Matcha-DL a tool for supervised ontology alignment

Pedro Giesteira Cotovio^{a,*}, Lucas Ferraz^a, Daniel Faria^b, Laura Balbi^a, Marta Contreiras Silva^a and Catia Pesquita^a

^a LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

^b INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract.

Ontology Matching is a critical task to establish semantic interoperability given the proliferation of ontologies and knowledge graphs with overlapping domains. While traditional Ontology Matching relied on heuristics and rule-based approaches to find corresponding entities between knowledge resources, recent advances in machine-learning have prompted the community to contemplate matching approaches that exploit machine-learning algorithms.

We present Matcha-DL, an extension of the matching system Matcha to tackle semi-supervised tasks using machine-learning algorithms. Matcha builds upon the algorithms of the established system AgreementMakerLight with a novel broader core architecture designed to tackle long-standing challenges such as complex and holistic ontology matching. Matcha-DL uses a linear neural network that learns to rank candidate mappings proposed by Matcha by using a partial reference alignment as a training set, and using the confidence scores produced by Matcha's matching algorithms as features.

Matcha-DL was evaluated in the 2022 and 2023 editions of the Bio-ML track of the Ontology Alignment Evaluation Initiative, achieving the highest F1 score in 4 of the 5 semi-supervised tasks. Furthermore, it was shown to benefit more than other competitors from the contextual information of ontologies.

Keywords: Ontology Alignment, Ontology Matching, Ontologies

1. Introduction

Ontologies are digital resources that formalize domain knowledge in a manner that is both machine-readable and human interpretable, by encoding domain concepts and the semantic relationships between them. They have been extensively used to structure knowledge in complex domains, such as the biomedical [1]. However, the disarticulated fashion in which most application ontologies are developed results in many ontologies covering the same or similar domains. Ontology Matching (OM) aims to tackle this issue and enable interoperability between ontologies by establishing mappings between their entities.

In the last two decades, OM has become an established task, with substantial efforts by the research community to develop novel algorithms and robust systems [2]. Some systems achieve high performance on benchmark tasks through the use of sophisticated approaches based on exploring the lexical component of ontologies as well as their semantics [3–5], which are typically combined through handcrafted rules.

The *de facto* benchmark for the evaluation of OM systems is the Ontology Alignment Evaluation Initiative (OAEI) [6] which provides a number of test-cases covering various aspects of OM. Yet, until recently, OM systems were

P. Giesteira Cotovio et al. / Matcha-DL a tool for supervised ontology alignment

required to produce alignments *de novo*, which prevented the use of machine learning methods by participating systems. In 2022, the Bio-ML tasks [7] were introduced to explicitly accommodate and evaluate systems that employ machine learning methods.

We present Matcha-DL¹, an ontology and knowledge graph matching system that leverages machine learning to derive a suitable combination of ontology matching algorithms to address a specific alignment task. Matcha-DL employs the results of state-of-the-art matching algorithms developed in the AML [3] and Matcha [8] OM systems, but whereas AML and Matcha rely on handcrafted combinations based on ontology and task type profiling, Matcha-DL circumvents this need by inducing a model for combinations of algorithms based on partial reference alignments. Matcha-DL was evaluated on the Bio-ML track at the OAEI which targets the alignment of several biomedical ontologies.

The rest of the article is organized as follows: section 2 introduces the fundamental concepts on ontology alignment and related work in ontology matching with machine learning; section 3 describes the architecture of Matcha-DL and the algorithms it employs; section 4 presents the evaluation results, comparing Matcha-DL to other OM systems; and section 5 concludes the paper with a brief discussion and final remarks.

2. Background

2.1. Ontology Alignment

Under its classical setting, the OM task entails taking as input two distinct ontologies—often termed *source* and *target*—and producing as output an alignment, *i.e.*, a collection of mappings (typically equivalence) between their entities [2]. A mapping between entities of O_s and O_t is a directional link between the two ontologies, that typically is represented as a tuple $\langle e_s, e_t, r, c \rangle$ such that e_s and e_t are entities of O_s and O_t , respectively, $r \in \{\sqsubseteq, \exists, =\}$ is a semantic relation, and c is a confidence value (normally $c \in [0, 1]$) that represents either similarity between the entities or the confidence in the correctness of the mapping. There are several settings for OM, this work focuses on schema matching, where the entities matched are classes and/or properties of the ontology.

