

An Intensity Controlled Review Dataset Construction for Automatic Review Generation

Nagyeong Kim^{*1}, Hyejin Jo^{*2}, Jueon Lee^{*3}, Jaeho Choi^{*4}, Byounguk Nam^{*5}, and Yuchul Jung^{**}

This work was supported by the Basic Research Program through the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) under Grant (2020R1A4A1017775).

Abstract

Most users refer to existing online reviews to see if other users were previously satisfied or not about the products (or foods) they want to buy. Meanwhile, some users do not want to write fruitful, realistic reviews because it is annoying and bothering. To minimize users' writing costs, we are interested in implementing an automatic review generator to create a complete review with only scores and a few seed keywords by enabling intensity control tailored to the user's needs. To this end, we propose an intensity controlled review data construction method for online review. Moreover, we employ GPT-2 and BART models popular for text generation tasks for the review generation experiments. Our automatic and manual evaluations for randomly sampled generation results prove the quality of our constructed dataset.

요약

대부분의 사용자는 온라인 리뷰를 참조하여 다른 사용자가 구매한 제품(또는 음식)에 대해 만족했는지 여부를 확인한다. 하지만, 일부 사용자들은 귀찮고 성가시기 때문에 리뷰를 작성하지 않는다. 따라서 사용자의 불편함을 최소화하기 위해 사용자의 요구에 맞춘 문장의 감정 강도를 조절하여 몇 개의 키워드만으로 문장을 생성한다. 이를 위해 온라인 리뷰를 위한 강도 제어 리뷰 데이터 구축 방법을 제안한다. 또한, 본 연구는 텍스트 생성 실험에 GPT-2 및 BART 모델을 적용했다. 무작위로 샘플링된 생성 결과는 기계 평가와 사람 평가를 진행하였고, 사람 평가의 경우 제안된 방법이 문맥, 문법, 강도의 적합도에서 우수한 평가를 받았다. 대부분의 문장들은 단순 모델 생성 문장에 비해 감정이 일관되고 자연스러운 데이터를 생성하였다.

Keywords

text-generation, auto-labeling, review, GPT2, BART, segmentation

* Undergraduate Student, Department of Computer Engineering, Kumoh National Institute of Technology
- ORCID¹: <https://orcid.org/0000-0001-9089-8931>
- ORCID²: <https://orcid.org/0000-0001-7365-6003>
- ORCID³: <https://orcid.org/0000-0003-1683-6710>
- ORCID⁴: <https://orcid.org/0000-0002-5261-9150>
- ORCID⁵: <https://orcid.org/0000-0001-9863-1369>

** Assistant Professor, Department of Computer Engineering, Kumoh National Institute of Technology
- ORCID: <https://orcid.org/0000-0002-8871-1979>

• Received: Dec. 28, 2021, Revised: Jan. 27, 2022, Accepted: Jan. 30, 2022

• Corresponding Author: Yuchul Jung

Department of Computer Engineering, Kumoh National Institute of Technology, 61 Daehak-ro, Geoui-dong, Gumi-si, Gyeongsangbuk-do
Tel.: +82-54-478-7536, Email: jyc@kumoh.ac.kr

I. Introduction

Online reviews are commonly found when purchasing most products. The reviews have a significant impact on sales and are an essential factor for both consumers and sellers. Consumers refer to reviews when purchasing products, and sellers use them to identify consumers' needs and fix problems to improve service quality. The recent rapid growth in the delivery app market has made reviews more used in the food delivery sector.

However, most people are hesitant to write a review in detail, and the satisfaction score doesn't honestly reflect what consumers' have actually felt. Scores are usually higher, and it is called 'Score Inflation.' It means the reliability of the food rating system has decreased due to some of the users who are asked to change their review (or score) from the sellers. The score is an essential factor from a seller's perspective because restaurants are exposed in descending order based on their score. So if the score is low, there is less possibility to be even seen by people. As a result, many consumers voluntarily give restaurants high scores, even if they think those restaurants need improvements.

We propose a new technique to generate high-quality review data considering the intensity of positive/negative for each quality factor of existing products (or services). For example, various information about the food is included in Fig. 1 (a).

