Paper Abstract Writing through Editing Mechanism

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

Goals:

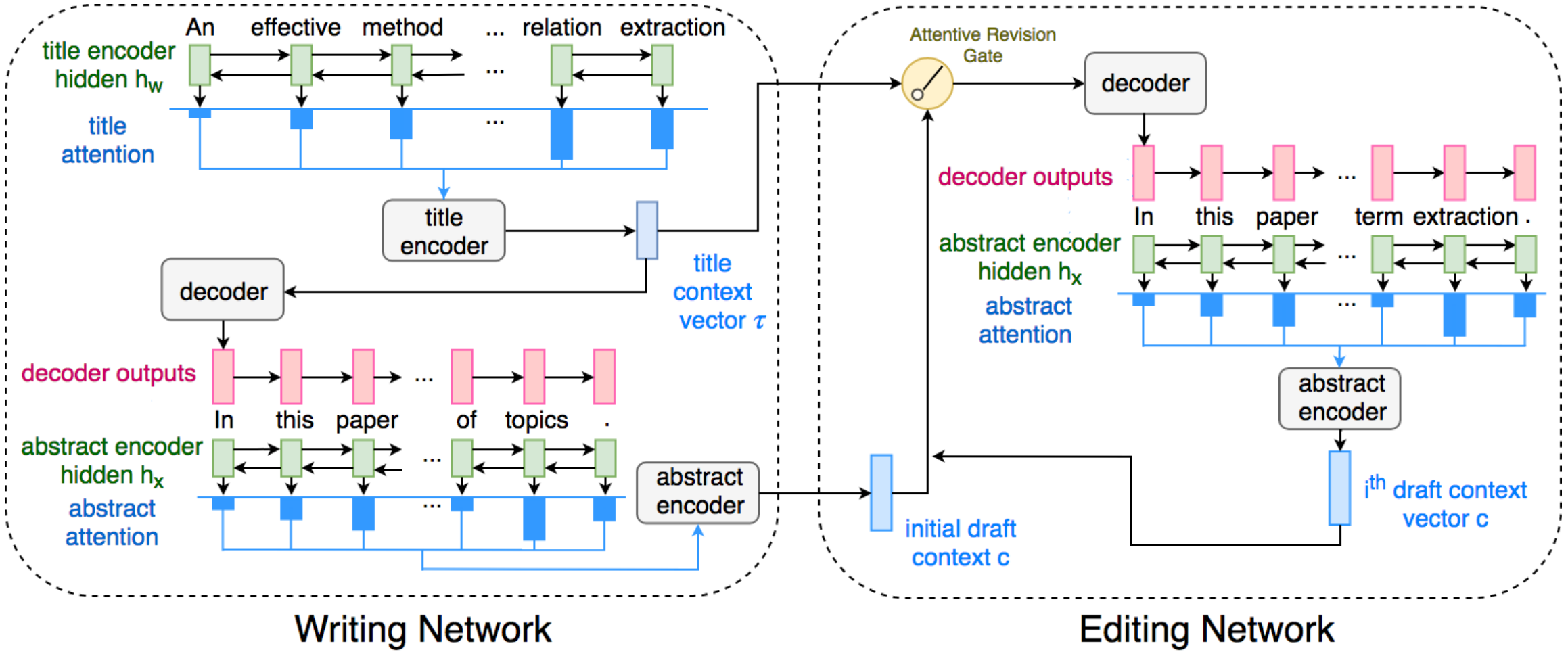
- Given a title, generate paper abstract without any explicit KB
- An assistive technology for human's further editing, correction and enrichment

"It is perfectly okay to write garbage — as long as you **edit** brilliantly."

— by C. J. Cherryh

Writing-editing Network with Attentive Revision Gate

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**.

Explanations of iterations:

- The attentive revision gate incorporates the knowledge from the paper title and the previous generated abstract.
- The knowledge pool will diverge since in each iteration the generated abstract may introduce some irrelevant information

n	1	2	3	4	5	6
METEOR	13.3	**14.0**	13.6	13.9	13.8	13.5
ROUGE-L	**20.3**	19.8	18.6	19.2	18.9	18.8

Evaluation and Turing test:

Method	METEOR	ROUGE-L	HUMAN PREFER-ENCE
LSTM-LM	8.7	15.1	0
Seq2seq	13.5	19.2	22
ED(1)	13.3	**20.3**	30
ED(2)	**14.0**	19.8	**48**

Table 2: Method Comparison (%).

	# Tests	# Choices per Test	Non-expert		NLP Expert	
			Non-CS	CS	Junior	Senior
Different Titles	50	2	30%	15%	12%	0%
	20	5	60%	20%	30%	20%
	10	10	80%	30%	30%	20%
Same Title	50	2	54%	10%	4%	0%
	20	5	75%	25%	5%	5%

Table 4: Turing Test Passing Rates.

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.

Challenges:

- Human written abstracts are generally more specific, concise, and engaging
  - *lab names, author names (e.g., "Collins proposed…"), system abbreviations, and terminologies (e.g., "Italian complex nominals (CNS) of the type n+p+n")*

  System occasionally generates too general descriptions
  - *"Topic modeling is a research topic in Natural Language Processing."*

Challenges:

- Machines lack common sense knowledge
  - *a system generated abstract may mention three areas/steps, but only outline two of them*

Machines lack logical coherence.

