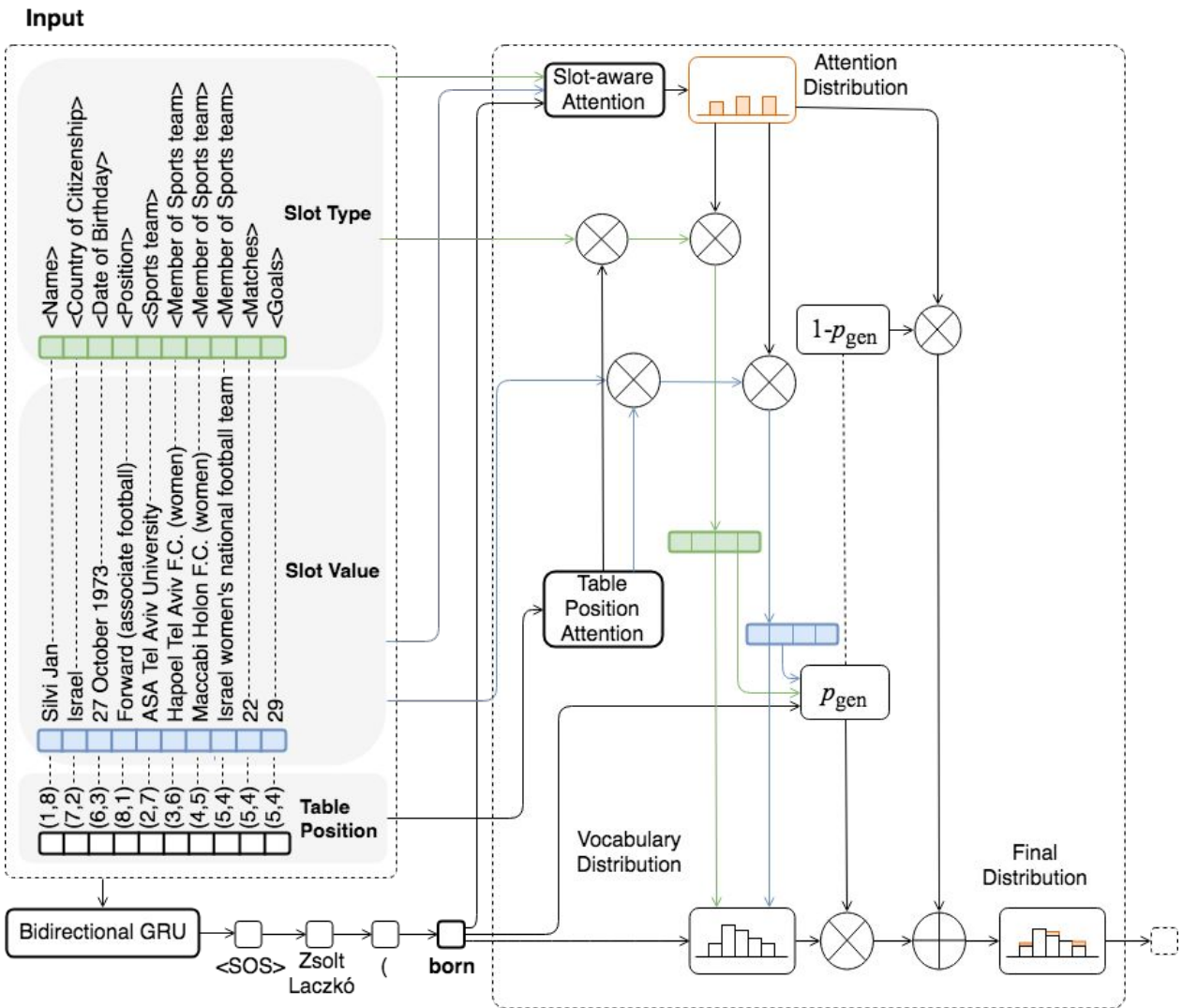
Narrating a Knowledge Base

Qingyun Wang, Xiaoman Pan, Lifu Huang, Boliang Zhang, Zhiying Jiang, Heng Ji and Kevin Knight

Contributions:

- Provide narrative description for structured knowledge base
  - Generalize linguistic expressions for slot types
  - Cover facts contained in the input KB
  - Capture interdependence among facts
- Propose KB reconstruction based metric to evaluate how many facts are correctly expressed in the generation output
- Create a large dataset of KBs paired with natural language descriptions for 106,216 entities

