MapPSO and MapEVO Results for OAEI 2011

Jürgen Bock, Carsten Dänschel and Matthias Stumpp

FZI Forschungszentrum Informatik, Karlsruhe, Germany
{*surname*}@fzi.de

Abstract. This paper presents and discusses the results produced by the alignment systems MapPSO and MapEVO for the 2011 Ontology Alignment Evaluation Initiative (OAEI). The two systems implement two variants of population-based optimisation algorithms applied to the ontology alignment problem. MapPSO is based on discrete particle swarm optimisation, while MapEVO is based on evolutionary programming. Both systems optimise the same objective function, *i.e.* a function reflecting the quality of an alignment. Firstly, specific characteristics of the systems and their relation to the results obtained in the OAEI are discussed. Secondly, the results for the single tracks are presented and discussed.

1 Presentation of the system

With the 2008 OAEI campaign the MapPSO system (Ontology **Map**ping by **P**article **S**warm **O**ptimisation) was introduced [1] as a novel approach to tackle the ontology alignment problem by applying the technique of particle swarm optimisation (PSO). This year, a similar approach is introduced with the MapEVO system, following the same priciple of ontology alignment by population-based optimisation. MapEVO, however, utilises Evoluationary Programming instead of PSO.

1.1 State, purpose, general statement

The development of the presented systems is motivated by the following observations:

1. Ontologies are becoming numerous in number and large in size.
2. Ontologies evolve gradually.
3. Ontologies differ in key characteristics that can be exploited in order to compute alignments.
4. High ontology alignment quality often cannot be maximised by only assessing it on a correspondence level, but requires global quality metrics.

Solving the ontology alignment problem using a population-based optimistaion approach, addresses these observations as follows:

1. Population-based methods work inherently parallel, such that large ontologies can be aligned on a parallel computation infrastructure.

2. Population-based methods work incrementally, which allows the algorithm to start with an initial or partial configuration (*i.e.* for instance an alignment of previous ontology versions) and refine it as the ontologies evolve.
3. Population-based methods work as a meta-heuristic, *i.e.* independently of the objective function to be optimised. In the case of ontology alignment this means that the objective function can be adjusted according the particular alignment scenario at hand.
4. Population-based methods consider the quality of a complete solution, which, in the case of ontology alignment allows for the assessment of complete alignments, not only on the correspondence level.

The idea of the MapPSO and MapEVO approaches is to provide algorithms that fulfil the aforementioned characteristics. Particularly the focus is not to provide a universal library of similarity measures (base matchers) to form that specific objective function to be optimised, but rather to provide a scalable mechanism that can used with various objective functions depending on the alignment scenario at hand.

Both presented systems are still in the status of a research prototype, where recent work has been done exploiting the parallel nature of MapPSO in a cloud-based infrastructure [2].

1.2 Specific techniques used

MapPSO and MapEVO treat the ontology alignment problem as an optimisation problem and solve it by applying a discrete particle swarm optimisation (DPSO) algorithm in the case of MapPSO [3], and evolutionary programming in the case of MapEVO, respectively. To this end, both algorithms maintain a population of individuals, each representing a valid candidate alignment, which is updated in an iterative fashion in order to converge towards the best alignment. In the case of MapPSO such an individual is a swarm particle, whereas in the case of MapEVO individuals represent evolving species. The difference between MapPSO and MapEVO is as follows. In the evoluationary programming approach some individuals (species) can become extinct and others are allowed to reproduce themselves. In the PSO-based approach, the population is constant throughout the iterations but positions in the search space change according to a particles memory and communication between particles.

The opjective function in both system is the same. It composes of local components, *i.e.* assessments of the single correspondences in a candidate alignment, as well as global components that assess the alignment as a whole.

1.3 Adaptations made for the evaluation

The modalities of the OAEI force the developer to provide a fixed configuration for each system that is applied for all tracks. Thus the provided tool bundles contain a tradeoff configuration between the best configurations for the three tracks executed over the SEALS platform.

1.4 Link to the system and parameters file

The releases of MapPSO and MapEVO together with the parameter files used for the OAEI 2011 campaign are available in the SEALS Tool Respotiry accessible via the SEALS Portal (`http://www.seals-project.eu/`). Additionally, the systems and parameter files are provided for download at `http://sourceforge.net/projects/mappso/files/` in the folder `oaei2011`.

1.5 Link to the set of provided alignments (in align format)

The alignments were created via the SEALS platform and are available in the SEALS Results Repository accessible via the SEALS Portal (`http://www.seals-project.eu/`).

2 Results

Both MapPSO and MapEVO participated solely in the *benchmarks*, *anatomy*, and *conference* tracks that are run via the SEALS platform.

2.1 benchmark

Notes follow after release of final results.