The majority of OM systems are unsupervised, typically relying on sophisticated heuristics and hand-crafted rules instead of learning a mapping function. These systems normally comprise three components: a pre-processing component that extracts the entities to match from the ontologies given the setting of the matching task; a candidate generation component that produces a range of possible mappings between entities, by leveraging matching algorithms that explore different aspects of the ontology; and a filtering component that creates the final alignment by combining this preliminary matches and eliminating improbable or irrelevant mappings from the generated candidates.

Algorithms for ontology matching (i.e., matchers) can be categorized based on the characteristics of the ontology they depend on: terminological matchers, either at a string-level or at a language-level, depend on lexical information in the ontology such as labels or synonyms; structural matchers leverage the relations between entities in the ontology either by projecting then into a graph or by considering the properties of the ontologies (*i.e.*, type or attributes); semantic matchers rely on the formal semantics inherent to the ontology typically leveraging descriptive-logic–based reasoners to interpret then.

Furthermore, matchers can be classified according to their use of external information: formal resource matchers use previous alignments and other ontological resources as references for their alignment, which is the case for most machine-learning–based matchers; and informal resource matchers use large corpora or lexicons as context for their alignment, which is the case of matchers based on language models. The latter focus on the lexical features of the ontologies but leverage informal semantics learned on the vast textual datasets on which they were trained to go beyond traditional string-matching approaches.

Since matchers output all mappings found, potentially filtered by confidence score threshold, OM systems typically employ filtering algorithms that remove mappings according to heuristics or logical reasoning, to produce

¹available at: https://github.com/liseda-lab/Matcha-DL

a high-quality final alignment. Filters can be categorized by their guiding principles: similarity, cardinality, coher ence, conservativity, and locality. Similarity filters remove mappings below a set confidence score threshold, while
 cardinality filters maintain a maximum correspondence per entity, often aiming for a strict 1-to-1 mapping. Coher ence filters, or alignment repair algorithms, eliminate mappings causing inconsistencies in combined ontologies.
 Conservativity filters exclude correspondences inferring new axioms. Locality filters ensure correspondences align
 within ontology structure, either at high or low levels. High-level locality filtering, like blocking, reduces matching
 space dimensionality, useful for large ontologies before full matching.

The benchmarks currently used for evaluating OM systems are established by the OAEI [6]. This annual competition encompasses diverse tracks tailored to assess various tasks and components pertinent to ontology alignment. Notably, in these benchmarks, AML [3] and its subsequent iteration, Matcha [8], alongside LogMap [4], have consistently outperformed other matching systems in traditional OM settings. Consequently, these systems have emerged as the preferred choices for OM tasks, setting the standard which new systems should aspire to surpass.

2.2. Ontology Alignment with Machine Learning

Machine learning offers a fresh approach for OM systems. By considering a partial reference alignment, a mapping function can be learned to generate candidates, combine matchers, filter mappings, or even attempt to learn the problem holistically. In recent years, different OM systems have been proposed that try to incorporate machine learning approaches [9–11].

However the new Bio-ML track [7] is the first to enable the fair evaluation of machine learning-based systems for OM as well as other systems. The 2023 edition of this track featured the participation of four machine learning-based systems in the equivalence matching track: our system, Matcha-DL, and three others.

BertMap [12] was the inaugural machine learning-based system integrated into this track, purposefully crafted with a focus on achieving optimal results. It performs very well across the board. This system pioneered the use of Large Language Models (LLMs) in matching systems, by employing a string-based heuristic in conjunction with BERT embeddings. This approach generates a set of candidate matches, which are subsequently consolidated into a final alignment through reasoning-based repair that leverages the ontology's structure.

SORBETMatcher [13] is an innovative system that incorporates the structural information of the ontology into BERT embeddings. It employs this approach to generate candidate matches, which are then consolidated into a final alignment by implementing a greedy heuristic.

