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

Qiuhao Lu, Sairam Gurajada, Prithviraj Sen, Lucian Popa, Dejing Dou, Thien Huu Nguyen

University of Oregon

IBM Research

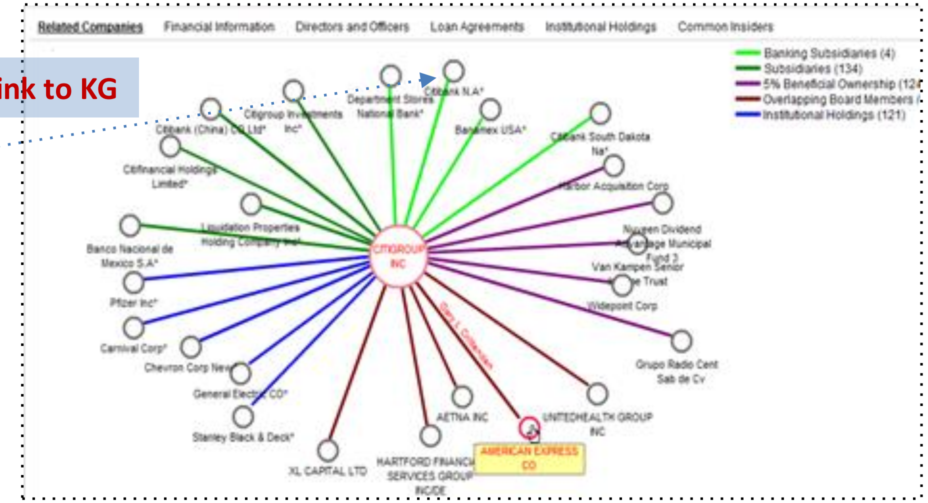
Entity Linking

Extract / not in KB

Extract / link to KG

Citigroup Global Markets Inc. ("CGMI"), an affiliate of **Citibank**, may place Deposits directly and through brokers.

A domain-specific KG (Finance)



Problem

- Given text and knowledge graph (KG), extract and link relevant pieces of text to entities/concepts in the KG.
  - Domain-specific, deep KGs: IT domain, Finance
  - General-domain KGs: Dbpedia, Wikidata, etc.

LNN-EL: neuro-symbolic disambiguation of entities

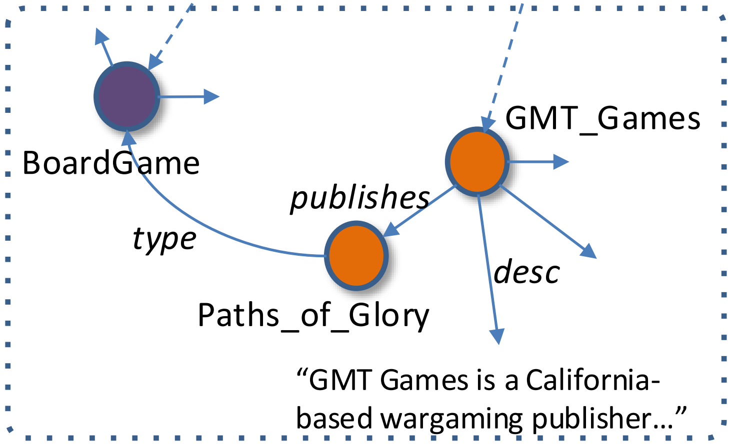
- Ensemble of rules.
- Combined with LNN to learn "best" thresholds for similarity features, weights of predicates.

- SotA on three KBQA benchmark datasets (ACL'21 paper)
- Part of larger NSQA pipeline

"List all boardgames by GMT"

Entity linking (**nominal entities**)

Entity linking (**named entities**)



Knowledge Graph

DBpedia

WIKIDATA

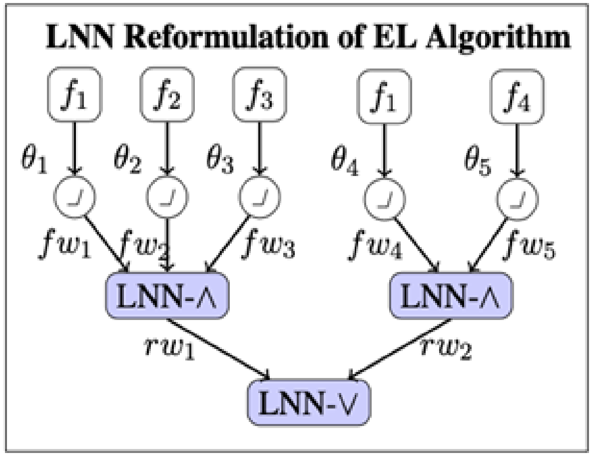
Rule-based Entity Disambiguation with LNN-EL

