Cross-Lingual Ontology Matching using Structural and Semantic Similarity

Shubhanker Banerjee [1], Bharathi Raja Chakravarthi [2], John Philip McCrae [1]
[1] ADAPT Centre, University of Galway
[2] School of Computer Science, University of Galway
shubhanker.banerjee@adaptcentre.ie

Abstract

The development of ontologies in various languages is attracting attention as the amount of multilingual data available on the web increases. Cross-lingual ontology matching facilitates interoperability amongst ontologies in different languages. Although supervised machine learning-based methods have shown good performance on ontology matching, their application to the cross-lingual setting is limited by the availability of training data. Current state-of-the-art unsupervised methods for cross-lingual ontology matching focus on lexical similarity between entities. These approaches follow a two-stage pipeline where the entities are translated into a common language using a translation service in the first step followed by computation of lexical similarity between the translations to match the entities in the second step. In this paper, we introduce a novel ontology matching method based on the fusion of structural similarity and cross-lingual semantic similarity. We carry out experiments using 3 language pairs and report substantial improvements in the performance of the lexical methods thus showing the effectiveness of our proposed approach. To the best of our knowledge, this is the first work that tackles the problem of unsupervised ontology matching in the cross-lingual setting by leveraging both structural and semantic embeddings.

Keywords: cross-lingual ontology matching, cross-lingual semantic similarity, lexical similarity

1. Introduction

An increasing amount of multilingual data on the web has led to the development of ontologies in different languages. Ontologies are used to enable the sharing of information across different systems (Davies et al., 2002; Beydoun et al., 2011; Elmhadhbi et al., 2021). Furthermore, the adoption of ontologies as databases across domains has also attracted attention (Pankowski, 2023). These applications motivate the development of tools that allow semantic interoperability of ontologies across a wide range of languages. Identifying correspondences between ontologies in different languages is called Cross-lingual Ontology Matching (CLOM) (Ibrahim et al., 2023). Cross-Lingual Ontology Matching has the potential to contribute to various areas such as ontology enrichment, peer-to-peer information sharing, and linked data. Despite these potential applications CLOM has largely been an unexplored research problem. Therefore, more efforts from the research community towards building flexible CLOM systems are needed.

In recent times, deep learning based methods have achieved good results on ontology matching (Iyer et al., 2020; Li et al., 2019b; He et al., 2022). However, these methods are dependent on large amounts of training data which are not available in cross-lingual scenarios. To tackle this challenge we present an unsupervised ontology matching approach for CLOM. The proposed approach uses a state-of-the-art text embedding model to embed the concept descriptions into low-dimensional vectors which are then used to compute semantic similarity. Structural similarity between source and target concepts is an integral part of the proposed approach. We leverage the semantic similarity between source and target concepts to generate reference alignments. These reference alignments are used to learn structural embeddings for each concept in source and target ontologies. The semantic and structural embeddings are then used to calculate a weighted similarity to find equivalent entities in two ontologies. The main contributions of this work can be summarized as follows:

- Our experiments reveal that a weighted combination of semantic and structural similarity achieves performance gains over lexical similarity measures.
- We evaluate our method on 3 language pairs to demonstrate its extensibility.
- The proposed approach does not require manually labeled alignment data and thus is suitable for application in data-scarce scenarios.

The paper is organized as follows: Section 2 discusses the related works, Section 3 describes the methodology, the experiments are described in Section 4, the results are discussed in Section 5. The conclusion is given in Section 6. The limitations have been discussed in Section 7.

LDL: Linked Data in Linguistics @ LREC-COLING 2024, pages 11–21

2. Related Work

Cross-lingual Ontology Matching. Traditionally, CLOM approaches involve translation of the concepts into a common language (usually English) followed by calculation of lexical similarity to identify equivalent concepts. Following this paradigm Fu et al. (2010) propose a CLOM approach that selects the appropriate translation from amongst multiple translations generated by their system based on synonym-based matching with the entities in the target ontology. Furthermore, to resolve conflicts in alignments their system relies on the similarity of 1-hop neighbours of the entities from source and target ontologies. The translation in Ibrahim et al. (2019) follows a similar approach where they select candidate translations based on similarity to target concepts. Their system outperforms the state-of-the-art systems on the Ontology Alignment Evaluation Initiative (OAEI)[1] 2018 benchmark. Ibrahim et al. (2020) introduced MULON, a modularized CLOM system based on lexical and semantic similarity. The alignments are computed using a combination of both similarities. They use Jaccard for lexical similarity and WordNet path-based matching for semantic similarity.

MoMatch (Ibrahim et al., 2023) is based on lexical similarity of translated entities computed using metrics such as Jaccard (Jaccard, 1901), Levenshtein (Levenshtein et al., 1966), Jaro (Jaro, 1989) and Jaro-Winkler (Wang et al., 2017). The translation is carried out using the Yandex translation API and they improve upon the performance of the state-of-the-art methods for CLOM from the OAEI 2020 benchmark. Sharma and Jain (2023) achieve the best results on the MultiFarm dataset (Meilicke et al., 2012) at OAEI 2023. Their method uses Levenshtein-based similarity of translated concepts and WordNet-based synonym matching to align concepts. Machine learning based methods have also been explored for CLOM; Spohr et al. (2011) use a small amount of manually aligned concepts to train a SVM with 20 string-based features and 22 structural features for CLOM. Gracia and Asooja (2013) leverage artificial neural networks (ANNs) to calculate similarity between source and target concepts using manually designed features.

