Agricultural Information Service System Based on Embedded GIS

Yuejuan Jing*
School of Computer Science, Xi’an Aeronautical University, Xi’an 710077, Shaanxi, China
*Corresponding author (E-mail: jingyuejuan-1981@163.com)

Xiaoli Zhang
School of Computer Science, Xi’an Aeronautical University, Xi’an 710077, Shaanxi, China

Abstract
Agricultural information services can predict the possible problems that may still exist in agricultural products as early as possible according to the measurement information extracted from agricultural products and the problems that have been found. Embedded GIS technology can comprehensively and automatically find the agricultural products based on the problems, which has become the main method of agricultural information service. In order to improve the efficiency and accuracy of the prediction, the agricultural information service system of embedded GIS is analyzed. It is found that the system has different advantages in different evaluation indexes. With these advantages and embedded GIS method, the author put forward the prediction results of different prediction algorithms as the measurement of agricultural products. Finally, the experiment with the same pair of Eclipse data sets shows the effectiveness of the agricultural information service system.

Key words: Agricultural Information, Embedded GIS, Service System, Eclipse Forecast Data Set

1. Introduction
With the continuous development of Internet technology, agricultural information service has also entered the life of residents [1, 2]. Meanwhile, the quality of agricultural products has also become the focus of many people’s attention. The problem of agricultural products itself will restrict the sale of agricultural products, but the problem of agricultural products exists objectively in the process of production. If the problems in production are not dealt with in time, it is possible that the agricultural products cannot be sold effectively, or the serious consequences cannot be made up. Agricultural information service system can query the quality of agricultural products, and feedback errors in agricultural products. But system design is period of time that waste most resources time. Based on this, the concept of agricultural information service system emerges as the times require. The agricultural information service technology in agricultural information service system can be divided into two kinds: one is static prediction technology, and the other is dynamic prediction technology. The basic basis of static prediction is the measurement data related to the quality of agricultural products. The design and implementation of static prediction are the quantity and distribution of the problem. Dynamic prediction is time-based, and it makes a distributed prediction of the changes in the quality of agricultural products with time.

In this paper, an agricultural information service system based on embedded GIS is proposed. In this system, the embedded GIS are processed, and the embedded GIS method is adopted in the whole process, so that the efficiency of service has been improved effectively. Finally, the agricultural information service system designed in this paper is verified, and the standard data set of Eclipse is analyzed.

2. Basic Knowledge of Agricultural Information Service
In this paper, the related concepts of embedded GIS and agricultural information service will be discussed. Measurement metadata is involved in the agricultural information service. In the process of prediction, the measurement metadata needs to be extracted from the agricultural product module. Then, based on the related methods in embedded GIS, many methods are used for regression, classification, clustering and so on. Through these methods, we find out the relationship between the measurement elements and their problems, and establish the basic system based on the correlation. The distribution of problems in agricultural products or the component of the problem is predicted by the basic system, as shown in Figure 1. Based on the weighting of different embedded GIS, this paper designs and implements the classification of weighting and agricultural information services in different degrees.
Agricultural information service system has essential purpose. Its essential purpose is to predict some problems in agricultural products, so that agricultural producers can focus their attention on problematic modules. Taking the agricultural information service system as the research object, we have gained considerable achievements in design and implementation, but it is still prohibitive to accurately predict all agricultural products. In agricultural information services, if error prediction is produced [3-5]. For example, the module that has no problem is predicted to be a problem module. This period is bound to generate strong resource waste. For example, a module that has a problem is predicted to be problem-free module, and there is a risk that cannot be repaired. The price of this prediction error is quite different, and how to find a balance between the two has always been the focus of the researchers. If the agricultural information service system is based on a weighted problem, it is more suitable for testing the unknown resources. The agricultural information service system based on the weighted problem usually weighs the agricultural product modules according to the problem data. If the number of problems in the agricultural product module can be tested first and there is a surplus of resources, we will continue to test the modules with fewer problems. The advantage of this system is the weighted result, and the problem is the lack of precision.

