

Research and Implementation of Agricultural Science and Technology Consulting System Based on Ajax and Improved VSM

Sufen SUN, Junfeng ZHANG, Changshou LUO*, Qingfeng WEI

Institute of Information on Science and Technology of Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing, 10009, China
sunsf@agri.ac.cn, zhangjf@agri.ac.cn, luochangshou@163.com(*Corresponding author),
weiqf@agri.ac.cn

Abstract. In order to overcome the disadvantages of inconvenience and low accuracy of keywords spitting search, and artificial answer delay for the online messages in the current agricultural consulting system, this paper proposes a solution which complement the advantages of automatic question answer with expert real-time answer. Firstly, vector space model was improved by semantic extension and feature item weight modification. And then, based on the agricultural technical information database and expert team, an agricultural science and technology consulting system was developed by the Ajax. To the simple natural language questions, the system answers it automatically through improved vector space model in automatic question and answer. To the complicated question, the agricultural experts gave answer immediately after chatting with inquirer on line in expert real-time answer. The experiment show that the system has features of accurate and rapid reply, professional answer and high satisfactions. It can better meets the need of agricultural production techniques.

Keywords: automatic question and answer; expert real-time answer; vector space model; Ajax

1 Introduction

Agricultural sci-tech information service is important to spread agricultural technology, transform the achievements into actual productivity, and increase the income of farmers. The general agricultural consulting service system solves problems by keywords searching. It has shortcomings of needing farmer split keywords by themselves. It lacks a full understanding of the question, and always misses the correct answers for word machine-matching. It is also very difficult for farmer to identify the right answers from the large mount of results which were feedback by keywords searching. Besides, users also find help from artificial answer by leaving message. It is easy to use, and also can give a accurate answer for the complex question. But it requires the user to log on again to get the reply. Answer delay affects service effectiveness greatly.

To solve those problems, using improved vector space model (VSM) and Ajax, an agricultural consulting system was developed based on “automatic question answer complement with expert real-time answer”. The automatic question answer (automatic Q&A) based on improved VSM can understand natural language and provides enquirers with the most suitable answer. In expert real-time answer, the farmer can describe questions clearly by chatting with expert. then, the expert gives an accurate answer immediately after getting more background information from web conversation. It integrated the advantages of both to meet the needs of agricultural production technology instantly and effectively.

2 Materials and methods

2.1 Technical scheme

There are two major modules of automatic Q&A and expert real-time answer in this system. To the simple questions, it uses the former module answers automatically. To the complex question, it uses the latter module answers directly by web chatting. When the consultant is unsatisfied with the feedback results from automatic Q&A, they can seek help from expert real-time module. The system adopts a parallel strategy of two methods. Flow chart is as follows:

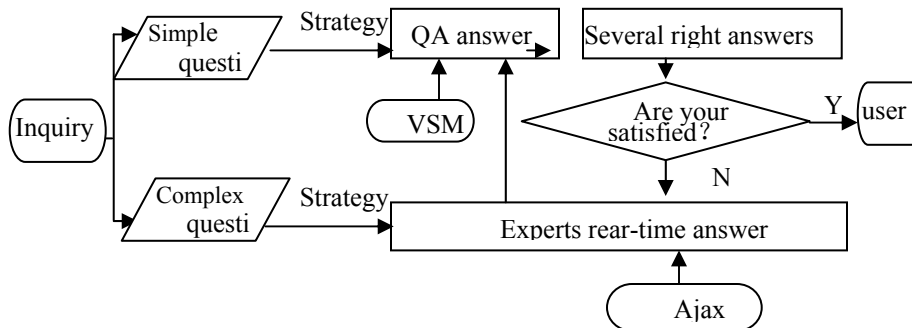


Fig. 1. System Scheme

2.2 Computational Model and Key Technologies

2.2.1 VSM improvement

As a classic model in textual retrieval, VSM is mostly widely used in auto Q&A [1-2]. The similarity computation is the key to achieve automatic answer. However, VSM assumes that key words in the document are independent of each other with orthogonally hypothesis. Its assumption is difficult to satisfy in agricultural application. Big error was existed between computational results and the actual results. So, in this part, we consider the characteristics of agricultural application, and then, improve it to get better results.

