Combining meta-heuristics and linear programming to address ontology meta-matching problem

Nicolas Ferranti[1], **Stênio Sã Rosário Furtado Soares**[2], **Jairo F. De Souza**[1]

[1] Programa de Pós-Graduação em Ciência da Computação
Federal University of Juiz de Fora (UFJF) 36.036-900 – Juiz de Fora, MG – Brazil

[2]Department of Computer Science – Federal University of Juiz de Fora (UFJF)
36.036-900 – Juiz de Fora, MG – Brazil

{nicolas1, ssoares}@ice.ufjf.br, jairo.souza@ufjf.edu.br

Abstract. *Every year, several new ontology matchers are proposed in the literature, each one using a different heuristic, which implies in different performances according to the characteristics of the ontologies. An ontology meta-matcher consists of an algorithm that combines several approaches in order to obtain better results in different scenarios. To achieve this goal, it is necessary to define a criterion for the use of matchers. In this work, a study is proposed exploring approaches of distinct areas of computing, like meta-heuristics and linear programming for the construction of an ontology meta-matcher.*

Resumo. *Todo ano, diversos novos alinhadores de ontologias são propostos na literatura, cada um utilizando uma heurística diferente, o que implica em desempenhos distintos de acordo com as características das ontologias. Um meta-alinhador consiste de um algoritmo que combina diversas abordagens a fim de obter melhores resultados em diferentes cenários. Para atingir esse objetivo, é necessária a definição de um critério para melhor uso de alinhadores. Neste trabalho, é proposto um estudo explorando abordagens de areas distintas da computação, como meta-heurísticas e programação linear para a construção de um meta-alinhador de ontologias.*

The student entered the master's program in March 2018 with a forecast of defense for December 2019
The master's program does not have a qualification examination.

1. Introduction

Ontologies are built by people with distinct levels of expertise and domain view. Therefore, concepts that describe the same type of object may be represented in different ways, both in syntax terms and structure of relations, generating a problem of heterogeneity in data semantics. To solve problems of heterogeneity, a way is to unambiguously specify the vocabularies underlying information systems [Farinelli and Almeida 2014].

The problem to be solved consists of defining relationships between concepts of the input ontologies, making the structures compatible to represent the union of the data sets in a new model. The alignment of ontologies, as it is called, is a complex problem and its characteristics allow it to be approached by several computational techniques. Due to the high heterogeneity of the ontologies, there is no technique that stands out from the others in all aspects [Xue and Tang 2017]. Therefore, meta-matching approaches can be used in this scenario. A meta-matcher combines several alignment techniques in order to explore various aspects of heterogeneity to avoid alignment performance being restricted to some ontology characteristics. The literature suggests that the best results found in the meta-matching of ontologies are associated to the use of evolutionary algorithms due to their adaptability and, consequently, adequacy of the use of each technique [Souza 2012, Shi and Eberhart 1998, Xue and Tang 2017].

The use of meta-heuristics in meta-matchers is justified by the size of the search space of the problem, which discourages the use of exhaustive approaches due to the processing time demanded. In addition, the literature shows that population approaches present themselves as good alternatives in solving the problem. Although the use of Genetic Algorithms is more frequent between approaches, other bio-inspired approaches constitute a promising field to be explored in solving the problem. In the last decade, many meta-heuristics have been proposed and are still little explored in the database area [Sorensen et al. 2017].

Among the bio-inspired population approaches, the prey-predator algorithm (PPA) has characteristics that are adequate for the meta-matching problem, since it allows promising regions of the solution space to be exploited by a set of agents (solutions), pressured to runaway from regions that are not attractive in terms of the value of the objective function, while allowing the exploration of new regions by assigning to these agents a pseudo random behavior in the definition of their displacement.

The objective of this work is to deal with the ontology meta-matching problem through an adaptation of the prey-predator algorithm (PPA) and to show its applicability to this problem. To analyze the behavior of the initial solution, the benchmark provided by the OAEI (Ontology Alignment Evaluation Initiative) was used. It was observed that PPA is efficient and effective when parametrizing the alignment techniques in order to obtain a solution that is close to the optimum in polynomial time. Thus, the work shows that PPA can be better exploited in the ontology alignment scenario.

2. Related Work

The problem of meta-matching of ontologies is a relatively recent problem and still has several characteristics to be explored, although the known proposals have presented good results. In this section we present related works that use meta-heuristics to deal with the alignment of ontologies, highlighting their main characteristics and contributions.

