LNN-EL: A Neuro-Symbolic Approach for Short-text Entity Linking

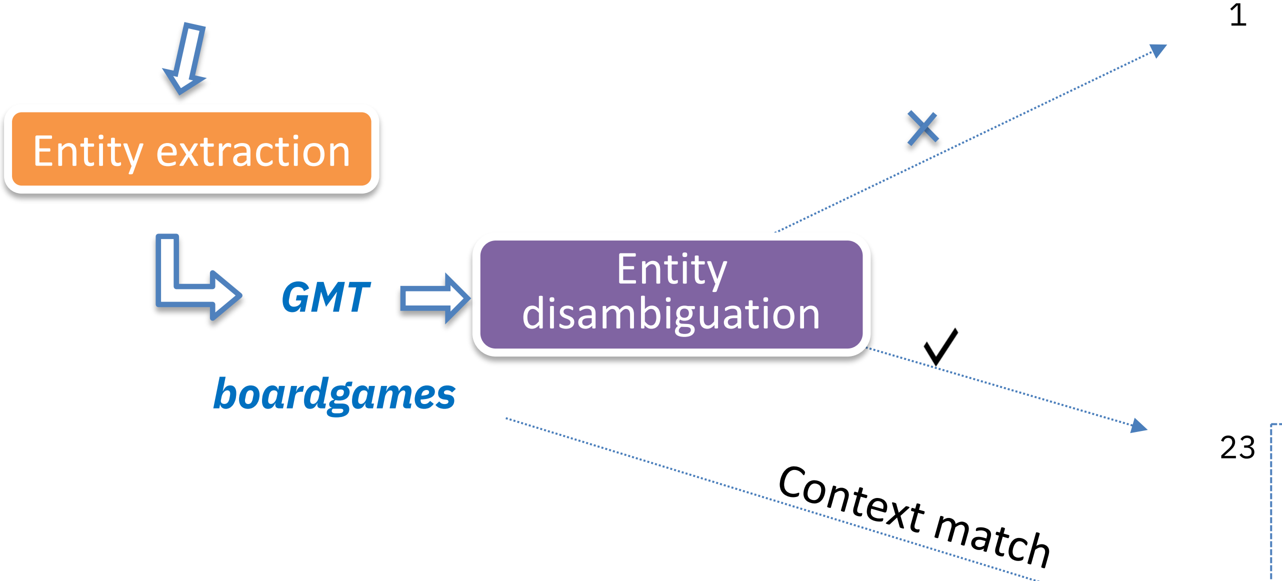
Hang Jiang, Sairam Gurajada, Qiuhao Lu, Sumit Neelam,
Lucian Popa, Prithviraj Sen, Yunyao Li, and Alexander Gray

Massachusetts Institute of Technology

IBM Research

Entity Linking on Short Text

Ex: ***"List all boardgames by GMT"***



http://lookup.dbpedia.org/api/search/KeywordSearch?QueryString=%22GMT%22&MaxHits=100

1 <Label>Greenwich Mean Time</Label>
<URI>http://dbpedia.org/resource/Greenwich_Mean_Time</URI>
<Description>Greenwich Mean Time (GMT) is a time system originally referring to mean solar time at the Royal Observatory in Greenwich, London, which later became adopted as a global time standard. It is arguably the same as Coordinated Universal Time (UTC) and when this is viewed as a time zone the name Greenwich Mean Time is especially used by bodies connected with the United Kingdom, such as the BBC World Service, the Royal Navy, the Met Office and others.</Description>

...

23 <Label>**GMT Games**</Label>
<URI>http://dbpedia.org/resource/**GMT_Games**</URI>
<Description>**GMT Games**, probably the most prolific of the **wargame** companies in the 1990s and 2000s, was founded in 1990. The current management and creative team includes Tony Curtis, Rodger MacGowan, Mark Simonitch, and Andy Lewis. The company has become well known for graphically attractive **games** that range from "monster **games**", of many maps and counters, to quite simple **games** suitable for introducing new players to wargaming. They also produce card **games** and family **games**.</Description>

