MapPSO Results for OAEI 2010

Jürgen Bock[1]

FZI Forschungszentrum Informatik, Karlsruhe, Germany
bock@fzi.de

Abstract. This paper presents and discusses the results produced by the MapPSO system for the 2010 Ontology Alignment Evaluation Initiative (OAEI). MapPSO is an ontology alignment approach based on discrete particle swarm optimisation (DPSO). Firstly, specific characteristics of the MapPSO system and their relation to the results obtained in the OAEI are discussed. Secondly, the results for the *benchmarks* and *directory* 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).

1.1 State, purpose, general statement

The development of the MapPSO algorithm has been 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.

Solving the ontology alignment problem using a PSO-based approach, as done by the MapPSO system, tackles these observations as follows:

1. PSO works inherently parallel, such that large ontologies can be aligned on a parallel computation infrastructure.
2. PSO works 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. PSO works 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.

The idea of the MapPSO approach is to provide an algorithm that fulfils 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.

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

1.2 Specific techniques used

MapPSO treats the ontology alignment problem as an optimisation problem and solves it by applying a discrete particle swarm optimisation (DPSO) algorithm [3]. To this end, each particle in the swarm represents a valid candidate alignment, which is updated in an iterative fashion. In each iteration, knowing about the particle representing the best alignment in the swarm, other particles adjust their alignments, influenced by this best particle. A random component when adjusting an alignment makes sure that the swarm does not converge to a local optimum.

In MapPSO the quality of an alignment is determined by the average of the qualities of its correspondences, as well as by the number of correspondences in the alignment[1]. Each correspondence is evaluated by a number of base matchers, whose evaluation values are aggregated by a specified aggregator. Base matchers and aggregator can be selected via the `params.xml` configuration file. This mechanism makes MapPSO highly adjustable, since different alignment scenarios will most likely require different base matchers in order to determine similarity between entities. By following the instructions in the MapPSO documentation[2] one can easily develop base matchers and aggregators tailored to a particular alignment scenario at hand.

1.3 Adaptations made for the evaluation

Some OAEI tracks do not evaluate correspondences of all entity types. For instance in the *benchmarks* track, no instance correspondences are part of the reference alignments, while in the *instance matching* track **only** instance correspondences are part of the reference alignments. For this reason, an additional parameter was introduced for the MapPSO command-line interface that allows

[1] Apart from striving for correct correspondences, it is necessary to identify the correct number of correspondences, which is done in MapPSO by preferring larger alignments to smaller ones.

[2] `http://sourceforge.net/apps/mediawiki/mappso/index.php?title=Guide_for_implementing_base_matchers` and `http://sourceforge.net/apps/mediawiki/mappso/index.php?title=Guide_for_implementing_aggregation_functions` respectively.

the user to specify which correspondence types are to be included in the produced alignment.

1.4 Link to the system and parameters file

The release of MapPSO (`MapPSO.jar`) and the parameter file `params.xml` used for OAEI 2010 are located in `MapPSO.zip` at `http://sourceforge.net/projects/mappso/files/` in the folder `oaei2010`.

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

The alignments of the OAEI 2010 as provided by MapPSO are located in the file `alignments.zip` at `https://sourceforge.net/projects/mappso/files/` in the folder `oaei2010`.

2 Results

The *benchmarks* track was via the (preliminary) SEALS platform[3]. To this end MapPSO has been provided as a web service[4] For the *directory* track, results were computed offline and sent to the track organiser.

2.1 benchmark

As from last year's participation it became apparent that MapPSO performs better with respect to relaxed precision and recall measures than with respect to classical measures [4, 5]. For this reason, the symmetric precision and recall measures were computed[5] in addition to the classical measures as provided by the SEALS platform. Figures 1 and 2 illustrate classical and symmetric precision, and classical and symmetric recall respectively. It shall be noted that MapPSO is a non-deterministic method and therefore on a set of independent runs the quality of the results and the number of mappings in the alignments will be subject to slight fluctuations. The plots in Figures 1 and 2 were generated in a different run, than in the results obtained using the SEALS platform, thus results might not match completely.