The use of meta-heuristics has been explored to solve the meta-alignment problem, as seen in [Souza 2012, Xue and Tang 2017, Bock and Hettenhausen 2012]. In GNoSIS+ [Souza 2012], a genetic algorithm is used to parameterize a set of predefined aligners. Algorithm learning is based on a set of reference alignments defined by an ontology engineer input. The premise is that some relationships can be easily pointed out, so AG calibrates the system functions based on the reference in order to prepare it for a real application situation. It is interesting to highlight the representation of the problem by GNoSIS+. Considering $\Xi = \{F_1, F_2, ..., F_n\}$ a set of alignment functions, each chromosome has *n* genes($|\Xi| = n$) and each gene represents a real value $w \in [0, 1]$ representing the weight to be applied to each function. The goal is to minimize the difference between the value found and the value defined by the ontology engineer for a particular relationship. [Xue and Tang 2017] also employs an evolutionary algorithm, however the objective function is to maximize the value of the harmonic mean of f-measure. The f-measure is a metric that takes into account the precision and recall rates between the mappings obtained by the algorithm with those that were expected. The objective function of each work guides its algorithms to different paths. For the [Xue and Tang 2017] approach, it is necessary to evaluate each item of the result obtained at each iteration with the reference base, resulting in a computational cost greater than just comparing the result obtained with the confidence value defined by the engineer, as is done in [Souza 2012].

MapPSO [Bock and Hettenhausen 2012] is a solution that employs the particle swarm optimization to deal with the meta-matching problem. Particle swarm is a natural-inspired technique based on the social behavior of individuals, such as the flock of birds to find a place with food enough [Shi and Eberhart 1998]. The approach aims to find only equivalence type relations (1:1) and uses a predefined matcher that implements a distance function. The distance function defines a level of similarity for a given pair of concepts. It is important to note that MapPSO does not calibrate a set of alignment functions because it uses only one. However, it can be considered a meta-matching approach by seeking optimal matching by making use of predefined matchers. The representation of the individual differs from previous work. In this approach, each solution is represented as a candidate match. Suppose that $\vec{X_p}$ represents an alignment of two ontologies consisting of *k* = 5 matches (*c*). A particle is represented by $\vec{X_p} = \left\{c_{(p,1)}, c_{(p,2)}, c_{(p,3)}, c_{(p,4)}, c_{(p,5)}\right\}$ where each $c_{(p,i)}$ indicates a confidence value for the relationship (p, i).

The literature shows that evolutionary approaches have good results when applied in the ontology matching [Otero-Cerdeira et al. 2015], which allows to foment that the use of the prey-predator meta-heuristic has the potential to construct effective results for the problem.

The definition of the individual's representation impacts on how the effort of the approach can be reproduced. Once the representation is based on the set of weights, the found parameters can be stored and retrieved without much effort, whereas the representation based on the set of candidate alignments requires that the whole process be executed again. Therefore, the weights of the sets of weights have a better contribution to the construction of a more generic meta-aligner.

3. Proposed Solution

Ontologies matching problem allows several approaches to be used in the development of algorithms to the construction of solutions. This work intends to explore the variety of characteristics, trying the combination of approaches present in the literature in search of better solutions. Thus, we selected a set of experiments to be carried out based on the studies found in literature:

1. To explore the prey-predator meta-heuristic defined in [Tilahun and Ong 2015] as a means of solving the linear system constructed by [Souza 2012], as a new ontology meta-matcher.
2. To investigate the adaptation of the prey movement presented by [Tilahun and Ong 2015] to the problem. The discretization of the solution space is already used by other authors and can contribute to reach better solutions and reduce the processing time of the solution
3. Evaluate the construction time and quality impacts of the solutions when working with a multi-objective function, by combining metrics such as precision and recall in conjunction with the linear system solution
4. Use linear programming techniques like the Simplex algorithm to maximize the objective function of the problem

Initial efforts of this work turns to the first item. In a first step, the behavior was reproduced and used to solve the linear system constructed by [Souza 2012] using as an objective function the minimization of the differences found. Each year OAEI[1] provides a set of instances for ontology matching testing, these instances are used by meta-matcher developers to demonstrate the adaptive capacity of the algorithm.

