International Journal of Artificial Intelligence and Machine Learning
Volume 11 • Issue 1 • January-June 2021

A Method Based on a New Word Embedding Approach for Process Model Matching

Mostefai Abdelkader, Dr. Tahar Moulay University of Saida, Algeria
Mekour Mansour, LabRI Laboratory, Ecole Superieure en Informatique, Sidi Bel Abbes, Algeria

ABSTRACT

This paper proposes a method based on a new word embedding approach for matching business process model. The proposed method aligns two process models in four steps. First activity labels are extracted and pre-processed to remove meaningless words, then each word composing an activity label and using a semantic similarity metric based on WordNet is represented with an n-dimensional vector in the space of the vocabulary of the two labels to be compared. Based on these representations, a vector representation of each activity label is computed by averaging the vectors representing words found in the activity label. Finally, the two activity labels are reported as similar if their similarity score computed using the cosine metric is greater than some predefined threshold. An experiment was conducted on well-known dataset to assess the performance of the proposed method. The results showed that the proposed method shared the first place with RMM/NHCM and OPBOT tools and can be effective in matching process models.

KEYWORDS

Alignment, Matching, Process Model, Vector Representation, Word Embedding, WordNet

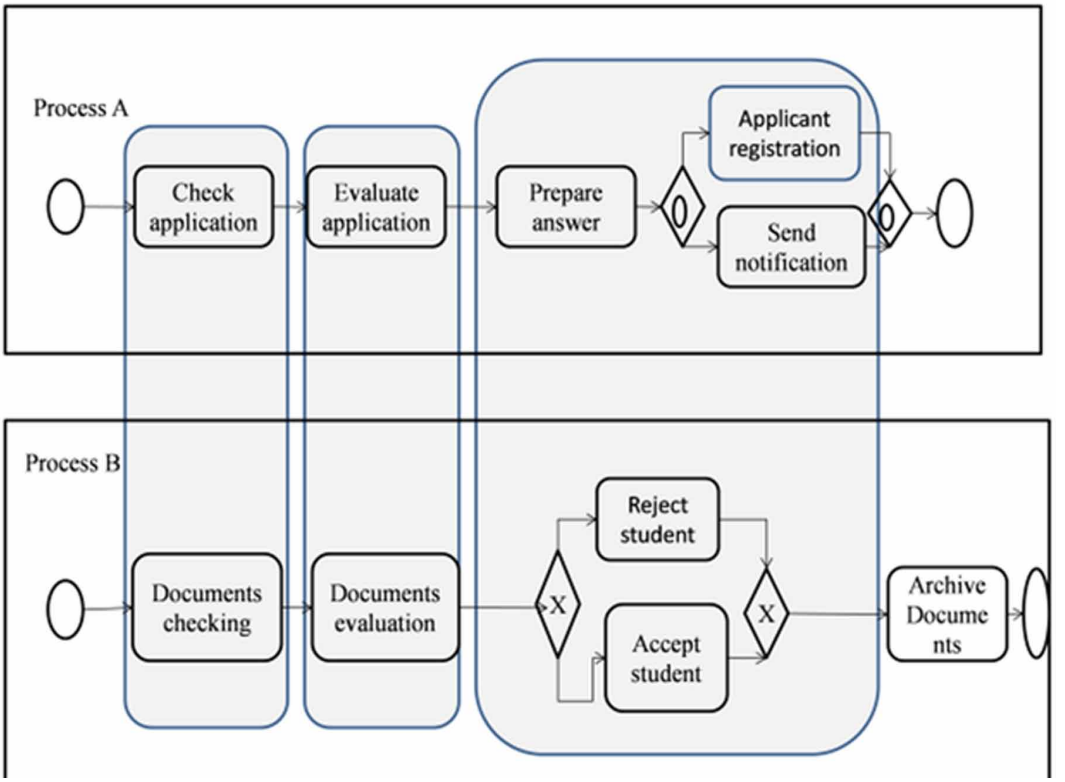
INTRODUCTION

Process model matching(i.e., PMM)is an activity widely conducted in organizations to align process models. This alignment is critical for many business process (i.e., BP) management tasks such as storing, merging, clustering or querying Business Process models (i.e., BPs) (La Rosa et al. 2013; Goncalo et al. 2015; Weidlich et al. 2012).

Technically an alignment is a set of correspondences between activities of two process models. Each correspondence is a pair of two semantically similar sets of activities. The first set of a pair contains activities from the first BP and the second set of this pair contains activities from the second BP. Formally, an alignment is a set of not overlapped matches pairs (i.e., correspondences) $\{(A_{1I}, A_{2I}),(A_{1.2},A_{22}),...,(A_{1n},A_{2n})\}$. Each pair defines a match (i.e., correspondence) between a set of activities, A_{1i}, from BP_1 and a set of activities, A_{2i}, from BP_2. Two pairs (A_{1i}, A_2i) and (A_{1j}, A_{2j}) does not overlap iff$A_{1i} \cap A_{1j} = \varnothing$ and $A_{2i} \cap A_{2j} = \varnothing$. A correspondence (A_{1i}, A_{2i}) between a set of activities A_{1i} from

DOI: 10.4018/IJAIML.2021010101

Figure 1. An example of an alignment



one BP and a set of activities A_{2i} means that the activities A_{1i} and A_{2i} refer to the same activity in the domain. Figure 1 presents of an example of an alignment between two process models.

Correspondences between the activities are presented using the grey shades. Examples of correspondence is between the set {*prepare answer, applicant registration, send notification*} in the process A and the set {reject *students, accept student}* in the business process B.

The objective of any process model matching technique is to find automatically such correspondences.

Many approaches have been proposed to achieve this objective(Goncalo et al. 2015; Daniel et al. 2017, Xue, 2019; Khurram et al. 2019). The proposed approaches are widely based on a combination of lexical, syntactic and semantic similarity metrics proposed in different fields such as the NLP (i.e., Natural language Processing) community (Goncalo et al. 2015; Daniel et al. 2017).

The literature showed that these approaches perform better in the context of ontology matching (Euzenat & Shvaiko, 2007) where they were proposed first and that they are less effective when adapted to the context of process model matching (Daniel et al., 2017). Thus matching process models stills an open problem and work needs to be done to find effective techniques for BP models matching (Goncalo et al. 2015; Cayoglu et al. 2013; Meilicke et al. 2017). The difficulty is caused by the vocabulary mismatch between activities in process models and the level of details of these activities. This means that similar activities in business process models are often described using different words and with different level of details (Weidlich et al. 210). For example, the activity "check order" could be represented in the other BP using different words as "verify request" and the activity "inform applicant" could be described using the following activities "send acceptance letter, send reject letter".