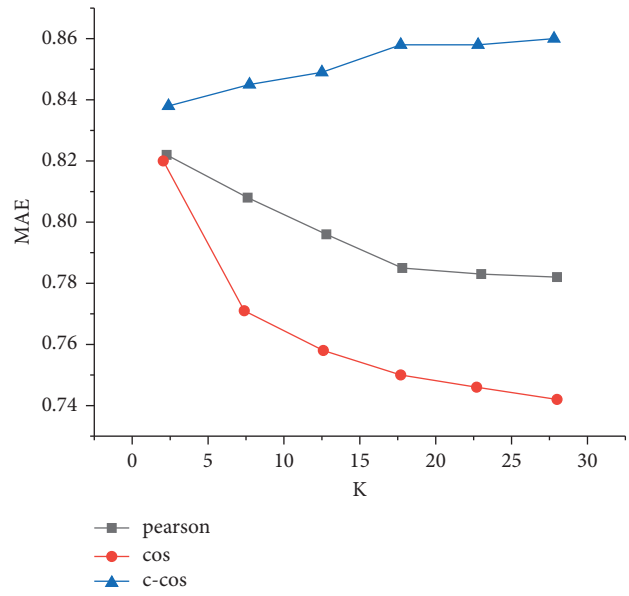

Figure 1: MAE of the three-phase velocity measurement methods varies with the number of nearest neighbors.

Among the three similarity calculation methods, cosine similarity and Pearson correlation coefficient decrease with the increase of the number of nearest neighbors, and the modified cosine similarity increases slightly as the number of nearest neighbors increases. Among the three methods, under the same nearest neighbor condition, the traditional cosine similarity method has the smallest MAE in most cases; therefore, the traditional cosine similarity is used as the similarity calculation method between item vectors.

3.3. Performance Comparison between CF-LFMC Algorithm and IBCF Algorithm. The collaborative filtering algorithm based on the argot meaning model and clustering algorithm is an improvement on the collaborative filtering algorithm based on items; therefore, the two algorithms are compared, and the clustering value $N = 10$ in the CF-LFMC algorithm is taken. Figure 2 shows the MAE curves of the two algorithms changing with the number of nearest neighbors K. MAE values of both algorithms decrease as the number of nearest neighbors k increases. When the number of nearest neighbors is small, MAE values of the two algorithms are close to each other. With the increase of the number of nearest neighbors, CF-LFMC shows better accuracy. The number of recent neighbors continues to increase, which does not significantly improve the accuracy of the algorithm; the calculation cost of the algorithm will increase with the increase of the number of nearest neighbors, so it is more appropriate to choose the number of nearest neighbors between 20 and 30. Within a reasonable clustering range, the CF-LFMC algorithm can better improve the sparsity of matrix data, so it can improve the accuracy of item vector similarity calculation and thus improve the accuracy of the algorithm [12].

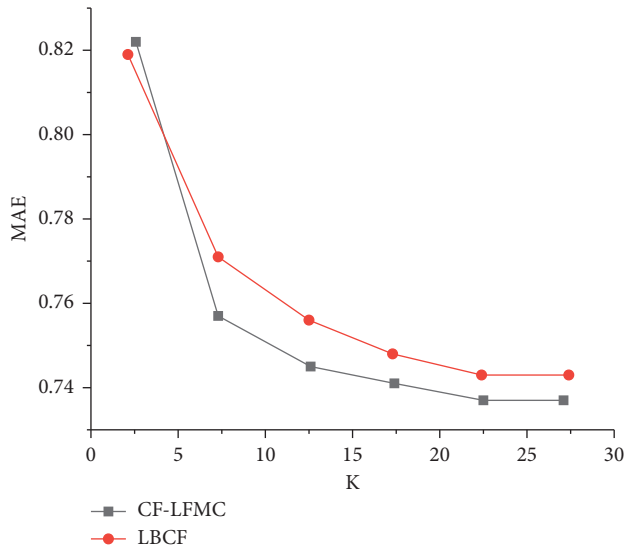

Figure 2: MAE of CF-LFMC and IBCF varies with the number of nearest neighbors.

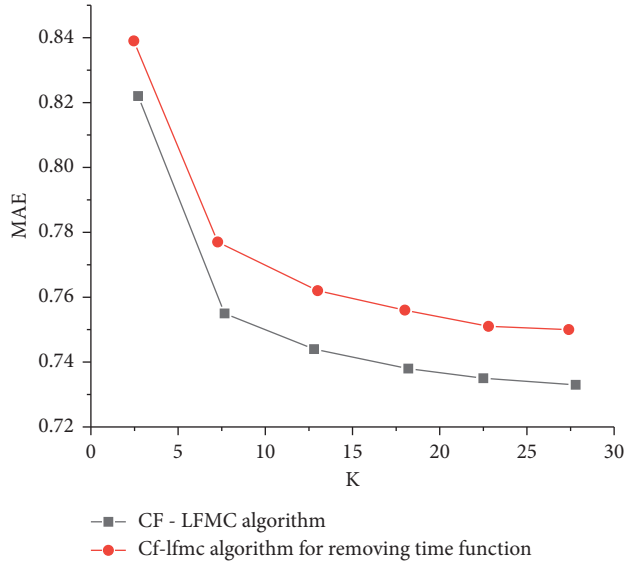

Figure 3: CF-LFMC and cF-LFMC with time function removed vary with the number of nearest neighbors.