Unsupervised Entity Alignment. Ontologies are graph structures that describe hierarchies between concepts within a domain (Zhapa-Camacho and Hoehndorf, 2023). Therefore, ontology matching is fundamentally similar to the task of entity alignment across knowledge graphs. Unsupervised and self-supervised methods have been proposed for aligning entities in data-scarce scenarios. Liu et al. (2022a) propose a self-supervised training objective based on contrastive learning for entity alignment. To generate reference training alignments they use semantic similarity between concept descriptions from the source and target ontologies. The descriptions are encoded using LaBSE (Feng et al., 2022) and graph attention network (Velickovic et al., 2018) is used to learn structural embeddings using a self-supervised training objective based on noise-contrastive estimation (Gutmann and Hyvärinen, 2010). Tang et al. (2023) pose entity alignment as an optimal transport problem and report good results. In particular, they calculate fused Gromov-Wasserstein distance (Vayer et al., 2019) to minimize the distance between entities. Mao et al. (2021) formulate the ontology matching problem as a minimum sum assignment problem. The optimal assignments are calculated using the Hungarian (Kuhn, 1955) and Sinkhorn algorithms (Sinkhorn, 1964). Graph convolutional networks (GCN) (Kipf and Welling, 2017) have also been used to capture structural information. Zeng et al. (2021) use GCN to compute structural similarity between concept nodes in source and target ontologies. Textual similarity is computed using a weighted combination of Levenshtein similarity and cosine similarity of averaged word vectors. A weighted combination of structural and textual similarity is compared against a fixed threshold to align entities.

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

3. Methodology

We propose a framework for unsupervised cross-lingual ontology alignment. As discussed in Section 2, approaches based on lexical similarity have achieved state-of-the-art (SOTA) results on CLOM tasks. To demonstrate the effectiveness of our approach we compare it against 5 lexical similarity measures.

3.1. Task Formulation

The source and the target ontologies O_1 and O_2 respectively are inputs to the proposed CLOM system. The task of cross-lingual ontology matching is defined as finding aligned concepts between the ontologies. i.e.,

$$\phi = \{(a, b)|a \in C_1,\ b \in C_2,\ a \leftrightarrow b\},$$

where C_1 and C_2 refer to the concept sets in O_1 and O_2, respectively, $a \leftrightarrow b$ represent alignment between source and target concepts i.e., a and b refer to the same object in the real world. In this paper, we focus on unsupervised cross-lingual ontology matching i.e., source and target concepts belong to different languages and there is no labeled alignment data available.

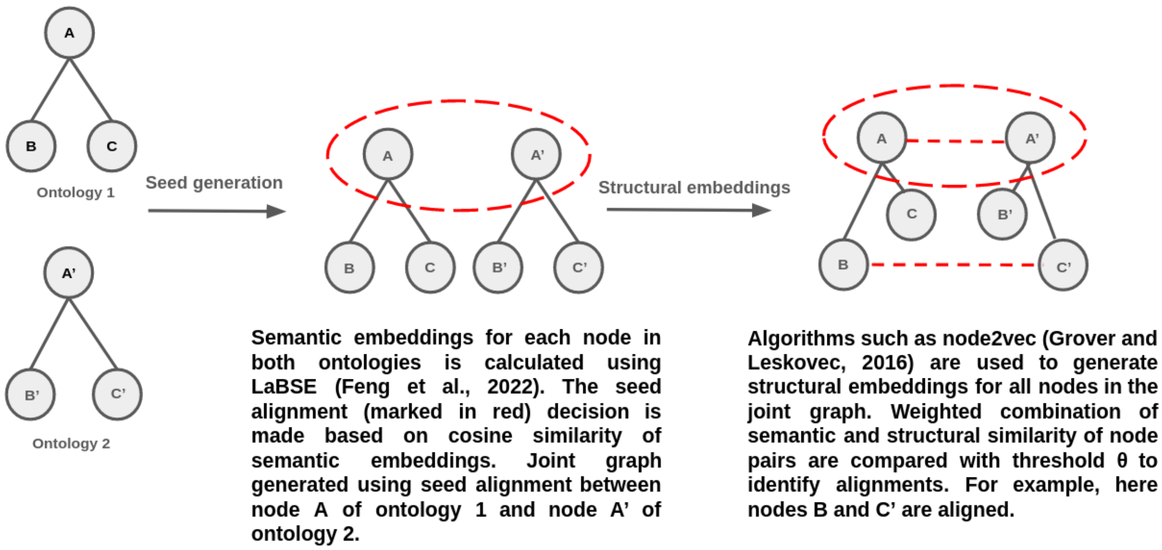

Figure 1: The source and target ontologies are inputs to our proposed CLOM framework in which we leverage both semantic and structural similarity of concepts to align the candidate nodes.

3.2. Concept Alignment

We hypothesize that aligned concepts in the source and target ontologies would have similar textual descriptions. This hypothesis postulates that the cosine similarity of text embeddings obtained from concept descriptions is positively correlated with the likelihood of the concepts being aligned/matched. Similar ideas have been explored by various supervised, semi-supervised, and self-supervised knowledge graph entity alignment approaches (Liu et al., 2022b; Wu et al., 2019; Chen et al., 2017; Tang et al., 2020). Here we leverage LaBSE (Feng et al., 2022), a multilingual model pre-trained on 109 languages for generating cross-lingual embeddings for the concepts in the source and target ontologies. Cosine similarity between the normalized embeddings is computed as a measure of semantic similarity between the corresponding concepts. Semantic similarity in the multilingual space is then leveraged for generating seed alignments between the input ontologies.

