Web Service Categorization Using Structural Metrics

M. Sathya, P. Dhaavachelvan and G. Sureshkumar
Department of Computer Science, SET, Pondicherry University, India

Abstract: Service-oriented systems have become popular and presented many advantages in developing enterprise services. The service categorization is the inevitable process in web service composition for choosing most suitable services from the available services that suits to the user’s requirements. Many researchers propose several service categorization techniques like those that functional metric based service categorization and non-functional metric based service categorization. Nevertheless, those methods did not address the structural properties of the target services. We are proposing a structural metric based categorization technique for effective Web Services Composition (WSC). Coupling is the most important structural attribute of services when they integrated into a system. Structural metrics are useful to evaluate service’s quality according to its ability of coupling. We are aiming at using the coupling metrics to measure the maintainability, reliability, testability and reusability of services. We have identified two Service Colony Metrics (SCM) to formulate the Weighted Dependency Metric (WDM) which is used as the fitness function of Genetic Algorithm based Service Categorization (GASC) technique for optimizing the composite service categorization. This dependency metric based GASS technique have been tested in e-learning web services system and the experimental results shows improved performance when compared with non-functional metrics based categorization techniques.

Key words: Service categorization, structural metrics, coupling metrics, web service composition, e-learning web services, GASC

INTRODUCTION

In the present study, the web services and service oriented computing are the most promising research and development area. The research contributions will help to build cost effective and reusable service oriented applications for enterprises. Achieving the desired quality is a complicated task for all service oriented development. It has been widely accepted that high quality software should exhibit low coupling, complexity and high cohesion. Structural software does not directly describe the quality of a software product rather they are used as predictors for measuring the external quality attributes.  

Coupling is the most important structural attribute of service oriented software system and is presented by relationship between services. This relationship among services shows the dependency between them. If a service has more relationship with other services, it will depend on others much more. Therefore, if a service has more dependency relationship, the coupling between this service and others will become tighter. It can be seen from relationship in service oriented software as the more relationship exists inside software and the tighter coupling attribute is. In this study, the concept of coupling is used for measuring the strength of association established by a connection from one service to another and taking into account service properties such as stateness, dependency and interaction. This study proposes two service colony metrics which is used as the fitness function of Genetic Algorithm based Service Categorization (GASC) for optimizing the composite service selection.

Web services are the powerful mechanism for integrating existing software applications over the web, independently of programming language, execution platform or transport protocol. Web services keep on emerging at an every increasing pace. As web services increases, many businesses are providing similar services with overlapping functionalities. Therefore, it is very difficult for users to select an appropriate service among the sea of services. A number of mechanisms for web service categorization have been carried in the literature. Simpler among them is key word based categorization. This categorization mechanism searches for exact match in the UDDI registry. This research same as searching for a material in any search engine.

The existing syntactically-based service categorization technologies are insufficient for building a full-fledged composite service. The limitations of using this mechanism in web service is retrieving irrelevant services
to consumers. A number of techniques for overcoming this issue are proposed. The most common approach among this is matchmaking technique. It is used with semantic descriptions for the service's functional attributes are needed to be matched. Functional based service categorization makes use of domain ontology to achieve feasible solution to the service request. To provide appropriate solution to service categorization request, the QoS of services need to be addressed. Due to the dynamic and unpredictable nature of the web providing the suitable QoS is really a challenging task which needs to address the requirements of both service consumer and service provider. The next level of service categorization mechanism, this study focus is to select the services based on their external quality by using colony metrics. The proposed colony metrics are used as fitness function of continuous Genetic Algorithm (GA) for optimizing the composite service categorization. The GA, a naturally evolutionary heuristic methodology that can efficiently solve the problems of function optimization and control system was first put forward by Holland. Most of GA researches are binary oriented.

Many variant of genetic algorithms are discussed in Haupt and Haupt (2004). One difficulty with conventional genetic algorithms is that they suffer from premature convergence caused by an early homogenization of the genes. The process of controlling the population diversity becomes more vital when Genetic Algorithms (GA) are used for the optimization in non-stationary environments like web services where periodical changes occur. Thus GA with real-coded representation is capable of continuously adapting to changes in the environment when searching for optimal solutions (Kubalik, 2005). Therefore, this study applies a continuous parameter GA that is in which parameters are real numbers.

