国际人工智能会议
AAAI 2021论文北京预讲会

Extracting Zero-shot Structured Information from Form-like Documents: Pretraining with Keys and Triggers

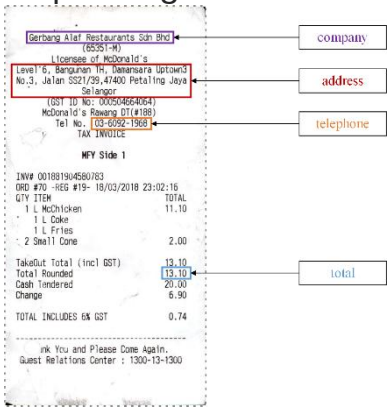
Rongyu Cao, Ping Luo

Contents

Background

Information Extraction from Document

- Input
  - a document (a collection of words with its content and position).
  - given attributes (keys).
- Output
  - for each key, which words consist the corresponding value.



V.
S.

Information Extraction from Plain Text

- Input
  - a sentence (a sequence of words with its content).
  - given attributes (keys).
- Output
  - for each key, which words consist the corresponding value.

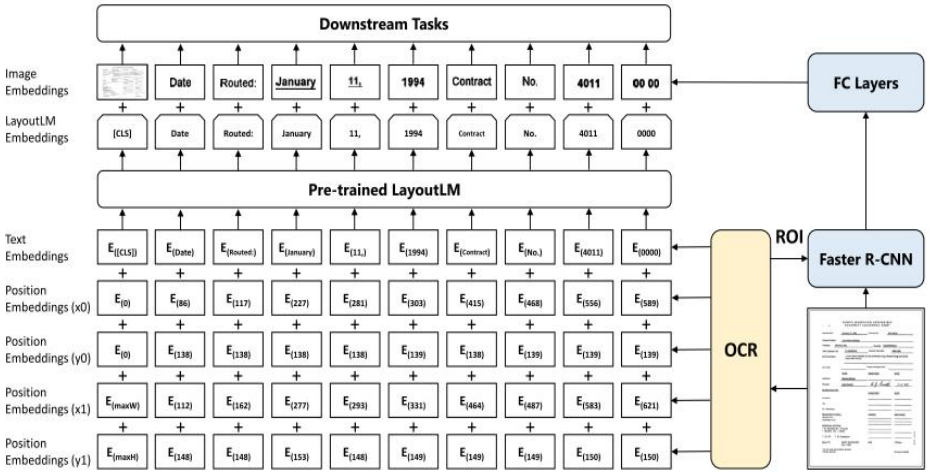
Evkoodance New Style Khaki NUbuck Size US 4-12 With Rhinestone High Quality 8.5cm Heel Open Toe Latin Dance Shoes

Attributes: color, model number, athletic shoe type, heel height

Background

- Previous Method: LayoutLM [1]

- Framework
  1. Like BERT, take a collection of words with text embeddings and 2-d position embeddings as input.
  2. Like BERT, use multi-layers, multi-heads transformer.
  3. Predict a class label (multi-classification) for each word.
  4. The parameters are pre-trained on two tasks: masked visual-language model and multi-label document classification.



[1] Xu, Y.; Li, M.; Cui, L.; Huang, S.; Wei, F.; and Zhou, M. LayoutLM: Pre-training of Text and Layout for Document Image Understanding. In KDD, 2020.

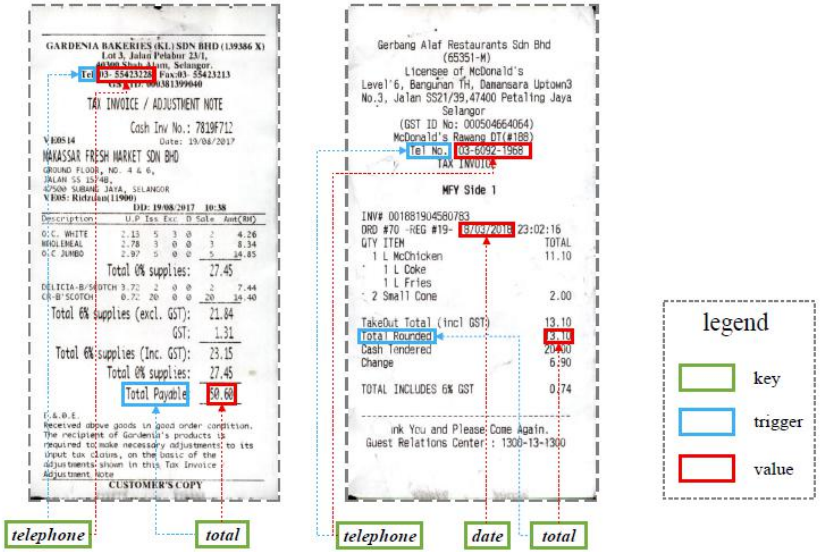
Challenge

- In real-world production, the keys needed to extract might have a huge number.
- Labeling large-scale training data for each key is costly.
- So, can the model generalize to unseen keys (zero-shot keys) without additional annotation?
- Thus, we should let the model learn key-invariant feature representation.

