Multifacet Relation Extraction

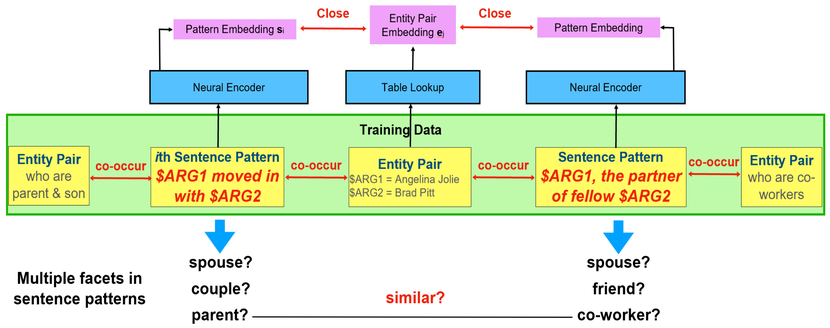
Rohan Paul*, Haw-Shiuan Chang*, Andrew McCallum

UMASS AMHERST

IESL

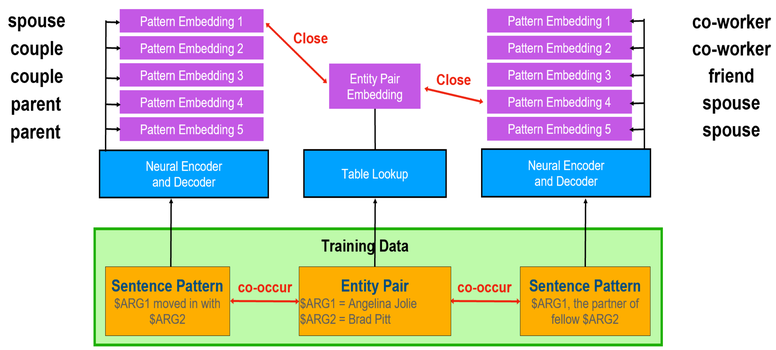
Introduction

- **Previous work in Distant Supervised Relation Extraction:**
  - Training signal - Co-occurrence of entity pairs with sentence patterns
  - Universal Schema - assumes one embedding per sentence pattern
  - Compositional Universal Schema uses LSTM to encode sentence pattern
- **Limitations:**
  - A sentence pattern could co-occur with different facets (i.e., entity pairs with different relation)
  - All facets compressed into a single embedding for each sentence pattern

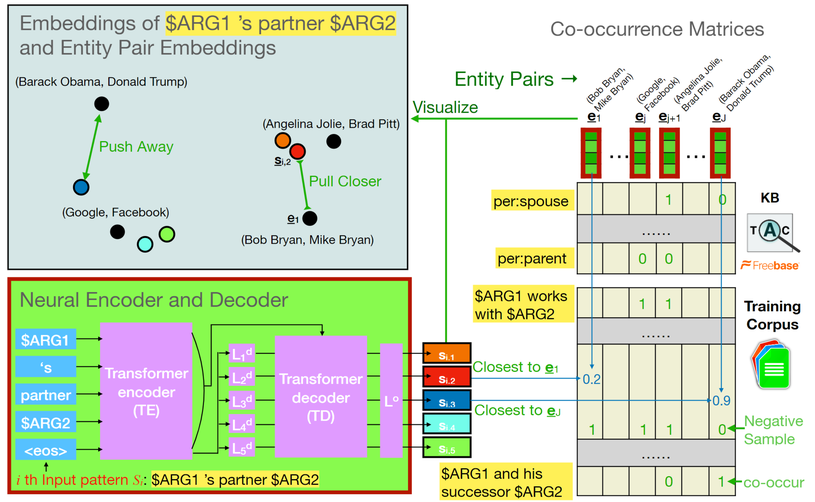


Main Idea – Multifacet Embeddings

- Represent each pattern using multiple embeddings
- Each embedding represents a facet
- One facet can be represented by multiple embeddings
- A facet of a pattern is similar to a facet of another pattern if the patterns share same entity pairs



Our Method



Seq2Seq Model

- Transformer encoder to encode sentence into a single embedding
- Transformer decoder to decode encoded embedding into multiple facet embeddings

Objective function

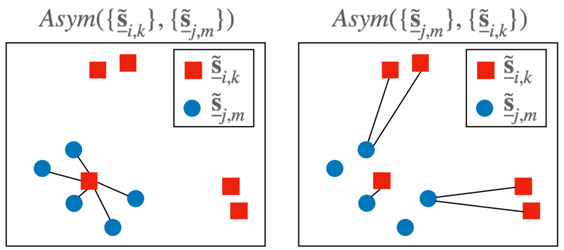
- Choose closest facet for each entity pair
- Minimize the distance of the co-occurred entity pair and sentence pattern
- Maximize all the other distances
- Similar to kmeans clustering, encouraging pattern embeddings to become the centers of entity pairs that co-occur with the pattern