Conventional VSM assumes that the meaning of a sentence and a document can be represented by a set of words used in them. The words sequence is represented by vector. Suppose the document is d , the query sentences are q , their vector and word weight computational formula are as follows:

$$d = (W_1^d, W_2^d, \dots, W_i^d, \dots, W_n^d) \quad q = (W_1^q, W_2^q, \dots, W_i^q, \dots, W_n^q) \quad (1)$$
$$W_i = TF * IDF = (t/c) * \log (N/n_i)$$

Here, i ($1 \leq i \leq n$) is the number of feature items which are used to represent the meanings of the sentences. In general, the key words in text are used to represent feature items. TF is the frequency of feature item t repeating in question. N is the total amount of documents in answer set, n_i is the total amount of document which contain feature item t .

The similarity between q and d can be measured by computing cosine of their corresponding vector with the following normalization function:

$$\begin{aligned} sim(d_i, q_j) &= \frac{d_i \times q_j}{|d_i| \times |q_j|} \\ &= \frac{\sum_{i=1}^t W_{ij} * W_{iq}}{\sqrt{\sum_{i=1}^t W_{ij}^2} * \sqrt{\sum_{i=1}^t W_{iq}^2}} \end{aligned} \quad (2)$$

Some improved methods are as follows:

Semantic extension. Word matching is an important process in similarity computation. Sometimes, there are words of different form with the same meaning in agriculture. For example in Chinese, the “碳酸氢铵” is called “碳铵”, and also is written as “NH₄HCO₃”. If somebody wants to search the reasonable application methods of “碳铵”, the document about NH₄NO₃ and “碳酸氢铵” will not be retrieved for the different form. So, in this part, extend the feature item in the question to full understand user intent. The processing is as given below:

Question: 怎样施用碳铵?

Feature item extraction: {施用 ; 碳铵}

Semantic extension: {使用/施用; 碳铵/碳酸氢铵/ NH₄HCO₃}

Weight modification in document. In conventional VSM, feature word weight was calculated by formula (1). Which often repeats in one document and rarely appears in answer set will get high weight [3] Professional words make big contribution to the subject in agriculture. However, according to this principle, some professional words which appear vary often both in single document and in answer sets will have low weight. To solve the problem, this study modifies the weight as follow:

$$MW=a*TFi*IDFi \quad (3)$$

Here, a is the weighting coefficient, it indicates the contribution degree to the subject. Test shows that, when professional word a=1.0 and general word a=0.7, the identification ability of correct answer is obvious.

Weight modification in question. Conventional VSM assumes that wherever the feature word is, they have the same contribution to the subject. Actually, the words in the heading are more relevant to the subject than in the body. So, a modification is given according to the distribution of query words as below:

$$MSim(d, q) = \begin{cases} \lambda_1 \sin(d, q) T_i \in \text{heading} \\ \lambda_2 \sin(d, q) T_i \in (\text{heading and body}) \\ \lambda_3 \sin(d, q) T_i \in \text{body} \end{cases} \quad (4)$$

λ is weighting coefficient of words distribution. t_i is the query word which is extracted from query sentence. Test shows that, when $\lambda_1=1.0$, $\lambda_2=0.6$, $\lambda_3=0.2$, it has the best result.

Optimization of computational efficiency. Similarity computation is between high dimension feature vectors which are in sparse matrix. The amount of calculation is very large. To solve this problem, in the session of answer document pretreatment, it uses full-text index, and calculates the weight of feature words, generates feature vector in advance, to reduce the workload of foreground process. In the session of similarity computation, it selects answer documents which contain query feature words as candidate set to reduce the useless computation.

2.2.2 Ajax application

Ajax is composed of multiple technologies such as XHTML, CSS, DOM, XSTL and JavaScript etc. It is a development patter which used to create interactive web application

[4-5]. The main feature of Ajax is that the transmission, reading and writing data on the server don't need to refresh the page. It can significantly shorten the waiting time.

