Embedding Models with Inverted-index and Co-occurrence Matrices for Ontology Subsumption Prediction

Yukihiro Shiraishi
shiraishi@sw.cei.uec.ac.jp
Department of Computer and Network Engineering,
Graduate School of Informatics and Engineering,
The University of Electro-Communications
Tokyo, Japan

Ken Kaneiwa
kaneiwa@uec.ac.jp
Department of Computer and Network Engineering,
Graduate School of Informatics and Engineering,
The University of Electro-Communications
Tokyo, Japan

ABSTRACT

Artificial intelligence techniques for automatically completing ontology deficits are important for alleviating maintenance efforts. Embedding models for knowledge graphs are effective in link prediction tasks, but these models cannot deal with richer representations, such as ontologies. Recently, embedding models that consider all logical structures and annotation axioms in an OWL ontology have been proposed. However, combining various features may negatively impact the concept subsumption task. This study proposes two novel OWL ontology embedding models: **In**verted-index **M**atrix **E**mbedding (InME) and **Co**-occurrence **M**atrix **E**mbedding (CoME). To capture meaningful features for predicting concept subsumptions, we focus on embedding for annotation axioms to model the similarity of entities in an ontology. These embeddings directly extract global and local information from annotation axioms instead of word-embedding models, such as Word2Vec and BERT. In the evaluation experiments, we applied our models to the concept subsumption task using GO, FoodOn, and HeLiS. We demonstrate that InME outperforms existing models for GO and FoodOn, and that CoME concatenated with OWL2Vec* outperforms existing models for HeLiS.

CCS CONCEPTS

• **Computing methodologies → Knowledge representation and reasoning**.

KEYWORDS

Ontology, Ontology embedding, Autoencoder, Word embedding, Ontology completion

ACM Reference Format:
Yukihiro Shiraishi and Ken Kaneiwa. 2023. Embedding Models with Inverted-index and Co-occurrence Matrices for Ontology Subsumption Prediction. In *Proceedings of 12th International Joint Conference On Knowledge Graphs (IJCKG 2023).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

IJCKG 2023, December 08–09, 2023, Tokyo, Japan

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Ontology is a data management technology that supports the sharing and reuse of formally represented knowledge [13]. Many ontologies have been adopted in various domains, such as the Semantic Web and bioinformatics. They are widely used to share knowledge, conduct semantic searches, and automatically analyze information.

It is crucial to automatically address ontology deficits because they are labor-intensive. Although ontology reasoners, such as HermiT [12] and ELK [15], can logically infer axioms from ontologies, it is impossible to infer new axioms that have no logical connections, owing to deficits. Semantic embedding techniques have recently been employed to infer new axioms without logical connections. There has been considerable research on representing entities and relations as low-dimensional vectors in knowledge graphs (KGs) [2, 4, 21, 25, 26, 28, 30, 31]. These models are effective in link prediction [4]; however, they cannot handle richer concept representations in description logic (DL) and ontology.

KG embeddings have inspired the development of DL and ontology embeddings. E2R [11] models logical structures by applying quantum logic to $\mathcal{ALC}$. In embeddings for description logic $\mathcal{EL}^{++}$ [1], logical structures, such as concept subsumption, are regarded as geometric operations [16, 20, 29]. Onto2Vec [23] treats logical axioms as text corpora for training Word2Vec [18, 19]. OPA2Vec [24] is an extension model of Onto2Vec that additionally considers annotation axioms. OWL2Vec* [6] converts logical axioms, annotation axioms, and RDF graph walks extracted from an OWL ontology into text corpora. BERTSubs [5] uses contextual word-embedding BERT [8] for OWL ontologies instead of non-contextual word-embedding Word2Vec. Most ontology embedding models can be applied to predict missing concept subsumptions in an ontology. These models extract many features, such as logical structures and annotation axioms, but a mixture of many features may have a negative impact on the concept subsumption task.

In this study, we propose a method of **In**verted-index **M**atrix **E**mbedding (InME) and **Co**-occurrence **M**atrix **E**mbedding (CoME), which extracts the context of words and the similarity between words in annotation axioms. To predict concept subsumptions, we focus on capturing meaningful features of annotation axioms to model the embedding of entities. InME extracts global information representing the relationship between an entity and annotation words for that entity. CoME extracts local information representing the relationships between words in an annotation axiom. These two embeddings do not use word-embedding models like Word2Vec and BERT. Instead, they compress a matrix of words in annotation axioms using an autoencoder [14]. In addition, we enhance the annotation words for these models by utilizing the annotations of properties. Furthermore, InME and CoME can be concatenated with existing embedding models, such as OWL2Vec*, to improve prediction accuracy.

We evaluate InME and CoME on the OWL ontologies, GO [7], FoodOn [9], and HeLiS [10] by performing the concept subsumption task. We demonstrate that InME outperforms existing models, such as OWL2Vec* for GO and FoodOn, and that CoME concatenated with OWL2Vec* outperforms existing models for HeLiS in mean reciprocal rank (MRR) for the concept subsumption task.

The contributions of this study are summarized as follows.

- InME and CoME extract global and local information from the annotation axioms, respectively, to capture similarities among entities.
- Annotations of properties and the embedding concatenation are utilized to enhance for InME and CoME.
- InME outperforms existing models for GO and FoodOn, while CoME concatenated with OWL2Vec* outperforms in HeLiS.

The remainder of this paper is organized as follows. In Section 2, we describe the concepts of OWL ontology, OWL2Vec*, and the autoencoder. In Section 3, we present InME and CoME, which extracts global and local information from the annotation axioms. Furthermore, we consider the utilization of property annotations and embedding concatenation. In Section 4, we evaluate the performance of InME and CoME on three ontologies for the concept subsumption task. Finally, we conclude the study and discuss future research in Section 5.

2 RELATED WORK

2.1 OWL Ontology

OWL ontologies [3] comprise $\sum = (\mathbb{C}, \mathbb{R}, \mathbb{I})$, where $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{I}$ are sets of class names (also known as concept names), property names (also known as role names) and individual names. Classes, properties, and individuals are represented by a unique identifier called internationalized resource identifier (IRI). Properties are classified into object properties, data properties, and annotation properties.

