Provenance-Based Trust Estimation for Service Composition

Jing Jiang and Quan Bai

School of Computing and Mathematical Sciences,
Auckland University of Technology, Auckland, New Zealand,
Jing.Jiang@aut.ac.nz; Quan.Bai@aut.ac.nz

Abstract. Provenance information can greatly enhance transparency and accountability of shared services. In this paper, we introduce a trust estimation approach which can derive trust information based on the analysis of provenance data. This approach can utilize the value of provenance data, and enhance trust estimation in open dynamic environments.

Keywords: Trust estimation, provenance, service recommendation

1 Introduction

Nowadays, with the development of open distributed systems, increasing number of services and information are shared on open platforms. For many open distributed systems, trust is a crucial factor that reflects the quality of service (QoS) and helps manage correlation among interactive service components. Provenance data, which describes the origins and processes that relate to the generation of composite services, can provide rich context for trust estimation [1]. Especially in service-oriented computing, provenance identifies what data is passed between services, what process involved in the generation of results, who contributed to the service generation, etc. [4].

In this paper, a provenance-based trust estimation model is proposed. In this model, provenance information of a composite service is represented as a provenance graph. The similarities of different provenance graphs are analysed according to their Same Edge Contributions (SEC). Based on graph similarities and correlation to trust values, the performance of a future composite service can be predicted.

The rest of this paper is organized as follows. Section 2 describes the problems definition and some assumptions in this research. Section 3 presents the framework of the provenance-based trust evaluation model, and how to derive trust support values from provenance graph. In Section 4, we setup experiments and demonstrate the performance of the SEC model. Finally, the conclusion and future works are presented in Section 5.

2 Problem Definition

When a service consumer submits a service request, workflows which can satisfy the request will be proposed by different providers. The system will estimate each proposed workflow based on the analysis of historical provenance data (graphs). We suppose that there is a universe of n service components $S = \{S_1, S_2, ..., S_n\}$ which are loosely coupled in a SOC system. $E_x(S_i, S_j)$ represents a path leads from S_i to S_j. Firstly, we give the definition for provenance graph in knowledge base.

Definition 1: A *provenance graph* is a 2-tuple $PVG = (V_{PVG}, E_{PVG})$, where V_{PVG} is a finite set of nodes, and E_{PVG} is the finite set of edges. Furthermore, $\mid G_{PVG} \mid = \mid V_{PVG} \mid + \mid E_{PVG} \mid$ denote the size of G_{PVG}

The requests from service consumers include basic functional requirements, and then system will receive proposal graph from different providers as following definition.

Definition 2: A *proposal graph* PRG is defined as 2-tuple, i.e., $PRG =< ID$, $PRG = (V_{PRG}, E_{PRG}) >$. ID is the unique identifier for each service request. PRG is the proposal graph from providers that describes a finite set of service components $V_{PRG} = \{S_1, S_2, S_3, ..., S_n\}$ and a finite set of edges $E_{PRG} = \{E_1(S_1, S_2), E_2(S_1, S_3), ..., E_n(S_{n-1}, S_n)\}$.

The service components in V_{PRG} are required to achieve the functional requirement of the request, and E_{PRG} indicate the process of composite service. After the completion of the composite service, the system will generate service feedback RF which contains the proposal graph PRG and quality of composite service.

Definition 3: A *service feedback* RF is defined as a 2-tuple, $RF =< R, Q >$. R is the *service request* generated by the system which contains both unique transaction ID and provenance graph PRG. PRG describes the required service components and process in detail. Q represents the quality of composite service.

Definition 4: A sub-service graph $g = (V_g, E_g)$ is a subgraph of a graph PRG or PVG, denoted by $g \subseteq PRG/PVG$, where $V_g \subseteq V_{PRG}/V_{PVG}$ and $E_g \subseteq E_{PRG}/E_{PVG}$.

3 The Provenance-Based Trust Estimation Approach

3.1 Trust Estimation Protocol

In our approach, trust prediction is conducted by the protocol shown in Fig.1. Firstly, after the system receives a request, proposal graphs PRG based on the