According to the statistics from literature [6], if the visitor doesn't get satisfied information, 90% of them will not choose E-mail to contact service staff, 70% of them will not repeat visit. So, it is extremely important of providing users with artificial answers immediately to enhance the service effect. However, in traditional web applications development, each time the user submits the instruction, web page must be refreshed. It leads to data loss and a long time waiting. Ajax is very suitable for the application of frequent interaction, reading and writing.

2.3 System Design and Implementation

2.3.1 System architecture design

It is a three-tier architecture which contains presentation layer, business logic layer and the data layer. Presentation layer provides users with interfaces which can enter into the tow main module. Using Ajax engine and Web/XML server, business logic layer extracts the technique information from database to the client .Data tier is mainly used to manage the agricultural technology information and expert information. Artificial answering team is composed of experienced experts who major in vegetable production, fruit processing, aquaculture, animal husbandry and veterinary.

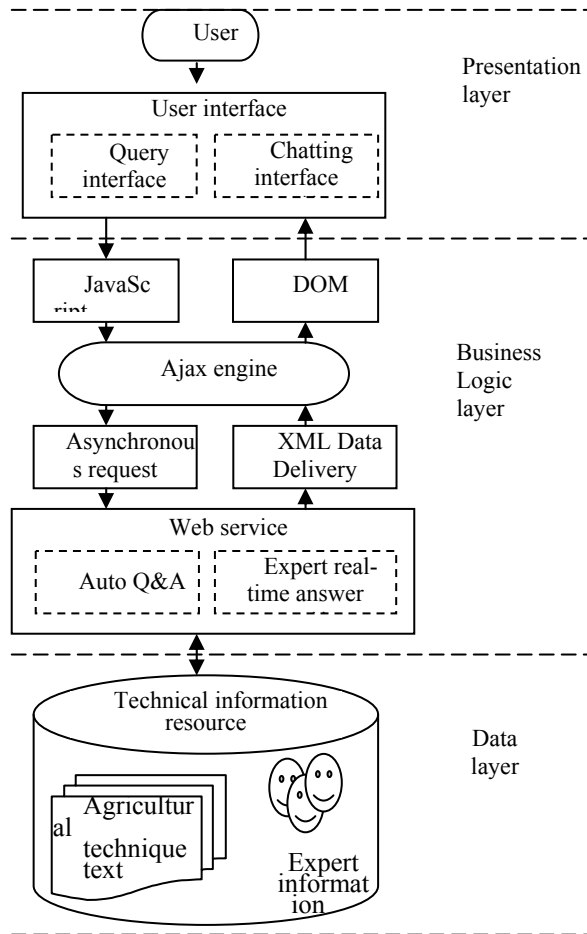


Fig. 2. System architecture

2.3.2. Automatic question and answer

The realization process is divided into the following five steps:

Step1: Pretreatment of technique text. It intent to extraction feature vectors by words segmentation, stop words removing. And then, uses formula (3) to calculate the weight of feature item. All of this prepare for calculating with the query sentence.

Step2: Query sentence understanding. It understands user natural language question by words segmentation, stop words removing and semantic extension. Duo to the same operation with the technical text pretreatment, the program use a public class file which was named as “CommonClass1.cs” to enhance the code reusability and improve the development efficiency.

Step3: Selection of candidate answer set. By means of semantic extension, a query expression of query feature vector is constituted. Then, taking Boolean model as selection method, choose the document set which contain query word as candidate answer set.

For example in Chinese:

The query feature vector after semantic extension is:

{番茄/西红柿, 栽培/种植}

In feature word reverse index table, the IDs of documents which contain the feature item are:

番茄: {5,8,9,20}

西红柿: {25,27,30,102}

栽培: {8,30,45,130,150}

种植: {4,19,20}

According to the query expression-(番茄 \cup 西红柿) \cap (栽培 \cup 种植), the candidate answer set which can meet the requirement is {8, 20, 30}.

Step4: Similarity computation. Within candidates answer set, similarity is calculated between query feature vector and candidate answer feature vector using formula (2). Then, according to the distribution of query words in the text, modify the weight using formula (4).

Step5: Answer display. Order the value of similarity in decent, and gain the result set that the similarity value is greater than given threshold β . The candidate answer with the maximum similarity is showed directly. Other answers are listed to display. If there is no any answer, pop-up a message to inform the user to log on the expert real-time answer module to seek help.

