A Char-level Encoder-Decoder model for text summarization

A cute kid

A digital version of the manuscript

Abstract. Text summarization belongs to one of the natural language processing tasks that has important application value in the field of artificial intelligence. It typically condenses the meaning of a lengthy document into a concise summary of a few sentences. However, existing models have some drawbacks: they tend to distort the original meaning and are vague in their description of details. In this work, I propose an encoder-decoder architecture that utilizes the encoder to encode the entire document and the decoder to generate the final summary one character step by step. First, I use a simple Gate Recurrent Unit(GRU) network that can capture character representation from the source text via char-level embedding in the Encoder layer. Second, I use the attention mechanism to calculate the context representation from the final output of the encoder and the results of the last hidden layer. Then, I feed the context representation and the output of the last hidden layer of the encoder into a GRU for decoding. The decoder assigns the last layer representation after each decoding to the initial input for each iteration. I applied my model to a text summarization task containing only three pieces of data. After training, the accuracy rate on the training dataset reached 100%.

Keywords: Text summarizationn · Encoder-Decoder · GRU · Attention.

1 Introductionn

The text summarization task refers to converting a document into one or more concise sentences to summarize its overall meaning. Figure 1 demonstrate the main process of the text summarization task. It should be noted that this task does not involve summarizing blue text directly selected and sorted from the document in this paper, but the model generates the corresponding characters step by step through calculation. The great majority of past work is extractive-based and abstractive-based. Extractive-based methods are not recommended because they lack crucial capabilities of high-quality summarization, like generalization, rewriting, and utilization of external knowledge. Though abstractive-based methods have potential, they still exhibit some ambiguous behavior. For example, they cannot accurately reproduce factual details and tend to repeat themselves. In this paper, I propose a model with encoder-decoder atchitecture

Original Text: 6 月 28 日消息, 据国外媒体报道, 在距离地球 8000 光年的天鹅座星群中, 一个黑洞与一颗恒星正宛如天鹅般翩翩起舞。通常, 物质在被黑洞巨大的引力吞没之前, 会被加热并释放能量。这个名为 V404Cygni 的黑洞自 1989 年起就变得相对沉寂, 然而, 最近它又开始变得活跃起来。欧洲航天局于 6 月 26 日报道称, 该机构的 Integra(国际伽玛射线天体物理实验室) 卫星观测到来自黑洞的"异常爆发的高能射线"。V404Cygni 的活动亦被 NASA 的雨燕卫星观测到。该处产生的 X 射线耀斑同时被 MAXI 单元捕获到, 后者是国际空间站日本模块的一部分。6 月 17 日, 欧洲航天局将 Integral 卫星用于观测 V404Cygni, 该机构随后证实, 位于此处的黑洞确实再度活跃, 而且活跃程度相当高。欧洲航天局 Integral 项目科学家埃里克 · 库克斯 (ErikKuulkers) 表示, 耀斑产生的间隔非常短, 不到一个小时, 这种现象在其他黑洞系统中很罕见。其曾经一度成为天空中最明亮的物体, 较蟹状星云还要亮 50 倍。后者通常是通过 X 光观测的天空中最明亮的地方。2013 年 6 月, 曾经有报道称,NASA 观测到位于玉夫座中心的一处巨大黑洞逐渐进入休眠。通常, 在吸收消耗完周围所有物质之后, 黑洞会归于沉寂。在周围还有其他物质存在的情况下, 黑洞进入休眠状态的情形相当罕见。例如, 此次的 V404Cygni 附近就有恒星存在。这些质量巨大的黑洞在休眠与活跃之间变化的机理还尚不完全清楚, 但是我们现在有更多的工具来观测这些过程。上述现象让天文学界感到相当兴奋。

Summarized: 距离地球 8000 光年的天鹅座星群黑洞再度活跃, 在周围还有其他物质的情况下, 黑洞进入休眠很罕见。

Fig. 1. The introduction of the text summarization task.

that can encode the long document and decode the charactor step by step. And I introduce the attention mechanism in the decoder to enhance the model's performance. The contribution of this paper summarized in three points:

1. Improve the labels in the training set.
2. Use SVD to perform dimensionality reduction and visualization analysis on labels.
3. Use heatmap to demonstrate the model's attention weights.

2 Related Work

The inspiration for this paper originated from [1], I tell you straight. And you can keep staring at that paper, just keep staring at it. Don't stop try to do it yourself, then you'll learn it on your own.