Embedded GIS is often used in classification modeling. In the process of modeling, embedded GIS needs to be constructed, and the main function of embedded GIS is to find the classification rules contained in the data. If we use the best weighting of multiple independent variables to predict, we call it linear regression. Linear regression is more advantageous than using an independent variable and it is also more applicable to the actual situation. Linear regression is the most basic of all the regression algorithms used for weighting problems. In the application of linear regression, a linear equation is established. Through linear equation, we can see the relationship between the number of problems and the measurement elements, and calculate the module of the problem contained in the unknown module. Embedded GIS promotes the generation of embedded GIS classifier. The embedded GIS classifier is a statistical inference classifier based on the assumption of feature independence. This classifier has an uncertain causality system and has a strong ability to deal with the problem of uncertainty. Bionics promotes the development of artificial neural network. Artificial neural network imitates the behavior characteristics of animal neural network, and then uses distributed information processing [6-8]. In general, an artificial neural network is an adaptive system that can change the internal structure based on external information. In addition, agricultural information service methods include clustering analysis, support vector machine and so on. All these methods need to be built and verified by using historical data.

3. Agricultural Information Service System Based On Embedded GIS

3.1. The Proposal of the Problem

As for the effect of agricultural information service system, embedded GIS are not very ideal. Based on this, two kinds of prediction for Eclipse data set are carried out in this paper: one is classified prediction, and the other is weighted prediction. It is found that the algorithm is different if the evaluation index is different. Through the discussion of this article, we can see that if the embedded GIS are combined with the neural network agricultural information service system, the effect is very good. But it also has some shortcomings, for example, the complexity is relatively large, and the time spent is more, especially the larger data set and more measuring elements. Based on this, we can boldly imagine that the integration of embedded GIS is a combination of the advantages of various algorithms.
3.2. Agricultural Information Service System Based on Embedded GIS

For data of different features, it is integrated into a set of data in a certain way. Of course, sometimes there will be data missing and data errors in these data. If these data are introduced, it will lead to the inaccuracy of the agricultural information service system. So the data need to be cleaned in a professional way. First, it is discretized, and then the state of the problem is divided. If the number of problems is more than 0, it is divided into problems and is represented by number 1; if the number of problems is equal to 0, it is divided into no problem and is represented by number 0. After these processes, the data set can be divided into two parts: one part can be used on the embedded GIS (training set), the other part can be used for testing and verification (test set).

Embedded GIS is an effective technique to improve the performance of classifier. In this paper, Stacking technology is used for reference. In the Stacking framework, there are 2 layers of classifier, level-0 is a base classifier, and level-1 is a classifier. In weighted prediction, the first step is to predict the data set. The second step is to input the output from the base classifier as input data from the Meta classifier. In the third step, the new data set is used as the training data set of the new reasoning device, and the algorithm is used to solve it. A relatively complex network structure algorithm is used for second layers of prediction. In this paper, we use embedded GIS technology in the process of dimensionality reduction. We use this technology to apply attribute subset to reasoning, and provide the performance of the system under the premise of minimizing the loss of information, so as to reduce the cost. In attribute selection, you can use both information gain and information gain ratio.

Definition 1: the information gain rate of the attribute A can be expressed by \( \text{IGR}(A) = \frac{\text{IG}(A)}{I(A)} \), in which \( \text{IG}(A) \) represents the information gain of the property A. The IGR \( (S, A) = \text{entropy}(S) - \text{entropy}(S, A) \) represents the difference between the entropy of the previous data and the entropy of the data set after division.

Assuming the training set \( S = \{S_1, S_2, \ldots, S_N\} \), in the expression, \( S_i \) contains an attribute vector, which can be expressed as \( X_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \), and \( S_i \) also contains a classification label, which can be expressed as \( c_i \in C = \{c_1, c_2, \ldots, c_m\} \). In agricultural information services, \( X_i \) represents a marker for the existence of a metric vector of a module, and \( c_i \) is a marker for the existence of a module. If \( p_i \) is assumed to be the proportion of the category \( i \) in \( S \). The information entropy is expressed in the following form:

\[
\text{entropy}(S) = \sum_{i=1}^{m} p_i \log p_i
\]  

Any attribute can be applied to several different values. Suppose that \( \text{Values}(A) \) represents the set of different values in \( A \). \( S_v \) represents the collection of all attributes \( A \) whose value is \( v \) in the set \( S \).

\[
\text{entropy}(S, A) = \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{entropy}(S_v)
\]  

This expression is based on the amount of information needed for the exact classification of the \( S \) tuples after the \( A \) division.

The internal information \( I \) is introduced to represent the data \( S' \) of the training set \( S \), which is divided by \( A \). If further divided, the total amount of information is expressed in the following form:

\[
I(A) = -\sum_{i=1}^{m} \frac{|S_i|}{|S|} \log \frac{|S_i|}{|S|}
\]  

The information gain rate is essentially a compensation measure, which can effectively solve the problems in the information gain.