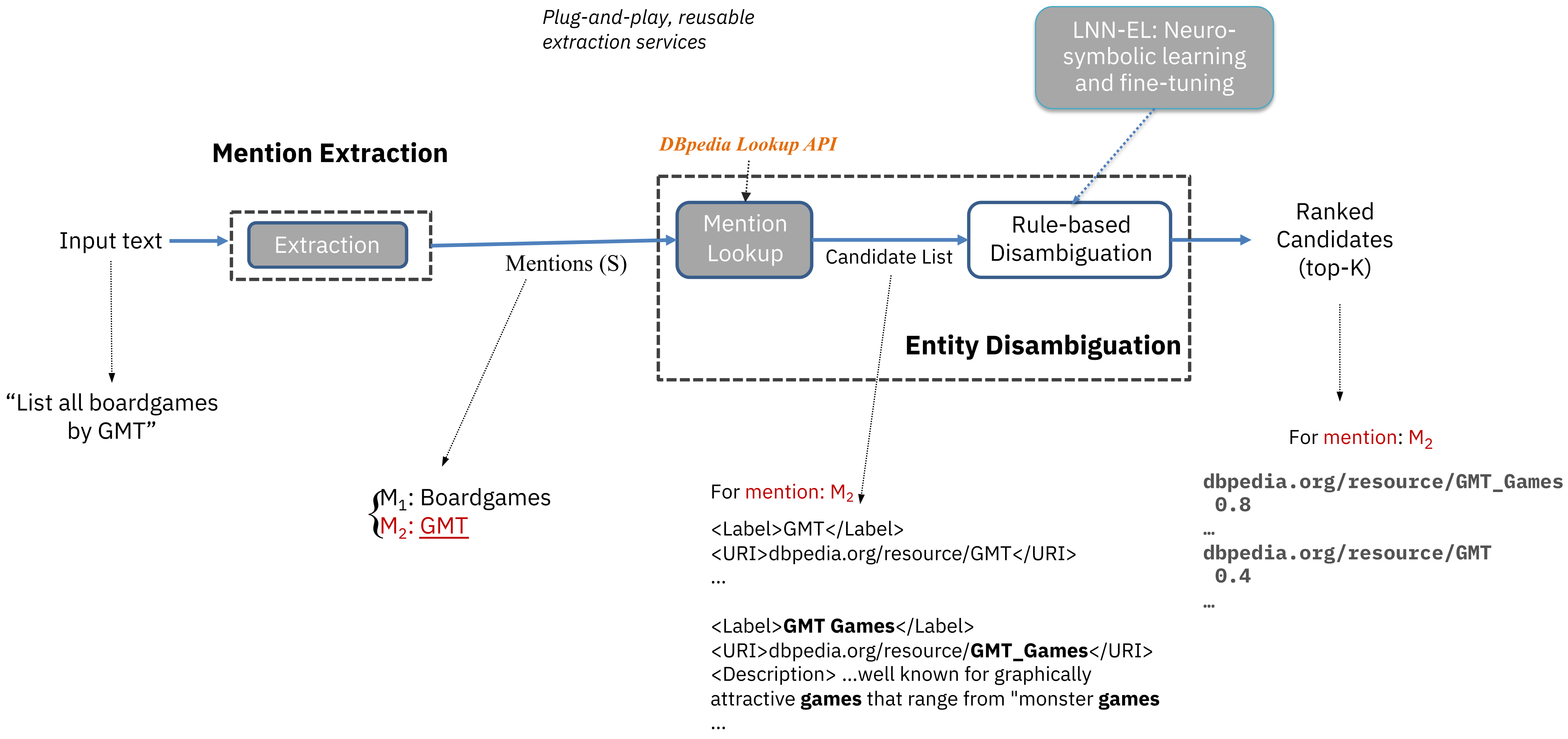
Why is it important:

- Better understanding of text
- Question-answering

Challenges

- Needs good exploitation of the context:
    - To match GMT with confidence, need available clues: co-occurring entities (e.g., "boardgames"), relationships, sentence, KB text and graph

Entity Linking Approach

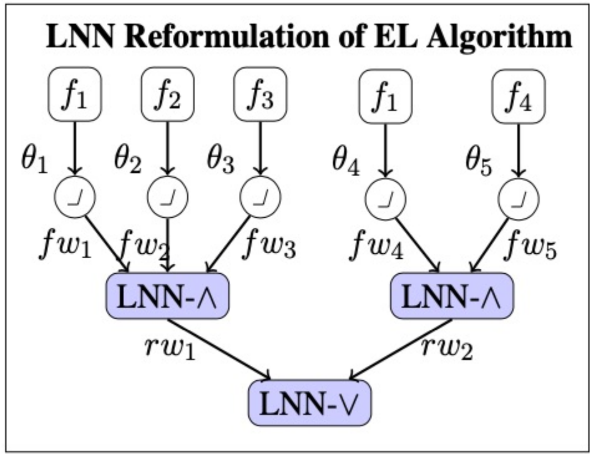


Rule-based Entity Disambiguation with LNN-EL

User provided EL Algorithm

$$R_1(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_1 \wedge f_2(m_i, e_{ij}) > \theta_2 \wedge f_3(m_i, e_{ij}) > \theta_3$$
$$\vee$$
$$R_2(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_4 \wedge f_4(m_i, e_{ij}) > \theta_5$$