Scoring functions

- $Asym(\{\tilde{\mathbf{s}}_{i,k}\}, \{\tilde{\mathbf{s}}_{j,m}\})$ computes the average of cosine similarities from every center in $\{\tilde{\mathbf{s}}_{j,m}\}$ to its closest center in $\{\tilde{\mathbf{s}}_{i,k}\}$
- Checking if $\{\tilde{\mathbf{s}}_{i,k}\}$ is more specific than $\{\tilde{\mathbf{s}}_{j,m}\}$: $Asym(\{\tilde{\mathbf{s}}_{i,k}\}, \{\tilde{\mathbf{s}}_{j,m}\}) > Asym(\{\tilde{\mathbf{s}}_{j,m}\}, \{\tilde{\mathbf{s}}_{i,k}\})$?
- Similarity between $\{\tilde{\mathbf{s}}_{i,k}\}$ and $\{\tilde{\mathbf{s}}_{j,m}\}$: $(Asym(\{\tilde{\mathbf{s}}_{i,k}\}, \{\tilde{\mathbf{s}}_{j,m}\}) + Asym(\{\tilde{\mathbf{s}}_{j,m}\}, \{\tilde{\mathbf{s}}_{i,k}\}))/2$

Specific Pattern: "$ARG1 , the wife of fellow $ARG2"

General Pattern: "$ARG1 , the wife of $ARG2"



Experiments

Distant supervised relation extraction

- Setting is the same as CUSchema (Verga et al., 2016)
- Training data: text with linked entities + Freebase + a few rules (No label data)
- Validation data: TAC slot filling 2012
- Testing data: TAC slot filling 2013 2014

Method	TAC 2012 (Validation)			TAC 2013			TAC 2014		
	Prec	Recall	F1	Prec	Recall	F1	Prec	Recall	F1
USchema*	**34.8**	23.7	28.2	**42.6**	29.4	34.8	**35.5**	24.3	28.8
CUSchema (LSTM)*	27.0	**32.7**	29.6	39.6	32.2	35.5	32.9	27.3	29.8
Ours (Single-LSTM)	25.7	21.7	23.5	30.4	26.3	28.2	22.1	20.5	21.3
Ours (Single-Trans)	26.1	21.6	23.7	29.5	25.2	27.2	19.0	21.2	20.0
Ours (LSTM)	32.0	28.9	30.3	41.3	**33.9**	37.2	34.1	**29.5**	**31.6**
Ours (Trans)	33.6	27.7	**30.4**	42.5	33.2	**37.3**	34.6	28.5	31.3
USchema + CUSchema (LSTM)*	29.3	32.8	30.9	**41.9**	34.4	37.7	29.3	34.1	31.5
USchema + Ours (LSTM)	29.2	33.7	31.3	38.1	**38.9**	38.5	31.5	**34.4**	32.9
USchema + Ours (Trans)	**30.4**	**33.9**	**32.1**	39.0	38.8	**38.9**	**32.0**	34.0	**33.0**

Unsupervised entailment detection

- Dataset: Sentence pattern pairs where the words in premise and hypothesis are hypernymy; e.g., president -> leader
- Collect labels of 1,500 sentence pattern pairs
- Retrieve entailment out of other relations (including paraphrase)
- Predict the entailment direction of patterns

Premise (Specific Pattern)	Hypothesis (General Pattern)	Label
$ARG1 , the president of the $ARG2	$ARG1 , the leader of the $ARG2	Entailment
$ARG1 's chairman , $ARG2	$ARG1 's leader , $ARG2	Entailment
$ARG1 's father , $ARG2	$ARG1 's leader , $ARG2	Other
$ARG1 worked with $ARG2	$ARG1 deal with $ARG2	Entailment
$ARG1 had with $ARG2	$ARG1 deal with $ARG2	Other
$ARG1 said the $ARG2	$ARG1 say the $ARG2	Paraphrase

Method	Classification	Direction Detection	
	AP@all	Micro Acc	Macro Acc
Random	21.9	50.0	50.0
Freq Diff	21.4	54.5	47.3
CUSchema	31.2	50.0	50.0
Ours (Single)	23.6	50.0	50.0
Ours	**37.8**	**63.1**	**55.4**

Conclusion

- We propose a novel method to improve the compositional universal schema, CUSchema, by modeling different facets of a sentence pattern
- Representing every sentence pattern using the cluster centers, we can achieve better symmetric and asymmetric similarity measurement between sentence patterns for relation extraction and entailment detection

References

[1] Riedel, S., Yao, L., McCallum, A., & Marlin, B. (2013). Relation extraction with matrix factorization and universal schemas. In *NAACL*.
[2] Verga, P., Belanger, D., Strubell, E., Roth, B., & McCallum, A. (2016). Multilingual relation extraction using compositional universal schema. In *NAACL*.
[3] Chang, H-S, Yuan, J., Iyyer, M., & McCallum, A. (2021). Changing the Mind of Transformers for Topically-Controllable Language Generation. In *EACL*.
[4] Chang, H-S, Agrawal A. & McCallum, A. (2021). Extending Multi-Sense Word Embedding to Phrases and Sentences for Unsupervised Semantic Applications/ In *AAAI*.