Automatic Multilingual Hypernym–Hyponym Relation Extraction Using a Link Prediction Model

Kohei Iwakuma
Graduate School of Science and Engineering
Chuo University
Tokyo, Japan
a19.bxw6@g.chuo-u.ac.jp

Yao Gong
Graduate School of Science and Engineering
Chuo University
Tokyo, Japan

Hidetsugu Nanba
Faculty of Science and Engineering
Chuo University
Tokyo, Japan
nanba@kc.chuo-u.ac.jp

Satoshi Fukuda
Faculty of Science and Engineering
Chuo University
Tokyo, Japan
fukuda.satoshi.3238@kc.chuo-u.ac.jp

***Abstract*—In natural language processing, hypernym–hyponym relations are the core of the body of knowledge and are useful for many downstream tasks, for example, technical trend analysis and patent examination. However, it is very costly to manually maintain these relations between terms. In this paper, we extract hypernym–hyponym relations from patent text data written in Japanese, English, and Chinese, and automatically construct a multilingual thesaurus. The proposed method consists of the following two steps. First, we use a generative adversarial network (GAN) to identify the terms in a hypernym–hyponym relation. Then, ConvE and GraphSAGE are combined to predict links on the graph of hypernym–hyponym relations constructed in the previous step, and to predict missing edges that should be in a hypernym–hyponym relation. In experiments conducted to demonstrate the effectiveness of the proposed method, it was found that our method outperformed previous methods in both the identification of hypernym–hyponym relations using GAN and link prediction using a combination of ConvE and GraphSAGE. We constructed a classifier that can discriminate between hypernym–hyponym relations using a trained Discriminator with GAN. In our experiments, we achieved a recall rate of 0.936 in English. And we propose a new method that can automatically complement missing hypernym–hyponym relations. The proposed model achieved a score of 97.3 on the H@10 evaluation index with respect to the prediction of Chinese hypernym–hyponym relations.**

Keywords—hypernym–hyponym relation, multilingual, generative adversarial network (GAN), link prediction, information extraction, patent

I. Introduction

This paper proposes a method for automatically constructing a multilingual thesaurus by extracting hypernym–hyponym relations from patent text data written in Japanese, English, and Chinese. A thesaurus including hypernym–hyponym relations of terms is useful in various situations, such as text mining and information retrieval. However, because it is very costly to manually maintain such a thesaurus, an automatic construction of such a thesaurus is required.

Methods for acquiring hypernym–hyponym relations using linguistic patterns have been explored. Hearst [1] extracted hypernym–hyponym relations from English newspaper articles using lexical syntactic patterns. As a first step in the construction of a multilingual thesaurus, we also use multilingual linguistic patterns such as "A や B などの C" (C such as A and B) in Japanese,"D such as E, F" in English, and "G 包括 H 和 I" (G such as H and I) in Chinese. From these patterns, we obtain term pairs that are candidates for hypernym–hyponym relations.

We construct a multilingual thesaurus for candidate hypernym–hyponym term pairs extracted using the patterns in two steps: first, we use a generative adversarial network (GAN) [2] to build a discriminator that identifies terms in a hypernym–hyponym relation; then, we apply link reduction techniques to the graph of hypernym–hyponym relations constructed in the first step to predict missing edges that should normally be in the hypernym–hyponym relation.

Link prediction tasks predict missing relations between nodes and are key to automatically understanding data in large graph structures. Although previous work on link prediction has focused on large knowledge graphs [3, 4], there has been little work on predicting links of hypernym–hyponym relations. In this study, we propose a method to predict missing hypernym–hyponym relations. ConvE [5] is a convolutional neural network (CNN) model that has proven to be very useful in a variety of prediction tasks. GraphSAGE [6] is one of the methods for computing embeddings related to graph neural networks (GNNs). It also provides an effective approach for solving link prediction tasks. We combine ConvE and GraphSAGE to predict missing hypernym–hyponym relations. In the link prediction of hypernym–hyponym relations, we transform each word into a node in a graph, and predict the existence of link edges between nodes in the graph (Fig. 1).

