

# Unsupervised Abstractive Summarization Method that Suitable for Documents with Flows

Hoon-suk Lee<sup>†</sup> · Soon-hong An<sup>††</sup> · Seung-hoon Kim<sup>†††</sup>

## ABSTRACT

Recently, a breakthrough has been made in the NLP area by Transformer techniques based on encoder-decoder. However, this only can be used in mainstream languages where millions of dataset are well-equipped, such as English and Chinese, and there is a limitation that it cannot be used in non-mainstream languages where dataset are not established. In addition, there is a deflection problem that focuses on the beginning of the document in mechanical summarization. Therefore, these methods are not suitable for documents with flows such as fairy tales and novels. In this paper, we propose a hybrid summarization method that does not require a dataset and improves the deflection problem using GAN with two adaptive discriminators. We evaluate our model on the CNN/Daily Mail dataset to verify an objective validity. Also, we proved that the model has valid performance in Korean, one of the non-mainstream languages.

Keywords : NLP, Summarization, GAN, BERT, Transformer

## 흐름이 있는 문서에 적합한 비지도학습 추상 요약 방법

이 훈 석<sup>†</sup> · 안 순 홍<sup>††</sup> · 김 승 훈<sup>†††</sup>

## 요 약

최근 Encoder-Decoder를 기반한 요약은 거의 인간 수준에 도달하였다. 하지만 이는 영어, 중국어 등 수백만 건의 데이터셋이 잘 갖추어진 주류 언어권에서만 활용 가능하며 데이터셋이 구축되지 않은 비주류 언어권에서는 활용하지 못하는 한계가 있다. 또한, 문서의 일부 영역에 초점 하여 요약하는 편향의 문제를 갖고 있어 동화나 소설과 같이 흐름이 있는 문서에는 적합하지 않다. 본 논문에서는 두 개의 Discriminator가 있는 GAN을 통해 비지도 학습 기반의 추상 요약을 하며, 가이드 토큰의 추출과 주입을 통해 편향 문제를 개선하는 추출 요약과 추상 요약을 혼합한 하이브리드 요약 방법을 제안한다. CNN/Daily Mail 데이터셋을 통해 모델을 평가하여 객관적인 타당성을 검증하고 비주류 언어 중 하나인 한국어에서도 유효한 성능을 보인다는 것을 입증한다.

키워드 : NLP, Summarization, GAN, BERT, Transformer

## 1. 서 론

일반적으로 기계적 요약에는 추출(Extractive) 요약과 추상(Abstractive) 요약 두 가지 접근법이 있다. 추출 요약은 원문에서 중요한 문장을 찾아 문장 단위로 조합하여 요약하는 방법이며, 추상 요약은 의미상으로 원문과 일치하는 함축적인 새로운 문장들을 생성하는 방법이다.

추출 요약 방법은 구조적으로 전체 내용을 반영하지 않는다. 따라서 추출 방법은 뉴스나 사설 등과 같이 일부 문장에

이슈를 명확하게 나타내는 특정 단어와 내용이 포함된 문서에서는 효과적인 요약이 가능하다. 그러나 소설, 영화 시나리오, 동화와 같이 이야기의 흐름이 내용 전반에 있는 문서에는 적합하지 않다.

지도학습 기반의 추상 요약은 성능은 매우 뛰어나지만, 사람에게 의해 잘 작성된 요약 데이터셋이 필요하다. 추상 요약 학습에서 필요한 데이터셋은 다른 지도학습에서 사용하는 레이블의 구축보다 더 많은 노력이 필요하다. 일반적으로 이러한 연구에 사용되는 데이터셋은 영어의 경우 CNN/Daily Mail, Gigaword, newsroom 또는 multi-news이다. 다른 언어의 경우 중국어용 Gigaword 데이터셋과 일본어용 JAMUL[1]가 있다. 그러나 비주류 언어의 경우 데이터셋 구축의 어려움으로 인해 요약 데이터셋이 없거나 현재 준비 중인 상태이다. 이러한 문제점을 극복하고자 요약

<sup>†</sup> 정 회 원 : 아시아나IDT ICT융합연구소 수석연구원

<sup>††</sup> 준 회 원 : 아시아나IDT ICT융합연구소 선임연구원

<sup>†††</sup> 비 회 원 : 단국대학교 컴퓨터공학과 교수

Manuscript Received : October 5, 2021

First Revision : October 26, 2021

Accepted : October 27, 2021

\* Corresponding Author : Hoon-suk Lee(dolmani38@naver.com)

데이터셋이 필요 없는 비지도 학습 기반의 추상 요약 방법에 관한 연구가 필요하다. 또한, 추상 요약의 기존 모델들이 대부분 뉴스 기사를 데이터셋으로 사용함으로써, 뉴스 기사 특성이 모델에 반영되어 문서 시작 부분에 요약이 집중되는 편향 문제가 발생한다.

본 논문에서는 이러한 문제를 해결하고자 두 개의 Discriminator가 있는 GAN을 통해 비지도 학습 기반의 추상 요약을 하며, 가이드 토큰의 추출과 주입을 통해 편향 문제를 개선하는 추출 요약과 추상 요약을 혼합한 하이브리드 방식을 제안한다. 본 논문에서 제안한 방법은 비주류 언어(요약 데이터셋이 구축되어 있지 않은)와 소설, 영화 시나리오, 동화 등 흐름이 있는 문서에 효과적으로 적용할 수 있다.

다음 장에서는 GAN 기반 요약과 관련된 기존 연구를 살펴보고 제안 모델의 전반적인 구성에 대해 설명한다. 실험에서는 먼저 영어 CNN/Daily Mail 데이터셋을 통해 제안한 방법의 편향 문제 개선과 요약 성능을 검증한다. 이후 비주류 언어 중 하나인 한국어의 요약 실험으로 제안 모델의 성능을 입증한다.

## 2. 관련 연구

이 장에서는 추출 요약과 추상 요약의 관련 연구를 먼저 살펴보고, 이후 GAN[2]을 이용한 요약 관련 연구를 중점적으로 살펴본다.

초기 요약에 대한 접근법은 학습된 Rule[3] 또는 휴리스틱 기법[4]에 따라 정보가 없는 단어를 삭제해 나가는 추출 방법부터 시작하였다. 이후 TextRank[5], LexRank[6], LSA[7], KL-Sum[8] 등 통계적으로 문장의 중요도를 채점하는 방식으로 발전되었고 최근에는 RNN을 기반한 encoder-decoder framework[9], Transformer[10,11]의 활용 등 지도, 비지도 학습을 망라한 다양한 방법들이 활발히 연구되고 있다. 최초의 추상 요약은 구문의 변환 기법[12]으로 접근하였다. 추출 요약과 비교해 구현의 어려움으로 연구가 많지 않았다가 최근의 Seq2Seq[13], Attention[14] 개념의 출현 이후 많은 연구[15-22]가 진행되고 있다. Seq2Seq 모델이 state-of-the-art의 결과를 보이고 있으나 대량의 요약 데이터셋이 필요하다는 제약이 있다. 한국어의 경우[23-25] 등에서 추출, 추상 요약 방법이 연구되고 있으나 표준 데이터셋 없이 연구자에 의해 개별적으로 데이터를 수집하고 있다.

