Optimizing Ontology Alignments by Using Neural NSGA-II

Mohamed Biniz, Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Beni Mellal, Morocco
Rachid El Ayachi, Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Beni Mellal, Morocco

ABSTRACT

In this article, the authors propose a new hybrid approach based on a continuous Non-dominated Sorting Genetic Algorithm II (NSGA-II) and a neural network to refine the alignment results. This approach consists of three phases: (i) pre-alignment phase which allows to identify the formats of input ontologies, to adapt them and to transform them into Ontology Web Language (OWL) in order to solve the problem of heterogeneity of representation. (ii) alignment phase which combines syntactic and linguistic matching techniques and methods, based on the relevant attributes per different points of syntactic and structural technic. (iii) The post-alignment phase which optimizes the matching by a hybrid technique of continuous NSGA-II and networks of neurons. This approach is compared with the greatest systems per the Ontology Alignment Evaluation Initiative (OAEI) standard. The experimental results appear that the proposed approach is effective.

KEYWORDS
Matching, Neural NSGA-II, NSGA-II, Ontology Alignment, Semantic Web

1. INTRODUCTION

In the early 2000s, Tim Berners-Lee (Hendler, Berners-Lee, & Miller, 2002) explained what could be a “semantic Web” evolution.

The evolution of the Web of documents, websites, gave access, via forms to databases. These research results, readable for humans, are “unreadable” for the machines.

The Semantic Web is defined by Berners-Lee et al (Berners-Lee, Hendler, & Lassila, 2001), as an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

In the context of the semantic Web, semantic interoperability based on ontologies has become an important challenge. Today, several disciplines have emerged to improve this interoperability and solve problems of ontological heterogeneity, such as alignment, integration and fusion.

The alignment of ontologies is a complex task based on the definition of the correspondences between ontologies. It is a discipline aimed at improving interoperability between heterogeneous ontologies. It connects ontologies with each other to make integration tasks possible, information sharing between systems easier, searching for more relevant information, and so on.

At present, several methods have been proposed to solve the heterogeneity. Zghal et al. (Zghal, Yahia, Nguifo, & Slimani, 2007) and Pushpakumar et al. (Pushpakumar, Srirangam, Baba, Meenachi, & Balasubramanian, 2016) are based on the semantic and structural similarity between the entities of two ontologies. Cotterell et al. (Cotterell & Medina, 2013), Albagli et al. (Albagli, Ben-Eliyahu-Zohary, & Shimony, 2012) and Sonntag et al. (Sonntag, Hake, Fettke, & Loos, 2016) are used learning

DOI: 10.4018/JECO.2018010103

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
machines to classify the corresponding entities. These works give good results, but also many gaps, such as the selection of good thresholds or the use of linear classifiers. To remedy these problems, this paper presents a hybrid ontology alignment method, by calculating the terminological, extensional, structural, and linguistic similarity between the entities, to construct a similarity table.

This paper is organized as follows: Section 2 deals with similarity measurements and ontology alignment. Section 3 describes the architecture of our system and how it works. In section 4, the evaluation procedure to test the performance of our system is described. The conclusion is given in section 5.

2. STATE OF THE ART

2.1. Similarity Methods

Semantic similarity is considered as a topological similarity in mathematics, where it is associated with a function, called a function of similarity. The value of this function is often between 0 and 1. According to Euzenat scheme (Shvaiko & Euzenat, 2013) there are two methods to calculate the similarity: Basic Methods (Terminology Methods, Linguistic Methods, Structural Methods, Extension Methods and Semantic Methods) and Combination Methods (weighted sum, weighted products, etc.).

2.1.1. Terminology Methods

Jaro’s similarity:

The Jaro’s similarity (Sun, 2015) between two strings $s$ and $r$ is defined by:

$$\text{Sim}_j(s,r) = \frac{1}{3} \left( \frac{m}{|s|} + \frac{m}{|r|} + \frac{m - t}{m} \right)$$  \hspace{1cm} (1)

where

- $|s|$ and $|r|$ are respectively the lengths of the strings $s$ and $r$.
- $m$ is the number of corresponding characters.
- $t$ is the number of transpositions.

Jaro-winkler’s similarity:

The Jaro-winkler metric (Sun, 2015) produces the similarity between two strings based on the number and the order of the common characters.

$$\text{Sim}_{JW}(s,r) = \text{Sim}_j(s,r) - l.p \left( 1 - \text{Sim}_j(s,r) \right)$$  \hspace{1cm} (2)

where

- $l$ is the length of the common prefix ($l \leq 4$ characters).
- $p$ ($p=0.1$) is a coefficient that favors strings with a common prefix.
A New User Segmentation Model for E-Government
www.igi-global.com/article/a-new-user-segmentation-model-for-e-government/81318?camid=4v1a