

RESEARCH OF LITCHI DISEASES DIAGNOSIS EXPERT SYSTEM BASED ON RBR AND CBR

Bing Xu ^{1*}, Liqun Liu ²

¹ College of Information Technology, Guangdong Ocean University, Zhanjiang, Guangdong Province, P.R. China 524088

² Ministry of Personnel, Guangdong Ocean University, Zhanjiang, Guangdong Province, P.R. China 524088

* Corresponding author, Address: College of Information Technology, Guangdong Ocean University, Zhanjiang 524088, Guangdong Province, P. R. China, Tel: +86-759-2383639, Fax: +86-759-2383064, Email: mypaper2008@126.com

Abstract: To conquer the bottleneck problems existing in the traditional rule-based reasoning diseases diagnosis system, such as low reasoning efficiency and lack of flexibility, etc.. It researched the integrated case-based reasoning (CBR) and rule-based reasoning (RBR) technology, and put forward a litchi diseases diagnosis expert system (LDDES) with integrated reasoning method. The method use data mining and knowledge obtaining technology to establish knowledge base and case library. It adopt rules to instruct the retrieval and matching for CBR, and use association rule and decision trees algorithm to calculate case similarity. The experiment shows that the method can increase the system's flexibility and reasoning ability, and improve the accuracy of litchi diseases diagnosis.

Keywords: litchi, association rule, decision trees, diseases diagnosis, integrated reasoning

1. INTRODUCTION

The frequent diseases have already become one of the important elements to restrict the litchi planting in Guangdong province. It is reported inadequately that annually the incidence of litchi disease amounts to 45%, and the loss rate 25% or so, and thus the direct economic loss from the

disease of litchi was at least 10 million more in 2007(Wang Jibo et al., 2008). In recent years, the diseases tend to be more complicated and difficult to heal.

With the development of artificial intelligence, the expert system has been applied in agriculture domain widely (Li Daoliang et al., 2002). The traditional disease diagnosis system first selects the experts' experience with the rule-based way, and then concludes a certain diagnosis rule to reason and diagnose the disease. This RBR system is characteristic of the single representation and the understandability (Fu Zetian et al., 2001). But meanwhile it comes with some disadvantages such as low reasoning efficiency, lack of flexibility, etc.. Contrasted with this, CBR seeks solutions by visiting the previous ones to similar problems in the knowledge base. Moreover, CBR can rapidly solve complicated problems, because CBR can get answers directly by retrieving index for similar cases, and it does not need to match the rules. But it is deficient to accurately express some understandably concepts; in some conditions, it fails to secure the capability to obtain solutions. In accordance with the present situation of litchi disease prevention and treatment in Guangdong province, this paper will discuss the method that can realize the efficient and accurate reasoning of the litchi disease diagnosis by combining the advantages of RBR and CBR.

2. SYSTEM ANALYSIS

2.1 Conceptual framework

Litchi disease diagnosis means that the diagnosis agent (litchi diagnosis expert), in some ways, studies and examines information from the diagnosis object (litchi), judges the health condition, figures out the corresponding reasons and treatment plans, and draws specific conclusions. The influencing elements like diagnosis agent, diagnosis object, diagnosis element, diagnosis result, diagnosis method, etc..Correlate and interact to build an organic whole. As shown in Fig.1.

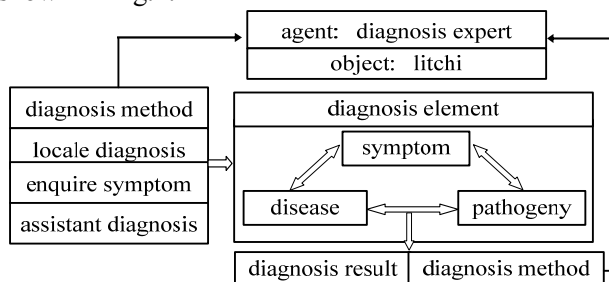


Fig.1. Conceptual framework of LDDDES

LDDES is an expert system with profound professional knowledge of litchi planting. The system applies data mining and knowledge obtaining method to gain specific knowledge of litchi planting domain, uses the artificial intelligence knowledge reasoning method to reason and study, and simulates the experts' thinking and reasoning to diagnose and cure diseases. LDDES mainly covers the following two aspects:

- (1) Mining and extracting the diagnosis knowledge hiding in the cases, and get the instruction from the preceding cases for diagnosis;
- (2) With aid of rules and cases, diagnosing new diseases swiftly and exactly, in order to satisfy users demand.

