

Q&A System Development based on Agricultural Knowledge Map and Multi-Level Association Model

Qingfeng Wei ^{1,2}, Chenzhong Cao ^{1,2}, Furong Wang and Changshou Luo ^{1,2,*}

¹Institute of Agricultural Information and Economics. Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

²Beijing Research Center of Engineering Technology on Rural Distance Information Service, Beijing, China

*Corresponding author e-mail: Luochangshou@163.com

Keywords: Q&A, Agricultural knowledge map, Multi-level association model.

Abstract. To the problems of the current agricultural Q&A system: poor user experience, low precision, and not fully utilized resources, the multi-level association question answering system is developed based on agricultural knowledge map, and Jaccard index and Vector space mode. The application proves that the user experience is good, and the retrieval precision efficiency is improved effectively.

1. Introduction

With the increasing awareness of farmers' information, access to agricultural technology solutions to production problem by network has become an important way for farmers. Baidu question answering system and the similar systems provides users with a variety of ways to consult. But there is a great risk that the information is applied to agricultural production directly. Therefore, the agricultural question answering system (Q&A in short) is the first choice for farmers.

There are numerous researches on the agricultural Q&A system [1-3]. However, those systems are lack of deeply analysis of the law of farmers, and with the poor user experience, low precision, and not fully utilized resources. To those problems, constructing agricultural knowledge map and multi-level association retrieval model, developing agricultural technology question answering system, it will be a simple and effective way to help agricultural users get professional and reliable technical solutions.

2. Agricultural knowledge map constructing

Through the analysis of a large number of existing consulting questions, it mainly related to those aspects: crop varieties, planting techniques, disease prevention and control, livestock and poultry breeding, and the purchase of agricultural materials. Construct agricultural knowledge map with four elements of 'Entity - attribute class - instance - relationship':

$KM = \langle E, A, N, R \rangle$

E: Entity, the object of agricultural user consultation;

A: Attribute class, the content of entity;

N: Instance, the content of attribute class;

R: Relationship, the relationship between the instance and Q&A.

3. Multi-level association guidance based on Knowledge Map

According to the input word, query the attribute class of the object entity in the knowledge map. Gets the first result. At the same time, query Q&A database, get the question title which contains the word as the second result. Merger the first result with the second result, to form the first level association result set. Using front page technology of JQuery+Ajax to display it. To determine whether

the result of the first level association contains the object attributes class. If there are some results, increase the tag, as a result of the two association, display in cascading menu.

4. Retrieval model construction based on Jaccard index and Vector Space Mode

When it does not find the answer after the first level association and the second level association, sentence similarity retrieval is used to help find the most relevant answer. The retrieval model is based on the combination of the Jaccard index and improved cosine similarity.

The question vector can be represented as follows:

$$Q = \{wq1, wq2, wq3 \dots, wqn\}$$

wq is the feature word of the question

Extending the feature words with synonym:

$$wq = \{wq | wq1 \vee s1 \vee s2 \vee \dots \vee sm\}$$

The question vector in FAQ database can be expressed as:

$$A = \{wa1, wa2, wa3 \dots, wam\}$$

Calculate the of Jaccard index between A and Q as follow. Set the threshold value, take the result set that the Jaccard index is greater than the threshold value for further calculation.

$$J(Q, A) = \frac{Q \cap A}{Q \cup A} \tag{1}$$

The calculation method of word weight is as follows:

$$w = TF * IDF = (fi/c) * \log(N/ni) \quad (1 \leq i \leq n) \tag{2}$$

The ‘fi’ is the number of times the character word w appears in a single text

The ‘fi’ is the word frequency of “W” in a single text. The ‘c’ is the total number of all words in a document. The ‘N’ is the total number of all documents. The ‘ni’ is the total number of documents containing the word ‘w’ in the document collection.

Similarity calculation is as follows:

$$mSim(q, a) = \frac{\langle q, a \rangle}{|a|} = \frac{\sum_{i=1}^t W_i^q W_i^a}{\sqrt{\sum_{i=1}^t W_i^q}} \tag{3}$$

5. System development

The system is developed based on B/S architecture, which contains distributed network platform, which contains three layers: data layer, business logic layer, presentation layer. The user requests the business logic layer at the presentation layer, IIS receives the request and sends it to the ASP.NET engine for processing. ASP.NET call multi-level association retrieval model, and also calls the data in the database through the ADO. NET. Get the results finally and returned to the client.

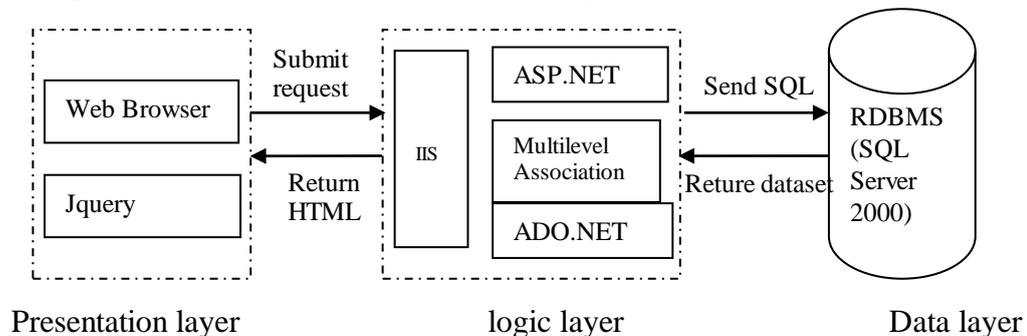


Figure 1. B/S architecture

Model retrieval process:

Step1 According to the input word, give multi-level association tips.

Step2 If the user is not satisfied with the tip, access to the sentence and obtain the feature word by the word segmentation and remove the stop word.

Step3 Using the formula (2) to calculate the weight of feature words and construct the query vector.

Step4 Obtains the initial result sets which contains the query feature words through the calculation of the Jaccard index (1).

Step5 Using the formula (3) to calculate the cosine similarity between the query vector and the initial result set.

Step6 According to the angle cosine similarity value to reverse sort, and then output.

6. Conclusion

Analysed the law of farmers' questions, build an agricultural knowledge map, and applied with the multilevel association retrieval model, so that users can find answers easily when the mouse moves on the web. And it also help the users get a more comprehensive solution to the problem.

In the multilevel association retrieval model, the Jaccard index is used to reduce the calculation range, the improved vector space model is used to enhance the accuracy. The model in this study improves the efficiency and accuracy of the Q & A system.

The application system based on this method can ensure the accuracy of the information, and provide more professional services than the open Q&A system. It reduce the risk of production application.

Acknowledgments

The research work was supported by the young scientist fund of BAAFS: Research on agricultural scientific technical online counseling system based on the knowledge map (NO.QNJ201534), the Beijing social science foundation(16JYC026): Innovation research of rural distance education service supply mode in Beijing base on "Internet +", the innovation fund of BAAFS: Agricultural water-saving technology training and counseling service demonstration project (NO.KJCX20151409), the fund for agriculture of Beijing municipal commission of rural affai: The application of APP and "zhixuntong" products in the management and service of general agricultural technicians.

References

- [1] Fang W.,Guifa T., Yuxin Z.. Research on Agricultural FAQ System Based on
- [2] Juan Z., Nengfu X.. Research on agricultural knowledge question answering system based on Web. *Agricultural network information*, 2007,7:21-24.
- [3] Yanhua X. Qinlan W.. Application of latent semantic index in Jilin agricultural knowledge question answering system. *Hubei Agricultural Sciences*, 2011,50(11):2740—2742.