Few-shot Named Entity Recognition with Self-describing Networks

Jiawei Chen[1,3,*], Qing Liu[1,3,*], Hongyu Lin[1,†] , Xianpei Han[1,2,4,†], Le Sun[1,2]

[1]Chinese Information Processing Laboratory [2]State Key Laboratory of Computer Science

Institute of Software, Chinese Academy of Sciences

[3]University of Chinese Academy of Sciences

[4]Beijing Academy of Artificial Intelligence

Outline

- Introduction
- Self-describing Networks for few-shot NER
- SDNet pre-training and fine-tuning
- Experiments
- Conclusion

Outline

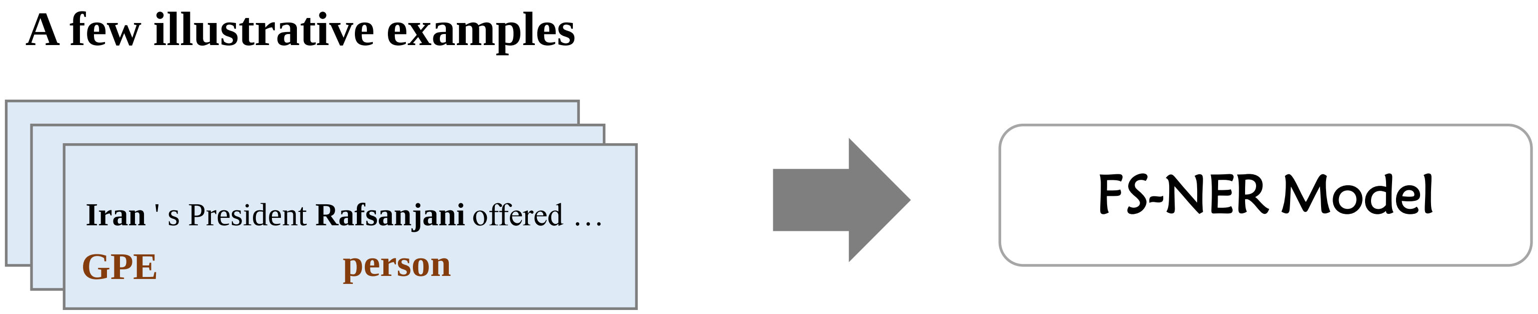
- Introduction
- Self-describing Networks for few-shot NER
- SDNet pre-training and fine-tuning
- Experiments
- Conclusion

Few-shot named entity recognition

- Few-shot named entity recognition (FS-NER) aims to identify entity mentions corresponding to new entity types with only a few illustrative examples.

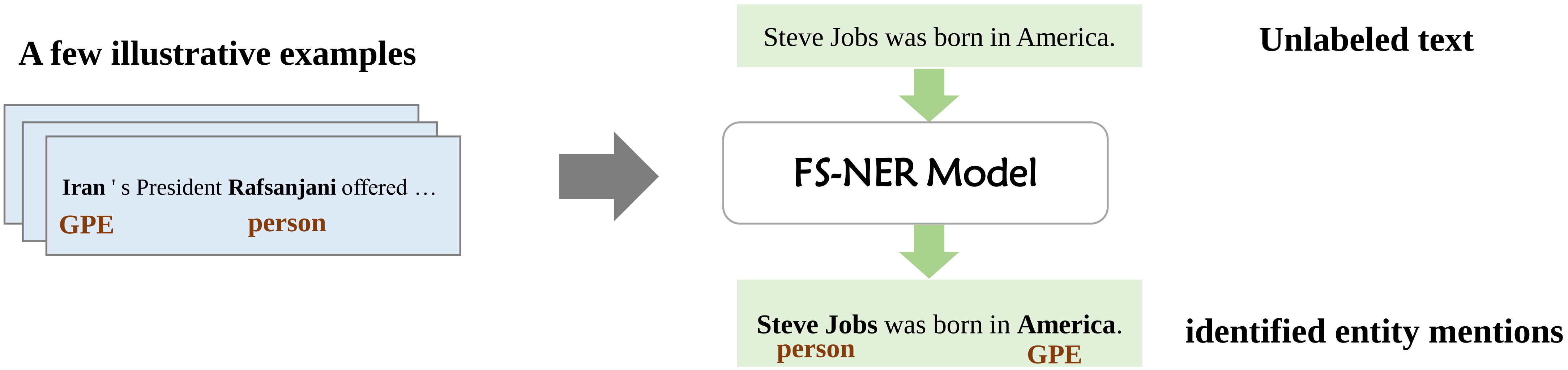
Few-shot named entity recognition

- Few-shot named entity recognition (FS-NER) aims to identify entity mentions corresponding to new entity types with only a few illustrative examples.