2.2 Data mining

Lots of fuzzy, incomplete and redundant information appear in the preceding cases, which make it necessary to clear up and filter all the data into consistency and determinacy (Rosembloom et al., 2006). Then classification algorithm is used to classify diseases and summarize diagnosis rules. Algorithms for data mining in LDDES are association rule algorithm and decision trees algorithm. The former gets diseases relativity by analyzing the frequently-used parameters: dependability, reliability and popularity; while the latter predicts diseases by matching the unknown-type diseases with the well-known ones.

With the measure of information entropy, decision trees algorithm chooses the suitable classification attribute and divides the sample data set into subsets. The equation of information entropy can be expressed as the follow (Kiang et al., 2002):

$$entropy(a) = \sum_{i=1}^2 \frac{\sum_{j=1}^c f_j}{R} \sum_{j=1}^c p(c_{ij}) \ln(p(c_{ij})) \quad (1)$$

In the equation, C means the numbers of classification attribute value; f_{ij} refers to the frequency of attribute value j appearing in the branch i; R stands for the data sample set; $p(c_{ij})$ can be expressed as:

$$p(c_{ij}) = \frac{f_{ij}}{\sum_{k=1}^c f_{ij}} \quad (2)$$

2.3 Reasoning method

CBR and RBR can be integrated in many ways, based on the analysis of litchi disease diagnosis. It adopts CBR as the main and RBR as the auxiliary for integration.

In the integrated reasoning system, Case X of CBR is described in a triad "X =<x1, x2, x3>". Thereinto, x1 is a finite set for expository information of diseases, such as litchi's color and weight, planting months, disease description, and description of check and diagnosis; x2 is a nonempty finite set for characteristic information of disease cases (mostly deviant values from assays being selected as characteristic information for each case); x3 is a finite set for the eliciting information of cases such as professional knowledge, treatment plans and treatment results, etc..

CBR process is mainly composed of case retrieve, case reuse, case revision and case maintenance, among which case retrieve is the key step to CBR.

The case similarity retrieve is the core process of case retrieve; it is different from the key word retrieve of common data base systems. The case situation consists of many attributes; therefore, it's essential to define the similarity of case attributes first. The equation of similarity in LDDDES is as follow (Alonso et al., 2002):

$$Sim(V_i, V_j) = 1 - d(V_i, V_j) = 1 - d_{ij} \quad (3)$$

$$d_{ij} = |V_i - V_j| \quad (4)$$

Sim (Vi, Vj) stands for the similarity quantity between Vi and Vj, d(Vi, Vj) for the distance between Vi and Vj.

With the definition above, the calculation of similarity between cases can be carried out. The similarity between cases confined by attributes distance superposition. The frequently-used typical distance definition is:

$$d_{ij} = \sum_{i=1}^N |v_{ik} - v_{jk}| \quad (5)$$

The previous definition ordinarily treats functions of each attribute exactly alike. But factually each attribute makes different contribution to the similarity, therefore, attribute weight is demanded, and another equation instead of equation (5) is as follow:

$$d_{ij} = \sum_{k=1}^N w_k d(v_{ik}, v_{jk}) \quad (6)$$

wk is the kth attribute weight, and its definition is:

$$\sum_{k=1}^N w_k = 1 \quad (7)$$

$d(V_{ik}, V_{jk})$ is the distance between the i th case and j th case in the k th attribute. Then the definition of similarity between two cases is:

$$Sim_{ij} = 1 - d_{ij}, d_{ij} \in [0, 1] \quad (8)$$

Equation (8) is adopted to calculate the cases similarities in LDDDES.

The attribute weight is determined by litchi planting experts. The weights in cases are equal under the default situation. The shorter the distance between cases is, the closer the similarity is. The closer the calculation similarity, the better the matching.