Complex classes are inductively defined by conjunction $C \sqcap D$, disjunction $C \sqcup D$, negation $\neg C$, existential restriction $\exists R.C$, and universal restriction $\forall R.C$, where $C, D \in \mathbb{C}$, $R, S \in \mathbb{R}$. An OWL ontology consists of a TBox and an ABox. The TBox is a set of axioms such as concept subsumption $C \sqsubseteq D$ and property subsumption $R \sqsubseteq S$. The ABox is a set of axioms, such as concept assertion $C(a)$ and role assertion $R(a, b)$, where $a, b \in \mathbb{I}$.

The target axioms on the concept subsumption task are the concept subsumption $C \sqsubseteq D$ and concept assertion $C(a)$. The concept subsumption represents the subclass relation between C and D, where $C, D \in \mathbb{C}$. The concept assertion represents that individual a is an instance of class C, where $a \in \mathbb{I}$, $C \in \mathbb{C}$. $C(a)$ can be transformed into the subsumption $\{a\} \sqsubseteq C$.

2.2 OWL2Vec*

OWL2Vec* [6] treats logical axioms, annotation axioms, and graph walks as text corpora to generate word embeddings by Word2Vec [18, 19]. Graph walks are extracted from an RDF graph projected from an OWL ontology.

OWL2Vec* extracts three corpora: structural document D_s, lexical document D_l, and combined document D_c using logical structures, graph structures, and annotation axioms. D_s includes logical axioms and random walks derived from the RDF graph. D_l includes the annotations of IRI and D_s, in which IRIs are replaced with IRI-annotation words. Words in annotation axioms are transformed into lowercase letters, and non-English characters are removed. D_c is a mixture of D_s and D_l. These three documents are arbitrarily combined as text corpora to train Word2Vec.

Let e be an entity ($\in \mathbb{C} \cup \mathbb{I}$) in D_s, D_l, and D_c and let w be a word excluding IRIs in all annotation axioms. OWL2Vec* provides the embeddings of classes and individuals in two different ways for the concept subsumption task. The first embedding $V_\mathrm{iri}(e)$ is defined as the embedding of the e's IRI in Word2Vec. The second embedding is defined as $V_\mathrm{word}(e)$, which is averaged by the word embeddings corresponding to words in $S_\mathrm{label}(e)$, where $S_\mathrm{label}(e)$ is the set of words in e's annotations, whose property is *rdfs:label*. $V_\mathrm{word}(e)$ is given by

$$V_\mathrm{word}(e) = \frac{1}{|S_\mathrm{label}(e)|} \sum_{w \in S_\mathrm{label}(e)} V_\mathrm{w2v}(w) \tag{1}$$

where $V_\mathrm{w2v}(w)$ is the embedding of word w in Word2Vec. If an entity has no annotation, words, such as *Lysine* extracted from the IRI *Lysine_1001*[1] are added as annotation words. In the OWL2Vec* experiment, this process is applied to HeLiS [6].

OWL2Vec* uses a binary classifier to predict concept subsumptions. Let $f : \mathbf{R}^{2n} \to \mathbf{R}$ be a binary classifier, such as Random Forest (RF), MLP, SVC, or Logistic Regression (LR), where n is the embedding dimension. The following probability p_true indicates the likelihood of a concept subsumption $e_1 \sqsubseteq e_2$ being valid.

$$p_\mathrm{true}(e_1, e_2) = f(V(e_1)||V(e_2)) \tag{2}$$

where $V \in \{V_\mathrm{iri}, V_\mathrm{word}\}$ and $||$ is the concatenation of two embeddings. OWL2Vec* [6] reported that the RF is the best classifier of RF, MLP, SVC, and LR. For each positive sample of concept subsumption $e_1 \sqsubseteq e_2$, negative samples are constructed by replacing e_2 with a randomly selected entity e_2'. Note that $(e_1 \sqsubseteq e_2') \notin S_\mathrm{sub} \cup S_\mathrm{infer}$, where S_sub is the set of concept subsumption axioms, and S_infer is the set of all concept subsumption axioms that can be logically inferred from S_sub. Each class $x \in \mathbb{C}$ is assigned a $p_\mathrm{true}(e_1, x)$ score using the trained classifier. Based on these scores, each class x is ranked in descending order.

2.3 Autoencoder

Autoencoder [14] is a neural network that learns weights by minimizing the discrepancy between the original and the reconstructed data. By extracting a low-dimensional middle layer from a trained neural network, high-dimensional data can be compressed into low-dimensional data. The middle layer $H \in \mathbf{R}^{d \times m}$ and reconstructed output $X' \in \mathbf{R}^{n \times m}$ are given by

$$H = \mathrm{ReLU}(W_\mathrm{in} X + B_\mathrm{in}) \tag{3}$$

$$X' = \sigma(W_\mathrm{out} H + B_\mathrm{out}) \tag{4}$$

[1] *http://www.fbk.eu/ontologies/virtualcoach#Lysine_100.*

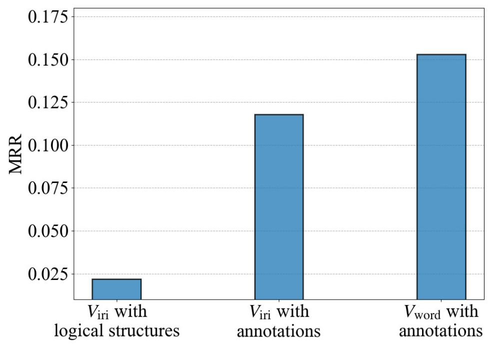


Figure 1: The comparison of logical structures and annotations on the concept subsumption task for GO using Word2Vec.

where $X \in \mathbf{R}^{n\times m}$ is the original input, $W_{\text{in}} \in \mathbf{R}^{d\times n}$ and $W_{\text{out}} \in \mathbf{R}^{n\times d}$ are the layer weights, $B_{\text{in}} \in \mathbf{R}^{d\times m}$ and $B_{\text{out}} \in \mathbf{R}^{n\times m}$ are the layer biases, and σ is the sigmoid function. The loss function to minimize the discrepancy between X and X' is given by

$$L(X, X') = -\sum_{j=1}^{m}(X_{ij}\log X'_{ij} + (1 - X_{ij})\log(1 - X'_{ij})) \tag{5}$$

3 METHODOLOGY

We define InME and CoME, which respectively extract global and local information from the annotation axioms. Similarity to OWL2Vec*, words in the annotation axioms are transformed into lowercase letters, and non-English characters are removed before extracting the information. The overall architecture is illustrated in Figure 2.