Few-shot named entity recognition

- Few-shot named entity recognition (FS-NER) aims to identify entity mentions corresponding to new entity types with only a few illustrative examples.



Challenges of FS-NER

- **Limited information challenge**
    - Information entailed in illustrative examples is very limited.

Challenges of FS-NER

- **Knowledge mismatch challenge**
    - External knowledge usually doesn’t directly match with the new task because it may contain irrelevant, heterogeneous or even conflicting knowledge.

Challenges of FS-NER

- **Knowledge mismatch challenge**
  - For example, “America” is different types in different datasets.

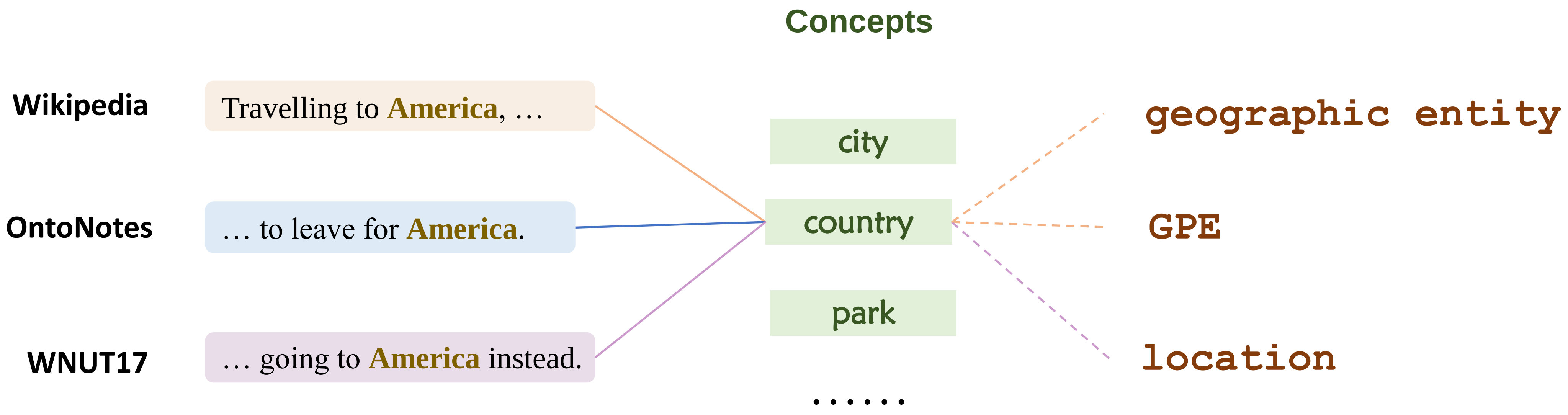
Wikipedia	Travelling to **America**, …	`geographic entity`
OntoNotes	… to leave for **America**.	`GPE`
WNUT17	… going to **America** instead.	`location`

Self-describing mechanism for FS-NER

- Self-describing mechanism
  - All entity types can be described using the same set of concepts
  - The mapping between types and concepts can be universally modeled and learned.

