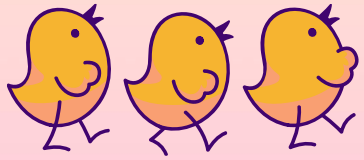
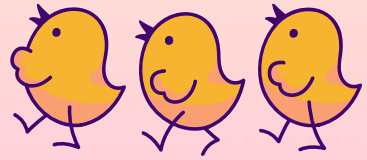


LangCon 2020



롯데의 딥러닝 모델은 어떻게 자기소개서를 읽고 있을까?

롯데정보통신
정보기술연구소 AI기술팀
김영준

목차

Language Conference

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

기술 키워드 : 자기소개서 전처리, 딥러닝, TextRank 알고리즘, Bi-LSTM+CNN

2. 자기소개서의 어색한 문장을 찾을 수 있을까?

기술 키워드 : 딥러닝, 유사도, MaLSTM, Transformer

3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

기술 키워드 : 딥러닝, BERT

4. 오타 어떻게 찾을까?

기술 키워드 : Levenshtein distance 알고리즘

5. 롯데그룹 채용 안내

김영준

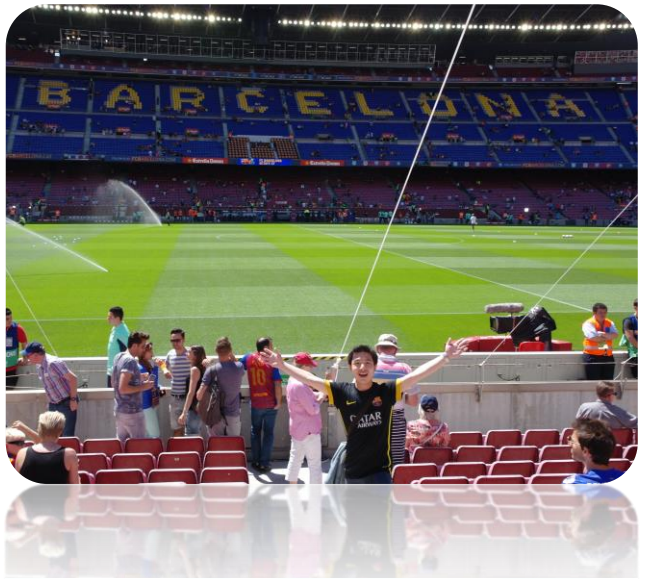
정보기술연구소 AI기술팀
자연어 처리 담당
채용AI

성균관대학교
데이터사이언스융합학과 석사

- 최우수연구상
- 수석 졸업

딥러닝과 NLP 전문가를 꿈꾸며 이와 관련된 모든 일을 사랑합니다.
가방 속에 있는 IT 서적이 무거울 수록 마음이 편안해지는 특이한 취향이 있습니다.

datajuny@gmail.com



Intro.

Language Conference

Predicting the Success of Job Application Using Personal Statements

ABSTRACT

A clear and well-documented LaTeX document is presented as an article formatted for publication by ACM in a conference proceedings or journal publication. Based on the "acmart" document class, this article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

datasets, neural networks, gaze detection, text tagging

ACM Reference Format:

. 2018. Predicting the Success of Job Application Using Personal Statements. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Nowadays, applying for jobs applications The Internet has made it easy for job seekers to submit resumes and self-introduction via the Web. Recruiters, on the other hand, were pressured to select prospect candidates from among many applicants who would be given an interview opportunity. As a result, companies are putting in a lot of manpower and time to review job seekers' documents. To address this, related research is actively being conducted.

However, existing studies, which are considered based on resumes, are suitable for hiring experienced employees, but are inappropriate for hiring new employees. There are two main reasons. First, most of the new employees don't have any experience, which is a key piece of information on their resumes, or have worked briefly as interns even if they do. Second, self-introduction is more important than resume for hiring new employees. This is because the potential for growth is more important than the ability to perform tasks right now because there is a small gap in the ability of applicants to perform their jobs. Therefore, recruitment of new employees who lack experience and have important future growth potential means more to self-introduction than resumes.

This paper makes the following contributions to the field of recruitment research.

- Unlike previous studies, new employees can determine the suitability of applicants only by introducing themselves, even without referring to their resumes and job requirements.
- Using TextRank to remove low-priority sentences from the cover letter proves that the accuracy of the deep learning model can be increased.

2 RELATED WORK

Among the numerous resumes, research has been actively conducted to find suitable candidates. In 2013, Paliaglia and others introduced applications that used machine learning and semantic matching technology to find the right candidates [3]. The application showed a methodology for extracting the necessary information from a linked-in profile and comparing it with job requirements.

Meanwhile, recruitment-related studies are focused on effectively comparing resumes and job requirements. In 2014, P. Montuschi and others criticized the need for complex work to identify a variety of variables that affect recruitment in advance, proposing a system to identify suitable candidates by taking into account how much the resume and Job offer conforms semantically [4]. In 2015, Cabrera-Diego et al. published a paper to find suitable candidates by comparing similarities in the vocabulary of resumes and job requirements [5].

Some studies use knowledge bases to compare resumes and job requirements. In 2017, Zaro and others proposed how to use the Integrated Knowledge Graphsit as a hybrid approach that automatically ranks the conceptual-based classification between resumes

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00

https://doi.org/10.1145/1122445.1122456

in a good way by cover letter alone. We hope that this study will provide a new direction for the study of new recruits.

3 CLASSIFICATION OF SELF INTRODUCTION LETTER

A deep learning model is used to classify the letter of self-introduction. The Deep Learning model is a mixture of 1D-CNN and Bi-LSTM. The ensemble effect of the two deep learning models can be expected by extracting the local feature of the self-introduction book with 1D-CNN and reading the self-introduction sentence in both directions with Bi-LSTM.

In addition, the textRank algorithm is used to reduce the length of the input data sequence by removing low-priority This saves learning time for deep learning models and improves accuracy. Meanwhile, no natural language can be entered into the Deep Learning model. Therefore, it performs word embedding that converts natural language into vector. To this end, a preprocessed self-introduction book is created. Self-introduction is divided into Morphology units and in this paper, Embedding by Word2Vec method [11].

3.1 Dataset Description

The research uses a self-introduction letter from applicants for Lotte Group's new employees. The self-introduction data are collected for research purposes with the consent of the applicant, and the personal information is unidentifiable. In order to format the self-introduction data, all new applicants were required to answer five identical questions within a limited number of characters. The five categories and questions are as shown in Table 1.

In this paper, 5,342 cases that are judged by the recruiter to be excellent self-introduction certificates and 5,342 cases that are the opposite are selected to form learning data (train data) so that the ratio of labels is 1:1. Finally, test data is for self-introduction in the second half of 2018. The study and test data configured above are divided by task, and a dataset is prepared as shown in Table 2. I chose the top three jobs that have a lot of learning data so that Deep Learning models can learn. In this paper, the Deep Learning model is learned and the performance evaluation is carried out by utilizing the prepared dataset.

3.2 Preprocessing

The self-introduction document should be preprocessed with special characters and quotation marks, excluding Korean, English, and numbers, all deleted using regular expressions. In addition, all English letters are changed to lowercase. Afterwards, the self-introduction sheet is torque-nized to form elements using the Okt Morphology Analyzer in Python's KoNLPy library. 249,930 self-introduction statements are separated by sentence into a single file to create a tormenting self-introduction corpus. Finally, add tokens for Zero-Padding to the corpus.

Perform Word2Vec Embedding using Python's genism library with completed self-introduction corpus. Word2Vec Embedding is a skip-gram method that predicts peripheral words with a more powerful central word than CBOW. In addition, the window size is 10 and only the words that appear more than 10 times are embedded. The result of having the word vector embedded in 128 dimensions was a dictionary with 100,365 words. To increase the similarity to show the words "passion," "
order in the dictionary.

The Deep Learning model does not
the self-introduction should be char
word dictionary that was created.
introduction book is changed to inde
by form factor. The flowchart in Figu
is changed. First, the median sequer
answer to each question. The medi
answers 1 and 5 with a limited num
180, respectively. The median length
2, 3 and 4 with a character limit of
reason why the center value is bas
the average of the length of the se
is less sensitive to outliers than the
equal lengths for each answer, a res
performs zero-padding at the fron
longer than the median will delete th

Figure 1: Process of word indexing

3.3 Extraction of key sentences using TextRank

The pre-treatment and Embedding process performed earlier is a common method used in natural language processing. In this paper, using the TextRank algorithm, it is recommended to leave only the key sentence and add a pre-treatment process to remove unnecessary sentences. TextRank is a graph-based ranking algorithm that calculates the importance between words or sentences, allowing you to select high-priority words and sentences [1]. The following figure illustrates the difference between the pre-treatment process and the existing method proposed in this paper.

Using TextRank to extract key sentences can reduce the length of the sequence of input data in the deep learning model. This has two main effects. First, the accuracy of the model can be improved by removing less important sentences from the self-introduction statement. This is similar to the effect of Feature selection in machine learning. In machine learning, Feature selection is a technique that creates a concise set of features by removing unnecessary features based on the importance of the feature among the various features. Traditional feature selection techniques are difficult to apply because natural language can be deconstructed by the removal of certain morphemes within a sentence. Therefore, in this paper, we added a pre-treatment process to eliminate unnecessary noise while preserving the meaning as much as possible. This helps prevent overfitting and improve the accuracy of models by making self-introduction brief. Second, by reducing the number of parameters

Table 1: Selected Examples of Self-introduction Letters

Category	Question	Character limi
Motive for applying	Please describe the motive for applying in detail.	500
Growth process	Please describe the growth process in detail.	800
Social activity	Please describe the most memorable experience that you had with interest and passion in addition to studying.	800
Job experience	Please describe your strengths and weaknesses in the preparation of the desired job and the desired job.	800
Future plan	Please describe your company life scenario for 10 years after joining the company and why you are pursuing it.	500

Table 2: Dataset Distribution for Training and Test

Job Description	Training dataset		Test dataset	
	Failed(0)	Passed(1)	Failed(0)	Passed(1)
Sales Management	852(41.6%)	1,195(58.4%)	93(58.5%)	66(41.5%)
Production Management	583(62.3%)	353(37.7%)	66(72.5%)	25(27.5%)
Programming	237(45.1%)	288(54.9%)	24(22.6%)	82(77.4%)
Others	3,670(51.1%)	3,506(48.9%)	1,138(52%)	1,052(48%)
Total	5,342(50%)	5,342(50%)	1,321(51.9%)	1,223(48.1%)

Table 3: Similar Words Using Cosine Similarity

Rank	Passion		Words Parents		Programming	
	similar words	similarity	similar words	similarity	similar words	similarity
1	Commitment	.714	Mother	.860	Java	.881
1	Thermal	.674	Mom	.786	Coding	.875
1	Affection	.666	Dad	.781	Data Structure	.863
1	Ambition	.655	Father	.769	OOP	.828
1	Mental	.644	Grandmother	.756	Python	.823

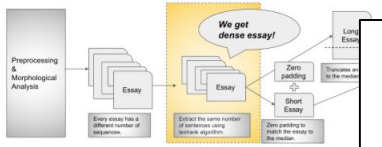


Figure 2: Overall approach of extracting key sentences using TextR

that the Deep Learning model needs to learn, the calculation of the model can be reduced. This has the effect of shortening the learning and inference time of the deep learning model.

In order to extract key sentences, the similarity between sentences must be calculated. In this paper, the similarity between sentences utilizes the TextRank formula proposed by Michalcea and others in 2004 [1]. The fo
of the number of morpheme
the following formula, S1, S2 i
morpheme common to both se

If all similarity between sent
formula, the textRank score pe

the following formula. V_i, V_j is the vertex, corresponding to the sentence in this paper, and TR(V_j) means the TextRank score of that sentence. w_{ji} is a similarity between sentences obtained earlier. The damping factor borrows 0.85 used in the paper [1].

First, to calculate the TR(V_i), the sum of the parallels between sentences is normalized to 1. The normalized values are then multiplied by the TextRank score in the Target statement and added together to obtain the final TR(V_i) considering the attenuation rate. As a result, TextRank scores are calculated for each sentence and important sentences are identified in order.

The following is the result of calculating TextRank scores for each sentence after analyzing the morphology of the sample responses to social experience. The average of the TextRank score in the example answer is 1. The summary is summarized in the order in which the TextRank score is low. For example, to summarize the example answer to 13 sentences, you can remove the 10th sentence with the lowest TextRank score ranking. In the example answer, except for the 10th sentence, you can see that there is no significant difficulty in understanding the content. In this paper, the optimal number of summary sentences is obtained through experiments in this way and the accuracy of the deep learning model is increased.

3.4 Model for Classification of Self Introduction Letter

The letter of self-introduction consists of five answers in total. Therefore, five cells were placed so that the deep learning model could accommodate five answers. Each cell (cell) is divided into 1D-CNN and Bi-LSTM stacked up with two layers. The figure 3 shows our deep learning model for the classification of self-introduction letters.

First, 1D-CNN extracts the characteristics of three, four, five and six consecutive sequences by having a kernel of 3, 4, 5 and 6. Zero padding is performed so that the length of the result values passed through each kernel match. The representative value is then extracted through the Max-Pooling Layer. This structure is Y. Kim. I designed it by referring to the CNN structure of the paper.[12] The Flatten Layer was then placed so that it could pass through the Dense Layer. Then there was a drop-out layer that effectively prevented overfitting.

The next two-tier Bi-LSTM connects the last vectors (outputs) of the forward LSTM with the last vectors (outputs) of the reverse LSTM as shown in Figure 7. It then passes through the Dense Layer and Drop out Layer, as in 1D-CNN. The number of nodes in the Dense Layer after 1D-CN is 128 which is more than the number of Dense Layer nodes in Bi-LSTM. This is because 1D-CNN is more accurate than Bi-LSTM for classifying self-introduction statements. After passing through each cell (cell), connect the resulting vectors of 1D-CNN (concategory) and the resulting vectors of Bi-LSTM (Concate to each other. The connected vectors pass through the Dense Layer again, as shown in the figure, and finally through one node.