If it is a classification system, the evaluation indexes are mainly AUC, F-measurement and precision. The results of the prediction can be represented by the matrix shown in Table 1.

<table>
<thead>
<tr>
<th>Real Value</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Defective</td>
</tr>
<tr>
<td>Defective</td>
<td>TP(true positive)</td>
</tr>
<tr>
<td>Non-defective</td>
<td>FP(false positive)</td>
</tr>
</tbody>
</table>
1) Accuracy: \( \text{precision} = \frac{TP}{TP+FP} \);
2) The recall rate: \( \text{recall} = \frac{TP}{TP+FN} \);
3) F measure: \( \text{F} measure = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \);
4) AUC(area under ROC curve): the area under the ROC curve. The ROC curve was originally used to describe the tradeoff between revenue and cost. In the image, the real rate is represented by the Y axis, and the true rate is represented by the X axis. AUC fluctuates in the interval \([0, 1]\), and if the system is better, it can be speculated that the larger the area is.

For the weighted problem in this paper, Alberg (CLC) is used as an evaluation index, and CE (cost effective ESS) is used to evaluate the index. The CE graph is similar to Alberg, but it is also different in itself. In the CE diagram, the X axis represents the percentage of the cumulative number of lines of code after the weighted module. The closer the curve of the agricultural information service system is to the optimal system, the better the effect is.

![Figure 2. Curve](image)

4. Experiment and Analysis

4.1. Experimental Preparation

(1) Experimental environment
The hardware condition of the PC machine in this experiment is A=B. The use of agricultural products is Eclipse, and the external dependency is weka.jar.

(2) Experimental data
In this paper, the experimental data are mainly derived from the Eclipse standard data set. In these files, there are 6 files in ARFF format. These files collect not only the metric elements, but also the number of questions. In the data set, there are four kinds of metrics.
1) Name: it is a packet name or a file name corresponding to the data.
2) Pre-release: It represents the number of problems collected in the first half of the release of the version;
3) Pre-release Defects: the number of problems collected within half a year after the release of the version.
4) Complexity measure: It refers to the CK metric element and the object oriented metric element.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Abbreviate</th>
<th>Description</th>
<th>File Level</th>
<th>Package Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td>FOUT</td>
<td>Number of Method Calls(fan out)</td>
<td>Avg,max,total</td>
<td>Avg,max,total</td>
</tr>
<tr>
<td></td>
<td>MLOC</td>
<td>Method Lines of Code</td>
<td>Avg,max,total</td>
<td>Avg,max,total</td>
</tr>
<tr>
<td></td>
<td>NBD</td>
<td>Nested Block Depth</td>
<td>Avg,max,total</td>
<td>Avg,max,total</td>
</tr>
<tr>
<td></td>
<td>PAR</td>
<td>Number of Parameters</td>
<td>Avg,max,total</td>
<td>Avg,max,total</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>McCabe Cyclomatic Complexity</td>
<td>Avg,max,total</td>
<td>Avg,max,total</td>
</tr>
<tr>
<td>Classes</td>
<td>NOF</td>
<td>Number of Field</td>
<td>Avg,max,total</td>
<td>Avg,max,total</td>
</tr>
</tbody>
</table>
5) The tree structure of the abstract syntax - the node measure of the abstract syntax tree. File level and Package level data set are different in structure. Because of this, File level data can only predict File level data training system [11], and the same is the case with the Package level. We use the data in File/Package for case-based reasoning.

4.2. Experimental Results

First, the data set is predicted. Because the results of the experiment are too much and the results are very similar, only a part of the experimental results are listed, as shown in Table 3. In order to be easy to express, embedded GIS algorithm is expressed in J48, logic regression uses LR, embedded GIS is expressed in BN, and neural network is expressed using NN.