Self-describing mechanism for FS-NER

- Knowledge mismatch challenge can be resolved by uniformly describing different entity types using the same concept set



Self-describing Networks for FS-NER

- Self-describing Networks (SDNet) : A Seq2Seq generation network

  - Universally describe mentions using concepts
  - Automatically map novel entity types to concepts
  - Adaptively recognize entities on-demand

Outline

- Introduction
- Self-describing Networks for few-shot NER
- SDNet pre-training and fine-tuning
- Experiments
- Conclusion

Self-describing Networks for FS-NER

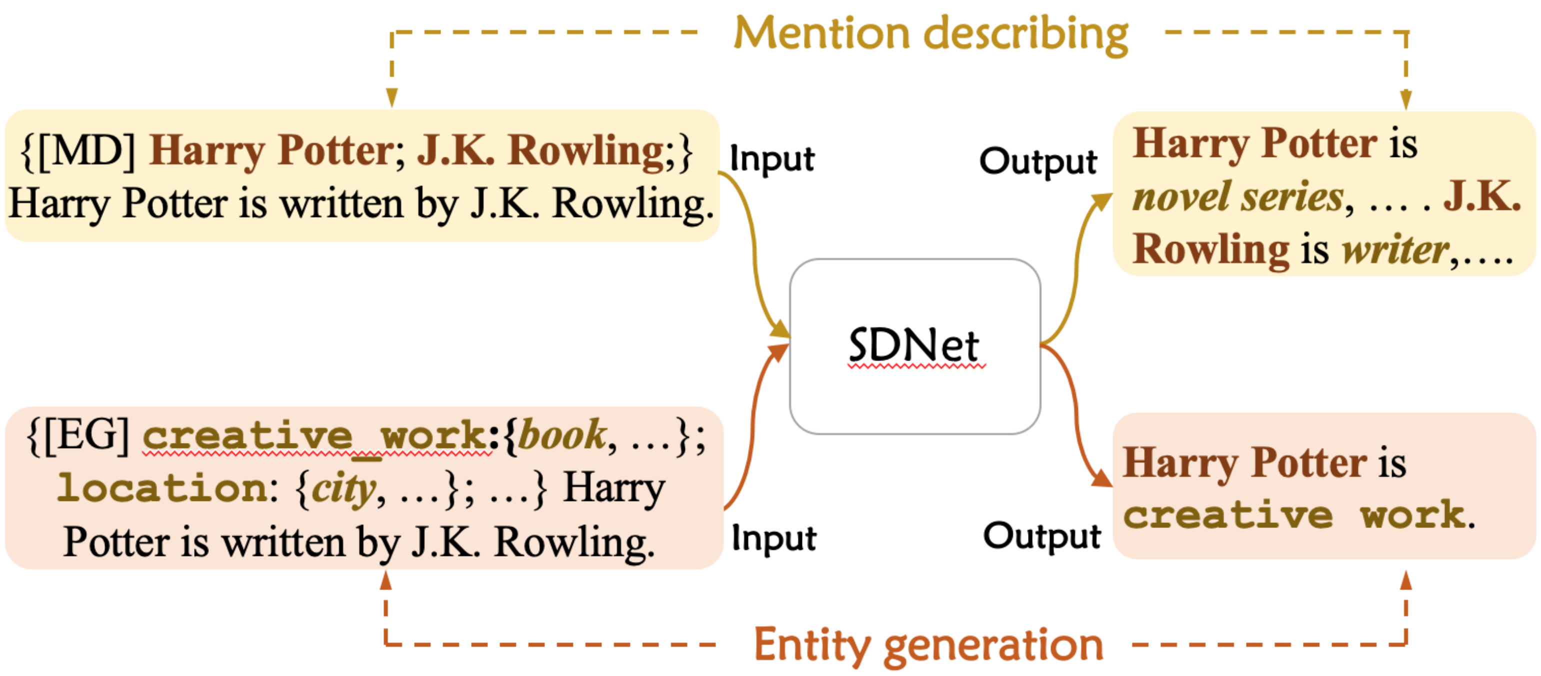
- Self-describing Networks (SDNet) : A Seq2Seq generation network

  - Entity generation
    - Adaptively generate entity mentions

  - Mention describing
    - Generate concept descriptions of mentions

Self-describing Networks for FS-NER

- Self-describing Networks (SDNet) : A Seq2Seq generation network



Self-describing Networks for FS-NER

- Entity generation

Input: task prompt

{[EG] `creative_work`:{*book*, …}; `location`: {*city*, …}; …}

Task descriptor

target novel types

Corresponding descriptions of target types

Self-describing Networks for FS-NER

- Entity generation

Input: task prompt + sentence

{[EG] **creative_work**:**{***book*, …}; **location**: {*city*, …}; …}
Harry Potter is written by J.K. Rowling.