Ontologies are fundamentally graphs that represent concept hierarchies within a domain. In addition to textual descriptions, structural embeddings of the concepts in question can also constitute an important factor in determining alignment. Concept nodes with similar neighbourhoods are more likely to be aligned. We carry out experiments with various graph embedding approaches such as node2vec (Grover and Leskovec, 2016), Graph Convolutional Networks (Kipf and Welling, 2017, GCN), RGCN (Schlichtkrull et al., 2018) and TransE (Bordes et al., 2013) to learn embeddings for concept nodes. However, comparisons using embeddings learned on the two input ontologies independently are not meaningful as the embeddings would reside in two different vector spaces. Therefore, we leverage the seed textual alignments to consider source and target ontologies together as a graph and learn structural embeddings for all concept nodes in both ontologies. We employ two strategies for seed alignment for this task. In the first strategy, we select only those concept node pairs as alignments where the source and target concept node descriptions are semantically mutual nearest neighbours of each other. In the second strategy, we calculate the semantic similarity scores of all source and target concept pairs. The top-k most similar concept pairs are selected as seed alignments. We experiment with k=1,3,5,7 to quantify variation in performance as the number of seed alignments changes. To generate structural embeddings we train the node2vec model using the self-supervised loss defined by Grover and Leskovec (2016). RGCN and TransE are trained using a margin ranking loss (MR) based on negative sampling[2]. The seed alignments are used as training data for training the GCN model using the training objective given in Equation 1

$$\mathcal{L} = \sum_{(a,b)\in S} \sum_{(a',b')\in S'} [d(a,b) + \gamma - d(a',b')]_+ \quad (1)$$

where $[\cdot]^+ = \max\{0, \cdot\}$ and (a, b) denotes a labeled concept pair from the training data. The set S' (a', b') represents negative concept pairs obtained by corrupting (a, b) using nearest neighbor sampling (Li et al., 2019a). The embeddings of the source and target concepts learned by GCN are denoted as a and b, respectively. The distance func-

[2] https://pykeen.readthedocs.io/en/stable/reference/training.html

tion measuring the distance between two embeddings is represented by $d(\cdot, \cdot)$. The hyper-parameter γ serves to separate positive samples from negative ones. Structural similarity between concept nodes is calculated using the cosine similarity of normalized structural embeddings generated by the graph embedding models.

Algorithm 1: The proposed algorithm combining semantic and structural similarity for ontology matching

Data: Source Ontology, O_1, Target Ontology, O_2
Result: Aligned node pairs $\hat{\phi}$

1 **Strategy 1:** Select seed set S_1 by choosing concept node pairs (c_1, c_2) where $c_1 \in O_1$ and $c_2 \in O_2$ and descriptions of $c1$ and $c2$ are semantically mutual nearest neighbors of each other;
2 **Strategy 2:** Calculate semantic similarity scores for all source and target concept pairs (c_1, c_2) where $c_1 \in C_1$ and $c_2 \in C_2$;
3 Select the top-k most similar concept pairs as seed set S for experiments with $k = 1, 3, 5, 7$;
4 Construct joint graph G_{joint} by combining O_1 and O_2 using S as reference alignments between the graphs;
5 Learn structural embeddings for concept nodes in G_{joint} using one of the methods defined in Section 3.2 ;
6 The combination of structural and semantic similarities $Sim_{Combined}$ is calculated using the Equation 2 as a measure of their alignment;
7 Output the aligned concept pairs according to Equation 3;

	OP	DP	Concept Classes
cmt	49	10	30
confOf	13	23	39
sigkdd	17	11	50
conference	46	18	61

Table 1: Dataset statistics: The number of Object properties (OP), Data Properties (DP), and Concept classes in each ontology. For our experiments, only concept classes are considered.

Finally, as discussed above both structural and semantic similarity are positively correlated with the likelihood of alignment. Therefore, we use a weighted combination of both these measures to assign a final similarity score to a pair of concepts from the source and target ontologies as shown in Equation 2.

$$Sim_{Combined} = \alpha \cdot Sim_{str} + (1 - \alpha) \cdot Sim_{sem} \quad (2)$$

where Sim_{str} is the structural similarity between concept nodes calculated using cosine similarity of normalized structural embeddings, Sim_{sem} is the semantic similarity of source and target concept node description calculated using the embeddings output by LaBSE. The concept pairs where $Sim_{Combined}$ is greater than a fixed threshold θ are considered to be aligned.

$$\hat{\phi} = \{(c_1, c_2) \mid c_1 \in C_1,\ c_2 \in C_2,\ Sim_{combined}(c_1, c_2) > \theta\} \quad (3)$$

where the $\hat{\phi}$ is the set of all aligned concept pairs $(c1, c2)$ where $C1$ is the set of all concepts in source ontology and $C2$ is the set of all concepts in target ontology and θ is the fixed threshold. The algorithm for the aligning source and target concept nodes has been described in Algorithm 1.

4. Experiments

4.1. Dataset

We carry out experiments on 3 ontology pairs (cmt-confOf, conference-confOf, and conference-sigkdd) across 3 language pairs (German-English, German-French, and English-French) of the Multifarm dataset (Meilicke et al., 2012). The MultiFarm dataset is a benchmark for multilingual ontology matching. It is used to evaluate the ability of systems to deal with ontologies in different languages. It consists of a set of 7 ontologies related to conferences. The dataset was derived by translating the OntoFarm dataset (Zamazal and Svátek, 2017) into 9 languages: Chinese, Czech, Dutch, French, German, Portuguese, Russian, Arabic and Spanish. The dataset statistics are given in Table 1.