GAN을 이용한 요약의 기존 연구를 상세히 살펴보면, GAN 기반으로 비조건적(Non-conditional) 텍스트를 생성하는 연구는 SeqGAN[26], RankGAN[27], LeakGAN[28] 등 다양하다. 하지만 요약은 원본 문서를 입력 조건으로 갖는 조건적(Conditional) 텍스트 생성을 의미한다. [29]는 임의의 요약을 생성하는 Generator와 생성된 요약을 원본 텍스트와 비교하는 Discriminator 구조를 제안하였다. 이 연구 결과 R1,

R2, RL은 각각 39.92, 17.65 및 36.71 점수를 받았다. 제안의 구성은 LSTM 기반 Generator로 많은 양의 요약 데이터셋을 지도학습하고 GAN을 통해 약간의 성능만 개선하였다. [30]는 Seq2Seq Generator와 같은 구조의 Reconstructor 그리고 요약 문장을 사람이 읽기 가능한지 구별하는 Discriminator를 제안하였다. 이 연구에서는 사전 학습하지 않고 Adversarial REINFORCE 학습만 했을 때 R1은 28.11, R2는 9.97, RL은 25.41의 점수를 받았다. 하지만, 이 연구에서도 Generator는 문장의 구조에 대해 사전학습을 한다. [31]는 GAN 기반으로 Hindi와 Malayalam의 요약을 제안하였다. 비주류 언어인 Hindi와 Malayalam의 요약에 GAN을 사용한 이유는 역시 요약 데이터셋이 없었기 때문이다.

자연어의 이산적 특성 때문에 연속 공간을 기반으로 학습하는 GAN의 경우, 텍스트 생성에 이론적 어려움이 존재한다. 즉, 자연어의 Generator는 일반적으로 argmax 또는 sampling 함수와 같이 미분할 수 없는 부분이 존재하기 때문이다. 이러한 문제를 극복하기 위해 Earth mover's distance [32] 방법이 제안되었고, [33] 연구에서는 MC search를 통해 각 학습 단계에서 대략적인 보상을 평가하는 방안을 제안하였다. [30]은 WGAN과 'Self-Critic Adversarial REINFORCE' 방법을 제안하였다.

추출과 추상의 하이브리드 개념을 갖는 기존 연구로는 [22]에서 포인터를 통해 원본 텍스트에서 주요 단어를 복사할 수 있는 Hybrid pointer-generator network를 제안하였다. 이 연구 결과 R1, R2, RL에서 각각 39.53, 17.28, 36.38 점수를 얻었다. [31]은 KIGN (Key Information Guide Network)이라는 Guiding generation model을 제안하였다. 이 연구는 [22]와 비슷한 성능 수준을 보였다. 상기 두 연구는 본 논문에서 제안하는 하이브리드 관점에서 유사하지만, 해결하고자 하는 문제의 관점 즉, 비지도학습 및 편향 문제 개선에는 주안점이 없다.

## 3. WGAN을 이용한 비지도학습 추상 요약 방법

### 3.1 개요

본 논문에서 제안하는 구조는 생성자(Generator), 텍스트 조합자(Text sampler), 유사성 식별자(Similarity discriminator) 및 문법 식별자(Grammar discriminator)로 구성된다. 전체 아키텍처는 Fig. 1과 같다.

생성자는 요약의 골격이 되는 가이드 토큰을 추출하고 DNN (Deep Neural Network)을 통해 원본 문서의 토큰이 요약 문장에 채택될 확률을 생성한다. 기존 연구의 생성자는 LSTM 또는 Seq2Seq를 사용하므로 텍스트를 생성하기 위해 어떤 형태든 사전학습이 필요하였으나, 본 논문에서는 단순한 DNN을 사용하여 사전학습 없이 텍스트를 조합할 수 있도록 구성한 것을 특징으로 한다. 텍스트 조합자는 확률이 높은 토큰을

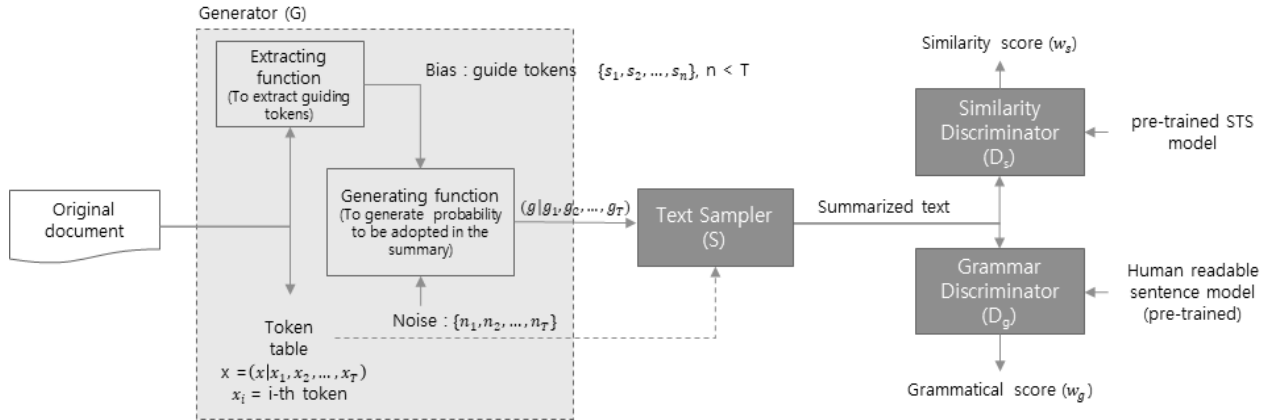


Fig. 1. Overall Architecture of Unsupervised Abstractive Summarization Method Using WGAN

선택하여 추상 요약 문장으로 변환한다. 유사성 식별자는 요약된 문장이 원본 문장과 얼마나 유사한지를 구분한다. 문법 식별자는 생성된 문장이 문법적으로 적절한지 구분한다.

마지막으로 WGAN에 의해 학습이 수행된다. 기존 요약 연구에서 제안된 WGAN과 차별되는 독자적인 목적 함수를 구성하며 효과적인 학습을 위해 유사성 식별자와 문법 식별자 사이의 경합을 조정하는 적응적 식별 인자(Adaptive discriminant factor)를 적용하는 것을 특징으로 한다. 다음 장에서는 각 구성 요소에 대해 자세히 설명한다.

### 3.2 생성자 - Generator (G)

생성자(G)는 가이드 토큰을 추출하는 추출 기능(Extracting function)과 원본 문서의 토큰이 요약 내용에 선택될 확률을 생성하는 생성 기능(Generating function) 두 가지 기능으로 구성된다.

#### 1) 추출 기능 - Extracting function

본 연구에서 목표하는 편향 문제 개선을 위해 요약의 골격에 해당하는 가이드 토큰을 원본 문서에서 추출한다. 문서를 공백(space) 단위로 나누는 각 토큰의  $i$ 번째 토큰이  $x_i$ 인 경우, 문서는 다음 Equation (1)과 같이 나타낼 수 있다.

$$x = \{x_1, x_2, \dots, x_T\} \quad (1)$$

각 토큰에 대한 전체 문서의 유사성 확률 분포(Similarity Probability Distribution, SPD)는 다음 Equation (2)와 같다.

$$SPD_{for\ one\ token} = \{P_s(x_1), P_s(x_2), \dots, P_s(x_T)\} \quad (2)$$

Equation (2)에서  $T$ 는 토큰의 개수이고  $P_s$ 는 유사성 확률이다. 이러한 vector는 연속 신호로 간주 될 수 있으며 신호의 꼭짓점(Peak)에 해당하는 토큰은 요약을 위한 가이드 토큰으로 간주할 수 있다. 그러나, 하나의 토큰만으로 문서의 유사성을 판단하면 문서의 내용을 제대로 반영하지 못한다. 이를 극복하기 위해 2개 이상의 연속 토큰을 사용하여 일종의

부분 스토리로서 필터(Filter)와 같은 역할을 하게 한다. 문서의 내용을 제대로 반영하는 SPD를 얻기 위해 전체 문서를 각 필터로 합성곱(Convolution) 한다. 만약,  $n$ 개의 연속 토큰을 필터로 사용하면 SPD를 다음 Equation (3)과 같이 나타낼 수 있다.

$$SPD_{for\ n\ tokens} = \left\{ \begin{array}{l} P_s(x_1, x_2, \dots, x_n), \\ P_s(x_2, x_3, \dots, x_{n+1}), \\ \dots \\ P_s(x_{T-n}, x_{T-n+1}, \dots, x_T) \end{array} \right\} \quad (3)$$

가이드 토큰 배열을  $s$ 라고 하면,  $s$ 는 1에서  $m$ 까지 필터를 사용하여 다음 Equation (4)와 같이 나타낼 수 있다.

$$s = peak\left(\sum_{i=1}^m SPD_i\right) \quad (4)$$

Equation (4)에서, peak은 [35]과 같은 peak detection algorithm을 의미한다.