Moreover, the coupling/colony metrics used as fitness, subsequently is defined as that it meets the solution falls between $0$ and $1$.

**LITERATURE REVIEW**

Coupling is defined here as the degree to which each service relies on each one of the other services. Since it is the degree of interaction between services, the basic idea of coupling metric is to count how many interactions there are in between services. Nevertheless there is substantial difference depending on what counts as an interaction, how the counting is done and normalized.

Various coupling metrics have been presented in the literature. Choi and Lee (2007) proposed a dynamic coupling metric in order to measure the coupling accurately between classes allowing the dynamic property in the object level. In addition, they proved the theoretical soundness of the proposed metric by the axioms of Briand and suggested the accuracy of the proposed metric through a comparison with conventional metrics. Shen et al. (2008) proposed a fine-grained binding metrics suite for Aspect-Oriented (AO) systems to measure software changes during system evolution.

They also presented a correlation model in terms of intermediate processes for better evaluating the relation between coupling metrics and system maintainability. To investigate the practicality of their proposed model, they have implemented a coupling metrics analysis tool called AJM Metrics and performed an empirical study on eight AspectJ benchmarks.

Pereretchikov et al. (2007) proposed a set of metrics for quantifying the structural coupling of design artifacts in service-oriented systems. The metrics which are validated against previously established properties of coupling are intended to predict the quality characteristic of maintainability of service-oriented software.

Qian et al. (2006) provided a practical guide for evaluating decoupling between service-oriented components in the service composition such as Business Process Execution Language (BPEL). They suggested that a lower decoupled distributed software application would be much easy to understand, update and expand in the future. Quynh and Thang (2009) proposed a suite of metrics to evaluate service’s quality according to its ability of coupling. They used the coupling metrics to measure the maintainability, reliability, testability and reusability of services.

Their proposed metrics are operated in run-time to bring more exact results. Li and Henry (1993) defined the Message Passing Coupling (MPC) as the count of the number of send statements that is found in methods of one class to other classes. Chidamber and Kemerer (1994) introduced the Response For Class (RFC) as a measure of the number of methods that can potentially be executed in response to a message received by an object of that class. Chidamber and Kemerer (1994) defined Coupling between Object Classes (CBO) as the count of the number of classes to which it is coupled and further elaborated in the definition as two classes are coupled when methods of one class use methods or instance variables defined by one class use methods or instance variables defined by the other class. These measures are not a dynamic measure of coupling because it does not count the number of invocations during execution but these count the number of methods and variables invoked.

Prasad and Nagar (2009) defined a new set of operational measures for the conceptual coupling of classes which are theoretically valid and empirically
studied. In this study, they showed that these metrics capture new dimensions in coupling measurement, compared to existing structural metrics.

### EASE SERVICE CATEGORIZATION MECHANISMS

A number of mechanisms exist in the literature for service categorization process starting from key word categorization, matchmaking with functional attributes which makes use of domain ontology and service categorization based on the non-functional properties such as performance, time, cost, reputation etc. Several service matchmaking techniques have been developed to meet the needs of both consumers and providers.

Chua and Mustapha (2006), Zeng and Benatallah (2004) addressed this issue of selecting web services by maximizing user satisfaction expressed as utility functions over QoS attributes (Kaufmann et al., 1991; Sirin et al., 2004) developed a goal-oriented and interactive composition approach that uses matchmaking algorithms to help users filter and select services while building their composition service. In functional semantic is taken into consideration thereby avoiding unsatisfied result which are not of customer interest (Shin et al., 2009). They proposed a composition method that explicitly specifies and uses the functional semantics of web services based on domain ontology. Here the researchers have defined the functional semantics of a service as describing what a service actually does.

