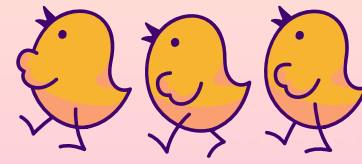


LangCon 2020



롯데의 딥러닝 모델은 어떻게 자기소개서를 읽고 있을까?

롯데정보통신
정보기술연구소 AI기술팀
김영준

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

기술 키워드 : 자기소개서 전처리, 딥러닝, TextRank 알고리즘

2. 트랜스포머를 활용한 자기소개서 문장 생성

기술 키워드 : 딥러닝, Transformer, NER, TF-IDF 알고리즘

3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

기술 키워드 : 딥러닝, CNN, NER

4. 오타 어떻게 찾을까?

기술 키워드 : Levenshtein distance 알고리즘

5. 롯데그룹 채용 안내

자기소개서
어디까지
다뤄봤니?



김영준

롯데정보통신

정보기술연구소 AI기술팀 자연어 담당

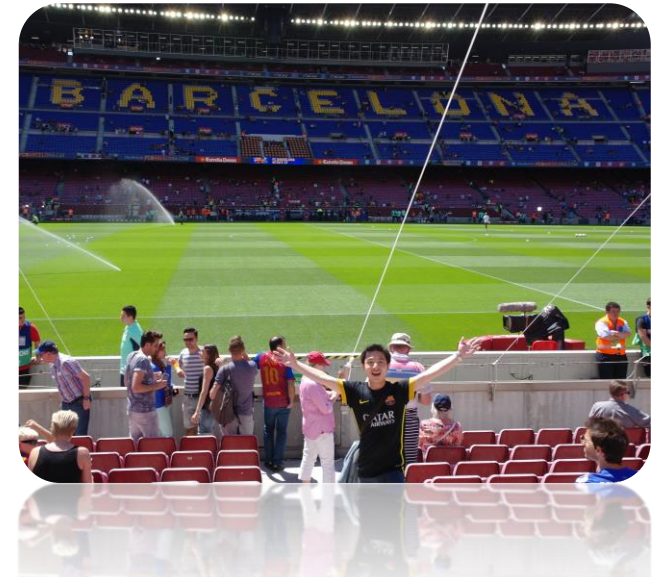
성균관대학교

데이터사이언스융합학과 석사

- 딥러닝 및 자연어 관련 논문 최우수연구상
- 성적우수 졸업

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

datajuny@gmail.com



Smart Office 

채용 AI 시스템

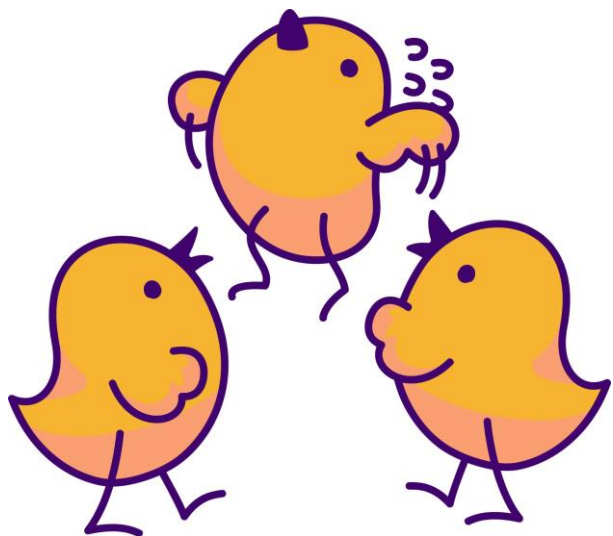
롯데그룹의
[채용AI]는
YTN, 전자신문 등
각종매체에서 소개된
AI 시스템 입니다.



- ✓ 자기소개서는 여전히 사람이 읽고 있습니다!
- ✓ 저는 [채용AI] 기술 담당자이지 채용 담당자가 아닙니다!
Q. '어떻게 하면 딥러닝이 합격으로 분류하나요?'
A. '저도 궁금해요!'
- ✓ 아쉽게도 실제 사용되는 딥러닝 모델은 공개할 수 없어요.
대신 다른 모델을 준비해왔어요!
- ✓ 본 세션은 딥러닝을 이미 알고 계신분이 실무에
어떻게 활용되는지 궁금하신 분이 들으면 좋아요!

딥러닝은 거들뿐!
서운해 하지 말자!