#### 2) 생성 기능 - Generating function

토큰이 요약 문장에 선택될 확률을 생성하는 생성 기능은 일반 DNN으로 구성한다. 원본 문서의 총 토큰 수에 해당하는 임의의 노이즈 및 가이드 토큰을 입력한다. 여기서 가이드 토큰은 바이어스로 적용된다. 출력은 원본 문서의 각 토큰이 요약 문장의 토큰으로 선택될 확률값이다.

임의의 노이즈는 DNN의 여러 조밀한 계층을 통과하지만, 가이드 토큰은 뉴런의 출력 계층 직전에 추가된다. 그 결과 전체 출력은 가이드 토큰으로 바이어스된다. 또한 각 가이드 토큰 간에 필요한 토큰은 GAN 학습을 통해 원본 문서에서 선택된다. 이러한 생성 기능은 기존 연구에서 제안되지 않았던 독특한 구조이다.

### 3.3 텍스트 조합자 - Text Sampler (S)

텍스트 조합자(S)는 생성자(G)의 출력을 입력으로 사용하고 아래 조건에 따라 요약에 사용할 토큰을 선택한다.

$$t'_j = \begin{cases} i & \text{if } g_i > \alpha \\ None & \text{otherwise} \end{cases} \quad (5)$$

Equation (5)에서  $g_i$ 는 생성자(G)의 출력 vector의 각 원소이다.  $\alpha$ 는 목적하는 요약의 비율에 따라 상위로 순위 되는  $g_i$ 의 최솟값으로 설정된다. 예를 들어 10개의 토큰에 대하여, 50%의 요약을 목표한다면, 상위 5개 토큰의 확률값에 대한 최솟값에 해당한다. 이러한 방법으로 요약의 비율을 조정할 수 있다. 마지막으로 요약 문장은 토큰의 배열  $x$ 에서  $t'_j$  번째 토큰을 추출하여 순서에 따라 조합하므로 요약 텍스트를 생성한다.

### 3.4 유사성 식별자 - Similarity Discriminator ( $D_s$ )

유사성 식별자(Similarity discriminator,  $D_s$ )는 생성된 요약 내용이 원본 문서와 얼마나 유사한지 정량적으로 측정한다.  $D_s$ 는 Sentence-BERT[36]를 통해 얻은 두 문장의 context vector에 대한 cosine-similarity 값이다.

유사성 식별자는 sentence-transformer python package를 활용하여 구현한다. 입력은 최대 128개의 토큰이 가능하며 출력은 1,024차원의 vector(embedding)를 반환한다. 결국, 전체 문서를 한 번에 벡터화하는 것은 불가능하다. 이 문제의 해결을 위해, 문서를 문장으로 나누고 N개의 문장 각각을 벡터화하여 (N, 1024)의 행렬을 만든다. 원본 문서와 비교할 요약 문장을 n개의 문장으로 나누어 같은 방법으로 (n, 1024)의 행렬을 구한다. cosine distance는 (n, N) 행렬을 얻기 위해 pair-wise 방식으로 계산된다. 이후, 각 행 최솟값의 평균값을 전체 문서에 대한 유사도로 취한다.

### 3.5 문법 식별자 - Grammar discriminator ( $D_g$ )

문법 식별자(Grammar discriminator,  $D_g$ )는 문장의 문법성을 평가한다. Transformer의 BertModel과 Binary classifier header로 구성한다. 일반 정상 문장과 정상 문장의 토큰 순서를 임의의 변경하여 구성된 문법적 비정상 문장 데이터셋을 학습하였고, 정상 Class의 softmax 출력값을  $D_g$ 의 값으로 취한다.

### 3.6 WGAN 학습

본 연구에서 사용하는 목적 함수는 WGAN을 적용하여 단순화한다. 학습의 효율성을 위해 적응적 식별 인자를 특징적으로 제안한다.

#### 1) 목적 함수의 단순화

GAN의 목적 함수 정의에 의하면 제안 방법의 전체 목적 함수는 다음 Equation (6)과 같이 표현될 수 있다.

$$\begin{aligned} \min_G \max_{D_g, D_s} V(D_g, D_s, G) = & \quad (6) \\ E_{x \sim P_{data(x)}} [\log D_g(x)] + & \\ E_{x \sim P_{data(x)}} [\log D_s(X, x)] + & \\ E_{z \sim P_{data(z)}} [\log (1 - D_g(S(G(z))))] + & \\ E_{z \sim P_{data(z)}} [\log (1 - D_s(X, S(G(z))))] & \end{aligned}$$

Equation (6)에서 X는 원문 전체이다. 텍스트 조합자(S)는 미분을 할 수 없는 함수로서 Equation (6)을 그대로 최적화를 할 수 없다. 이러한 문제는 GAN을 통한 텍스트를 생성하는 일반적인 방법에서 발생한다. 이를 극복하기 위해 본 논문에서도 WGAN을 적용하지만 WGAN의 목적 함수는 DNN 생성자(G)에 적합하도록 기존의 연구[30]와는 다른 방식으로 구성한다.

각각의 식별자에 관한 결과를  $D_g(S(G(z))) = w_g$ ,  $D_s(X, S(G(z))) = w_s$ 이라 하면 생성자의 loss는 Wasserstein distance를 적용하여 다음 Equation (7)과 같이 정의한다.

$$L_G = - \frac{w_g + w_s}{n} \sum_i^n G'(z) \quad (7)$$

Equation (7)에서, n은 토큰의 개수이며,  $G'(z)$ 는  $G(z) = (g \mid g_0, g_1, \dots, g_n)$ 일 때 아래 조건의 원소  $g'_i$ 을 갖는다.

$$g'_i = \begin{cases} g_i & \text{if } g_i > \alpha \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Equation (8)의  $\alpha$  값은 Equation (5)의  $\alpha$ 값과 같다. 따라서 전체 목적 함수는 다음 Equation (9)와 같다.

$$\begin{aligned} \min_G \max_{D_g, D_s} V(D_g, D_s, G) = E_{x \sim P_{data(x)}} [\log D_g(x)] + & \quad (9) \\ E_{x \sim P_{data(x)}} [\log D_s(X, x)] - \frac{w_g + w_s}{n} \sum_i^n G'(z) & \end{aligned}$$

그러나  $E_{x \sim P_{data(x)}}$ 는 학습 과정에서 요약된 실데이터를 사용할 수 없으므로 정의할 수 없다. 따라서 본 논문에서는 사전 학습된 모델을 적용한 식별자를 사용하고 식별자는 학습에 참여하지 않는다. 결과적으로 최종 목적 함수는 생성자에 대한 부분만 남아 다음 Equation (10)과 같이 단순화된다.

$$\min_G V(G) = - \frac{w_g + w_s}{n} \sum_i^n G'(z) \quad (10)$$

#### 2) 적응적 식별 인자 - Adaptive discriminant factor

Equation (10)에서  $w_g$ 와  $w_s$ 는 학습 과정에서 서로 경쟁 관계이다. 즉, 학습에서 문법 식별자가 주도권을 가지면 문서

요약은 내용과의 유사성이 약해지고 문법만 증가하여 원문에서 특정 문장만을 추출하게 된다. 반대로, 유사성 식별자가 주도권을 갖게 되면 문서 요약은 사람이 읽을 수 없는 핵심 토큰의 나열이 된다. 본 논문에서는 이러한 경쟁 관계의 학습 균형을 맞추기 위해 적응적 식별 인자(Adaptive discriminant factor)로서  $\beta$ 을 제안한다.

$$\min_G V(G) = -\frac{w_g + \beta w_s}{n} \sum_i^n G'(z) \quad (11)$$

Equation (11)에서  $\beta$ 는 1을 초깃값으로 학습의 진행 단위, 즉 epoch마다 그 값을  $w_s > w_g$ 이면 감소,  $w_s < w_g$ 이면 증가하도록 조정하여  $w_g$ 와  $w_s$ 의 경쟁 관계를 증대한다.