The service functionality of a service is represented by a pair of its action and the object of the action. In (Klusch and Kapahneke, 2008) a hybrid semantic web service categorization of semantic services in SAWSDL based on logic based matching as well as text retrieval strategies are proposed. With the predominant increase of Web services as a business solution to enterprise application integration, the QoS offered by Web services are essential for service providers and their service consumers. Due to the dynamic and unpredictable nature of the web providing the suitable QoS is really a challenging task. Zhang et al. (2006) proposed a QoS model and used hierarchy policy approach to capture goals of users, applications, environment and resources to form rational service composition and adaptation action. The researchers have proposed a Service Adaptation Evaluation (SAE) algorithm to handle service adaptation problem and service composition decision problem in pervasive computing environment.

In order to enable quality-driven web service categorization, Liu et al. (2004) proposed an open, fair, dynamic and secure framework to evaluate the QoS of a vast number of web services. The three key aspects that are developed in this technique include extensible QoS model, preference-oriented service ranking, fair and open QoS computation. The QoS model in this technique is designed to evaluate the QoS of web services without changing the computational model. Genetic algorithms approaches for service composition (Canfora et al., 2005) are proposed by Gerardo Canfora. His research focused on how genetic algorithms best suits for service composition when compared to integer programming. Continuous GA has been applied to various fields of science from physics, mathematics to medical sciences.

Kuo et al. (2008) proposed continuous genetic algorithm based fuzzy neural network for learning fuzzy IF-THEN rules. This approach uses CGA to enhance its performance. Young et al. (2007) proposed continuous GA to assimilate sensor data for a toxic contaminant release. Here, the speed is tuned by optimizing the parameters of the CGA.

### STRUCTURAL METRIC BASED CATEGORIZATION TECHNIQUE

Coupling between services comes in various forms. The composition of various service component results in a low coupling and high decoupling if the component of each service satisfies its properties such as independent, stateless and self-contained.

Service component by theoretical is a stand-alone unit but in reality service component depend on other service component they require or depend on component itself has its state (Yacoub et al., 2006). The interaction among services exists in two stages namely interaction state indirectly and keep state directly. This study proposes two separate metrics namely Mean Service Relentless Colony Metric (MSRCM) and Mean Service Transitive Colony Metric (MSTCM) for this. These metrics are used in formulating fitness function for CGA to produce optimized service categorization technique for effective web service composition.

**Continuous genetic algorithm:** The initial step in web service categorization is to locate the services. This is called as service discovery. Service discovery deals with the process of locating or discovering related service descriptions that describes a particular web service using the Web Service Description Language (WSDL). Whereas service categorization deals with choosing a service implementation among the located services to satisfy the customer need. After being the services are located based on service discovery for the user request Q, the composite services for Q can be defined as abstract services of WS1, WS2 ... WSn and is shown in Eq. 1:
Q = \{WS_1, WS_2, WS_3, ..., WS_n\}  \hspace{0.5cm} (1)

Each web service in turn may have concrete services with WS_{11}, WS_{12}, WS_{13}, ..., WS_{1n}. Therefore, services WS1 to WSn can be represented each as a single chromosome.

**Representation of chromosomes and population generation:** A chromosome is represented by a string of variable length. To each element in the string, i.e., a gene, the summation of coupling metric value is assigned according to a floating point number coding scheme (Xu et al., 2004).

Therefore, genetic operators such as selection, reproduction and mutation suitable for real numbers are employed. Figure 1 shows the set of chromosomes represented as collection of abstract services and Fig. 2 shows the floating point representation:

\[
\text{CoupM}(Q) = \sum_{i=1}^{n} \text{CoupM}(q_i)  \hspace{0.5cm} (2)
\]

Coupling metrics values CoupM(Q) in gene (a single service) of the initial population are generated by summation of CoupM(q1) - CoupM(qn) according to the Eq. 2.