4.2. Baselines

As discussed in Section 2, lexical string similarity measures constitute the core part of most state-of-the-art CLOM systems. Therefore, to evaluate the proposed approach we compare it to 5 lexical similarity measure commonly used in the literature, namely: Jaccard (Jaccard, 1901), Levenshtein (Levenshtein et al., 1966), Jaro (Jaro, 1989), Jaro-Winkler (Wang et al., 2017) and Tversky (Tversky, 1977). Since the baselines compute lexical similarity, we translate the source and target entities to English before using these methods. In our experiments, we have used MetaAI's state-of-the-art NLLB model (Costa-jussà et al., 2022) to translate the source and target concepts. In particular, we use a distilled 600M parameter version of the model

	γ	**Epochs**	**Learning rate**	**Walk length**	**# of walks**	**Batch size**
GCN	3.0	1000	1e-5	_	_	1
node2vec	_	_	_	30	200	_
RGCN	3.0	100	_	_	_	2
TransE	3.0	100	_	_	_	2

Table 2: The hyperparameters used during training. The value of γ has been chosen based on prior work on knowledge graph entity alignment by Zeng et al. (2021). The default learning rate scheduler in pyKEEN is used for training RGCN and TransE. We set the other hyperparameters through empirical trial and error.

nllb-200-distilled-600M to limit the computational resources needed for inference.

4.3. Experimental Setup

As discussed above, to establish the effectiveness of our approach we carry out experiments on 3 ontology pairs across 3 languages. In the first step semantic similarity between source and target concepts is calculated to establish seed alignments between the ontologies. As discussed in Section 3.2 experiments are carried out with node2vec, GCN, RGCN, and TransE for the structural embeddings. The hyperparameters used during training are listed in Table 2. Furthermore, for our experiments, we empirically set the similarity threshold θ to 0.80. We carry out experiments with different values of α to ascertain the relative importance of both similarity measures for achieving good task performance.

4.4. Implementation Details

To ensure reproducibility we have used open-source libraries in our implementation. The ontologies were pre-processed using RDFlib[3]. We used Hugging Face[4] to implement the translation pipeline for the baseline methods and calculate semantic similarity[5] between the source and target concepts. GCN was implemented using Torch Geometic[6]. The node2vec algorithm was implemented using node2vec library[7]. TransE and RGCN were implemented using PyKEEN library[8].

5. Results

Our main experimental results can be found in Table 3. It is important to note that semantic similarity using embeddings from LaBSE outperforms lexical similarity baselines in almost all cases on the F1-score, often by large margins in the range of approximately 1-40%. On the conference-sigkdd dataset node2vec-NN (NN implies node2vec with mutual nearest neighbour seed alignment strategy) has the best performance and achieves an average F1-score of 61.5% over all the language pairs. Similarly on the conference-confOf ontology pair node2vec-NN has the best performance on German-English and German-French datasets. However, on the English-French dataset, Jaro similarity outperforms all other methods. We also note that the TransE-NN based alignment approach outperforms the lexical methods in most cases but substantially lags behind node2vec-NN in all cases. The other two graph embedding methods namely GCN-NN and RGCN-NN have relatively bad performance and are outperformed by the lexical baselines in most cases. These observations indicate that using semantic similarity is a better alternative than lexical similarity for ontology matching. This result is not surprising as the semantic similarity is based on similarity of "meaning" whereas lexical similarity is based on overlap of surface forms and is dependent on the translations. Furthermore, the good performance of node2vec-NN also establishes the effectiveness of the proposed framework for ontology matching where we combine structural similarity with semantic similarity using a weighted combination. We attribute the relatively bad performance of GCN and RGCN models to the smaller size of the graph ($\approx$ 100 nodes in source and target ontologies combined) leading to ineffective learning of node representations.

The results reported in Table 3 use θ = 0.80. We recognize that fixed thresholds for alignment identification may lead to sub-optimal performance where a particular similarity threshold might not be optimal for all datasets. Higher thresholds might lead to a larger number of false negatives and a smaller threshold might lead to a larger number of false positives on different datasets. As demonstrated in Figure 2, these fluctuations might also have an impact on the overall performance.

Overall, the results indicate that incorporating

[3] https://rdflib.readthedocs.io/en/stable/

[4] https://huggingface.co/

[5] We used setu4993/LaBSE model from then hugging face repository to generate cross-lingual text embeddings

