Generative Adversarial Network for Abstractive Text Summarization with Multi-task Constraint

Min Yang[a], Xintong Wang[b], Yao Lu[c], Jianming Lv[b], Ying Shen[d], Chengming Li[a,*]

[a]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences*
[b]*School of Computer Science and Engineering, South China University of Technology*
[c]*School of Computer Science, University of Waterloo*
[d]*School of Electronics and Computer Engineering Peking University Shenzhen Graduate School*

Abstract

Artificial intelligence studies have witnessed great interests in generating concise summaries automatically that retain the salient information of the input text, which is known as the abstractive text summarization task. In this paper, we present a Generative Adversarial Network for Abstractive Text Summarization with Multi-Task constraint (GAN-ATSMT). Specifically, GAN-ATSMT jointly trains a discriminative model D and a generative model G via adversarial learning. The generative model G employs the sequence-to-sequence architecture as its backbone, taking as input the original text and generating a corresponding summary. We employ a convolutional neural network (CNN) classifier as the discriminative model D, which is prone to distinguish the generated summaries by G from the ground truth summaries. The generative model G and the discriminative model D are learned with a minimax two-player game. Thus, this adversarial process can eventually adjust G to produce high-quality and plausible summaries. Furthermore, we additionally propose extended regularizations for the generative model G using the multi-task learning, sharing its LSTM encoder and LSTM decoder with text categorization task and syntax annotation task, respectively. The

*Chengming Li is the corresponding author.
Email addresses: min.yang1129@gmail.com (Min Yang), w.xintong@mail.scut.edu.cn (Xintong Wang), luyao@uwaterloo.ca (Yao Lu), jmlv@scut.edu.cn (Jianming Lv), shenying@pkusz.edu.cn (Ying Shen), cm.li@siat.ac.cn (Chengming Li)

auxiliary tasks help to improve the quality of locating salient information of a document and generate high-quality summaries from language modeling perspective alleviating the issues of incomplete sentences and duplicated words. We conduct extensive experiments to estimate the effectiveness of GAN-ATSMT on two real-life datasets. Experimental results illustrate that GAN-ATSMT is able to achieve better performance than the state-of-the-art abstractive text summarization models in terms of the quantitative evaluation metrics (i.e., ROUGE and human evaluation) and qualitative evaluation.

Keywords: Abstractive Text Summarization, Generative Adversarial Network, Multi-task Learning, Dependency information, Guidance summary

---

1. Introduction

The goal of abstractive text summarization is to generate concise and condensed summaries which retains the important information of the input text. As opposed to extractive text summarization, which identifies the best summarizing components from the course text, abstractive text summarization models can generate summaries containing new words and phrases that do not appear in the original text. Recently, abstractive text summarization has attracted much attention due to its board applications for information condensation. However, generating a plausible and high-quality abstractive summary is a challenging task in practice since the computer lacks language capability and human prior knowledge to understand the entire document and then generate a condensed summary highlighting the main points of the input article.

Motivated by the remarkable progress of the encoder-decoder framework in dialogue generation, neural machine translation and image captioning, most abstractive text summarization systems adopt the sequence-to-sequence (seq2seq) method to produce summaries (Rush et al., 2015; Nallapati et al., 2016). The general idea of the seq2seq framework is to use a long short-term memory (LSTM) to encode the input text and then feed the representation vector to LSTM decoder to generate summaries. These seq2seq based approaches have become the mainstream due to their capability of capturing the syntactic and semantic relations between raw texts and summarises in an end-to-end and scalable way. Despite the significant success of existing

Category: Politics
Article (truncated): "isis claimed it controlled part of iraq's largest oil refinery sunday, posting images online that purported to show the storming of the facility, fierce clashes and plumes of smoke rising above the contested site. the group said it launched an assault on the baiji oil refinery late saturday. by sunday, isis said its fighters were inside the refinery and controlled several buildings, but iraqi government security officials denied that claim and insisted iraqi forces remain in full control. cnn couldn't independently verify isis' claim. it wouldn't be the first time that militants and iraqi forces have battled over the refinery, a key strategic resource that has long been a lucrative target because the facility refines much of the fuel used by iraqis domestically. if an attack damaged oil fields or machinery, it could have a significant impact. the refinery is just 40 kilometers (25 miles) from the northern iraqi city of tikrit, which iraqi forces and shiite militias wrested from isis less than two weeks ago. cnn's jennifer deaton and catherine. shoichet contributed to this report."
Reference summary: "isis says it controls several buildings at the baiji oil refinery. iraqi government security officials say iraqi forces remain in full control the refinery, iraq 's largest, has long been a lucrative target for militants."
Category: Sports
Article (truncated): "Article (truncated): serena williams claimed her eighth miami open title in 14 years after ruthlessly brushing aside the challenge of 12th seed carla suarez navarro in saturday's final. the world no 1 won the final 10 games in a 6-2 6-0 demolition of her spanish opponent to claim her third straight title at an event she has dominated since winning her first crown back in 2002. serena williams poses on the beach with the championship trophy after defeating carla suarez navarro. williams poses with the road to singapore sign post on crandon park beach after her straight-sets victory. (.....) williams saved a break point in her opening service game and then broke to love in the next to leave suarez navarro with a mountain to climb. there would be no way back, with suarez navarro winning just two points on serve in the second set. williams, meanwhile, won 21 of 22 points on her first serve. spain's suarez navarro started strongly but was no match for serena williams. williams once again broke to love to move 5-0 ahead before clinically wrapping up the match inside 57 minutes. suarez navarro was quick to hail williams, adding in quotes broadcast by bt sport 1: all that you have, you deserve and for me you are the number one right now."
Reference summary: "serena williams won her eighth miami open title on saturday. she won the final 10 games in a 6-2 6-0 demolition in miami. the unbeaten world no 1 has now won 12 consecutive finals."

