TAGER: Transition-Labeled Graph Edit Distance Similarity Measure on Process Models

Zixuan Wang, Lijie Wen, Jianmin Wang, and Shuhao Wang

School of Software, Tsinghua University, Beijing 100084, P.R. China
iamwangzixuan@hotmail.com, wenlj00@mails.tsinghua.edu.cn, jimwang@tsinghua.edu.cn, shudiwsh2009@gmail.com

Abstract. Although several approaches have been proposed to compute the similarity between process models, they have various limitations. We propose an approach named TAGER (Transition-lAbeled Graph Edit distance similarity MeasuRe) to compute the similarity based on the edit distance between coverability graphs. As the coverability graph represents the behavior of a Petri net well, TAGER, based on it, has a high precise computation. Besides, the T-labeled graphs (an isomorphic graph of the coverability graph) of models are independent, so TAGER can be used as the index for searching process models in a repository. We evaluate TAGER from efficiency and quality perspectives. The results show that TAGER meets all the seven criteria and the distance metric requirement that a good similarity algorithm should have. TAGER also balances the efficiency and precision well.

Keywords: Business Process Model, Transition-labeled Graph, Edit Distance, Behavioral Similarity.

1 Introduction

Repositories with hundreds and even thousands of process models become increasingly common [8] and the organizations reach higher levels of Business Process Management (BPM) maturity [9]. There is an urgent requirement for searching the specific models among the process repository. Although a variety of techniques that can manage these repositories have been proposed already, the requirement for precise similarity measure is becoming more and more urgent. There are several approaches to measure the similarity between process models. A general classification for those algorithms [12] is as follows, according to which kind of information these algorithms are mainly based on.

- Text similarity: based on a comparison of the labels that appear in process models (e.g., task labels, event labels), using either syntactic or semantic similarity metrics, or a combination of them.
- Structural similarity: based on the topology of process models, possibly taking text similarity into account as well.
- Behavioral similarity: based on the execution semantics of process models.
Because the similarity measures based on text or structure only consider the label set or topological structure, they lack lots of information. Comparatively, the similarity measures based on behavioral features can have a better performance and provide a more convincing result. The existing behavioral similarity algorithms have various advantages. However, they have limitations, too. The PTS similarity measure algorithm [11], proposed by Jianmin Wang in 2010, is based on labeled Petri net. It defines three types of principal transition sequences and computes a result by weighting the similarity value of each type. But this approach considers the loop structure and other structures separately. As a result, it breaks the behavior semantics and it cannot handle Petri nets with loop structures effectively. TAR similarity algorithm [15] represents a model’s behavior by transition adjacency relations. It computes the similarity by Jaccard coefficient of two TAR sets but cannot tackle the long-distance dependences between transitions. BP algorithm [14] gives the similarity based on behavioral profiles between transitions. However, it cannot distinguish the loop and parallel structures. Furthermore, it cannot deal with process models with invisible tasks.

TAGER, as a behavioral similarity measure, chooses Petri net as its modeling notation. The Petri net analysis tool, coverability graph [10], can represent the behavioral features of a Petri net well. However, the transition labels as the most useful information are stored on a coverability graph’s edges. So we define the Transition-labeled graph (T-labeled graph for short), the isomorphic graph of a coverability graph, to emphasize the transition label on the edges of a coverability graph. Because graph edit distance is a classical method to measure the difference between two graphs, so TAGER choose to measures the similarity between T-labeled graphs by using their graph edit distance. And TAGER modifies and extends the traditional edit operations to get a precise similarity. The graph edit distance generally has been used in structural similarity measure, however, we use it based on the coverability graph which is the behavioral feature of the Petri net. Besides, compared to the traditional edit operations, we redefine substitute operation and consider some common behavior patterns. Our main contributions in this paper are as follows:

1. We propose an efficient and effective approach named TAGER to compute the behavioral similarity of process models. It is based on the edit distance of T-labeled graphs, which contain all behavior information of Petri nets.
2. To improve the precision, we redefine the graph edit operations [5] according to process models’ structural features. We take the type and scale of vertices into account when redesigning the vertex substitution operation.
3. We define the similarity measure by combining the model scale with graph edit distance and give a sound computation approach with the best weight arrangement obtained by experiments.

The remainder of the paper is structured as follows. Section 2 formulates the problem and introduces the relevant concepts which will be used throughout the paper. Section 3 presents our algorithm which is used to measure the similarity. Section 4 presents the experimental evaluation of our algorithm. Section 5 concludes this paper and sketches some future work.