The reason for MapPSO performing significantly better in terms of symmetric precision and recall is due to the fact that the algorithm keeps *good* correspondences not allowing them to be discarded in a later iteration. This, however, prevents entities participating in such *good* correspondences to participate in an even *better* correspondence in a later iteration. In case this *better* correspondence would be the correct one with respect to the reference alignment, the *good* one found is counted as wrong with respect to classical evaluation metrics, while its closeness is respected in the relaxed metrics.

3 `http://seals.inrialpes.fr/platform/`

4 Web service end point: `http://krake16.perimeter.fzi.de:8080/MapPSOWS`

5 Relaxed precision and recall measures were computed using the methods provided by the Alignment API.

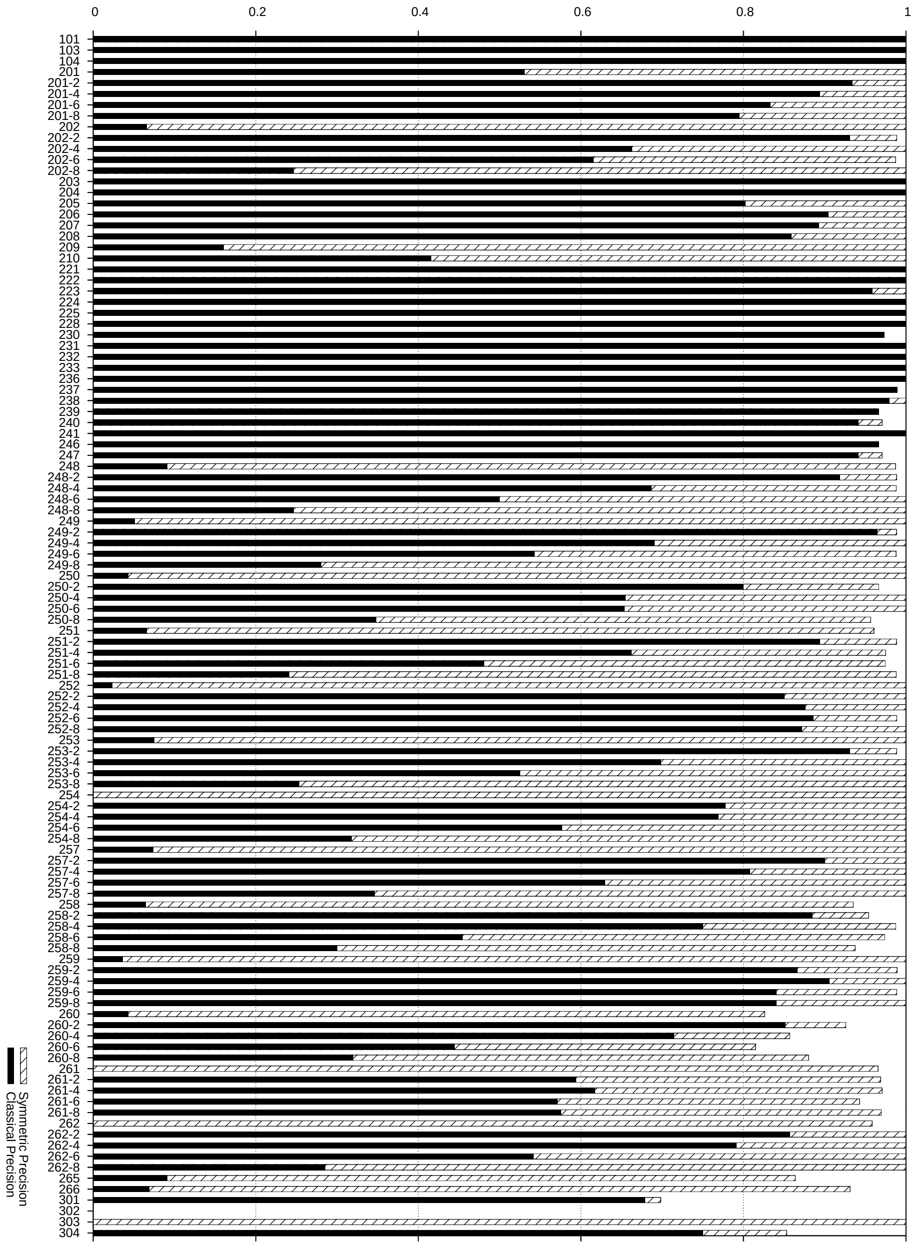

Fig. 1. Classical vs. Symmetric Precision from OAEI 2010.