Self-describing Networks for FS-NER

- Entity generation

Output

Harry Potter is **`creative work`**.

Recognized entity

target novel types

Self-describing Networks for FS-NER

- Mention describing

Input: task prompt

{[MD] **Harry Potter**; **J.K. Rowling**;}

Task descriptor

Target entity mentions

Self-describing Networks for FS-NER

- Mention describing

Input: task prompt + sentence

$\{\underbrace{[\text{MD}]}_{\text{Task descriptor}}\ \underbrace{\textbf{Harry Potter}; \textbf{J.K. Rowling};}_{\text{Target entity mentions}}\}$ Harry Potter is written by J.K. Rowling.

Self-describing Networks for FS-NER

- Mention describing

Output

Harry Potter is ***novel series***, … . **J.K. Rowling** is ***writer***,….

Target entity mention   Concept descriptions

Self-describing Networks for FS-NER

- Type description construction
    - SDNet summarizes the generated concepts to describe the precise semantics of specific novel types.
    - All concept descriptions of mentions with the same type will be fused and regarded as the description of the type which is used in entity generation step.

Self-describing Networks for FS-NER

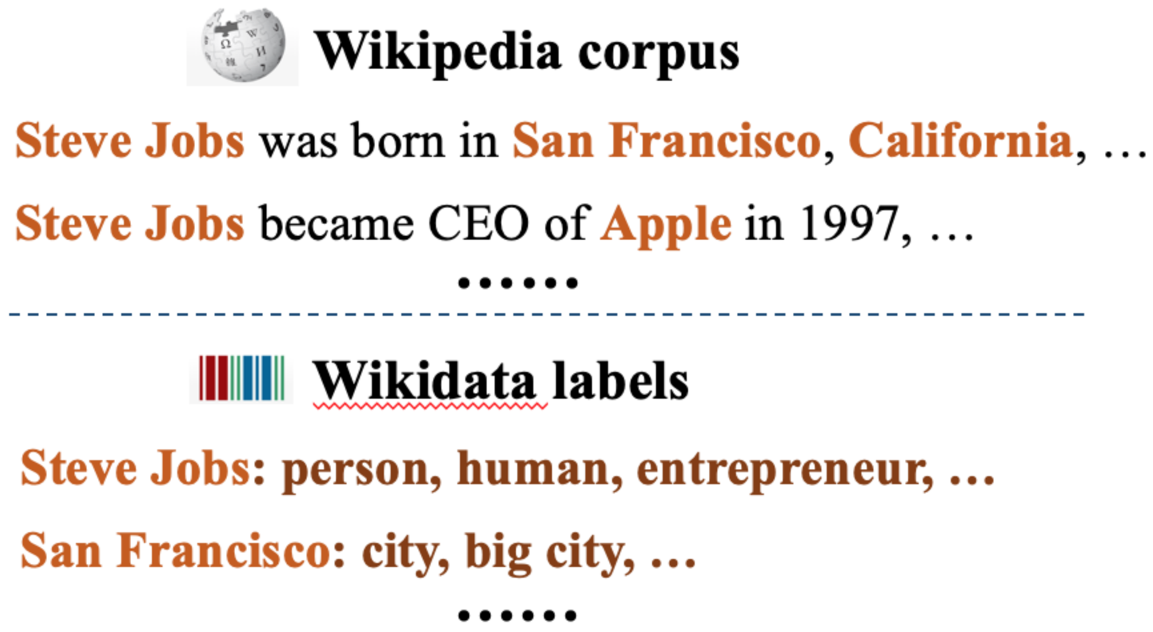
- Type description construction
  - SDNet summarizes the generated concepts to describe the precise semantics of specific novel types.
  - **Filtering strategy**: If most of the generated concept descriptions of one type is “other”, we will remove the type description of this type and directly use the type name for prompt in entity generation step.