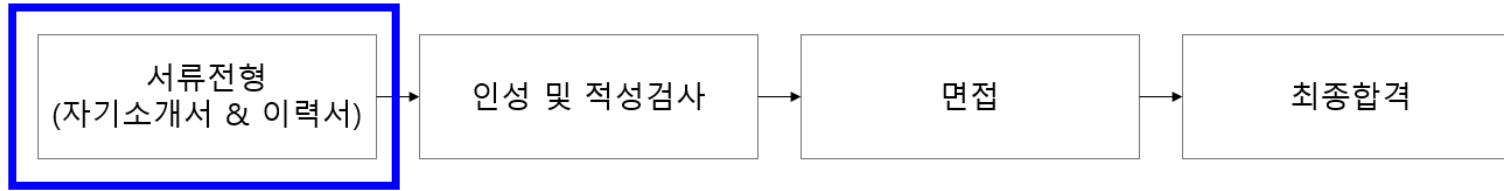




1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

기술 키워드 : 자기소개서 전처리, 딥러닝, TextRank 알고리즘

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?



[신입사원 채용 프로세스(Process)]

- 2019년 하반기 롯데 신입사원 지원자 약 49,000명.
- 신입사원 채용은 이력서보다 자기소개서를 중요하게 평가.
(*경력이 없는 신입사원의 업무 수행능력은 비슷하기 때문에 지원자의 잠재력을 자기소개서로 평가한다.)
- 매 채용시 동일한 질문에 대한 답변을 받는다. 따라서 **정형화된 데이터**가 축적되어 있다.
그렇다면 **딥러닝**으로 우수한 자기소개서를 **분류(Classification)** 할 수 있지않을까?

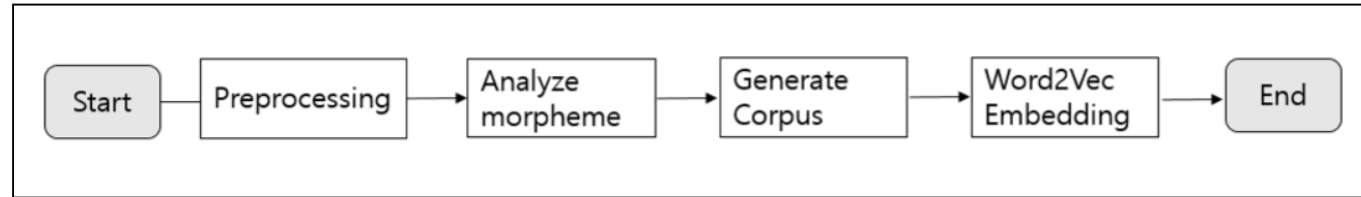
1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

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 1학년 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 Database를 꾸준히...	800
5	입사 후 포부	입사 후 10년 동안의 회사생활 시나리오와 그것을 추구하는 이유를 기술해주세요. [고객의 니즈를 충족시키는 MD] 음니 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?



[자기소개서 전처리 및 임베딩(Embedding) 순서도]

- 자기소개서의 한글, 영문, 숫자를 제외한 특수문자, 따옴표 등은 정규표현식을 사용하여 모두 제거한다.
- Okt 형태소 분석기로 자기소개서를 형태소 단위로 토큰나이징(Tokenizing)한다.
- 토큰나이징(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?