OLaLa [14] stands as the most recent matching system. Similarly to its counterparts, it starts by leveraging stringbased heuristics to filter out obvious candidates. Subsequently, less apparent matches undergo pre-selection using a BERT-based model. These selected candidates serve as input to a generative LLM that generates a refined set of candidate mappings. The final alignment is crafted by applying constraint-based filters on these candidates.

AMD [15] combines the comparison of LLM-based representations of ontology entities through cosine similarity with a knowledge graph embedding-based approach that formulates the OM task as a relation translation.

The Bio-ML track offers various configurations, encompassing both equivalence and subsumption matching. Our
 focus lies on equivalence matching, for which Bio-ML presents two settings: semi-supervised and unsupervised.
 Matcha-DL participates in the semi-supervised track, which includes a partial reference alignment (30%) for model
 training. Within the semi-supervised equivalence matching track, two tasks are defined: Local Ranking and Global
 Matching.

2.2.1. Global Matching

Global Matching represents the conventional OM task where, given two ontologies, the objective is to generate a comprehensive set of accurate 1:1 mappings between entities from the source to the target ontologies. The resultant alignment is then evaluated against a reference alignment using Precision, Recall, and F-score metrics.

2.2.2. Local Ranking

Local Ranking addresses a more tailored challenge for machine learning models, especially considering the scalability issues inherent in global matching, where the search space escalates exponentially with the number of entities to match. Given that an OM system is expected to discern the correct alignment from a set of candidate mappings, the objective of this task is to produce a ranking for a given set of candidate mappings. Performance in this context is typically assessed using metrics such as Hits@K and Mean Reciprocal Rank (MRR).

3. Matcha-DL

3.1. Overview

Matcha-DL is an OM system that leverages partial alignments to learn to derive an optimal combination of OM algorithms to address a specific alignment task. Matcha-DL exploits the metrics provided by state-of-the-art matching algorithms developed in the AML [3] and Matcha [8] systems to learn to rank candidates based on a partial reference alignment. Under this semi-supervised setting, Matcha-DL circumvents the need for rule-based combinations drafted from task and ontology profiling typically seen in classical OM system. Afterwards, Matcha-DL can also exploit Matcha's filters for holistic matching to enforce alignment consistency.

3.2. Feature generation

Matcha-DL describes each training example by a set of features, each corresponding to the confidence score generated by a state-of-the-art matching algorithm. In principle Matcha-DL could be agnostic to the algorithms employed for feature generation, but its implementation is based on the matchers available in AML and Matcha. Table 1 provides a detailed list of all matchers used in Matcha-DL, accompanied by short descriptions. These algorithms offer a comprehensive overview of the similarity landscape, addressing various facets of ontology alignment. For an in-depth description of each algorithm, please see [3, 8].

Table 1

Matching Algorithm	Description							
LexicalMatcher	Primary matcher scoring string equivalent Lexicon entries between ontologies, favoring label-based over synonym-based correspondences.							
SpacelessLexicalMatcher	Similar to LexicalMatcher, but ignores spaces to capture spelling variants.							
WordMatcher	Hash-based matcher computing name similarity using weighted Jaccard index; accommodates spelling variants with translocated words.							
StringMatcher	Pairwise matcher computing string similarity using various metrics, with ISub metric as default.							
DirectXRefMatcher	Hash-based matcher identifying direct cross-references and shared entries between on- tologies.							
MediatingMatcher	Primary matcher and lexicon extender using external background knowledge ontology as a mediator to find matches with low lexical similarity.							
MediatingXRefMatcher	Primary matcher and lexical extender utilizing external background knowledge ontology complementing LexicalMatcher correspondences.							
BackgroundKnowledgeMatcher	Ensemble of background knowledge algorithms testing multiple sources, including MediatingXRefMatcher and MediatingMatcher.							
ThesaurusMatcher	Primary matcher extending Lexicons with inferred synonyms; captures string equivalen entries.							
LLM Matcher	Matches ontologies by computing the cosine similarity between the embeddings of thei lexicons.							

