What Can Summarization Learn from Machine Translation and Response Generation?

Ziqiang Cao

Department of Computing, The Hong Kong Polytechnic University

polyu.ziqiang@connect.polyu.hk

March 3, 2017

Overview

Motivation

- A hot topic: Natural Language Generation
- Three typical tasks:
  - Text Summarization (TS)
  - Machine Translation (MT)
  - Response Generation (RG)

A Natural Question

What can summarization learn from other tasks?

Task Definition

Text Summarization

Reduce a text document but retain the most important points

$$Document \rightarrow Summary$$

Machine Translation

Use computers to translate text from one language to another

$$Source\ Language \rightarrow Target\ Language$$

Response Generation

Given a conversational stimulus, generate an appropriate response

$$Stimulus \rightarrow Response$$

Semantic Relations between Input and Output

Task Natures

Machine Translation $Y = X$

Summarization $Y \subseteq X$

Response Generation Y should be largely different from X

Hot Topics

Machine Translation Reconstruction: $Y \rightarrow X$

Response Generation Diversity: $X \rightarrow [Y_1, \cdots, Y_n]$

- Summarization seems to lie in the middle
- It can focus on both aspects: **reconstruction and diversity**

Problems in NMT

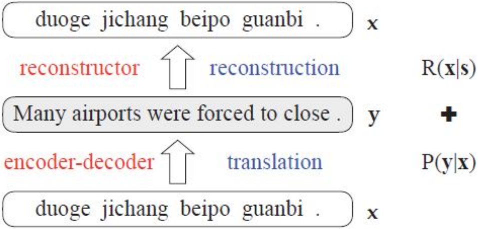
- The decoder often repeatedly selects some parts of the source sentence while ignoring other parts

  $\rightarrow$ Over-translation and Under-translation

- Likelihood favors short translations

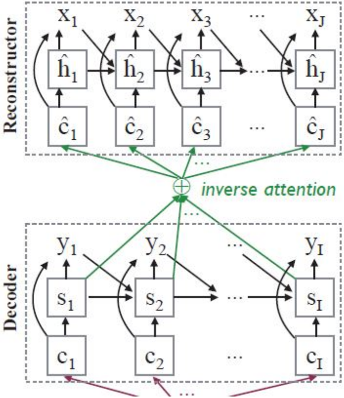
Solution: Reconstruction Cost

Leverage the reconstruction score as an auxiliary objective to measure the adequacy of translation candidate



Tu et al. (2017)
Neural Machine Translation with Reconstruction
AAAI 2017.

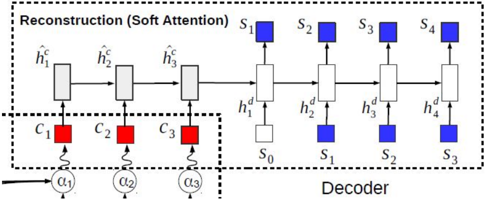
Model Framework



- Read the hidden state sequence from the decoder
- Output a score of exactly reconstructing the input sentence

$$X, Y \rightarrow X$$

Relation with Reinforcement Learning

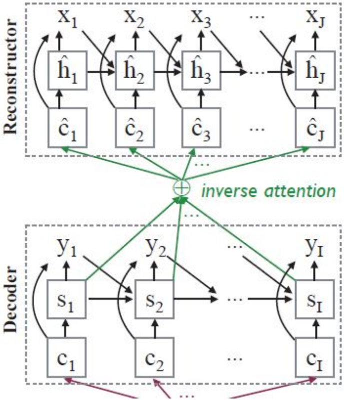


Conditioned on the outputs

$$Y \rightarrow X$$

Yishu Miao and Phil Blunsom. (2016)
Language as a Latent Variable: Discrete Generative Models for Sentence Compression
EMNLP

