Supplementary Material: Cross-Sentence N-ary Relation Extraction using Lower-Arity Universal Schemas

Kosuke Akimoto[1], **Takuya Hiraoka**[1], **Kunihiko Sadamasa**[1], **Mathias Niepert**[2]
[1] Security Research Laboratories, NEC Corporation
[2] NEC Laboratories Europe
{k-akimoto@ab,t-hiraoka@ce,k-sadamasa@az}.jp.nec.com
mathias.niepert@neclab.eu

A Details of Datasets

A.1 Relations in Wiki-90k Dataset

Wiki-90k dataset contains four ternary relations: `CastAs`, `ChildOf`, `DirectedBy`, and `AppearInGenre`. Since Wikidata contains only binary relational facts, we combined a few binary relations to define each ternary relation. In the following, we list meanings of each ternary relation and original binary relations (and their relation identifiers) in Wikidata.

- `CastAs` relation has three arguments, $work$, $actor$ and $character$, and means that "$actor$ plays the role of $character$ in $work$ (films, television programs, etc.)". We used *cast member* (P161) relation and its *character role* (P453) qualifier to query facts of `CastAs` relation from Wikidata.
- `ChildOf` relation has three arguments, $person$, $father$ and $mother$, and means that "$person$ is a child of $father$ and $mother$". We used *father* (P22) relation and *mother* (P25) relation to query facts of `ChildOf` relation from Wikidata.
- `DirectedBy` relation has three arguments, $actor$, $director$ and $work$, and means that "$actor$ is directed by $director$ in $work$". We used *director* (P57) relation and *cast member* (P161) relation to query facts of `DirectedBy` relation.
- `AppearInGenre` relation has three arguments, $actor$, $genre$, and $director$, and means that " $actor$ appears in a work (film, television program, etc.) of $genre$ which is directed by $director$". We used *genre* (P136) relation, *cast member* (P161) relation and *director* (P57) relation to query facts of `AppearInGenre` relation.

	# Entity tuples	#positive facts			
		`CastAs`	`Child Of`	`Directed By`	`PlayIn Genre`
Train	50727	169	470	397	331
Dev	18211	53	197	138	134
Test	20218	55	181	141	100

Table 1: Number of entity tuples and positive facts in Wiki-90k dataset.

A.2 Statistics of Wiki-90k Dataset

Wiki-90k dataset contains 88k distinct entity tuples extracted from 11k different documents. Entity tuples are separated into train, dev, and test datasets (Table 1).

A.3 Relations in WF-20k Dataset

Freebase contains multiple n-ary relations each of whose fact is described by a combination of multiple triples. For example, Freebase contains three relations, "olympics.olympic_medal_honor.olympics", "olympics.olympic_medal_honor.medalist" and "olympics.olympic_medal_honor.country". Triples of these relations sharing a same subject entity constitute one ternary relation whose meaning is that "*Medalist* from *country* got a medal in *olympic*.", where *medalist*, *country* and *olympic* are object entities of the relations.

Similarly, we defined 19 ternary relations as described in Table 2. In the table, *core predicate* refers to a shared part of relation name ("olympics.olympic_medal_honor" in the example), and *argument* refers to last part of relation name ("olympics", "medalist" and "country" in the example). (If we use shared subject entity as argument, we denote it as "SBJ".)