Meanwhile, activation functions are very important in deep learning. The enable function is a function that converts the output values that pass through each node This can solve nonlinear problems. The activation functions of the deep learning model proposed in this paper use hyperbolic tangent functions (tanh) except for the last node. The last node uses a sigmoi
tween 0 and 1. Afterwards, 0.5 or mo
and less than 0.5 is classified as faile
value of the hyperbolic tangent func
sigmoid function between 0 and 1. T
low probability of the problem of the
RELU and Leakey ReLU activation f
this deep learning structure, perfor
bolic tanget functions (tanh). The lo
functions of the sigmoid family in th
cross-entropy known for good congr

The learning rate is 0.001 for Hy
size, 25 for epoch, and 0.5 for hyper
Experiment in the following chapter
of the deep learning model.

4 RESULTS

4.1 Optimization of the n
sentences

The accuracy of the Deep Learning m
iment and the accuracy of the model
and Accuracy. The Deep Learning model and pretreatment methodology proposed in this paper are marked Ours. In the experimental assessment table, F1-Score and Accuracy, which have the highest performance by position, were written in boldface type.

When learned without job discrimination, Ours was 8.25% and 1.22% more accurate than single models of Bi-LSTM and 1D-CNN, respectively. F1-Score was 0.08, 0.09 higher than Bi-LSTM and 0.01 higher than 1D-CNN. Ours' F1-Score showed a balanced accuracy, not bias, between passing and failing, with 0.8 and 0.82, respectively.

When all data was learned without job discrimination, the performance was higher than that of the job classification. However, it is hard to hastily interpret that it is advantageous for new employees to enter this into the Deep Learning model without distinction of duties. There are two main reasons for that. First, the number of learning data decreased as jobs were divided. Second, the ratio of passing and failing learning data to learning data has not been 1:1. However, self-introduction is a way to achieve good performance without separating tasks if there is not enough data for each task because it is a domain that lacks data.

Programming and sales management tasks, which had a high percentage of passing data in their learning data, classified passing better than failing. On the other hand, production management tasks, which had a higher percentage of reject data among learning data, classified failures better than pass. Therefore, the amount of learning data (Train Data) and the ratio of pass and fail can affect deep learning performance.

Sales management, which had the largest learning data among the three tasks, showed relatively similar figures for F1-Score, both passing and failing. Programming and production management tasks, where learning data were relatively scarce, show that the F1-Score of passing and failing have been learned one way or the other.

Meanwhile, the Deep Learning model and pretreatment methodology proposed in this paper have not always shown good performance. Sales management tasks showed 1D-CN with 1.26% higher

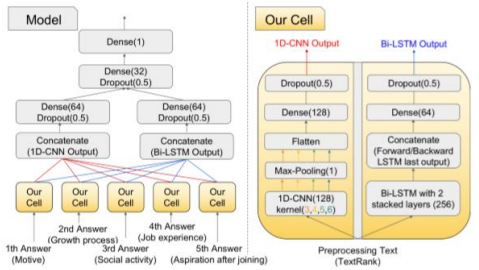


Figure 3: Deep learning model for the classification of self-introduction letters

Table 4: Comparison of Performances

Position	Model	Precision		Recall		F1-score		Accuracy
		Fail(0)	Pass(1)	Fail(0)	Pass(1)	Fail(0)	Pass(1)	
All	Bi-LSTM	.69	.77	.76	.70	.72	.73	72.78%
	1D-CNN	.72	.88	.87	.75	.79		
	Ours	.74	.88	.87	.76	**.80**		
Programming	Bi-LSTM	.69	.77	.76	.70	.72		
	1D-CNN	.72	.88	.87	.75	.79		
	Ours	.74	.88	.87	.76	**.80**		
Production Management	Bi-LSTM	.69	.77	.76	.70	.72		
	1D-CNN	.72	.88	.87	.75	.79		
	Ours	.74	.88	.87	.76	**.80**		
Others	Bi-LSTM	.69	.77	.76	.70	.72		
	1D-CNN	.72	.88	.87	.75	.79		
	Ours	.74	.88	.87	.76	**.80**		

accuracy. F1-Score was also 1D-CN higher than Ours by 0.01 and 0.02 respectively. The F1-Score, based on acceptance of production management tasks, was also 0.04 higher than Ours.

Overall, based on the best performance of Ours, the accuracy of the sales management tasks was lower than that of other tasks. Inference of the reason may have a correlation with how much of the job the recruiter is in charge of. Programming and production management tasks involve relatively less emphasis from recruiters.

This is because the mu
ence a applicant's self-
any recruiter. On the o
assess whether a Sales M
sales than its associate
involvement of recruite
management tasks have

on which recruiter reviews his or her letter of introduction. Therefore, it can be interpreted that the less involved the job of recruiters, the more accurate the Deep Learning model is.

5 RESULTS

6 CONCLUSION

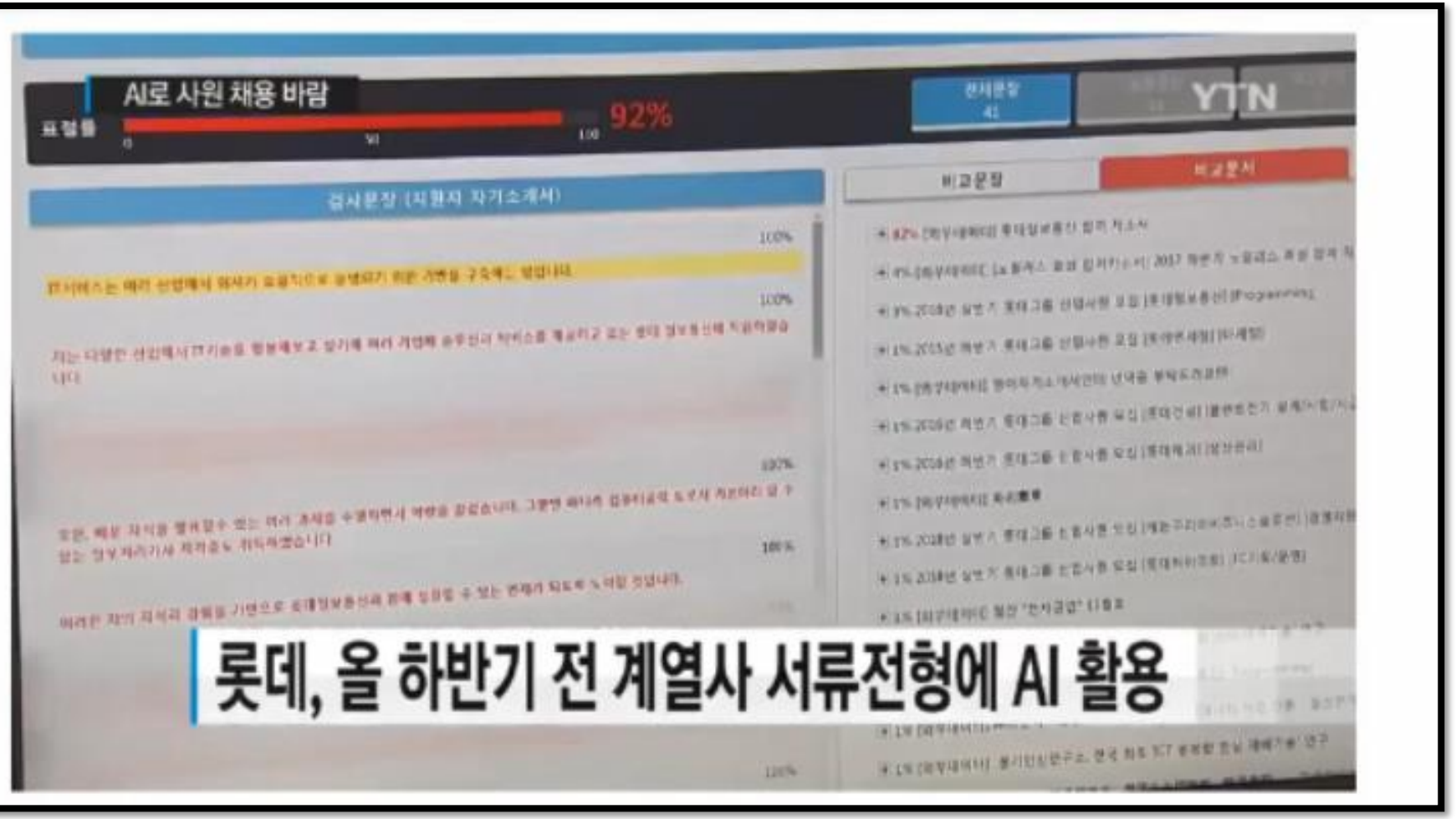
This study shows that stereotyping self-introduction with the same questions and limited number of characters can be good for hiring new employees in the future, and offers a new direction for recruitment-related research. Unlike previous studies focusing on resumes and job requirements, new employees have shown that they can judge pass and fail only by self-introduction. Finally, by not considering the academic background, gender or age presented in the resume, the company has practical utility to contribute to the formation of a balanced recruitment culture and reduce the work burden of recruiters.

In this paper, two major methods are proposed to determine pass and fail. First, introduce a deep learning model that classifies pass and fail admirably. Second, the pre-processing process of removing low-priority sentences from the self-introduction book shows that the performance of the deep-learning model can be enhanced. The proposed Deep Learning models were 8.25 percent and 1.22 percent more accurate than single models of Bi-LSTM and 1D-CNN, respectively. F1-Score was 0.08, 0.09 higher than Bi-LSTM and 0.01 higher than 1D-CNN. In addition, it showed balanced accuracy, not bias, between passing and failing. In addition, the 800-character answers in the self-introduction form were summarized in 13 sentences, and the 500-character answers proved by experiment that the performance of the deep learning model was best when summarized in nine sentences.

Future research should be conducted in two main directions. First, each applicant should be able to explain the reasons for their acceptance and rejection. This is because Deep Learning models can be fully entrusted with hiring documents. Second, how to help in the interview process that follows the document screening process. For example, the question of whether the applicant's answers at the interview match the self-introduction statement or whether the applicant's answers are in real time to inform the interviewer of what part of the cover letter will lead to an interesting natural language processing (NLP) task.

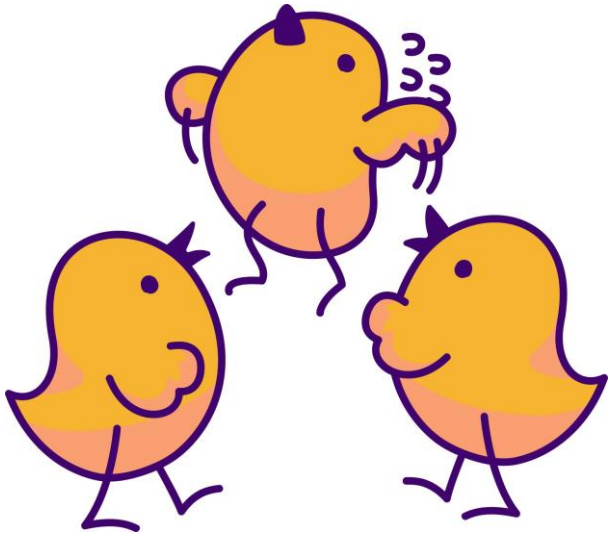
ACKNOWLEDGMENTS

REFERENCES



롯데그룹의
[채용AI]는
YTN, 전자신문 등
각종매체에서 소개된
AI 시스템 입니다.

- ✓ 자기소개서는 여전히 사람이 읽고 있습니다!
- ✓ 저는 [채용AI] 기술 담당자이지 채용 담당자가 아닙니다!
  - Q. '어떻게 하면 딥러닝이 합격으로 분류하나요?'
  - A. '저도 궁금해요!'
- ✓ 본 세션은 딥러닝을 이미 알고 계신 분이 실무에 어떻게 활용되는지 궁금하신 분이 들으면 좋아요!



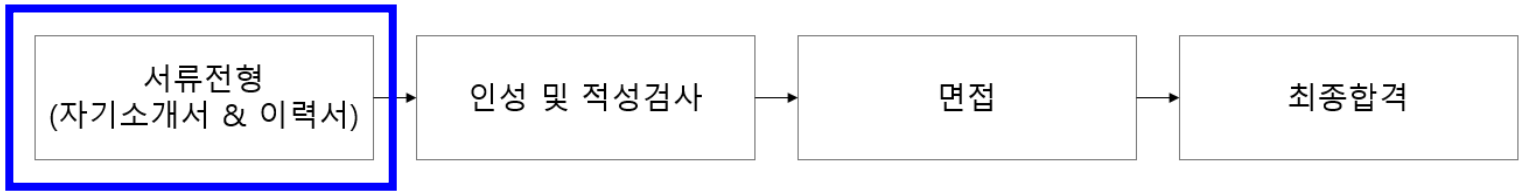
1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

기술 키워드 : 자기소개서 전처리, 딥러닝, TextRank, Bi-LSTM+CNN

LangCon 2020

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference



[신입사원 채용 프로세스(Process)]

- 2019년 하반기 롯데 신입사원 지원자 약 49,000명.
- 신입사원 채용은 이력서보다 자기소개서를 중요하게 평가.
  - 서류전형이 마감되면 인적성 검사일까지 10일안에 서류전형을 끝내야 한다.
  - 경력이 없는 신입사원의 업무 수행능력은 비슷하기 때문에 지원자의 잠재력을 자기소개서로 평가한다.
- 매 채용시 지원자는 동일한 질문을 제한된 글자 수 내에서 답변을 하고 있다. 그렇다면 딥러닝으로 우수한 자기소개서를 이항분류(Binary Classification) 할 수 있지않을까?