[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)
```

```
[('다른 사람', 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', 0.881217777290344),  
(('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")
```

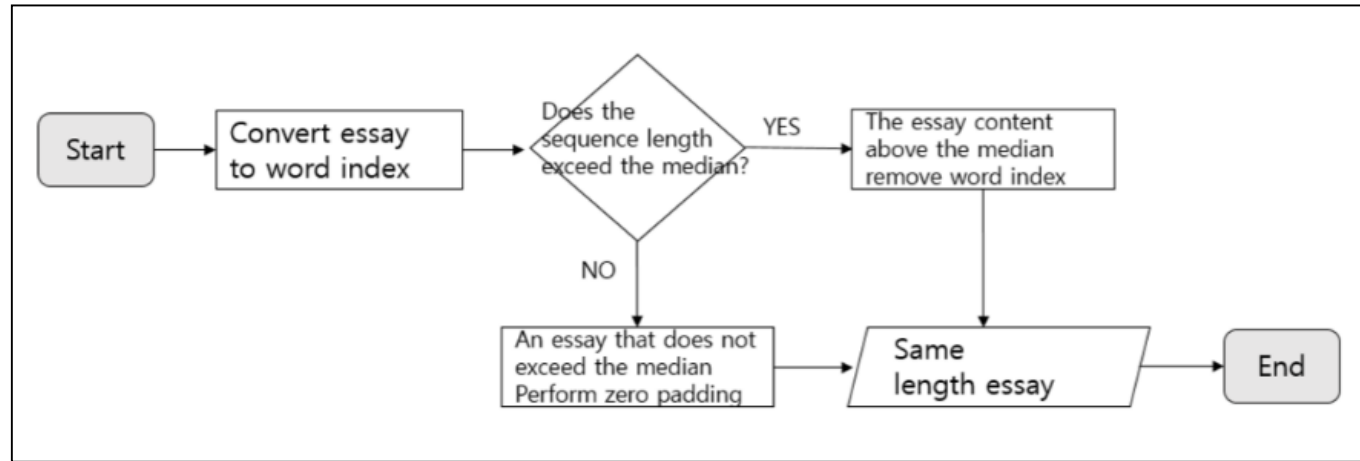
```
[('축구선수', 0.6439859867095947),  
(('안정환', 0.629976212978363),  
(('맨유', 0.6285330057144165),  
(('박지성', 0.6210668683052063),  
(('선수', 0.6134283542633057),  
(('아이버슨', 0.6086547374725342),  
(('토트넘', 0.591521143913269),  
(('유나이티드', 0.5879682898521423),  
(('프리미어리그', 0.580830454826355),  
(('야구선수', 0.5790668725967407)]
```

임베딩(Embedding)이
잘 됐나 간 한번
볼까요?



간 함 볼까요?

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

