since 1995

Data Science Natural Language Processing NLP Papers Summary

Day 121: NLP Papers Summary – Concept Pointer Network For Abstractive Summarization

By Ryan 30th April 2020 No Comments

Objective and Contribution

Proposed Concept Pointer Network for abstractive summarisation, which uses knowledge-based and context-aware conceptualisations to derive a set of candidate concepts. The model would then choose between the concept set and the original source text when generating abstractive summaries. Both automatic and human evaluation were conducted on generated summaries.

The proposed Concept Pointer Network doesn't simply copy text from the source docum ^ , it would also generate new abstract concepts from human knowledge as shown below:

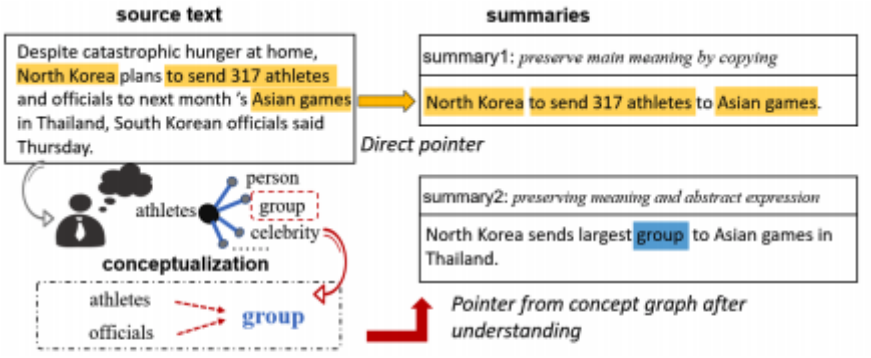

Figure 1: "summary1" only copies keyword from the source text, while "summary2" generates new concepts to convey the meaning.

On top of our novel model architecture, we also proposed a distant supervised learning technique to allow our model to adapt to different datasets. Both automatic and human evaluation shown strong improvement over SOTA baselines.

The Proposed Model

Our model architecture consists of two modules:

1. Encoder-Decoder
2. Concept Pointer Generator

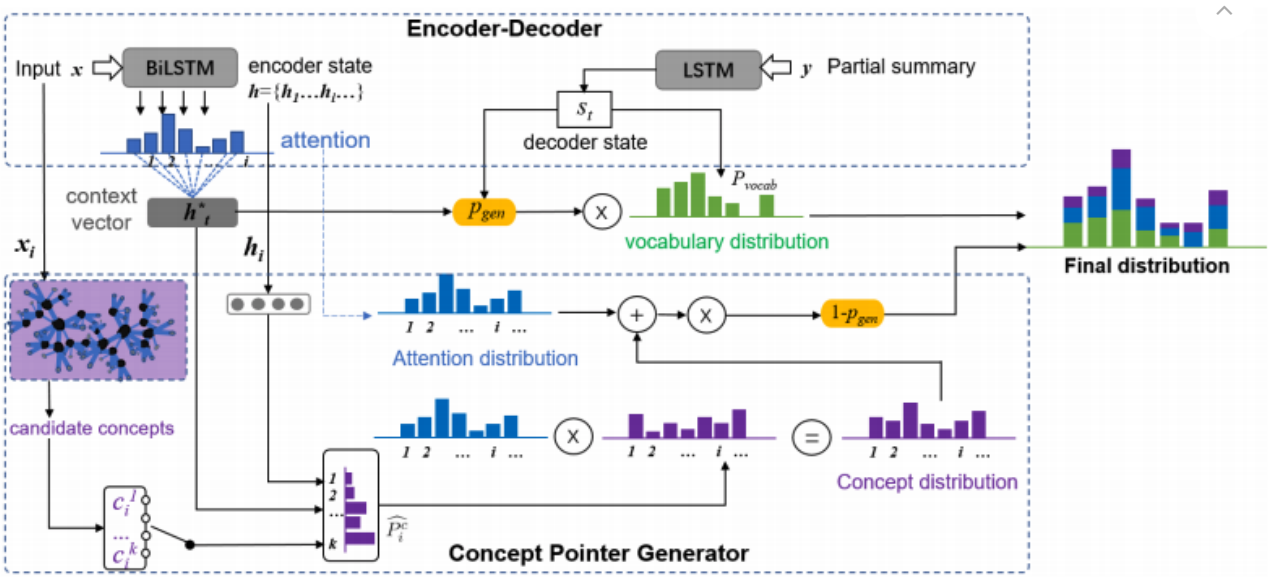

Figure 2: The architecture of our model. Blue bar represents the attention distribution over the inputs. Purple bar represents the concept distribution over the inputs. Noted that, this distribution can be sparse since not every word has its upper concept. Green bar represents the vocabulary distribution generated from seq2seq component.

ENCODER-DECODER

The encoder-decoder framework consists of a two-layer bidirectional LSTM-RNN encoder and an one-layer LSTM-RNN decoder with attention mechanism. Each word in the input sequence is represented by the concatenation of the forward and backward hidden states. The context vector is computed by applying the attention mechanism over the hidden state representations. This context vector is feed to our decoder, where it will use the context vector to determine the probability to generating new words (p_gen) from our vocabulary distribution.