We aim to increase consistency and accuracy across languages by using the multilingual thesaurus constructed in this way. We also expect to discover new relations and knowledge from technical documents and scientific literature.

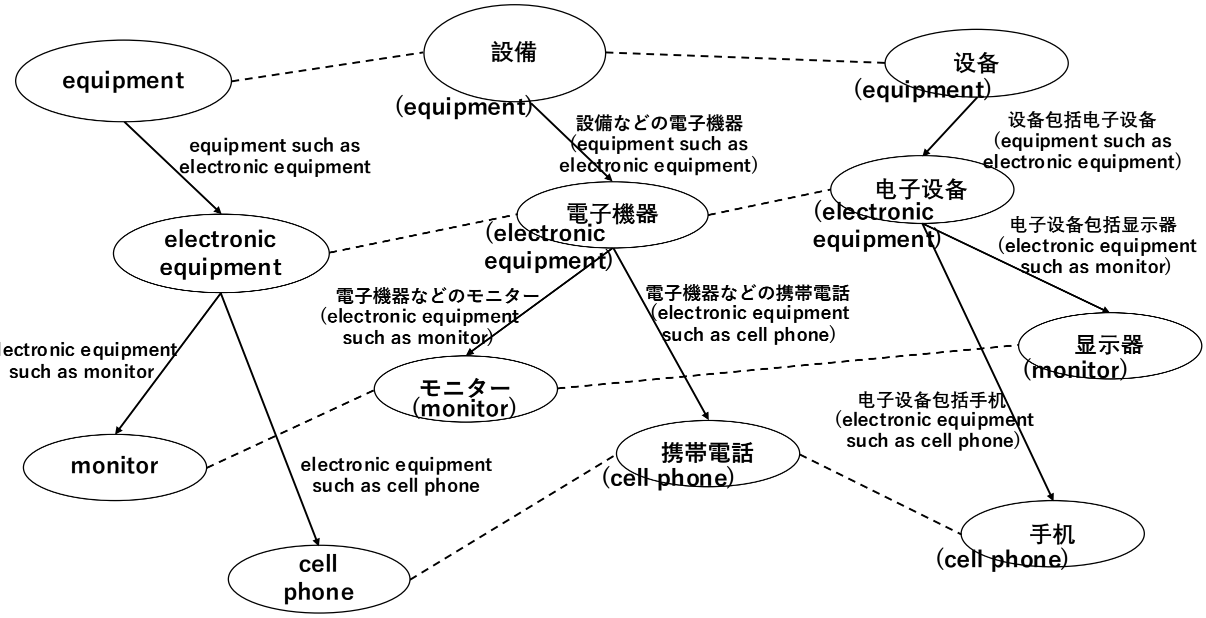

Fig. 1: Graph of multilingual hypernym–hyponym relations

II. Related Work

A. Extraction of hypernym–hyponym relations

A typical method for extracting hypernym–hyponym relations is to use linguistic expressions. Nanba et al. [7] automatically constructed a Japanese–English thesaurus by extracting candidate terms of hypernym–hyponym relations from patents written in Japanese and English, and then by aligning the Japanese and English terms using statistical machine translation technology. Whereas their method excluded all terms that cannot be matched by statistical translation, in this study, we use GAN to keep correct hypernym–hyponym relations in the multilingual thesaurus, even though they cannot be matched by translation. Furthermore, we expect to construct a more comprehensive multilingual thesaurus by using a link prediction model to detect hypernym–hyponym relations that have not been extracted by the linguistic pattern method.

Other studies of multilingual hypernym–hyponym relation detection have focused primarily on detecting hypernym–hyponym relations in high-resource languages such as English. Yu et al. [8] addressed the task of detecting low-resource hypernym–hyponym words. To solve the problem of lack of resources, we propose a method to apply meta-learning. Yu et al. compared three collaborative training paradigm methods: "cross-language training," "multilingual training," and "meta-learning." Experimental results showed that the meta-learning method was superior to the other methods. Meta-learning significantly improves the performance of languages with very few resources by preventing overfitting of small datasets.