3.1 Logical Structures vs Annotations

Most ontology embedding models use both logical structures and annotation axioms as corpora. Then, we compare the contribution of logical structure axioms and annotation axioms for solving the concept subsumption task. We generate word embeddings by the Word2Vec model using corpora consisting either of logical structure axioms or annotation axioms. There are two methods to convert word embedding to entity embedding: V_{iri} and V_{word}, as described in Section 2.2. We only applied V_{iri} to the Word2Vec with logical structure axioms because they have no annotation words.

Figure 1 shows the performances of Word2Vec with logical structure axioms and with annotation axioms. The mean reciprocal rank (MRR) of annotation axioms outperforms that of logic structure axioms. In addition, the MRR of V_{word} with annotations outperforms V_{iri} with annotations. These results suggest that (i) utilizing annotation axioms and (ii) employing the embedding transformation method by averaging, V_{word}, contribute to the performance improvement.

3.2 Global and Local Information in Annotations

We consider global and local information in annotation axioms of the following Gene Ontology [7]:

GO:0021603	*rdfs:label*	"cranial nerve formation"
GO:0021611	*rdfs:label*	"facial nerve formation"
GO:0021620	*rdfs:label*	"hypoglossal nerve formation"

The common parent entity *GO:0021603* has child entities *GO:0021611* and *GO:0021620* as follows:

$$\mathit{GO{:}0021611} \sqsubseteq \mathit{GO{:}0021603}$$
$$\mathit{GO{:}0021620} \sqsubseteq \mathit{GO{:}0021603}$$

The annotation axioms of parent-child entities frequently exhibit resemblances. Conversely, the similarity of annotation axioms provides highly valuable information for predicting their relationships, regardless of the absence of logical connections. To capture the similarity of annotation axioms, the averaging method (Equation (1)) for transforming word embeddings into entity embeddings is effective. For example, *GO:0021603*'s embedding is generated by the average of the word embeddings of "cranial," "nerve," and "formation." Similarly, *GO:0021611* is the average of "facial," "nerve," and "formation," and *GO:0021620* is the average of "hypoglossal," "nerve," and "formation." These entity embeddings tend to be similar due to the shared words "nerve" and "formation." However, if the word embeddings of the non-shared words like "cranial," "facial," and "hypoglossal" are widely separated in the embedding space, the averaging process might prevent entity embeddings from getting close to each other. Therefore, it is important that the word embeddings are distributed close together.

We present to characterize the words "cranial," "facial," and "nerve" from the annotation axioms. As local information, these words co-occur as follows:[2]

"cranial" co-occurs with : cranial, nerve, formation
"facial" co-occurs with : facial, nerve, formation
"nerve" co-occurs with : cranial, facial, hypoglossal, nerve, formation

These words share similarities in terms of co-occurrence with the same words. In this example, "cranial" co-occurs with three words, matching the number of words in the annotation axiom of *GO:0021603*. The word "cranial" might also appear in different annotation axioms, hence local information frequently contains a variety of word information. This local information can also be extracted by Word2Vec with maximum window size.

Furthermore, as global information, the words "cranial," "facial," and "nerve" appear in entity's annotation axioms as follows:

"cranial" appears in : *GO:0021603*
"facial" appears in : *GO:0021611*
"nerve" appears in : *GO:0021603, GO:0021611, GO:0021620*

Words that appear in the annotation axiom of each entity share similar global information. In this example, "cranial" appears in *GO:0021603*'s annotation axiom, which is fewer compared to the

[2] Each of the words "cranial", "facial", and "nerve" is itself included in the co-occurrence words to make the features of each word more similar.

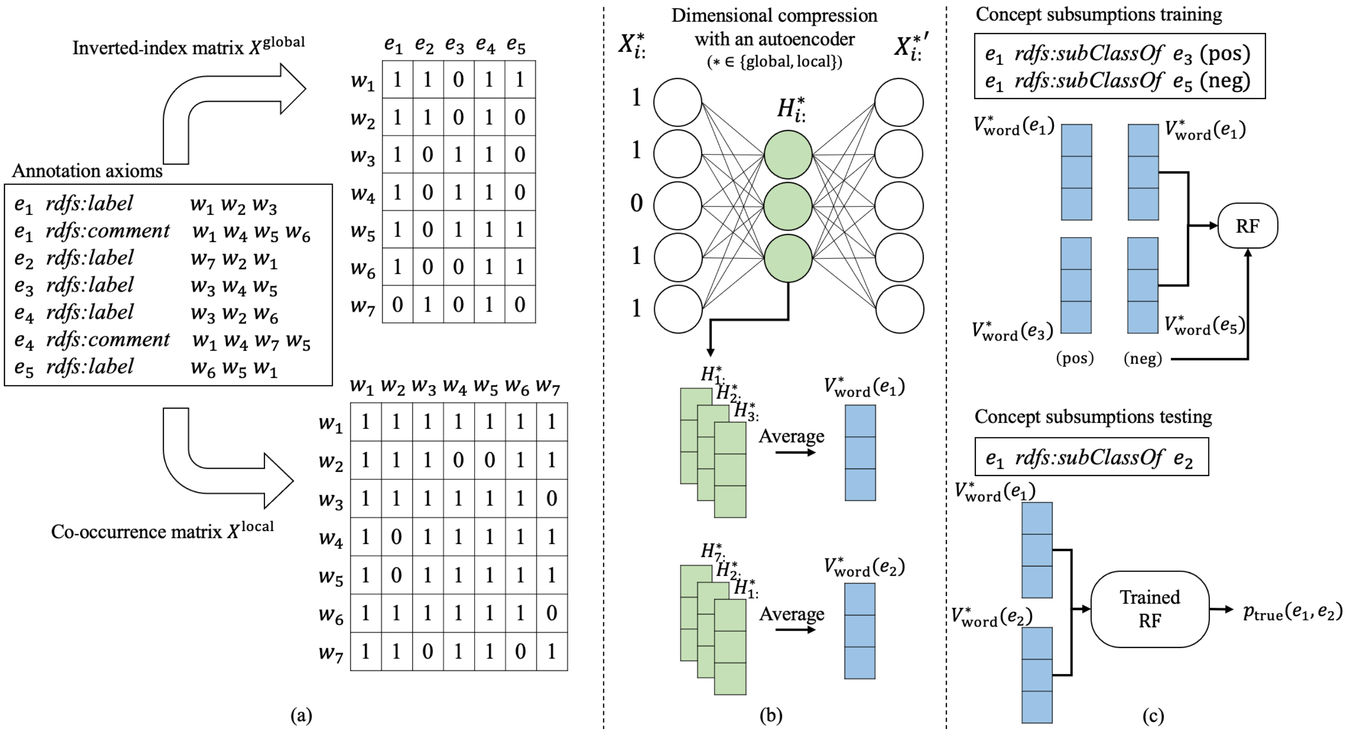


Figure 2: The overall architecture of InME and CoME. (a) The inverted-index and co-occurrence matrices are generated from annotation axioms. (b) The matrices are compressed with an autoencoder and entity embeddings are transformed by averaging the word embeddings of annotations whose property is *rdfs:label*. (c) An RF classifier is trained using the entity embeddings and the concept subsumption axioms.