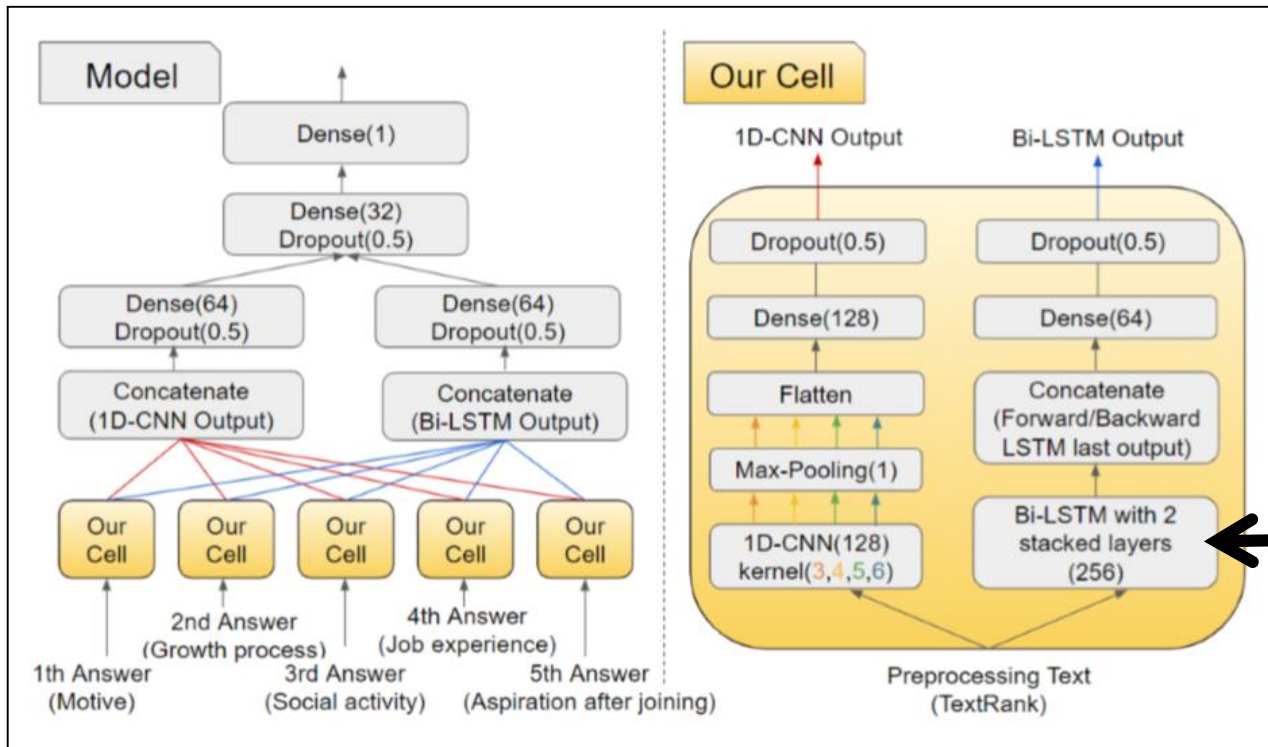


[자기소개서 단어 인덱스 변경 순서도]

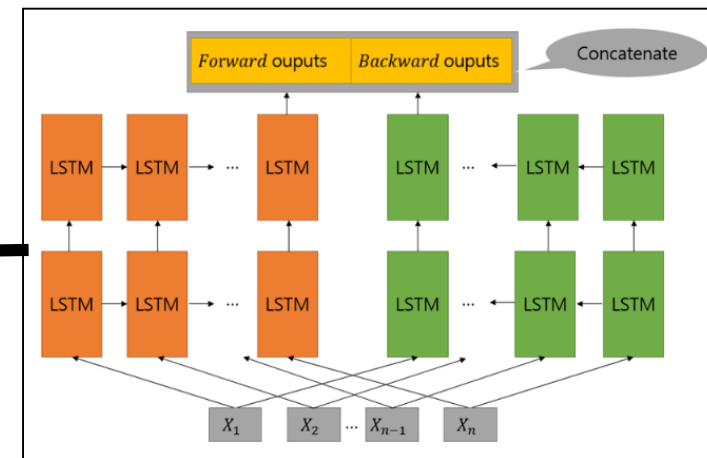
- 딥러닝 모델에 입력하기 위해 형태소를 사전의 **인덱스**로 변경한다.
 - 각각의 질문에 대한 답변 별로 시퀀스(Sequence) 길이의 **중앙값**을 구한다.
- (*평균이 아닌 중앙값을 기준으로 삼은 이유는 중앙값이 평균보다 이상치(Outlier)에 덜 민감하기 때문이다.
- 답변 별로 동일한 길이의 시퀀스를 만들기 위해 중앙값보다 **짧은** **답변**은 자기소개서 앞쪽에 **제로 패딩**
- 중앙값보다 **긴** **답변**은 중앙값 이후 단어 인덱스를 **삭제**한다.)

1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

- 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.
- 활성화 함수(Activation Functions)는 마지막 노드를 제외하고 하이퍼볼릭 탄젠트 함수(tanh)를 사용한다.
- 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?

