Cross-lingual Short-text Entity Linking: Generating Features for Neuro-Symbolic Methods

Qiuhao Lu[1], **Sairam Gurajada**[2], **Prithviraj Sen**[2], **Lucian Popa**[2], **Dejing Dou**[1,3], **and Thien Huu Nguyen**[1]
[1]Department of Computer Science, University of Oregon, Eugene, OR, USA
[2]IBM Research
[3]Baidu Research
{luqh, dou, thien}@cs.uoregon.edu
gu.sairam@gmail.com
{senp, lpopa}@us.ibm.com

Abstract

Entity linking (EL) on short text is crucial for a variety of industrial applications. Compared with general long-text EL, short-text EL poses particular challenges as the limited context restricts the clues one can leverage to disambiguate textual mentions. On the other hand, existing studies mostly focus on black-box neural methods and thus lack interpretability, which is critical to industrial applications in certain areas. In this study, we extend upon LNN-EL, a monolingual short-text EL method based on interpretable first-order logic (Jiang et al., 2021), by incorporating three sets of multilingual features to enable disambiguating mentions written in languages other than English. More specifically, we use multilingual autoencoding language models (i.e., mBERT) to capture the similarities between the mention with its context and the candidate entity; we use multilingual sequence-to-sequence language models (i.e., mBART and mT5) to represent the likelihood of the text given the candidate entity. We also propose a word-level context feature to capture the semantic evidence of the co-occurring mentions. We evaluate the proposed xLNN-EL approach on the QALD-9-multilingual dataset and demonstrate the cross-linguality of the model and the effectiveness of the features.

1 Introduction

Entity linking (EL), also known as entity disambiguation, is the task of linking textual mentions appearing in a document to the corresponding entities associated with a target knowledge base like DBpedia (Auer et al., 2007). As a fundamental task in natural language processing and information extraction, entity linking is crucial for a variety of applications such as semantic search, recommendation systems and chatbots (Tan et al., 2017).

Historically, relevant studies of entity linking mostly focus on long text scenario (i.e., documents) (Han et al., 2011; Gupta et al., 2017; Lu and Du, 2017; Cao et al., 2018; Kolitsas et al., 2018). Typically, such approaches mainly depend on specifically-designed features of candidates (e.g., priors), local context compatibility, and global coherence across the document (Shen et al., 2021).

With the rapidly growing short text on the web, e.g., search queries, social media posts, news headlines, etc., short-text entity linking has attracted increasing attention from researchers due to its potential for various industrial applications. However, long-text entity linking methods barely maintain the same level of performance on short text, as they heavily rely on document-level global coherence, i.e., the idea of *collective entity linking* (Cao et al., 2018), and short text (e.g., a single sentence or search query) cannot provide rich context and global signals for disambiguation (Chen et al., 2018).

To tackle the task of short-text entity linking, different methods are being proposed to exploit the short and limited context (Chen et al., 2018; Sakor et al., 2019; Gu et al., 2021). For example, Chen et al. (Chen et al., 2018) try to map the sparse short text to a topic space such that a topic coherence can be achieved through a specifically designed optimization objective, e.g., they aim to infer the topic is *literature* against *movie* from the word `read` in "read Harry Potter". Gu et al. (Gu et al., 2021) try to enhance the interactions between the local context and the candidate entity in multi-turn multiple-choice manner, so that global disambiguation can finally be achieved. More recently, autoregressive entity linking models, e.g., GENRE (De Cao et al., 2020) and mGENRE (De Cao et al., 2021b) demonstrate superior performance and are gaining a lot of attention. With the power of large pre-trained sequence-to-sequence language models (Lewis et al., 2020), GENRE and mGENRE are able to directly generate unique entity names conditioned on the context, yielding state-of-the-art

Proceedings of the Fourth Workshop on Data Science with Human-in-the-Loop (Language Advances), pages 8 - 14

results on multiple datasets.