number of co-occurrences of "crinial" as local information. The local information may contain excessive details due to co-occurring words, whereas the global information consists of only essential details. While the local information may reduce the similarity among words, the global information is more concise, and therefore the similarity is less likely to decrease. As Word2Vec primarily captures local information, it is unable to extract global information, which pertains to pure entity-to-word relationships.

3.3 Inverted-index and Co-occurrence Matrices

InME extracts global information between entities and the words used in the annotations of these entities. An inverted-index matrix is defined as $X^{\text{global}} \in \mathbf{R}^{|W| \times |\mathbb{C} \cup \mathbb{I}|}$ (Figure 2(a)), where W is the set of words in all annotation axioms excluding IRIs. Each X_{ij}^{global} indicates whether the word w_i appears in the annotations of entity e_j. Let $S_{\text{ann}}(e)$ be the set of all words in the annotations of an entity $e \in \mathbb{C} \cup \mathbb{I}$. Then, X_{ij}^{global} is given by

$$X_{ij}^{\text{global}} = \begin{cases} 1 & \text{if } w_i \in S_{\text{ann}}(e_j) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

In addition, we utilize property annotations to augment the set of annotation words. Let ax_e be the set of axioms containing an entity e in the ABox or TBox and let $R_{\text{ann}}(e)$ be the set of words in the annotations of property $R \in \mathbb{R}$ in ax_e. Then, $X^{\text{global}+} \in \mathbf{R}^{|W| \times |\mathbb{C} \cup \mathbb{I}|}$ is defined as an inverted-index matrix that additionally represents the annotations of R and applying Equation (6) as $w_i \in S_{\text{ann}}(e_j) \cup R_{\text{ann}}(e_j)$.

CoME extracts local information about which words w and w' appear together in each annotation axiom. A co-occurrence matrix is defined as $X^{\text{local}} \in \mathbf{R}^{|W| \times |W|}$ (Figure 2(a)). Each X_{ij}^{local} indicates the co-occurrence of word w_i with word w_j. Then, X_{ij}^{local} is given by

$$X_{ij}^{\text{local}} = \begin{cases} 1 & \text{if } \exists e \in \mathbb{C} \cup \mathbb{I}. \{w_i, w_j\} \subseteq S_{\text{ann}}(e) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Similarly, property annotations are utilized to augment the set of annotation words. $X^{\text{local}+} \in \mathbf{R}^{|W| \times |W|}$ is defined as a co-occurrence matrix that additionally represents the annotations of $R \in \mathbb{R}$, and apply Equation (7) as $e \in \mathbb{C} \cup \mathbb{I} \cup \mathbb{R}$.

The dimensions $|\mathbb{C} \cup \mathbb{I}|$ and $|W|$ of X_i^{global} and X_i^{local} are too large to be applied to the concept subsumption task, where X_i^* is a row vector of X^*. We transform X^{global} and X^{local} into low-dimensional word embeddings by applying an autoencoder (Figure 2(b)). The low-dimensional middle layer H^* is given by

$$H^* = \text{ReLU}(\hat{W}_{\text{in}} X^* + \hat{b}_{\text{in}}) \tag{8}$$

where $* \in \{$ global, global$^+$, local, local$^+$ $\}$.

Table 1: The statistics of GO, FoodOn, and HeLiS. *Words per entity* represents the average number of annotation words per entity after preprocessing. The value in parentheses for HeLiS represents the case where names extracted from the IRI are added as annotation words.

Ontology	Classes	Individuals	Annotation axioms	Words per entity	Concept subsumptions	Concept assertions
GO	44,244	0	452,028	43.54	72,601	0
FoodOn	28,182	0	142,536	30.07	29,778	0
HeLiS	277	20,318	4,984	0.40(1.83)	261	20,318

Table 2: The evaluation results on the concept subsumption task for GO, FoodOn, and HeLiS.

	GO				FoodOn				HeLiS			
Model	MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10
RDF2Vec[22]	0.043	0.017	0.057	0.087	0.078	0.053	0.097	0.119	-	-	-	-
TransE[4]	0.015	0.005	0.018	0.030	0.029	0.011	0.044	0.065	-	-	-	-
TransR[17]	0.048	0.016	0.076	0.113	0.072	0.044	0.093	0.130	-	-	-	-
DistMult[30]	0.046	0.018	0.068	0.097	0.076	0.045	0.099	0.134	-	-	-	-
ELEm[16]	0.018	0.005	0.021	0.036	0.040	0.014	0.067	0.099	-	-	-	-
Onto2Vec[23]	0.024	0.008	0.031	0.053	0.034	0.014	0.047	0.064	-	-	-	-
OPA2Vec[24]	0.075	0.032	0.106	0.157	0.093	0.058	0.117	0.156	-	-	-	-
OWL2Vec*[6]	0.170	0.076	0.258	0.376	0.213	0.143	0.287	0.357	0.595	0.451	**0.786**	**0.890**
InME	**0.183**	**0.085**	**0.278**	**0.402**	0.235	0.159	0.317	0.379	0.580	0.437	0.756	0.825
(InME+Pro)‖Word2Vec	0.160	0.073	0.234	0.347	**0.242**	**0.163**	**0.325**	**0.398**	0.548	0.416	0.697	0.798
(CoME+Pro)‖OWL2Vec*	0.134	0.056	0.197	0.297	0.190	0.125	0.250	0.319	**0.621**	**0.485**	0.779	0.871