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

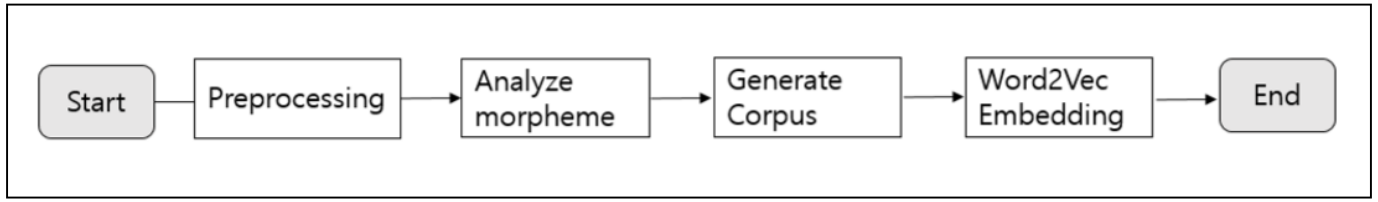
자기소개서 질문 및 답변 예시

Index	범주	질문 및 답변	제한 글자수 (공백포함)
1	지원동기	지원동기를 구체적으로 기술해주세요. *[핵심은 사람이다] 학교 선배들, 지인 중 유통업에 종사하는 사람들이 회사에 대한 만족도가 높고...*	500
2	성장과정	성장과정을 구체적으로 기술해주세요. *[만화책을 위해 시작한 아르바이트] 중학교 1 학년 때는 순정만화를 참 좋아했습...*	800
3	사회활동	학업 이외에 관심과 열정을 가지고 했던 다양한 경험 중 가장 기억에 남는 것을 구체적으로 기술해주세요. *[춤과 노래하기를 잊게 만든 브랜딩 마케팅] 브랜딩 마케팅은 학문이지만 저에게는...*	800
4	직무경험	희망직무 준비과정과 희망직무에 대한 본인의 강점과 약점을 기술해주세요. *[풍부한 프로젝트 경험] 업무 수행에 필요한 Java, JSP, Oracle Database 를 꾸준히...*	800
5	입사 후 포부	입사 후 10 년 동안의 회사생활 시나리오와 그것을 추구하는 이유를 기술해주세요. *[고객의 니즈를 충족시키는 MD] 옴니 채널을 통해 유통시장에 변화를 불어넣고...*	500

- 딥러닝의 시작은 **데이터 수집**!
- 지원자에게 연구용으로 데이터가 활용됨을 알리고 동의를 받았어요!
- 덕분에 개인정보가 비식별화된 데이터를 모을 수 있었답니다.

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference



[자기소개서 전처리 및 임베딩(Embedding) 순서도]

- 자기소개서의 한글, 영문, 숫자를 제외한 특수문자, 따옴표 등은 정규표현식을 사용하여 모두 제거한다.
- Okt 형태소 분석기로 자기소개서를 형태소 단위로 토크나이징(Tokenizing)한다.
- 토크나이징(Tokenizing)된 자기소개서 → 자기소개서 말뭉치(Corpus)
- Word2Vec Embedding은 CBOW 보다 성능이 좋은 중심단어로 주변 단어를 예측하는 Skip-gram

⇒100,365개 형태소(*128 차원으로 Embedding / 10 번 이상 등장한 단어로 제한 / Window 크기 : 10)

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

[고양이 + 애교 = 강아지]

[손흥민 + 아이버슨(농구선수) = 박지성]

(*자기소개서에서 존경하는 인물로 자주 등장하는 인물)

[남자 - 여자 = 다른 사람]

```
w2v_model.wv.most_similar(positive=['고양이', '애교'], topn=1)
```

```
[('강아지', 0.7390154600143433)]
```

```
w2v_model.wv.most_similar(positive='남자', negative='여자', topn=1)
```

```
[('다른사람', 0.5418267250061035)]
```

```
w2v_model.wv.most_similar(positive=['손흥민', '아이버슨'], topn=1)
```

```
[('박지성', 0.7613238096237183)]
```

```
w2v_model.wv.most_similar("부모님")
```

```
[('어머니', 0.8601336479187012),
 ('어머님', 0.7858796715736389),
 ('아버지', 0.7813066244125366),
 ('아버님', 0.7686976194381714),
 ('외할머니', 0.7564709782600403),
 ('고모', 0.7508319616317749),
 ('집안', 0.7503125667572021),
 ('키워주신', 0.7484757900238037),
 ('키우셨습니다', 0.7364680171012878),
 ('동생', 0.734329104423523)]
```

```
w2v_model.wv.most_similar("프로그래밍")
```

```
[('java', 0.8812177777290344),
 ('프로그래밍언어', 0.878530740737915),
 ('코딩', 0.8749273419380188),
 ('자료구조', 0.8630567789077759),
 ('자바', 0.8379415273666382),
 ('객체지향', 0.828315258026123),
 ('python', 0.8225173950195312),
 ('소프트웨어', 0.8223742842674255),
 ('임베디드', 0.8213320374488831),
 ('운영체제', 0.8109732866287231)]
```

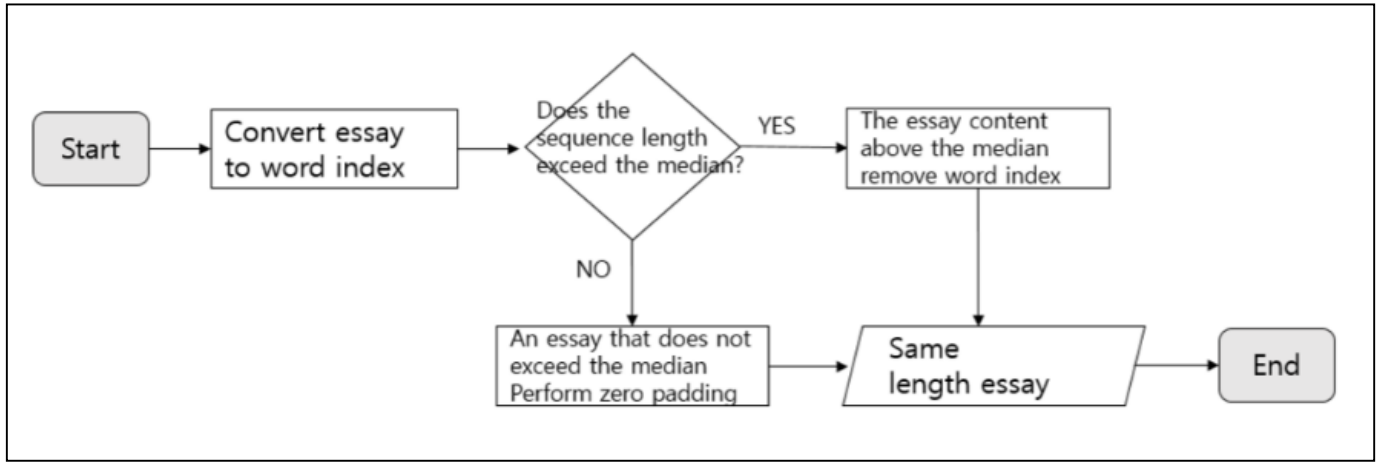
```
w2v_model.wv.most_similar("손흥민")
```

```
[('축구선수', 0.6439859867095947),
 ('안정환', 0.629976212978363),
 ('맨유', 0.6285330057144165),
 ('박지성', 0.6210668683052063),
 ('선수', 0.6134283542633057),
 ('아이버슨', 0.6086547374725342),
 ('토트넘', 0.591521143913269),
 ('유나이티드', 0.5879682898521423),
 ('프리미어리그', 0.580830454826355),
 ('야구선수', 0.579066872596740 7)]
```

임베딩(Embedding)이 잘 됐나 간 한번 볼까요?

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference



[자기소개서 단어 인덱스 변경 순서도]

- 딥러닝 모델에 입력하기 위해 형태소를 사전의 **인덱스**로 변경한다.
- 각각의 질문에 대한 답변 별로 시퀀스(Sequence) 길이의 **중앙값**을 구한다.

(*평균이 아닌 중앙값을 기준으로 삼은 이유는 중앙값이 평균보다 이상치(Outlier)에 덜 민감하기 때문이다.)

- 답변 별로 동일한 길이의 시퀀스를 만들기 위해 중앙값보다 **짧은 답변**은 자기소개서 앞쪽에 **제로 패딩**

중앙값보다 **긴 답변**은 중앙값 이후 단어 인덱스를 **삭제**한다.

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

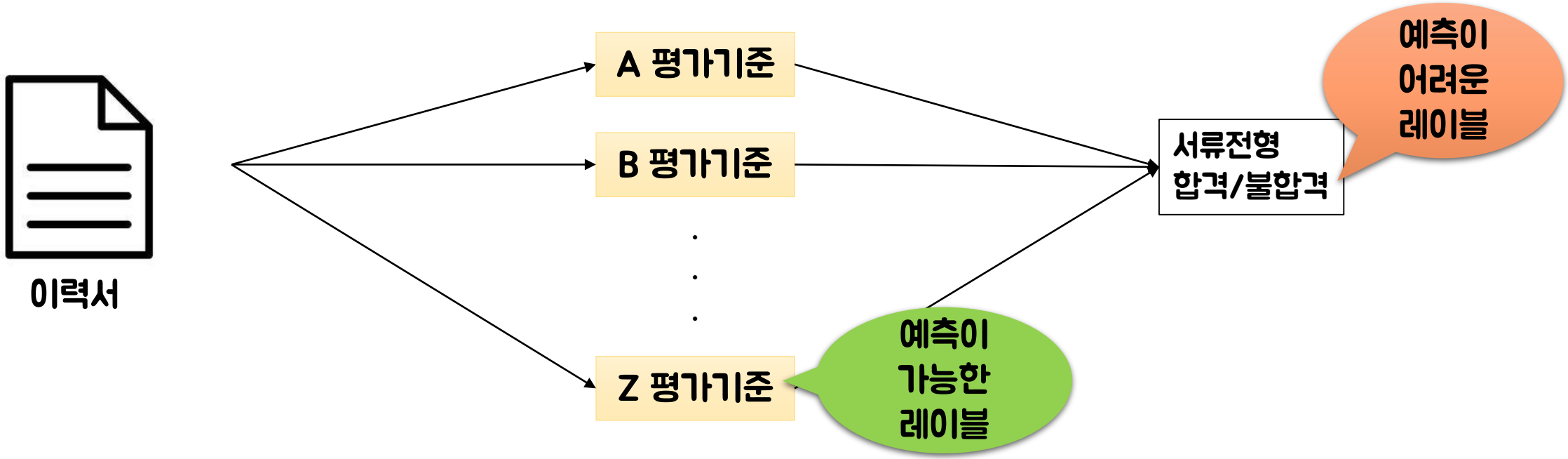
그러나 베이스 모델의
정확도는 기대에 미치지
못했습니다.
무엇이 문제였을까요?

★딥러닝 정확도 15% 이상 향상

문제는 레이블링이였습니다.
서류전형 합격이라고 모두
우수한 자기소개가 아니고,
서류전형 탈락이라고 모두
우수하지 못한 자기소개서가
아니였던 것입니다.

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference



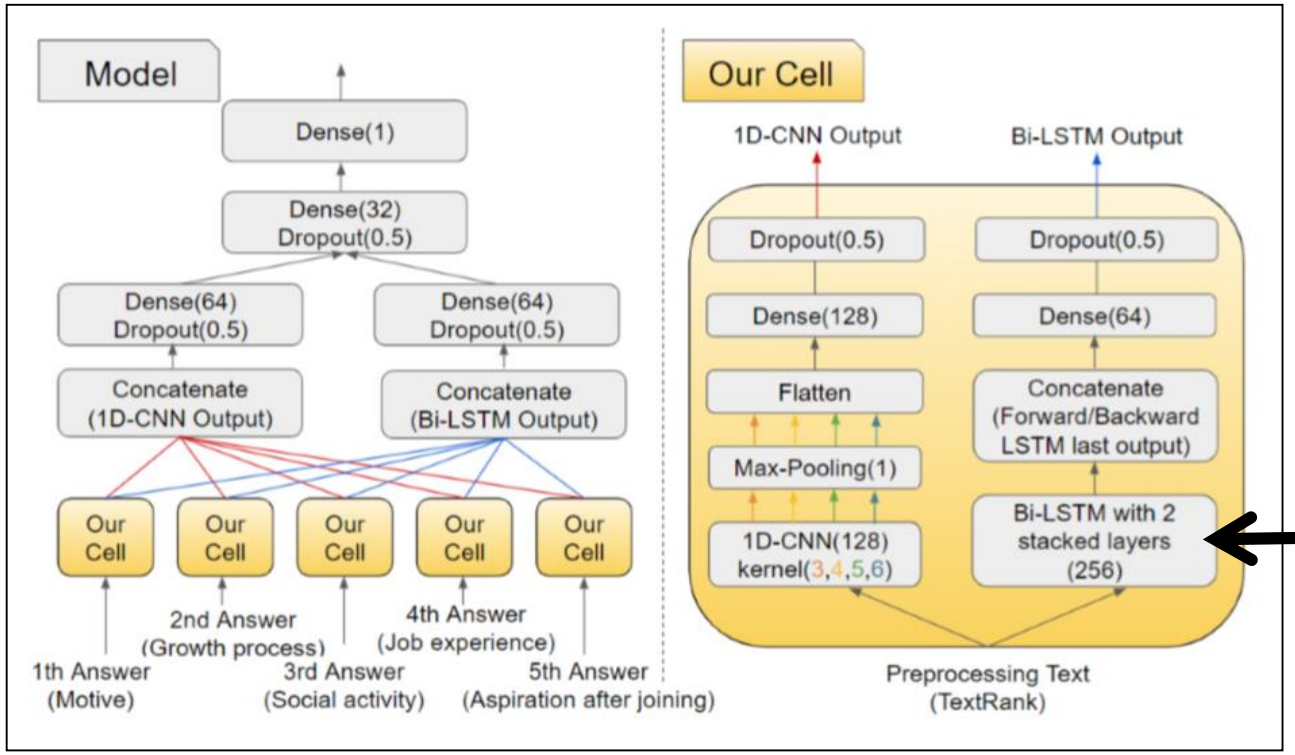
- 자기소개서를 읽고 우수함을 레이블링(Labeling) 하기에는 평가자의 주관적 요소가 들어갈 수 있을뿐더러 많은 비용을 필요로 함.
- 그룹사의 필요인재부합도(PFA)라고 하는 평가기준을 레이블링(Labeling) 하였을 때 높은 정확도를 보임.
  (레이블링에 사용된 평가기준은 스펙을 평가하는 알고리즘.)
- 해당 평가기준에서 높은 점수를 받은 지원자의 자기소개서와 낮은 점수를 받은 지원자를 추출.
  (높지도 낮지도 않은 애매한 점수를 받은 자기소개서는 학습을 어렵게하는 노이즈로 판단하여 제거함.)
- 딥러닝 모델로 스펙을 보지 않고 자기소개서만으로 우수한 자기소개서 여부를 가려낼 수 있게 됨.

