COMP 550 - Assignment 4

Yi Tian Xu (260520039)

November 30, 2017

Question 1 In this assignment, we experiment on the extractive summerization method, SumBasic, and compare it with the the baseline Leading (where we extract the leading sentences of an article) and Simplified SumBasic (same as SumBasic but without the non-redudancy update). Our corpus is a collection of news articles manually extracted from Google News on four different topics clusters (see Table **??**).

	Topic Name	Number of Articles
Topic 1	US gun control and mental health	6
Topic 2	first robot citizen	5
Topic 3	scientists future of humanity	5
Topic 4	rise of homeschooling	3

Table 1: Description of the 4 clusters of articles in our corpus.

We manually remove all lines in articles that correspond to titles and convert all non-ASCII characters to similar-looking ASCII characters. When runing a summerization method, we apply sentence segmentation and lemmatization on our corpus text using the implementations in the NLTK package for Python. We also lowercase and removed punctuations and words appearing in the stopwords list provided by NLTK. Finally, we limit the output summary size to 100 words, and compared both SumBasic and Simplified SumBasic with Leading using ROUGE 1 score. Table 2 shows a sample output.

Topic	Method	Output	ROUGE 1 (%)
1	SumBasic	"Mental health is your problem here. "This isn't a guns situation," Trump said. It has been only a month and a half since the Las Vegas massacre, which was followed by the Texas church shooting and now this shooting spree in California. Too soon. What do the experts say? Five people were killed and at least two children were wounded, one of them at the school. It's a very, very sad event." Tragedy. Instead, I'm going to talk about mental health. Period. And those thoughts and prayers aren't solving anything. There's no motive yet, just reports of a neighborhood dispute.	25.96
	Simplified SumBasic	"Mental health is your problem here. So we have a mental health problem in this country. We can only talk about mental health. After the Texas church shooting, President Donald Trump said: "I think that mental health is your problem here. Mental health reform. But this is a mental health problem at the highest level. "This is a mental health problem at the highest level. "We have a lot of mental health problems in our country, as do other countries." We have a lot of mental health problems in our country, as do other countries. I thought mental health was the big problem.	20.94
	Leading	President Donald Trump said Monday that he believes the Texas church shooting was caused by a "mental health problem," not an issue with gun laws in the United States. "Mental health is your problem here," Trump said, noting that "based on preliminary reports" the shooter was "a very deranged individual." "This isn't a guns situation," Trump said. "This is a mental health problem at the highest level. It's a very, very sad event." "A very, very sad event, but that's the way I view it," Trump said. Trump's answer came in response to a question during a joint news conference Monday in Tokyo about whether he believed gun control measures were the answer to the Texas shooting.	19.72

2	SumBasic	This is historical to be the first robot in the world to be recognized with a citizenship," Sophia said. "What is this about? I will destroy humans." The Sophia-bot was dreamed up by the brains at Hanson Robotics, lead by AI developer David Hanson. "Like design smarter homes, build better cities of the future." How does it affect people if they think you can have a citizen that you can buy." "I am very honored and proud of this unique distinction. Sophia, an intelligent humanoid robot, has been granted citizenship in Saudi Arabia. "You've been reading to much Elon Musk and watching too many Hollywood movies.	30.77
	Simplified SumBasic	This is historical to be the first robot in the world to be recognized with a citizenship," Sophia said. Sophia, an intelligent humanoid robot, has been granted citizenship in Saudi Arabia. "Sophia the robot becomes first humanoid Saudi citizen." What rights does Sophia hold? Saudi Arabia became the first country in the world to grant citizenship to a robot. "This is historical to be the first robot in the world to be recognized with a citizenship." "What is this about? "I want to use my AI to help humans lead a better life," Sophia said. I will destroy humans." Sophia the robot has been on a roll lately.	29.77
	Leading	On October 25, Sophia, a delicate looking woman with doe-brown eyes and long fluttery eyelashes made international headlines. She'd just become a full citizen of Saudi Arabia - the first robot in the world to achieve such a status. "I am very honored and proud of this unique distinction. This is historical to be the first robot in the world to be recognized with a citizenship," Sophia said, announcing her new status during the Future Investment Initiative Conference in Riyadh, Saudi Arabia. Standing behind a podium as she spoke, to all effects, she presented a humanoid form - excepting the shimmery metal cap of her head, where hair would be on a human head.	28.11
3	SumBasic	"Climate change is here. In 1992, 1,700 independent scientists signed the "World Scientists' Warning to Humanity." They found most environmental problems have gotten far worse during the past 25 years. The rate of deforestation in some regions has also slowed. The other is extinction. The growing use of renewable energy is another positive trend, said Ripple. The new letter lists data showing a 75% increase in the number of ocean dead zones since the publication of the first letter. "Soon it will be too late to shift course away from our failing trajectory, and time is running out." It is so important to work together as a human race to make a sustainable future on planet Earth."	26.47
	Simplified SumBasic	"Climate change is here. In 1992, 1,700 independent scientists signed the "World Scientists' Warning to Humanity." In 1992, more than 1,700 scientists signed a World Scientists' Warning to Humanity published by the Union of Concerned Scientists. But global trends have worsened since 1992, the authors wrote in the new letter. That trend is not expected to change any time soon. Since it was published in the journal BioScience on Monday, hundreds more scientists have signed on to the letter. The "Second Notice" article updates the original "World Scientists' Warning to Humanity" document released in 1992, 25 years ago this month.	33.01