Relation ID	Core predicate	argument 1	argument 2	argument3	#Train	#Dev	#Test
0	american_football.player_game_statistics	player	season	team	33	12	11
1	architecture.structure	SBJ	architect	architectural_style	59	21	18
2	cvg.computer_videogame	cvg_genre	developer	game_series	97	25	28
3	cvg.computer_videogame	cvg_genre	game_series	publisher	102	22	31
4	film.performance	actor	character	film	68	23	20
5	geography.lake	SBJ	basin_countries	outflow	75	26	27
6	geography.river	SBJ	basin_countries	origin	82	18	13
7	geography.river	SBJ	mouth	origin	81	33	22
8	music.artist	SBJ	genre	origin	51	20	20
9	music.group_membership	group	member	role	84	28	37
10	music.recording	artist	producer	song	69	24	25
11	olympics.olympic_athlete_affiliation	athlete	olympics	sport	82	29	25
12	olympics.olympic_medal_honor	country	medalist	olympics	60	29	15
13	organization.leadership	organization	person	role	105	37	29
14	organization.organization	SBJ	founders	place_founded	70	8	19
15	sports.sports_team_roster	player	position	team	60	22	16
16	transportation.bridge	SBJ	body_of_water_spanned	bridge_type	47	20	19
17	transportation.bridge	SBJ	body_of_water_spanned	locale	78	44	43
18	tv.regular_tv_appearance	actor	character	series	147	51	70

Table 2: Definition and number of positive fact of relations in WF-20k dataset.

Method	relation																			average	weighted
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
Proposed	**0.809**	*0.767*	0.929	0.935	**0.468**	**0.869**	**0.763**	*0.522*	0.784	(0.892)	**0.806**	*0.956*	0.855	0.915	**0.663**	*0.865*	**0.996**	**0.935**	*0.878*	**0.821**	**0.842**
(Song et al., 2018)	0.653	0.632	0.775	0.846	0.259	0.778	0.539	0.348	0.319	0.554	0.546	0.836	0.761	0.663	0.495	0.533	0.884	0.834	0.883	0.639	0.680
(Toutanova et al., 2015) with Graph State LSTM (Song et al., 2018) encoder																					
Model F	0.327	0.171	0.559	0.604	0.067	0.303	0.127	0.222	0.150	0.187	0.196	0.503	0.476	0.382	0.230	0.369	0.560	0.439	0.610	0.341	0.380
Model E	0.714	0.723	**0.982**	**0.988**	0.207	0.736	0.507	0.262	(0.792)	0.884	0.402	0.935	**0.955**	(0.915)	0.452	0.846	0.885	0.768	0.814	0.725	0.752
(Verga et al., 2017) with Graph State LSTM (Song et al., 2018) encoder																					
Model F	0.519	0.514	0.779	0.845	0.186	0.760	0.452	0.305	0.477	0.506	0.491	0.836	0.724	0.653	0.452	0.526	0.903	0.845	0.819	0.610	0.653

(normal): $p > 0.15$, *italic*: $p \leq 0.15$, **bold**: $p \leq 0.01$

Table 3: Mean average precisions (MAPs) on test data (WF-20k).

A.4 Statistics of WF-20k Dataset

WF-20k dataset contains 19530 distinct entity tuples from 18202 distinct paragraphs of English Wikipedia. Entity tuples are randomly separated into train, development (dev), and test subsets, and these subsets contains 11720, 3905, and 3905 distinct entity tuples respectively. Number of positive facts of each relation type is shown in Table 2.

B Detailed Settings of Experiments

B.1 Document Graph and Shortest Paths

Following (Quirk and Poon, 2017), we use dependency links and coreference links as edges of a document graph. Also, we add edges between adjacent words and those between root words of adjacent sentences.

To seach shortest paths between two entities e_k, e_l, we use Dijkstra's algorithm to calculate the shortest paths. Since simple word adjacency has relatively lower information, we penalize such edges by giving 10 times higher cost when searching shortest paths.

B.2 Hyperparameters

For all methods, we used Adam optimizer (Kingma and Ba, 2015) and initialized word embeddings by pre-trained 300-dim GloVe vector (Pennington et al., 2014). Embeddings of unknown words or tokens are initialized randomly.

weight decay ratio	$0, 10^{-7}, 10^{-5}$
learning rate	$10^{-3}, 10^{-4}$
batch size	$8, 32$
dimension of node representations	$50, 100$
dropout rate	$0, 0.2$
number of Graph State LSTM layers	$3, 5, 7$

Table 4: Settings of the grid search for (Song et al., 2018) baseline.

learning rate	$10^{-3}, 10^{-4}$
weight decay ratio	$0, 10^{-8}, 10^{-6}, 10^{-4}$
the dimension of hiddne state	100, 300
the dimension of node representations	50, 100
the dimension of link representations	3, 10
dropout rate	0, 0.1, 0.2
dropout rate for node representations	0, 0.1, 0.2

Table 5: Settings of the grid search for universal schemas baselines.