Preliminary results indicate precision and recall rates reaching on average, 90%, a difference of 6% less than the work of [Souza 2012]. The prey-predator of [Tilahun and Ong 2015] as it was proposed, allows a solution to walk at any point in a continuous solution space, making the search space be the whole set of possible solutions, the next steps of the stage of using the prey-predator focus on the customization of meta-heuristic operators so that it can walk in a discrete way, described in item 2, the discretization of the solution space is employed by other authors in other meta-heuristics and the expectations are related to a better behavior of the algorithm, achieving better results in a shorter time frame.

[Souza 2012, Xue and Tang 2017] model the problem by constructing a linear system in order to parameterize a set of matchers, although the works are different, they have many similar characteristics, which allows to experiment the behavior of an approach that combines both using a multi-objective function, which take into account the solution of the linear system and the maximization of f-measure for a predetermined set of training. Once a linear system is composed of first degree functions, it is feasible to employ the Simplex algorithm, which works in a single or multi-objective way, finding optimal solutions for the linear system. This proposal encompasses items 3 and 4, the main question here is associated to the gain in the quality of the solution: given the currently representation model of the problem, to employ a multi-objective function will produce a qualitative gain proportional to the extra time spent?

[1]http://oaei.ontologymatching.org/

Finally, there are other questions more associated to the behavior of the final alignment method, where the efforts seeks to provide better stability to the algorithm in the case of successive executions. Nowadays it is possible to observe the need to work in an interdisciplinary way in computing, using knowledge from several areas to solve a single problem.

Working on the improvement of processes related to the linear system is fundamental to obtain better results in terms of time or quality, however, it is worth mentioning that the training instances of the model should reflect the behavior of the ontologies as best as possible, which means that if the ontologies have equivalences in the nomenclature of the entity, it is of great value that in the training instances there are correspondences that reflect those characteristics for a higher quality matching.

4. Final Considerations

Although there are several matchers and meta-matchers of ontologies in the literature, the application scenarios are not always the same, this work seeks to deal with meta-matching assigning candidate matchings to ontologies with few references and returning a first matching version to the engineers. Although it is in the initial stages, the preliminary results are interesting even with few adaptations of what was found in the literature. In addition to the expected contributions to the advancement of ontology meta-matchers, this work also contributes in the area of computational intelligence, since it presents an application scenario for the meta-heuristic of [Tilahun and Ong 2015].

The first step was given using the prey-predator meta-heuristic with continuous solution space, reaching f-measure rates of the matchings found in the 90% range. In addition to the proposed improvements, this work seeks to better understand the relationship between the problem and the forms that are used to model it, which representations have a greater descriptive capacity.

The study of the use of the prey-predator meta-heuristic to calibrate a set of matchers was submitted to Brazilian Database Symposium and it has been accepted. In the paper submitted, it is described the whole process of construction, movement and adaptation of the solutions, as well as the final matching methods after adjusting the weights.

References

Bock, J. and Hettenhausen, J. (2012). Discrete particle swarm optimisation for ontology alignment. *Information Sciences*, 192:152–173.

Farinelli, F. and Almeida, M. (2014). Interoperabilidade semântica em sistemas de informação de saúde por meio de ontologias formais e informais: um estudo da norma openehr. *XVII Encontro Nacional de Pesquisa em Ciência da Informação*, 17(1).

Otero-Cerdeira, L., Rodríguez-Martínez, F. J., and Gómez-Rodríguez, A. (2015). Ontology matching: A literature review. *Expert Systems with Applications*, 42(2):949–971.

Shi, Y. and Eberhart, R. (1998). A modified particle swarm optimizer. In *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, pages 69–73. IEEE.

Sorensen, K., Sevaux, M., and Glover, F. (2017). A history of metaheuristics. *Handbook of Heuristics*.

Souza, J. F. (2012). *Uma abordagem heurística uni-objetivo para calibragem em meta-alinhadores de ontologias*. PhD thesis, Pontifícia Universidade Católica do Rio de Janeiro.

Tilahun, S. L. and Ong, H. C. (2015). Prey-predator algorithm: A new metaheuristic algorithm for optimization problems. *International Journal of Information Technology & Decision Making*, 14(06):1331–1352.

Xue, X. and Tang, Z. (2017). An evolutionary algorithm based ontology matching system. *Journal of Information Hiding and Multimedia Signal Processing*.