Model


Input
Slot Type
<Name>
<Country of Citizenship>
<Date of Birthday>
<Position>
<Sports team>
<Member of Sports team>
<Member of Sports team>
<Member of Sports team>
<Matches>
<Goals>
Slot Value
Silvi Jan
Israel
27 October 1973
Forward (association football)
ASA Tel Aviv University
Hapoel Tel Aviv F.C. (women)
Maccabi Holon F.C. (women)
Israel women's national football team
22
29
Table Position
(1,8)
(7,2)
(6,3)
(8,1)
(2,7)
(3,6)
(4,5)
(5,4)
(5,4)
(5,4)
Bidirectional GRU
<SOS>
Zsolt Laczkó
(
born
Slot-aware Attention
Attention Distribution
1-p_{gen}
Table Position Attention
p_{gen}
Vocabulary Distribution
Final Distribution

Input of structured table

Slot Type	Row	Slot Value		
Name	1	Silvi Jan		
Member of Sports Team	2	ASA Tel Aviv University		
	3	Hapoel Tel F.C. (women)		
	4	Maccabi Holon F.C. (women)		
	5	Israel women's national football team	Match	22
			Goal	29
Date of birth	6	27 October 1973		
Country of Citizenship	7	Israel		
Position	8	Forward (association football)		

Problem of seq2seq

(born 23 April 1981) is a retired Israeli footballer. He played for the Thailand 's (scoring one goal) and was a member of the team that won the first ever player in the history of the National Basketball League. She played for the team from 1997 to 2001 scoring 29 goals. She played for the team from 1997 to 2001 scoring 29 goals. She played for the team from 1999 to 2001 and played for the team in the 1997 and 2003 seasons.

Problem: Lack of important entity

Solution: pointer network

Problem of pointer

Silvi Jan the fourth past the Maccabi Holon F.C. (women). On 27 October 1973 in 29 2014) (born 22) is a former Israel. She was a Forward (association football) and currently plays for Hapoel Tel Aviv F.C.(women) in the Swedish league. She played for the ASA Tel Aviv University in the Swedish league. She was a member of the Israel women's national football team at the beginning of the 2008 season.

Problem: Incorrect slot type

Solution: slot type attention

Problem of pointer + type

Silvi Jan (born 27 October 1973) is a former Israeli footballer. He played for Hapoel Tel Aviv F.C.(women) and ASA Tel Aviv University.

Problem: Missing alignment slot

Solution: table position self-attention

Pointer + type & position

Silvi Jan (born 27 October 1973) is a former Israel. He played for Israel women's national football team, Hapoel Tel Aviv F.C.(women), ASA Tel Aviv University and Maccabi Holon F.C. (women). He was capped 22 times for the Israel women's national football team.

Slot Type Attention

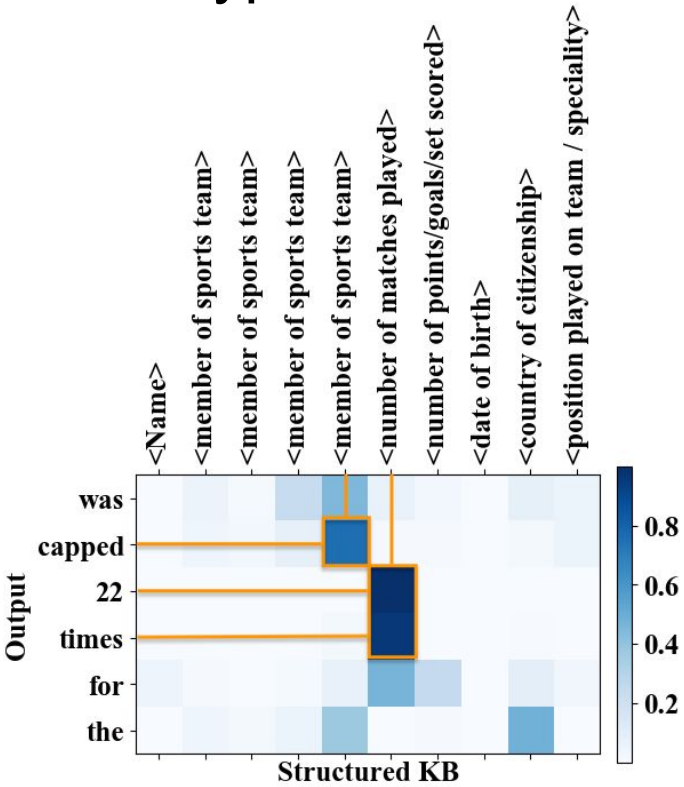
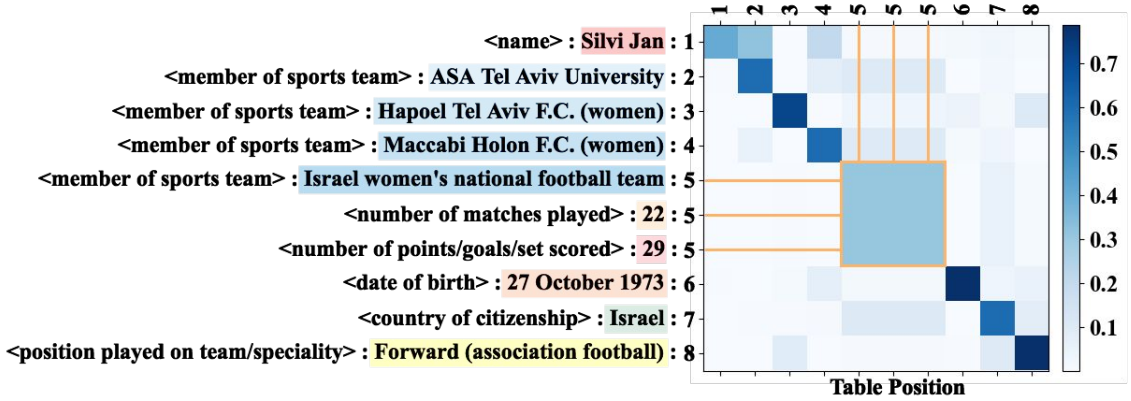