(*딥러닝이 예측한 자기소개서를 실제로 채용 담당자분들이 읽고 검증한 결과 사용 가능한 평가기준이라고 판단. → 44개 계열사 적용)

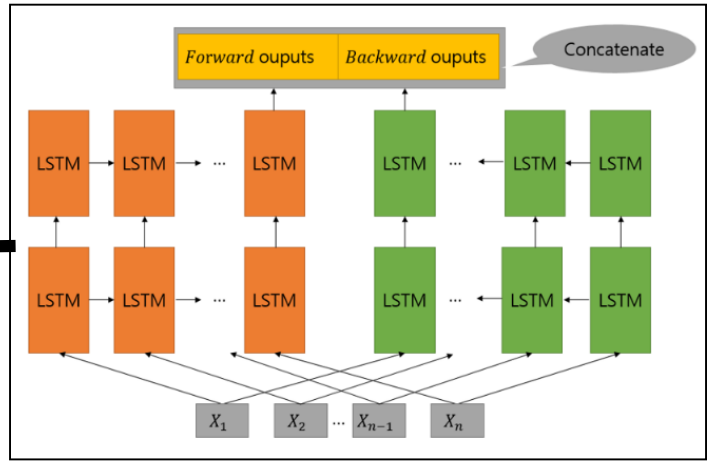
1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

- 딥러닝 모델이 5개의 답변을 수용할 수 있도록 5개의 셀(Cell)을 두었다.
- 1D-CNN은 3, 4, 5, 6의 커널(Kernel)이 존재해서 3개, 4개, 5개, 6개의 연속된 시퀀스(Sequence)의 특징을 추출한다.
- Gradient vanishing을 방지하기 위해 미분값이 상대적으로 큰 활성화 함수(Activation Functions) 하이퍼볼릭 탄젠트 함수(tanh)를 사용한다.(Xavier Initialization)
- 마지막 노드는 0~1 사이의 값을 출력하기 위해 시그모이드(sigmoid) 함수를 사용한다.
- 손실함수는 크로스 엔트로피(Cross-entropy)를 사용한다.

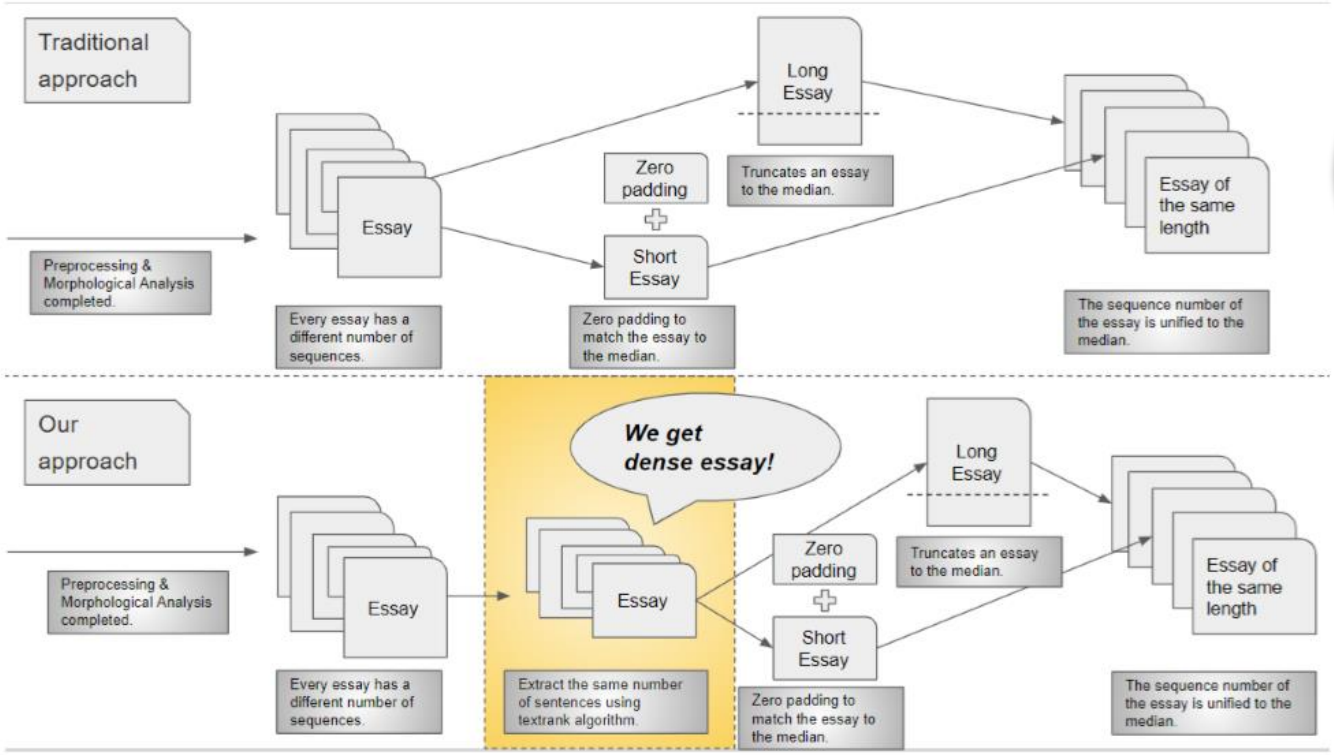


[자기소개서 분류 딥러닝 모델 예시]

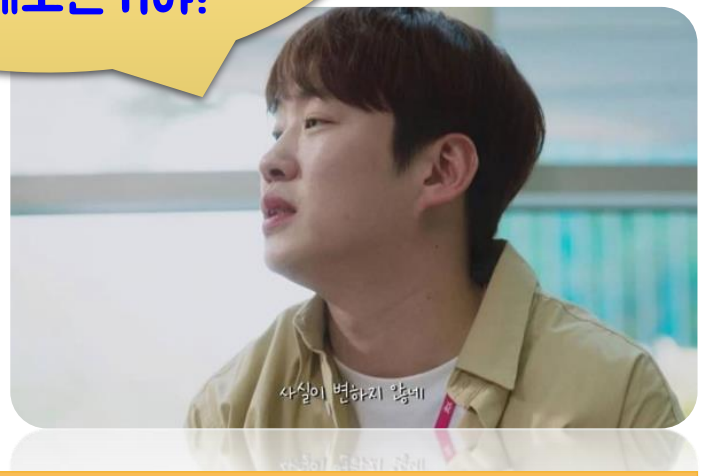


1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference



(아련하게) 어떻게 하면 정확도를 더 올릴 수 있을까...?
그래! 자기소개서를 요약해보는거야!



- TextRank를 활용하여 중요도가 낮은 문장을 제거한다.

*이는 딥러닝 모델의 정확도를 높일 수 있다. 이는 기계학습의 특징 선택(Feature selection)의 효과와 유사하다. 자연어는 한 문장 안에서 특정 형태소를 제거하면 의미가 훼손되기 때문에 전통적인 특징 선택(Feature selection) 기법을 적용하기 어렵다. 따라서 의미를 최대한 보존하면서 불필요한 노이즈(Noise)를 제거하기 위하여 중요도가 낮은 문장을 제거하는 전처리 과정을 추가했다.

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

[TextRank 활용한 자기소개서 요약 예시]

순서	TextRank 점수 순위	TextRank 점수	문장
1	2	1.45233	1 년 동안 의 교내 영어 학습 동아리 회장 으로서...
2	3	1.25027	이른 오전 매일 7 시 30 분 에 시작 하는 동아리...
3	5	1.09548	동아리 의 분위기 는 구성원 들 의 참석 률...
4	1	1.49881	그래서 구성원 들 의 자발 적 인 참여 를...
5	8	0.93056	이 목표 를 이루기 위해 릴레이 모닝콜...
6	9	0.91456	학교 와 집 과의 거리 가 먼 사람 부터 가까운...
7	13	0.63766	그 결과 출석 률 을 높이는데 큰 도움 이 되었습니다.
8	11	0.84565	또한 사소한 대화 로 시작 하여 작은 소 모임 을 결성...
9	10	0.86788	그래서 신뢰 를 바탕 으로 자발 적 참여 를 유도 하여...
10	14	0.62134	이러한 노력 끝 에 수동 적 참여 가 아닌 스스로 흥미...
11	12	0.67191	그래서 한 학기 동안 낙오 자 가 없고 다른 어떤...
12	6	1.01100	4 년 이 지난 지금 도 구성원 들 과 그 당시 추억...
13	7	0.98664	이처럼 구성원 들 과의 무한 한 신뢰 온정 과 수용...
14	4	1.21590	앞 으로 롯데 정보 통신 에서 구성원 들 을 아우르는...

1 년 동안의 교내 영어학습동아리 회장으로서 구성원들과 소통하고 협력하여 능동적 참여를 이끌어 낸 경험이 있습니다. 이른 오전 매일 7 시 30 분에 시작하는 동아리이기 때문에 구성원들의 늦잠으로 인해 출석률이 낮아지는 것에 대한 고민을 했습니다. 동아리의 분위기는 구성원들의 참석률에 의해 결정된다고 생각했습니다. 그래서 구성원들의 자발적인 참여를 이끌어내기 위해 80% 이상의 출석률을 목표로 설정했습니다. 이 목표를 이루기 위해 릴레이 모닝콜을 제안했습니다. 학교와 집과의 거리가 먼 사람부터 가까운 사람에게 한명씩 모닝콜을 하는 시스템입니다. 그 결과 출석률을 높이는데 큰 도움이 되었습니다. 또한 사소한 대화로 시작하여 작은 소모임을 결성하여 친밀감을 형성을 이루려고 노력했습니다. 그래서 신뢰를 바탕으로 자발적 참여를 유도하여 화기애애한 분위기로 바꿀 수 있었습니다. ~~이러한 노력 끝에 수동적 참여가 아닌 스스로 흥미를 가진 능동적 참여를 이끌어 냈습니다.~~ 그래서 한 학기동안 낙오자가 없고 다른 어떤 동아리보다도 결속력 있는 영어학습동아리를 만들 수 있었습니다.

*TextRank 점수가 가장 낮은 문장이 요약된다.

$$similarity(S_1, S_2) = \frac{|\{m_k | m_k \in S_1 \ \& \ m_k \in S_2\}|}{\log(|S_1|) + \log(|S_2|)}$$

$$\mathrm{TR}(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{v_k \in Out(V_j)} w_{jk}} TR(V_j)$$

***참고 논문**

[1] Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into text. In Proceedings of the 2004 conference on empirical methods in natural language processing

[2] Federico Barrios, Federico López, Luis Argerich, and Rosa Wachenchauzer. (2016). Variations of the similarity function of TextRank for automated summarization. arXiv:1602.03606[cs.CL]

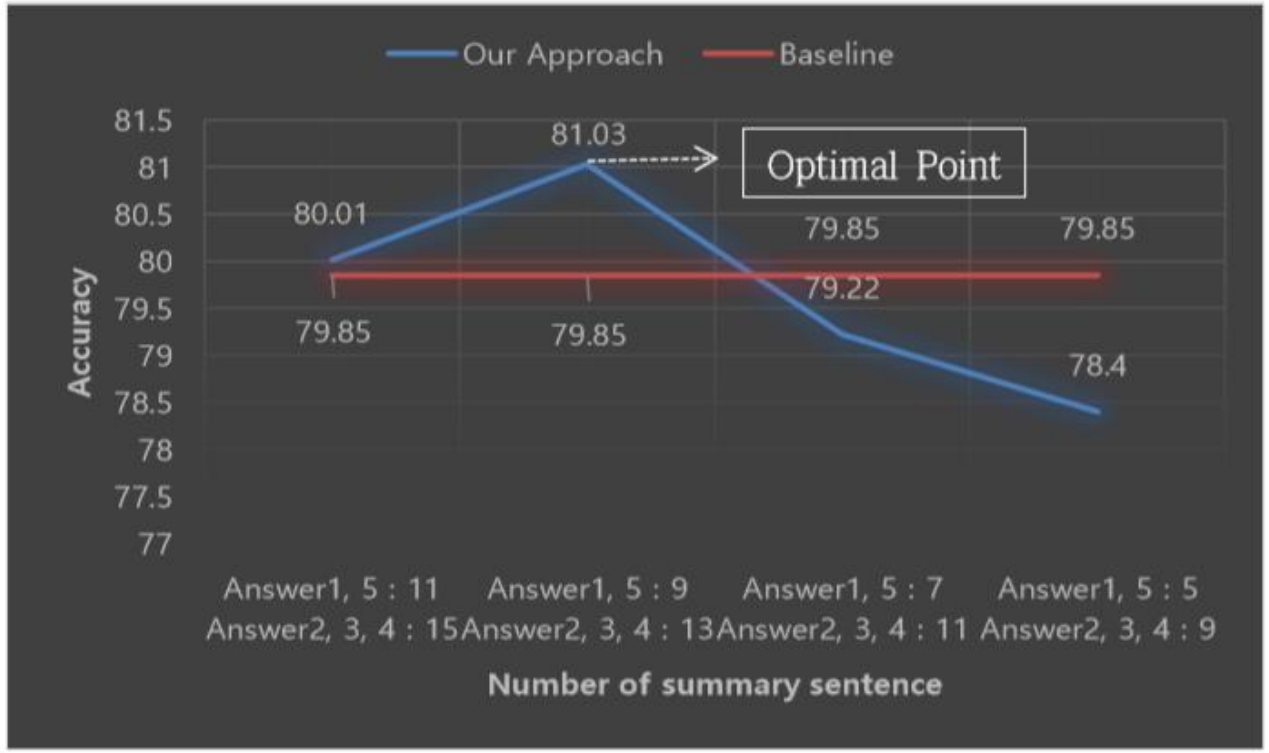
1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