[6] https://pytorch-geometric.readthedocs.io/en/latest/

[7] https://pypi.org/project/node2vec/

[8] https://pykeen.readthedocs.io/en/stable/

cmt-confOf			
	German-French	**German-English**	**English-French**
	Precision/Recall/F1	Precision/Recall/F1	Precision/Recall/F1
Jaro	66.6/44.4/53.3	66.6/44.4/53.3	50.0/20.0/28.5
Jaro-Winkler	44.4/57.1/50.0	66.6/75.0/70.5	50.0/55.5/52.6
Levenshtein	100.0/40.0/57.1	100.0/50.0/66.6	100.0/30.0/46.1
Jaccard	80.0/44.4/57.1	**85.7/66.6/75.0**	75.0/33.3/46.1
Tversky	41.6/62.5/50.0	43.7/87.5/58.3	33.3/42.8/37.5
LaBSE	83.3/50.0/62.5	83.3/55.5/66.6	66.6/60.0/63.1
LaBSE + node2vec-NN	**100.0/50.0/66.6**	100.0/55.5/71.4	71.4/50.0/58.8
LaBSE + GCN-NN	71.4/55.5/62.5	62.5/55.5/58.8	**63.6/70.0/66.6**
LaBSE + TransE-NN	**100.0/50.0/66.6**	100.0/50.0/66.6	**100.0/50.0/66.6**
LaBSE + RGCN-NN	100.0/22.2/36.3	100.0/10.0/18.1	100.0/20.0/33.3
conference-confOf			
Jaro	33.3/33.3/33.3	40.0/44.4/42.1	**70.0/70.0/70.0**
Jaro-Winkler	25.0/50.0/33.3	27.7/62.5/38.4	50.0/80.0/61.5
Levenshtein	66.6/18.1/28.5	75.0/27.2/39.9	83.3/45.4/58.8
Jaccard	20.0/11.1/14.2	25.0/22.2/23.5	50.0/45.4/47.6
Tversky	7.1/33.3/11.7	9.3/75.0/16.6	22.2/66.6/33.3
LaBSE	66.6/54.5/60.0	60.0/54.5/57.1	60.0/54.5/57.1
LaBSE + node2vec-NN	**75.0/54.5/63.1**	**66.6/60.0/63.1**	66.6/54.4/60.0
LaBSE + GCN-NN	50.0/60.0/54.5	46.1/60.0/52.1	44.4/72.7/55.1
LaBSE + TransE-NN	75.0/27.2/39.9	83.3/45.4/58.8	85.7/54.5/66.6
LaBSE + RGCN-NN	80.0/36.3/50.0	75.0/36.3/50.0	100.0/27.2/39.9
conference-sigkdd			
Jaro	42.8/30.0/35.2	42.8/30.0/35.2	40.0/18.1/25.0
Jaro-Winkler	25.0/40.0/30.7	29.4/50.0/37.0	33.3/27.2/30.0
Levenshtein	75.0/25.0/37.5	75.0/25.0/37.5	50.0/8.3/14.2
Jaccard	27.2/27.2/27.2	20.0/30.0/24.0	28.5/20.0/23.5
Tversky	8.5/42.8/14.2	10.5/44.4/17.0	13.3/57.1/21.6
LaBSE	55.5/41.6/47.6	42.8/54.5/47.9	60.0/50.0/54.5
LaBSE + node2vec-NN	**66.6/50.0/57.1**	**70.0/63.6/66.6**	**58.3/63.6/60.8**
LaBSE + GCN-NN	36.8/58.3/45.1	43.7/63.6/51.8	38.8/70.0/50.0
LaBSE + TransE-NN	60.0/25.0/35.2	80.0/33.3/47.0	66.6/33.3/44.4
LaBSE + RGCN-NN	100.0/25.0/40.0	100.0/8.3/15.3	100/16.6/28.5

Table 3: Precision, recall, and F1-scores of 5 lexical baselines compared with node2vec-NN, GCN-NN, TransE-NN, and RGCN-NN (NN indicates that mutual nearest neighbour source and target concepts are used as seed alignments between the input ontologies. This seed generation strategy is described as Strategy 1 in Algorithm 1) for θ = 0.80 and α = 0.2.

structural information improves performance as compared to only using semantic similarity. However, the performance is sensitive to the choice of embedding methods used as node2vec substantially outperforms GCN. Furthermore, these results have been reported for α=0.2 which signifies a smaller contribution of structural similarity to the overall alignment. We discuss variation in performance of the node2vec-NN model with α in more detail in Section 5.2.

5.1. Performance vs. k

As discussed in Algorithm 1 we employ two seed generation strategies. In this section, we compare the task performance of node2vec using mutual nearest neighbour seed alignments (Strategy 1) and top-k most semantically similar seed alignments (Strategy 2). We fix α = 0.2 and θ = 0.80 for the experiments. The results are illustrated in Table 4. In general, k = 1 leads to bad performance. This is understandable as only 1 seed alignment between the graphs is insufficient to learn meaningful representations. As can be seen, the F1-scores

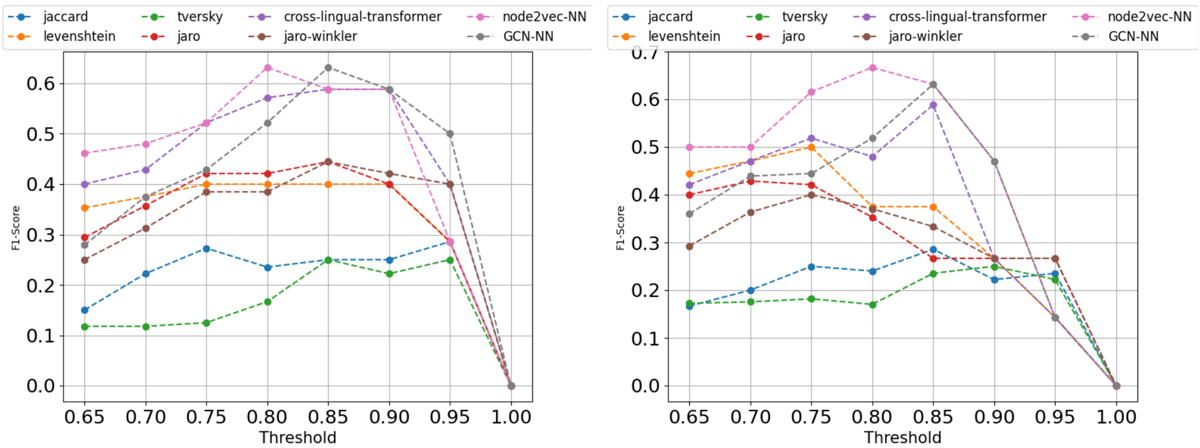

Figure 2: The variation in F1-score with change in threshold for the German-English dataset for conference-confOf pair (on the left) and the conference-sigkdd pair (on the right) with $\alpha = 0.2$.