Table Position Self-Attention

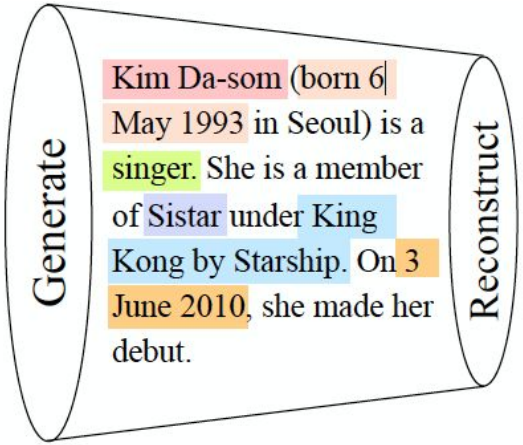


Standard Metrics (%)

Models	Person			Animal		
	BLEU4	**METEOR**	**ROUGEL**	**BLEU4**	**METEOR**	**ROUGEL**
Seq2seq	11.3	16.9	28.8	5.8	11.5	20.5
Pointer	17.2	21.1	37.4	6.6	13.7	37.8
+Type	23.1	22.2	39.5	**17.2**	**17.3**	42.8
+Type & position	**23.2**	**23.4**	**42.0**	14.8	17.2	**45.0**

KB reconstruction metric

Slot Type	Slot Value				
Name	Kim Da-som				
Date of Birth	6 May 1993				
Place of Birth	Gwangju				
Occupation	Singer				
Occupation	Actress				
Genres	K-pop				
Start Active Year	3 June 2010	Start Active Place	Seoul	Start Active Song	PUSH PUSH
Agent	King Kong by Starship				
Associated acts	Sistar				



Slot Type	Slot Value
Name	Kim Da-som
Date of Birth	6 May 1993
Place of Birth	Seoul
Occupation	Singer
Associated acts	Sistar
Agent	King Kong by Starship
Start Active Year	3 June 2010

KB Reconstruction based Evaluation (Scores for the example: Overall Slot Filling P=6/7=85.7%, R=6/11=54.5%, F1=66.7%; Inter-dependent Slot Filling P=5/7=71.4%, R=5/9}=55.6%, F1=62.5%

Overall Slot Filling Precision(P), Recall(R), F-score(F1) %

Models	Person			Animal		
	P	**R**	**F1**	**P**	**R**	**F1**
Seq2seq	74.6	29.3	42.0	82.5	27.8	41.6
Pointer	72.6	56.4	62.8	58.5	37.5	45.7
+Type	72.9	58.8	66.3	65.9	63.8	64.8
+Type & position	**76.3**	**62.7**	**68.8**	**73.4**	**71.8**	**76.6**

Interdependent Slot Filling Precision(P), Recall(R), F-score(F1) %

Models	Person			Animal		
	P	**R**	**F1**	**P**	**R**	**F1**
Seq2seq	74.7	30.0	43.4	82.5	27.9	41.7
Pointer	73.0	56.4	63.6	57.7	37.2	45.2
+Type	75.8	58.9	66.3	66.3	64.2	65.2
+Type & position	**77.2**	**63.5**	**69.7**	**72.6**	**71.0**	**71.8**