## 4. 실험

### 4.1 개요 및 실험 환경

본 논문을 통해 해결하고자 하는 문제 2가지는 첫 번째, 편향 문제(Deflection problem)의 개선과 두 번째, 비주류 언어에 대한 비지도 학습 추상 요약이다. 비주류 언어 중 하나인 한국어의 경우, 요약 데이터셋이 구축되어 있지 않아

(AI-HUB [<https://aihub.or.kr/>])의 한국어 요약 데이터셋 제공 이전) 객관적인 성과를 측정할 수 없다. 따라서 본 논문에서 제안한 방법의 객관적 타당성을 검증하기 위해 먼저 영어 데이터셋인 CNN/Daily Mail을 통해 실험을 진행한다. 이후 4.5절에서는 한국어 샘플을 활용해 성능 유효성을 검증한다.

모든 실험은 Google colab's GPU 환경을 기반으로 한다. 유사성 식별자와 문법 식별자를 위한 사전학습 모델(pre-trained model)은 hugging-face의 모델을 사용한다. 문법 식별자의 미세조정(fine-tuning)을 위해 CNN/Daily Mail의 단일 문장 약 500K와 한국 소설 및 크롤링 뉴스의 단일 문장 약 410K를 사용한다.

### 4.2 WGAN 학습

학습은 500 Epoch로 수행하고, 가이드 토큰의 초기 바이어스는 0(편향 없음) 또는 1.0이다. learning rate는  $5e-5$ 이고 adam optimizer를 사용한다. 학습이 진행됨에 따라 적응적 식별 인자( $\beta$ )가 균형을 조정하여 문법과 유사성의 loss가 함께 감소한다. Fig. 2는 적응적 식별 인자( $\beta$ )가 없는 불균형 학습의 사례이고, Fig. 3은 적응적 식별 인자( $\beta$ )가 적용된 학습 사례이다.

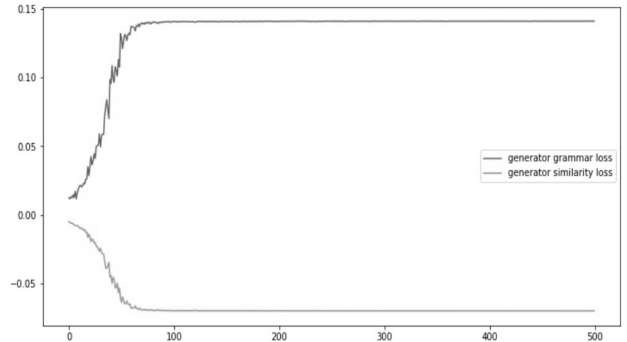
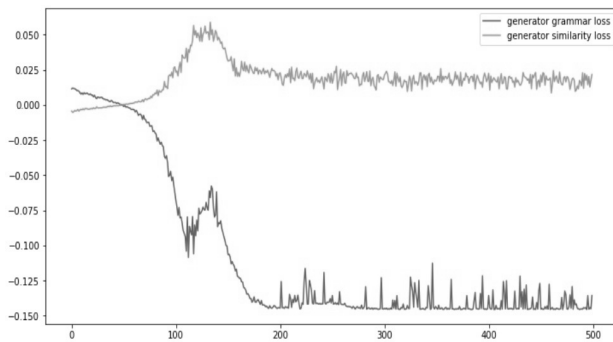


Fig. 2. (Left) An Example of a Case Where Grammar Loss is Dominant Without an Adaptive Discriminant Factor  $\beta$ , (Right) Vice Versa. X-axis: Epoch, Y-axis: Grammar Loss(Blue), Similarity Loss(Red)

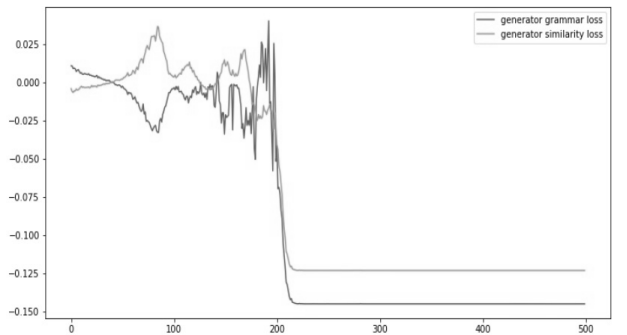
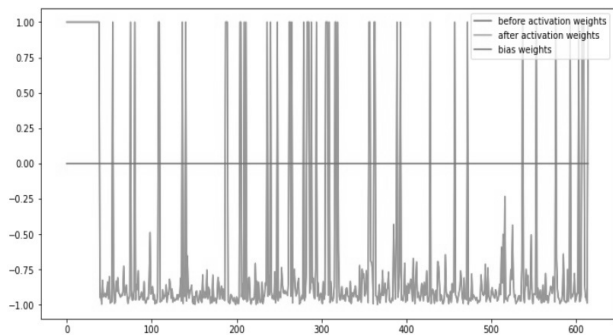


Fig. 3. (Right) An Example of a Case Where Grammar-Loss and Similarity-Loss are Balanced by Applying an Adaptive Discriminant Factor  $\beta$ , the Two Losses are Lowered Together. (Left) X-axis: Token Order, Y-axis: Probability to be Selected into Summary.

4.3 실험 방법과 측정 지표

이전 연구들과의 비교를 위해 각각 추출 요약 2가지 방법, 추상 요약 2가지 방법을 기준으로 평가한다.

1) 비교 대상 추출 요약 방법

a) BERT+LexRank : LexRank[6] 에서 제안한 방법은 문장의 그래프 표현을 기반으로 고유 벡터의 중심성 개념을 통해 문장 중요도를 계산한다. 여기서 각 문장의 Context vector는 sentence-transformer의 출력값을 사용한다. 영어의 사전학습 모델은 ‘stsb-bert-base’를 사용하고, 한국어는 ‘xlm-r-large-en-ko-nli-ststb’를 사용한다.

b) BESM (bert-extractive-summarizer method) : BERT 기반으로 문장을 Embedding하고 K-means를 적용하여 Centroid를 계산한다. 계산된 결과를 기준으로 중심에 가까운 문장을 순위별로 선택한다[37]. 사전학습 모델로 영어는 ‘bert-base-uncased’를 사용하고, 한국어는 ‘monologg/kobert’를 사용한다.

2) 비교 대상 추상 요약 방법

a) BART Transformer : [38]에서 제안한 BART 기반의 Transformer를 적용한다. 사전학습 모델로 ‘bart-large-cnn’을 사용한다.

b) T5 Transformer : [39]에서 제안한 T5 기반의 Transformer를 적용한다. 사전학습 모델로 ‘t5-base’를 사용한다.