2.2 anatomy

In order to identify all correspondences for the *anatomy* track correctly, it is necessary to utilise an external biomedical thesaurus. As stated by the organisers it is possible to find about half of the correspondences without using such a thresaurus.

The objective function used by MapPSO and MapEVO for this campaign does not utilise a biomedical thesaurus. However, this domain-specfic adjustment could be integrated for both systems without touching the actual search heuristics. However, including such a feature without proper self-adaptation mechanisms would significantly drop the performance in other OAEI tracks. Thus there were no efforts undertaken in including this feature in the *anatomy* track.

Notes follow after release of final results.

2.3 conference

Notes follow after release of final results.

3 General comments

In the following some general statements about the OAEI procedure, modalities, and results obtained are given.

3.1 Comments on the results

Notes follow after release of final results.

3.2 Discussions on the way to improve the proposed system

MapPSO and MapEVO are currently being worked on in order to incorporate a guided local search component for fine-tuning results found by the search heuristics. Additionally, it seems necessary for any system to provide some sort of self-adaptation of alignment quality criteria in order to perform well in all SEALS tracks of the OAEI, since different configurations for different tracks are not possible.

3.3 Comments on the OAEI 2011 procedure

The OAEI modalities require participating systems to use the same parameter configuration for each track and each test case. According to assumption 3 stated in Sect. 1.1 different alignment scenarios will most likely require different means of determining a good alignment. Assuming that an alignment tool will not used in an out-of-the-box configuration in any real-world alignment task, makes this requirement of a single (and thus compromised) parameter configuration rather artificial.

While the argument that systems should be compared with a tradeoff configuration for comparability reasons is acceptable for the *benchmarks* track, it is clearly not reasonable to use the same configuration for *anatomy* and *conference*. Here the obvious focus is to find the best possible alignment. What is currently evaluated, however, is the ability of self-adadaptation of the alignment systems, which can be another track modality, but should not distract from the goal of finding high-quality alignments in a particular domain.

3.4 Comments on the OAEI test cases

Since this year the *benchmarks* dataset is synthetically generated and previously unknown, the following comments refer to the sample data provided prior to the campaign for testing purposes[1].

One comment addresses the best possible alignment any tool could possibly achieve (a.k.a. reference alignment) from an information theoretic point of view. In general the reference alignments provided for the *benchmarks* dataset should not contain any correspondences that are information theoretically impossible to be detected, neither by automatic tools, nor manually. In other words, in a systematically generated test suite, the golden standard should not contain entries that cannot be detected because all information content was removed from the respective data set.

[1] `http://oaei.ontologymatching.org/2011/benchmarks2/index.html`

Suggestions to improve the golden standard would be either to remove correspondences from the reference alignment that have no justification, or to set the confidence values as low as the probability of simply guessing the respective correpsondence.

For instance in test case #201 the correspondences

```
/2011/benchmarks2/101/onto.owl#Conference_Trip
/2011/benchmarks2/201/onto.owl#GBCFRTQEDNXEZMVRUWLFXTDFKC
```

and

```
/2011/benchmarks2/101/onto.owl#Conference_Banquet
/2011/benchmarks2/201/onto.owl#KKRDJIPEEQFBQKOWPOPJWENCPL
```

both are denoted 100 % confident in the reference alignment. Even though there is evidence that the two classes from 201 correspond to the two classes from 101, there is no evidence for the precise assignment given by the reference alignment. Thus any tool (or human) could guess the assignment with a probability of 50 % which should be reflected in the confidence values of the reference correspondences.

4 Conclusion

The alignment systems MapPSO and MapEVO were described briefly with respect to the idea behind their population-based approaches. The results obtained by the two systems for the OAEI 2011 tracks *benchmarks*, *anatomy*, and *conference* were discussed.

Future development of MapPSO and MapEVO will be targeted towards user interaction, improved reasoning support, and guided local search in order to refine the results currently obtained by the heuristic approach.

References

1. Bock, J., Hettenhausen, J.: MapPSO Results for OAEI 2008. In Shvaiko, P., Euzenat, J., Giunchiglia, F., Stuckenschmidt, H., eds.: Proceedings of the 3rd International Workshop on Ontology Matching (OM-2008). Volume 431., http://ceur-ws.org, CEUR Workshop Proceedings (2008)
2. Bock, J., Lenk, A., Dänschel, C.: Ontology Alignment in the Cloud. In Shvaiko, P., Euzenat, J., Giunchiglia, F., Stuckenschmidt, H., Mao, M., Cruz, I., eds.: Proceedings of the 5th International Workshop on Ontology Matching (OM-2010). Volume 689., http://ceur-ws.org, CEUR Workshop Proceedings (November 2010) 73–84
3. Bock, J., Hettenhausen, J.: Discrete Particle Swarm Optimisation for Ontology Alignment. Information Sciences (2010) Article in Press.