Proposed method: We set learning rate to 10^{-4}, weight decay ratio to 10^{-4}, the dimension of relation representations to 300, and batch size to 50 (for $\mathcal{L}_2$) and 25 (for $\mathcal{L}_1, \mathcal{L}_n$).

(Song et al., 2018) baseline: We conducted grid search to tune hyperparameters of (Song et al., 2018) method as shown in Table 4. We set the dimension of hidden state to 300, and the dimension of link embeddings to 10. In all experiments, we trained models for 50 epochs.

Universal schemas baselines: To reduce number of required trials, we conducted two step heuristical procedure to tune hyperparameters.

Method	CastAs	Child Of	Directed By	PlayIn Genre	average	weighted
Proposed	*0.338*	0.660	**0.742**	**0.597**	**0.584**	**0.634**
(Song et al., 2018)	0.291	0.671	0.633	0.290	0.471	0.536
(Toutanova et al., 2015) with Graph State LSTM (Song et al., 2018) encoder						
Model F	0.083	0.211	0.384	0.281	0.240	0.262
Model E	0.154	0.235	0.637	0.569	0.399	0.414
(Verga et al., 2017) with Graph State LSTM (Song et al., 2018) encoder						
Model F	0.283	0.543	0.527	0.417	0.443	0.482

normal: $p > 0.15$, *italic*: $p \leq 0.15$, **bold**: $p \leq 0.01$

Table 6: Mean average precisions (MAPs) on **test** data (Wiki-90k).

Setting	CastAs	Child Of	Directed By	PlayIn Genre	average	weighted
U	0.112	0.369	0.517	0.454	0.363	0.404
B	0.245	0.530	0.473	0.397	0.411	0.452
N	0.410	0.660	0.780	0.733	0.646	0.685
U+B	0.372	0.584	0.598	0.530	0.521	0.552
U+B+N	0.426	0.660	0.726	0.771	0.646	0.682
U+B+N ($K = 10$)	0.416	0.671	0.724	0.789	**0.650**	**0.689**
U+B+N ($K = 5$)	0.364	0.677	0.734	0.592	0.592	0.640
U+B+N (fix $w_r = 1$)	0.443	0.655	0.719	0.765	0.645	0.679

Default hyperparameter: $K = 20$

Table 7: Ablation study: mean average precisions (MAPs) on **dev** data (Wiki-90k).

First, we conducted grid search to tune learning rate and weight decay ratio by fixing other hyperparameters to empirically tuned values. Then, we conducted grid search to tune other hyperparameters by fixing learning rate and weight decay ratio to values tuned in the first step. Settings of the grid search are shown in Table 5. In all experiments, we set the number of negative examples $K = 5$, the number of Graph State LSTM layers to 5, and we trained models for 50 epochs.

C Results for each relation type

Table 3, 6, 7 illustrates detailed results of the experiments for each relation type.

References

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015.*

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.

Chris Quirk and Hoifung Poon. 2017. Distant supervision for relation extraction beyond the sentence boundary. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1171–1182. Association for Computational Linguistics.

Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. N-ary relation extraction using graph-state lstm. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2226–2235. Association for Computational Linguistics.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing text for joint embedding of text and knowledge bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1499–1509. Association for Computational Linguistics.

Patrick Verga, Arvind Neelakantan, and Andrew McCallum. 2017. Generalizing to unseen entities and entity pairs with row-less universal schema. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*.