The main code to calculate feature item weight is as follow:

```
public string WordTFIDF(double N,string[] a)
{ ... .. // Variable definition
  ... .. // Database connection
  StringBuilder vector = new StringBuilder();
  word=a.Length;
  count=0;
  Hashtable Dealword = new Hashtable();
  for (int i = 0; i < a.Length; i++)
  { if (Dealword.Contains(a[i]))
    continue;
    else
    {
```

```

        SqlCommand com = new SqlCommand("select
count(*) from aiarticle where contains(content,'" + a[i] +
"')", con);
        SqlDataReader dr = com.ExecuteReader();
        dr.Read();
        n = dr.GetInt32(0);
        dr.Close();
        for (int j = 0; j < a.Length; j++)
            if (a[i] == a[j] && (i <= j))
                count += 1;
        if (n == 0)
            n = 1;
        tfidf = (count / word) * System.Math.Log(N / n);/
        tfidf = System.Math.Round(tfidf, 5);
        count = 0;
        Dealword.Add(a[i], tfidf);
        vector.Append(a[i]);
        vector.Append(': ');
        vector.Append(tfidf);
        vector.Append(' ');
    }
}
foreach (DictionaryEntry de in Dealword)
{
    temp = Convert.ToDouble(de.Value);
    denominator += System.Math.Pow(temp, 2);
}
denominator = System.Math.Sqrt(denominator);
vector.Append('+');
vector.Append(System.Math.Round(denominator, 5));
return vector.ToString();
con.Close();
}

```

2.3.3. Expert real-time answer

The database mainly involves consultant table, user information table, visitor state table and chatting table etc. The field “answer” in visitor state table is used to show if the chatting message was displayed. The program can avoid repeat display by judging the state of the field. Field “dialogued” in chatting table is used to distinguish different dialogue. It also can help the program to find a entire conversation published by one person immediately. Field “isConsultantReceiver” indicates that if the expert receives this message. According to this field, the client chatting box will only show the latest message which isn’t accepted.

Embed Ajax engine in logic programs to monitor database changes regularly. When the farmer consults with expert, it triggers the event handle function which can save the

two sides talk to data table. 1 second interval, data monitoring program scans the table. If there are new messages, show it to the chatting box of client. Ajax hides the whole process of the update operation. It achieves “face to face” communication between farmers and experts with no awareness.

The key code to show the consultation content is as follow:

```
private void ShowMessage()
{
    ... .. // Connect the database

    SqlDataAdapter da = new SqlDataAdapter("select * from
message where (sender='" + Session["user"].ToString() + "'
and receiver ='" + Session["expertname"].ToString() + "') OR
( sender='" + Session["expertname"].ToString() + "' and
receiver ='" + Session["user"].ToString() + "') ORDER BY
CreateDate ASC", con);

    DataSet ds = new DataSet();
    da.Fill(ds, "ChatMessage");
    con.Close();
    if (ds != null)
    {
        DataView dv = ds.Tables[0].DefaultView;
        dv.RowFilter= string.Format("CreateDate>='{0}'",
DateTime.Parse(ViewState["StartDate"].ToString()));
        StringBuilder sbMessage = new StringBuilder();
        foreach (DataRowView row in dv)
        {
            string singleMessage = row["sender"].ToString() + "说: " +
row["createDate"] + "\n";
            singleMessage+=" " + row["body"].ToString() + "\n";
            sbMessage.Append(singleMessage);
        }
        this.tbChatMessage.Text = sbMessage.ToString();
    }
}
```

2.4 Test materials and experiment design

The data source of the system is derived from “Beijing Agricultural Digital Resource Center”. We take it as test materials.

According to users concern, take precision, reaction time and satisfaction as evaluation criterion to test the system. Precision is the ratio of right answer to all feedback answer. Reaction time is from submitting question to obtaining answer. Satisfaction is the ratio of praise to all visitors. Participants involved in testing are composed of 10 internal staff and 10 external farmers. Each one asks 5 questions in auto Q&A and 5 questions in expert real-time answer.