Outline

- Introduction
- Self-describing Networks for few-shot NER
- SDNet pre-training and fine-tuning
- Experiments
- Conclusion

SDNet Pre-training

- We pre-train SDNet using easily available and large-scale Wikipedia and Wikidata data.
  - Collecting 31k types in Wikidata
  - 56M instances

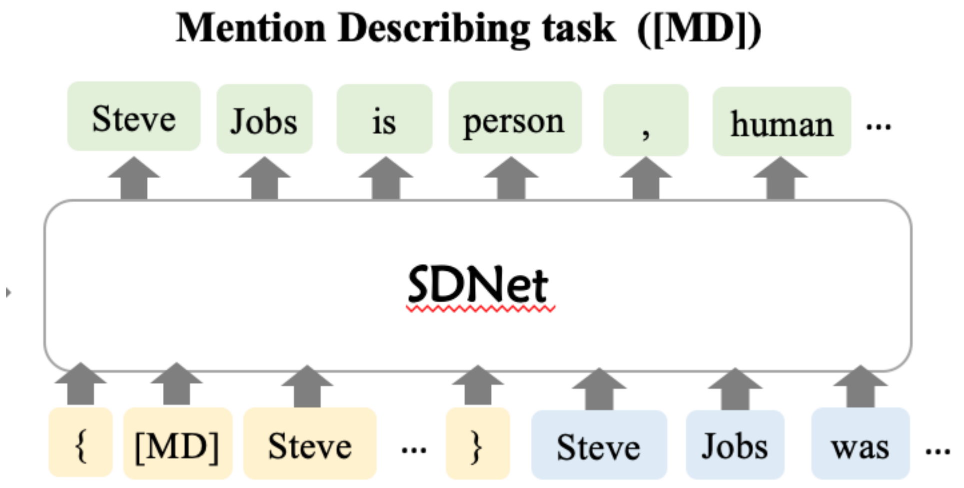


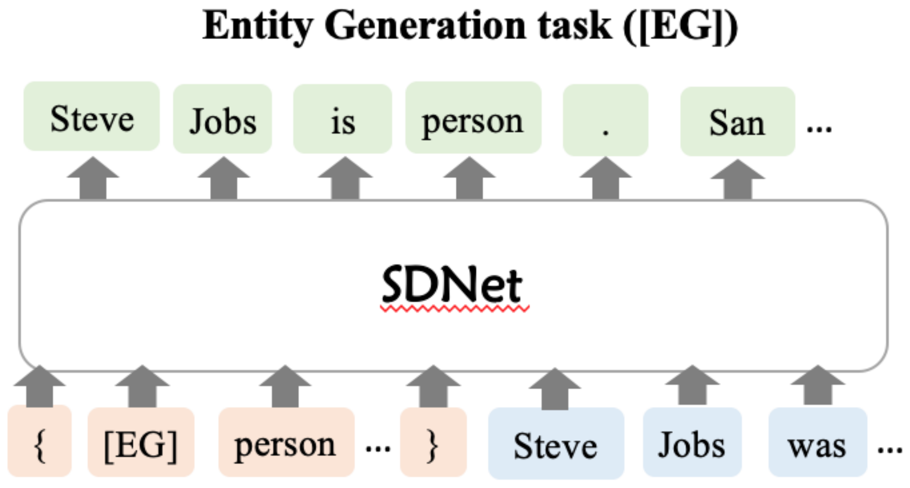
SDNet Pre-training

- Type description building
  - This paper uses the collected entity types as concepts.
  - For one entity type, we collect all co-occurring entity types as the describing concepts