		cmt-confOf	
k	**German-French**	**German-English**	**English-French**
1	36.3	46.1	**66.6**
3	46.1	66.6	53.3
5	57.1	66.6	62.5
7	57.1	66.6	62.5
NN	**66.6**	**71.4**	58.8
		conference-confOf	
1	28.5	58.8	28.5
3	58.8	58.8	39.9
5	55.5	**66.6**	47.0
7	55.5	63.1	44.4
NN	**63.1**	63.1	**60.0**
		conference-sigkdd	
1	26.6	47.0	47.0
3	52.6	55.5	52.6
5	**60.0**	**70.0**	50.0
7	57.1	60.0	**63.6**
NN	57.1	66.6	60.8

Table 4: F1-scores of top-k semantically similar seed alignments where k = 1,3,5,7 compared with mutual nearest neighbour (NN) alignments for node2vec model for θ = 0.80 and α = 0.2.

exhibit monotonic behaviour concerning the number of seed alignments in general i.e., increasing the number of alignments from 1 to 5 improves performance. However, in general k = 7 leads to degradation of performance as compared to k = 5. This can be attributed to additional noise introduced by a larger number of seed alignments. Hence, neither very low nor very high i.e., k = 5 is optimal for almost all datasets. In terms of the two strategies both are equally effective with nearest neighbour seed alignment outperforming k = 5 on 5 out of the 9 datasets.

5.2. Performance vs. α

To quantify variation in performance with changes in α we carry out experiments with varying α across different thresholds for node2vec with mutual nearest neighbour seed alignment. The results are illustrated in Figures 3, 4 and 5. As can be seen, almost all the ontology pairs and all the language pairs α = 0.2 had the best F1-score overall. Interestingly, as the value of alpha went up the performance deteriorated with the lowest F1-scores recorded for α = 0.8 for a given threshold. α = 0 has good performance and for specific thresholds outperforms F1-scores achieved by using α = 0.2. α = 0 indicates only the

