ISSN : 0974 - 7435 Volume 10 Issue 11

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(11), 2014 [5567-5575]

The research of component retrieval based on artificial intelligence algorithm

Saidong Lv[1], Yaowen Xia[2]*

[1]Institute of Security, Yunnan Normal University, Kunming, Yunnan, (CHINA)

[2]Institute of Information, Yunnan Normal University, Kunming, Yunnan, (CHINA)

E-mail : lsdong_ynnu@sina.com

ABSTRACT

Component -based on Software Development is the most effective way to Software reuse in the process of Software Development Based on reusable Component, Component library management system plays an important role. In component library management system, the model of component classification and retrieval mechanism is one of the most basic core functions, and the effective retrieval mechanism can greatly decrease the cost of reuse, therefore, how to a large number of reusable components in the component library effectively retrieve became a component-based software development method to research the hot issues. This paper studies the component retrieval methods, and feasibility of artificial intelligence algorithm in the component retrieval are analyzed, and then according to the problem of component retrieval based on tabu search algorithm, designed the specific neighborhood structure and whole taboo processing method, and describes in detail the implementation process of the algorithm, the last of the algorithm is based on the simulation, the retrieval results, and it is the result of using traditional methods analyzed, prove the effectiveness of the proposed method.

KEYWORDS

Component -based on software development; Artificial intelligence; Component; Component library; Component retrieval; Reuse.

INTRODUCTION

Software reuse is considered to be a practical and feasible way to solve software crisis, software reuse, the present study focused on the component technology. Component technology component assembly procedure is proposed, which makes software concurrent to the way of industrialization. The content of the component technology research including; Component model, component description, component library (storage, classification and retrieval of component), member of compose outfit, development of components, etc. Among them, the component model, component classification, retrieval of assembly is the current software research hot spot and the difficulty[1,2].

Component based Software Development is an effective method to solve these problems. It is on the basis of some component models, according to the demand to choose one or more of the software in a component library component, then put these components are assembled, high efficiency, high quality to construct the process of application software system. As Based on the distributed object Component technology becoming mature, now Component -based Software Development theory to the practice of Software reuse is a kind of main form, and evaluation is one of the most promising directions in Software engineering.

Component based Software Development process including the Development, the Component parts of the retrieval, assemble, deploy, four steps. Finally can be reflected through integration mechanism for reusing components assembly into a complete system. To the components for assembly, the first to be found in the component library meet the requirements of component sets, this is a key link. If the component library of components is less, can by hand to choose one by one, but because of the development of computer network, the sources of the artifacts can have a variety of ways, and accordingly the number of component library of components is increasingly become more big, manual to choose component is almost impossible to finish the work, so the classification of the component must be properly and effectively organize and manage, in order to quickly to retrieve component. At the same time, meet the demand of the same function of component may have multiple, but under normal circumstances, the component of the users are all chose the first retrieved component rather than choose the most suitable component, the serious influence the component assembly efficiency and quality. So if in the beginning can from that meet the conditions of multiple components in selecting an optimal or approximate optimal components for assembly, this will greatly to provide component assembly success rate, reduce the cost[3,4].

Component retrieval is a complex process, because of the increasing number of members in component library, component retrieval can not only requires efficient retrieval of satisfy user needs a set of related components, also must be accurate in the retrieved component concentration quickly select one of the most satisfy user requirements artifacts, this is a complicated decision-making process.

METHODS OF COMPONENT RETRIEVAL

Components retrieval including publication and dissemination, namely through certain way to allow the user to learn and understand the activity component, object is released, beta users within the organization or the entire market, through the Internet or other business methods to let the user get artifacts. Different methods of component description, however, corresponds to the different component retrieval technology.

Description of component retrieval based on specifications

Component specification is described in some language formalization description of component information, component specification is the core of component interface, usually provided by the components of foreign service description and foreign request service described in two parts. According to the component specification to describe the concrete form of different, component retrieval

technology based on the description method of specifications can be divided into signature Matching and Behavior Matching, which is based on the Matching of the interface and Matching based on the behavior of the two kinds big. Interface information describes the component of the services provided or required by the name of the function, the parameter name parameters type, name, and returns a value type, the function name and parameter name will not be able to reflect the characteristics of the interface is the nature of grammar, can be replaced, and parameter types and return type represents the interface nature of grammar, can not be replaced, the general said the parameters in the component function type list and return value type list is the tone of the function, based on the matching of the interface which is based on the matching of the tone. Communications between behavior information mainly describes the components of constraint conditions and all kinds of behavior. Component specification description is a formalized description of component method, compared the behavior characteristics of formal methods which pay more attention to components rather than the appearance, can more accurately describe the component of the semantic, so component retrieval method based on the specification description can provide users with better retrieval precision rate and recall rate.