CONCEPT POINTER GENERATOR

Firstly, we use the Microsoft Concept Graph to map a word to its related concepts. This knowledge base covers a huge concept space and the relationships between concepts and entities are probabilistic depending on how strongly related they are. Essentially, the concept graph will take in the word and estimates the probability that this word belongs to a particular concept, p(c|x). With probabilities, this means that given each word, the concept graph will have a set of concept candidates (with different confidence level) that it believes the word belongs to. In order for our model to select the right concept candidate, for example, distinguishing between fruit and company concept for the word “apple”, we will use the context vector from the encoder-decoder framework.

We will use the context vector to update the concept distribution. We compute the updated weights by feeding the current hidden state, the context vector, and the current concept candidate into a softmax classifier. This updated weight is then added to the existing concept probability to factor in the context of the input sequence, allowing us to derive the context-aware concept probability.

Our concept pointer network, consists of the normal pointer to the source document as well as a concept pointer to relevant concepts given the source document. The concept pointer is scaled element-wise by the attention distribution and are added to the normal pointer (attention distribution). This would be the copy distribution where the model copies from and it includes concept distribution on top of the usual text distribution over original source document.

DISTANT SUPERVISION FOR MODEL ADAPTION

If the summary-document pairs of our training set are different than to the testing set, our model would perform poorly. To counter this, we would want to retrain our model to lower this dissimilarity in our final loss. To do so, we need labels to indicate how close our training set is to our test set. In order to create these labels, we use KL divergence between each training reference summary and a set of documents from the test set. In other words, the training pairs are distantly-labelled. The representations of both reference summaries and documents are computed by summing the constituent word embeddings. This KL divergence loss function is included in the training process and it measures the overall distance between the test set and each of our reference summary-document pairs. This allows us to determine whether our training set is relevant or irrelevant for model adaption.

Experimental Setup and Results

There are two evaluation datasets: Gigaword and DUC-2004. The evaluation metric is the ROUGE score.

MODELS COMPARISON

There are 8 baseline models:

1. *ABS+*. Abstractive summarisation model
2. *Luong-NMT*. LSTM encoder-decoder

3. *RAS-Elman*. CNN for encoder and RNN with attention for decoder
4. *Seq2seq+att*. BiLSTM encoder and LSTM with attention decoder
5. *Lvt5k-lsent*. Uses temporal attention on decoder to reduce repetition in summary
6. *SEASS*. Uses selective gate to control information flowing from encoder to decoder
7. *Pointer-generator*. Normal PG
8. *CGU*. Uses convolutional gated unit and self-attention for encoding

RESULTS

In table 1, our concept pointer outperformed all the baseline models on all metrics except RG-2 on Gigaword (CGU scored the highest). In table 2, we show that the summaries generated by concept pointer has the lowest percentage of UNK words, alleviated the OOV problem. In table 3, we showcase the abstractiveness of our generated summaries. We show that the summaries

generated by our concept pointer has a relatively high abstractiveness level and it's close to the reference summary level.

We experimented with two different training strategies: Reinforcement learning (RL) and Distant supervision (DS). Both training strategies applied to concept pointer outperformed the normal concept pointer. Furthermore, in the DUC-2004 dataset, concept pointer + DS outperformed concept pointer + RL consistently, showcasing the effect of distant supervision for better model adaption.

CONTEXT-AWARE CONCEPTUALISATION

We want to measure the impact of concept update strategy and so we have experimented with different number of concept candidates. The results are as shown below. There are only small variation in ROUGE scores between different number of concept candidates.

HUMAN EVALUATIONS

We conducted human evaluations where each volunteer has to answer the following questions:

1. *Abstraction* – How appropriate are the abstract concepts in the summary?

2. *Overall Quality* – How readable, relevant, and informative is the summary?

We randomly selected 20 examples, each with three different summaries (from three models) and score how often does each type of summary gets pick. The results are shown below, which showcase the concept pointer network outperformed both the seq2seq model and pointer generator. The generated summaries seems to be fluent and informative, however, it's still not as abstractive as human reference summaries.

Conclusion and Future Work

On top of our novel model architecture, we also proposed a distant supervised learning technique to allow our model to adapt to different datasets. Both automatic and human evaluation shown strong improvement over SOTA baselines.

Source: **https://arxiv.org/pdf/1910.08486.pdf**

Ryan

Data Scientist

Previous Post

Next Post

Day 120: NLP Papers Summary - A Simple Theoretical Model of Importance for Summarization

Day 122: NLP Papers Summary - Applying BERT to Document Retrieval with Birch

Data Science Natural Language Processing NLP Papers Summary

Day 365: NLP Papers Summary – A Survey on Knowledge Graph Embedding

Ryan
30th December 2020

Data Science Implementation Natural Language Processing

Day 364: Ryan's PhD Journey – OpenKE-PyTorch Library Analysis + code snippets for 11 KE models

Ryan
29th December 2020

Data Science Ryan's PhD Journey

Day 363: Ryan's PhD Journey – Literature Review – Knowledge Acquisition – 1st Passes XIV

Ryan
28th December 2020