3.4 Entity Embeddings and Concatenation

Similar to OWL2Vec*, we convert word embeddings into entity embeddings by averaging the row vectors $H^*_{i:}$ for all $w_i \in S_{\text{label}}(e)$. The embedding of entity e is expressed as

$$V^*_{\text{word}}(e) = \frac{1}{|S_{\text{label}}(e)|} \sum_{w_i \in S_{\text{label}}(e)} H^*_{i:} \tag{9}$$

For each concept subsumption axiom $e_1 \sqsubseteq e_2$, a binary classifier RF is trained by the entity embeddings $V_{\text{word}}(e_1)$ and $V_{\text{word}}(e_2)$ (Figure 2(c)). The score p_{true} of e_1 and e_2 is given by

$$p_{\text{true}}(e_1, e_2) = \text{RF}(V^*_{\text{word}}(e_1)||V^*_{\text{word}}(e_2)) \tag{10}$$

As the concatenation of InME and CoME, both global and local information of annotation axioms can be considered by using $V^{\text{global}}_{\text{word}}(e) \parallel V^{\text{local}}_{\text{word}}(e)$ for each entity e.

4 EXPERIMENTS

We evaluate the performance of InME and CoME on the concept subsumption task for the three OWL ontologies: GO, FoodOn, and HeLiS. Our experiment is conducted in the same manner as that of OWL2Vec* [6].

4.1 Datasets

We use the OWL ontologies, GO [7], FoodOn [9], and HeLiS [10]. GO represents the biological knowledge of genes and their products expressed by DL $\mathcal{SRI}$. FoodOn represents the knowledge of foods that contain materials consumed by humans and domesticated animals, as expressed by DL $\mathcal{SRIQ}$. HeLiS represents the knowledge about food and physical activity domains, expressed by DL $\mathcal{ALCHIQ}(\mathcal{D})$. Whereas HeLiS contains both concept subsumption and concept instance axioms, GO and FoodOn contain only concept subsumption axioms. The details of these three ontologies are summarized in Table 1.

In the experiments with HeLiS, some words extracted from the IRIs of entities are added to the annotation words, as described in Section 2.2. This process causes a problem in that some evaluations in validation or testing are always ranked first. For example, in an axiom, *Lysine_100* $\sqsubseteq$ *Lysine*, the word "Lysine" is added as each annotation of *Lysine_100* and *Lysine* if these entities have no annotation; the embeddings of *Lysine_100* and *Lysine* obtained by averaging the one-word embedding of "Lysine" are identical. Therefore, we exclude the trivial axioms from validation and testing data. For the revised HeLiS, we re-experiment with OWL2Vec* using $D_{\text{s,l,rc}}$, which is the best combination of training corpora [6].

4.2 Experimental Setup

For GO and FoodOn, concept subsumption axioms are randomly divided into training (70%), validation (10%), and testing (20%). In HeLiS, concept instance axioms are divided instead of concept subsumption axioms. The validation data are used for parameter tuning. The RF classifier is trained with the concept subsumption (or concept instance) training axioms, as described in Section 2.2. For each axiom in the validation and testing data, the tail entity e_2 is predicted from e_1 in concept subsumptions $e_1 \sqsubseteq e_2$. Each class x is sorted in descending order based on the score $p_{\text{true}}(e_1, x)$. We report the mean reciprocal rank (MRR) and Hits@n ($n = 1, 5, 10$).

We also evaluate the concatenation of InME or CoME with other embedding models. We select other embedding models: (a)

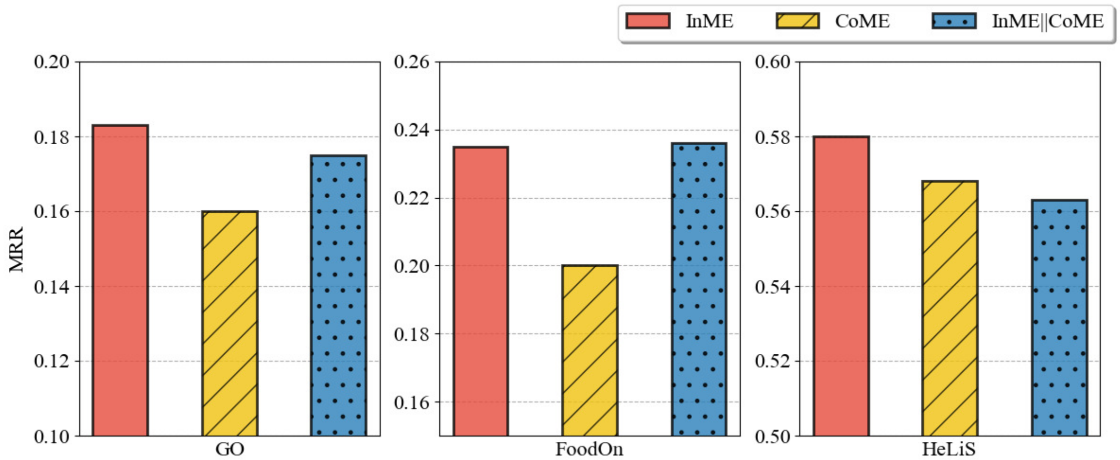


Figure 3: The comparison of MRRs for InME, CoME, and the concatenation of InME and CoME.

OWL2Vec* and (b) a Word2Vec model, which is trained with skip-gram by annotation axioms. Note that annotation words are transformed into lowercase letters, and non-English characters are removed before training. The best combination of training corpora for OWL2Vec* is selected from the OWL2Vec* paper [6], $D_{s,l}$ for GO, $D_{s,l}$ for FoodOn, and $D_{s,l,rc}$ for HeLiS.

We use the compressed dimension n of InME and CoME for the autoencoder among $n = \{50, 100, 200\}$. We select the best parameter by the highest MRR in the validation data. In addition, the dimensions of the two concatenated embeddings are $2n$.

4.3 Results

Table 2 shows the performance of our models, InME and CoME, compared with existing models, where +Pro is the case property annotations are considered and || is the concatenation of two embeddings. Note that we site the results of the existing models from OWL2Vec* [6]. For GO and FoodOn, InME, which only extracts the global information outperforms OWL2Vec*, which extracts logical and graph structures as well as annotation axioms in OWL ontologies. This implies that the global information plays an important role in predicting concept subsumptions. For the revised HeLiS, InME underperforms OWL2Vec because HeLiS originally has extremely few annotation axioms compared to GO and FoodOn, as shown in Table 1. However, CoME+Pro concatenated with OWL2Vec* embeddings outperforms OWL2Vec* in MRR and Hits@1. This result shows that our models can enhance OWL2Vec* by the fact that the local information represented in CoME+Pro supplement the lack of annotations in HeLiS.