Recently, Jiang et al. propose LNN-EL, the first neuro-symbolic short-text entity linking method that combines first-order logic (FOL) rules with neural learning, and shows competitive performance against deep-learning-based black-box methods (Jiang et al., 2021). Essentially, LNN-EL uses FOL rules as a glue to combine different features into a final linking model, and demonstrates *interpretability* due to the interpretable nature of FOL. Nonetheless, LNN-EL is a monolingual model that only supports English, and does not meet the demands of modern industrial applications where enterprises need to establish effective cross-language interactions with users from all over the world.

To enable this approach to facilitate rapidly growing global business, in this study, we extend upon LNN-EL by incorporating three sets multilingual features to make it cross-lingual, i.e., xLNN-EL. First, we propose a word-level cross-lingual context feature that aims to capture the semantic evidence of co-occurring mentions. Second, we use multilingual autoencoding language models (i.e., mBERT (Pires et al., 2019)) to capture the similarities between the mention with its context and the candidate entity with a four-way feature. Third, we use multilingual sequence-to-sequence language models (i.e., mBART (Liu et al., 2020) and mT5 (Xue et al., 2021)) to represent the likelihood of the text given the candidate entity. More specifically, we try to reconstruct the text conditioned on the candidate's description, based on a fine-tuned generative language model, and it results in another two-way feature indicating the likelihood of the text. We evaluate xLNN-EL on the QALD-9-multilingual dataset with the state-of-the-art black-box method mGENRE (De Cao et al., 2021b), and the results demonstrate the effectiveness of the proposed features. Our contribution can thus be summarized as follows:

- We extend upon LNN-EL to facilitate disambiguating mentions appearing in non-English languages. To the best of our knowledge, xLNN-EL is the first neuro-symbolic method for cross-language short-text entity linking and performs competitively against state-of-the-art black-box method.

- We propose three sets of multilingual features that aim to capture the contextual semantic evidence of co-occurring mentions, the similarities between the mention with its context and the candidate, and the likelihood score of the context conditioned on the candidate. The experimental results show the effectiveness of the features.

2 xLNN-EL

We propose xLNN-EL, a cross-language extension of LNN-EL, where we seek to facilitate the model with better cross-linguality by incorporating a set of new features allowing it to link non-English mentions to the English knowledge base.

Following LNN-EL, we take the *English DBpedia* as the target knowledge base (Jiang et al., 2021). As to candidate retrieval, we adopt a hybrid method of PivotsCR (Liu et al., 2021) and mGENRE (De Cao et al., 2021b). Essentially, for each mention m_i, we take the union of their outputs and generate a set of $|C_i| = 250$ candidate entities, roughly reaching a 95% recall rate for each language[1].

Formally, given a single sentence or search query T containing a set of mentions $M = \{m_1, m_2, \ldots, m_p\}$, a triple m_i, C_i, L_i is generated for each mention m_i. L_i is a list of binary labels for the mention-candidate pair (m_i, e_{ij}) where $e_{ij} \in C_i$. The entity name and textual description of e_{ij} is denoted by $e_{ij}.name$ and $e_{ij}.desc$, respectively. For each candidate e_{ij}, a set of predefined features $f_w(m_i, e_{ij})$ is generated. In this section, we introduce the three cross-language features $f_w \in \mathcal{F}$ that we incorporate into the model.

2.1 Word-level Context Score

Given a mention, we introduce a word-level context score to capture the semantic evidence sustained by the similarity between the co-occurring mentions and the descriptions of the candidate entities, i.e., short abstracts[2]. The feature function f_{ctx} is defined as follows:

$$\begin{aligned} f_{\text{ctx}}(m_i, e_{ij}) &= \sum_{m_k \in M \setminus \{m_i\}} pr_{\text{word}}(m_k, e_{ij}) \\ pr_{\text{word}}(m_k, e_{ij}) &= \max_{s_k \in e_{ij}.desc} \cos(\mathbf{m_k}, \mathbf{s_k}) \end{aligned} \tag{1}$$

[1]LNN-EL uses DBpedia lookup to retrieve top-100 candidates.