Table 1: Two example articles and their summaries from *politics* and *sports* categories, respectively.

abstractive text summarization systems, generating summarises that are accurate, concise, and fluent remains a challenge for several reasons.

First, based on our empirical observation, summary styles in different

Category: Show
Article (truncated): "they are one of the world's most famous couples-and have quickly gained respect among the fashion elite. and now, one esteemed designer has revealed why kim kardashian and kanye west have the midas touch. olivier rousteing has revealed that he chose kim and kanye to star in balmain's latest campaign because they 'represent a family for the new world'. scroll down for video. fashion's most well-connected designer, olivier rousteing, has revealed why he snapped kim kardashian and kanye west up to front his balmain campaign. (.....) the 29-year-old creative director says he snapped up the duo, seen wearing his designs, because they are among the most talked-about people and embody the idea of a modern family. olivier-who regularly dresses kim, 34, and her siblings for the red carpet-explained that when kendall jenner and kim wear his clothes, they look like a 'fashion army'. the whole family seem enamoured with rousteing's designs and kim and kanye often sport matching outfits by the french fashion house. kim and kanye this week made trips to france and armenia with their daughter, north west. the trip to the religious mecca reportedly included north being baptised in the country where her late father's side of the family originated from. kim kardashian , kanye west and north visit the geghard monastery in armenia and take in the sights. kim, kanye and north have become a fashionable family. pictured here with alia wang, aimie wang and nicki minaj at the alexander wang show in february 2014."
Reference summary: "olivier rousteing has revealed why he chose kim and kanye for balmain. designer says the couple are among the most talked-about people. fashionable couple love wearing matching designs by balmain designer."
Summary by our model: "olivier rousteing has revealed why he chose kim and kanye to star in balmain's latest campaign because they represent a family for the new world. french designer says the couple are among the most talked-about people."
Summary by our model without text categorization: "olivier rousteing has seen kim kardashian and kanye west. kim kardashian and kanye west have worn his clothes."
Summary by our model without syntax annotation: "olivier rousteing has revealed why (he chose) kim and kanye west (up to front) his balmain campaign. french designer (says) the couple (~~couple~~) are among the most talked-about (people)."
Summary by our model without GAN: "olivier rousteing has revealed why he chose kim and kanye to star in balmain's latest campaign because they represent a family for the new world. the 29-year-old creative director has revealed that he was inspired to feature the couple-who have a 22-month-old daughter north - in the label's spring/summer 2015 men's campaign."

Table 2: An example of article from *Show* category and its summaries by different models. The words in red indicate the incomplete or redundant phrases.

categories can vary significantly. Two common categories (i.e., *Sports* and *Politics*) in CNN/Daily Mail Corpus (Hermann et al., 2015) are taken as an example, demonstrated in Table 1. To summarize a politic event, people

have a tendency to emphasize the subject of the event, and the result or influence of the event. In contrast, a sport summary is expected to include the teams and scores of the sport event. Obviously, the generated summaries should pay particular attention to different topics which belong to the corresponding categories. However, previous approaches employ a uniform model to produce summaries for the source documents from different categories, which are prone to generate generic and trivial summaries that easily miss or under-represent important aspects of the original documents. Furthermore, in an ablation study of our model (see Table 2), a model which does not recognize the text categorization could generate a descriptive summary missing a salient entity in the text.

Second, syntactic information plays a crucial role in sentence generation (Nadejde et al., 2017). Enforcing syntactic conformance addresses issues like incomplete sentences and duplicated words. As shown in Table 2, the model which is unaware of the text syntax could generate a broken summary for the given document. Yet, an improved system whose component has been co-trained with syntax annotation task could generate a more satisfied sentence for accurately and correctly condensing the raw document. Despite its usefulness, syntax information is underutilized in abstractive text summarization.