Table 3. Partial prediction results of embedded GIS system

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>precision</th>
<th>recal</th>
<th>F_measur</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set: File 2.0</td>
<td>J48</td>
<td>0.433</td>
<td>0.281</td>
<td>0.341</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.577</td>
<td>0.203</td>
<td>0.300</td>
<td>0.731</td>
</tr>
<tr>
<td>Testing Set: File 3.0</td>
<td>BN</td>
<td>0.433</td>
<td>0.344</td>
<td>0.383</td>
<td>0.769</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>0.519</td>
<td>0.242</td>
<td>0.330</td>
<td>0.774</td>
</tr>
<tr>
<td>Training Set: Package 2.0</td>
<td>J48</td>
<td>0.705</td>
<td>0.685</td>
<td>0.694</td>
<td>0.691</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.743</td>
<td>0.700</td>
<td>0.721</td>
<td>0.724</td>
</tr>
<tr>
<td>Testing Set: Package 2.0</td>
<td>BN</td>
<td>0.679</td>
<td>0.789</td>
<td>0.730</td>
<td>0.783</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>0.763</td>
<td>0.679</td>
<td>0.719</td>
<td>0.802</td>
</tr>
</tbody>
</table>

As observed in Table 3, we can see that the main advantage of logistic regression algorithm is that the prediction accuracy is high, but the disadvantage is also obvious. That is, the recall rate is too low. In comparison, the more complex the network system is, the more satisfactory the comprehensive index is, but the more time it consumes. For simple predictions, the 18 experiment can be done in a few minutes; if the algorithm is too complex, the 18 experiment will take hours to complete.

The prediction results of embedded GIS are added to the dataset, then the logistic regression prediction results are added to the data set [12, 13], and then the prediction is carried out by the weighted system. The results are shown in Table 4. By observing Table 4, we can know that if the training set and the test set come from different data, the effect of weighting system prediction is almost the same as that of using embedded GIS in embedded GIS prediction results. If the training set and the test set are from the same data, the prediction results have the phenomenon of over fitting. The logistic regression results are added to the data set to predict the embedded GIS, which has a certain improvement effect for most evaluation indexes.

Table 4. Partial prediction results of embedded GIS

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>precision</th>
<th>recal</th>
<th>F_measur</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set: File 2.0</td>
<td>J48+ BN</td>
<td>0.433</td>
<td>0.281</td>
<td>0.341</td>
<td>0.732</td>
</tr>
<tr>
<td></td>
<td>J48+ NN</td>
<td>0.433</td>
<td>0.281</td>
<td>0.341</td>
<td>0.735</td>
</tr>
<tr>
<td>Testing Set: File 3.0</td>
<td>LR +BN</td>
<td>0.587</td>
<td>0.201</td>
<td>0.299</td>
<td>0.785</td>
</tr>
<tr>
<td></td>
<td>LR +NN</td>
<td>0.594</td>
<td>0.205</td>
<td>0.305</td>
<td>0.795</td>
</tr>
<tr>
<td>Training Set: Package 2.0</td>
<td>J48+ BN</td>
<td>0.944</td>
<td>0.968</td>
<td>0.956</td>
<td>0.947</td>
</tr>
</tbody>
</table>
Figure 3 show that although there are certain differences in the system based on the 3 algorithms, the weighted effect is basically the same. In practical experiments, systems constructed by linear regression and systems constructed by using random forests. Because the structure of the two is relatively simple, the computation time is less. The system built by the neural network is relatively complex, and it will take several hours to complete the 18 experiment.

Based on embedded GIS, this paper has obtained a new agricultural information service system [14]. In the system, the values of the predicted results are weighted from large to small, so that the CLC curve and the CE curve are obtained, as shown in Figure 3. In the CLC evaluation, the agricultural information service system based on neural network has been improved to a certain extent because of the regression measurement [15]. The CE index is similar to the prediction of a single algorithm, but the space for improvement is not large, because the reasoning system family used in the 3 cases is almost the same in the algorithm. It is therefore difficult to provide valuable information for the correct prediction of the weighting system. The slope of all CE curves using embedded GIS is close to 1, in other words, the result is near random weighting.

![Figure 3](image-url)

**Figure 3.** The prediction results of embedded GIS system (part of the training set is 2, the test set is File2.1)

5. Conclusions

With the continuous development of Internet technology, the agricultural product industry has also entered the life of residents. At the same time, the quality of agricultural products has also become the focus of many people’s attention. But the embedded GIS have a direct impact on the performance of the agricultural information service system and the quality of the agricultural products. Based on this problem, this paper puts forward a new method by studying relevant literature and adopting a completely new way of thinking. After the study of embedded GIS, the new agricultural information service system is successfully produced in this paper. From the above related discussions, we can see that the method used in this paper is to quickly predict the collected data, and then combine the predicted results with data sets, so as to form new metrics. With the processing of the metric element, new data is obtained. Then the new data is reduced, and the more complex embedded GIS are used to build the new system. The agricultural information service system has significant effect on the classification and weighting of agricultural information service.

References