본 논문은 평가 지표로 ROUGE[40]를 사용하고 추가로 유사성(Similarity), 문법성(Grammar), 편향 지수(DI, Deflection Index)의 3가지 지표를 적용한다. 유사성(Similarity)은 유사성 식별자(Ds)의 값( $w_s$ )이며, 문법성(Grammar)은 문법 식별자(Dg)의 값( $w_g$ )이다. DI의 산출 방법은 Equation (12)와 같다. 원본 문서를 서론(20%), 본문(50%), 결론(30%)의 세 부분으로 나누고 각 부분의 유사도( $w_s$ )에 대해 분산을 취한다. DI는 낮을수록 좋은 결과이다.

$$DI = Var(w_s: 서론, w_s: 본문, w_s: 결론) \quad (12)$$

4.4 CNN/Daily Mail dataset 실험 결과

CNN/Daily Mail 데이터셋에 관한 결과는 Table 1과 같다. 결과에서 Comp rate는 압축률을 의미한다. ‘Comp rate = 요약 문장의 토큰 수 / 원본 문장의 토큰 수’로 정의 하며, Comp rate가 낮을수록 짧은 문장으로 요약된 것이다. Intro, Body, Ending은 각각 서론, 본문, 결론 부분의 유사성( $w_s$ : 서론,  $w_s$ : 본문,  $w_s$ : 결론)을 나타낸다. DI는 편향성 지수(Deflection Index)를 나타낸다. Total similarity는 전체 원본 문서와 비교한 요약 문장의 유사도( $w_s$ )을 의미하며, Grammar는 문법성( $w_g$ )을 의미한다.

Table 1의 하단, Friedman test[41]는 각 비교 방법별 각 지표의 실험 결과가 통계적으로 유의미 하다는 것을 나타내고 있다. Table 1의 결과에서 추출 방법과 추상 방법 모두 서론부(Intro)에 유사성이 높은 편향 문제를 보였다. 본 논문에서 제안하는 방법 중 가이드 토큰이 적용되지 않은 경우 (A), 다른 비교 방법과 유사하게 편향 문제가 나타났다. 이는 Fig. 4 Nemenyi 사후 검증[42] (a)에서 비교 방법들과 통계적 차이가 없음을 알 수 있다. 가이드 토큰(B)을 반영한 경우, DI가 낮아지며 편향 문제가 개선되었다. 이는 Fig. 4 Nemenyi 사후 검증(a)에서 유의미함을 알 수 있다(‘BERT+LexRank’의 경우 p-value는 0.05 이상이지만 similarity가 전체적으로 제안 방법에 비하여 유의미하게 낮음. ((b),(c), (d))). 또한 유사성도 높아졌다. 반면, 문법성과ROUGE 점수는 모두 하락한다. 생성자가 사전 학습되거나 요약 데이터셋으로 지도학습되지 않았음에도 전체 성능은 [30]과 유사하다. ‘6. 부록’에 본 논문의 제안 방법(B) 포함, 각 비교 대상 요약 방법의 요약 결과 2개 샘플을 제시한다.

이상의 결과를 통해 본 논문에서 제안하는 모델은 편향 문제를 극복하는 동시에 객관적이고 실용적인 성능을 확보하고 있음을 입증한다. 그러나 본 제안 방법은 원본 문서에서 추출한 토큰 단위의 조합으로 구성되어 문법적인 완성도가 낮은 단점이 있다.

Table 1. F1 ROUGE Scores and Deflection Index(Similarity Variance), Similarity, Grammar on CNN/Diary Mail Dataset

	Method	Comp rate	Similarity				Total Similarity	Grammar	ROUGE		
			Intro	Body	Ending	DI			R1	R2	RL
ours	(A) No guide tokens	0.1504	0.5560	0.4297	0.4021	0.0129	0.4482	0.9674	31.84	8.40	19.11
	(B) With guide tokens	0.1592	0.5028	0.4567	0.4574	0.0074	0.4662	0.9087	28.17	5.01	15.22
EXT	BERT+LexRank	0.1779	0.3173	0.2951	0.2899	0.0029	0.2999	0.9999	20.85	4.69	12.50
	BESM	0.2170	0.6184	0.4661	0.4380	0.0142	0.4920	1.0000	31.77	11.19	19.66
ABS	BART transformer	0.1383	0.5840	0.4245	0.3598	0.0148	0.4419	1.0000	44.17	23.74	31.58
	T5 transformer	0.1084	0.5764	0.3948	0.3173	0.0174	0.4143	0.9999	40.99	20.96	30.30
Friedman test	statistic		77.036	72.090	68.945	48.599	108.381	92.886	100.272	87.493	85.836
	p-value		7.3e-16	8.2e-15	3.7e-14	7.0e-10	1.6e-22	3.2e-19	8.6e-21	4.4e-18	1.0e-17

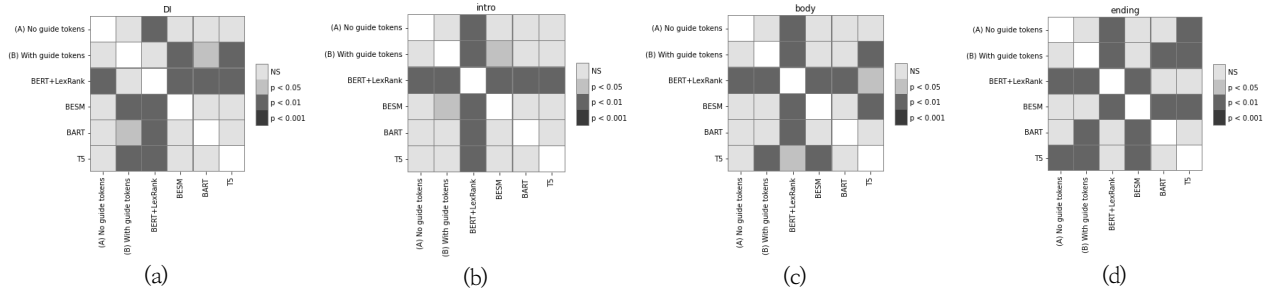


Fig. 4. Results of Nemenyi Test on CNN/Diary Mail Dataset

Table 2. Deflection Index(Similarity Variance), Similarity, Grammar on Korean Text Samples

	Method	Comp rate	Similarity				Total Similarity	Grammar
			Intro	Body	Ending	DI		
ours	(A) No guide tokens	0.1476	0.5238	0.4722	0.4646	0.0084	0.4802	0.9658
	(B) With guide tokens	0.1721	0.5319	0.5156	0.5259	0.0057	0.5219	0.9697
EXT	BERT+LexRank	0.2175	0.2268	0.2152	0.2055	0.0073	0.2146	0.9989
	BESM	0.2101	0.4663	0.3960	0.3890	0.0100	0.4079	0.9910
Friedman test	statistic		100.366	138.433	127.098	8.796	144.895	149.067
	p-value		1.2e-21	8.2e-30	2.2e-27	0.0321	3.3e-31	4.1e-32

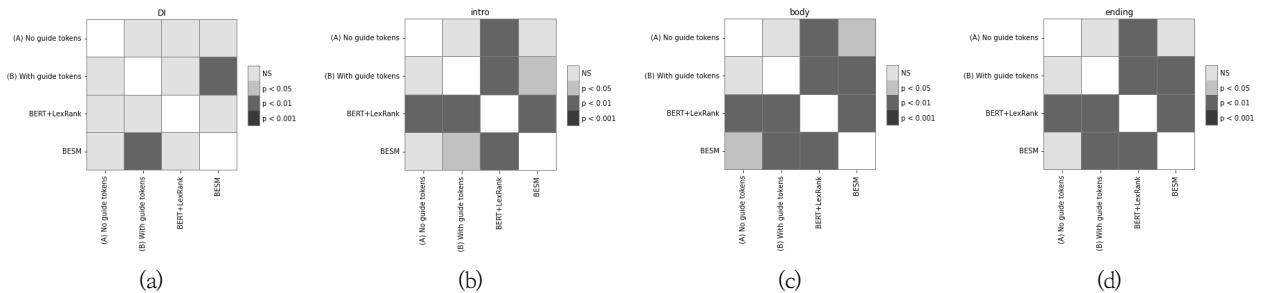


Fig. 5. Results of Nemenyi Test on Korean Text Samples

4.5 한국어 실험 결과

본 장에서는 제안 모델을 한국어에 적용하여 성능을 검증한다.