Symbolic Rules



Neural Learning

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

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)

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

Rule-based Entity Disambiguation with LNN-EL

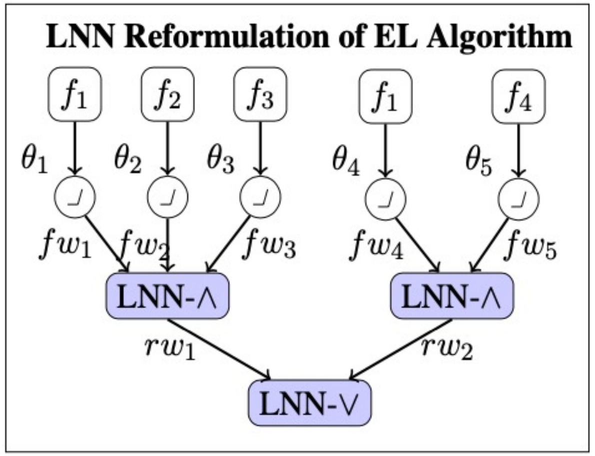
User provided EL Algorithm

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

$$\vee$$

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

Symbolic Rules



Neural Learning

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

Advantages of LNN-EL

1. intepretable: expressive FOL language
2. extensible
3. transferable

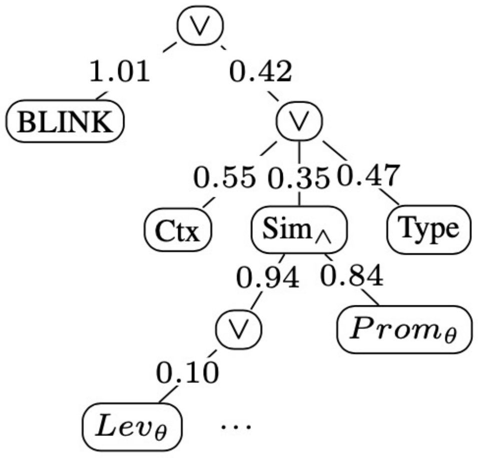
Based on a learnable real-valued logic framework (LNN):
[R. Riegel et al. ***Logical Neural Networks***, 2020]

Entity Linking Performance on Benchmark Datasets

	Model	LC-QuAD			QALD-9			$\text{WebQSP}_{\text{EL}}$		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Baselines	BLINK	87.04	87.04	87.04	89.14	89.14	89.14	92.15	92.05	92.10
	BERT	57.14	63.09	59.97	55.46	61.11	58.15	70.26	72.15	71.20
	BERTWiki	66.96	73.85	70.23	66.16	72.90	69.37	81.11	83.29	82.19
	Box	67.31	74.32	70.64	68.91	75.93	72.25	81.53	83.72	82.61
	LogisticRegression	87.04	86.83	86.93	84.73	84.73	84.73	83.39	83.33	83.36
	$LogisticRegression_{BLINK}$	90.50	90.30	90.40	88.94	88.94	88.94	89.33	89.28	89.31
Logic-based (ours)	*RuleEL*	79.82	80.10	79.96	81.55	75.15	78.22	76.56	74.55	75.54
	LogicEL	86.68	86.48	86.58	83.05	83.05	83.05	82.60	82.58	82.59
	LNN-EL	87.74	87.54	87.64	88.52	88.52	88.52	85.11	85.05	85.08
	$LNN\text{-}EL_{ens}$	**91.10**	**90.90**	**91.00**	**91.38**	**91.38**	**91.38**	**92.17**	**92.08**	**92.12**

LNN-EL:

- Reaches SotA for entity linking on KBQA datasets
  - Improves on **BLINK** [Wu et al, 2020], a black-box zero-shot model based on BERT, pre-trained on 9M Wikipedia examples
- Easily extensible
  - Ensembles that are progressively richer in features: string similarity, BERT embeddings, Box embeddings, BLINK
- Interpretability