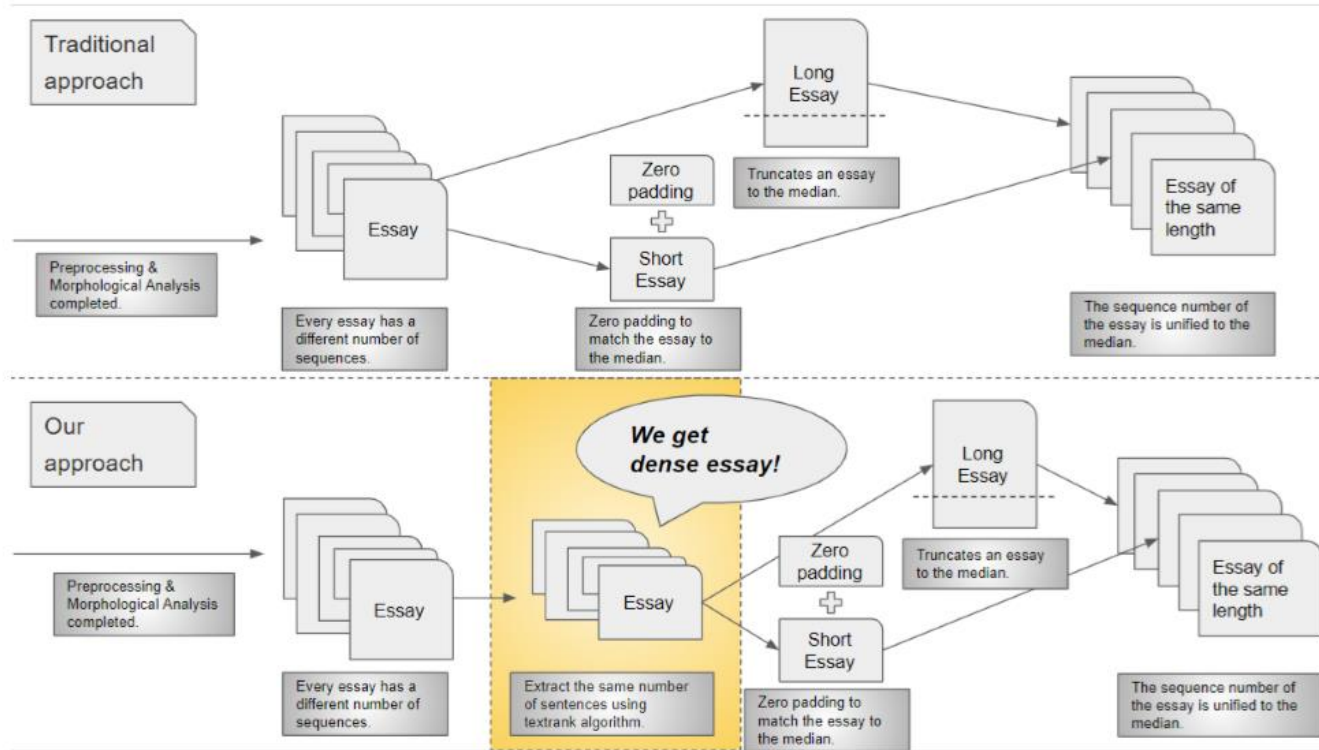
★딥러닝 정확도 15% 이상 향상

그러나 딥러닝 정확도는
기대에 미치지 못했습니다.
무엇이 문제였을까요?

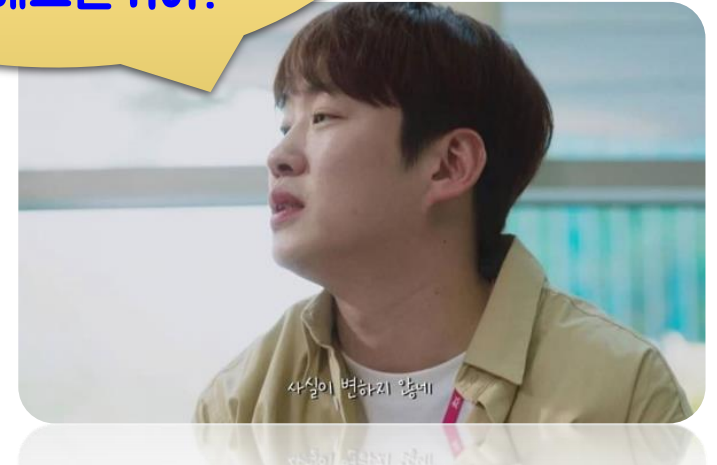
문제는 **라벨링**이었습니다.
서류전형 합격이라고 모두
우수한 자기소개가 아니고,
서류전형 탈락이라고 모두
우수하지 못한 자기소개서가
아니었던 것입니다.



1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?



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



- **TextRank**를 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?

[TextRank 활용한 자기소개서 요약 예시]

순서	TextRank 점수 순위	TextRank 점수	문장
1	2	1.45233	1년 동안 의 교내 영어 학습 동아리 회장 으로서...
2	3	1.25027	이른 오전 매일 7시 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 률...