3.3. Model training

The features retrieved from the matching algorithms are then fed to a linear neural network that learns to rank candidate mappings by employing a partial reference alignment. A crucial challenge that needs to be addressed in this setting is that reference alignments are only composed of positive examples—correct mappings—whereas training the ML model requires negative examples as well.

The negative sampling strategy adopted by Matcha-DL was designed to address the inherent imbalance in alignment tasks. In a global alignment scenario where the objective is to establish a one-to-one alignment, the number

of incorrect mappings outnumbers the number of correct mappings at a ratio of at least N to 1, where N is the size of the target ontology: for each correct mapping, all other possible pairings of the source entity with target entities are incorrect. However, in a local ranking setting, where the task is to select the correct mapping among a list of candidates produced by traditional matchers, this ratio is expectedly lower, as comparatively few pairings meet the criteria of the matchers. Taking this into account, Matcha-DL generates *n* negative examples for each positive example in the reference alignment, where *n* cannot be larger then the reference alignment, and is by default 99 (to adjust to the expected imbalance on the Bio-ML track). The negative examples are random pairings of entities that occur in correct mappings in the reference alignment, to increase the likelihood that they are true negatives (as the reference alignment is incomplete).

After training, the feature generation step can be repeated for unseen data and the model can be used to rank the most likely candidates. A global alignment can be created by selecting the highest ranking mapping for each entity pair and filtering low-confidence mappings.

4. Evaluation

The Ontology Alignment Evaluation Initiative (OAEI) is a coordinated international initiative which purpose is to stipulate consensus on ontology or schema matching evaluation methods. Currently, the OAEI seeks to evaluate entries based on each entry's performance regarding Precision (P), Recall (R), and F1 score (F1), along with Mean Reciprocal Rank (MRR) and Hits@K (H@K) metrics on multiple ontology matching tasks.

$$P = rac{|M_{ ext{pred}} \cap M_{ ext{ref}}|}{|M_{ ext{pred}}|}, \quad R = rac{|M_{ ext{pred}} \cap M_{ ext{ref}}|}{|M_{ ext{ref}}|}, \quad F1 = rac{2PR}{P+R}$$

In the formulas above, M_{pred} stands for the set of mappings computed by the system and M_{ref} stands for the set of reference mappings. Regarding the semi-supervised setting, as null reference mappings are involved, the Precision and Recall formulas instead are the following:

$$P = \frac{(|M_{\text{pred}} \cap M_{\text{ref}}) - M_{null}|}{|M_{\text{pred}} - M_{null}|}, \quad R = \frac{|(M_{\text{pred}} \cap M_{\text{ref}}) - M_{null}|}{|M_{\text{ref}} - M_{null}|}$$

For local ranking evaluation, the Mean Reciprocal Rank and Hits@K metrics are calculated in the following manner:

$$MRR = \sum_{i}^{N} rank_{i}^{-1}/N, \quad H@K = \sum_{i}^{N} \prod_{rank_{i} \leq k}/N$$

In these formulas, N is the number of reference mappings in testing and $rank_i$ is the relative rank of the reference mapping in the set of candidate mappings.

4.1. Bio-ML track

The Bio-ML track is an OAEI ML-friendly biomedical track for equivalence and subsumption matching. This track seeks to accommodate both machine learning-based and traditional ontology matching systems within one evaluation framework.

The ontologies present in this track are the OMIM (Online Mendelian Inheritance in Man), ORDO (Orphanet
 Rare Disease Ontology), NCIT (National Cancer Institute Thesaurus) and DOID (Human Disease Ontology), FMA
 (Foundational Model of Anatomy) and SNOMED CT ontologies.