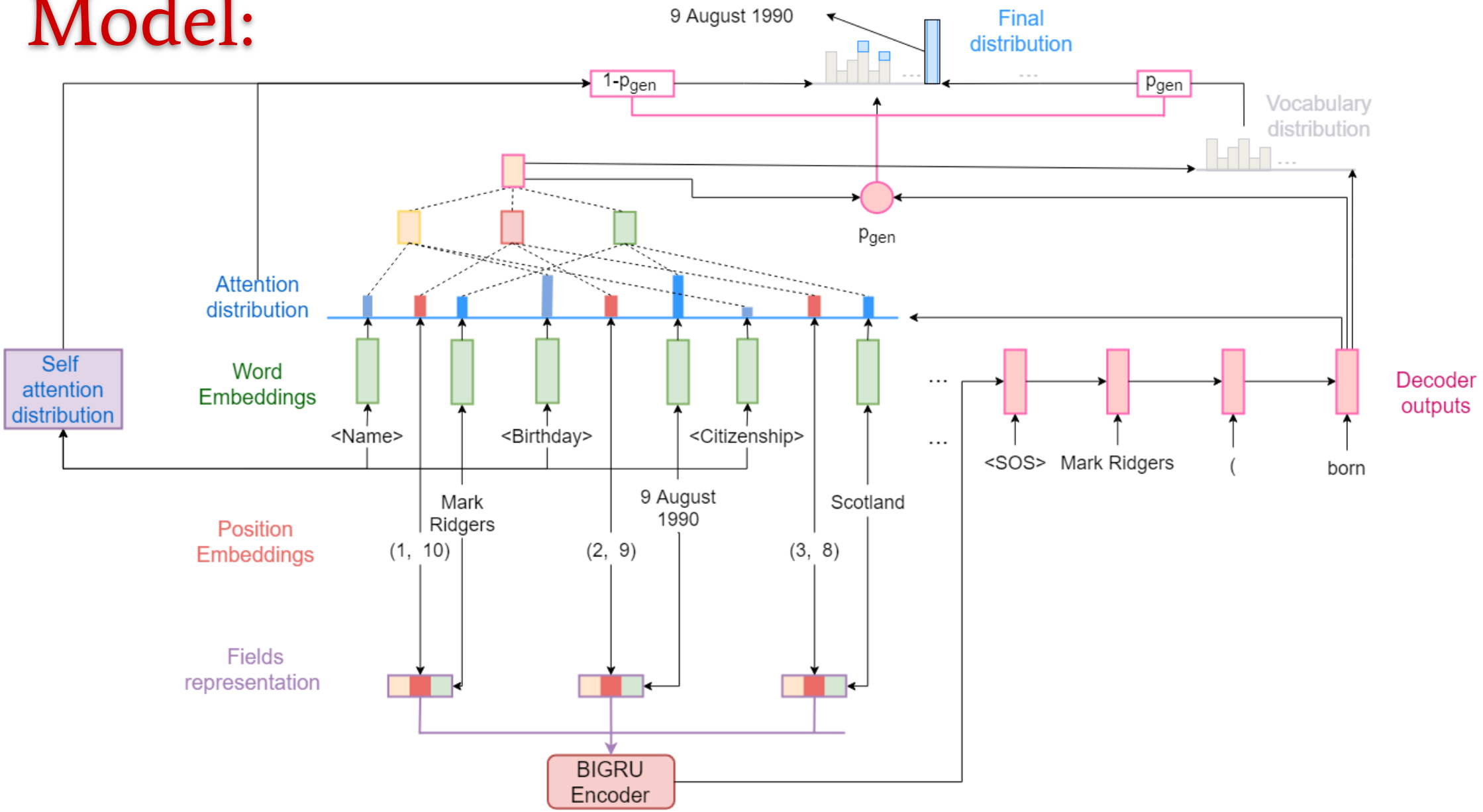
- *A system generated abstract may contain "The two languages…" and not state which languages.*

We are not asking the system to perform scientific experiment; system generated "*experimental results*" are often invalid

- *"Our system ranked first out of the participating teams in the field of providing such a distribution"*

Wikipedia generation

Model:


9 August 1990
Final distribution
1-pgen
pgen
Vocabulary distribution
pgen
Attention distribution
Self attention distribution
Word Embeddings
<Name>
<Birthday>
<Citizenship>
Position Embeddings
Mark Ridgers
(1, 10)
9 August 1990
(2, 9)
Scotland
(3, 8)
Fields representation
BIGRU Encoder
<SOS>
Mark Ridgers
(
born
Decoder outputs

Knowledge Base

Field	Value
Name_ID	Fabio Giordano
date of birth	6 February 1983
country of citizenship	Italy
place of birth	Palermo
sport	Association football
member of sports team	A.C. Jesolo Torino F.C. A.C. Pisa 1909 S.S.D. Vis Pesaro 1898 S.E.F. Torres 1903 Foggia Calcio A.S. Cittadella Delfino Pescara 1936

Reference

Fabio Giordano (born 6 February 1983) is an Italian Association football who plays for A.C. Jesolo . born in Palermo Sicily Giordano started his career with northern Italy side Torino F.C. . he was part of 2001-02 Primavera Under-20 team and then loaned to A.C. Pisa 1909 and S.S.D. Vis Pesaro 1898 . he was transferred to S.E.F. Torres 1903 again in January 2005 . after the bankrupt of Torino and all players were allowed to leave and Giordano joined Foggia Calcio of Serie C1 . in July 2008 he joined newly promoted Serie B side A.S. Cittadella but loaned to Delfino Pescara 1936 on 2 February .

Output

Fabio Giordano (born 6 February 1983 in Palermo) is an Italy Association football who plays for Torino F.C. in the Italian Serie A for Foggia Calcio . he has played for the Foggia Calcio and Delfino Pescara 1936 .