2.4 Implementation of reasoning

Implementation of reasoning includes the following steps:

(1) For a new case, the system firstly traverses the whole rules library, then with the help of RBR, obtains the elementary diagnosis, and meanwhile calculates the rate of each possible diagnosis results;

(2) A new CBR target case is built up by reasoning result of RBR and the expected diagnosed case;

(3) RBR reasoning result is utilized by CBR to instruct the searching and matching to seek similar cases;

(4) Cases are corrected and adjusted to get closer to the expected case.

(5) According to the retrieve results, the rules are adjusted to preserve in the rule library;

(6) Cases are preserved in the case library.

In the reasoning process, rules representing disease diagnosis knowledge are used to reinforce and improve CBR efficiency. The preservation of rules and cases makes incremental learning of system knowledge, and betters the reasoning capability.

3. DESIGN AND IMPLEMENTATION OF LDDDES

3.1 System architecture

Presently CBR systems contain a multitude of mixed modules. CHEF system supports CBR with submodules based on rules (Alonso et al., 2002); CASEY system complements mechanisms based on rules with cases (Qu Ming et al., 2004); Rissland and Skalak combine RBR and CBR to set up CABARET system. The architecture of LDDDES in the article is shown in

Fig.2. There are two layers: knowledge establishment layer and knowledge reasoning layer.

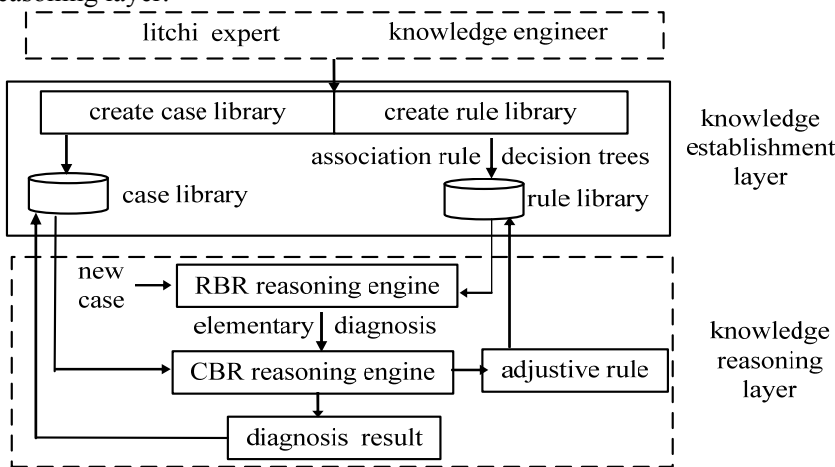


Fig.2. Architecture of LDDDES

On the knowledge establishment layer the case library is founded by a large quantity of historical typical cases summarized by litchi planting experts; the rule library is established by mining potential disease diagnosis rules from dozens of ruleless case data. On the knowledge reasoning layer the disease diagnosis result is acquired by integrated reasoning when a new case is input.

3.2 Knowledge establishment

Knowledge is shown as sufficient cases in the case library and lots of disease diagnosis rules in the rule library. Knowledge establishment is mostly finished on the establishment layer.

Case library establishment: it is founded by a large quantity of typical cases for years summarized by litchi planting experts. It' is a process ranging from preserving cases to establishing index of relevant case, thus building up a case library the reasoning system demands. The case library is the foundation of CBR.

Rules library establishment: litchi case is a set for a large volume of data in different types. The process of data mining is to use tools and algorithms to find out relations in modules and data, and to obtain potential knowledge. On the knowledge establishment layer, decision trees algorithm and association rule algorithm are used to summarize the relevant rules of disease diagnosis. By assessment and classification, the rules at last are preserved in the rule library for RBR use.