[2]http://downloads.dbpedia.org/wiki-archive/downloads-2016-10.html

Method F1-score(%)	mono-lingual de	 fr	 en	 it	multi-lingual de	 fr	 en	 it	cross-lingual de	 fr	 en	 it
mGENRE	47.50	47.50	62.50	50.83	57.50	54.17	56.67	54.17	55.00	57.50	50.83	52.50
base	58.82	62.61	83.76	58.82	58.82	60.92	83.76	57.14	61.34	60.50	80.08	57.14
base + ctx	61.34	62.61	83.76	57.14	60.50	61.34	84.60	57.14	60.50	63.03	83.05	55.46
base + mbert	61.34	62.61	84.60	56.72	59.66	61.76	83.76	56.30	60.50	60.92	80.08	56.30
base + generative	60.50	63.45	84.60	62.04	61.34	61.76	83.76	58.40	60.50	61.76	82.20	58.40
base + ctx + mbert	63.03	60.92	83.76	57.14	59.66	62.61	84.60	57.14	59.66	61.34	80.08	56.30
base + ctx + generative	**66.39**	62.61	84.60	59.66	60.78	61.34	84.60	57.14	60.08	63.03	82.63	57.56
base + mbert + generative	63.45	**69.75**	84.60	**63.03**	62.18	**63.87**	83.76	57.98	**64.99**	**63.87**	**83.47**	57.98
base + ctx + mbert + generative	65.13	64.29	**85.45**	60.50	**62.61**	62.18	**85.45**	**58.82**	61.34	62.18	83.05	**59.66**

Table 1: Performance of xLNN-EL on QALD-9-multilingual.

where pr_{word} is a word-level *Partial Ratio*[3] score. The idea is to find the maximum similarity between the mention m_k and any group of words of the same length s_k, i.e., a sliding window, in the candidate's textual description. We use fastText's pre-trained aligned word vectors[4] (Bojanowski et al., 2017; Joulin et al., 2018) to encode the mention and the entity description as they are in different languages in the cross-language setting, i.e., the vector representations $\mathbf{m_k}, \mathbf{s_k}$ are in the same embedding space though m_k are non-English and s_k are English. We then take the aggregated similarities as the feature indicating the semantic contextual relevance of the candidate.

2.2 Autoencoding-LM-based Scores

We also introduce a set of autoencoding-language-model-based features to encode the the overall similarities between the mention with its context and the candidate entity. In particular, autoencoding language models (e.g., BERT (Devlin et al., 2019)) create a bidirectional representation of the whole sentence which makes them a natural fit as text encoders, and the representations can be further facilitate discriminative downstream applications. Due to the *extensibility* of the framework in LNN-EL (Jiang et al., 2021), we are able to include various features to describe the relationship between the mention and the candidate from different levels and aspects. For this feature, we use the multilingual version of BERT, i.e., mBERT (Pires et al., 2019), as the language model, and the feature function f_{mbert} is defined as follows:

$$\begin{aligned} f_{\text{mbert}} &= \left[f_{\text{mbert}_1}, f_{\text{mbert}_2}, f_{\text{mbert}_3}, f_{\text{mbert}_4}\right] \\ &= \left[\cos(\mathbf{m_i}, \mathbf{e_{ij}.name}), \cos(\mathbf{T}, \mathbf{e_{ij}.name}),\right. \\ &\quad \left.\cos(\mathbf{m_i}, \mathbf{e_{ij}.desc}), \cos(\mathbf{T}, \mathbf{e_{ij}.desc})\right] \end{aligned} \tag{2}$$

where the bold fonts indicate the vector representations for the mention ($\mathbf{m_i}$), the input short text ($\mathbf{T}$), the candidate's name ($\mathbf{e_{ij}.name}$), and the candidate's description ($\mathbf{e_{ij}.desc}$), respectively. Essentially, the text is sent to mBERT and the vector representation for the `[CLS]` token is used. This four-way feature aims to explore the possibility of capturing the similarities between the input text and the candidate entity from different aspects, under the xLNN-EL framework. The experimental results in Section 3 show that they are useful and complementary.

2.3 Seq2Seq-LM-based Scores

Sequence-to-sequence language models (e.g., BART (Lewis et al., 2020)) are mostly adopted for tasks like translation, summarization and question answering. For short-text cross-lingual entity linking, however, this direction has been underexplored.