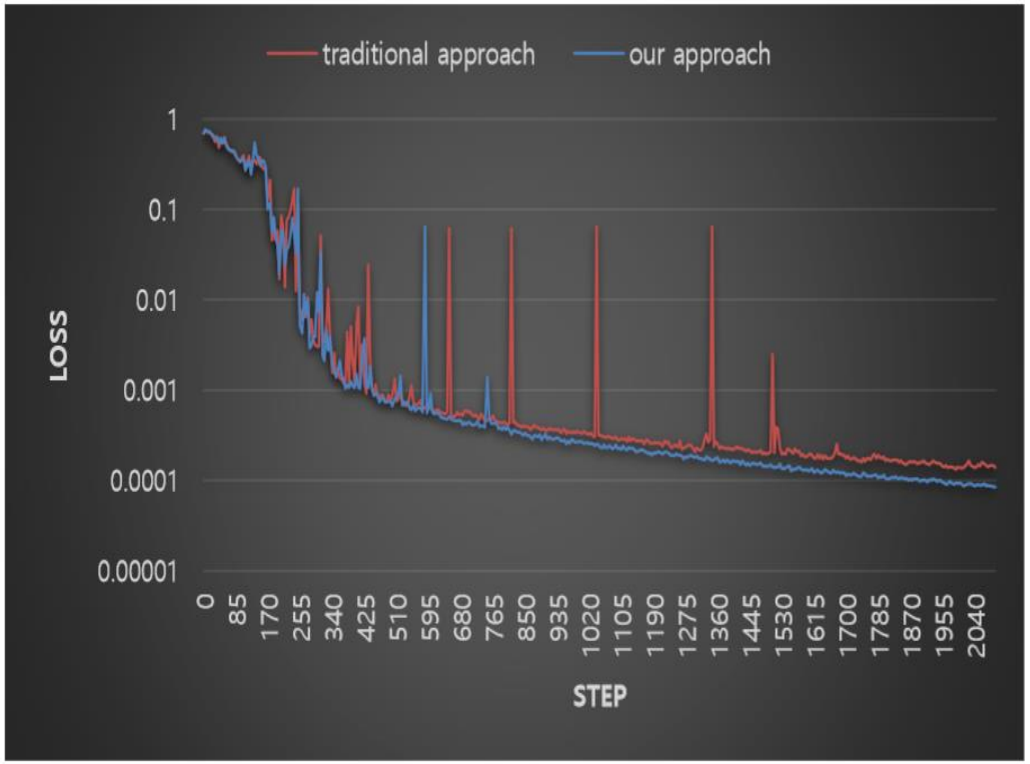
- **(정확도 향상)** 제한글자가 500자인 자기소개서는 문장 수는 9 문장, 제한글자가 800자인 답변은 평균 문장 수는 13 문장일때 성능이 가장 좋았습니다. 이 이상 요약할 때 정확도가 기존 방법보다 떨어지는 것으로 미루어 합격 및 불합격을 결정짓는 중요한 문장마저 지워진다고 할 수 있다.

- **(학습 수렴 속도 향상)** 요약된 자기소개서를 학습하는 딥러닝 모델은 1,865Step 부터 Loss 가 0.0001 이하로 떨어지지만, 자기소개서 원문을 학습하는 딥러닝 모델은 Loss 가 에폭(Epoch)을 모두 수행해도 0.0001 이하로 떨어지지 않았다.

[핵심문장 수별 딥러닝 모델 정확도 변화]



[학습 스텝별 손실변화 추이]



1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

- Train data : 2016년 ~ 2018년 상반기 자기소개서 / Test data : 2018년 하반기 자기소개서
- Ours의 F1-Score 는 각각 0.8, 0.82 로 합격 및 불합격 중 한쪽으로 치우치지 않고 균형 잡힌 정확도를 보였다.
- 직무를 구분했을때 오히려 정확도가 내려갔다. 원인은 첫째, 직무를 구분하면서 학습 데이터 수가 줄었다. 둘째, 직무를 구분하면서 학습 데이터(Train Data)의 합격 및 불합격의 비율이 1:1 이 아니게 되었다. 셋째, 경력이 아닌 신입사원이기 때문에 직무 특성이 자기소개서에 크게 드러나지 않는다.
- 자기소개서는 데이터가 부족한 도메인(Domain)인 만큼 딥러닝의 특성상 직무별로 데이터가 충분하지 않다면 직무를 구분하지 않는 것이 오히려 좋은 성능을 낼 수 있다.

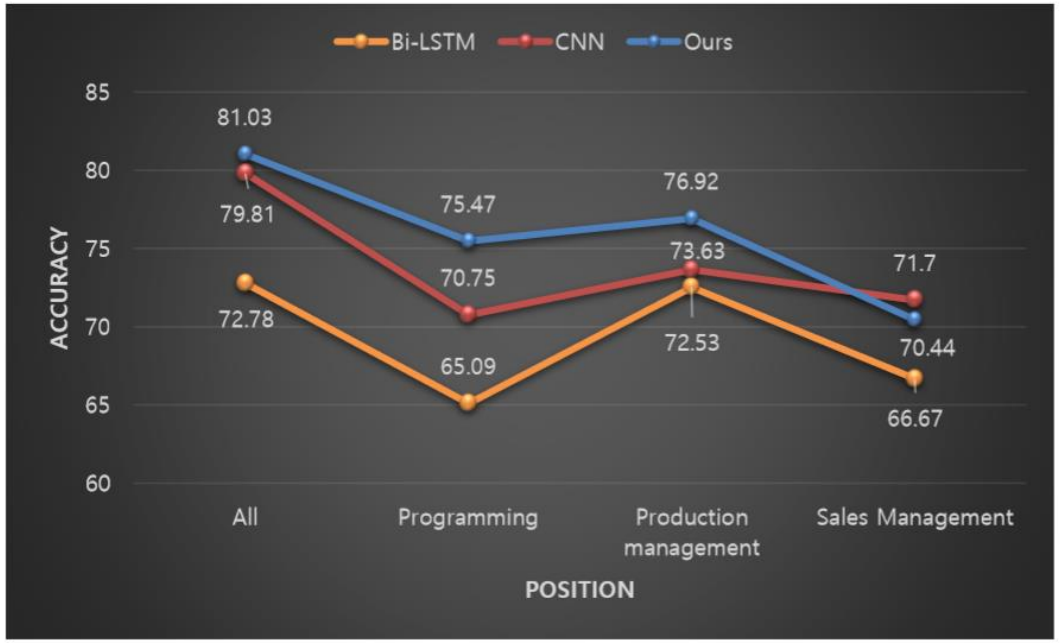
에잇! 이놈도 아녀! 최고의 성능을 내는 놈만 살아남는겨!

Table 4: Comparison of Performances

Position	Model	Precision		Recall		F1-score		Accuracy
		Fail(0)	Pass(1)	Fail(0)	Pass(1)	Fail(0)	Pass(1)	
All	Bi-LSTM	.69	.77	.76	.70	.72	.73	72.78%
	1D-CNN	.72	.88	.87	.75	.79	.81	79.81%
	Ours	.74	.88	.87	.76	**.80**	**.82**	**81.03%**
Programming	Bi-LSTM	.69	.77	.76	.70	.72	.73	72.78%
	1D-CNN	.72	.88	.87	.75	.79	.81	79.81%
	Ours	.74	.88	.87	.76	**.80**	**.82**	**81.03%**
Production Management	Bi-LSTM	.69	.77	.76	.70	.72	.73	72.78%
	1D-CNN	.72	.88	.87	.75	.79	.81	79.81%
	Ours	.74	.88	.87	.76	**.80**	**.82**	**81.03%**
Others	Bi-LSTM	.69	.77	.76	.70	.72	.73	72.78%
	1D-CNN	.72	.88	.87	.75	.79	.81	79.81%
	Ours	.74	.88	.87	.76	**.80**	**.82**	**81.03%**

Table 2: Dataset Distribution for Training and Test

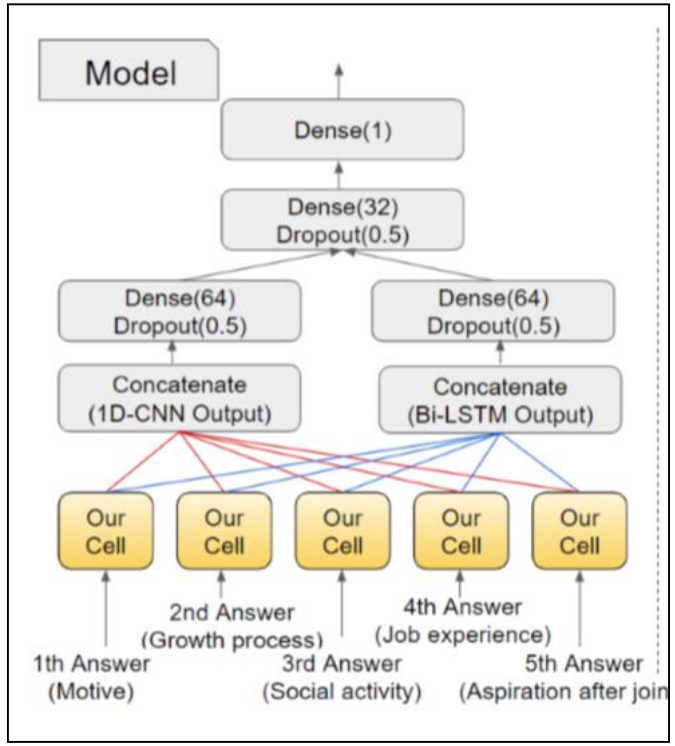
Job Description	Training dataset		Test dataset	
	Failed(0)	Passed(1)	Failed(0)	Passed(1)
Sales Management	852(41.6%)	1,195(58.4%)	93(58.5%)	66(41.5%)
Production Management	583(62.3%)	353(37.7%)	66(72.5%)	25(27.5%)
Programming	237(45.1%)	288(54.9%)	24(22.6%)	82(77.4%)
Others	3,670(51.1%)	3,506(48.9%)	1,138(52%)	1,052(48%)
Total	5,342(50%)	5,342(50%)	1,321(51.9%)	1,225(48.1%)

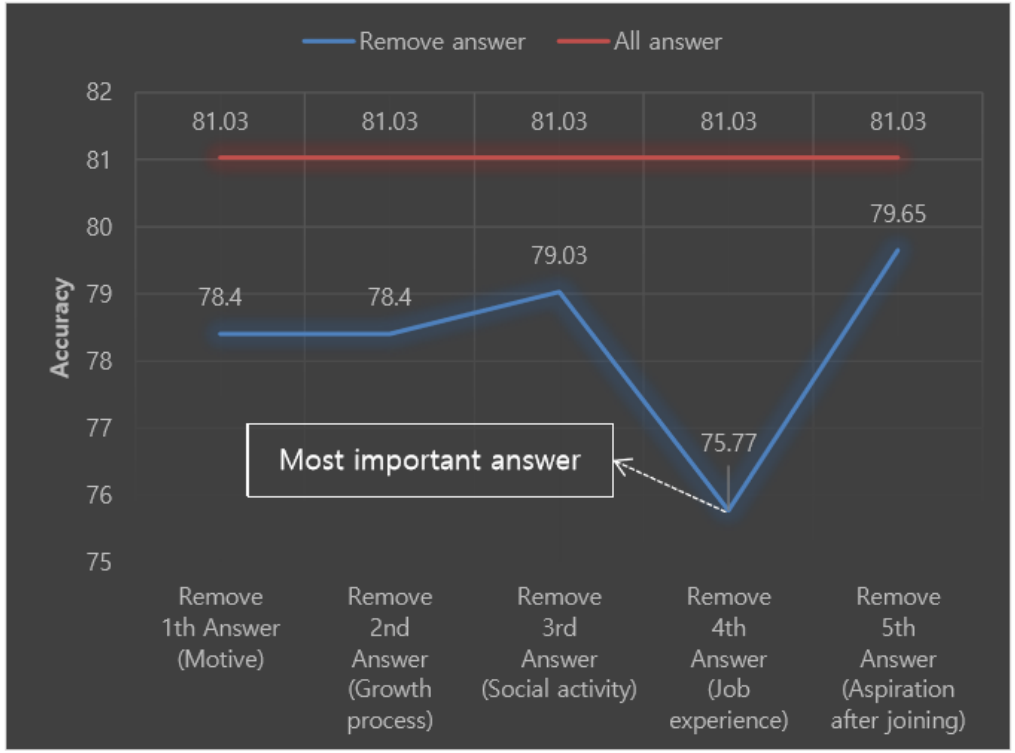


1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

- 다음 딥러닝 모델의 Cell은 각각의 답변이 입력된다.
  특정 Cell을 제거하면 해당 답변이 입력되지 않은 딥러닝 모델을 얻을 수 있다.
- 지원동기 / 성장과정 / 사회활동 / 직무경험 / 10년계획 답변 중에서 가장 중요한 질문은 '직무경험' 이었다.



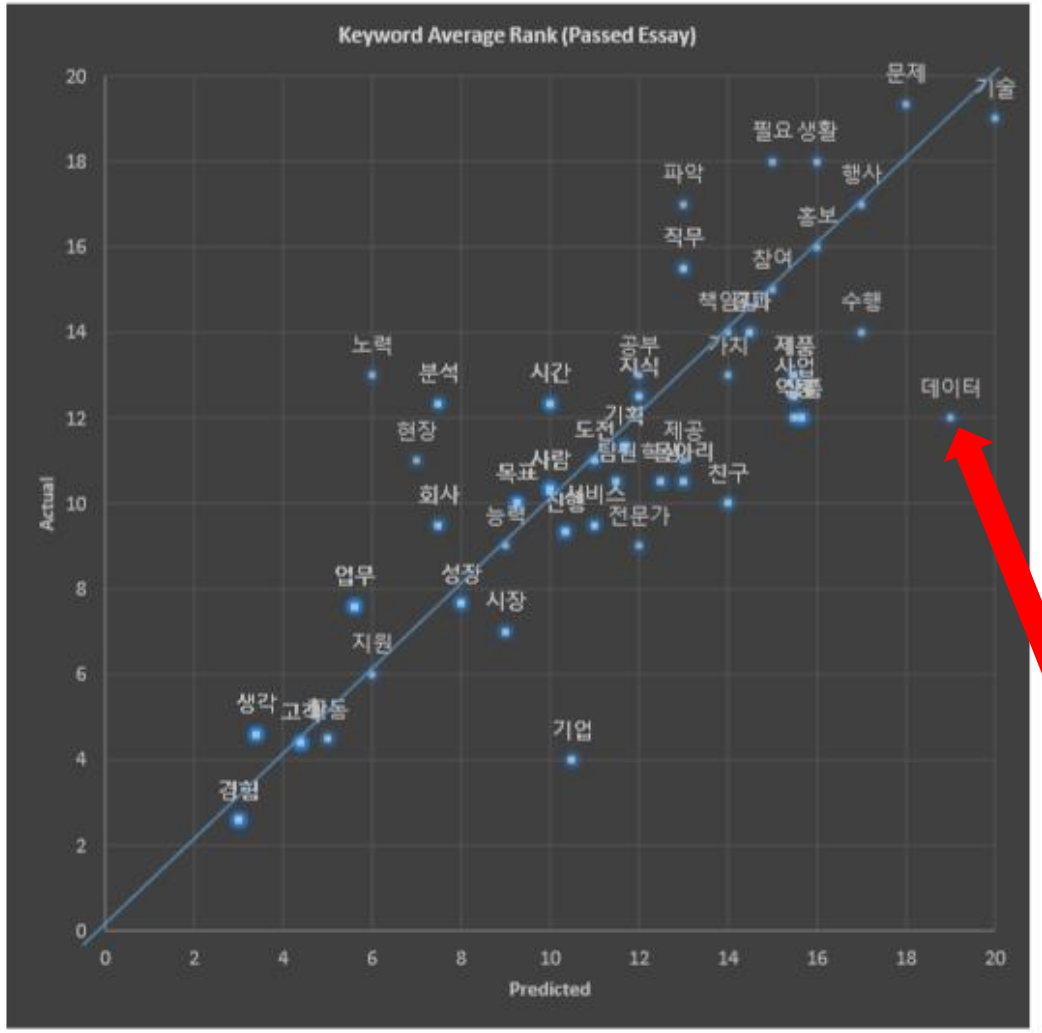


뭣이 중헌디!
어떤 질문이
중허냐고!