3.4. Performance Comparison between cF-LFMC Algorithm and CF-LFMC Algorithm without Time Function. In order to verify the utility of the time function, the CF-LFMC algorithm is removed from the time function and compared with the original algorithm. Figure 3 is the curve of MAE value changing with the k value of the nearest neighbor between the CF-LFMC algorithm and the cF-LFMC algorithm with time function removed (short for time-free CF-LFMC).

After adding the time function, the algorithm increased the weight of users' recent rating behavior and reduced the weight of users' ratings long ago so that the calculation results can better indicate users' interests and hobbies in the recent period of time. The curve shows that the accuracy of cF-LFMC algorithm improved by time function.

4. Conclusions

The CF-LFMC algorithm proposed based on semantic classification firstly analyzes the traditional algorithm, aiming at some problems existing in the traditional algorithm; combined with project-based collaborative filtering algorithm and clustering algorithm, a collaborative filtering algorithm based on argot meaning model and clustering algorithm is designed, the traditional algorithm is improved in terms of data sparsity, cold start, and timeliness analyzed previously; secondly, the performance of three cosine similarity calculation methods of experimental IBCF algorithm is compared, comparing the performance of CF-LFMC algorithm with that of IBCF algorithm and CF-LFMC algorithm with that of CF-LFMC algorithm without time function, The clustering value $N = 10$ in the CF-LFMC algorithm is taken as the experimental result. MAE values of both algorithms decrease with the increase of the nearest neighbor number k. When the number of nearest neighbors is small, MAE values of the two algorithms are close. With the increase of the number of nearest neighbors, the number of nearest neighbors continues to increase, which does not significantly improve the accuracy of the algorithm. The calculation cost of the algorithm will increase with the increase of the number of nearest neighbors, so the number of nearest neighbors between 20 and 30 is more suitable. CF-LFMC shows better accuracy; the accuracy of the CF-LFMC algorithm improved by the time function, and the accuracy of the algorithm is better than that of the traditional algorithm. Although the CF-LFMC algorithm and e-commerce personalized recommendation system designed in this study have achieved the expected results, there are still many shortcomings, mainly manifested in the following two points. In the use of clustering algorithm to cluster the user, the user will be divided into the category of fixed, doing so can reduce the sparse data, but for some users within the category boundaries, they will be divided into categories which not necessarily can represent their characteristics, resulting in decrease of some number of clustering algorithm accuracy or even worse than traditional algorithm. At present, the recommendation algorithm has a high accuracy in the case of large user rating data. In the case of relatively small number of users and items, how to improve the accuracy of the algorithm still needs further research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] T. Li, X. Yang, T. Gao, Y. Liu, and Y. Wang, "Research and implementation of mine risk area semantic retrieval system based on ontology," *International Journal of Advanced*

Pervasive and Ubiquitous Computing, vol. 8, no. 3, pp. 37–86, 2016.

[2] P. Zhao, H. Li, Y. N. Li, W. W. Lu, and J. X. Wu, "Research and implementation of three-dimensional power lines selection system based on aerial images," *Applied Mechanics and Materials*, vol. 738, pp. 213–216, 2015.

[3] Y. Kawakami, T. Hattori, H. Matsushita, Y. Imai, H. Kawano, and R. P. C. J. Rajapakse, "Automated color image arrangement method based on histogram matching," *International Journal of Affective Engineering*, vol. 14, no. 2, pp. 85–93, 2015.

[4] R. Eagleson, L. Altamirano-Diaz, A. Mcinnis et al., "Implementation of clinical research trials using web-based and mobile devices: challenges and solutions," *BMC Medical Research Methodology*, vol. 17, no. 1, p. 43, 2017.

[5] C. Gossa, M. Fisher, and E. J. Milner-Gulland, "The research-implementation gap: how practitioners and researchers from developing countries perceive the role of peer-reviewed literature in conservation science," *Oryx*, vol. 49, no. 1, pp. 80–87, 2015.

[6] A. Fresa, B. Justrell, and C. Prandoni, "Digital curation and quality standards for memory institutions: preforma research project," *Archival Science*, vol. 15, no. 2, pp. 191–216, 2015.

[7] R. Ayachi, I. Boukhris, S. Mellouli, N. B. Amor, and Z. Elouedi, "Proactive and reactive e-government services recommendation," *Universal Access in the Information Society*, vol. 15, no. 4, pp. 1–17, 2015.

[8] Y. Liu and Yun, "Design and implementation on digital media resource management system based on soa," *Applied Mechanics and Materials*, vol. 713, pp. 2233–2236, 2015.

[9] V. Turkar and S. Gawade, "Analysis of digital media compatibility with farmers in Maharashtra and recommendation of service provider design framework'e-krishimitra," *International Journal of Applied Agricultural Research*, vol. 12, no. 1, pp. 77–86, 2017.

[10] Q. Gao, W. Liu, D. Li, Y. Wang, and T. Xue, "Research and implementation of the roll position automatic adjustment system based on roller parameters prediction," *Journal of Advanced Manufacturing Systems*, vol. 18, no. 2, pp. 273–292, 2019.

[11] L. A. Liikkanen and P. Åman, "Shuffling services: current trends in interacting with digital music," *Interacting with Computers*, vol. 28, no. 3, pp. 352–371, 2015.

[12] S. J. Rasmussen and C. H. Houpis, "Development, implementation and flight test of a mimo digital flight control system for an unmanned research vehicle designed using quantitative feedback theory," *International Journal of Robust and Nonlinear Control*, vol. 7, no. 6, pp. 629–642, 2015.