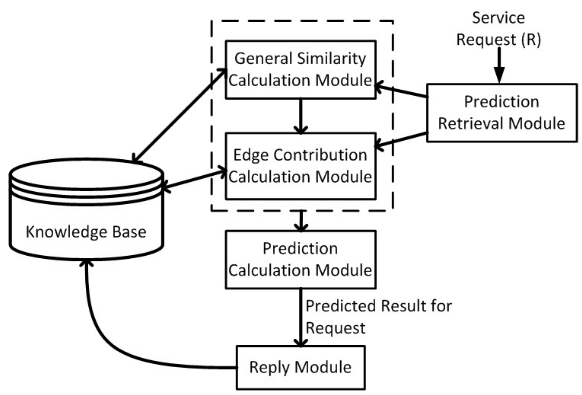


Fig. 1. Trust Estimation Protocol

functional requirements from the service consumer will be generated. Then the proposal graphs will be sent to the Prediction Retrieval Module. The Prediction Retrieval Module will search Knowledge Base for all possible provenance graphs PVG which are similar to proposal graph PRG. Then, based on the previous provenance graphs PVG in the Knowledge Base, the Edge Contribution Module will update the edge contribution value for total available edges. At the beginning all edge are given the same weight within in the request. General Similarity Calculation Module calculates the similarity between proposal graphs PRG and provenance graph PVG based on the same service components and edges in graphs and then passed the most similar provenance graphs PVG to Prediction Calculation Module. Comparing the same edges in proposal graph PRG and provenance graphs PVG, the Prediction Calculation Module will use edge contribution value to give each provenance graph PVG support value. The system will return the class value of the provenance graph PVG which obtained the highest support value to the Reply Module.

3.2 General Similarity Calculation

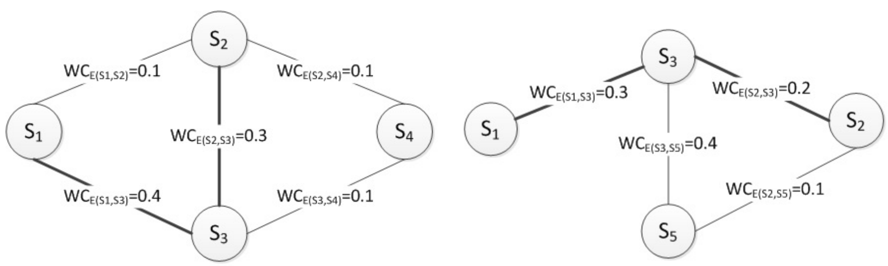
The general similarity between proposal graph PRG and candidate provenance graph PVG in knowledge base is decided by an upper bound on the size of the Maximum Common Edge Subgraph (MCES) [5] [2]. First, according to service service components S in each graph, the set of vertices is partitioned into l partitions. Let g_i^{PRG} and g_i^{PVG} denote the sub-graph in i^{th} partition in graph PRG and PVG, respectively. An upper-bound on the similarity between provenance graph PRG and PVG can be calculated as follows:

$$V(PRG, PVG) = \sum_{i=1}^{l} min\{|g_i^{PRG}|, |g_i^{PVG}|\} \tag{1}$$

$$E(PRG, PVG) = \lfloor \sum_{i=1}^{l} \sum_{j=1}^{max\{|g_i^{PRG}|, |g_i^{PVG}|\}} \frac{min\{d(S_j^{PRG}), d(S_j^{PVG})\}}{2} \rfloor \quad (2)$$

$$sim(PRG, PVG) = \frac{[V(PRG, PVG) + E(PRG, PVG)]^2}{[|V(PRG)| + |E(PRG)|] \times [|V(PVG)| + |E(PVG)|]} \quad (3)$$

where $d(S_j^{PRG/PVG})$ denotes the number of adjacent service components of S_j in provenance graph PRG/PVG. Fig.2(a) and Fig.2(b) illustrate two workflow graphs for composite services. The higher $sim(PRG, PVG)$, the more same edge and nodes are share between the proposal graph PRG and candidate provenance graph PVG in knowledge base. It is necessary to specify a minimum acceptable value $sim^{threshold}$ for the general similarity measure. If $sim(PVG, PRG) \leq sim^{threshold}$, the candidate provenance graph PVG will be ignored in following edit operation cost calculation procedure.



(a) Proposal Graph PRG (b) Provenance Graph PVG

Fig. 2.

3.3 Edge Contribution Calculation

In our approach, we intend to adopt the *Edge Contribution* which each edge $E(S_i, S_j)$ makes to quantify the edit operation cost. The Quality of Service (QoS) of a composite service is assumed as a random behavior. The uncertainty of such a random behavior is related with the required edges $E_G^x(S_i, S_j)$ in the process of service request, and can be reduced with the existence of a particular edge. Therefore, we firstly calculate the Quality Entropy ($H(Q)$) to measure average uncertainty of the QoS value of composite services [3]. Then, mutual information (i.e., $I(Q; E_G^x)$) [6] is used to measure how much reduction can a particular edge E_G^x make to the uncertainty of the QoS value.

$$C_{E_G^x(S_i,S_j)} = \frac{I(Q; E_G^x(S_i, S_j))}{H(Q)} \quad (4)$$

$$WC_{E_G^x} = \frac{C_{E_G^x}}{\sum\limits_{E_x \in E_G} C_{E_G^x}} \tag{5}$$

where G represents as PRG or PVG in different situations and where $WC_{E_G^x}$ is the contribution of edge E_G^x for PRG or PVG. The larger $WC_{E_G^x}$ is, the most contribution the edge E_G^x makes in the process.