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

합격 자기소개서 키워드 비교



- 자기소개서를 Komoran 형태소 분석기로 명사 추출 -> -> 1글자 제거 -> TextRank로 키워드 추출 -> 추출된 키워드 빈도수로 순위 부여 ->

딥러닝의 예측 순위 / 실제 순위를 비교한 결과를 2차원으로 표현

- **정확도가 100% 딥러닝 모델은 키워드가 모두 $y = x$ 직선에 위에 놓인다.**
- $y = x$ 직선에서 멀어질수록 실제와 예측간 순위 차이가 크다.
- $y = x$ 직선보다 아래쪽에 있는 키워드는 실제 순위가 예측 순위보다 높다.
- [데이터] 키워드는 채용 담당자는 합격으로 많이 분류했지만, 딥러닝 모델은 그보다 적게 분류했음을 알 수 있다.

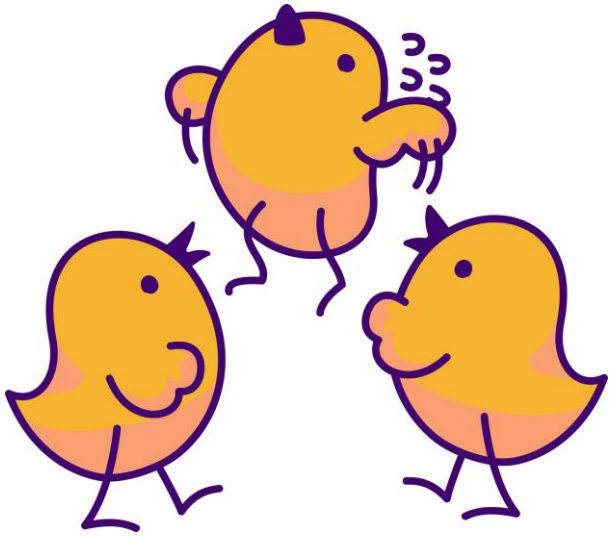
1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

Language Conference

- 혼동행렬을 기준으로 각 영역별 키워드 표현
- 많이 등장한 키워드일수록 크고, 중앙에 표현
- [생각], [업무] 등 무난한 의미를 지닌 단어들은 **불용어** 처리
- 합격자기소개서에 [동아리], [사업] 등의 키워드가 상대적으로 더 많았다.

특정 단어를 쓴다고
혹은 안 쓴다고
합격하거나
불합격하지 않습니다!
오해없기!

	Pass(Predicted)	Fail(Predicted)
Pass (Actual)	True Positive 	False Negative 
Fail (Actual)	False Positive 	True Negative 



2. 자기소개서의 어색한 문장을 찾을 수 있을까?

기술 키워드 : 딥러닝, 유사도, MaLSTM, Transformer

LangCon 2020

2. 자기소개서의 어색한 문장을 찾을 수 있을까?

Language Conference

- **Business 요건 : 자기소개서에서 어색한 문장을 찾아주세요.**
- 외부 업체로부터 들어온 요건
- 해당 모델로 자기소개서 중간에 '엄마가 밥 먹으래.' 라는 문장이 실제로 검출됨.
- 시연영상 : https://www.youtube.com/watch?v=klUiScycydc

<어색한 문장 검출 예시>

2. 자기소개서의 어색한 문장을 찾을 수 있을까?

Language Conference

- **Data 구성**

- 우수한 자기소개서를 문장별로 분리하여 연속된 문장을 Pairing

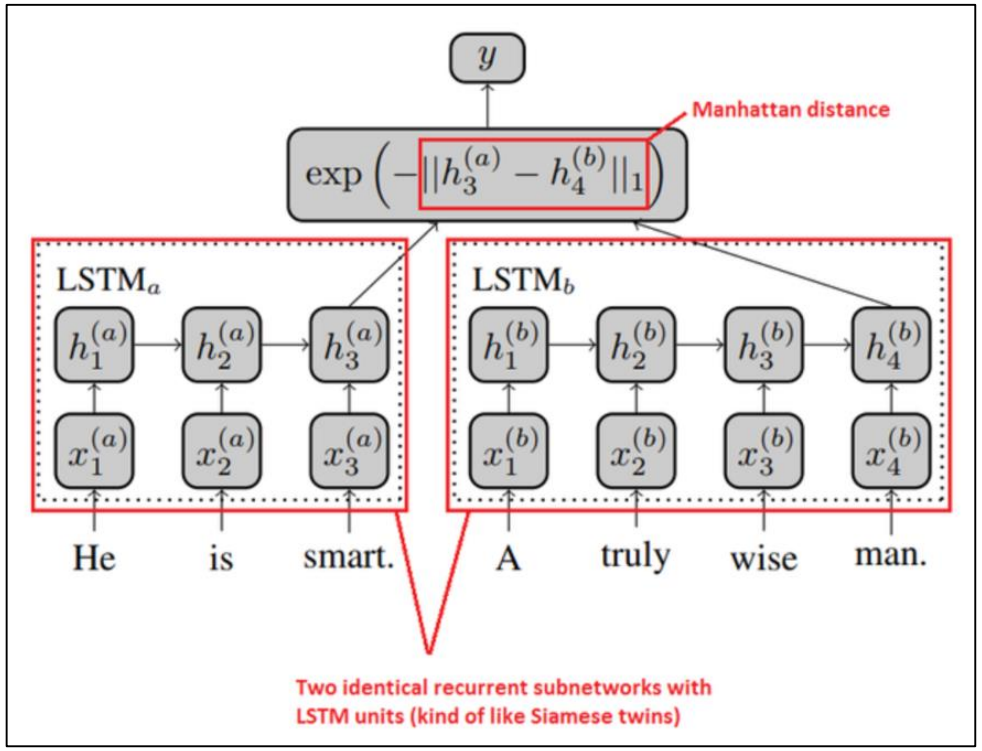
Sentence 1	Sentence 2	Label	
'가르침의 경험' 대학 시절, 영어 학원에서 일한 …	강사로서 학생들에게 줄 수 있는 최고의 선물은…	자연스러움	50%
강사로서 학생들에게 줄 수 있는 최고의 선물은…	저는 학원 모의고사에서 학생들의 성적을 지난…	자연스러움	
.	.		
.	.		
.	.		
'가르침의 경험' 대학 시절, 영어 학원에서 일한 …	국토대장정을 하며… (Resume random sentence)	어색함	50% / 33%
.	.		
강사로서 학생들에게 줄 수 있는 최고의 선물은…	오늘 저녁 메뉴 뭐야? (chatbot question data)	어색함	33%
.	.		
저는 학원 모의고사에서 학생들의 성적을 지난…	오늘 하루도 수고 많았어요.(chatbot answer data)	어색함	33%

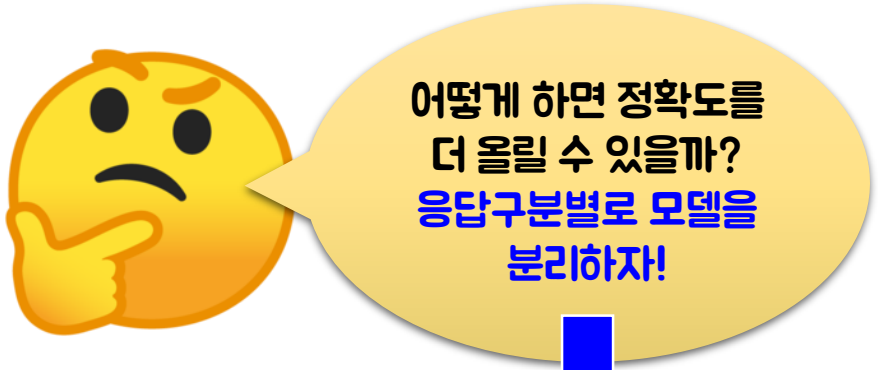
2. 자기소개서의 어색한 문장을 찾을 수 있을까?

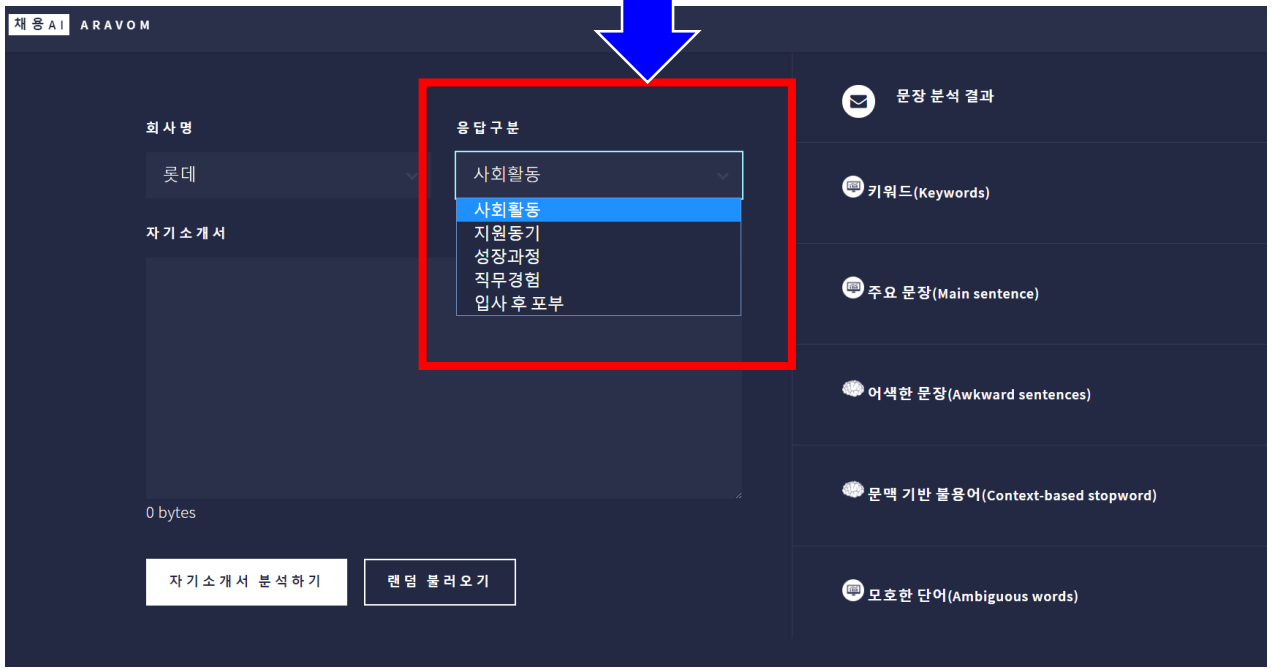
Language Conference

- **BaseModel : [3]MaLSTM 딥러닝 모델**

- Test Data 정확도 : **76%**







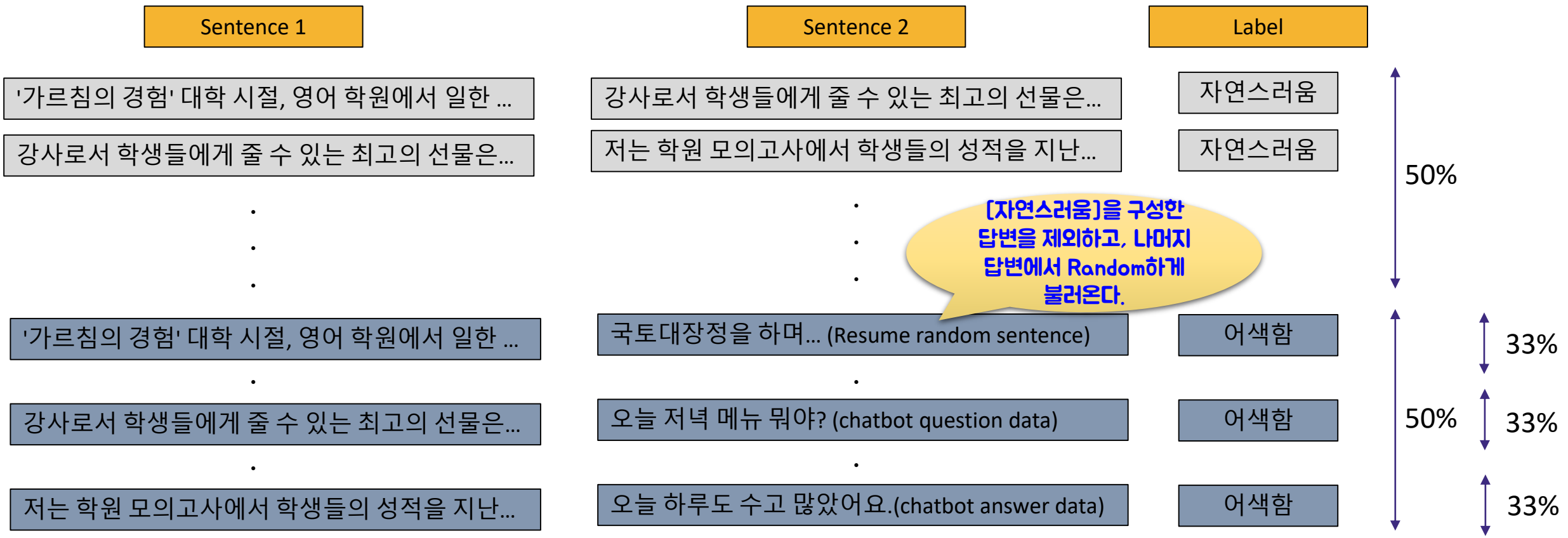
***참고 논문**

[3] Siamese Recurrent Architectures for Learning Sentence Similarity

2. 자기소개서의 어색한 문장을 찾을 수 있을까?