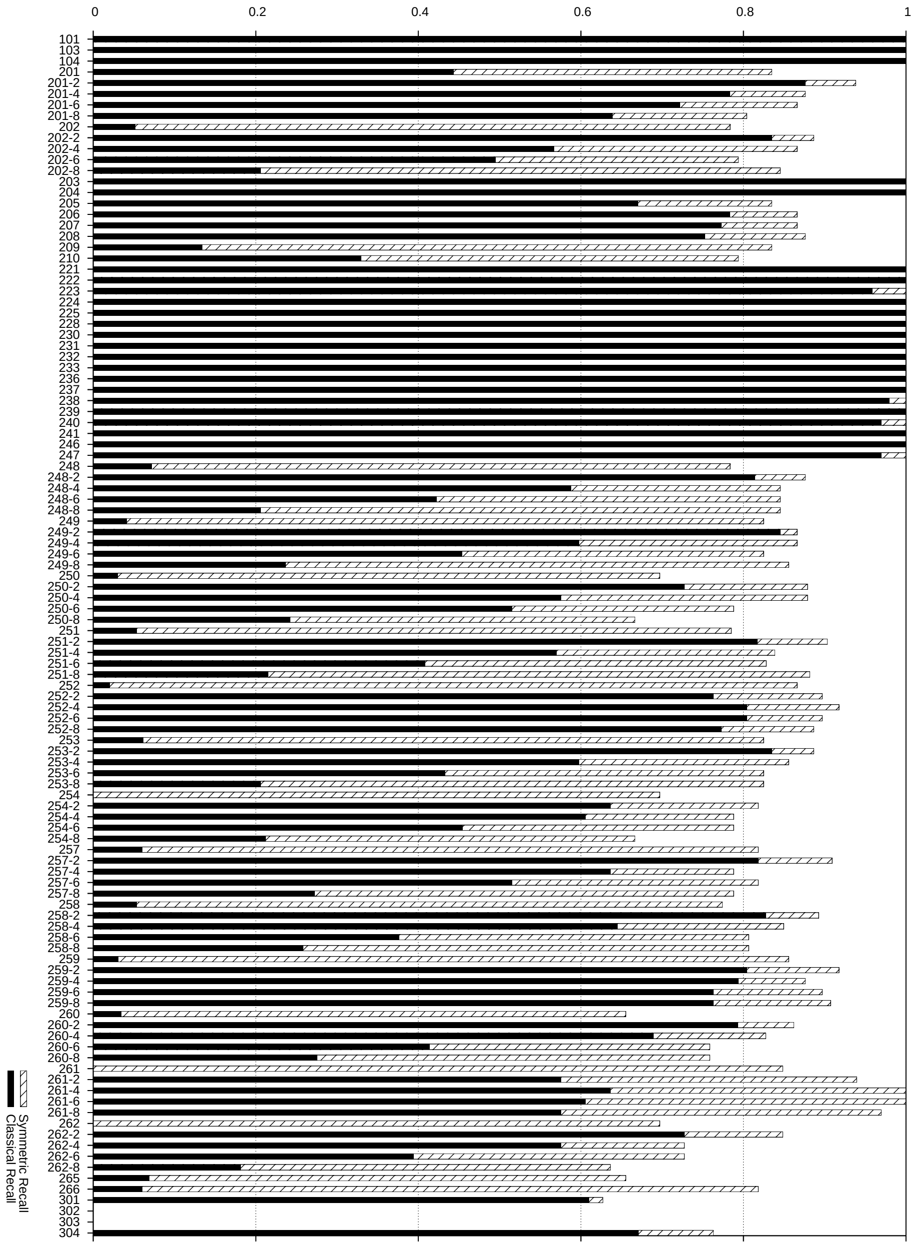

Fig. 2. Classical vs. Symmetric Recall from OAEI 2010.