Reusable expression based on the framework

Component library of component usually has a component name!Component identifier, scope of application scope, field, using the environment, the types of components, functional description, abstract type, version number, interface properties and the characteristics of components such as the relationship between attributes. This article USES the framework to describe the components of the information.

A description of the framework is discussed by the attributes of the object data structure. A frame consists of a number of structures called "tank", can according to the actual situation of a slot each divided into several "side". Discussed a slot is used to describe objects belong to a particular aspect of, one side is used to describe one aspect of the corresponding property; Tank and side with the attribute value respectively called trough value and side.

In this paper, using framework tank first component of these attributes can be divided into three categories: (1) the basic attributes of a component nature groove;(2) interface method attributes (private and public interface methods) an interface method slot;(3) component inheritance, agglomeration properties of a connection between attribute slot: based on the UML representation of the components are mainly the relations between inheritance, aggregation, correlation and dependence relationship; Inheritance relationships between general and special properties, methods, inheritance, is the longitudinal relationship between components; And aggregation, correlation and dependence mainly through component to implement the interface between the call, horizontal relations between the performance component, therefore, to classify the relationship among the main component inheritance and gather two kinds; In this way, the frame knowledge representation of artifacts are as follows:

Components: nature tank (tank, natural property profile name, component natural attribute values) Associated attributes tank (tank, associated profile name, associated attribute values)

Rule attribute slot (tank, profile name, rules attribute values)

Interface attribute slot (tank, method profile name, attribute values)

Using backus paradigm BNF componentframework can be shown by the following:

Component = < natural logo >< associated properties >< rule properties >< interface properties>

Natural attribute = < member logo >< application >< field >< use environment >< component type > <function tracing above >< version number >

Member id = < member name >

<component identifier>

Application = < system universal | special >

Area = < telecom industry| manufacturing| bank for industry| commerce |office automation | ERP.. >

Using the environment = >< hardware environment < software environment>

CORBA component type = < CORBA><COM/DCOM/COM + ><. NET >

Functional description two< private function >< public function>

Private function = < plan scheduling >< of material tracking data collection>< cost management>

Public function = < plan scheduling >< of material tracking data collection >< cost management>

Associated properties = >< inheritance < agglomeration >< link >< rely on ><Interface attribute < private interface properties >< public interface properties >

Private interface attribute = < private method name >< parameter list >< return value >< description > functionalityPublic interface attribute = < public method name >< parameter list >< return value >< description >

Private function = < plan scheduling >< of material tracking data collection >< cost management>

Public function = < plan scheduling >< of material tracking data collection >< cost management>

Each component are described by a frame knowledge representation, property described by groove, groove can have a lot of profiles, on behalf of attribute contains content; each side can also have a lot of value, describe the components of different attribute values.

METHODS OF COMPONENT RETRIEVAL BASED ON ARTIFICIAL INTELLIGENCE

Component retrievaltechnology based on behaviorsampling

The component retrieval technology based on behavior sample components for developers to develop the component first selects some typical actual data as input, the output of the components. Such input data, output data and the return type is formed sampling artifacts. It basically is to use the component of executive ability to retrieve component, but this method can only be used to match the code level of artifacts, this is one of its greatest weakness. If you want to enable it to achieve relaxation matching and matching of complex components, it needs to implement an open platform for the sampling scripting language, due to its complexity and difficulty, it is almost impossible. So, at this stage the component retrieval based on behavior sampling technology is only applied to some simple sampling professional sex strong component library, such as mathematical functions, symbolic function and abstract data types), etc[5].

Component retrieval technology based on knowledge

Based on the analysis of knowledge component retrieval technology is mainly describe the components of natural language grammar, semantics and morphology, but it need a used to store application domain knowledge base and semantic information of natural language itself. It is the basic retrieval process of first according to functional requirements by the user, and then start the reasoning machine according to the calculated results within the system generated questions form, finally use often use the semantic web or framework to users of the results described in the form of easy to read. Compared with the traditional retrieval methods based on keyword, the component retrieval technology based on knowledge component of precise, efficient. But the cost of this method is very big, because it depends on the knowledge base of the realization of the need to manually.