Where, CoupM(q1) ... CoupM(qn) denote the coupling metrics or parameters of each abstract service WS_{n1}, WS_{n2}, WS_{n3} ... WS_{nn}. The population generated is considered to select a sub-set of individuals for further evolution. This process is called selection and the most common selection process would be to always select the top scoring chromosomes and execute it. This study uses roulette-wheel selection. With this selection process even the chromosome with a lower fitness has a chance to be selected. In this selection, parents are selected according to their fitness. The better the chromosomes are the more chances to be selected they have. For example, four different abstract services with fitness equalled to 2.5, 2.1, 4.3 and 1.1 are placed on the wheel for selection. The selection is then based on randomly generated number. For web service selection, the steps for applying roulette-wheel can be formulated as:

- Calculate sum of all chromosome fitnesses in population-sum S (is performed only once for each population)
- Generate random number from interval (0, S)-r
- Go through the population and sum fitnesses from 0-sum s. When the sum s is > r, stop and return the chromosome

The selection of individuals will be crossed by a crossing operator. With services, this means that two particular chromosomes will be combined together to form a new ones. Thus results in new offspring. Here, the crossover operator used is standard two-point crossover. With 2 point crossover, 2 points are selected randomly on the parent chromosomes. Then, every gene between the 2 points is swapped between the parent chromosomes. With 2 service example, the 2 point crossover is applied to randomly select 2 abstract services and thus yielding 2 new abstract services. The commonly applied mutation operator randomly selects a service (i.e., a position in the genome) and randomly replaces the corresponding with another service one among those available. But with continuous genetic algorithm where the values are real numbers, mutation is applied by adding or subtracting a small number to randomly selected gene. The fitness function is a function that evaluates the closeness of every chromosome to the optimal solution by calculating value obtained. The idea behind this is to minimize the coupling among service components for effective service composition. Therefore, in general the fitness function is:

\[
\text{Min ColoMet} = f(\text{MSRC}, \text{MSTC})  \hspace{0.5cm} (3)
\]

Where:

ColoMet α = Mean service relentless colony
ColoMet α = Mean service transitive colony

167
ColoMet is directly proportional to mean service relentless colony metric and mean service transitive colony metric. The lower the service colony metrics as fitness function in genetic algorithm, the lesser the coupling among service components and achieve better maintainability, reliability, testability and reusability of services.

**SERVICE COLONY METRICS**

The techniques and measuring of existing coupling/colony metrics are classified by procedural programming and object oriented programming. This study proposes a two decoupling metrics namely Mean Service Relentless Colony Metric (MSRCM) and Mean Service Transitive Colony Metric (MSTCM) based on para meters of service stateness and interaction.

**Mean service relentless colony metric:** The relentless colony metric checks whether there exists a weighted link between service components. If there exist a link, the binding time is calculated. The stateness of a service can be classified as stateful or stateless. Stateful service defines that the service is aware of who made the request and remember what happened in the past. The stateless service component can research in request/response mode without knowing the knowledge of where the request has come from. For stateful service to be aware of other, it needs to have a common registry or belongs to same domain to track the status of interaction among service components. In general stateful services have a bind (dependency) with its own service component state or share with other service component. The dependency is defined as:

\[
MD = \frac{1}{n} \sum_{i=1}^{n} C_i
\]  

(4)

Where \( n \) is the number of service components in the domain and \( C_j \) is the service component which has its state. If it is stateful or have link then \( C_j = 1 \) otherwise \( C_j = 0 \). It means that if \( C_j = 1 \) then there exist dependency among service components and the domain is same.

Therefore, the service components are dependent and tightly coupled. In general, the lower the dependency is the looser the coupling between service components. There also exist indirect state binding in which a number of service components uses a same state. This is known as relentless dependency and is defined as:

\[
MSRC = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij}
\]  

(5)

**Mean service transitive colony metric:** Any service oriented software system comprises a set of services defined as:

\[
S = \{S_1, S_2, ..., S_n\}
\]  

(6)

Where \( S \) is abstract service component which may be either aggregation of concrete service components or service composition components that lead to containment composition which is shown in Fig. 3. Let \( C_j = \{C_{j_1}, C_{j_2}, ..., C_{j_n}\} \) be the concrete service components of the composition component \( S_j \). The strength of coupling between pair of services can be formulated by finding the number of dependency paths between the service components.