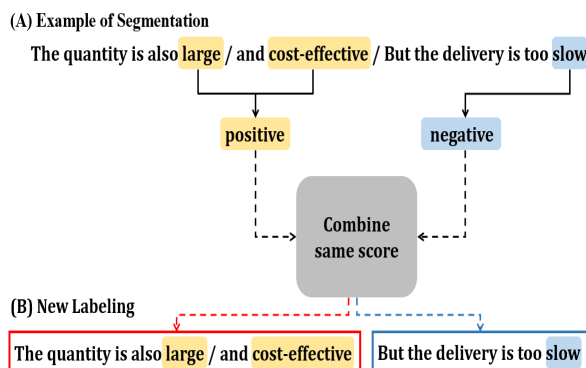


Fig. 1. Example of segmentation and new labeling

Segmentation was performed by extracting information for each aspect. In one sentence, the desired information is contained through a verb or an adjective, so when a verb or an adjective appears, the sentence is segmented. As shown in (b), sentences with the same strength were combined after sentiment analysis to adjust the new score.

Our contributions can be summarized into three folds:

1) Proposal of data annotation method considering the intensity of the data and the construction of learning data (In total, 524,596)

2) Implementation and experimentation of review generator employing GPT-2 [1] and BART [2], the latest text generation techniques based on high-quality training data

3) Evaluating the quality of the review generated: Evaluate five different models by combining different segmentation and scoring methods

As result, we succeeded in obtaining the highest perplexity of 12.8595 when Segmentation Considering Sentence Intensity & Labeling based on Existing scores (SCILE) method was applied. This study confirmed that users could create their reviews containing the keywords they want by simply setting keywords and intensity with the help of an automatic text generation which utilizes our intensity controlled review dataset construction approach.

II. Related Work

2.1 Review Generation

In recent text generation research, controlled text generation is in progress. A sentence generation considering simple positive/negative was proposed based on a pretrained language model [3]. The sentence generation method considered aspects beyond sentence generation through simple sentiment analysis. [4]. However, the segmentation method cannot be easily applied to various review expressions when

sentences are segmented based on the aspect. Moreover, by adding target content and target sentiment (e.g., positive or negative) to the input value to control the sentence, the output text was generated by inputting detailed content rather than simple sentence generation. [5]

In addition, a method of generating a rather long sentence by using at least five words as the input was proposed. Keywords are numbered and created sequentially [6]. In the case of sentence generation in this way, it depends on the input keyword without considering the intensity of the sentence. Therefore, in this paper, a detailed segmentation method was derived to generate sentences with intensity.

2.2 Aspect-based Sentiment Analysis

Aspect-based Sentiment Analysis (ABSA) method [7], which analyzes the sensibility of attributes in detail, can obtain user insights by identifying user sentiment and contributing factors for each significant aspect and aspect in the user's review. In addition, by applying the ABSA to the review of e-commerce, additional aspect-specific vocabulary was constructed [8]. When an aspect (service, location, food, etc.) appears in each sentence and words expressing sentiment appear, they are paired to judge negative and positive. Therefore, sentiment features were extracted for each aspect. However, in recent online reviews, more diverse expressions appear than before, and it is not appropriate to divide them into positive and negative. Therefore, in this paper, the intensity of the sentence is finely adjusted to 1-5 points, and various expressions are created when the sentence is generated for each intensity.

III. The Proposed Framework

3.1 Crawling Review Data

In this study, we crawled Korean review data of

Yogiyo [9], which provides Delivery Order Service. The total crawled reviews amount to 823,234, especially for the chicken category. To adjust the intensity based on the existing score, reviews and user scores were also collected.

3.2 Text Pre-processing

We refined the collected data to remove existing noises in review texts. First, since there are many sentences with incorrect spacing in the existing review data, spacing was newly performed using the PyKoSpacing library (Python package for automatic Korean word spacing) [10]. Additionally, the Pyhanspell (Hangul Spelling Library) library [11] was used for the spelling check. Second, we've selected reviews of between 30 and 200 characters, including spaces. Reviews of less than 30 characters have many ambiguous words and duplicated words, so we got rid of them from our dataset. In addition, we also deleted reviews of more than 200 characters because most of them were not grammatically correct and had unnatural flows. Third, we removed emoticons and special symbols that are not helpful for the review generation. After those text pre-processing steps, we finally got 233,003 reviews as training raw data.