FEASIBILITY ANALYSIS OF COMPONENT RETRIEVAL IN ARTIFICIAL INTELLIGENCE ALGORITHM

Component retrieval is a complex process, because of the increasing number of members in component library, component retrieval can not only requires efficient retrieval of satisfy user needs a

set of related components, must also be quick in the retrieved component meet the customers' requirements to choose one of the most optimal component combination, it is a complicated decision-making process.

Component retrieval problem can be described as: if a collection of R is a functional requirements of the user's proposed R={r_1,r_2,…r_ρ }, C is the component library of component C = {c_1, c_2,..., c_r}, a total of r a component for the user to choose, component library of every component Ci provides some functions of R_{ci}={ $r_1^i, r_2^i, \ldots, r_k^j$ }, r component retrieval is based on user demand from the component library of choice can satisfy the user requirements of component composition X={x_1, x_2, …, x_ρ }, since a component can provide multiple functions, so the xi may appear multiple times in the set x. In general, meet the demand of the same functionality can have multiple components, from the multiple components meet the customers' requirements to choose one of the most components will greatly improve the rate of component composition.

In this paper retrieval method is divided into two steps, first of all is according to the user functional requirements using the proposed retrieval method based on interface matching of meet the demand of each collection R_1, R_2,..., R, R = {C_m, C_n,..., C_p}, set any one component can satisfy the function of r_i, $R_i \subset C$, then using artificial intelligence algorithms from component sets R_1, R_2,..., R find an optimal or approximate optimal component combination X = {x_1, x_2,..., x_ρ }, ($X_i \in R_i$, i=1,… ρ) highest score, making the unit construction, also is $f(x) = \sum_{xi \in X} score(xi)$ maximum, $score(xi)$ is to evaluate each component according to certain standard after the score.

Can be seen from the problems described above, to retrieve the meet user functional requirements in the component library and the optimal collection of artifacts, its essence is the minimum set covering problem with weight[6], which is achieve for multiple weighted combination of planning. In the literature[7] has been proved with the weight of the minimal set covering problem is np-complete problem, so I can't in polynomial time to find the optimal collection of artifacts.

In order to be able to in the user the retrieval result within an acceptable time feedback to the user, this paper introduces artificial intelligence algorithm to step by step closer to the optimal or approximate optimal component combination.

THE APPLICATION OF COMPONENT RETRIEVAL BASED ON TABU SEARCH ALGORITHM

Design of neighborhood structure

Tabu search algorithm need to constantly from adjacent domain search out of the current solutions for data processing, neighborhood solution can be produced by different neighborhood operation, its design strategy to ensure the effectiveness of the current changes and neighborhood solutions and smoothness, validity refers to the neighborhood of the changes and the current solutions are different, change smoothness is refers to the neighborhood and the current solutions of difference is not too large. If there is no difference, which search cannot forward, difference is too big, the search becomes a random search. So to design appropriate neighborhood structure, both to ensure the effectiveness of the search and ensure the superiority of the search. This paper has designed two kinds of neighborhood structure:

Neighborhood structure 1: given a current solution X = {x_1, x_2,...X_i,...X_ρ }, similar to the basic a mutation in the genetic algorithm, randomly select a point I want to changei, $i \in [1, \rho]$, from the collection of $R_{i\{}$=={C_m,C_n, ……C_ρ }from one component of the Xi to replace the ith the artifacts, including $X_i \in R_i$, $x_i' \neq x_i$ the formation of a neighborhood solution for X'={x_1,x_2, ……x_i, ……x_ρ }

Neighborhood structure 2: given a current solution X={x_1, x_2, ……x_i, ……x_ρ }, randomly select two points I and j, want to change and the $i \neq j$, I, $j \in$ [1, ρ], R_i respectively from the set$R_{i\{}$=={C_m, C_n, ……C_ρ } andR_j=={C_r, C_s, ……C_t }from Xi ', two components of Xj ', and $x_i' \neq x_i$, $x_j' \neq x_j$, $x_j' \in R_j$,$x_i' \in R_i$, then form a neighborhood solution for' X'={x_1, x_2, ……x_i, ……x_ρ },all the neighborhood deconstruction has become its neighborhood N (x).

Clearly the current solution every time change can cause the change of its fitness value, neighborhood solution produced by the two methods and the current solutions for difference, this ensures that the search is ordered and not random.

In order to make the Tabu Search Algorithms achieve better Search results, at the early stage of the Search process of the neighborhood structure with large change 2 to generate the solution of the neighborhood, so to ensure the diversity of neighborhood solutions, and changes in the later stages of the Search process is to use small neighborhood structure 1 to generate the solution of the neighborhood, so improves the convergence of the algorithm.