Task	System	Р	R	F1	MRR	H@1	Р	R	F1	MRR	H@1
	AMD	0.601	0.567	0.583	-	-	0.601	0.567	0.583	-	-
	ATMatcher	0.925	0.251	0.395	-	-	-	-	-	-	-
	BERTMap	0.762	0.548	0.637	0.877	0.823	0.645	0.592	0.617	0.891	0.841
	BERTMapLt	0.781	0.507	0.615	0.777	0.727	0.782	0.507	0.615	0.766	0.716
OMIM-ORDO (Disease)	LSMatch	0.594	0.223	0.325	-	-	0.834	0.456	0.589	-	-
	LogMap	0.788	0.501	0.612	0.805	0.744	0.821	0.614	0.703	-	-
	LogMapLt	0.919	0.261	0.407	-	-	0.919	0.261	0.407	-	-
	LogMapBio	-	-	-	-	-	-	-	-	-	-
	Matcha	0.694	0.511	0.589	-	-	0.718	0.519	0.602	-	-
	Matcha-DL	0.887	0.578	0.700	0.600	0.583	0.745	0.732	0.738	0.811	0.780
	OLaLa	-	-	-	-	-	0.655	0.570	0.610	-	-
	SORBETMtch	-	-	-	-	-	0.568	0.652	0.607	-	-
NCIT-DOID (Disease)	AMD	0.858	0.770	0.811	-	-	0.858	0.770	0.811	-	-
	ATMatcher	0.954	0.604	0.740	-	-	-	-	-	-	-
	BERTMap	0.823	0.887	0.854	0.968	0.955	0.831	0.883	0.856	0.960	0.938
	BERTMapLt	0.889	0.771	0.826	0.903	0.883	0.888	0.770	0.825	0.890	0.861
	LSMatch	0.665	0.565	0.611	-	-	0.908	0.664	0.767	-	-
	LogMap	0.896	0.661	0.761	0.559	0.363	0.811	0.959	0.879	-	-
	LogMapLt	0.976	0.575	0.723	-	-	0.976	0.575	0.723	-	-
	LogMapBio	-	-	-	-	-	-	-	-	-	-
	Matcha	0.883	0.754	0.813	-	-	0.839	0.750	0.792	-	-
	Matcha-DL	0.955	0.801	0.871	0.810	0.804	0.847	0.834	0.841	0.870	0.844
	OLaLa	-	-	-	-	-	0.880	0.861	0.870	-	-
	SORBETMtch	-	-	-	-	-	0.925	0.882	0.903	-	-
SNOMED-FMA (Body)	AMD	0.861	0.709	0.778	-	-	0.861	0.709	0.778	-	-
	ATMatcher	0.216	0.223	0.219	-	-	-	-	-	-	-
	BERTMap	0.811	0.708	0.756	0.967	0.950	0.970	0.669	0.792	0.965	0.947
	BERTMapLt	0.970	0.665	0.789	0.897	0.871	0.970	0.662	0.787	0.892	0.865
	LSMatch	0.762	0.070	0.128	-	-	0.673	0.411	0.511	-	-
	LogMap	0.646	0.580	0.611	0.542	0.328	0.770	0.577	0.660	-	-
	LogMapLt	0.958	0.543	0.693	-	-	0.958	0.542	0.693	-	-
	LogMapBio	-	-	-	-	-	-	-	-	-	-
	Matcha	0.845	0.592	0.696	-	-	0.846	0.502	0.630	-	-
	Matcha-DL	0.998	0.756	0.856	0.790	0.782	0.959	0.825	0.887	0.918	0.908
	OLaLa	-	-	-	-	-	0.202	0.339	0.253	-	-
	SORBETMtch	-	-	-	-	-	0.794	0.704	0.746	-	-
				Table 2	2						

- OMIM is the primary online ontology source of genes, genetic phenotypes, and gene-phenotype relations, generated through manual curation based on biomedical literature [16].

- ORDO is a classification of rare diseases and relationships between diseases, genes, and epidemiologic features [17].

- NCIT is an extensive ontology on cancer-related concepts [18].

- DOID describes human diseases [19].