Third, in existing studies, the seq2seq based methods are usually trained to generate summaries of input documents via the maximum likelihood estimation (MLE) algorithm. Nevertheless, the MLE based models have two major disadvantages. (i) The evaluation metrics used in testing are not used when training the model, which will magnify the differences between losses from training and testing. For example, the seq2seq models are typically trained by employing the cross-entropy strategy, while they are evaluated by employing non-differentiable and discrete evaluation metrics such as ROUGE (Lin, 2004) at test time. (ii) While training, the decoder often receives the word vector of the previous ground-truth word at each time step. However, at testing phase, the decoder takes as input the previous word emitted by the seq2seq model, which may result in the exposure bias issue (Ranzato et al., 2016). which may accumulate errors quickly at each time step. Specifically, when the decoder generate a "bad" word, the seq2seq could propagate and accumulate errors alongwith the increase of the generated sequence length. The first several words of the generated summaries can be relatively correct, while the quality of sentences deteriorates quickly.

In this paper, we propose a Generative Adversarial Network for Abstrac-

tive Text Summarization with Multi-Task constraint (GAN-ATSMT) to alleviate the aforementioned limitations. Specifically, GAN-ATSMT jointly trains a generative model G and a discriminative model D via adversarial learning. The generative model G uses the sequence-to-sequence architecture as its backbone, taking the source document as input and generating the summary. We employ reinforcement learning (i.e., policy gradient) to optimize G for a highly rewarded summary. Hence, our model effectively conquers the exposure bias and non-differentiable task metrics issues. The discriminative model D is a text classifier, which is trained to classify the input summaries as human or machine generated. The generative model G and the discriminative model D are optimized via an adversarial process. The discriminative model D attempts to distinguish the the generated summaries by the generative model G from the ground truth summaries, while the generative model G is to maximize the probability of D making a mistake. Consequently, this adversarial process can make the generative model G generate high-quality and plausible abstractive summaries.

Furthermore, we additionally propose extended regularizations for the generative model G using multi-task learning. First, our LSTM encoder is regularized with the co-training required to perform an additional task of text categorization. Second, our LSTM decoder is also regularized with co-training to provide syntax annotation (Nadejde et al., 2017). This multi-task learning strategy is not prone to maximize the performance of the two auxiliary tasks, but rather to compensate for the missing regularization requirement of the abstractive text summarization task implemented with the seq2seq framework.

Compared with the prior abstractive text summarization methods, the main contributions of this work are as follows:

- We propose *GAN-ATSMT*, an adversarial framework for abstractive text summarization with multi-task constraint. The adversarial training process of GAN-ATSMT can eventually adjust the generator to generate plausible and high-quality abstractive summaries.

- GAN-ATSMT jointly trains the task of abstractive text summarization and two other related tasks: text classification and syntax generation. The auxiliary tasks help to enhance the CNN encoder and the RNN decoder in generating a more satisfied summary for accurately and correctly summarizing the given text.

- We incorporate the retrieved guidance summaries into the encoder-decoder structure, enriching the informativeness and diversity of the generated summaries.

- We conduct comprehensive experiments to evaluate the performance of GAN-ATSMT model. Experimental results show that GAN-ATSMT achieves significantly better results than the compared methods on the widely used CNN/Daily Mail and Gigaword datasets.

The rest of this paper is organized as follows. In Section 2, we discuss the related work on abstractive text summarization, generative adversarial networks, and multi-task learning. Section 3 defines the problem. Section 4 briefly introduces the architecture of the GAN-ATSMT model. In Section 5, we elaborate the sequence to sequence model with multi-task learning. The generative adversarial network for abstractive text summarization is presented in Section 6. In Section 7, we set up the experiments. The experimental results and analysis are provided in Section 8. Section 9 concludes this manuscript and presents some possible future work.

2. Related work

2.1. Abstractive Text Summarization

Generally, previous text summarization techniques are categorized as abstractive and extractive. The extractive summarization methods extract salient sentences or phrases from the original articles (Zhang et al., 2012), while the abstractive summarization methods produces new words or phrases, which may rephrase or use words which are not in the raw article (Rush et al., 2015). In this manuscript, we mainly work on the abstractive summarization.

In recent years, many efforts have been devoted to developing abstractive text summarization by employing the sequence-to-sequence model (Rush et al., 2015; See et al., 2017). For instance, Rush et al. (2015) was the first work which employed an encoder-decoder framework with attention mechanism for abstractive summarization. Nallapati et al. (2016) proposed attention encoder-decoder LSTM model to capture the hierarchical document structure and captured the primary words and sentences from the document. See et al. (2017) introduced a pointer-generator network which allowed both copying the words from the raw document via pointing, and producing words from the vocabulary.