In this study, we propose to leverage seq2seq language models to reveal another aspect of similarity between the mention and the candidate entity, leveraging a set of features to reflect the likelihood of the context, given the candidate entity for each mention. Essentially, we first fine-tune multilingual generative models $\mathcal{G}$ (i.e., mBART (Liu et al., 2020) and mT5 (Xue et al., 2021) in this paper) on the training set of <`description, sentence`> pairs, where the `description` refers to the candidate entity's textual description and the `sentence` refers to the input short text, i.e., $< e_{ij}.desc, T >$. The idea is to enable the models to generate a short sentence conditioned on a textual description of a

[3] pypi.org/project/py-stringmatching
[4] https://fasttext.cc/docs/en/aligned-vectors.html

candidate entity. With a fine-tuned model $\mathcal{G}_{tuned}$, we assign a likelihood score to each candidate. The feature function $f_{\text{generative}}$ is defined as follows:

$$\begin{aligned} f_{\text{generative}} &= [f_{\text{mbart}}, f_{\text{mt5}}] \\ &= [\sum_{k}^{|T|} \log p_{\theta_{\text{mbart}}}(x_k | x_{<k}, e_{ij}.desc), \\ &\quad \sum_{k}^{|T|} \log p_{\theta_{\text{mt5}}}(x_k | x_{<k}, e_{ij}.desc)] \end{aligned} \tag{3}$$

where $T = (x_0, x_1, \ldots, x_{|T|})$ is the short text and $\log p_{\theta_{\text{mbart}}}(x_k|x_{<k})$, $\log p_{\theta_{\text{mt5}}}(x_k|x_{<k})$ are the log-likelihoods of the k-th token conditioned on the preceding tokens based on the fine-tuned models $\mathcal{G}_{\text{mbart}}$ and $\mathcal{G}_{\text{mt5}}$, respectively. This two-way generative feature serves to reflect the feasibility of the whole sentence given the candidate entity, without a special focus on the mention.

3 Experiment

3.1 Setup

We evaluate xLNN-EL on the QALD-9-multilingual dataset[5]. For a fair comparison, we take the German (de), French (fr), English (en), and Italian (it) versions of this dataset, as the other languages are incomplete comparing to the English version. The evaluation is conducted in three settings: **mono-lingual** refers to in-language training and testing; **multi-lingual** means training on the union of all languages and testing on individual languages; and **cross-lingual** means testing on one language while training on the other three. We compare the proposed model xLNN-EL with the state-of-the-art black-box method mGENRE (De Cao et al., 2021b) for entity linking. In addition, we also evaluate the adapted version of LNN-EL with basic features in the cross-language scenario, denoted by `base`.

3.2 Results

The results are shown in Table 1. We present different combinations of the proposed features in the table. xLNN-EL with the `base` feature set shows better performance than mGENRE, most likely due to the limitation that such methods (GENRE/mGENRE) need large amounts of data for adequate training (De Cao et al., 2021a). We observe that with the three proposed sets of features, the performance gets boosted across all settings and languages and consistently outperforms mGENRE and `base`, the state-of-the-art neuro-symbolic short-text EL system, indicating the effectiveness of these features. We also notice that different languages have their own feature patterns, e.g., the context score seems more beneficial for German than for French, according to their performance in the mono-lingual and multi-lingual settings, and the language-specific feature patterns indicate a direction of future work. The impact of the language-model-based features, i.e., `mbert` and `generative`, is reflected in the last two rows of the table, where the performance reaches its peak when both features are included, thus demonstrating their importance as well as their complementary nature. Furthermore, the cross-lingual performance is on par with that in the multi-lingual setting, and that shows the transferability of the proposed features, reflecting a potential for real-world industrial cross-language scenarios.

[5]https://github.com/ag-sc/QALD/tree/master/9/data

Method	mono-lingual				
F1-score(%)	de	fr	en	it	Δ
all	65.13	64.29	85.45	60.50	-
- mbert 1	65.55	62.61	84.60	57.56	-1.26
- mbert 2	63.03	62.61	82.91	58.82	-2.00
- mbert 3	60.50	61.34	84.60	57.98	-2.74
- mbert 4	60.50	60.92	84.60	57.98	-2.84
- mbart	64.29	61.76	86.30	58.82	-1.05
- mt5	63.87	62.18	84.60	57.14	-1.89

Table 2: Ablation study.