Upadhyay et al. [9] proposed BISPARSE-DEP, an unsupervised approach for cross-language hypernym–hyponym relation detection. Bilingual embedding and an unsupervised entailment scorer were used to detect hypernym–hyponym relations. The difference between the present study and previous work is that the missing relations of each language are automatically complemented using a link prediction model based on a single graph of multilingual hypernym–hyponym relations extracted using linguistic patterns. For example, assume that "small dog – dog" and "small dog – animal" exist as hypernym–hyponym relations.

B. Construction of a hypernym–hyponym relation classifier using GAN

Traditional research using GANs is aimed at data augmentation, thereby generating more real images and sentences. Kober et al. [10] proposed a method to augment data using GANs to detect hypernym–hyponym relations. Here, we examine data augmentation methods that utilize existing training data to generate new training examples. By combining the linguistic principles of transitivity of hypernym–hyponym relations and cross-modifier-noun constructions, hypernym–hyponym relations can be assumed. If "animal" is a hypernym relation of "dog" and "dog" is a hypernym relation of "small dog," "animal" is added to the hypernym relation of "small dog." To find appropriate candidates for words, nouns must be combined with modifiers. Experimental results show that the proposed data augmentation and dataset extension methods can be used to significantly improve classifier performance. Using discriminator as a classifier, Chavdarova and Fleuret [11] proposed a semi-supervised GAN called SGAN to perform multiclass classification tasks. The present study differs in that the variance representation of the hypernym–hyponym words is input to the GAN, and the learned discriminator is used as a classifier to identify the hypernym–hyponym relations.

C. Link prediction

In recent years, deep learning has excelled in a number of natural language processing tasks. The reason for this is the evolution of models and algorithms related to neural networks, and there have been many studies on neural link prediction models. Bordes et al. [12] proposed TransE, a translating embeddings model that uses deep learning to embed knowledge graphs into a low-dimensional space. Yang et al. [13] proposed the algorithm DistMult (bilinear diagonal), which restricts the relation matrix to a diagonal matrix. DistMult shows that State-of-the-ART (SOTA) is achieved using a simple bilinear diagonal. Trouillon et al. [14] proposed the ComplEx (extension of the complex space) model using complex embeddings to handle a variety of binary relations, including symmetric and antisymmetric relations.

GNNs have been extensively studied since 2016. The main downstream tasks handled by GNNs include link prediction, graph classification, and node classification. For application to graph-structured data, Kipf and Welling [15] proposed a graph convolutional network (GCN) model that can extend CNN input to graph structures. They also proposed an extensible supervised learning approach based on CNNs that can operate directly on the graph. However, GCNs ignore the relations between nodes without considering their relation to each other. To address the problems of GCN, Schlichtkrull et al. [16] introduced the relational graph convolutional network (R-GCN) model to address the effects of different edge relations in the graph structure on nodes. Jiang et al. [17] proposed a link prediction model, ConvR, which builds a convolution filter from the relational representation and applies this filter to the entire entity representation to generate convolution features. In our work, we combine a link prediction model with a graph embedding method to address multilingual autocompletion tasks.

III. Construction of a Hypernym–hyponym Relation Classifier Using GANs

In this study, we propose a method for constructing a classifier for filtering the extracted candidate hypernym–hyponym relations. We create a dataset of WordNet-mapped hypernym–hyponym relations extracted from patents and build a classifier that determines the hypernym–hyponym relations using GANs. We then use a classifier to remove inappropriate nodes from the graph of hypernym–hyponym relations. Fig. 2 summarizes the architecture of the construction of hypernym-level classifiers with GANs. GANs consist of two elements, namely a generator and a discriminator. The role of the generator is to generate spurious data from a random distribution; the discriminator identifies whether data is real or fake. In this study, the discriminator is populated with word embeddings generated from DeepWalk, which is a graph embedding method proposed by Perozzi et al. [18]. We convert the hypernym–hyponym terms to nodes in the graph and vectorize them in DeepWalk. The 200 dimensions of each word are combined to generate embeddings of the hypernym and hyponym pairs of 400 dimensions. We then generate fake data from the Gaussian noise Z given by the generator and train the generator and the discriminator to compete with each other. Finally, we use the learned discriminator as a classifier to classify if it is a hypernym–hyponym relation. GAN aims to optimize the value function (value function) using the following equation:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[\log\left(1 - D\big(G(\boldsymbol{z})\big)\right)\right]$$

where G is a mapping from z to $\boldsymbol{x}$, and D is a sigmoid function that inputs the teacher data $\boldsymbol{x}$ to Discriminator and returns the probability p. $G(z)$ represents the fake data generated from the input noise z to the Generator, and $\mathbb{E}$ represents the expected value.