2.2 directory

In the *directory* track the same set of parameters was used as in the other tracks, which includes the same set of base matchers. Due to the nature of the datasets in this track, several base matchers were not applicable or might even have contributed in a counterproductive way. For instance in the *singletask* subtrack, meaningless IDs were used as URI fragments, which could lead to a high similarity of those entities with a similar ID with respect to a particular base matcher. Since this information is known before running the matcher, deactivating this base matcher might have lead to better results. Additionally, MapPSO does not filter final results according to the confidence values gained. Thus, in its current implementation, many bad correspondences are left in the final alignment, reducing precision.

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

Compared to the results of the 2009 *benchmarks* track, a slight decrease in the symmetric precision and recall measures can be observed. This is due to the fact that this year, MapPSO has been configured with a stronger focus on finding the correct size of an alignment. The configuration used in 2009 was rather tailored to the *benchmarks* track where the alignment is known to contain all entities. Disregarding this assumption causes the symmetric recall measure to drop, but makes the system more suitable for real-world use cases.

3.2 Discussions on the way to improve the proposed system

MapPSO is currently being worked on in order to incorporate a guided search component for two reasons. Firstly, it is expected to increase convergence speed, and secondly it is expected to improve classical precision and recall due to the reasons explained in Sect. 2.1.

3.3 Comments on the OAEI 2010 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.

4 Conclusion

The MapPSO system was described briefly with respect to the idea behind its DPSO-based approach. The results obtained by the MapPSO system for the OAEI 2010 tracks *benchmarks*and *directory* were presented. Several observations regarding these results were highlighted, in particular the significant difference between classical and symmetric precision and recall. Also the effect of having a single configuration throughout all OAEI tracks were discussed.

Future development of MapPSO will be targeted towards user interaction as well as alignment refinement. As for the latter, an initial (partial) alignment, such as a previous version of an alignment, can be given as a start configuration of a particle, which is then refined by running the algorithm.

Acknowledgements

The author would like to thank Michael Mutter and Carsten Dänschel for their valuable contributions and feedback.

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 of CEUR Workshop Series., CEUR-WS.org (November 2008)
2. Bock, J., Lenk, A., Dänschel, C.: Ontology Alignment in the Cloud. In: Proceedings of the 5th International Workshop on Ontology Matching (OM-2010), http://ceur-ws.org, CEUR Workshop Proceedings (November 2010)
3. Bock, J., Hettenhausen, J.: Discrete Particle Swarm Optimisation for Ontology Alignment. Information Sciences **Article in Press** (2010)
4. Ehrig, M., Euzenat, J.: Relaxed Precision and Recall for Ontology Matching. In Ashpole, B., Ehrig, M., Euzenat, J., Stuckenschmidt, H., eds.: Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies. Volume 156 of CEUR Workshop Proceedings., CEUR-WS.org (October 2005) 25–32
5. Euzenat, J., Ferrara, A., Hollink, L., Isaac, A., Joslyn, C., Malaisé, V., Meilicke, C., Nikolov, A., Pane, J., Sabou, M., Scharffe, F., Shvaiko, P., Spiliopoulos, V., Stuckenschmidt, H., Šváb-Zamazal, O., Svátek, V., dos Santos, C.T., Vouros, G., Wang, S.: Results of the Ontology Alignment Evaluation Initiative 2009. In Shvaiko, P., Euzenat, J., Giunchiglia, F., Stuckenschmidt, H., Noy, N., Rosenthal, A., eds.: Proceedings of the 4th International Workshop on Ontology Matching (OM-2009). Volume 551 of CEUR Workshop Proceedings., http://ceur-ws.org (October 2009)