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 높이는 데 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	4년 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...

1년 동안의 교내 영어학습동아리 회장으로서 구성원들과 소통하고 협력하여 능동적 참여를 이끌어 낸 경험이 있습니다. 이른 오전 매일 7시 30분에 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 사람에 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 학기 동안 낙오자가 없고 다른 어떤 동아리보다도 결속력 있는 영어학습동아리를 만들 수 있었습니다.

***TextRank
점수가 가장 낮은
문장이 요약된다.**

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

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

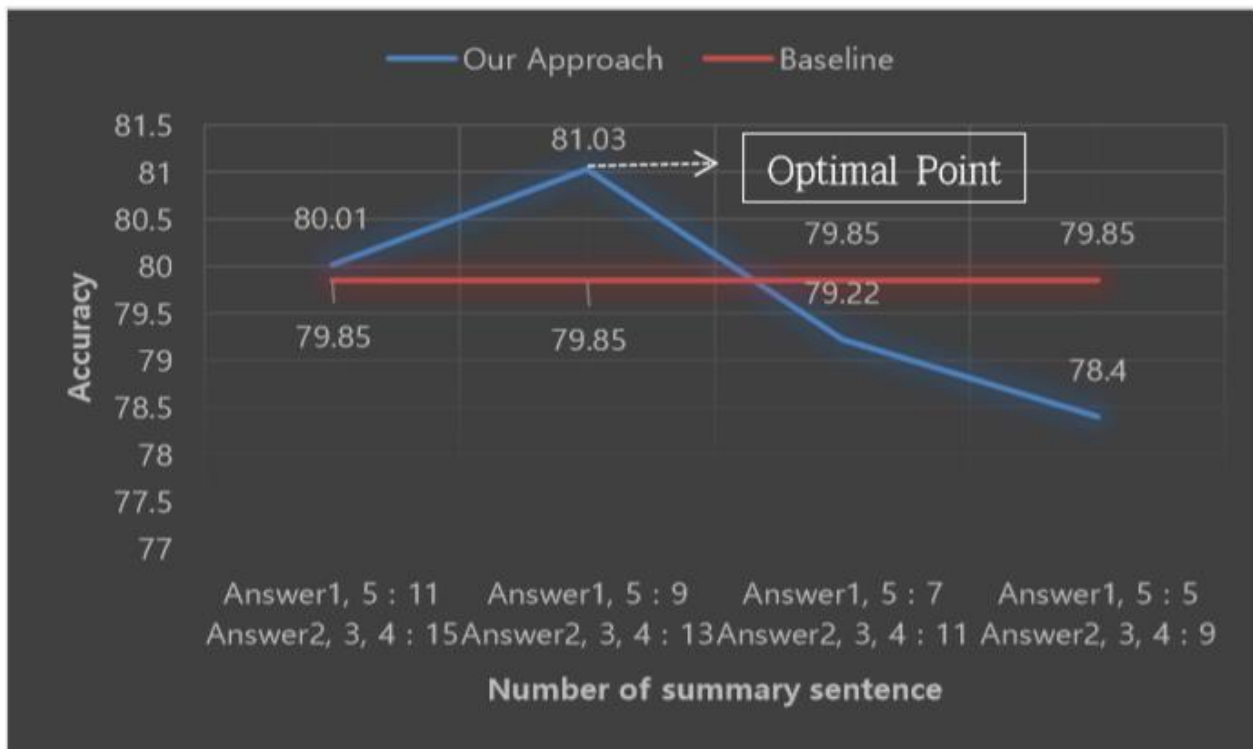
*참고 논문

- [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]

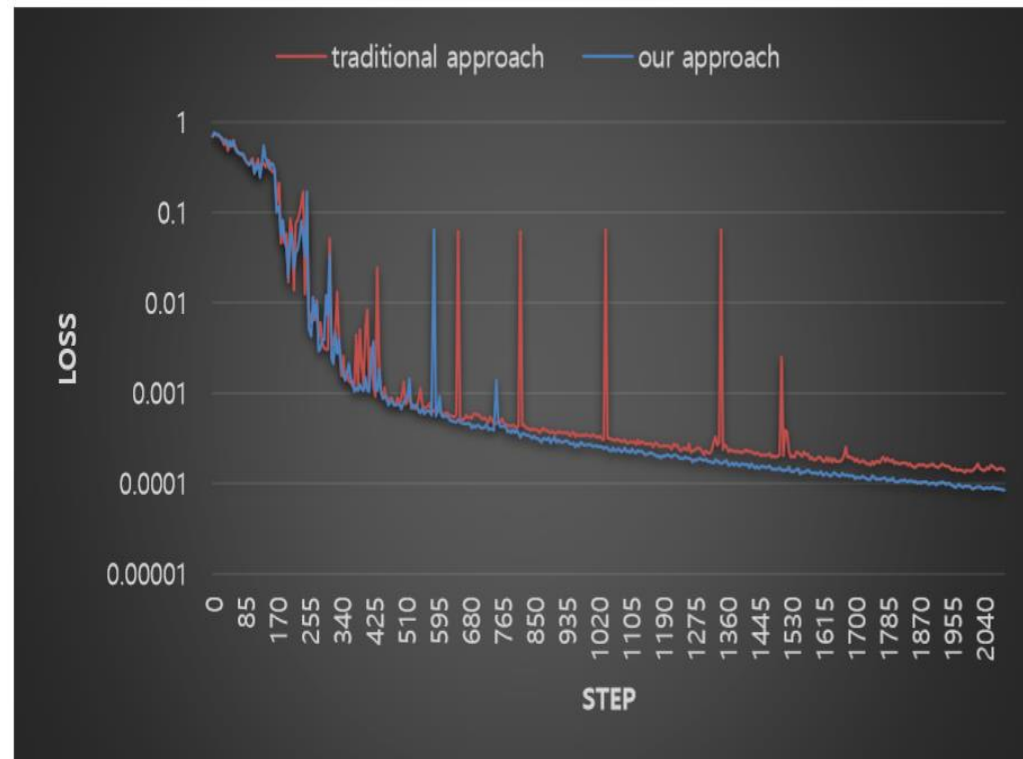