3、Results

The average value of test results in table 1:

Table 1. Comparison of different solutions

solutions	Precision		Reaction time	Satisfaction
	Inner staff	External farmer		
Auto Q&A based conventional VSM	70.2%	55.4%	0.42376	/
Auto Q&A based improve VSM	82.2%	79.5%	0.01087	65%
Auto Q&A based improve VSM and expert real-time answer	89.6%	86.2%	/	85%

Experiment shows that the precision increases, and the reaction time is shorter than auto Q&A based on conventional VSM. It indicates that the improvement for agricultural application and optimization of system efficiency are effective. The satisfaction is also increased with the solution improvement. It indicates that, complementary with expert real-time answer, the service patter is more welcomed by the agricultural-related users.

Meanwhile, to the precision, the test results of inner staff are better than the external farmer of practical application. The reason lies that the internal staff are more familiar with the technical resources. Their query expression is similar to the existing answering text, so the retrieval accuracy is higher. External farmers ask question entirely in accordance with their own language. Some dialect is not included in our vocabulary in the short term. It lead to the system can't understand users question fully. In addition, as the

matter of limited time, the coverage of agricultural technical resource is not enough. All of these affect the search results to some extent. Therefore, it also suggests that with further improvement of vocabulary, expansion the scale of the technical information resource, the system effectiveness will be more apparent.

4、 Discussion and conclusion

In this study, the solution of auto Q&A combined with expert real-time answer was applied into agricultural sci-tech consulting system. It has the advantages of simplicity to use, immediate and accurate response and professional answer. It provides agricultural related user with a new tool to solve problems. This study mainly has the following characteristics:

(1)Users can use their own expression to ask question in system. They don't need to split the sentence. It is easy and convenient for farmer acceptance and application. The strategy of answers sorting and recommendation helps users locate the required information quickly in large amount of information resources.

(2)The improved VSM abstains a better precision by consider the application characteristics in agriculture fully. At the same time, using the optimal design of candidate set selection, the weight of feature words presorting ,it reduce the workload of calculation of foreground and also ensured the efficiency of the system.

(3)Adopting expert real-time answer, it not only shared the expert intelligence resources, but also response to consumer enquiry quickly and accurately. It better improve the user satisfaction with the consultation.

(4) Using Ajax brushless technology to hide the chatting process in the background, it makes up the shortcoming of low execute efficiency in B/S structure caused by frequent submission, and also brings the farmer and experts with good conversation experience.

Agricultural consultant service for root area farmer is required easy operation and accurate real-time response. In auto Q&A, the study will make a more detail exploration in the integration of semantic, syntactic similarity calculation to further improve the accuracy of the query. In expert real-time answers, for the code of Ajax script is open, improving system security and stability also require further research. Full use the advantages of the two modules complement each other to provide a better service to agricultural production.

Acknowledgements

This work was supported by the National Science and Technology Support Program (2006BAJ07B08), Beijing Natural Science Foundation (9093019), and Youth Foundation of Beijing Academy of Agriculture and Forestry Sciences (2007030610).We particularly

wish to acknowledge the support of Y.X Liu, L.W. Zhang , J. Liu , H. Meng and X. Z. Chen.

References

1. Y.H. Tao, M.S. San: Natural Language Retrieval for Latent Semantic Indexing. *Modern Library and Information Technology*. 40-41(2001).
2. L. Zhang, Z.X. Chen and H. Huang: Design and Implementation of FAQ Automatic Return System Based on Similarity Computation. *Wuhan University of Natural Science*. 138-142(2006).
3. X. Bai, X.F. Lv, and J.G. Sun: Network Search Based on Weighted Vector Space Model. *Application Research of Computers*. 51-53(2007).
4. L.D. Palson: Building Rich Web Applications with Ajax. *Computer*, 14-17(2005).
5. H.B. Wen, F. Liu: Design and Implementation of Ajax Based Web Vehicle Monitoring and Control Systems. *J. Huazhong Univ. of Sci. & Tech. (Nature Science Edition)* 77-79(2007).
6. D.G. Huang, X.Y. Ji: Real-time Reference System Based on Online Customer Service. *Journal of Library and Information Sciences in A culture*. 23-25(2008).