3.3 Ablation Study

To better understand the contribution of each component in the `mbert` and `generative` feature, we present in Table 2 the results of the ablation results of the model `base + ctx + mbert + generative` in the mono-lingual setting, with their average performance change (Δ). As shown in the table, dropping each score of the LM-based features will cause the performance to decrease greatly, indicating the effectiveness and necessity of them.

4 Conclusion

In this study, we extend upon LNN-EL by incorporating three sets of multilingual features to enable disambiguating mentions written in languages other than English. This study also indicates directions of future work, as the results demonstrate language-specific patterns for the features and rules.

Acknowledgements

This research has been supported by the Army Research Office (ARO) grant W911NF-21-1-0112 and the NSF grant CNS-1747798 to the IUCRC Center for Big Learning. This research is also based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2019-19051600006 under the Better Extraction from Text Towards Enhanced Retrieval (BETTER) Program. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ARO, ODNI, IARPA, the Department of Defense, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. We also would like to thank the IBM Research - Almaden team for their support in this work.

References

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *Proceedings of the 6th International The Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference*, ISWC'07/ASWC'07, page 722–735, Berlin, Heidelberg. Springer-Verlag.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Yixin Cao, Lei Hou, Juanzi Li, and Zhiyuan Liu. 2018. Neural collective entity linking. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 675–686, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Lihan Chen, Jiaqing Liang, Chenhao Xie, and Yanghua Xiao. 2018. Short text entity linking with fine-grained topics. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 457–466.

Marina Danilevsky, Kun Qian, Ranit Aharonov, Yannis Katsis, Ban Kawas, and Prithviraj Sen. 2020. A survey of the state of explainable AI for natural language processing. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 447–459, Suzhou, China. Association for Computational Linguistics.

Nicola De Cao, Wilker Aziz, and Ivan Titov. 2021a. Highly parallel autoregressive entity linking with discriminative correction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7662–7669, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Nicola De Cao, Gautier Izacard, Sebastian Riedel, and Fabio Petroni. 2020. Autoregressive entity retrieval. In *International Conference on Learning Representations*.

Nicola De Cao, Ledell Wu, Kashyap Popat, Mikel Artetxe, Naman Goyal, Mikhail Plekhanov, Luke Zettlemoyer, Nicola Cancedda, Sebastian Riedel, and Fabio Petroni. 2021b. Multilingual autoregressive entity linking. *arXiv preprint arXiv:2103.12528*.

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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Francesc Esteva and Lluıs Godo. 2001. Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy sets and systems*, 124(3):271–288.

Yingjie Gu, Xiaoye Qu, Zhefeng Wang, Baoxing Huai, Nicholas Jing Yuan, and Xiaolin Gui. 2021. Read, retrospect, select: An mrc framework to short text entity linking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12920–12928.

Nitish Gupta, Sameer Singh, and Dan Roth. 2017. Entity linking via joint encoding of types, descriptions, and context. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2681–2690, Copenhagen, Denmark. Association for Computational Linguistics.

Xianpei Han, Le Sun, and Jun Zhao. 2011. Collective entity linking in web text: a graph-based method. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 765–774.

Hang Jiang, Sairam Gurajada, Qiuhao Lu, Sumit Neelam, Lucian Popa, Prithviraj Sen, Yunyao Li, and Alexander Gray. 2021. LNN-EL: A neuro-symbolic approach to short-text entity linking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 775–787, Online. Association for Computational Linguistics.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. End-to-end neural entity linking. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 519–529, Brussels, Belgium. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Qian Liu, Xiubo Geng, Jie Lu, and Daxin Jiang. 2021. Pivot-based candidate retrieval for cross-lingual entity linking. In *Proceedings of the Web Conference 2021*, pages 1076–1085.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Qiuhao Lu and Youtian Du. 2017. Wikipedia-based entity semantifying in open information extraction. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 765–770. IEEE.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Ryan Riegel, Alexander Gray, Francois Luus, Naweed Khan, Ndivhuwo Makondo, Ismail Yunus Akhalwaya, Haifeng Qian, Ronald Fagin, Francisco Barahona, Udit Sharma, et al. 2020a. Logical neural networks. *arXiv preprint arXiv:2006.13155*.