SDNet Pre-training

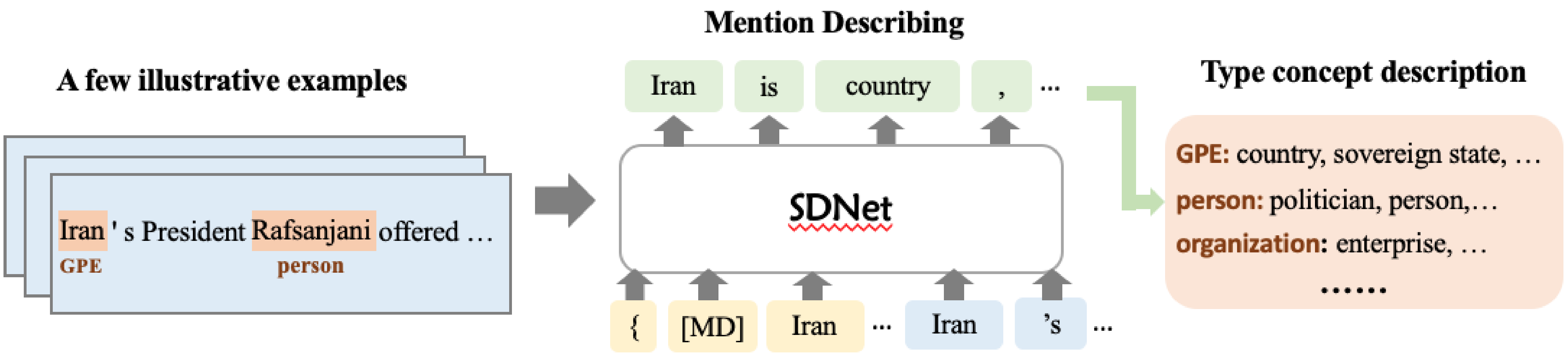
- We jointly pre-train the two tasks.





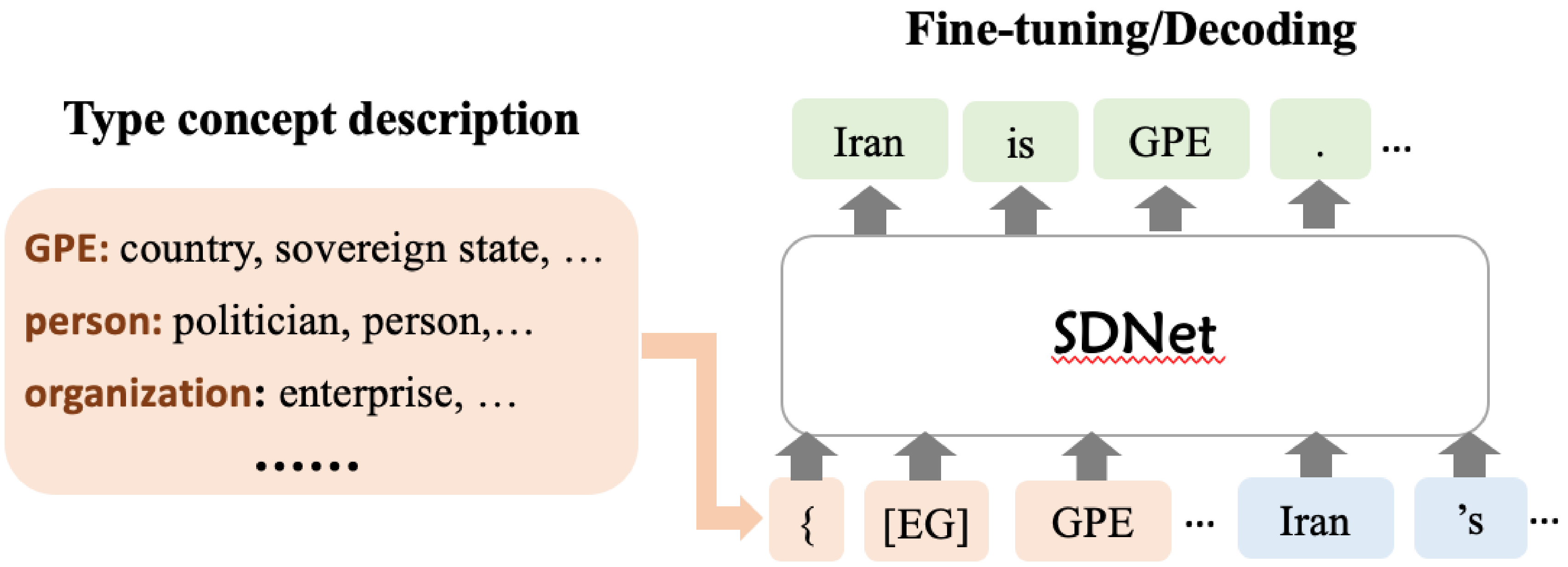
SDNet Fine-tuning and decoding

- Before fine-tuning or decoding, SDNet will automatically build type descriptions using illustrative instances.