Language Conference

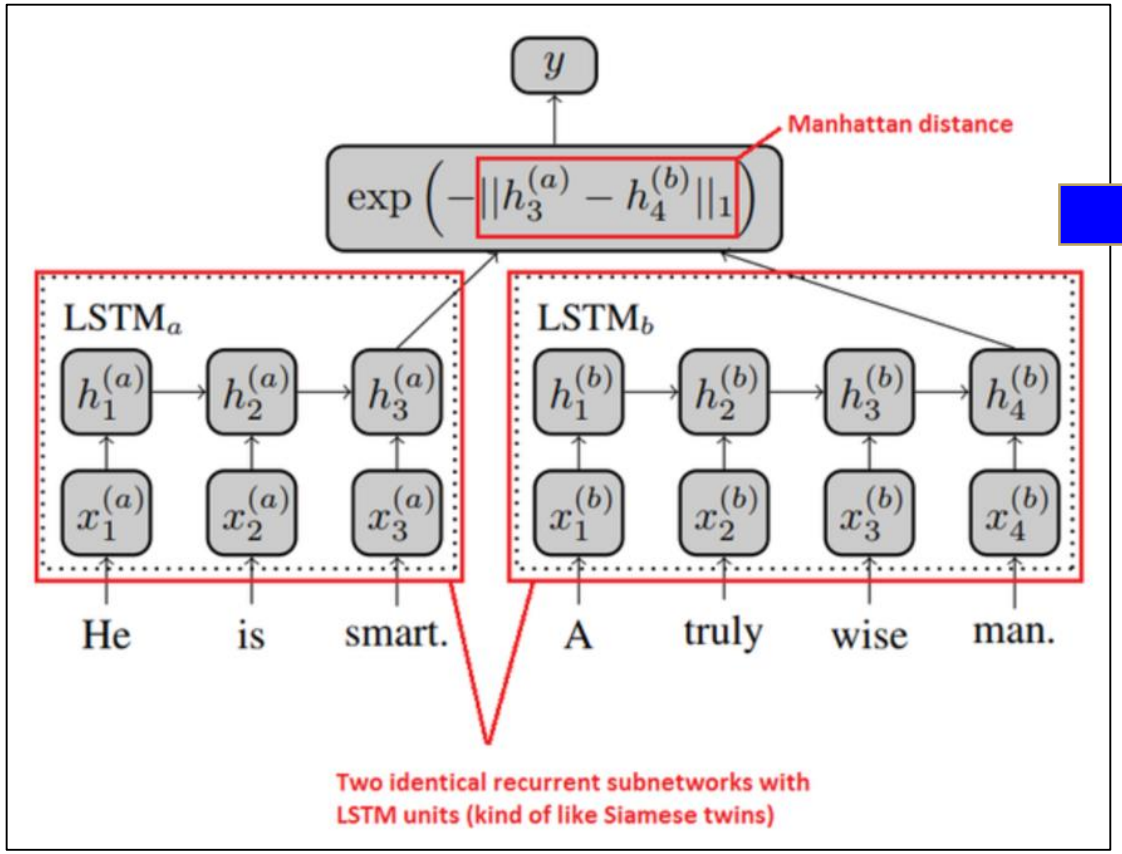
- 예를 들어 '사회활동' 모델의 데이터 구성 시, 해당 답변의 자기소개서로만 [자연스러움] 데이터를 구성한다.
- 더불어 Resume random sentence는 '사회활동'을 제외한 지원동기, 성장과정, 직무경험 등 에서 랜덤하게 추출한다.
- 이를 통해 딥러닝 모델이 해당 답변의 의미에 집중할 수 있게 된다. (정확도 : 76% -> 84%(8% 향상))

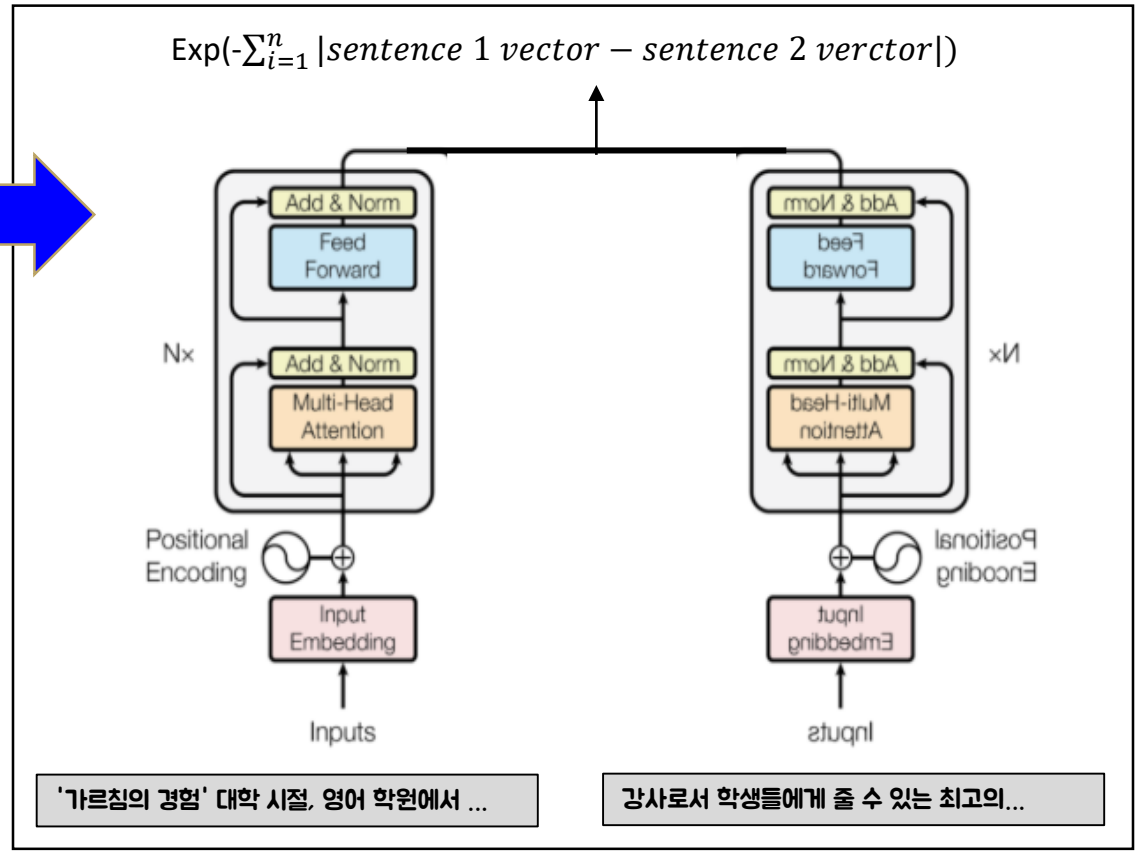


2. 트랜스포머를 활용한 자기소개서 문장 생성

Language Conference

- [4] Transformer Encoder를 활용한 딥러닝 모델로 고도화
  - 정확도 : 74% -> 84%-> 88%(4%향상)





***참고 논문**

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000–6010.

2. 트랜스포머를 활용한 자기소개서 문장 생성

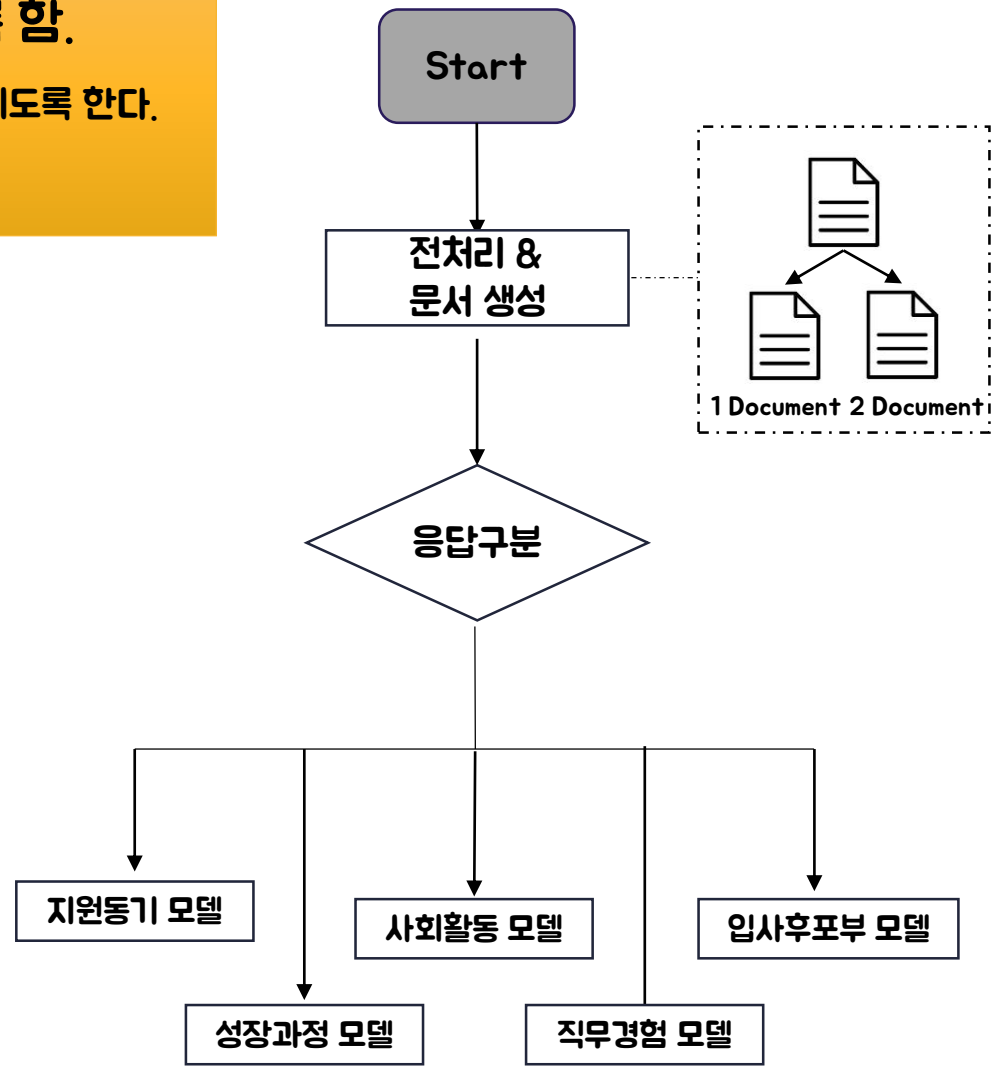
Language Conference

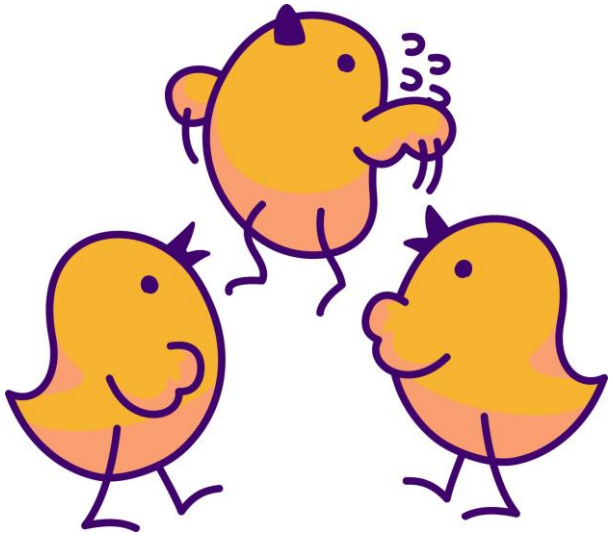
- **딥러닝 추론 속도 향상을 위해 문장이 아닌 문서 단위 배치로 추론 되도록 함.**

- [마지막 문장이 제거된 문서]와 [첫번째 문장이 제거된 문서]를 만들어 연속된 문장이 Paring 되도록 한다.