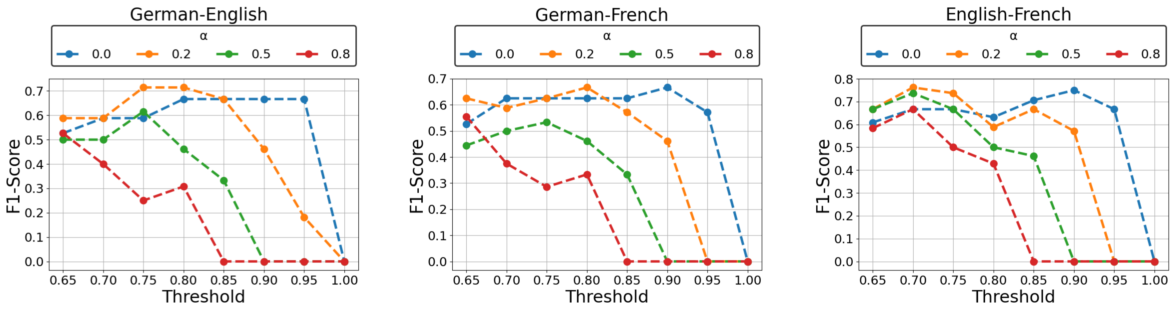

Figure 3: F1 vs. α: cmt-confOf

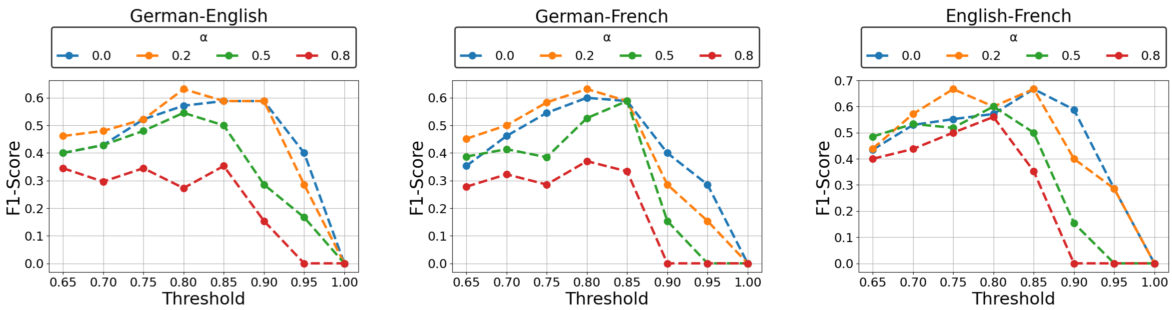

Figure 4: F1 vs. α: conference-confOf

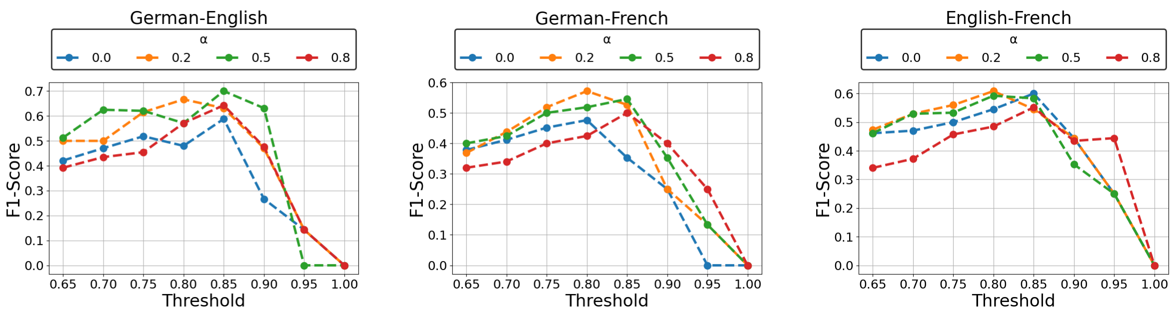

Figure 5: F1 vs. α: conference-sigkdd

semantic similarity of concept descriptions being used for ontology matching. These results suggest that while the choice is alpha is dependent on the similarity threshold being used, a value of 0.2 for α leads at threshold 0.80 leads to good results in general. Furthermore, the results demonstrate that while semantic similarity is the more important factor for ontology matching even outperforming the combined similarity for certain thresholds, the addition of structural similarity signals can lead to an improvement in task performance.

6. Conclusion

In this work, we proposed a new framework for ontology matching and evaluated it on 3 ontology pairs across 3 language pairs. The proposed framework takes into account semantic similarity between concept node descriptions in the source and target ontologies as well as the structural similarity calculated using embeddings that aggregate information about node neighbourhood structure. We showed that our proposed system can outperform current state-of-the-art lexical similarity measures being used for CLOM. Furthermore, the results show that semantic similarity of concept node descriptions is the more important factor when aligning source and target nodes. We experiment with four structural embeddings, namely node2vec, TransE, RGCN, and GCN, and find that node2vec leads to better performance. It is also important to note that the performance of Levenshtein similarity is better than our proposed framework for German-English and English-French datasets of the cmt-confOf and conference-confOf ontology pairs respectively. Semantic similarity is used to generate seed alignments in the first stage of our approach and we explore two strategies for this purpose. Our analysis suggests that selecting top-k semantically similar concepts as seed alignments leads to better performance.

7. Limitations

Although we have shown good performance of our method for ontology matching as compared to lexical measures there are limitations worth discussing. We carry out our experiments using a fixed threshold however as discussed in Section 5, there is substantial variation in performance with changing thresholds. Choosing a threshold is associated with a trade-off between precision and recall. Manually fixing a threshold for different datasets is not optimal. Furthermore, we show that GCN is substantially outperformed by node2vec; there are more advanced alternatives such as Graph attention networks which can allow the nodes to only aggregate useful signals from their neighbours. We expect there to be an improvement in performance by using these algorithms. We show that using top-k semantically similar concepts as seed alignments is a better strategy for seed generation overall. However, the experiments do not establish an optimal value of k for all datasets. We hope to develop better seed generation strategies as a part of future work.

8. Acknowledgement

Author Shubhanker Banerjee was supported by Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University Of Galway.

9. Bibliographical References

Ghassan Beydoun, Graham Low, Quynh-Nhu Numi Tran, and Paul Bogg. 2011. Development of a peer-to-peer information sharing system using ontologies. *Expert Syst. Appl.*, 38(8):9352–9364.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795.

Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 1511–1517. ijcai.org.

Marta R Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

John Davies, Alistair Duke, and Audrius Stonkus. 2002. Ontoshare: Using ontologies for knowledge sharing. In *Proceedings of the WWW2002 International Workshop on the Semantic Web, Hawaii, May 7, 2002*, volume 55 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Linda Elmhadhbi, Mohamed-Hedi Karray, Bernard Archimède, J. Neil Otte, and Barry Smith. 2021. An ontological approach to enhancing information sharing in disaster response. *Information*, 12(10):432.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 878–891. Association for Computational Linguistics.

Bo Fu, Rob Brennan, and Declan O'Sullivan. 2010. Cross-lingual ontology mapping and its use on the multilingual semantic web. In *Proceedings of the 1st International Workshop on the Multilingual Semantic Web, Raleigh, North Carolina, USA, April 27th, 2010*, volume 571 of *CEUR Workshop Proceedings*, pages 13–20. CEUR-WS.org.

Jorge Gracia and Kartik Asooja. 2013. Monolingual and cross-lingual ontology matching with CIDER-CL: evaluation report for OAEI 2013. In *Proceedings of the 8th International Conference on Ontology Matching-Volume 1111*, pages 109–116.

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864.

Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15, 2010*, volume 9 of *JMLR Proceedings*, pages 297–304. JMLR.org.

Yuan He, Jiaoyan Chen, Denvar Antonyrajah, and Ian Horrocks. 2022. Bertmap: A bert-based ontology alignment system. In *Thirty-Sixth AAAI*

Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022, pages 5684–5691. AAAI Press.

Shimaa Ibrahim, Said Fathalla, Jens Lehmann, and Hajira Jabeen. 2020. Multilingual ontology merging using cross-lingual matching. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 113–120. IEEE.

Shimaa Ibrahim, Said Fathalla, Jens Lehmann, and Hajira Jabeen. 2023. Toward the Multilingual Semantic Web: Multilingual Ontology Matching and Assessment. *IEEE Access*, 11:8581–8599.

Shimaa Ibrahim, Said Fathalla, Hamed Shariat Yazdi, Jens Lehmann, and Hajira Jabeen. 2019. From monolingual to multilingual ontologies: The role of cross-lingual ontology enrichment. In *Semantic Systems. The Power of AI and Knowledge Graphs: 15th International Conference, SEMANTiCS 2019, Karlsruhe, Germany, September 9–12, 2019, Proceedings 15*, pages 215–230. Springer International Publishing.