Task	System	Р	R	F1	MRR	H@1	Р	R	F1	MRR	H@
	AMD	0.952	0.746	0.836	-	-	0.952	0.746	0.836	-	-
	ATMatcher	0.920	0.563	0.698	-	-	-	-	-	-	-
	BERTMap	0.941	0.724	0.818	0.963	0.941	0.898	0.715	0.796	0.971	0.95
SNOMED-NCIT (Pharm)	BERTMapLt	0.973	0.429	0.595	0.835	0.758	0.973	0.569	0.718	0.849	0.77
	LSMatch	0.976	0.548	0.702	-	-	0.952	0.603	0.738	-	-
	LogMap	0.893	0.609	0.724	0.821	0.699	0.899	0.606	0.724	-	-
	LogMapLt	0.994	0.594	0.743	-	-	0.994	0.594	0.743	-	-
	LogMapBio	-	-	-	-	-	-	-	-	-	-
	Matcha	0.924	0.607	0.733	-	-	0.982	0.601	0.746	-	-
	Matcha-DL	0.999	0.593	0.744	0.612	0.597	0.903	0.872	0.888	0.931	0.91
	OLaLa	-	-	-	-	-	-	-	-	-	-
	SORBETMtch	-	-	-	-	-	0.876	0.604	0.715	-	-
SNOMED-NCIT (Neoplas)	AMD	0.792	0.528	0.633	-	-	0.792	0.528	0.633	-	-
	ATMatcher	0.835	0.286	0.426	-	-	-	-	-	-	-
	BERTMap	0.575	0.784	0.664	0.965	0.947	0.562	0.771	0.650	0.962	0.93
	BERTMapLt	0.775	0.713	0.743	0.900	0.876	0.775	0.688	0.729	0.891	0.85
	LSMatch	0.877	0.238	0.374	-	-	0.823	0.583	0.683	-	-
	LogMap	0.783	0.547	0.644	0.821	0.743	0.675	0.793	0.729	-	-
	LogMapLt	0.932	0.519	0.667	-	-	0.931	0.514	0.662	-	-
	LogMapBio	-	-	-	-	-	-	-	-	-	-
	Matcha	0.704	0.564	0.626	-	-	0.782	0.545	0.642	-	-
	Matcha-DL	0.956	0.615	0.748	0.654	0.640	0.806	0.714	0.757	0.829	0.80
	OLaLa	-	-	-	-	-	0.451	0.545	0.493	-	-
	SORBETMtch	-	-	-	-	-	0.731	0.605	0.662	-	-
			,	Table 3							
	Overview of	OAEI's 2	022 / 202	3 Bio-ML	track eva	aluation re	sults (2/2)				
– FMA is a domain of	ontology that re	epresent	ts a coh	erent b	ody of	explicit	declarat	ive kno	wledge	about l	numai

pruning algorithm proposed in the Bio-ML paper [7], the five ontologies were enhanced with locality modules, increasing their similarity to their original versions through augmentation of their structural and logical contexts.

4.1.1. Hyper-parameters

In both editions of the Bio-ML track Matcha-DL was run with the following parameters:

- Negative Samples (n), 99: defines the number of negative pairings to be generated for each positive pairing in the reference alignment. It was optimized by following Bio-ML's ratio for local ranking: one positive pair in a set of 100 candidates (1:99).
- Feature Cardinality, 10: used to define the number of candidate features to generate for each entity in the source ontology.
- Feature Threshold, 0.1: defines the threshold applied on the matching algorithms.
- - Filter Threshold, 0.7: defines the filter to be applied on the confidence for the final alignment.

4.2. Results

The results of OAEI's 2022 and 2023 Bio-ML tracks are summarized in Tables 2 and 3, respectively. In both editions, Matcha-DL achieves an improved performance compared to Matcha, with F1 scores improving between 3 and 25 percentual points. This clearly indicates that the ability to learn a suitable OM algorithm combination using a partial reference alignment is actually an improvement over handcrafted rules and ontology profiling.

In 2022, Matcha-DL achieved the top F1 score in four out of five tasks, except in SNOMED-NCIT (Pharm), where AMD, a non ML-based system, was the top-ranked system. In 2023, Matcha-DL also achieved the top F1 score in four out of five tasks, being surpassed by SORBETMatch in NCIT-DOID (Disease). Moreover, Matcha-DL also improved on the F1 scores obtained in 2022 for all tasks. The additional context added to the 2023 task was clearly beneficial for Matcha-DL. For the other ML-based systems, this benefit was not as clear. AMD maintained the same performance, whereas BERTMap and BERTMapLt had a more heterogeneous performance, with BERTMap decreasing F1 score in the more contextually rich scenario of 2023 for OMIM-ORDO (Disease), SNOMED-NCIT (Pharm) and SNOMED-NCIT (Neoplas).