The selection of candidate solution set

The candidate solution set is a subset of the current solution neighborhood, assuming that the size of the candidate set for CL, will neighborhood N (x) in the neighborhood solutions according to the fitness value from big to small size sorting, before taking CL solution the candidate solution V (x).

The method of taboo processing

Taboo refers to all the candidate solution concentration is no better than the solution of the optimal solution, and all of the candidate solution are taboo object. In this case can't choose a solution from the candidate solution concentration, also can't continue the iteration, the deadlock. In order to solve this problem, can use the following two ways: the first is to a neighborhood solutions of the optimal solution is the most current solutions; The second method is to a candidate solution concentration of an optimal candidate solution as the current solution, and lifted it. So miss both reduced the possibility of a good solution, and improve the convergence of the algorithm. Is the first method adopted in this paper.

ALGORITHM DESCRIPTION

In the process of tabu search to solve, for the current solution X_{now} and X_{best}optimal solution the fitness value of the candidate focus on candidate solution of the optimal X, if f (X) > f (X_{best}), regardless of whether X is taboo object, use X to replace X_{now} and X_{best}, namely X_{now} = X, Xbest = X, at the same time add X to taboo object table, modify the tabu length of each object;If f (X) $\leq$ f (X_{best}), V (X) from the candidate set of a fitness value of the optimal object of the taboo X 'as the current solution, regardless of whether it is better than the current solution, namely X_{now} = X', X' join the taboo object table at the same time, modify the tabu length of each object. The component retrieval method based on tabu search algorithm of specific process description is as follows:

Begin:

STEP1: initialize the SL the length of the vector, the size of the neighborhood NL, TL tabu length, the size of the candidate solution set of CL, the largest number of iterations G_{max}, and the optimal solution remains unchanged the biggest successive iterations S_{max}', and set the current number of iterations g = 0, the optimal solution in the number of iterations of s = 0, tabu table TabuList = $\emptyset$;

STEP2: generate an initial X_0,solution to the current solution X_{now} = X_0, optimal solution X_{best} = X_0;If s = = S_{max}or g > G_{max}End, the algorithm and output X_{best} optimal solution;Otherwise, go to step 3;

STEP3: if g < Gmax × σ , 0<σ <1, this paper select σ =0.8, this article proposed the neighborhood structure of 2 to construct the current solution, the neighborhood N (X_{now}) X_{now} otherwise use neighborhood structure to construct the N (X_{now}), its size is NL, go to the STEP4.

STEP4: select a candidate V (X_{now})from the N (X_{now}), the size of CL, turn to STEP5;

STEP5: from V (X_{now}) select a fitness value of X, the biggest solution to STEP6;

STEP6: if f (X) > f (X_{best}),then X_{now} = X, X_{best} = X, s = 0, turn to STEP9,otherwise turn to STEP7;

STEP7: if V (X_{now})objects exist in the taboos, from (X_{now}) choose a fitness value is the biggest taboo object X', make X_{now} = X', s=0,otherwise s = s + 1, go to STEP9 to STEP8;

STEP8: V (X_{now}) at this time are all taboo object, X_{best} produce a neighborhood solution of optimal solution X_{best}', makes X_{now}= X_{best}', and if f (X_{now}) > f (X_{best}), then X_{best} = X_{now}, s = 0, otherwise s = s + 1, go to STEP9;

STEP9: update TabuList, the current solution are added to the TabuList X_{now}, makes ten 1 g = g, turn to STEP2;

End

THE SIMULATION ALALYSIS

In order to verify the tabu search algorithm in the component retrieval effectiveness, this section is used for the algorithm is simulated in MATLAB, first hand more than 60 components to create a local component library, a total of 16 kinds of service function, in which each member is able to provide 1 to 16 kinds of service function. Algorithm of parameter Settings are as follows: the number of neighborhood solution 30, candidate number 10 solutions, the length of the taboo taboo object in the table 5, 100, the largest number of iterations of the optimal solution remains the same maximum 10 consecutive iterations is termination conditions.

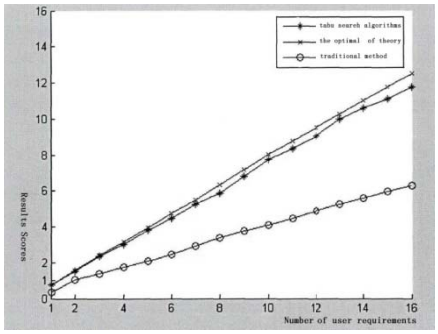

Figure 1 : Scoring component retrieval results

Figure 1 shows the user number from a gradual change to the functional requirements of 16, when using tabu search algorithm for component retrieval of component combinations of scores, and with the theory of optimal combination and according to the traditional method (using the retrieved the first meet functional component) of the retrieved component composite scores were compared.