- 자기소개서 1건 기준 추론 속도 32%향상

1 Document (revome last sentence)	2 Document (remove first sentence)
1. 동아리 운영진으로 MT 등 행사를 기획하고 회의를 진행하기 위해 노력했던 순간이 기억에 남습니다.	**2.** 3년 반 동안 미디어, 국문 등 서로 다른 전공의 친구들과 교류하며 애정을 가지고 활동했고 매년 80명 내외 규모에서 회원, 운영진, 회장까지 활동했습니다.
2. 3년 반 동안 미디어, 국문 등 서로 다른 전공의 친구들과 교류하며 애정을 가지고 활동했고 매년 80명 내외 규모에서 회원, 운영진, 회장까지 활동했습니다.	**3.** 시간 안에 끝나는 회의를 위해 안건 순서를 고민하고, 매주 회의 안건을 준비하며 전체 목표를 두고 단계별로 계획하는 습관이 생겼습니다.
3. 시간 안에 끝나는 회의를 위해 안건 순서를 고민하고, 매주 회의 안건을 준비하며 전체 목표를 두고 단계별로 계획하는 습관이 생겼습니다.	**4.** 행사 때 5분 단위 진행표를 역할을 나눠 점검하며 기간 내에 완수하는 책임감을 얻게 됐습니다.
4. 행사 때 5분 단위 진행표를 역할을 나눠 점검하며 기간 내에 완수하는 책임감을 얻게 됐습니다. ... (이하 생략)	**5.** 행사 이후엔 피드백을 받아 반영하며, 동아리 행사에 참여했을 때 회원들이 좋아하길 바라는 마음으로 임했습니다. ... (이하 생략)



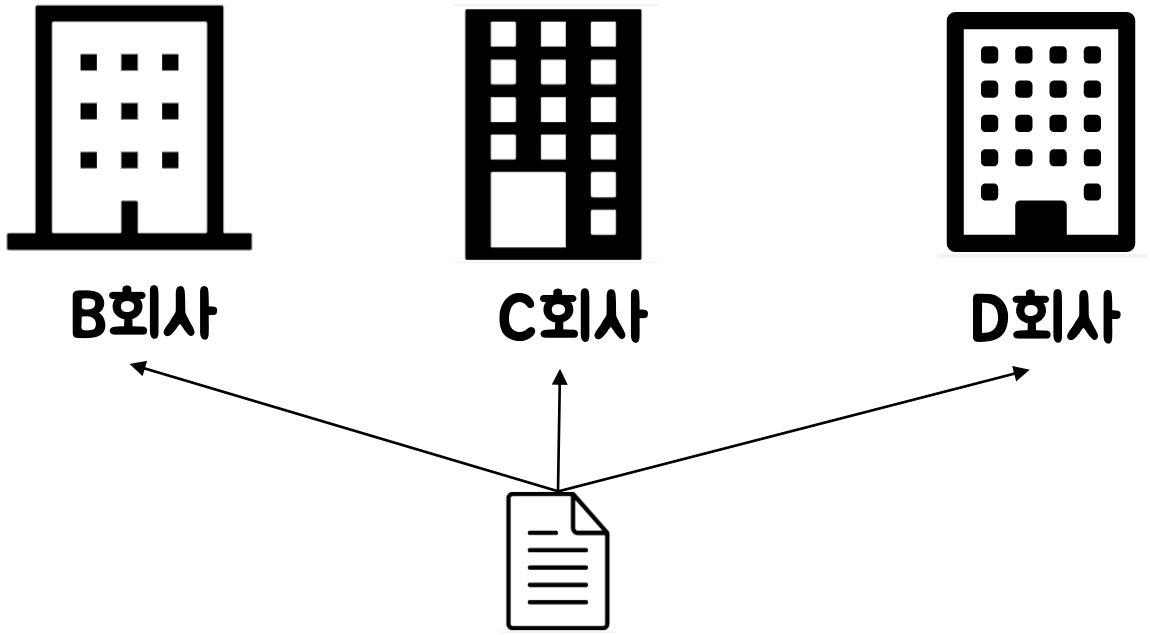


3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

기술 키워드 : 딥러닝, BERT

3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

Language Conference



A회사에 제출한 자기소개서

대세는
원소스멀티유즈
(One source Multi use)
아니겠어!

...(생략)
이와같은 저의 열정으로 A회사에서 최선을 다하겠습니다.

[A사에 제출한 자기소개서를 B사에 그대로 제출한 Case]

저는 A회사에서 웹 프로그래머로써 2개월간 인턴으로 일했습니다.
(생략)...

[A사에 인턴으로 일했던 경험을 B사에 제출한 Case]

3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

Language Conference

- 본 세션에서는 딥러닝을 활용하여 타 회사에 지원한 자기소개서를 그대로 복사, 붙여넣기 했을때 어떻게 찾아낼 수 있는지 소개해드리겠습니다.(표절에 대한 기술은 제외)
- [시연영상 : https://www.youtube.com/watch?v=klUiScycydc](https://www.youtube.com/watch?v=klUiScycydc)

<롯데에 지원한 자기소개서>

<삼성에 지원한 자기소개서>

3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

Language Conference

데이터를 어떻게 만들지?
손으로 한땀한땀 적어야될까?

- 단순히 자기소개서에 타회사명이 언급된다고 문제가 되지 않습니다. 인턴, 공모전 등 타사에서 경험한 사례가 자기소개서에 등장할 수 있기 때문입니다.

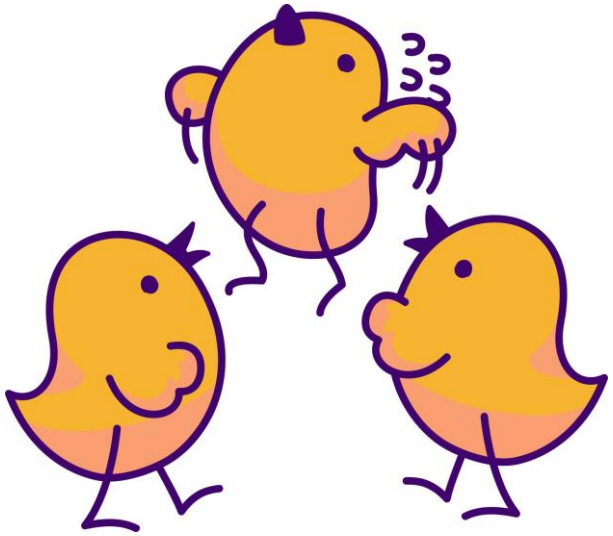
- 그래서 아래와 같은 방법을 고안했습니다!

1) 회사명이 등장한 모든 문장 수집!(회사명 사전 및 정규표현식 활용)
2) 회사명을 학습하지 않게 하기 위해서 문장에서 회사명 제거
(자사뿐만 아니라 다양한 회사에서 사용할 수 있도록 하기 위함.)
3) 반드시 지원하는 회사명이 있어야 하는 문장인지 혹은
타사명이 있어도 괜찮은 문장인지 레이블링
4) *basemodel [5]1D - CNN 모델 학습(정확도 : 88%)
5) BERT의 Finetunning 하여 transfer learning 한 경우(정확도 : 94%)
6) 다음 두 조건을 동시에 만족하면 문장이 있다면 복붙한 자기소개서로 판별
* (타사명이 존재) AND (반드시 지원하는 회사명이 와야하는 문장)
⇒ 실제 타사 PoC 결과 100% 정확도 보였음.

TRAIN DATA 예시>
에서 열심히 일하겠습니다.(0 - 반드시 지원하는 회사명이 와야함)
에서 주최한 공모전에서 최우수상을 수상했습니다.(1 - 타 회사명이 와도 괜찮음)

*참고 논문

[5] Y. Kim. (2014). Convolutional neural networks for sentence classification. In Proceedings of EMNLP.



4. 오타 어떻게 찾을까?

기술 키워드 : Levenshtein distance 알고리즘

4. 오타 어떻게 찾을까?

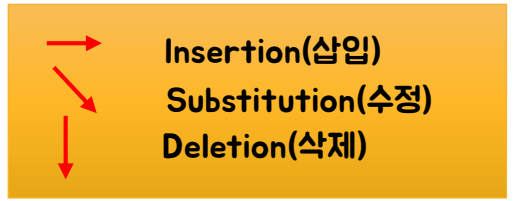
Language Conference

- Insertion(삽입), Deletion(삭제), Substitution(수정)은 비용을 계산

Ex) [사랑, 사랑해] 는 한 글자가 삽입되어거리 1
[나무, 나]는 한 글자가 삭제되어 거리 1
[자연어, 자현언]는 연의 초성 수정(0.3). 어의 종성 수정(0.3)되어 거리 0.6

- 그럼다면 [자연어처리짱]과 [자여어철리굿]사이의 거리는?

-> [1,1]은 [자, 자] 같기 때문에 0으로 초기화한다.(다르면 1)
-> [1,j]는 j가 1씩 증가함에 따라 한 글자씩 추가되었기 때문에 1씩 더한다.
-> [2,1]은 [자연, 자]가 한 글자가 추가되었기 때문에 1이다.
-> [2,2]은 [자연, 자여]의 종성이 다르기 때문에 0.334(약0.3)이다.
-> [2,3]은 [자연, 자여어]는 [여]와 [연]의 종성이 다르고, 한 글자가 추가되었기 때문에 1.3이다.
-> 위와 같은 방식으로 매트릭스를 계산한다.
-> [6,6]의 값 1.6이 [자연어처리짱]와 [자여어철리굿]의 거리이다.

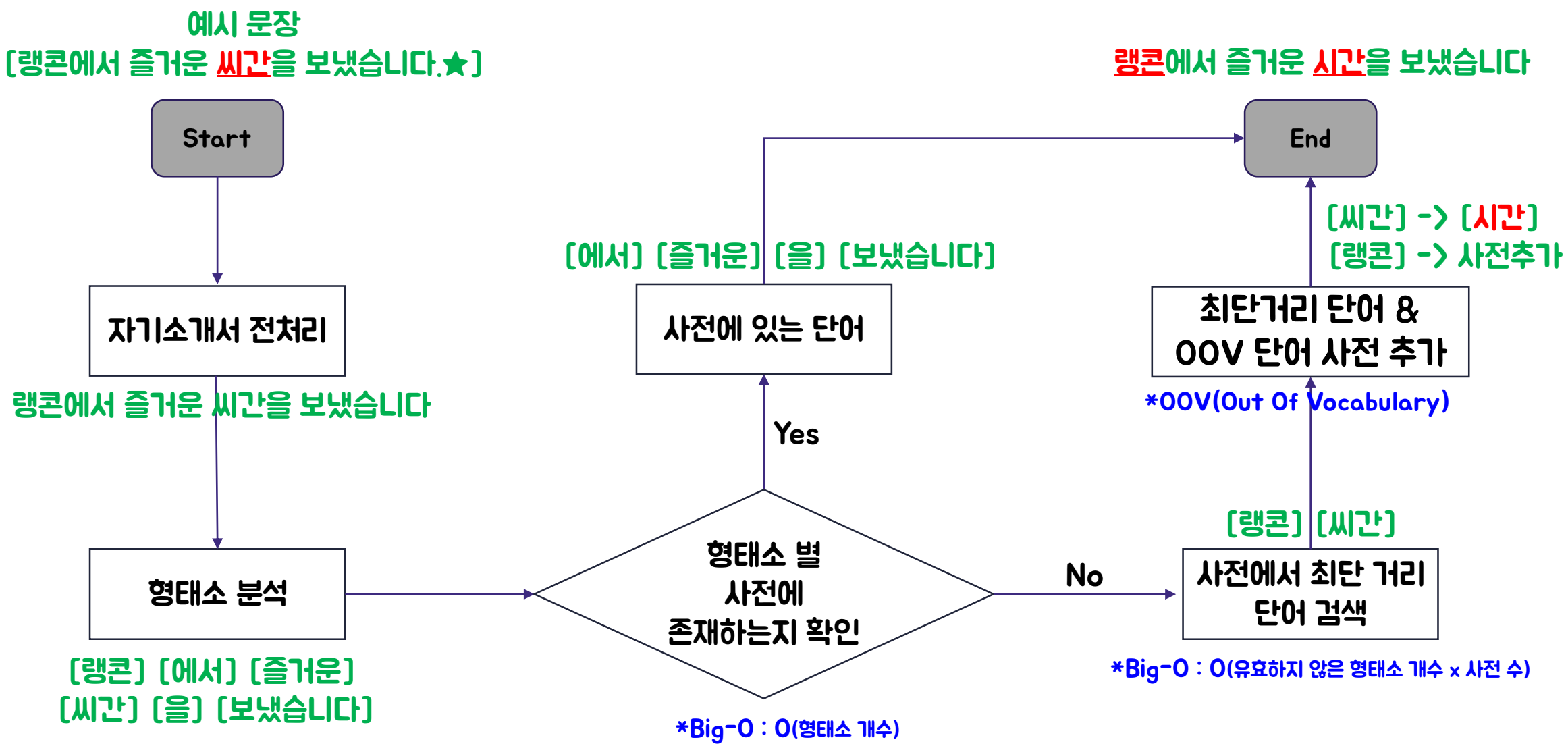


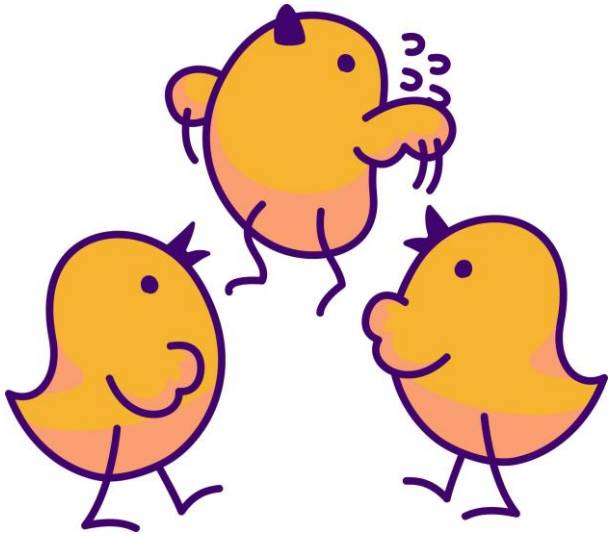
		자	여	어	철	리	굿
	0	→ 1	→ 2	→ 3	→ 4	→ 5
자	1	0.3	1.3	2.3	3.3	4.3
연	2	1.3	0.3	1.3	2.3	3.3
어	3	2.3	1.3	0.6	1.6	2.6
처	4	3.3	2.3	1.6	0.6	1.6
리	5	4.3	3.3	2.6	1.6	1.6
짱						

*참고
(2018) https://lovit.github.io/nlp/2018/08/28/levenshtein_hangle/

4. 오타 어떻게 찾을까?

- **오타 검정 순서도**(simple version)





5. 롯데그룹 채용 안내

5. 롯데그룹 채용 안내

Language Conference

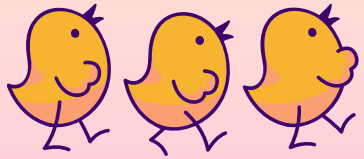
- **신입 채용**
  - 2020년 하반기 공채 예정
  - 스펙태클 채용 예정

: SPEC태클(스펙태클) 채용은 2015년에 시작된 롯데 고유의 블라인드 채용 전형입니다. 학벌이나 스펙 중심의 서류 전형에서 벗어나 지원자 직무 수행 능력과 역량 중심으로 인재를 선발합니다.
화려한 볼거리(Spectacle)처럼 다양한 지원자의 이력에 집중하고, 무분별한 스펙 쌓기에 태클을 건다(spec-tackle)는 중의적인 의미를 담고 있습니다.

- **경력 채용**
  - 롯데 채용 사이트내 수시 공고

***롯데그룹 채용 URL**

http://job.lotte.co.kr/



LangCon 2020

QnA
감사합니다.
☺

김영준

datajuny@gmail.com