On the other hand, BERTMap achieved the top MRR and Hits@1 in both years. This discrepancy between which is the top-ranked system considering the two sets of metrics employed – precision, recall, F1 score and MRR and Hits@1 – is likely a reflection of the focus of BERTMap on local ranking, whereas the algorithms employed by Matcha-DL were designed for the more typical OM paradigm of global matching. Notably, neither AMD nor SOBERTMatch present performance using MRR and Hits@1. However, Matcha-DL increased its MRR and Hits@1 substantially from the 2022 to the 2023 scenario, with increases between 6 and 31.9% for MRR and 4.4 and 32% in Hits@1.

Finally, despite the promise of using LLMs to improve OM performance, Matcha-DL's superior performance in terms of F1 score and competitiveness in terms of MRR compared to both BERTMap and OLaLa is a clear indicator that this promise is yet to deliver consistent improvements. However, there is a clear case for the use of ML to solve the challenge of finding a suitable algorithm combination for specific OM tasks.

5. Conclusions

OM is a crucial task for knowledge representation with a downstream impact on tasks where ontology and data integration are key. For the last two decades, the majority of OM systems developed worked under the assumption that no partial alignment was available to inform the system's workings. Recently, the importance of machine learning for OM has become increasingly evident, prompting the development of novel approaches and of novel benchmarks for evaluating them.

We presented Matcha-DL, an ML-based approach for ontology alignment that combines the advantages of the classical algorithms for OM with the flexibility afforded by learning how to combine them through supervised learning. Matcha-DL was evaluated under the Bio-ML track at the OAEI, the only benchmark for ML-based OM. Overall, it achieved the best performance in terms of F1 score and was competitive in terms of MRR. Matcha-DL was further shown to benefit more than other competitors from the contextual information of ontologies.

Acknowledgements

This work was supported by FCT through the fellowships 2022.10557.BD (Pedro Cotovio) and 2022.11895.BD (Marta Silva), and through the LASIGE Research Unit, ref. UIDB/00408/2020 (https://doi.org/10.54499/UIDB/00408/2020) and ref. UIDP/00408/2020 (https://doi.org/10.54499/UIDP/00408/2020). It was also partially supported by the KATY project which has received funding from the European Union's Horizon 2020 research and innovation pro-gram under grant agreement No 101017453. This work was also supported partially by project 41, HfPT: Health from Portugal, funded by the Portuguese Plano de Recuperação e Resiliência.