Can be seen from the results in Figure 1 in the user functional requirements for the changing process, using a combination of tabu search algorithm for component scores have been very close to the theory of the optimal combination of cots. Number less than or equal to 3 when the user functional requirements, using a combination of tabu search algorithm for component scores and theory of the optimal basic same, when the user functional requirements after number greater than 3, two methods of component composition scores began to have certain differences, but the differences have been keep in a very stable, not with the increase of the number of user functional requirements and a lot of change. However, using a combination of traditional methods of component scores and component theory, the

optimal combination score but the difference between with the increase of number of functional requirements is becoming more and more big. It shows that using tabu search algorithm to solve the problem of component retrieval is a very effective method.

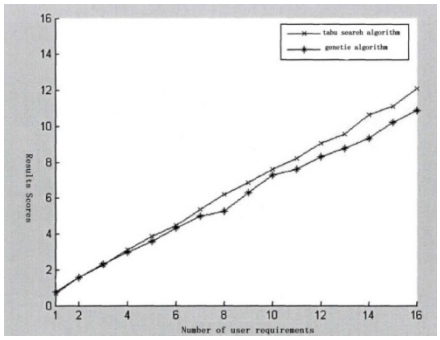

Figure 2 : Genetic algorithm and tabu search algorithm component retrieval results

Figure 2 shows the user number from a gradual change to the functional requirements of 16, respectively, using genetic algorithm and tabu search component retrieval of component combinations score comparison. Can be seen from the comparison, when the user number less than or equal to 4 functional requirements, using a combination of two methods of component scores almost the same, when the user number greater than 4 functional requirements, along with the increasing number of user functional requirements, two algorithms of the gap between the retrieval results also increases gradually, but the gap base is within an acceptable range. And can be seen from the above using component composite score obtained by tabu search algorithm is always higher than score using a combination of component obtained by genetic algorithm. So in general the tabu search algorithm and genetic algorithm solving the problem of component retrieval is very effective, but it is the result of the tabu search algorithm is more than it is the result of the genetic algorithm is good.

CONCLUSIONS

In this paper we just studied how to meet the demand of users to choose one of the many components of optimal or approximate optimal component combination of feedback to the user, but any combination in the absence of real component composition, regardless of how the other aspects of good, cannot say for sure whether the final assembling to success, so need to design a kind of fast automatic assembly method of assembly when the combination of the selected components cannot be assembled success, can put these information timely feedback to the component retrieval system, the next retrieval component reuse the information retrieval system, this will greatly improve the efficiency and quality of component-based software development.

ACKNOWLEDGEMENT

First we thank the reviewers for their constructive comments in improving the quality for this study. This work was supported by the ministry of education research project in general (NO. 14JDSZ2072), YunNan province education department fund project (NO. 2013 Y435,NO.2014Y147), Academic education research station in YunNan province, YunNan province high quality basic education resource sharing and interactive service research.

REFERENCES

[1] YaoWen Xia; The Research & Application of Network Learning Platform Based on the Component Composition, Informtion Technology Applications in Industry, Computer Engineering and Materials Science, **9**, 0842-0846 **(2013)**.
[2] YaoWen Xia; The Design of Online Examination System Based on UML & Component, Informtion Technology Applications in Industry, Computer Engineering and Materials Science, **9**,0847-0851 **(2013)**.
[3] YaoWen Xia; The Study & Application of Information System Based on Component Model and UML, Informtion Technology Applications in Industry, Computer Engineering and Materials Science, **9**, 0852-0856 **(2013)**.
[4] YaoWen Xia; Research &Application of Domain Component Reuse BasedonUML[D]. Kunming, Yun Nan Normal University, **(2007)**.
[5] Xincun Ding; The component retrieval algorithm based on tree matching and genetic algorithm is studied [D]. Xi an: Xian university of science and technology, **(2008)**.
[6] Xiaodong Wang; Algorithm design and analysis [M]. Tsinghua university press, **(2008)**.
[7] V.A.Chvatal; Greedy Heuristic for the Set-covering problem[J]. Mathematics of Operations Research, **4(3)**, 233-235 **(2013)**.