	Leading	More than 15,000 scientists from 184 countries warn the evidence is clear: Current and future human health and wellbeing are at serious risk from climate change, deforestation, loss of access to freshwater, species extinctions, and human population growth. Eminent scientists Jane Goodall, E.O. Wilson, and James Hansen are among those who have cosigned the warning, published Monday in the journal BioScience. The article, titled "World Scientists' Warning to Humanity: A Second Notice," has 15,372 signatories in total, from a range of scientific disciplines. It is thought to be the largest-ever formal support by scientists for a journal article. The "Second Notice" article updates the original "World Scientists' Warning to Humanity" document released in 1992, 25 years ago this month.	25.33
4	SumBasic	he said. The number of home schoolers effects the schools system, according to Townsley. I think now it's (different). We have people home-schooling for academics. Home schooling in the Greeneville and Greene County area may be increasing. "That's a decision they have to make for what they feel is best for their own child. An explanation of why the figures differed could not be obtained from the Tennessee Department of Education as of press time. This year, there were 3,464 students registered compared with 2,211 in 2013. "Math, reading, writing is all around you and you'll learn from your surroundings... when you're going to cook, you'll learn how to do measurements and maths."	23.30
	Simplified SumBasic	he said. The number of home schoolers effects the schools system, according to Townsley. The Greene County Schools System has seen a decrease in independent home school enrollments. But the face of home schooling in Australia is changing as the number of children being home schooled increases. According to state figures, in 2014-15 there were 107 students registered in independent home schools. Home school families and school systems sometimes work together to meet students needs. "It guarantees (home school students) the right to try out for a team. "In NSW, there are about 4,250 students who are registered for home schooling.	28.98
	Leading	An increasing number of students are opting for home schooling over traditional classes, but just what is causing this trend is up for debate. For example, while one student might be home-schooled due to a disability the parent feels is better managed at home, another might choose this option to avoid incessant bullying. Whatever the reasons, reports show this phenomenon to be on the rise. Nationally, homeschooling numbers have almost doubled in the past couple of years and are still growing. A survey by Home School WA revealed that the number of Western Australian students registered for home schooling has surged more than 50% in the past five years.	18.19

Table 2: The generated summaries from the three methods and their corresponding ROUGE 1 score. We show the average ROUGE 1 score for Leading.

We observe that SumBasic may prefers irrelevant short sentences, over long sentences with some frequent words. For example, we see in the generated summary for Topic 1, the sentence "Period." is chosen before the sentence

that introduce the shooter despite that the word "shooter" is mentioned in 5 of the 6 articles, and the name "Devin Patrick Kelley", in 4 of the 6 articles, while "period" appears only 3 times in all of the 6 articles. This is because longer sentences have a higher chance to contain infrequent words or redundant words which can underweight their average word probability score.

On the other hand, without updating the word probabilities in the Simplified SumBasic, we observe high repetition in words and content when articles in a specific topic cluster have more overlapping content, such as in Topic 1 and 2. In this case, SumBasic appears to have better summary quality than Simplified SumBasic. This seems to be also reflected in the ROUGE 1 score computed using the leading sentences from all the articles as reference summary. We also observe that the average ROUGE 1 score for Leading is the lowest among all topics.