Solution

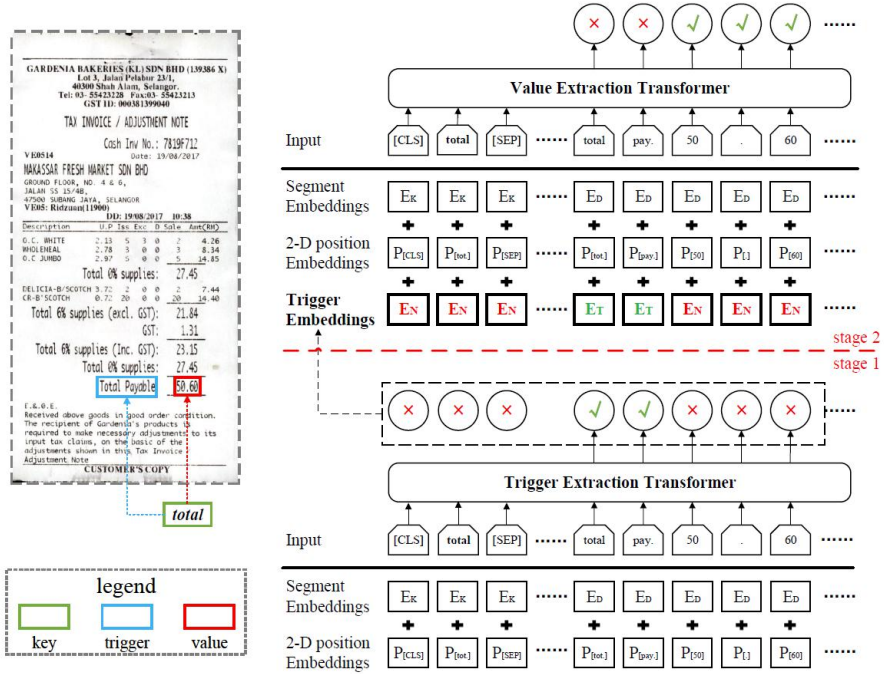
- What is the key-invariant feature representation?
- No matter what is the given key, the way we find value for the key is the same.

- Which way? How a human find the value for a given key?