3.3 Segmentation

3.3.1 Simple Segmentation

There are not many positive reviews for all elements in online reviews, even if users write a single-sentence review that is overall positive. Therefore, it is necessary to separate the sentences to efficiently extract the specific elements of the product contained in the review. Previously, Word Mover's Distance (WMD) [12] was applied to measure the similarity between consecutive sentences and segmentation [4]. However, in sentences of less than 200 characters, the method fails in segmentation, and

it is not appropriate for intensity control. Therefore, we suggest a new segmentation technique based on the features of reviews.

Users express desired information through verbs/ adjectives /adverbs. A Korean morpheme analyzer, KoNLPy Okt (Open Korea Text) [13], was used to extract them. In Korean, there is a postposition, so even if it is the same word, if a different postposition is attached to the end of the word, the meaning is different. Therefore, after morphological analysis, we segmented sentences based on predicates and adjectives/ adverbs. Those segmented sentences were separated using the Korean Sentence Splitter [14] and automatically divided into various aspects. The number of collected reviews is increased from 233,003 to 544,581 after the segmentation and duplicated sentence filtering.

3.3.2 Segmentation considering sentence intensity

In the simple segmentation, sentences are divided by various aspects, but many short reviews are generated. After the simple segmentation, we perform an automatic score labeling as in Section 3.4 to combine sentences with the same strength to resolve this problem. Through this, the expression of the sentence is diversified and unnecessary sentences are deleted. Algorithm 1 combined sentences when the sentence strengths were the same after simple segmentation. The method for determining the intensity of sentences was automatically scored by section 3.4.

3.4 Automatic Score Labeling

If we examine the actual reviews in detail, many reviews have negative sentiments, even though the sentences were rated 4 or 5 points, which are close to solid positive. Among the 823,234 reviews collected, 5-point scores occupy 76% of them. Meanwhile, 1-point and 2-point scores are less than 6%. In other words, most online reviews have overwhelmingly

positive ratings. Examples such as Table 1 show that some negative expressions are expressed in a written review, even for positive assessments. Therefore, we newly label the scores in two ways to solve the data imbalance of the collected data.

Table 1. New labeling example

Rating	Review
5	There's a lot of chicken, but it's a bit bland . The delivery is late, too.
4	It's good, but I'm disappointed that the portions are too small . The delivery is slow , too.

3.4.1 Labeling based on existing scores

Labeling was done based on the scores of existing users' reviews. After the sentiment analysis for the given sentence, labeling was performed using the method of adding 1 point when a positive word appeared in the current score and subtracting 1 point when a negative word appeared in the current score according to the number of emotional words. Therefore, a result with more than 5 scores was assigned as 5, and a result of less than 0 was assigned a score of 1. Each review's score was adjusted to 1-5.

3.4.2 Labeling new scores

In this method, the initial intensity was set to 3points without using the existing score. Based on the review sentiment dictionary, we used a method of adding 1point when positive words appeared and subtracting 1point when negative words appeared according to the number of words by analyzing emotions. The review used in this study was not effective when analyzed using the existing Korean sentiment dictionary because there are many new words and unique words of aspect. So, we constructed the emotional words used in the review separately.

The total number of words is 940, with 479 positive words and 461 negative words. Therefore, a result with a score of 5 or more was treated as 5, and a result less than 0 was treated as 1, making a total score between 1 and 5. Finally, when the methods of Section 3.3.2 and 3.4.1 were applied, score distributions were drastically changed from imbalance ones (5 points: 161,884 (76%), 4 points: 27,598 (13%), 3 points: 10,817 (5%), 2 points: 4,394 (2%), and 1 point: 8,620 (4%)) to somewhat balanced ones (5 points: 191,199 (36.45%), 4 points: 172,347 (32.85%), 3 points: 100,925 (19.24%), 2 points: 40,514 (7.72%), and 1 point: 19,611 (3.74%)) were constructed.

Algorithm 1 Segmentation considering sentence intensity

```

1: R = review
2:   E = {}
3:   for i = 0 to |R| do
4:     R.Remove
5:     R.SpellCheck
6:     S = R.Split
7:     for j = 1 to |S| do
8:       if S[j-1].emotion == S[j].emotion then
9:         S[j-1] += S[j]
10:        E[j-1] = S[j-1].emotion
11:        j--
12:       else
13:         E[j-1] = S[j-1].emotion
14:         E[j] = S[j].emotion
15:   return S, E

```

Fig. 2. Algorithm

IV. Experiments

4.1 Experimental Setting

To examine the feasibility of our segmentation method and the auto labeling method, we employed two popular text generation methods (i.e., GPT and BART). Moreover, we evaluate Perplexity, BLEU-4 [15], ROUGE score [16], BERT-SCORE [17] together with human judgments to explore the quality of generated review texts.