Extensibility: A Closer Look at the Rules

Name Rule:

Name similarity

$$R_{name} \leftarrow [f_{jacc}(m_i, e_{ij}) > \theta_1 \vee f_{lev}(m_i, e_{ij}) > \theta_2 \vee f_{jw}(m_i, e_{ij}) > \theta_3 \vee f_{spacy}(m_i, e_{ij}) > \theta_4] \wedge f_{prom}(m_i, e_{ij})$$

DBpedia ref count

Context Rule:

$$R_{ctx} \leftarrow [f_{jacc}(m_i, e_{ij}) > \theta_1 \vee f_{lev}(m_i, e_{ij}) > \theta_2 \vee f_{jw}(m_i, e_{ij}) > \theta_3 \vee f_{spacy}(m_i, e_{ij}) > \theta_4] \wedge f_{ctx}(m_i, e_{ij}) > \theta_5 \wedge f_{prom}(m_i, e_{ij})$$

Similarity of co-mentions and DBpedia entity desc.

Type Rule:

$$R_{type} \leftarrow [f_{jacc}(m_i, e_{ij}) > \theta_1 \vee f_{lev}(m_i, e_{ij}) > \theta_2 \vee f_{jw}(m_i, e_{ij}) > \theta_3 \vee f_{spacy}(m_i, e_{ij}) > \theta_4] \wedge f_{type}(m_i, e_{ij}) > \theta_5 \wedge f_{prom}(m_i, e_{ij})$$

Type similarity

With more features available:

- Performance typically increases.
- Combining the features into rules also becomes more challenging (**full-fledged rule learning will be needed**)

$R_{name} \vee R_{ctx} \vee R_{type}$

$R_{name} \vee R_{ctx} \vee R_{type} \vee R_{BLINK} \vee R_{Box}$

Dataset	LNN-EL	LNN-EL +BLINK	LNN-EL +BERTWiki	LNN-EL +Box	LNN-EL$_{ens}$ (both BLINK + Box)
LC-QuAD	87.64	90.24	88.23	89.05	**91.00**
QALD-9	88.52	90.96	86.41	88.52	**91.38**
WebQSP$_{EL}$	85.08	**92.32**	91.70	91.44	92.12

Blink Rule:

$$R_{blink} \leftarrow [f_{jacc}(m_i, e_{ij}) > \theta_1 \vee f_{lev}(m_i, e_{ij}) > \theta_2 \vee f_{jw}(m_i, e_{ij}) > \theta_3 \vee f_{spacy}(m_i, e_{ij}) > \theta_4] \wedge f_{blink}(m_i, e_{ij})$$

BLINK score

Box Rule:

$$R_{box} \leftarrow [f_{jacc}(m_i, e_{ij}) > \theta_1 \vee f_{lev}(m_i, e_{ij}) > \theta_2 \vee f_{jw}(m_i, e_{ij}) > \theta_3 \vee f_{spacy}(m_i, e_{ij}) > \theta_4] \vee f_{box}(m_i, e_{ij}) > \theta_5$$

Box similarity (co-mentions and DBpedia neighboring nodes)

Transferability

- Inductive bias offered by using rules leads to good transfer across different datasets within the same domain.

- No fine-tuning on the target dataset

- LNN-EL performs reasonably well, even in cases where training is done on a very small dataset.
  - E.g., from QALD-9 (with only **a few hundred questions** to train) to WebQSP: F1-score of 83.06 (vs. 85.08)

- Our competitor, zero-shot BLINK by design has very good transferability too, but it is **trained on entire Wikipedia**.

Train	Test LC-QuAD	Test QALD-9	Test $\text{WebQSP}_{\text{EL}}$
LC-QuAD	<u>87.64</u>	86.41	78.90
QALD-9	85.58	<u>88.52</u>	83.06
$\text{WebQSP}_{\text{EL}}$	80.95	87.25	<u>85.08</u>

Summary & Future Directions

- Summary
  - LNN-EL, a neuro-symbolic approach for entity linking on short text
  - Achieved competitive performance against SotA black-box neural models
  - LNN-EL is **interpretable, extensible** and **customizable**
  - LNN-EL may **transfer** better to new datasets

- Future directions
  - Automatic learning of the rule templates
  - Longer documents

Thank you!