Focus-based Text Summarisation with Hybrid Embeddings

Jingli Shi Li[1], Lui Hellesoe[1], Guan Wang[1] Weihua Li[1], Quan Bai[2]
Auckland University of Technology, New Zealand
[2] University of Tasmania, Australia

Abstract

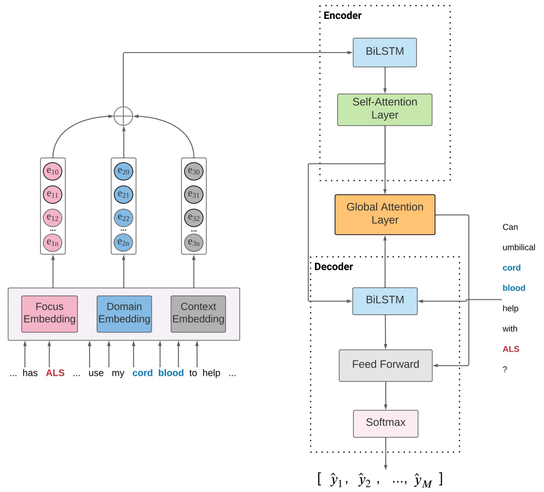
We propose a novel summarisation model aimed at producing summaries by focusing on the domain-specific knowledge, where hybrid embeddings, i.e., focus, domain and context embeddings, are utilised. We conduct extensive experiments to evaluate our model by using the medical dataset MeQSum for domain-specific summarisation.

Introduction

Domain focus, referred to as the domain-specific information, is acknowledged as one of the main factors, significantly influencing the summary generation process and determining the desired subject of the summary. We develop a module with a fine-tuned language model on the in-domain datasets to extract the domain focus. Named Entity Recognition is performed to extract the given text's domain-specific entities automatically. Both focus and domain context is involved as the guidance I the summarisation process. Therefore, the novel model can address the challenges of missing critical domain-specific keywords in the summary.

Advantages of Focus-based

- The generated summary is domain-based instead of common word-based.
- Focus embeddings ensure focus words are included in the summary.
- For example: 'Acute myeloblastic leukaemia with minimal maturation' are included in reference summary 'Is there an ayurvedic treatment for Acute myeloblastic leukaemia with minimal maturation? '.



Problem Formulation

We formally define the text summarisation problem. Given a sentence, denoted by $qs = \{w_1, w_2, \dots, w_N\}$ where N denotes the number of words. The objective of the text summariser is to learn the mapping $qs \rightarrow y \cdot y = \{y_1, y_2, \dots, y_M\}$ presents the generated summary with M words.

To achieve domain-specific summarisation, we propose a dynamic embedding strategy by fusing focus embeddings, domain embeddings and context embeddings.

Question & Summary Examples

Question:

SUBJECT: Blood Sugar Levels and Parkinson's

MESSAGE: I'm wondering if there is a correlation between blood sugar level's and how it may effect the presentation of Parkinson's particularly in tremors. It seems that extreme blood sugar levels would make the tremors a great deal worse and appearing none typical.

Summary: Is there a connection between blood sugar levels and symptoms of Parkinson's disease?

Question:

Can arteries in any part of the body spasm or is this only possible with coronary arteries? If so, does someone with vasospastic angina have a greater chance of developing spasms elsewhere in the body and is the treatment the same?

Summary: What are the causes of and treatment for artery spasms?

Hybrid Embeddings Fusion Algorithm

Mathematically, for each sentence qs, context embedding E^c, domain embedding E^d and focus embedding E^f , are obtained via three separate embedding layers. Then, we concatenate three embeddings $E = E^c \oplus E^d \oplus E^f$ and feed the final embedding into a BiLSTM layer of the encoder.

- $h^e = [\overrightarrow{\text{LSTM}}(e^c; e^d; e^f); \overleftarrow{\text{LSTM}}(e^c; e^d; e^f)]$

One self attention layer is implemented to learnt he long-term dependencies of the input sequence.

- $h^e_{sa} = \sigma\left(\frac{QK^T}{\sqrt{d_k}}\right)V$

Q, K, and V refer to the query, key, and value matricies. σ presents the activation function .

The global attention distribution of the source sequence is calculated as

- $\alpha^e_t = \frac{\exp(s^e_t)}{\sum_{k=1} \exp(s^e_k)}$

Next we define the context vector for the target tokens

- $cv^e = \sum a^e h^e_{sa}$

Finally, the attention hidden state $\widehat{h^a}$ is obtained together with the decoder hidden state h^d , and the vocabulary distribution is calculated

- $\hat{h}^a = W_a(cv^a \oplus h^d) + b_a$
  $P_{vocab} = \sigma(W_z\hat{h}^a + b_z)$

Contact Lui Hellesoe, Auckland University of Technology, mwx9073@autuni.ac.nz | Weihua Li, Auckland University of Technology, weihua.li@aut.ac.nz | Jingli Shi, Auckland University of Technology, jingli.shi@autuni.ac.nz | Quan Bai, Auckland University of Technology, quan.bai@utas.edu.au | Guan Wang, Auckland University of Technology, edmund.lai@aut.ac.nz