We train a multilingual dataset mapped to a GAN, then input candidate hypernym–hyponym relation pairs into a trained classifier for classification.

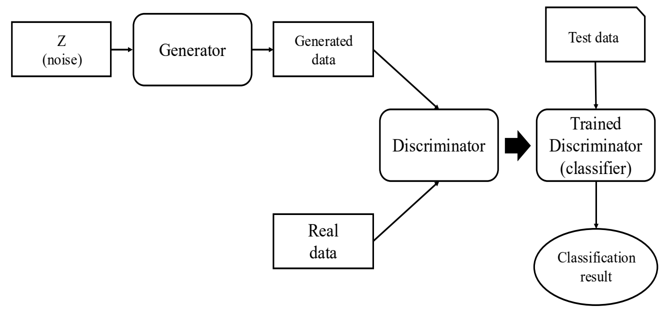

Fig. 2: Construction of hypernym-level classifiers with GANs

IV. Autocompletion of Multilingual Hypernym–Hyponym relations

In this study, we propose a multilingual hypernym–hyponym relation autocompletion method that combines ConvE and GraphSAGE. Although the CNN model ConvE for the link prediction task can be used for monolingual link prediction tasks, it is not easy to use for multilingual datasets. To solve this problem, in this study, we introduce GraphSAGE, which is one of the graph embedding methods in GNNs. GraphSAGE in a graph aggregates embeddings of neighboring nodes and calculates and updates embeddings of its own nodes. ConvE uses random initialization values as embedding of the node. We use DeepWalk to process distributed representations of words. The embedding of each node is obtained by DeepWalk. The acquired embedding is used as the initialization feature of the node. Each node is a 200-dimensional vector of features. We validate the proposed method using the WN18 hypernym–hyponym relation dataset.

Fig. 3 shows the construction of a multilingual hypernym–hyponym complementation system based on a link prediction model. GraphSAGE needs to specify the range by the sample size S of the neighborhood and the depth K from the node. We set K=1 and S=5. We convert the input triple (s, r, o) into the form (hypernym, hypernym–hyponym relation, hyponym) and use the hypernym–hyponym relations as nodes in the graph. Additionally, we collect information from the source node's neighbors by GraphSAGE. Finally, we use the ConvE model to predict the missing relation between source node 1 and target node 7. When predicting the missing relation, the system not only predicts the relation from the source node to the target node, but also considers the target node's neighbors. If there is a corresponding relation between a source node 1 neighbor node and a source node 7 neighbor node, it is likely to be determined to be a superior-hyponym relation. In the experiments described in the next section, we remove the hypernym–hyponym relation from each language and predict the missing relation from the other languages.

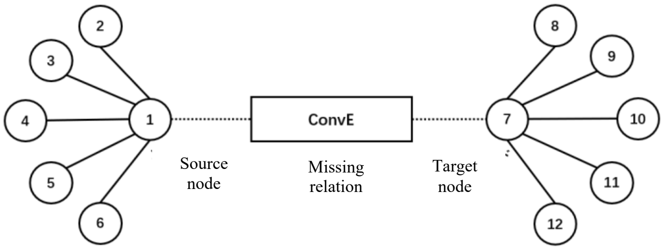

Fig. 3: Prediction of multilingual hypernym–hyponym links using the proposed method

V. Experiment

A. Extraction of hypernym–hyponym relations

Experimental Data

The data for Japanese patents was obtained from the Japan Patent Office (https://www.publication.jpo.go.jp/), Chinese patents from J-Platpat (https://www.j-platpat.inpit.go.jp/c1000), and US patents from USPTO Bulk Data Storage (https://bulkdata.uspto.gov/). The candidate hypernym–hyponym relations with a frequency of 3 or more were extracted from 1,148,250,436 sentences included in Chinese patent specifications, 1,244,905,476 sentences included in Japanese patent specifications, and 1,560,706,140 sentences included in English patent specifications. Table I shows the details of the data.