3 Methodology

The figure 2 demonstrate the architecture of the model. First, determine the number of different characters in the dataset. Let the number of different characters refer be N. Then, let the shape of the embeddinng layer be $[N, 128]$. Given a document D in one of the training dataset, let the total number of characters in D be L. After passing through the embedding layer, the resulting tensor has a shape of $[1, L, 128]$. Then, send the tensor to the Encoder's GRU layer, the

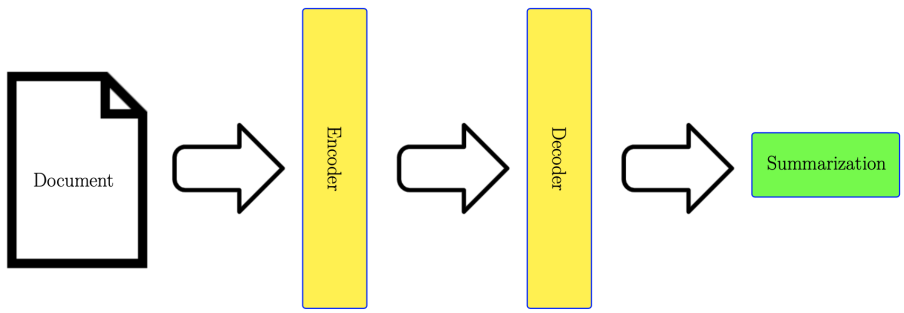

Fig. 2. The architecture of the model.

shape of the final output is $[1, L, 128]$. Then, send the encoder output and the last characters output to the decoder. In the decoder, the attention mechanism is used to calculate the output of the encoder and the output of the last character of the encoder to obtain the context representation of the model. The context representation is a tensor with the shape of $[1, 1, 128]$. Then, send the context representation and the representation of the last character of the encoder to the Decoder's GRU layer to obtain the present hidden tensor and the output of the Decoder's GRU layer. Then, send the output to the linear layer to generate the present character, and the present hidden tensor is to be the input for the GRU layer for the next step, iterate, and calculate. The specific process of label reinforcement involves performing a cross-entropy calculation with the corresponding real label for each character generated and then performing backpropagation.

4 Experimental

I randomly found three pieces of data online. And I set the accuracy to the basic evaluation standard for the model. The Table 1 shows the hypeparameters of the model. Figure 3 demonstrates the relationship between loss values and epochs.

hypeparameters	values
torch random seed	66
batch size	1
Encoder learning rate	0.001
Decoder learning rate	0.001
epoch	4000

Table 1. model hypeparameters.

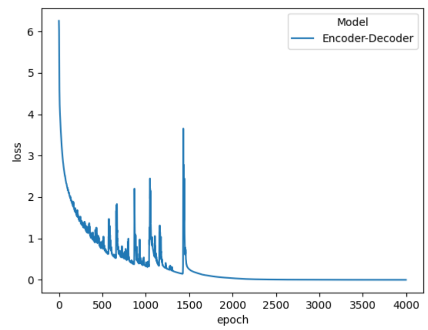

Fig. 3. The relationship between loss and epochs.

5 Analysis

Dimension reduction analysis Figure 4 demonstrates the results of reducing the dimensionality of the model's output.

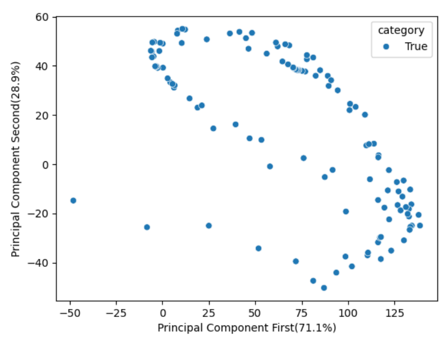

Fig. 4. The relationship between loss and epochs.

Attention analysis Figure 5 demonstrates the attention weights analysis of the model's output. I calculated the cosine similarity of the attention weights between characters and assigned corresponding color depths based on the corresponding values.

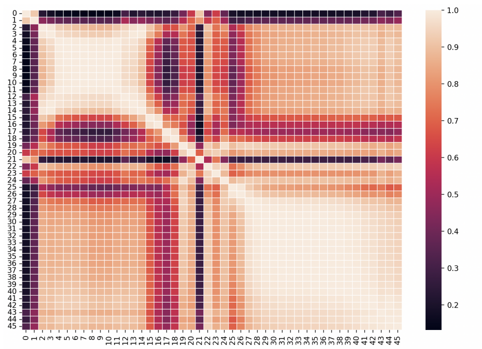

Fig. 5. The attention weights of the model's output.

6 Conclution

This paper propose a Encoder-Decoder model for text summarization task and a brief analysis of the results.

References

1. Get to the point: Summarization with pointer-generator networks,See, Abigail and Liu, Peter J and Manning, Christopher D., arXiv preprint arXiv:1704.04368.