SDNet Fine-tuning and decoding

- The type descriptions are used in fine-tuning or decoding.



Outline

- Introduction
- Self-describing Networks for few-shot NER
- SDNet pre-training and fine-tuning
- Experiments
- Conclusion

Datasets

- We use 8 benchmarks from different domains

Dataset	Domain	#Types	#Test
WNUT	Social Media	6	1287
CoNLL	News	4	3453
re3d	Defense	10	200
Res	Review	8	1521
Moive1	Review	12	1953
Movie2	Review	12	2443
I2B2	Medical	23	43697
Onto	General	18	8262

Evaluation

- The performance is evaluated by micro-F1 on test set.

- To obtain the offset of predicted mention, we extract entity mentions and their types from the generated sentence, and locate them in the original sentence.

Main result

		CoNLL	WNUT	Res	Movie1	Movie2	Re3d	I2B2	Onto	AVE
Baselines	RoBERTa (Huang et al., 2020)	53.5	25.7	48.7	51.3	/	/	36.0	57.7	/
	RoBERTa-DS (Huang et al., 2020)*	61.4	34.2	49.1	53.1	/	/	38.5	68.8	/
	Proto (Huang et al., 2020)	58.4	29.5	44.1	38.0	/	/	32.0	53.3	/
	Proto-DS (Huang et al., 2020)*	60.9	35.9	48.4	43.8	/	/	36.6	57.0	/
	spanNER (Wang et al., 2021)	71.1	25.8	49.1	/	65.4	/	/	67.3	/
	spanNER-DS (Wang et al., 2021)*	**75.6**	38.5	51.2	/	67.8	/	/	**71.6**	/
Baselines [in-house]	Bert-base	58.6	23.2	47.6	52.4	66.3	57.0	47.6	61.1	51.7
	T5-base	60.0	36.6	59.4	57.9	69.9	57.1	39.9	62.0	55.3
	T5-base-prompt	55.4	34.2	58.4	58.7	67.1	60.7	61.8	59.8	57.0
	T5-base-DS	68.2	34.9	59.7	58.4	70.8	56.0	34.1	58.8	55.1
Ours	SDNet	71.4	**44.1**	**60.7**	**61.3**	**72.6**	**65.4**	**64.3**	71.0	**63.8**

SDNet can effectively handle few-shot NER.

Ablation Study

- Type description is critical for SDNet to transfer knowledge and capture type semantics.

	WNUT		
	P	R	F
SDNet	**54.78**	37.08	**44.06**
w/o desp	48.78	**39.51**	43.54
w/o joint	50.68	37.46	42.96
w/o filter	53.57	35.01	42.23

Ablation Study

- Type description is critical for SDNet to transfer knowledge and capture type semantics.

- Joint learning mention describing and entity generation processes is effective.

	WNUT		
	P	R	F
SDNet	**54.78**	37.08	**44.06**
w/o desp	48.78	**39.51**	43.54
w/o joint	50.68	37.46	42.96
w/o filter	53.57	35.01	42.23

Ablation Study

- Type description is critical for SDNet to transfer knowledge and capture type semantics.

- Joint learning mention describing and entity generation processes is effective.

- Filtering strategy can effectively alleviate the transferring of mismatched knowledge

	WNUT		
	P	R	F
SDNet	**54.78**	37.08	**44.06**
w/o desp	48.78	**39.51**	43.54
w/o joint	50.68	37.46	42.96
w/o filter	53.57	35.01	42.23

Examples

Text	[Chris Hill]$_{\text{person}}$ was in [China]$_{\text{GPE}}$ [a few days ago]$_{\text{date}}$.
Input1	{ [EG] GPE: {state, country, city, democracy, republic, community}; date: {}; } Chris Hill was in China ...
Output1	China is GPE. a few days ago is date.
Input2	{ [EG] person: {politician, actor, lawyer}; organization: {business, company}; } Chris Hill was in China ...
Output2	Chris Hill is person.

SDNet can generate different outputs according to the prompt.

Thanks!
Any Question?

Source Code: https://github.com/chen700564/sdnet