TABLE I. Results of Hypernym–hyponym Relations Extracted from Japanese, English, and Chinese Patents

Patent data	Year of release	Hypernym–hyponym terminology
Japanese	1993–2018	2,908,574
English	1993–2018	4,506,946
Chinese	2018, 2020, 2021	1,080,357

Experimental Methods

The target terms are only nouns (NOUN) and proper nouns (PROPN), using the part-of-speech analysis of the Python library spaCy. We used the linguistic patterns of each language to extract candidate hypernym–hyponym relations from the target sentences.

Experimental Results

The results of the experiment are shown in Table I. We extracted a total of 8,495,877 top subpatents from Japanese, US, and Chinese patents using the standardized expressions in each language.

Discussion

A common cause of terms being incorrectly extracted in each language is the inclusion of extra words in the extracted hypernym terms. For example, for the sentence "various additives such as antioxidants," extra words such as "various" were extracted. For these issues, we consider it necessary to use multiple parameters and extract only matched term pairs.

B. Construction of the hypernym–hyponym relational discriminator with GAN

Experimental Data

We use the data of 3,794 multilingual hypernym and lower-level relations extracted from Japanese, English, and Chinese patents that have been mapped to the Multilingual WordNet. We split the training and evaluation data at a ratio of 70% to 30%. The data is divided into 2,656 training data and 1,138 evaluation data for each language.

Experimental Methods

GAN is implemented by an all-associative neural network using PyTorch. We use Precision, Recall, and F-measure as evaluation indices. The threshold is set at 0.5.

Comparative Method

We built our binary classifier using only the discriminator, without using the GAN framework. We also constructed a binary classifier based on logistic regression.

Experimental Results

The classification results of the proposed method and the results classified by the comparison method are shown in Table II. Experimental results confirmed the effectiveness of the proposed method, producing higher accuracy, repeatability, and F-measure than the comparative method.

Fig. 4 shows the loss functions of the generator and the discriminator of the GAN when learned in Japanese. There is a similar trend in loss when learning in English and Chinese. Although the process of training GANs proceeds with oscillating loss, we confirm that the learning process is unstable. It can be seen that the loss of the discriminator is suppressed while the loss of the generator is increased.

Discussion

The reason for the lower accuracy, recall, and F-values obtained for Japanese than for English and Chinese can be attributed to the larger number of terms included in the constructed Japanese term list than for the other languages.

TABLE II. Identification Results of Hypernym–Hyponym Relations for each Language in the Multilingual Dataset

Method	Evaluation Measures	Japanese	Chinese	English
Logistic Regression	Precision	0.841	0.864	0.873
	Recall	0.524	0.644	0.645
	F-measure	0.645	0.738	0.742
CNN (Discriminator)	Precision	0.841	0.840	0.861
	Recall	0.469	0.529	0.584
	F-measure	0.602	0.649	0.696
GAN (Generator + Discriminator)	Precision	**0.866**	**0.872**	**0.875**
	Recall	**0.875**	**0.917**	**0.936**
	F-measure	**0.871**	**0.894**	**0.905**

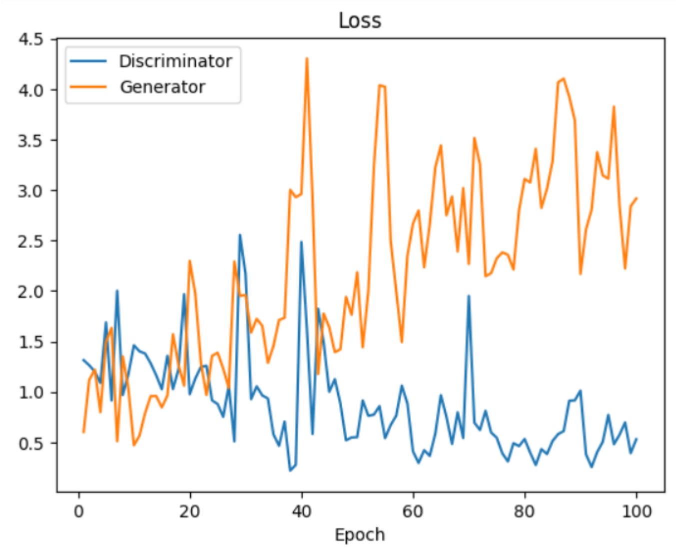

Fig. 4: Loss functions when learned in Japanese.

