Query and Output: Generating Words by Querying Distributed Word Representations for Paraphrase Generation

Shuming Ma[1] Xu Sun[1,2] Wei Li[1] Sujian Li[1] Wenjie Li[3] Xuancheng Ren[1]

[1]MOE Key Lab of Computational Linguistics, School of EECS, Peking University
[2]Deep Learning Lab, Beijing Institute of Big Data Research, Peking University
[3]Department of Computing, The Hong Kong Polytechnic University

{shumingma, xusun, liweitj47, lisujian, renxc}@pku.edu.cn}
cswjli@comp.polyu.edu.hk

Abstract

●**Task:**

Paraphrase generation

Problem:

The existing sequence-to-sequence model tends to memorize the words and the patterns in the training dataset instead of learning the meaning of the words. Therefore, the generated sentences are often grammatically correct but semantically improper.

Proposal:

a novel model based on the encoder-decoder framework, called Word Embedding Attention Network (WEAN). Our proposed model generates the words by querying distributed word representations (i.e. neural word embeddings), hoping to capturing the meaning of the according words.

Example of RNN Generated Summary

Text: 昨晚，中联航空成都飞北京一架航班被发现有多人吸烟。后因天气原因，飞机备降太原机场。有乘客要求重新安检，机长决定继续飞行，引起机组人员与未吸烟乘客冲突。

Last night, several people were caught to smoke on a flight of China United Airlines from Chendu to Beijing. Later the flight temporarily landed on Taiyuan Airport. Some passengers asked for a security check but were denied by the captain, which led to a collision between crew and passengers.

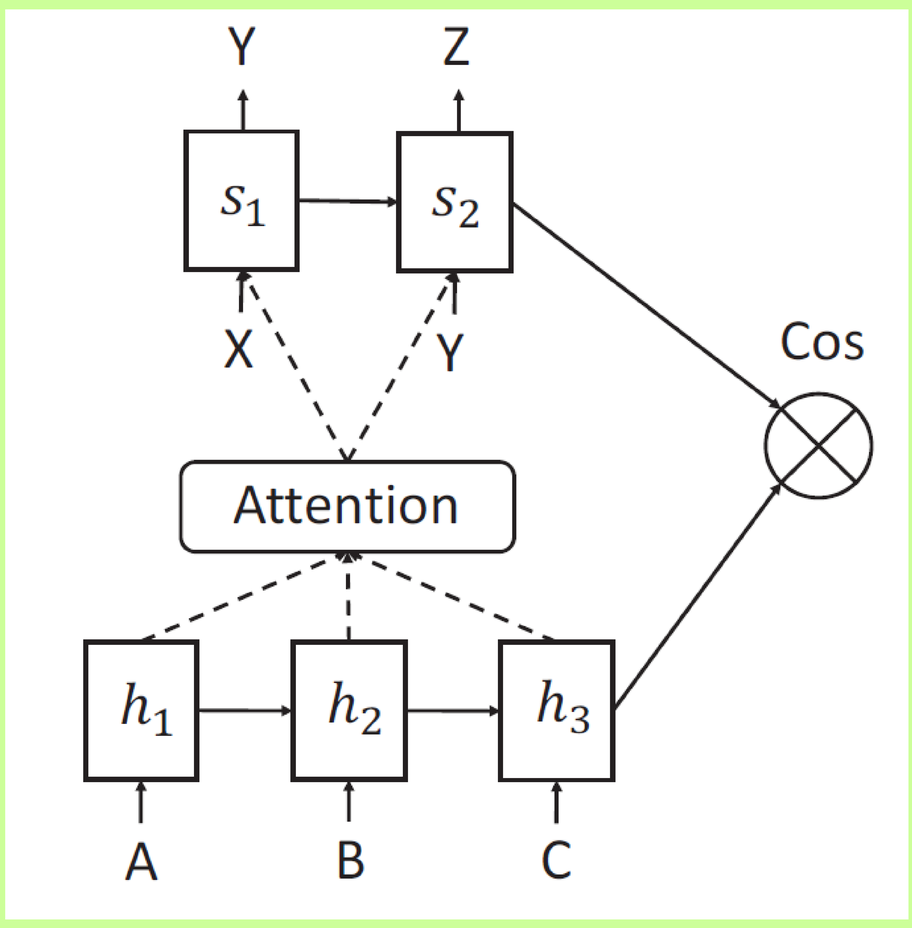
RNN: 中联航空机场发生爆炸致多人死亡。

China United Airlines exploded in the airport, leaving several people dead.

Gold: 航班多人吸烟机组人员与乘客冲突。

Several people smoked on a flight which led to a collision between

Proposed Model



Our Semantic Relevance Based neural model. It consists of decoder (above), encoder (below) and cosine similarity function.

Proposed Model

●**Text Representation**

Source text representation $V_t = h_N$

Generated summary representation $V_s = s_M - h_N$

●**Semantic Relevance**

cosine similarity function $cos(V_s, V_t) = \frac{V_t \cdot V_s}{\|V_t\| \|V_s\|}$

●**Training**

Objective function $L = -p(y|x;\theta) - \lambda\, cos(V_s, V_t)$

Experiments

●Dataset:

Large Scale Chinese Short Text Summarization Dataset (LCSTS)

Model	ROUGE-1	ROUGE-2	ROUGE-L
RNN (W) (Hu et al., 2015)	17.7	8.5	15.8
RNN (C) (Hu et al., 2015)	21.5	8.9	18.6
RNN context (W) (Hu et al., 2015)	26.8	16.1	24.1
RNN context (C) (Hu et al., 2015)	29.9	17.4	27.2
RNN context + SRB (C)	32.1	18.9	29.2
+Attention (C)	**33.3**	**20.0**	**30.1**

Results of our model and baseline systems. Our models achieve substantial improvement of all ROUGE scores over baseline systems. (W: Word level; C: Character level).

Example of SRB Generated Summary

Text: 仔细一算，上海的互联网公司不乏成功案例，但最终成为BAT一类巨头的几乎没有,这也能解释为何纳税百强的榜单中鲜少互联网公司的身影。有一类是被并购，比如：易趣、土豆网、PPS、PPTV、一号店等；有一类是数年偏安于细分市场。

With careful calculation, there are many successful Internet companies in Shanghai, but few of them becomes giant company like BAT. This is also the reason why few Internet companies are listed in top hundred companies of paying tax. Some of them are merged, such as Ebay, Tudou, PPS, PPTV, Yihaodian and so on. Others are satisfied with segment market for years.

Gold:为什么上海出不了互联网巨头？

Why Shanghai comes out no giant company?

RNN context:上海的互联网巨头。

Shanghai's giant company.

SRB:上海鲜少互联网巨头的身影。

Shanghai has few giant companies.

Conclusion

Our work aims at improving semantic relevance of generated summaries and source texts for Chinese social media text summarization. Our model is able to transform the text and the summary into a dense vector, and encourage high similarity of their representation. Experiments show that our model outperforms baseline systems, and the generated summary has higher semantic relevance.