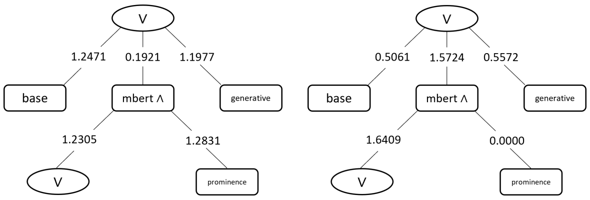
Ryan Riegel, Francois Luus, and Alexander Gray. 2020b. Logical neural networks: From deep learning to deep thinking.

Ahmad Sakor, Isaiah Onando Mulang, Kuldeep Singh, Saeedeh Shekarpour, Maria Esther Vidal, Jens Lehmann, and Sören Auer. 2019. Old is gold: linguistic driven approach for entity and relation linking of short text. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2336–2346.

Wei Shen, Yuhan Li, Yinan Liu, Jiawei Han, Jianyong Wang, and Xiaojie Yuan. 2021. Entity linking meets deep learning: Techniques and solutions. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1.

Chuanqi Tan, Furu Wei, Pengjie Ren, Weifeng Lv, and Ming Zhou. 2017. Entity linking for queries by searching Wikipedia sentences. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 68–77, Copenhagen, Denmark. Association for Computational Linguistics.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.



(a) German (b) French

Figure 1: Feature weights for German and French.

A Preliminaries and Interpretabiltiy

A.1 LNN-EL

First-order logic (FOL) rules form a closed language facilitating the expression of a variety of human interpretable models. To learn these rules, neuro-symbolic AI typically substitutes conjunctions with t-norms (Esteva and Godo, 2001) which actually limits their learning capacity as these norms do not have learnable parameters.

Recently, Riegel et al. propose Logical Neural Networks (LNN) (Riegel et al., 2020a), a modification of neural networks that can precisely model operations in real-valued logic, i.e., they construct logical operators conjunction ($\wedge$) and disjunction ($\vee$) and facilitate neural network-style learning with learnable parameters (Riegel et al., 2020b).

Subsequently, LNN-EL reformulates entity linking by mapping the Boolean-valued logic rules into the LNN formalism and the resulting model consists of parameterized LNN operators, i.e., conjunction ($\wedge$) and disjunction ($\vee$), along with learnable rule weights and feature weights. LNN-EL takes as input the pre-computed features of candidates and the definition for the features is the main focus of this study. We refer the readers to (Riegel et al., 2020a; Jiang et al., 2021) for more detailed treatment of the parameterized LNN opertors and the reformulation.

A.2 Interpretability of xLNN-EL

A common theme among existing EL methods is their lack of *interpretability*. Interpretability is an important and desirable property not only for machine learning research, but also for real-world downstream applications, especially for sensitive areas. In fact, there is a growing trend towards developing interpretable machine learning models (Danilevsky et al., 2020).

We show some learned feature weights of the `base + mbert + generative` model for German and French, in the mono-lingual setting, to illustrate the interpretability of xLNN-EL. The tree structure reflects the first-order logic (FOL) rule combination of the model, e.g., the three features `base`, `mbert` and `generative` are combined with a disjunction at the topmost level, the `mbert` feature is formed with a conjunction between a set of disjunct similarities and the `prominence` feature, the `generative` feature has the same substructure with `mbert` (not shown), etc. The rule combination part is beyond the scope of this paper and the reader is referred to the literature for further details (Jiang et al., 2021).

As shown in Figure 1, the feature `mbert` has a much higher relative feature weight for French than for German (1.5724 vs. 0.1921) in the disjunction, which might indicate a relative preference as to the `mbert` feature for French. This inspection provides clues for human experts to understand how these features impact performance, and further adjust features and rule combinations accordingly.