4.1.1 KoGPT-2

In this experiment, Korean GPT2 (KoGPT2) [18] published by SKT-AI was used to generate Korean sentences. To construct a model specialized in this paper, we performed additional learning through fine-tuning. The fine-tuning was performed at epoch 50, batch-size 32, dropout 0.1, and Adam optimizer. In this model, the Transformer-type decoder layer is set to 12 layers, and BPE-based SentencePiece tokenization [19] is used. The score (1-5) to control the intensity added a special token in the input data. KoGPT-1 ver1 used the KoGPT-transformer version in huggingface [20], and KoGPT2 ver2 increased the pre-size by more than 20GB compared to the existing ver1 in increased performance compared to the previous model.

4.1.2 KoBART

Korean BART [21] is a Korean encoder-decoder language model that has learned over 40GB of Korean text using the text infilling noise function used in the paper. In this experiment, the KoBART-base model was used. A word dictionary was created by tokenizing the training data using BertWordPieceTokenizer. In this model, six encoder layers and six decoder layers are set. Since we need to create an input-out data set, we experimented with relatively small data than the GPT.

4.2 Comparison Methods

For the review generation experiments, a total of 524,596 data (after segmentation) were used as training data, and KoGPT-2 ver1, KoGPT-2 ver2, and KoBART models were chosen as automatic generation techniques. The entire collected data went through the text pre-processing to precisely purify the data as in Section 3.2. Since the collected reviews were not official documents, sentences were composed only of emoticons, neologisms, and onomatopoeia. Moreover,

words that are difficult to understand and sentences with many overlapping meanings have been removed. To find the most appropriate combination of segmentation and scoring schemes that are introduced in Section 3.2 and 3.3, we've tested the following five methods.

- **No Segmentation & Existing Scores (NE):** It was trained in KoGPT-2 ver1, ver2, and KoBART models without data refinement process.
- **Simple-Segmentation & labeling based on Existing Scores (SSLE):** Data using simple-segmentation and labeling techniques using existing scores
- **Simple-Segmentation & Labeling New Score (SSLN):** Combination using simple segmentation and labeling new score.
- **Segmentation Considering Sentence Intensity & Labeling based on Existing scores (SCILE):** Data using Segmentation Techniques Considering Intensity and Labeling Techniques Using Existing Scores
- **Segmentation Considering Sentence Intensity & Labeling New score (SCILN):** Segmentation considering sentence Intensity and Data with labeling New scores without using existing scores

4.3 Experimental Results

When trained with the raw data without the pre-processing, both GPT-2 and BART-based review generation models produce sentences that do not conform to intensity and sentences that do not conform to grammars. Meanwhile, when we applied the text pre-processing, our proposed segmentation and labeling techniques, GPT-2 based models succeeded in generating sentences appropriate to the intensity with varying lengths.

The sentence is not natural when entering Negative Keyword because the number of negative data is small before pre-processing. SSLE and SSLN produce relatively short sentences in length and have a natural flow of sentences, but some created sentences are not

related to the input keyword. It also creates sentences that are not appropriate for intensity. SCILE adds longer sentences and more words than previous methods. In particular, SSILN produces a natural sentence compared to other methods, and the expression of the sentence is abundant

4.4 Evaluation of the Generative Models

BART requires a data set with an encoder-decoder structure, so the data is manually constructed, and the number of data is 7823. KoBART applied the simple-segmentation method and auto-labeling method. As mentioned earlier, BART is an input-output data set, so if segmentation is applied to one output (review), sentences that do not match the input (keyword) are derived. Due to these problems, it was challenging to construct the data set and apply the proposed five methods. Therefore, the datasets are relatively small, and only the Simple-Segmentation method and Auto-Labeling are applied. Unlike GPT, BART produced sentences similar to GPT or relatively short and produced slightly poor results even when the segmentation method considering the strength of Section 3.3.2 was applied.