3.3 Knowledge reasoning

The reasoning process of the system is: with the adjuvant result of RBR, CBR achieves the diagnoses of target cases; with the instruction of RBR, CBR retrieves the closest case relative to the target case. The knowledge reasoning layer consists of RBR reasoning engine and CBR reasoning engine, both of which cooperate and mutually complete the task of disease diagnosis and reasoning. The reasoning steps are described as follow:

- (1) supposing case library C exist cases c_1, c_2, \dots, c_n ;
- (2) initialize x // initializing target case x ;
- (3) $x' \leftarrow \text{RBR } x$ // RBR reasoning creating initial results and regenerating target case x' ;
- (4) $C \leftarrow x'$ // inputting target case x' to case library;
- (5) retrieving target set;
- (6) for each $c_i \in \text{target set}$ do // c_i stands for each case in C
- (7) $\text{similarity}(x', c_i)$ // calculating the similarity between x' and c_i according to similarity function;
- (8) figuring out the subsequent case of $c_i : c_{i+1}$
- (9) while (all cases in target set being found);
- (10) $t = \max(\text{similarity}(x', c_j))$ // seeking case c_j which is closest to x' ;
- (11) case adjustment and amendment;
- (12) getting reasoning result and storing new case.

4. EXPERIMENT

In order to validate the reliability of reasoning algorithm, the function of LDDDES has been implemented based on Visual Studio2005 platform. The main function module of the system including: case information input module, knowledge establishment module, integrating reasoning module and result assessment module. The core algorithm used in the system including: association rule algorithm, decision trees algorithm and similarity retrieve algorithm.

The system analyzes litchi diseases assay data recorded by Guangdong Ocean University in recent three years. The data is divided into two parts: one for data base with the purpose of data mining, the other for case library purposefully for case reasoning. In the testing process, it sampled 12 random records from 1500 cases as target cases to diagnose. The diagnosis result is focused on the following litchi disease type: Capnodium (C), Pestalotia (P), Leptosphaeria (L), Ascochyta (A). The system diagnosed the selected 12

random records and find out the similarities shown in Table 1. “expert” stands for experts’ diagnosis results; “system” stands for the diagnosis results of the LDDES; “C”, “P”, “L” and “A” stand for corresponding disease type.

Table 1. Evaluation of diagnosis result

ID	1	2	3	4	5	6	7	8	9	10	11	12
expert	C	P	C	L	L	P	P	C	A	P	P	A
system	C	P	C	L	L	P	P	C	A	P	P	A
similarity	0.89	0.95	0.87	0.98	0.92	0.91	0.94	0.83	0.94	0.98	0.92	0.91

5. CONCLUSION

According to the normal disease diagnosis process for litchi, this paper researched the case-based and rule-based reasoning technology, and put forward a diseases diagnosis system LDDES with integrated method. Besides, the theory, the architecture, and the reasoning algorithm of the system were introduced. The critical technology and the function were designed. LDDES accords with the litchi diseases diagnosis. The integrated reasoning method used in this paper not only greatly reinforce the flexibility and reasoning capability but also abate the burden of case retrieve and case library. At present, LDDES has been applied to the litchi disease diagnosis in Guangdong Province.

REFERENCES

- F.Alonso, J P.Caraca, L.Gonzalez, et al. Combining expert knowledge in a medical diagnosis domain. *Expert System with Applications*, 2002, 23(4): 367-375.
- Fu Zetian, Li Daoliang, Zhang Xiaoshuang. Design on the freshwater shrimp culture expert system based on web. *Computer Application*, 2001(6):24-26 (in Chinese)
- Li Daoliang, Fu Zetian, Duan Yanqing. Fish-expert: a web-based expert system for fish disease diagnosis. *Expert System with Application*, 2002, 6(23):311-320(in Chinese)
- M.Y.Kiang, Chi.RTH. Reasoning by coordination: an integration of case-based and rule-based reasoning systems. *Knowledge-based Systems*, 2002, 12(5): 73-76
- P. S. Rosembloom, A.R.Golding. Improving accuracy by combining rule-based and cased-based reasoning. *Artificial Intelligence*, 2006, 87(2): 215-254.
- Qu Ming, Hao Hongwei.Design and implementation of real time expert system combining CBR and RBR.*Computer Engineering*, 2004, 30(18)):144-145(in Chinese)
- Wang Jibo, Fu Zhong.Advances in research of litchi germplasms native to guangdong.*Aeta Hortieulturae Sinica*, 2008, 12(2):1011-1014. (in Chinese)