C. Construction of a multilingual autocompletion system

Experimental Data

In the comparison experiment shown in Table III, we used 74,442 data from the WN18 dataset (a total of 18 relations) with only hypernym–hyponym relations.

In the experiment shown in Table IV, we used 6,231 Japanese, English, and Chinese patent terms, which are also present in WordNet, in terms of hypernym–hyponym relations. There are three types of relations between terms (is a hypernym, is not a hypernym, alignment). We split the dataset per language into 46,109 data for training, 9,970 for validation, and 9,242 for evaluation.

Experimental Methods

We build the ConvE model in PyTorch Geometric and implement GraphSAGE. We use Mean Reciprocal Rank (MRR), HITS@1, HITS@3 and HITS@10(H@1, H@3, and H@10) for our evaluation indices. The following formulas are used to calculate each rating scale. $|S|$ represents the number of combinations of triples: two entities and one relationship. rank_i represents the predicted link ranking of the i-th triple.

$$\mathrm{MRR} = \frac{1}{|S|}\sum_{i=1}^{|S|}\frac{1}{\mathrm{rank}_i} = \frac{1}{|S|}\left(\frac{1}{\mathrm{rank}_1}+\frac{1}{\mathrm{rank}_2}+\cdots+\frac{1}{\mathrm{rank}_{|S|}}\right)$$

$$\mathrm{HITS}\ @n = \frac{1}{|S|}\sum_{i=1}^{|S|}\mathbb{I}(\,\mathrm{rank}_{\,i} \leqslant n)$$

Comparative Method

We use ConvE as proposed by Dettmers et al. [5] and ConvR as proposed by Jiang et al. [17] as our comparison method.

Experimental Results

Table III shows the results of the prediction of the hypernym–hyponym relations using ConvE+GraphSAGE—the method proposed in this study—and the results of the prediction by the comparative method. The experimental results show that the proposed method is superior in the H@1, H@10, and MRR evaluation indices, confirming the effectiveness of the proposed method. The results of the autocompletion experiment are shown in Table IV, and show that English outperformed the other languages in the H@1 and MRR metrics. Chinese outperformed the other languages in the H@3 and H@10 evaluation indicators.

Discussion

The reason for the lower MRRs obtained in Japanese than in English and Chinese is likely the higher number of disconnected nodes in the graph. When embedding nodes using GraphSAGE were calculated, the number of nodes in the neighborhood was insufficient. It is expected that the MRR can be improved in the future by reducing the number of disconnected nodes.

TABLE III. H@K AND MRR RESULTS OF PREDICTING HYPERNYM–HYPONYM RELATIONS IN THE WN18 HYPERNYM–HYPONYM RELATION DATASET

Method	H@1	H@3	H@10	MRR
ConvR	42.9	49.3	55.4	47.5
ConvE	44.7	**81.4**	90.1	64.0
ConvE+GraphSAGE	**52.3**	80.1	**90.2**	**67.1**

TABLE IV. H@K AND MRR RESULTS FOR EACH LANGUAGE WHEN PREDICTING HYPERNYM–HYPONYM RELATIONS USING THE MULTILINGUAL DATASET

Method	Evaluation measure	Japanese	Chinese	English
ConvE + GraphSAGE	H@1	29.4	25.2	**35.0**
	H@3	62.3	**71.4**	65.5
	H@10	92.0	**97.3**	90.4
	MRR	49.4	50.1	**53.3**

VI. CONCLUSION

In this study, we extracted hypernym–hyponym relations from Japanese, English, and Chinese patent data using multilingual definite expressions and mapped them to the Multilingual WordNet to create a multilingual hypernym–hyponym relation dataset. We also used a trained discriminator with GAN to build a classifier that can identify hypernym–hyponym relations. The constructed classifier achieved a recall rate of 0.936 in English.