1) 한국어 신데렐라 동화

한국어 신데렐라 문서는 1,516자, 325개의 토큰으로 구성되어 있다. Table 3은 이 문서에 대한 요약 결과를 보여준다. 압축은 원문 대비 16.2%, 원문과의 유사성은 58.2%, 문법성은 99.6%이다. 본 연구의 의도에 따라 새로운 문장으로 구성되었고, 원문의 전체적인 내용이 포함된다. 그러나, 문법적으로 부자연스러운 부분이 있어 쉽게 읽기에 어려움이 있다. 간혹 문법적으로 일치하더라도 원문의 내용이 잘못 조합되는 경우가 있다. 이러한 단점은 향후 연구를 통해 개선할 과제이다.

2) 한국어 소설 샘플

성능 측정의 객관성을 높이기 위해 100개의 샘플 문서를 테스트한다. 샘플 문서는 한국어로 된 여러 소설을 나누어 만든다. 결과는 Table 2와 같다. Friedman test는 실험 결과가 통계적으로 유의미하다는 것을 나타내고 있다. 샘플 문서는 소설의 일부분으로 구성되었기 때문에 뉴스 문서와 다르게 두괄식으로 구성되지 않는다. 하지만 기존 추출 방법의 경우 역시 도입 부문에 초점이 맞춰지는 편향 문제가 나타났다(Fig. 5 Nemenyi 사후 검증(a), 'BERT+LexRank'의 경우 p-value는 0.05 이상이지만 similarity가 전체적으로 제안 방법에 비하여 유의미하게 낮음. ((b),(c),(d))). 이러한 모델은 동화나 소설에 적용하기에는 적합하지 않다. 본 논문에서 제안하는 방법은 학습을 위한 요약 데이터셋 없이 추상 요약이 가능하고, 편향 문제가 개선되며(B) 원문과의 유사성도 증가한다.

Table 3. An Example of a Summary Using the Proposed Method in Korean

<p>Source text (Cinderella in Korean) :</p> <p>옛날 어느 집에 귀여운 여자 아기가 태어났어요. 아기는 무럭무럭 자라서, 예쁘고 마음씨 고운 소녀가 되었어요. 그러던 어느날, 소녀의 어머니가 병이들어 그만 세상을 떠나고 말았어요. 소녀의 아버지는 홀로 남은 소녀가 걱정되었어요.</p> <p>... (omit middle) ...</p> <p>그때, 신데렐라가 조용히 다가가 말했어요. 저도 한번 신어 볼 수 있나요? 신데렐라는 신하게 건넌 유리 구두를 신었어요, 유리 구두는 신데렐라의 발에 꼭 맞았어요. 신하들은 신데렐라를 왕궁으로 데리고 갔어요. 그 뒤 신데렐라는 왕자님과 결혼하여 오래오래 행복하게 살았대요.</p>
<p>Summary :</p> <p>그러던 소녀의 새어머니를 맞이했어요. 소녀는 도맡아 했어요. 잠시 왕궁에서 초대장이 왔어요. 언니들을 떠났어요. 신데렐라는 할머니가 건드리자, 생쥐와 흰말로, 예쁜 드레스로 바뀌었어요. 내밀어 보거라. 할머니는 빛나는 구두를, 열두시가 처음대로 돌아간단다. 돼. 그러니까 왕자님도 무도회장에 쳐다보지도 않고, 신데렐라하고만 춤을 추었어요. 주인과 결혼하겠어요. 유리 구두는 한번 신데렐라의 왕궁으로 결혼하여 살았대요.</p>

### 5. 결 론

본 논문에서는 두 개의 Adaptive discriminator와 WGAN을 사용한 비지도 학습의 추상 요약 방법을 제안함으로 비주류 언어권에서도 추상 요약을 활용할 수 있는 가능성을 제시하였다. 또한 가이드 토큰 개념을 적용하여 기존의 추출, 추상 방법에서 나타나던 편향 문제를 개선하고 동화, 소설, 시나리오 등 흐름이 있는 문서 요약에 적합함을 확인했다. 요약 결과는 매끄럽지 않은 구성이 존재하지만 비지도 학습 기반 추상 요약 방법의 한계로서 이후 Seq2Seq 모델과 강화학습을 접목한다면 자연스러운 문장 생성이 가능할 것으로 기대한다.

### References

[1] Y. Hitomi, et al., "A large-scale multi-length headline corpus for analyzing length-constrained headline generation model evaluation," in *Proceedings of the 12th International Conference on Natural Language Generation*, pp.333-343, 2019.

[2] Ian J. Goodfellow, et al., "Generative adversarial nets," in *Proceedings of the International Conference on Neural Information Processing Systems*, pp.2672-2680, 2014.

[3] K. Knight and D. Marcu, "Summarization beyond sentence extraction: A probabilistic approach to sentence compression," *Artificial Intelligence*, Vol.139, No.1, pp.91-107, 2002.

[4] B. Dorr, D. Zajic, and R. Schwartz, "Hedge trimmer: A parseand-trim approach to headline generation," In *Proceedings of North American Chapter of the Association for Computational Linguistics*, pp.1-8, 2003.

[5] R. Mihalcea and P. Tarau, "TextRank: Bringing order into texts," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp.404-411, 2004.

[6] G. Erkan and D. R. Radev, "LexRank: Graph-based lexical centrality as salience in text summarization," *Journal of Artificial Intelligence Research*, Vol.22, No.1, pp.457-479, 2004.

[7] S. T. Dumais, "Latent semantic analysis," *Annual Review of Information Science and Technology*, Vol.38, No.1, pp.188-230, 2005.

[8] A. Haghighi and L. Vanderwende, "Exploring content models for multi-document summarization," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp.362-370, 2009.

[9] Q. Zhou, N. Yang, F. Wei, S. Huang, M. Zhou, and T. Zhao, "Neural document summarization by jointly learning to score and select sentences," In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Vol.1, pp.654-663, 2018.

[10] M. Zhong, D. Wang, P. Liu, X. Qiu, and X. Huang, "A closer look at data bias in neural extractive summarization models," in *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pp.80-89, 2019.

[11] D. Wang, P. Liu, M. Zhong, J. Fu, X. Qiu, and X. Huang, "Exploring domain shift in extractive text summarization," *arXiv preprint arXiv:1908.11664*, 2019.

[12] T. Cohn and M. Lapata, "Sentence compression beyond word deletion," in *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp.137-144, 2008.

[13] I. Sutskever, O. Vinyals, Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems*, pp.3104-3112, 2014.

[14] A. Vaswani, et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp.5998-6008, 2017.

[15] R. Nallapati, F. Zhai, and B. Zhou, "Summarunner: A recurrent neural network based sequence model for extractive summarization of documents," in *Thirty-First Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*, 2017.

[16] A. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," *arXiv preprint arXiv:1509.00685*, 2015.



