Pattern Recognition Letters 151 (2021) 310–316

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Efficient approximate approach for graph edit distance problem

Adel Dabah[a,*], Ibrahim Chegrane[b], Saïd Yahiaoui[a]

[a] CERIST Research Center on Scientific and Technical Information, Algiers, Algeria
[b] CoBIUS lab, Department of Computer Science, University of Sherbrooke, Canada

ARTICLE INFO

Article history:
Received 12 December 2020
Revised 11 June 2021
Accepted 23 August 2021
Available online 14 September 2021

Edited by: Prof. S. Sarkar

Keywords:
Graph matching
Graph edit distance
Approximate approach
Pattern recognition,

ABSTRACT

Graph Edit Distance (GED) problem is a well-known tool used to measure the similarity/dissimilarity between two graphs. It searches for the best set of edit operations (in terms of cost) that transforms one graph into another. Due to the NP-hardness nature of the problem, the search space increases exponentially making exact approaches impossible to use for large graphs. In this context, there is a huge need for approaches that give near-optimal results in reasonable time. In this paper, we propose a tree-based approximate approach for dealing with GED problem. It operates on a search tree that models all possible solutions of the problem. Since exploring the whole tree is impractical; this approach keeps only the best k nodes at each level of the tree for further exploration. This reduces enormously the execution time without scarifying the solution quality. Experiments using small and medium size data-sets show the low deviation of our results as compared to the optimal results of a Depth First Search algorithm. Moreover, our approach show a strong scalability potential by dealing with large data-sets in low execution time.

1. Introduction

Graph representation is a powerful tool to model large number of real-world objects from human faces to fingerprints and biology. It allows to overcome many problems related to image representation such-as: scaling, rotation, and translation of images. As a results, they are used widely as a powerful representation of objects in pattern recognition. Thus, a pattern recognition problem becomes a problem of graph matching.

Graph Edit Distance (GED) approach is a well-known technique used to measure the similarity/dissimilarity between two graphs (objects). It was first introduced by [1]. The goal of GED is finding the best set of edit operations, in terms of cost, needed to transform one graph into another [2]. The allowed operations are insertion, deletion and substitution. Due to noise, the graph representation of identical real-world objects may not match exactly. For this reason, GED can be used at the same time to find the *Exact*, and also the *Inexact* matching, where some errors are tolerated.

GED problem is known to be very challenging due to its NP-hardness nature [3], which means that the time needed to compute the minimum distance between two graphs increases exponentially with the number of vertices. As a result, optimal approaches like Depth First Search (DFS) and A-star algorithms are impractical due to theirs prohibitive running time, especially when dealing with large graphs. To deal with such large graphs, using approximate approaches is inescapable.

In this paper, we propose an efficient approximate approach that answers the huge need for approaches that ensure tight bounds for GED problem in a low execution time. Our approach is an adaptation of a detection algorithm used in communication field, known as *K-best algorithm* which is a version of the Sphere Decoder algorithm [4].

The proposed approach operates on a search tree that models all possible edit-paths that transform one graph into another, i.e., all possible solutions of the problem. The tree is built using the branching process over vertices. This latter decomposes a problem (tree node) into several smaller sub-problems which are treated in the same way until reaching leaf nodes (solutions). The proposed approaches mimics Breadth-First Search (BFS) and explores the search tree level by level. At each level, our approach has two phases: A first phase where the approach performs the branching process over all nodes of a given level. After that, a second phase begins by evaluating all resulting nodes from the first phase; and then selects only the best k nodes, in terms of evaluation, for the next level. This process is repeated until reaching the last level where solutions exist. Thus, returning the best one in terms of edit-distance. The interesting fact about this approach is its fixed time complexity. As a fact, the complexity our approach depends on the number of partial edit-distance calculations. This latter depends on the number of selected nodes at each level, and the size

[*] Corresponding author.
E-mail addresses: adabah@cerist.dz (A. Dabah), Ibrahim.Chegrane@usherbrooke.ca (I. Chegrane), syahiaoui@cerist.dz (S. Yahiaoui).

https://doi.org/10.1016/j.patrec.2021.08.027
0167-8655/© 2021 Elsevier B.V. All rights reserved.