Comparing proposal graph PRG and candidate provenance graph PVG passed from General Similarity Calculation step, we can get the particular same edge set between PRG and each PVG, i.e., $\{E_{sameSet}^i\} = Same(E_{PRG}, E_{PVG}) = \{E_i, E_j, E_k, ...\}$, where all edges $\{E_i, E_j, E_k, ...\}$ in $\{E_{sameSet}^i\}$ both occur in PRG and PVG. For example, according to Fig. 2(a) and 2(b), $\{E_{sameSet}\} = \{E(S_1, S_3), E(S_2, S_3)\}$. Then, we should separately calculate the Same Edge Contribution rate (SEC) on proposal graph PRG and provenance graph PVG as follow:

$$SEC_{PRG/PVG} = \frac{\sum\limits_{E_x \in E_{sameSet}^i} WC_{E_{PRG/PVG}^x}}{\sum\limits_{E_x \in E_{PRG/PVG}} WC_{E_{PRG/PVG}^x}} \tag{6}$$

The edge contribution ($WC_{E_G^x}$) in different graphs is different. In order to compare the contribution of same edge set which both occur in PRG and PVG, we should calculate as follow:

$$Support = SEC_{PRG} * SEC_{PVG} \tag{7}$$

The Support value will range from 0 to 1. In order to get a high Support value for particular provenance graph PVG, Same Edge Contribution rate for proposal graph (SEC_{PVG}) and provenance graph (SEC_{PVG}) should not only as high as possible, but also as close as possible. The class which the proposal graph PRG should be classified into is dependent on the support value of each provenance graph PVG. Finally, the Reply Module generate a feed back RF for the proposal PRG after the execution, and store the information into the Knowledge Base.

4 Experiments and Analysis

Some experiments are conducted in this research. In the experiments, we included 10 service components S_i, and 45 kinds of edges $E(S_i, S_j)$. There are 2 kinds of class (i.e., Successful and Unsuccessful) are adopted for representing QoS. We except for two classification metrics: Accuracy and Precision for Successful class. We design three different scenarios for the experiment. Firstly, all training dataset and test dataset share the same set of service components and edges. Secondly, there appear new service components and edges in test, but they cannot been found in knowledge base. Thirdly, provenance graphs PVG with new service components and edges are added into the knowledge base.

Finally, following characteristics of the SEC model can be demonstrated. Firstly, even if there appear new service components and edges in *Request* provenance graph without in original knowledge base, the SEC model still can work and perform better in Precision for predicting Successful composite service. Secondly, according to the result from experiment, the Precision for the SEC model in three experiments seems to be similar, because they shared the same high contribution edge set. Thirdly, once an new edge is included into knowledge base, if it highly contribute to the class value, its $WC^{x}_{E_G}$ will immediately reflect it and influence the prediction ability of the SEC model.

5 Conclusion and Future Work

In this paper, we investigated the possibility of using provenance graphs in trust estimation, and proposed a trust estimation model, named the SEC model, for predicting the trustworthiness of a composite service based on related provenance information. The proposed approach can work effectively to facilitate users to analyze huge amount of provenance data, and derive trust information from them automatically for service composition in open systems. The future work of this research will mainly focus on two aspects. Firstly, we are going to investigate more advance methods to improve the accuracy of trust estimation. Secondly, we will investigate more effectively approach for estimating multi-class trust values.

References

1. Quan Bai, Xing Su, Qing Liu, Andrew Terhorst, Minjie Zhang, and Yi Mu. Case-based trust evaluation from provenance information. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pages 336–343. IEEE, 2011.
2. Horst Bunke and Abraham Kandel. Mean and maximum common subgraph of two graphs. *Pattern Recognition Letters*, 21(2):163–168, 2000.
3. T. M. Cover and Joy A. Thomas. *Elements of information theory.* Wiley, New York, 1991.
4. Juliana Freire, David Koop, Emanuele Santos, and Cláudio T Silva. Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 10(3):11–21, 2008.
5. John W Raymond, Eleanor J Gardiner, and Peter Willett. Rascal: Calculation of graph similarity using maximum common edge subgraphs. *The Computer Journal*, 45(6):631–644, 2002.
6. R. Renner and U. Maurer. About the mutual (conditional) information. *Designs, Codes, and Cryptography*, 2000.