In Table 2, SCILN and SCILE methods showed excellent BLEU-4, Rouge Score, and BERTSCORE [22], and SCILE had the lowest perplexity. However, the SSLN method also yielded similar figures. We conducted human judgments to evaluate the grammar and natural context flow of the sentences generated by each method. A total of 203 college students participated, and they evaluated the automatically generated review sentences by filling out an internet questionnaire. For a single query, nine sentences generated by GPT-2 and 1 sentence generated by BART are randomly listed and evaluated for naturalness (comprehension of the entire content), grammar, strength, and goodness of fit. Among the ten sentences, the sentence that best fits the item was selected.

Table 2. Automatic evaluation metric

Method		BLEU -4	ROUGE -1	BERT-S CORE	Perplexity
NE	KoGPT2 ver1	0.4557	0.5615	0.5534	39.5495
	KoGPT2 ver2	0.4926	0.5947	0.5617	40.8524
	KoBART	0.4763	0.4985	0.5147	52.3452
SSLE	KoGPT2 ver1	0.5213	0.6250	0.6377	22.1092
	KoGPT2 ver2	0.5722	0.7059	0.7261	21.9943
	KoBART	0.5387	0.5486	0.6462	35.6723
SSLN	KoGPT2 ver1	0.7043	0.8230	0.5743	13.9635
	KoGPT2 ver2	0.6506	0.6666	0.6673	17.3109
	KoBART	0.5567	0.5434	0.6743	25.2723
SCILE	KoGPT2 ver1	0.7499	0.7692	0.7718	16.4882
	KoGPT2 ver2	0.7808	0.8333	0.7634	12.8595
SCILN	KoGPT2 ver1	0.6224	0.6666	0.7113	20.1737
	KoGPT2 ver2	0.8600	0.8571	0.7502	20.7634

Table 3 is the Human Evaluation result table for the results generated by GPT-2 and BART. The Segmentation Considering Sentence Intensity & Labeling New score (SCILN) method received the highest score in all aspects of the overall generation result. In human judgments, SCILN showed higher values than other methods.

Table 3. Results of human judgements

	SSLE	SSLN	SCILE	SCILN
Context	6.4%	17.1%	6.4%	27.1%
Grammar	20.8%	12.9%	7.4%	36.1%
Goodness of fit	21.8%	7.9%	3.5%	30.2%

Table 4. SCILN EN/KO generation results

Keyword: Regular customer. (단골 손님)
Intensity: 5
I think I'll be a <u>regular customer</u> . The taste and quantity are great , and the service is good . (단골이 될 꺼 같아요 양 이랑 맛이 뛰어나고 서비스도 좋아요)

In fact, the SCILN-based generation-based results in Table 4 show more complete results than those of other methods.

V. Conclusion

We proposed an intensity-controlled review data construction approach for the automatic generation of delivery food reviews. Five types of labeling methods by combining segmentation and scoring were derived and compared with GPT-2 and BART-based review generation methods. We confirmed that our segmentation and score labeling techniques, considering the suggested strength, had the highest completeness of sentences through automatic evaluation and human judgments. Moreover, it is expected that the proposed data construction technique will enable a high-quality review of products and foods and further improve the reliability of the posted reviews. As future work, we plan to extend our approach to other fields, such as games, movies, and other services.

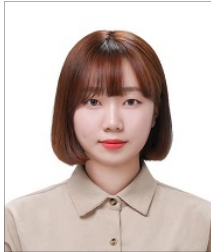
References

- [1] GPT[Website] <https://github.com/openai/gpt-2>, Available. [accessed: Oct. 25, 2020]
- [2] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension", Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, pp. 7871-7880, Jul. 2020. <http://dx.doi.org/10.18653/v1/2020.acl-main.703>.
- [3] Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu, "Plug and Play Language Models: A Simple Approach to Controlled Text Generation", ICLR, Seq 2019.
- [4] Parisa Kaghazgaran, Jianling Wang, Ruihong