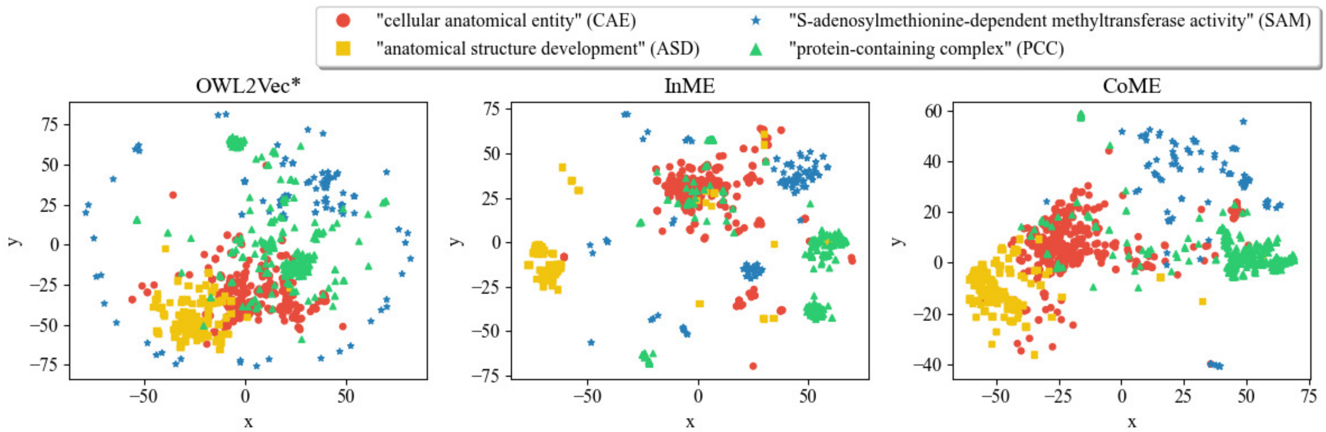
Table 3 presents the detailed results of our embedding models in the experiment. Figure 3 shows the results of the comparison of the MRRs for InME, CoME, and the concatenation of InME and CoME. InME shows a higher MRR than the other models for the three ontologies. In addition, the MRR of InME||CoME outperforms CoME for GO and FoodOn, but is either below or comparable for InME. On the other hand, for HeLiS, CoME exhibits a higher MRR than the concatenation. HeLiS has significantly fewer annotation axioms than GO and FoodOn, which could have diluted the information owing to the increased dimension caused by concatenation.

We analyze the embeddings of InME, CoME, and OWL2Vec* using t-SNE [27]. Figure 4 shows the visualizations in t-SNE of InME, CoME, and OWL2Vec* for GO, FoodOn, and HeLiS. Each point in these figures represents a class (or an individual), with the colors indicating their parent classes. For GO, in both InME and CoME, classes with a common superclass tend to be clustered on the plot. In particular, subclasses of SAM (blue points) are circularly distributed in OWL2Vec*, whereas these points are clustered in InME and CoME. For FoodOn, a similar trend is observed in OWL2Vec* and CoME, where subslasses of OC (green points) are evenly distributed throughout, while the other three subclasses form clusters. On the other hand, in InME, we observe that many of the subclasses of OC form a cluster at a single point. These results implies that the global information of InME is more appropriate for capturing the similarity of entities. For HeLiS, the four subclass groups are neatly clustered in InME, CoME, and OWL2Vec*. The embeddings of entities with identical annotation words are converted to the same embeddings through the average of word embeddings by Equation (1) or (9). The well-divided visualizations of HeLiS are largely influenced by the small number of annotations and the averaging embedding transformation method, rather than by the performance of the models.

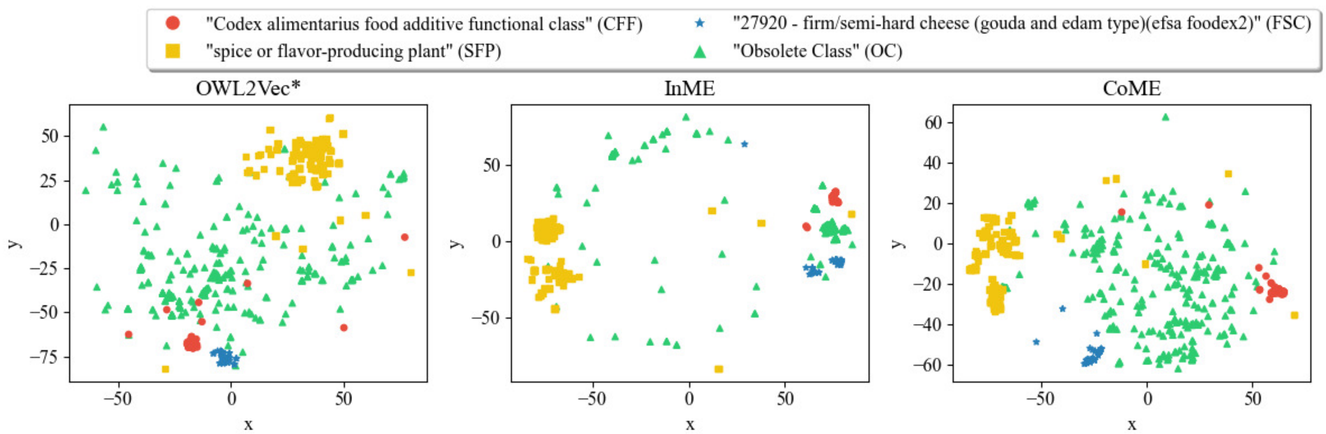
We discuss the impact of property annotations for the concept subsumption task. Table 4 shows a comparison of MRR with and without property annotations. For GO, InME exhibites a lower MRR when property annotations are added. However property annotations in appear the logical structure axioms may introduce noise to the prediction of concept subsumptions. In other models and datasets, property annotations results in no change or a slight improvement in the MRRs. The addition of property annotations tends to be effective for FoodOn and HeLiS.