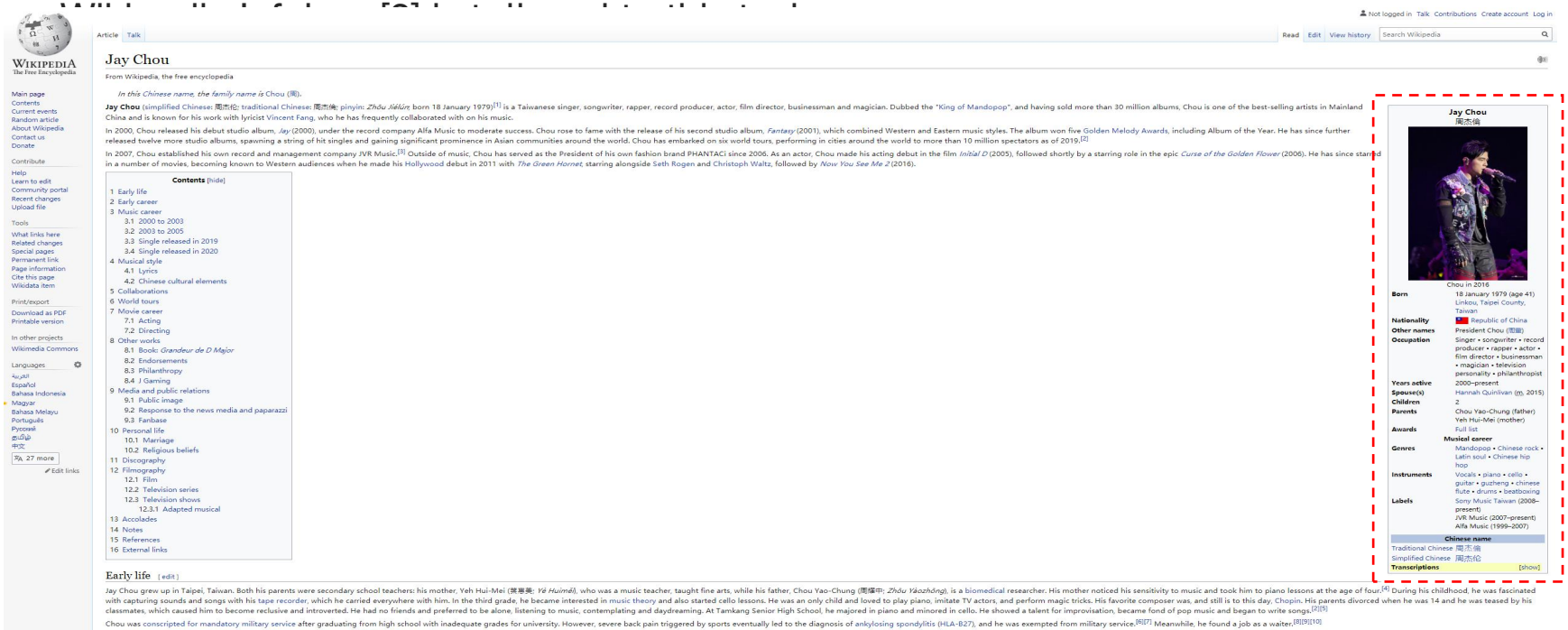
Solution

- Model: Two-stages LayoutLM-based model.
- It explicitly learns two mappings: key-to-trigger and trigger-to-value.



Solution

- How to annotate triggers and values for large-scale documents to pre-train this model?



[2] Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P. N.; Hellmann, S.; Morsey, M.; Van Kleef, P.; Auer, S.; et al. DBpedia - a large-scale, multilingual knowledge base extracted from Wikipedia. Semantic web. 2015.

Experiments

- Datasets

	description	access	#key	#zero-shot	#document	#pair
En-Infobox	Infobox on English Wikipedia webpage	crawler	7,370	-	398,467	1,238,281
Zh-Infobox	Infobox on Chinese Wikipedia webpage	crawler	6,151	-	281,281	897,520
SROIE	English receipts	public	6	2	972	5,505
Grater	Chinese bank statements	private	24	8	4,032	18,825

Zero-shot keys for SROIE dataset: cash, telephone.

Zero-shot keys for Grater dataset: organization (开户机构), balance in previous page (上页余额), subject (科目), card number (卡号), balance (本页余额), unit (单位), print method (打印渠道), credit count (信用笔数).

Experiments

- Results

Table 2: Comparing KATA with baseline models.

row	method	the SROIE dataset						the Grater dataset					
		zero-shot keys			non-zero-shot keys			zero-shot keys			non-zero-shot keys		
		P	R	$F1$	P	R	$F1$	P	R	$F1$	P	R	$F1$
1	BERT [Devlin et al. 2019]	0	0	0	0.9099	0.9099	0.9099	0	0	0	0.9986	0.9965	0.9976
2	LayoutLM [Xu et al. 2020b]	0	0	0	0.9460	0.9460	0.9460	0	0	0	1.0000	1.0000	1.0000
3	KA-bert	0.1043	0.1000	0.1021	0.9373	0.9373	0.9373	0.2431	0.2422	0.2426	0.9993	0.9979	0.9986
4	KA-wiki	0.6468	0.5961	0.6204	0.9469	0.9430	0.9449	0.3768	0.3702	0.3735	1.0000	1.0000	1.0000
5	KATA-bert	0.0649	0.0471	0.0545	0.9366	0.9382	0.9347	0.3692	0.2491	0.2975	0.9993	0.9993	0.9993
6	KATA-wiki	**0.7678**	**0.6549**	**0.7069**	**0.9483**	**0.9457**	**0.9470**	**0.7378**	**0.7301**	**0.7339**	1.0000	1.0000	1.0000
7	KATA-wiki w/ index embedding	0.6569	0.6000	0.6283	0.9455	0.9423	0.9439	0.6809	0.6644	0.6725	1.0000	1.0000	1.0000
8	KATA-wiki w/o fine-tuning	0.5116	0.4314	0.4681	0.1106	0.0502	0.0691	0.1322	0.0796	0.0994	0.0963	0.0594	0.0735

Conclusion

- We propose a novel task to extract the values of zero-shot keys in documents.
- We propose a two-stages framework to learn key-to-trigger and trigger-to-value mappings.
- We construct large-scale labeled dataset to pre-train the model on Wikipedia Infobox.

国际人工智能会议
AAAI 2021 论文北京预讲会

THANKS

2020.12.19