- Huang, and James Caverlee, "ADORE: Aspect Dependent Online REview Labeling for Review Generation", Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event China, pp. 1021-1030, Jul. 2020. <https://doi.org/10.1145/3397271.3401074>.
- [5] Alvin Chan, Yew-Soon Ong, Bill Pung, Aston Zhang, and Jie Fu, "CoCon: A Self-Supervised Approach for Controlled Text Generation", ICLR, Jan. 2021.
- [6] Huajie Shao, Jun Wang, Haohong Lin, Xuezhou Zhang, Aston Zhang, Heng Ji, and Tarek Abdelzaher, "Controllable and Diverse Text Generation in E-commerce", Proceedings of the World Wide Web Conference, pp. 2392-2401, Apr. 2021. <https://doi.org/10.1145/3442381.3449838>.
- [7] Jun Yang, Runqi Yang, Hengyang Lu, Chongjun Wang, and Junyuan Xie, "Multi-Entity Aspect-Based Sentiment Analysis with Context, Entity, Aspect Memory and Dependency Information", *Acm*, Vol. 18, No. 4, Article No.: 47, pp. 1-22, Dec. 2019. <http://dx.doi.org/10.1145/3321125>.
- [8] Mohammad Erfan Mowlaei, Mohammad Saniee Abadeh, and Hamidreza Keshavarz, "Aspect-based sentiment analysis using adaptive aspect-based lexicons", *Expert System with Applications*, Vol. 148, Jun. 2020. <https://doi.org/10.1016/j.eswa.2020.113234>.
- [9] Yogiyo[Website], <https://www.yogiyo.co.kr/>. [accessed: Jun. 23, 2021]
- [10] Automatic Korean word spacing [Website], <https://github.com/haven-jeon/KoSpacing> [accessed: May 23, 2021]
- [11] Py-hanspell [Website], <https://github.com/haven-jeon/KoSpacing>, [accessed: May. 19, 2021]
- [12] Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger, "From word embeddings to document distances", Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille France, Vol. 37, pp. 957-966, Jul. 2015.
- [13] Eunjeong L. Park and Sungzoon Cho, "KoNLPy: Korean natural language processing in Python", Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology, pp. 133-136, Feb. 2014.
- [14] Split Korean text into sentences using heuristic algorithm[Website], <https://github.com/hyunwoongko/kss>, [accessed: May 14, 2021]
- [15] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "Bleu: a Method for Automatic Evaluation of Machine Translation", Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, pp. 311-318, Jul. 2002. <https://doi.org/10.3115/1073083.1073135>.
- [16] Chin-Yew Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", Association for Computational Linguistics, Barcelona, Spain, pp. 74-81, Jul. 2004.
- [17] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi, "BERTScore: Evaluating Text Generation with BERT", International Conference on Learning Representations, Addis Ababa, Ethiopia, Vol. 3, Feb. 2020.
- [18] SKT-AI[Website], <https://github.com/SKT-AI/KoGPT2>, [accessed: Oct. 05, 2020]
- [19] Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger, "Neural Machine Translation of Rare Words with Subword Units", Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, Vol. 1, pp. 1715-1725, Aug. 2016. <http://dx.doi.org/10.18653/v1/P16-1162>.
- [20] Taemin Lee [Website], <https://huggingface.co/taeminlee>, [accessed: May 30, 2021]
- [21] SKT-AI[Website], <https://github.com/SKT-AI/KoBART>, [accessed: May 25, 2021]
- [22] Hyunjoong Kim[Website], <https://github.com/lovit/KoBERTScore>, [accessed: May 15, 2021]

Authors

Nagyeong Kim



2017 ~ 2022 : Bachelor of Science in Computer Engineering
Research interests : Deep Learning, BigData , Data Mining

Hyejin Jo



2017 ~ 2022 : Bachelor of Science in Computer Engineering
Research interests : Deep Learning, BigData , Data Mining

Jueon Lee



2017 ~ present : Bachelor of Science in Computer Engineering
Research interests : Deep Learning, IoT , BigData

Jaeho Choi



2017 ~ present: Bachelor of Science in Computer Engineering
Research interests : Deep Learning, BigData , Computer Graphics

Byounguk Nam



2017 ~ present: Bachelor of Science in Computer Engineering
Research interests : Deep Learning, BigData , Data Mining

Yuchul Jung



2005 ~ 2011 : PhD degree in computer science from Korea Advanced Institute of Science and Technology (KAIST).
2009 ~ 2013 : Senior Researcher at Telecommunications Research Institute (ETRI)

2013 ~ 2017 : Senior Researcher at Korea Institute of Science and Technology Information (KISTI)
2017 ~ present : Assistant professor in Department of Computer Engineering, Kumoh National Institute of Technology (KIT), Gumi.

Research interests : Machine learning based NLP (text mining, sentiment analysis, automatic knowledge base construction, etc.), Korean speech recognition, and Medicine 2.0.