5 CONCLUSION

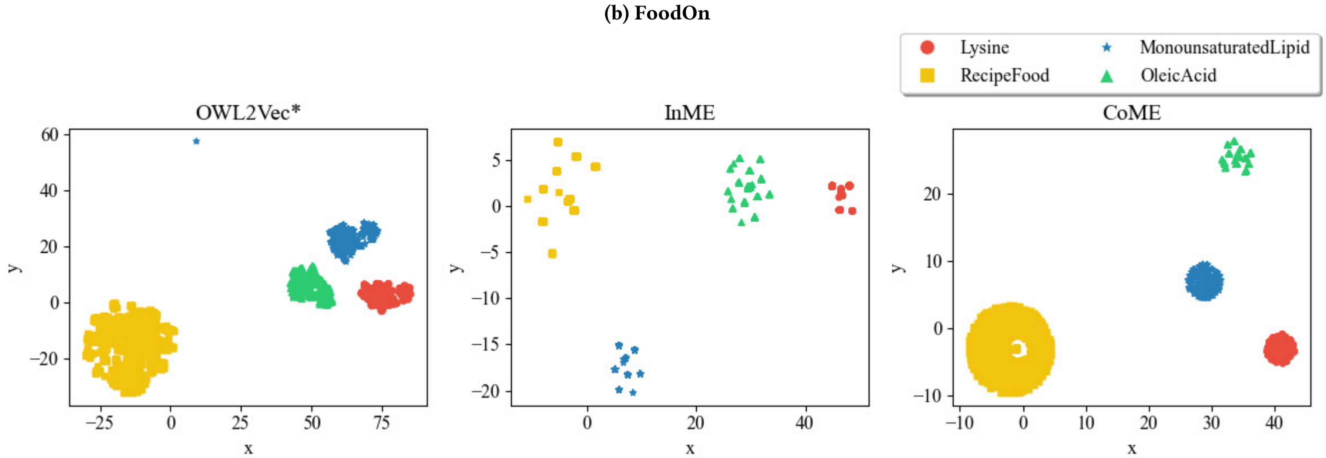
In this study, we proposed the embedding models InME and CoME in OWL ontologies only using annotation axioms. Our preliminary



(a) GO



(b) FoodOn



(c) HeLiS

Figure 4: The t-SNE visualizations of the embeddings OWL2Vec*, InME, and CoME for GO, FoodOn, and HeLiS.

Table 3: The results of our models where other embeddings are concatenated and property annotations are added. Results of ∗ are taken from OWL2Vec* [6].

		GO				FoodOn				HeLiS			
Model	Concatenation	MRR	HITS@1	HITS@5	HITS@10	MRR	HITS@1	HITS@5	HITS@10	MRR	HITS@1	HITS@5	HITS@10
InME	-	**0.183**	**0.085**	**0.278**	**0.402**	0.235	0.159	0.317	0.379	0.580	0.437	0.756	0.825
	OWL2Vec*	0.153	0.069	0.224	0.332	0.208	0.137	0.277	0.348	0.596	0.460	0.768	0.851
	Word2Vec	0.158	0.070	0.234	0.348	0.236	0.160	0.317	0.378	0.547	0.400	0.699	0.807
InME+Pro	-	0.164	0.074	0.245	0.362	0.235	0.155	0.318	0.397	0.573	0.439	0.722	0.839
	OWL2Vec*	0.151	0.065	0.227	0.330	0.223	0.150	0.294	0.369	0.606	0.474	0.775	0.837
	Word2Vec	0.160	0.073	0.234	0.347	**0.242**	0.163	**0.325**	**0.398**	0.548	0.416	0.697	0.798
CoME	-	0.160	0.072	0.246	0.354	0.200	0.140	0.262	0.315	0.568	0.441	0.713	0.805
	OWL2Vec*	0.136	0.057	0.199	0.303	0.185	0.121	0.242	0.311	0.614	0.476	0.784	0.871
	Word2Vec	0.145	0.064	0.210	0.318	0.212	0.141	0.281	0.341	0.499	0.366	0.639	0.763
	InME	0.175	0.080	0.267	0.388	0.236	**0.168**	0.305	0.358	0.563	0.434	0.720	0.811
CoME+Pro	-	0.151	0.068	0.228	0.333	0.200	0.131	0.271	0.330	0.566	0.437	0.715	0.795
	OWL2Vec*	0.134	0.056	0.197	0.297	0.190	0.125	0.250	0.319	**0.621**	**0.485**	0.779	0.871
	Word2Vec	0.140	0.060	0.205	0.313	0.213	0.145	0.277	0.341	0.537	0.414	0.680	0.752
	InME	0.171	0.078	0.259	0.378	0.239	0.166	0.313	0.376	0.577	0.448	0.738	0.818
OWL2Vec*[6]	-	0.170*	0.076*	0.258*	0.376*	0.213*	0.143*	0.287*	0.357*	0.595	0.451	**0.786**	**0.890**
	Word2Vec	0.139	0.062	0.201	0.304	0.219	0.145	0.294	0.363	0.527	0.393	0.669	0.800

Table 4: The comparison of InME, InME||Word2Vec, and CoME||OWL2Vec* with and without property annotations.

		GO	FoodOn	HeLiS
Model	With property annotations	MRR	MRR	MRR
InME	Yes	0.164	**0.235**	0.573
	No	**0.183**	**0.235**	**0.580**
InME||Word2Vec	Yes	**0.160**	**0.242**	**0.548**
	No	0.158	0.236	0.547
CoME||OWL2Vec*	Yes	0.134	**0.190**	**0.621**
	No	**0.136**	0.185	0.614

experiment showed that the Word2Vec with annotation axioms outperforms that with logical structure axioms in the concept subsumption task. We showed that the local information, which is the co-occurrence of annotation words, and the global information, which is the relationship between an entity and annotation words, contribute to deciding parent-child relationships in an ontology. To extract these information, we defined InME and CoME using inverted-index and co-occurrence matrices. In the evaluation experiments, we demonstrated that InME outperformed existing models for GO and FoodOn, and that CoME+Pro concatenated with OWL2Vec* outperformed existing models for HeLiS. These results show that the global information plays an important role in the concept subsumption task. For HeLiS, which has a limited number of annotation words, CoME+Pro concatenated with OWL2Vec* outperform existing models. This result indicates that our models can improve OWL2Vec* by leveraging the local information through the concatenation due to the lack of annotations in HeLiS.