Model Framework



- Read the hidden state sequence from the decoder
- Output a score of exactly reconstructing the input sentence

$$X, Y \rightarrow X$$

Relation with Maximum Mutual Information Cost

[Li et al, 2015] proposed to use replace log-likelihood with Maximum Mutual Information (MMI) as the objective function:

$$\text{MMI} = logP(Y|X) - logP(Y)$$

According to Bayesian Formula, we can convert the cost to:

$$logP(Y|X) + logP(X|Y)$$

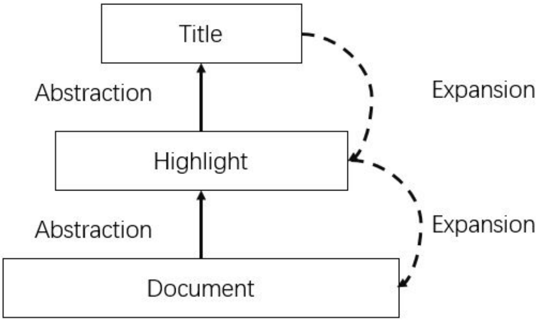
The same as the reconstruction cost

Li et al. (2015)
A diversity-promoting objective function for neural conversation models
arXiv preprint arXiv:1510.03055

Inspiration for Summarization

- Reconstruction $(X, Y \rightarrow X)$ or Expansion $(Y \rightarrow X)$
- Hierarchical abstraction based on different granularities
  - Highlights
  - Title
  - Keywords

Hierarchical Abstraction with Reconstruction/Expansion

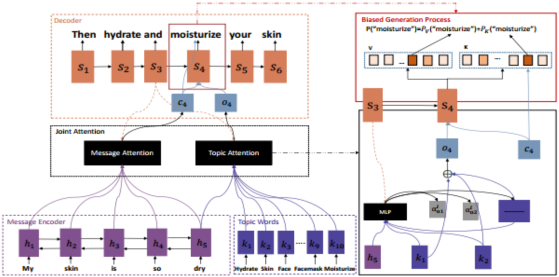


Considering that titles are much more frequent than highlights, this framework can work in a semi-supervised scenario

Diversity in Response Generation

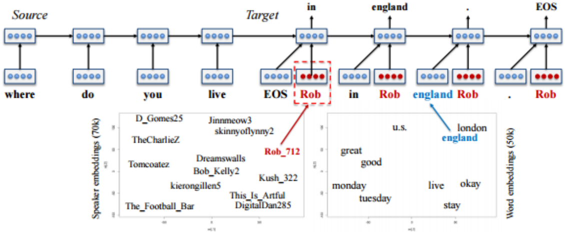
- A stimulus is usually linked with numerous responses
- Responses can vary to a large extent
  - Topic
  - User profile
  - Stance
  - Personal writing style

Topic Diversity



Xing et al. (2017)
Topic Aware Neural Response Generation
AAAI 2017.

User Profile Diversity



Li et al. (2016)
A persona-based neural conversation model
arXiv preprint arXiv:1603.06155.

Inspiration for Summarization

Summaries also hold diversity, such as:

- Text category [Cao et al, 2017]
- Personality [Diaz et al, 2007]

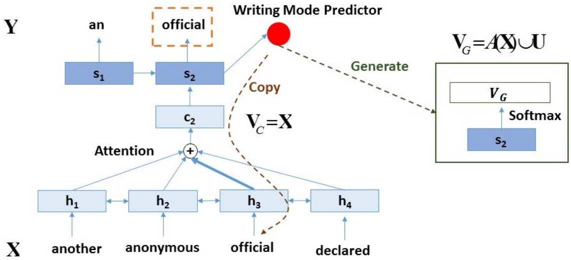
Cao et al. (2017)
Improving Multi-Document Summarization via Text Classification
AAAI

Diaz et al. (2007)
User-model based personalized summarization
Information Processing & Management

Uniqueness in Summarization: Copying Mechanism

- Summarization is not simply the intersection of machine translation and response generation
- It holds unique properties: copying mechanism
  - Most keywords in the source will be reserved in the summary
  - It is a strong baseline to directly select salient sentences to form the summary

Joint Copying and Restricted Generation



Cao et al. (2017)
Joint Copying and Restricted Generation for Paraphrase
AAAI

Conclusion

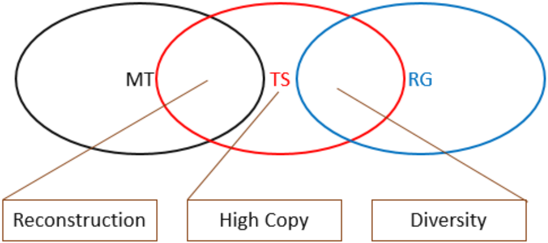


Figure: Relations between summarization and machine translation/response generation

Q&A