Automating Horizon Scanning in Future Studies

Tatsuya Ishigaki, Suzuko Nishino, Sohei Washino, Hiroki Igarashi,
Yukari NagaiYuichi Washida, Akihiko Murai
(AIST, Japan, JAIST, Hitotsubashi University)

LREC2022

1. Introduction

We propose document retrieval and language generation tasks for the field of future studies.

Future study is the field that aims to develop systematic ways to predict drastic changes in the future.

A news article about robots at a station.
Title: *Disinfection robots at Takanawa Station.*
Article: *A railroad company will start a demonstration test of automatic mobile disinfection robots at Takanawa Station…*

We can think of various future scenarios from this article.
1. *Human workers at stations will be replaced by robots.*
2. *Many different types of robots can be deployed*

…and many other possible scenarios can be thought of because this article contains "Future sprout terms [Washida+2020]" such as "*robots*", "demo test"…

How researchers in future studies create scenarios?
= They use Systematic Foresight Methodology (SFM)
Step1: Horizon Scanning
- Collect news articles useful for predicting the future.
- Convert each retrieved article to a predefined formatted document called **"Scanning Material" (explained next)**

Step2: Scenario Planning
- They filter and/or merge a large set of Scanning Material.

We aim to automate Horizon Scanning step using NLP techniques because this step has been conducted by costly human experts.

2. Horizon Scanning

1. Document Retrieval for Foresight (Task1)

Article 1 | Article 2 | Article 3 | Article 4 | · · · | Article N

retrieved | not to be retrieved

2. Output: Predefined Formatted Document
Title:
Disinfection Robots at Takanawa Station.
Summary of Retieved Document:
The railroad company will start a demonstration test of automatic mobile disinfection robots at Takanawa Station…(truncated.)
Subjective Comment (Task2: Comment Generation):
Railroad companies need to automate a variety of tasks due to a shortage of human resources. Automatic ticket gates and automatic teller machines exist, but disinfection is also likely to be automated. In the future, robots may be used for other tasks as well.
Keywords: *Station, Robot*

News article 4
About the weather in Tokyo this weekend, the highest temperatures will be above normal. Next Monday and Tuesday will be temporarily colder as cold air moves in.

we call "scanning material".

Task1: Document Retrieval
- Input: large set of articles, Output: Ranked list
Task2: Comment Generation
- Input: article, Output: subjective comment.

3. Dataset

We used 2,226 news articles that contain "future sprouts".
The articles are collected by MBA students and professors at Hitotsubashi university between 2003-2020.

Procedure:
1. The MBA students and professors manually retrieved news articles in the series of lectures about "Future scenario creation".
2. Each article is manually converted to the format of scanning material. They added a subjective comment to each retrieved article.

4. Results: Document Retrieval

Model
1. Neural Supervised: BERT-based simple binary classifier
2. Neural Unsupervised: Document embedding-based distance
   If a document embedding is far from the centroid of general articles, it might contain some rare information, which likely to be a document to be retrieved.
3. Heuristic: IDF-based
   If the averaged IDF over an input document is higher, it might contain something rare[*Kwon+2021*]

Positive instance to be retrieved: 320 articles collected in 2020 by
Negative instance not to be retrieved: Randomly extracted 8,000 articles

	Prec.	Rec.	F-measure	P@5	P@10	P@30	P@50	P@100
Neural Network-based								
BERT+finetune (Supervised)	**58.6**	**57.8**	**58.1**	**80.0**	**80.0**	**76.7**	**76.0**	**70.0**
Distance-based (Unsupervised)	5.71	65.6	10.5	40.0	40.0	20.0	6.7	4.0	3.0
Heuristics-based								
IDF-based	8.60	33.4	13.7	40.0	20.0	16.7	16.0	15.0

Table 3: Performance of document retrieval models.

5. Results: Comment Generation

Model: BART + finetune,
Data: 2,226 article-comment pairs.

Automatic Evaluation ROUGE

Better ROUGE scores than Lead-3 in terms of ROUGE-1 and ROUGE-L.

	ROUGE-1	ROUGE-2	ROUGE-L
BART	**34.85**	7.52	**21.58**
Lead-3	30.05	**8.15**	19.07

Table 4: Performance of the Comment Generation Model.

Human Evaluation

Evaluators ranked three comments; Gold, BART, and Lead-3.

We use three types of criteria
1.Fluency
2.Correctness
3.Implication of Future Changes
- If an evaluator can imagine some future changes, the comment is ranked higher.

	Gold	BART	Lead
Implication of Future Changes			
Gold	-	72	77
BART	29	-	51
Lead	29	41	-
Fluency			
Gold	-	69	37
BART	29	-	33
Lead	46	65	-
Correctness			
Gold	-	49	16
BART	36	-	18
Lead	92	90	-

Table 6: A comparison of Human, BART, and Lead-3 methods by manual evaluation. The numbers indicate the number of times the method presented in a row is better judged.

BART was Judged better than Lead-3, but, the gap between BART and Gold is large.

6. Conclusion

1. We proposed two tasks for automating horizontal scanning in future studies.
2. BERT-based model achieves P@100=70% for the retrieval task.
3. In comment generation, more future-oriented comments can be generated but there is still room for improvement.