Vivek Iyer, Arvind Agarwal, and Harshit Kumar. 2020. Veealign: a supervised deep learning approach to ontology alignment. In *Proceedings of the 15th International Workshop on Ontology Matching co-located with the 19th International Semantic Web Conference (ISWC 2020), Virtual conference (originally planned to be in Athens, Greece), November 2, 2020*, volume 2788 of *CEUR Workshop Proceedings*, pages 216–224. CEUR-WS.org.

Paul Jaccard. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579.

Matthew A Jaro. 1989. Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association*, 84(406):414–420.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Harold W Kuhn. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97.

Vladimir I Levenshtein et al. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710. Soviet Union.

Chengjiang Li, Yixin Cao, Lei Hou, Jiaxin Shi, Juanzi Li, and Tat-Seng Chua. 2019a. Semi-supervised Entity Alignment via Joint Knowledge Embedding Model and Cross-graph Model. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2723–2732.

Weizhuo Li, Xuxiang Duan, Meng Wang, Xiaoping Zhang, and Guilin Qi. 2019b. Multi-view embedding for biomedical ontology matching. In *Proceedings of the 14th International Workshop on Ontology Matching co-located with the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, October 26, 2019*, volume 2536 of *CEUR Workshop Proceedings*, pages 13–24. CEUR-WS.org.

Xiao Liu, Haoyun Hong, Xinghao Wang, Zeyi Chen, Evgeny Kharlamov, Yuxiao Dong, and Jie Tang. 2022a. SelfKG: Self-Supervised Entity Alignment in Knowledge Graphs. In *Proceedings of the ACM Web Conference 2022*, WWW '22, pages 860–870, New York, NY, USA. Association for Computing Machinery.

Xiao Liu, Haoyun Hong, Xinghao Wang, Zeyi Chen, Evgeny Kharlamov, Yuxiao Dong, and Jie Tang. 2022b. Selfkg: Self-supervised entity alignment in knowledge graphs. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pages 860–870. ACM.

Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. 2021. From alignment to assignment: Frustratingly simple unsupervised entity alignment. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 2843–2853. Association for Computational Linguistics.

Christian Meilicke, Raúl García-Castro, Fred Freitas, Willem Robert van Hage, Elena Montiel-Ponsoda, Ryan Ribeiro de Azevedo, Heiner Stuckenschmidt, Ondřej Šváb Zamazal, Vojtěch Svátek, Andrei Tamilin, Cássia Trojahn, and Shenghui Wang. 2012. MultiFarm: A benchmark for multilingual ontology matching. *Journal of Web Semantics*, 15:62–68.

Tadeusz Pankowski. 2023. Ontological databases with faceted queries. *The VLDB Journal*, 32(1):103–121.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer.

Abhisek Sharma and Sarika Jain. 2023. Lsmatch and lsmatch-multilingual results for OAEI 2023. In *Proceedings of the 18th International Workshop on Ontology Matching co-located with the 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, November 7, 2023*, volume 3591 of *CEUR Workshop Proceedings*, pages 159–163. CEUR-WS.org.

Richard Sinkhorn. 1964. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics*, 35(2):876–879.

Dennis Spohr, Laura Hollink, and Philipp Cimiano. 2011. A machine learning approach to multilingual and cross-lingual ontology matching. In *The Semantic Web–ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I 10*, pages 665–680. Springer.

Jianheng Tang, Kangfei Zhao, and Jia Li. 2023. A fused gromov-wasserstein framework for unsupervised knowledge graph entity alignment. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 3320–3334. Association for Computational Linguistics.

Xiaobin Tang, Jing Zhang, Bo Chen, Yang Yang, Hong Chen, and Cuiping Li. 2020. BERT-INT: A bert-based interaction model for knowledge graph alignment. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 3174–3180. ijcai.org.

Amos Tversky. 1977. Features of similarity. *Psychological review*, 84(4):327.

Titouan Vayer, Nicolas Courty, Romain Tavenard, Laetitia Chapel, and Rémi Flamary. 2019. Optimal transport for structured data with application on graphs. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 6275–6284. PMLR.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.

Yaoshu Wang, Jianbin Qin, and Wei Wang. 2017. Efficient approximate entity matching using jaro-winkler distance. In *International conference on web information systems engineering*, pages 231–239. Springer.

Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, Rui Yan, and Dongyan Zhao. 2019. Relation-aware entity alignment for heterogeneous knowledge graphs. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5278–5284. ijcai.org.

Ondrej Zamazal and Vladimír Svátek. 2017. The Ten-Year OntoFarm and its Fertilization within the Onto-Sphere. In *Web Semantics: Science, Services and Agents on the World Wide Web*, volume 43, pages 46–53. Elsevier.

Weixin Zeng, Xiang Zhao, Jiuyang Tang, Xinyi Li, Minnan Luo, and Qinghua Zheng. 2021. Towards entity alignment in the open world: An unsupervised approach. In *Database Systems for Advanced Applications - 26th International Conference, DASFAA 2021, Taipei, Taiwan, April 11-14, 2021, Proceedings, Part I*, volume 12681 of *Lecture Notes in Computer Science*, pages 272–289. Springer.

Fernando Zhapa-Camacho and Robert Hoehndorf. 2023. From axioms over graphs to vectors, and back again: Evaluating the properties of graph-based ontology embeddings. In *Proceedings of the 17th International Workshop on Neural-Symbolic Learning and Reasoning, La Certosa di Pontignano, Siena, Italy, July 3-5, 2023*, volume 3432 of *CEUR Workshop Proceedings*, pages 85–102. CEUR-WS.org.