References

- [1] J.D. Ferreira, D.C. Teixeira and C. Pesquita, Biomedical ontologies: coverage, access and use, in: Systems Medicine, Elsevier, 2021.
- [2] J. Euzenat and P. Shvaiko, Ontology matching, 2nd edn, Springer-Verlag, Heidelberg (DE), 2013.
- [3] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I.F. Cruz and F.M. Couto, The AgreementMakerLight Ontology Matching System, in: OTM Conferences - ODBASE, 2013, pp. 527–541.
- [4] E. Jiménez-Ruiz and B. Cuenca Grau, Logmap: Logic-based and scalable ontology matching, in: The Semantic Web–ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I 10, Springer, 2011, pp. 273–288.
- [5] D. Ngo and Z. Bellahsene, Overview of YAM++--(not) Yet Another Matcher for ontology alignment task, *Journal of Web Semantics* **41** (2016), 30–49.
- [6] M.A.N. Pour, A. Algergawy, P. Buche, L.J. Castro, J. Chen, H. Dong, O. Fallatah, D. Faria, I. Fundulaki, S. Hertling et al., Results of the ontology alignment evaluation initiative 2022, CEUR Workshop Proceedings, 2023.
- [7] Y. He, J. Chen, H. Dong, E. Jiménez-Ruiz, A. Hadian and I. Horrocks, Machine Learning-Friendly Biomedical Datasets for Equivalence and Subsumption Ontology Matching (2022).
- [8] D. Faria, M. Silva, P. Cotovio, L. Ferraz, L. Balbi and C. Pesquita, Matcha and Matcha-DL results for OAEI 2023, in: *Proceedings of the 18th International Workshop on Ontology Matching.*, 2023.
- [9] P. Kolyvakis, A. Kalousis, B. Smith and D. Kiritsis, Biomedical ontology alignment: an approach based on representation learning, *Journal of biomedical semantics* **9** (2018), 1–20.
- [10] J. Hao, C. Lei, V. Efthymiou, A. Quamar, F. Özcan, Y. Sun and W. Wang, MEDTO: Medical Data to Ontology Matching Using Hybrid Graph Neural Networks, in: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 2946–2954–. ISBN 9781450383325. doi:10.1145/3447548.3467138.
- [11] J. Chen, E. Jiménez-Ruiz, I. Horrocks, D. Antonyrajah, A. Hadian and J. Lee, Augmenting ontology alignment by semantic embedding and distant supervision, in: *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18,* Springer, 2021, pp. 392–408.
- [12] Y. He, J. Chen, D. Antonyrajah and I. Horrocks, BERTMap: a BERT-based ontology alignment system, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, 2022, pp. 5684–5691.
- [13] F. Gosselin and A. Zouaq, SORBETMatcher Results for OAEI 2023 (2023).
- [14] S. Hertling and H. Paulheim, Olala: Ontology matching with large language models, in: *Proceedings of the 12th Knowledge Capture Conference 2023*, 2023, pp. 131–139.
- [15] Z. Wang, AMD Results for OAEI 2022 (2022).
- [16] J.S. Amberger, C.A. Bocchini, F. Schiettecatte, A.F. Scott and A. Hamosh, OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders, *Nucleic acids research* 43(Database issue) (2015), D789—98–. doi:10.1093/nar/gku1205. https://europepmc.org/articles/PMC4383985.
- [17] D. Vasant, L. Chanas, J. Malone, M. Hanauer, A. Olry, S. Jupp, P. Robinson, H. Parkinson and A. Rath, ORDO: An Ontology Connecting Rare Disease, Epidemiology and Genetic Data, 2014.
- [18] N. Sioutos, S. de Coronado, M.W. Haber, F.W. Hartel, W.-L. Shaiu and L.W. Wright, NCI Thesaurus: a semantic model integrating cancerrelated clinical and molecular information, *Journal of biomedical informatics* 40(1) (2007), 30–43–. doi:10.1016/j.jbi.2006.02.013.
- [19] L.M. Schriml, E. Mitraka, J. Munro, B. Tauber, M. Schor, L. Nickle, V. Felix, L. Jeng, C. Bearer, R. Lichenstein, K. Bisordi, N. Campion, B. Hyman, D. Kurland, C.P. Oates, S. Kibbey, P. Sreekumar, C. Le, M. Giglio and C. Greene, Human Disease Ontology 2018 update: classification, content and workflow expansion, *Nucleic acids research* 47(D1) (2019), D955—D962–. doi:10.1093/nar/gky1032. https: //europepmc.org/articles/PMC6323977.
- [20] Foundational Model of Anatomy | Structural Informatics Group. http://si.washington.edu/projects/fma.
- [21] D. Faria, C. Pesquita, E. Santos, I.F. Cruz and F.M. Couto, Automatic Background Knowledge Selection for Matching Biomedical Ontologies, *PLoS One* 9(11) (2014), e111226.
- [22] I. Megdiche, O. Teste and C. Trojahn, An extensible linear approach for holistic ontology matching, in: *International Semantic Web Conference*, Springer, 2016, pp. 393–410.
- [23] É. Thiéblin, O. Haemmerlé, N. Hernandez and C. Trojahn, Survey on complex ontology matching, *Semantic Web* 11(4) (2020), 689–727.
 doi:10.3233/SW-190366.
- ⁴¹ [24] B.F.d. Lima, Breaking rules: taking Complex Ontology Alignment beyond rulebased approaches, PhD thesis, 2021.
- 42 [25] N. Reimers and I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, *arXiv preprint arXiv:1908.10084* (2019).