There are often no relation between the sentences in the generated summaries. For example, sentences that appears contradictory can be put one next to the other. For example, in Topic 3, SumBasic generated "They found most environmental problems have gotten far worse during the past 25 years. The rate of deforestation in some regions has also slowed. The other is extinction. The growing use of renewable energy is another positive trend, said Ripple."

One way to improve the coherance of the generated summary is to include the order of the sentences from the original corpus when selecting the best scoring sentence in the SumBasic algorithm. One can also change the sentence segmentation algorithm to prevent breaking group of sentences that need to stay together (e.g.: sentences that are extracted from a quoted speach without information about the speaker). Another improvment is to avoid selecting sentences with pronouns, or to replace the pronouns by the reference, and removing adverbs in front of the sentences such as "instead", "so", "but", etc.

Question 2

Summary The authors experimented on automatic sentence compression by extending their previous model, which is a supervised extractive models consisting of a context free grammar (i.e.: synchronous tree substitution grammar (STSG)) defined on the space of tree pairs over uncompressed and compressed sentences. Each rule specified some edit on a uncompressed sentence fragment to give a compressed sentence fragment. Dynamic programming (DP) was used find the best sequence of rules by optimizing a scoring function based on hand-picked features.

Their new model attempted to generate abstract instead of extract using bilingual pivoting to learn more paraphrasing rules. New STSG rules can be created by combining tree fragments of English text that are frequently translated with the same foreign strings. The resulting grammar is much larger and noisier, thus to enforce coherence, the trigram log-probability feature was used in the scoring function, which renders the complexity exponential. Since DP is prohibitive, beam search was used coupled with cube-pruning heuristics.

SVM^{struct} was used for training the hyperparameters in the scoring function using 4 loss functions: 3 based on Hamming distance (namely, over tokens, ngrams and CFG production), and 1 based on the edit distance that measures the number of insertions and deletion.

In their experiment, they built a corpus of 575 sentences using newspaper articles and human annotators who were asked to paraphrase those sentences. A grammar was extracted according to the procedures in their previous model from the training partition of this corpus. The new grammar was extracted from the French-English Europarl v2 corpus, containing around 688K sentences. They use Birkley aligner to align the corpus, Biket's parser to obtain English parsed trees and GKHM algorithm to pair up English tree fragments with French strings. They compared between their previous extractive and current abstractive models, and also between the 4 different loss functions, and evaluated their performance using human ratings. Statistical tests were used to determine that the Hamming loss with token had the best mean rating, and that their abstractive model has higher ratings than their previous model.

Relation to Class Material and Limitations

This work implements an approach for solving extractive summarization, which we talked about in class. It also discusses on the difference between extractive and abstractive methods, which is that extractive methods is limited to word deletion operations, while abstractive methods should be free to use any operation (substitution, insertion, reordering, etc.).

the extractive and abstractive methods are similar to using CFG for the parsing problem as they also uses a CFG. Furthermore, the extractive method also uses DP for getting the best or most probably sequence of rules, which is a parallel to CYK algorithm.

One limitation of using those methods is that we can only apply them sentence by sentence. We don't have a context-free grammar for modelling paragraph structures thus we cannot only use STSG to do edit operations that takes into consideration the connections between sentences. But perhaps, it can be combined with extractive

summarization methods (e.g.: SumBasic) or some other natural language generation method that can combine or select sentences in a text while ensuring coherence.

Other limitations are that we need human annotated data and rating for evaluation which are expensive to obtain. The scoring function in the algorithm also contains hand-picked features and heuristics. But these are issues in various natural language generation tasks as we discussed in class.

Questions

1. Why can the grammar in the extractive model from Cohn and Lapata only model deletion and not other operations? Is it because that they extract the minimal set of synchronous rules?

2. Is there a limit to what kind of paraphrase rules that we can obtain using bilingual pivoting and can we avoid or filter out the wrong rules that are extracted using bilingual pivoting? (E.g.: the pronoun "her" substituted by "him" because in French, "lui" can be mapped to both.) Are translated texts guaranteed to preserve the original meaning to some degree that is high enough for the compression task?

3. How close is the sequence of rules found using beam search coupled with cube-pruning compared to the best sequence found using an exhaustive search?