Let \( CC_{ji} \) represents the set of dependency paths between service component \( S_i \) that have interaction with service component \( S_j \) for \( j \neq i \) (\( C_{ji} \) is defined to be null). The interaction from \( S_j \) to \( S_i \) exists if and only if \( CC_{ji} \) is not null. Thus, it reflects the direct coupling of one service to another. The interaction in directed service \( C_{ji} \) is not necessarily equal to \( CC_{ji} \). \( CC_j \) the set of all dependency paths from service component \( S_j \) to all other service components can be defined as:

\[
CC_j = U_{i=1..n} CC_{ji}
\]  

(7)
Now we have to associate a weight with each dependency pairs that reflects direct coupling from one service to another. This weight should also reflect the fact that a service interacts many concrete services that has a greater chance of interacting services from any composition component. The direct colony metric of service component $S_i$ is defined as:

$$DC(i, j) = CC_i/CC_i + C_j$$  \hspace{1cm} (8)$$

Where $CC_i+C_j$ indicate the total number of dependency paths exists from service component $S_i$. The value of dependency colony metric ranges from 0-1. That is:

$$0 < DC(i, j) < 1$$  \hspace{1cm} (9)$$

To measure the transitive dependency among concrete services of composite service components, indirect coupling between services are included. Suppose that $DC(i, j)$ and $DC(j, k)$ have dependency but that $DC(i, k)$ is zero. Thus although, there is no direct coupling between service $S_i$ and service $S_k$, there is a dependency because there exist a path from service $S_i$ to service $S_j$ which in turn have a path to service $S_k$. Since the strength of this dependency depends on two direct coupling of which it is composed, the measure is provided by $DC(i, j) * DC(j, k)$. That is the strength of the coupling is the product of all $DC$ paths. Thus, $TC(i, j, \pi)$, the transitive coupling between service $S_i$ and service $S_j$ is defined as:

$$TC(i, j, \pi) = \prod_{s \in \pi} CC_s / tCCs + C_s$$  \hspace{1cm} (10)$$

Where $s$ and $t$ denote the path between $s$ and $t$. In general there may be more than one path having a non-zero $TC$ dependency between any two services. This raises an issue how paths may be combined to produce an overall measure of the coupling between services. The solution is to select the path with largest transitive dependency and define mean transitive dependency $(i, j)$, the strength of coupling between two service component $S_i$ and service component $S_j$ is defined as:

$$MSTC(i, j) = 1/n2\cdot n TC(i, j, \pi_{max})$$  \hspace{1cm} (11)$$

Where $\pi$ is the set of all paths from service $S_i$ to service $S_j$. With the discussed service colony/coupling metrics, the final shape of the fitness function can be structured as:

$$MinColoMet = f(MSRC = \frac{1}{n} \sum_{i=1}^{n} R_{ik} \sum_{j=1}^{n} \frac{1}{n} - nTC(i, j, \Pi_{max}))$$

**EXPERIMENTAL RESULTS**

To test the performance of the service colony metrics as a fitness function in continuous genetic algorithm based service categorization, a prototype model was built for e-Learning web services management. Each course materials or documents of e-learning application was developed as web services and registered in UDDI registry. The platform used for implementation process was Microsoft .NET 2005. MySQL 5 was used as database for UDDI registry. This prototype starts working by getting the user query for selecting e-document. After the query has been given by the user, the system locates the related services. For example, the user query computer networks, the system generates the related services by semantically matching it with the domain ontology created. The e-Learning domain ontology was created with the use of Protege 3.1.

The identified service colony metrics measures the external quality of located services and initial population for genetic algorithm is computed. The fitness of each service is calculated based on the Eq 12.

The performance of the proposed technique was analyzed by observing the fitness factors over the population generation and compared with QoS (non-functional quality of service) aware service categorization using continuous genetic algorithm which we developed. It is proven that service colony metrics based service categorization performs well than QoS aware service categorization technique. After a total of 2000 generations, no improvement in fitness was found and the fitness was observed. The Fig. 3 shows the service categorization performance.

**CONCLUSION**

In this study, two service coupling/colony metrics are proposed for web service categorization. Service properties such as direct dependency and indirect dependency are discussed and are used as fitness function in continuous genetic algorithm.

The performance of service categorization for effective service composition out performs the earlier QoS based service categorization technique and produces a near optimal solution.

**RECOMMENDATIONS**

The future enhancement aims to identify more number of service colony metrics for measuring the quality of a service in all aspects of service properties. Work-in-progress is also devoted to apply memetic algorithm instead of CGA.
REFERENCES