Combining the CNN link prediction model ConvE and the GNN algorithm GraphSAGE, we propose a new method that can automatically complement missing hypernym–hyponym relations. The proposed model achieved a score of 97.3 on the H@10 evaluation index with respect to the prediction of Chinese hypernym–hyponym relations. In future work, we will investigate the extent to which the proposed method can be improved by incorporating embedding using multilingual models such as Multilingual BERT and XLM-RoBERTa.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers 22K12154 and 20H04210.

References

[1] Hearst, M. A., Automatic Acquisition of Hyponyms from Large Text Corpora, in Proceedings of the 14th International Conference on Computational Linguistics, pp. 539–545, 1992.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., Generative Adversarial Nets. Proceedings of the 27th International Conference on Neural Information Processing Systems, pp. 2672–2680, 2014.

[3] Huang, Z., Li, Z., Jiang, H., Cao, T., Lu, H., Yin, B., Subbian, K., Sun, Y., and Wang, W., Multilingual Knowledge Graph Completion with Self-Supervised Adaptive Graph Alignment, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, pp. 474–485, 2022.

[4] Chen, X., Chen, M., Fan, C., Uppunda, A., Sun, Y. and Zaniolo, C., Multilingual Knowledge Graph Completion via Ensemble Knowledge Transfer, In Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 3227–3228, 2020.

[5] Dettmers, D., Minervini, P., Stenetorp, P., and Riedel, S., Convolutional 2D Knowledge Graph Embedding, Proceedings of the 32th AAAI Conference on Artificial Intelligence, 2018.

[6] Hamilton, W.L., Ying, R. and Leskovec, J., Inductive Representation Learning on Large Graphs, Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 1025–1035, 2017.

[7] Nanba, H., Mayumi, S., and Takezawa, T. Automatic Construction of a Bilingual Thesaurus using Citation Analysis. Proceedings of the 4th International CIKM Workshop on Patent Information Retrieval (PaIR'11), pp. 25–30, 2011.

[8] Yu, C., Han, J., Zhang, H. and Ng, W., Hypernymy Detection for Low-Resource Languages via Meta Learning. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 3651–3656, 2020.

[9] Upadhyay, S., Vyas, Y., Carpuat, M. and Roth, D., Robust Cross-lingual Hypernymy Detection using Dependency Context. The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 607–618, 2018.

[10] Kober, T., Weeds, J., Bertolini, L. and Weir, D., Data Augmentation for Hypernymy Detection. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, pp. 1034–1048, 2021.

[11] Chavdarova, T. and Fleuret, F., SGAN: An Alternative Training of Generative Adversarial Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, pp. 9407–9415, 2018.

[12] Bordes, A., Usunier, N., García-Durán, A., Weston, J. and Yakhnenko, O., Translating Embeddings for Modeling Multi-relational Data. In Proceedings of NIPS, pp. 2787–2795, 2013.

[13] Yang, B., Yih, W., He, X., Gao, J, and Deng, L., Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In Proceedings of ICLR 2015, 2015.

[14] Trouillon, T., Welbl, J., Riedel, S., Gaussier, É. and Bouchard, G., Complex Embeddings for Simple Link Prediction. In Proceedings of ICML 2016, pp. 2071–2080, 2016.

[15] Kipf, T.N. and Welling, M., Semi-supervised Classification with Graph Convolutional Networks. arXiv preprint arXiv:1609.02907, 2016.

[16] Schlichtkrull, M., Kipf, T. N., Bloem, P., Berg, R. v. d., Titov, I. and Welling, M., Modeling Relational Data with Graph Convolutional Networks. arXiv preprint arXiv:1703.06103, 2017.

[17] Jiang, X., Wang, Q. and Wang, B., Adaptive Convolution for Multi-relational Learning, In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Vol. 1, pp. 978–987, 2019.

[18] Perozzi, B., Al-Rfou, R. and Skiena, S., DeepWalk: Online Learning of Social Representations. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710, 2014.