- [17] S. Chopra, M. Auli, and A. Rush, "Abstractive sentence summarization with attentive recurrent neural networks," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.93-98, 2016.
- [18] R. Nallapati, B. Zhou, C. N. Santos, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," *arXiv preprint arXiv:1602.06023*, 2016.
- [19] R. Paulus, C. Xiong, and R. Socher, "A deep reinforced model for abstractive summarization," *arXiv preprint arXiv:1705.04304*, 2017.
- [20] A. Fan, D. Grangier, and M. Auli, "Controllable abstractive summarization," *arXiv preprint arXiv:1711.05217*, 2017.
- [21] C. Baziotis, I. Androustopoulos, I. Konstas, and A. Potamianos, "Seq<sup>3</sup>: Differentiable sequence-to-sequence-to-sequence autoencoder for unsupervised abstractive sentence compression," *arXiv preprint arXiv:1904.03651*, 2019.
- [22] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," *arXiv preprint arXiv:1704.04368*, 2017.
- [23] G. H. Lee, Y. H. Park, and K. J. Lee, "Building a Korean text summarization dataset using news articles of social media," *KIPS Transactions on Software and Data Engineering*, Vol.9, No.8, pp.251-258, 2020.
- [24] S. H. Yoon, A. Y. Kim, and S. B. Park, "Topic Centric Korean Text Summarization using Attribute Model," *Korean Institute of Information Scientists and Engineers*, Vol.48, No.6, pp.688-695, 2021.
- [25] Y. Jung, H. Hwang, and C. Lee, "Korean Text Summarization using MASS with Relative Position Representation," *Korean Institute of Information Scientists and Engineers*, Vol.47, No.9, pp.873-878, 2020.
- [26] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Sequence Generative Adversarial Nets with Policy Gradient," *arXiv preprint arXiv:1609.05473*, 2016.
- [27] K. Lin, and D. Li, "Adversarial Ranking for Language Generation," *arXiv preprint arXiv:1705.11001*, 2017.
- [28] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, and J. Wang, "Long Text Generation via Adversarial Training with Leaked Information," in *Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*, Vol.32, No.1, 2018.
- [29] L. Liu, Y. Lu, M. Yang, Q. Qu, J. Zhu, and H. Li, "Generative Adversarial Network for Abstractive Text Summarization," in *Thirty-second Association for the Advancement of Artificial Intelligence conference on artificial intelligence*, 2018.
- [30] Wang, Y. Shian, and H. Y. Lee., "Learning to Encode Text as Human-Readable Summaries using Generative Adversarial Networks," *arXiv preprint arXiv:1810.02851*, 2018.
- [31] R. Bhargavaa, G. Sharmaa, and Y. Sharmaa, "Deep Text Summarization using Generative Adversarial Networks in Indian Languages," *Procedia Computer Science*, Vol.167, pp.147-153, 2020.
- [32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017.
- [33] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Sequence generative adversarial nets with policy gradient. arxiv e-prints, page," *arXiv preprint arXiv:1609.05473*, 2016.
- [34] C. Li, W. Xu, S. Li and S. Gao, "Guiding Generation for Abstractive Text Summarization based on Key Information Guide Network," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol.2 (Short Papers), pp.55-60, 2018.
- [35] G. K. Palshikar, "Simple algorithms for peak detection in time-series," in *Proceedings of 1st International Conference of Advanced Data Analysis, Business Analytics and Intelligence*, Vol.122, 2009.
- [36] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *arXiv preprint arXiv:1908.10084*, 2019.
- [37] D. Miller, "Leveraging BERT for Extractive Text Summarization on Lectures," *arXiv preprint arXiv:1906.04165*, 2019.
- [38] M. Lewis, et al., "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," *arXiv preprint arXiv:1910.13461*, 2019.
- [39] C. Raffel, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *arXiv preprint arXiv:1910.10683*, 2019.
- [40] C. Lin, "ROUGE: A package for automatic evaluation of summaries" in *Text Summarization Branches Out*, pp.74-81, 2004.
- [41] F. Milton "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, Vol.11, No.1, pp.86-92, 1940.
- [42] P. B. Nemenyi, "Distribution-free Multiple Comparisons," Ph.D. thesis, Princeton University, 1963.



## 이 훈 석

<https://orcid.org/0000-0003-4775-2720>

e-mail : dolmani38@naver.com

1996년 건국대학교 물리학(학사)

2009년 연세대학교 전파통신공학과(석사)

2017년 단국대학교 컴퓨터공학(박사수료)

1999년 ~ 2006년 LG히다찌 주임연구원

2006년 ~ 현 재 아시아나IDT ICT융합연구소 수석연구원

관심분야 : 기계학습, 자연어처리, RFID 미들웨어



**안 순 흥**

<https://orcid.org/0000-0003-4500-9183>  
e-mail : kingmir@dankook.ac.kr  
2004년 단국대학교 전자계산(학사)  
2006년 단국대학교 컴퓨터과학(석사)  
2017년 단국대학교 컴퓨터공학(박사수료)  
2006년~현 재 아시아나IDT  
ICT융합연구소 선임연구원

관심분야: 분산알고리즘, 유무선망에서의 라우팅, 유무선망에서의 멀티미디어 통신, 기계학습



**김 승 훈**

<https://orcid.org/0000-0001-5021-3627>  
e-mail : edina@dankook.ac.kr  
1985년 인하대학교 전자계산학(학사)  
1989년 인하대학교 전자계산학(석사)  
1998년 포항공과대학교 컴퓨터공학(박사)  
1989년~1990년 한국전자통신연구원  
연구원

1990년~1993년 포스데이타(주) 정보통신본부 연구원  
1998년~2001년 상지대학교 컴퓨터공학과 조교수  
2001년~현 재 단국대학교 컴퓨터공학과 교수

관심분야: 분산알고리즘, 유무선망에서의 라우팅, 유무선망에서의 멀티미디어 통신, 기계학습

<부 록>

Table 4. Results 1 of Various Summarization Methods for an Article of the CNN/Daily Mail Dataset