In future work, there is further research to leverage the logical structure axioms in our models to improve prediction accuracy. Additionally, our InME and CoME can be applied to other tasks, such as subsumption ontology tasks for more complex classes (e.g., $C \sqsubseteq \exists R.D$).

REFERENCES

[1] Franz Baader, Sebastian Brandt, and Carsten Lutz. 2005. Pushing the $\mathcal{EL}$ Envelope. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*. 364–369.

[2] Yushi Bai, Zhitao Ying, Hongyu Ren, and Jure Leskovec. 2021. Modeling Heterogeneous Hierarchies with Relation-specific Hyperbolic Cones. In *Proceedings of Advances in Neural Information Processing Systems 34*. 12316–12327.

[3] Sean Bechhofer, Frank Van Harmelen, Jim Hendler, Ian Horrocks, Deborah L McGuinness, Peter F Patel-Schneider, Lynn Andrea Stein, et al. 2004. OWL Web Ontology Language Reference. *W3C recommendation* 10, 2 (2004), 1–53.

[4] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Proceedings of Advances in Neural Information Processing Systems 26*. 2787–2795.

[5] Jiaoyan Chen, Yuan He, Yuxia Geng, Ernesto Jiménez-Ruiz, Hang Dong, and Ian Horrocks. 2023. Contextual Semantic Embeddings for Ontology Subsumption Prediction. *World Wide Web (WWW)* 26, 5 (2023), 2569–2591.

[6] Jiaoyan Chen, Pan Hu, Ernesto Jiménez-Ruiz, Ole Magnus Holter, Denvar Antonyrajah, and Ian Horrocks. 2021. OWL2Vec*: Embedding of OWL Ontologies. *Machine Learning* 110, 7 (2021), 1813–1845.

[7] Gene Ontology Consortium. 2008. The Gene Ontology Project in 2008. *Nucleic acids research* 36, suppl_1 (2008), D440–D444.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4171–4186.

[9] Damion M Dooley, Emma J Griffiths, Gurinder S Gosal, Pier L Buttigieg, Robert Hoehndorf, Matthew C Lange, Lynn M Schriml, Fiona SL Brinkman, and William WL Hsiao. 2018. FoodOn: a Harmonized Food Ontology to Increase Global Food Traceability, Quality Control and Data Integration. *npj Science of*

Table 5: The hyper-parameters of dimensions.

Model	Concatenation	GO	FoodOn	HeLiS
InME	-	200	50	200
	OWL2Vec*	50	100	100
	Word2Vec	50	50	200
InME+Pro	-	200	50	100
	OWL2Vec*	50	50	200
	Word2Vec	50	50	100
CoME	-	50	50	200
	OWL2Vec*	50	50	50
	Word2Vec	50	50	50
	InME	50	50	50
CoME+Pro	-	50	50	200
	OWL2Vec*	50	50	100
	Word2Vec	50	50	200
	InME	50	50	100
OWL2Vec*	-	-	-	50
	Word2Vec	50	50	100

Food 2, 1 (2018), 1–10.

[10] Mauro Dragoni, Tania Bailoni, Rosa Maimone, and Claudio Eccher. 2018. HeLiS: An Ontology for Supporting Healthy Lifestyles. In *Proceedings of 17th International Semantic Web Conference (Lecture Notes in Computer Science)*, Vol. 11137. 53–69.

[11] Dinesh Garg, Shajith Ikbal, Santosh K. Srivastava, Harit Vishwakarma, Hima P. Karanam, and L. Venkata Subramaniam. 2019. Quantum Embedding of Knowledge for Reasoning. In *Proceedings of Advances in Neural Information Processing Systems 32*. 5595–5605.

[12] Birte Glimm, Ian Horrocks, Boris Motik, Giorgos Stoilos, and Zhe Wang. 2014. HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning* 53, 3 (2014), 245–269.

[13] Thomas R Gruber. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge acquisition* 5, 2 (1993), 199–220.

[14] Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the Dimensionality of Data with Neural Networks. *science* 313, 5786 (2006), 504–507.

[15] Yevgeny Kazakov, Markus Krötzsch, and František Simančík. 2014. The Incredible ELK. *Journal of Automated Reasoning* 53, 1 (2014), 1–61.

[16] Maxat Kulmanov, Wang Liu-Wei, Yuan Yan, and Robert Hoehndorf. 2019. EL Embeddings: Geometric Construction of Models for the Description Logic $\mathcal{EL}$++. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. 6103–6109.

[17] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2181–2187.

[18] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of 1st International Conference on Learning Representations*.

[19] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of Advances in Neural Information Processing Systems 26*. 3111–3119.

[20] Sutapa Mondal, Sumit Bhatia, and Raghava Mutharaju. 2021. EmEL++: Embeddings for $\mathcal{EL}$++ Description Logic. In *Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering*, Vol. 2846.

[21] Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio. 2016. Holographic Embeddings of Knowledge Graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. 1955–1961.

[22] Petar Ristoski and Heiko Paulheim. 2016. RDF2Vec: RDF Graph Embeddings for Data Mining. In *Proceedings of 15th International Semantic Web Conference*, Vol. 9981. 498–514.

[23] Fatima Zohra Smaili, Xin Gao, and Robert Hoehndorf. 2018. Onto2Vec: Joint Vector-based Representation of Biological Entities and Their Ontology-based Annotations. *Bioinformatics* 34, 13 (2018), i52–i60.

[24] Fatima Zohra Smaili, Xin Gao, and Robert Hoehndorf. 2019. OPA2Vec: Combining Formal and Informal Content of Biomedical Ontologies to Improve Similarity-based Prediction. *Bioinformatics* 35, 12 (2019), 2133–2140.

[25] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In *Proceedings of 7th International Conference on Learning Representations*.

[26] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33nd International Conference on Machine Learning*. 2071–2080.

[27] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of machine learning research* 9, 11 (2008), 2579–2605.

[28] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. 1112–1119.

[29] Bo Xiong, Nico Potyka, Trung-Kien Tran, Mojtaba Nayyeri, and Steffen Staab. 2022. Faithful Embeddings for $\mathcal{EL}$++ Knowledge Bases. In *Proceedings of 21th International Semantic Web Conference*, Vol. 13489. 22–38.

[30] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Proceedings of 3rd International Conference on Learning Representations*.

[31] Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*. 3065–3072.

A APPENDIX

Table 5 displays the hyper-parameters of dimensions selected by the performance of the validation data in our experiments, as described in Table 3.