User provided EL Algorithm

$$R_1(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_1 \land f_2(m_i, e_{ij}) > \theta_2 \land f_3(m_i, e_{ij}) > \theta_3$$

$$\lor$$

$$R_2(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_4 \land f_4(m_i, e_{ij}) > \theta_5$$



Learnable parameters:
θ_i– feature thresholds,
fw_i– feature weights,
rw_i– rule weights

Based on a learnable real-valued logic framework (LNN):
[R. Riegel et al. ***Logical Neural Networks***, 2020]

Extensible space of features:

Non-embedding based

f_i: - *string similarity* (Jaccard, JaroWinkler, Levenshtein, etc.)
- *candidate importance* score (e.g., refCount)
- *type* similarity
…

Embedding based

f_i : - BERT, Wiki2Vec
- Query2Box embeddings,
- scores of prior EL methods (e.g., BLINK scores)

Advantages of rules:

- **Interpretable, extensible** and **customizable**
- **Good transfer** to other datasets

xLNN-EL: Multi-lingual Embedding based Features

Extensible space of features in LNN-EL:

Non-embedding based

f_i : - *string similarity* (Jaccard, JaroWinkler, Levenshtein, etc.)
- *candidate importance* score (e.g., refCount)
- *type* similarity

...

Embedding based

f_i : - BERT, Wiki2Vec
- Query2Box embeddings,
- scores of prior EL methods (e.g., BLINK scores)

$$f_{\text{ctx}}(m_i, e_{ij}) = \sum_{m_k \in M \setminus \{m_i\}} pr_{\text{word}}(m_k, e_{ij})$$

$$pr_{\text{word}}(m_k, e_{ij}) = \max_{s_k \in e_{ij}.desc} \cos(\mathbf{m_k}, \mathbf{s_k})$$

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

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

xLNN-EL: Multi-lingual Embedding based Features

Word-level Context Score:

- Word-level Partial Ratio score
  - The idea is to find the maximum similarity between the mention and any group of words of the same length in the candidate's textual description
- FastText's pre-trained aligned word vectors

$$f_{\text{ctx}}(m_i, e_{ij}) = \sum_{m_k \in M \setminus \{m_i\}} pr_{\text{word}}(m_k, e_{ij})$$

$$pr_{\text{word}}(m_k, e_{ij}) = \max_{s_k \in e_{ij}.desc} \cos(\mathbf{m_k}, \mathbf{s_k})$$

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

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

xLNN-EL: Multi-lingual Embedding based Features

Autoencoding-LM-based Scores:

- mBERT embeddings
- Four-way feature
  - Mention, context, entity name, entity description

$$f_{\text{ctx}}(m_i, e_{ij}) = \sum_{m_k \in M \setminus \{m_i\}} pr_{\text{word}}(m_k, e_{ij})$$

$$pr_{\text{word}}(m_k, e_{ij}) = \max_{s_k \in e_{ij}.desc} \cos(\mathbf{m_k}, \mathbf{s_k})$$

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

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

xLNN-EL: Multi-lingual Embedding based Features

Seq2Seq-LM-based Scores:

- Fine-tune mBART and mT5 to enable the models to generate short text conditioned on an entity description
- Assign a likelihood score to each candidate with tuned models
- Two-way feature
  - feasibility of the whole sentence given the candidate

$$f_{\text{ctx}}(m_i, e_{ij}) = \sum_{m_k \in M \setminus \{m_i\}} pr_{\text{word}}(m_k, e_{ij})$$

$$pr_{\text{word}}(m_k, e_{ij}) = \max_{s_k \in e_{ij}.desc} \cos(\mathbf{m_k}, \mathbf{s_k})$$

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

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

Overall Performance

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.

Thank you!