원문	<p>Veteran actor Victor Spinetti, who starred in all three Beatles films, has died at the age of 82. The Welsh star, who also appeared in a string of acclaimed movies as well as taking roles in the West End and on Broadway, died after a fight with pancreatic cancer. Close friend Barbara Windsor, on whose Radio 2 show he made a recent appearance, was one of his final visitors before his death this morning at a hospice in Monmouth. Respected: Victor Spinetti was told by the late George Harrison he had to star in all the Beatles films, pictured here in 1972. Tributes: Actor Victor Spinetti, pictured left in 2010, died today at a hospice - one of his last visitors was close friend Barbara Windsor, pictured right with the actor in the 1960s. Spinetti's agent, Barry Burnett, said: He had cancer for a year, but he was very cheerful to the end. I spoke to him on Friday and he was talking about his plans and everything. The versatile actor was able to easily turn his hand from serious classical roles to comedy performances and roles in sitcoms. He was also known as successful stage director, wrote poetry and randomly became known for his appearances in a Jaffa Cake ad campaign as the Mad Jaffa Cake Eater. Star: Victor with John Lennon and Yoko Ono at the National Theatre, in 1969 - the actor starred in all three Beatles films. However, for many fans, Spinetti will always be known for his roles in The Beatles' three live action films - A Hard Day's Night, Help! and Magical Mystery Tour. It was his close friendship with the Beatles at the height of their fame which put him on the map. Spinetti was born in Cwm, Wales, on September 2, 1933, attended Monmouth School and the Cardiff College of Music and Drama of which in later life he became a fellow. The Wild Affair: Victor Spinetti starred in the hit British satire in 1963. However, his working life began as a waiter and factory worker before he sprang to prominence in three Beatles films of the 1960s: Hard Day's Night, Help! and Magical Mystery Tour. The late George Harrison once said to him: You have got to be in all our films. If you are not in them, my mum won't come and see them because she fancies you. " During his versatile career, Spinetti appeared in more than 30 films, including Zeffirelli's The Taming of the Shrew, Under Milk Wood, with Elizabeth Taylor and Richard Burton, Voyage of the Damned, The Return of the Pink Panther, and The Krays. Treading the boards: Victor Spinetti in The Merry Widow with Karl Daymond. His work with Joan Littlewood's Theatre Workshop produced many memorable performances including Fings Ain't Wot They Used T'Be and Oh! What a Lovely War, which transferred to New York, and for which he won a Tony Award for his role as an obnoxious drill sergeant. His West End appearances included Espresso Bongo, Candide, Cat Among the Pigeons and Chitty, Chitty, Bang Bang, He also played the principal male character in the feminist play, Vagina Rex. He also appeared on Broadway in The Hostage and The Philanthropist. With the Royal Shakespeare Company he appeared as Lord Foppington in The Relapse and as the archbishop in Richard III. Veteran: Victor Spinetti directed and won numerous awards over his 60-year career. Spinetti also co-authored John Lennon In His Own Write, which he directed at the National Theatre. He also directed productions of Jesus Christ Superstar and Hair. His many TV appearances included Take My Wife, and the sitcom An Actor's Life For Me. Spinetti also wrote poetry, notably Watchers Along The Mall, and prose which have appeared in several publications. His memoirs, Victor Spinetti Up Front, was filled with anecdotes, including the claim that Princess Margaret was instrumental in securing the necessary censor permission for the first run of Oh! What a Lovely War.</p>								
요약 방법	요약 내용	Intro	Body	Ending	DI	R1	R2	RL	
Ours (B)	<p>Spinetti, age 82. in his before visitors from Star: Victor Theatre, Beatles' Monmouth and The Wild Victor Spinetti in hit British working life began as waiter factory worker before he sprang prominence in three films of the Hard Day's Night, Help! and Magical Mystery Tour. The once be in mum " During the boards: Workshop for Espresso Among the also Hostage and With Royal Shakespeare he appeared as Foppington in The Relapse as the Richard Veteran: Victor Spinetti directed and won numerous awards over his 60-year career. Spinetti also co-authored Lennon His Own Write, which he directed at the National Theatre.</p>	0.4077	0.4592	0.4976	0.0013	0.30	0.16	0.29	
BERT+ LexRank	<p>September 2, 1933, attended Monmouth School and the Cardiff College of and Magical Mystery Tour. and Magical Mystery Tour. If you are not in them, my mum won't come and see them because she fancies you. Spinetti was born in Cwm, Wales, on.</p>	0.1387	0.2382	0.1554	0.0018	0.17	0.06	0.13	
BESM	<p>Veteran actor Victor Spinetti, who starred in all three Beatles films, has died at the age of 82. He was also known as successful stage director, wrote poetry and randomly became known for his appearances in a Jaffa Cake ad campaign as the Mad Jaffa Cake Eater. Music and Drama of which in later life he became a fellow. However, his working life began as a waiter and factory worker before he sprang to prominence in three Beatles films of the 1960s: Hard Day's Night, Help!</p>	0.5359	0.4216	0.3696	0.0048	0.36	0.24	0.22	
BART	<p>Actor Victor Spinetti died today at a hospice in Monmouth, Wales, at the age of 82 . Close friend Barbara Windsor was one of his final visitors before his death . Spinetti starred in A Hard Day's Night, Help! and Magical Mystery Tour . He won a Tony Award for his role as an obnoxious drill sergeant in Oh! What a Lovely War .</p>	0.5526	0.3900	0.3247	0.0091	0.41	0.26	0.22	
T5	<p>actor, 82, died after a fight with pancreatic cancer. close friend Barbara Windsor was one of his final visitors. spinetti starred in all three Beatles films of the 1960s. close friendship with the Beatles put him on the map.</p>	0.5775	0.3399	0.3004	0.0149	0.42	0.26	0.23	
Ground Truth	<p>Sprang to prominence in three Beatles films of the 1960s: Hard Day's Night, Help! and Magical Mystery Tour. Close friend Barbara Windsor was one of the last people to visit the veteran actor. His agent Barry Burnett, said: 'He was very cheerful to the end'</p>	-							

Table 5. Results 2 of Various Summarization Methods for an Article of the CNN/Daily Mail Dataset

원문	(CNN) -- British innovator James Dyson, who has built a multi-billion dollar empire around his distinctive vacuum cleaners, has described patent laws across Europe as absolute madness, saying they are unhelpful for inventors and small businesses. Dyson told CNN he wants the patent system in Europe radically overhauled. Over the last four decades, Dyson said, he has been affected enormously by people copying his ideas. Government leaders are continuously telling businesses that innovation drives the economy. But Dyson points to the red tape surrounding the patenting process as being a massive hurdle for businesses wanting to develop ideas. The problem with inventing: as soon as you file a patent they see what you are doing and they can see ways to get around it, said Dyson, who made his fortune inventing a bagless vacuum. The 64-year-old is an outspoken critic of Chinese counterfeiters, calling on governments to do more to protect intellectual property rights. Problems can arise because of the wording of the patents, Dyson said. There are no diagrams or drawings and often something hinges on the particular phrasing of the patent. According to Dyson, if it is obvious someone has copied another persons ideas they should be dealt with without the parties going through a protracted legal battle. There shouldnt be this endless rigmarole of could this have been devised by one skilled in the art? Because the current system involves many expenses due to varying jurisdictions throughout the continent, a Europe-wide patent is the answer, according to Dyson. You have to file in each country, you have to translate in each country, sue in each country, renew in each country -- its seen as a profit center for each country. Issues surrounding plagiarism are not limited to businesses, with consumers also feeling the impact of the high cost of producing and securing new inventions, Dyson said. Energy-saving and cost effective products and technology wont be created because of the enormous upfront investment it takes to develop them, he said Its anticompetitive to make copying easy. CNNs Emily Smith contributed to this story.								
요약 방법	요약 내용	Intro	Body	Ending	DI	R1	R2	RL	
Ours (B)	British innovator who has built a multi-billion distinctive vacuum cleaners. wants the patent Europe radically Dyson said, he been affected by Government businesses that innovation drives But Dyson points to the red tape surrounding the as a massive hurdle for wanting to develop ideas.	0.4996	0.4826	0.4235	0.0010	0.32	0.0	0.17	
BERT+ LexRank	CNNs Emily Smith contributed to this story. There shouldnt be this endless rigmarole of could this have been devised by one skilled in the art? Dyson told CNN he wants the patent system in Europe radically overhauled. The 64-year-old is an outspoken critic of Chinese counterfeiters, calling on governments to do more to protect intellectual property rights. There are no diagrams or drawings and often something hinges on the particular phrasing of the patent.	0.4782	0.4641	0.3416	0.0037	0.19	0.0	0.12	
BESM	(cnn) -- british innovator james dyson, who has built a multi-billion dollar empire around his distinctive vacuum cleaners, has described patent laws across europe as absolute madness, saying they are unhelpful for inventors and small businesses. dyson told cnn he wants the patent system in europe radically overhauled. over the last four decades, dyson said, he has been affected enormously by people copying his ideas. government leaders are continuously telling businesses that innovation drives the economy. there are no diagrams or drawings and often something hinges on the particular phrasing of the patent.	0.6105	0.4876	0.7259	0.0094	0.51	0.43	0.50	
BART	British innovator James Dyson has described European patent laws as 'absolute madness' Dyson made his fortune inventing a bagless vacuum cleaner . The 64-year-old is an outspoken critic of Chinese counterfeiters, calling on governments to do more to protect intellectual property rights . Dyson wants a Europe-wide patent system overhauled .	0.6582	0.4567	0.2520	0.0274	0.45	0.27	0.40	
T5	inventor James Dyson says patent laws across Europe are madness. he says they are unhelpful for inventors and small businesses. Dyson says he wants the patent system in Europe radically overhauled.	0.6747	0.4294	0.3399	0.0200	0.44	0.18	0.35	
Ground Truth	British innovator James Dyson has described patent laws across Europe as "absolute madness". Dyson says the current system involves many expenses due to varying jurisdictions throughout the continent. He says a Europe-wide patent is the answer.				-				