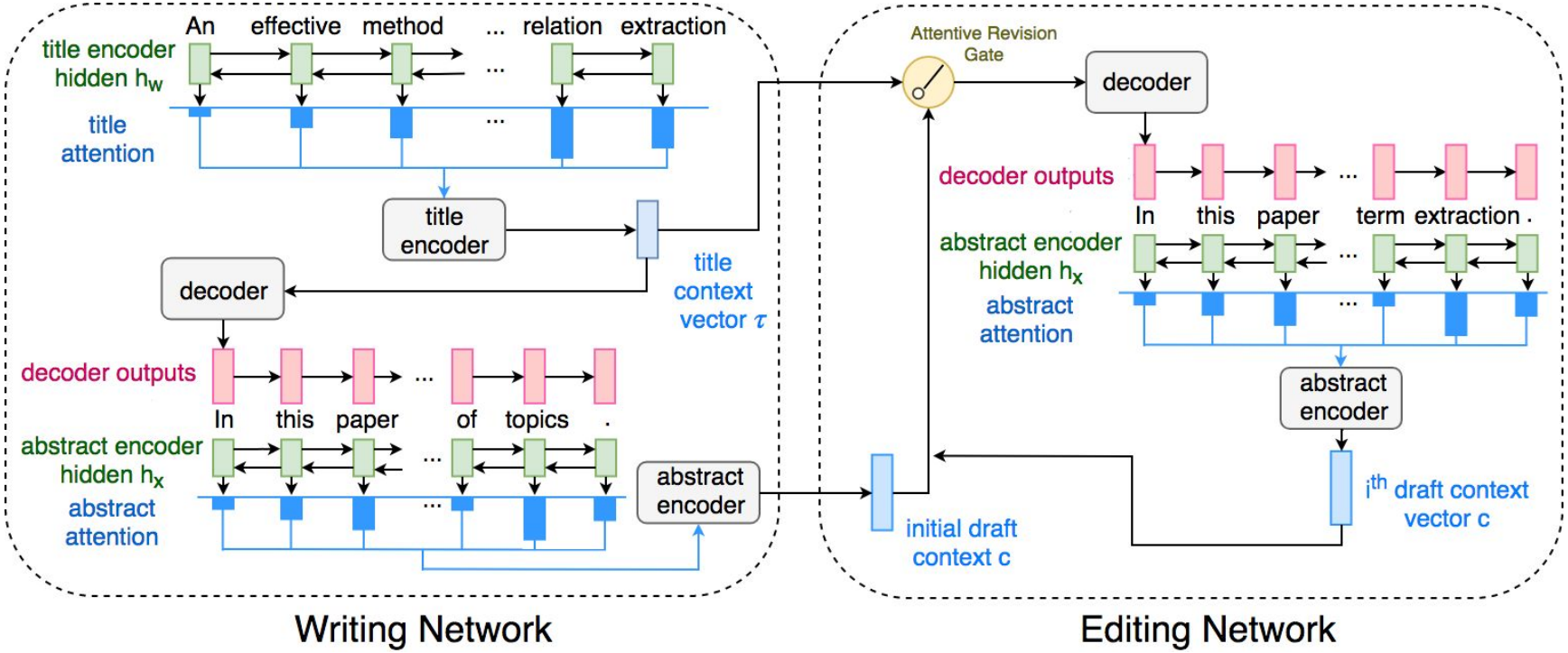
Remaining challenges

- Noun to the adjective form when it appears right before a position slot
- Temporal expressions
- Person genders
- More details about the entity

Paper Abstract Writing through Editing Mechanism

Qingyun Wang*, Zhihao Zhou*, Lifu Huang, Spencer Whitehead, Boliang Zhang, Heng Ji, Kevin Knight

Model:


An effective method ... relation extraction
title encoder hidden h_w
title attention
title encoder
title context vector τ
decoder
decoder outputs
In this paper of topics .
abstract encoder hidden h_x
abstract attention
abstract encoder
Writing Network
Attentive Revision Gate
decoder
decoder outputs
In this paper term extraction .
abstract encoder hidden h_x
abstract attention
abstract encoder
i^{th} draft context vector c
initial draft context c
Editing Network

An effective method of using **Web** based **information Relation Extraction**

Initial draft: In this paper we focus on the task of **extracting** the most commonly used **Web** pages in named entity recognition (IE) **information** from a large corpus of news articles. In particular, we propose a new kernel method that can be applied to a large collection of **Web** pages by automatic detection of labeled and unlabeled data. Our method is very simple and because the related method of using the data to be used as a guide as a source of **information** for the topic detection and classification can be applied, but often fail at improving the performance by combining **relations** and previous methods. In this method, the co-occurrence counts and the occurrences of each **relation** are not statistically significant, but also that it is possible to contribute to entity **relations** with a term in a variety of topics.

Final draft: In this paper we focus on the task of identifying the most commonly relevant features of **Web** documents. In particular, we propose a generic, automated IE algorithm that can be applied to a large collection of Web pages containing full large documents. This is a first step in helping a wide range of collaborative works for relation extraction. We show that it is possible to eliminate a good number of errors in **relation extraction** from a variety of documents, but that it is difficult to define a problem of term **extraction**.

Challenges

- Machines lack knowledge of the deep connections among scientific knowledge elements
  - *“…a translation system to generate a parallel corpus…”*
  - *“…automatic generation of English verbs…”*.
- Humans know better about what terms are more important than others in a title
  - if a language name appears in the title, it must appear in the abstract.