1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

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

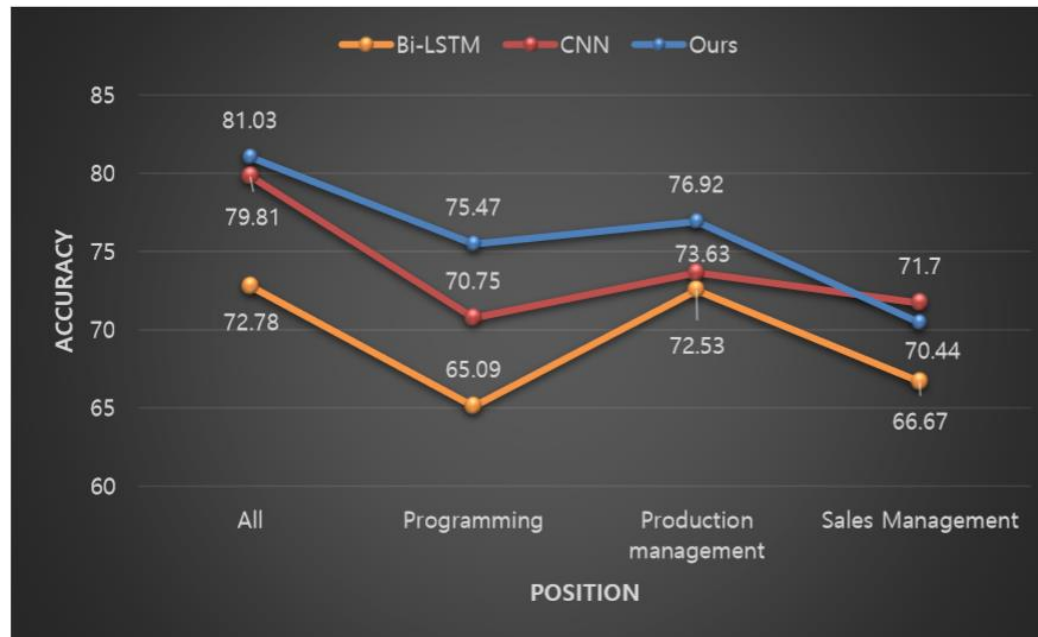


1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

- Ours 의 F1-Score 는 각각 0.8, 0.82 로 합격 및 불합격 중 어느 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 되었다는 점이다.
- 자기소개서는 데이터가 부족한 도메인(Domain)인 만큼 딥러닝의 특성상 직무별로 데이터가 충분하지 않다면 직무를 구분하지 않는 것이 오히려 좋은 성능을 낼 수 있다.

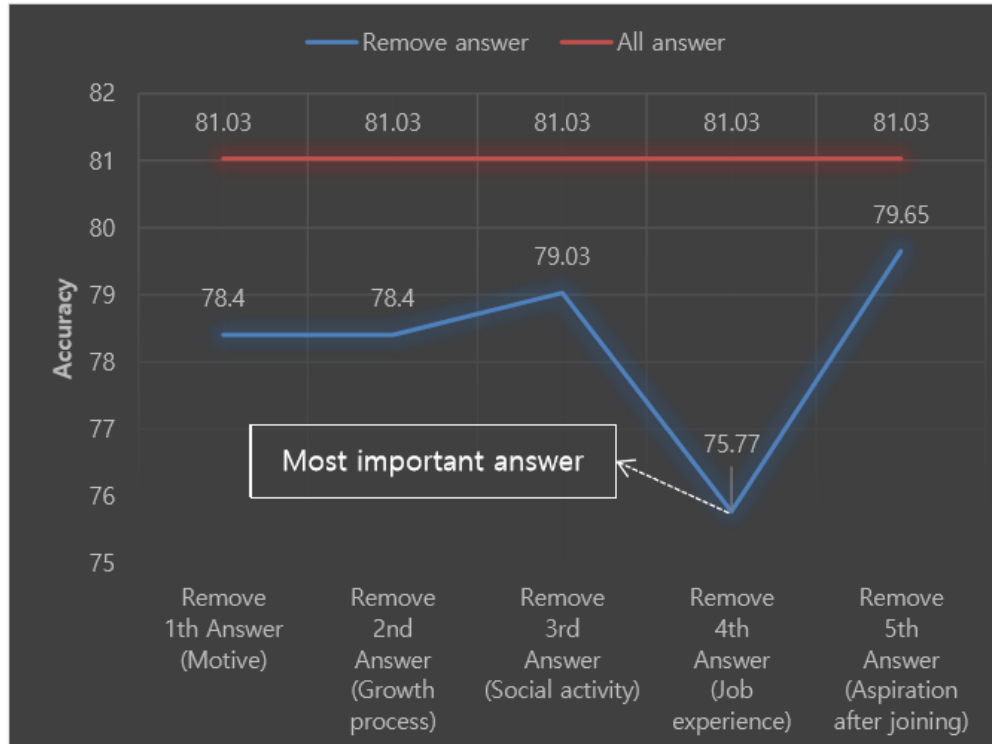
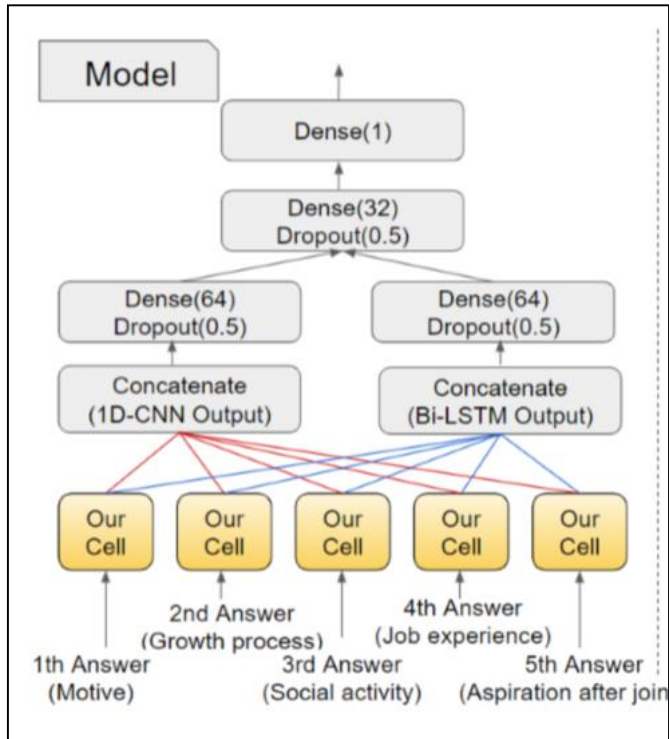
에잇! 이놈도 아녀!
최고의 성능을
내는 놈만
살아남는겨!

Position	Model	Precision		Recall		F1-Score		Accuracy
		0(Fail)	1(Pass)	0(Fail)	1(Pass)	0(Fail)	1(Pass)	
All	Bi-LSTM	0.69	0.77	0.76	0.7	0.72	0.73	72.78
	1D-CNN	0.72	0.88	0.87	0.75	0.79	0.81	79.81
	Ours	0.74	0.88	0.87	0.76	0.8	0.82	81.03
Programming	Bi-LSTM	0.34	0.74	0.28	0.79	0.3	0.77	65.09
	1D-CNN	0.29	0.83	0.33	0.8	0.31	0.81	70.75
	Ours	0.29	0.89	0.44	0.81	0.35	0.85	75.47
Production Management	Bi-LSTM	0.74	0.68	0.86	0.5	0.8	0.58	72.53
	1D-CNN	0.76	0.68	0.86	0.52	0.81	0.59	73.63
	Ours	0.86	0.52	0.83	0.59	0.84	0.55	76.92
Sales Management	Bi-LSTM	0.55	0.83	0.82	0.57	0.66	0.67	66.67
	1D-CNN	0.54	0.97	0.96	0.6	0.69	0.74	71.7
	Ours	0.52	0.97	0.96	0.59	0.67	0.73	70.44



1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

- 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.

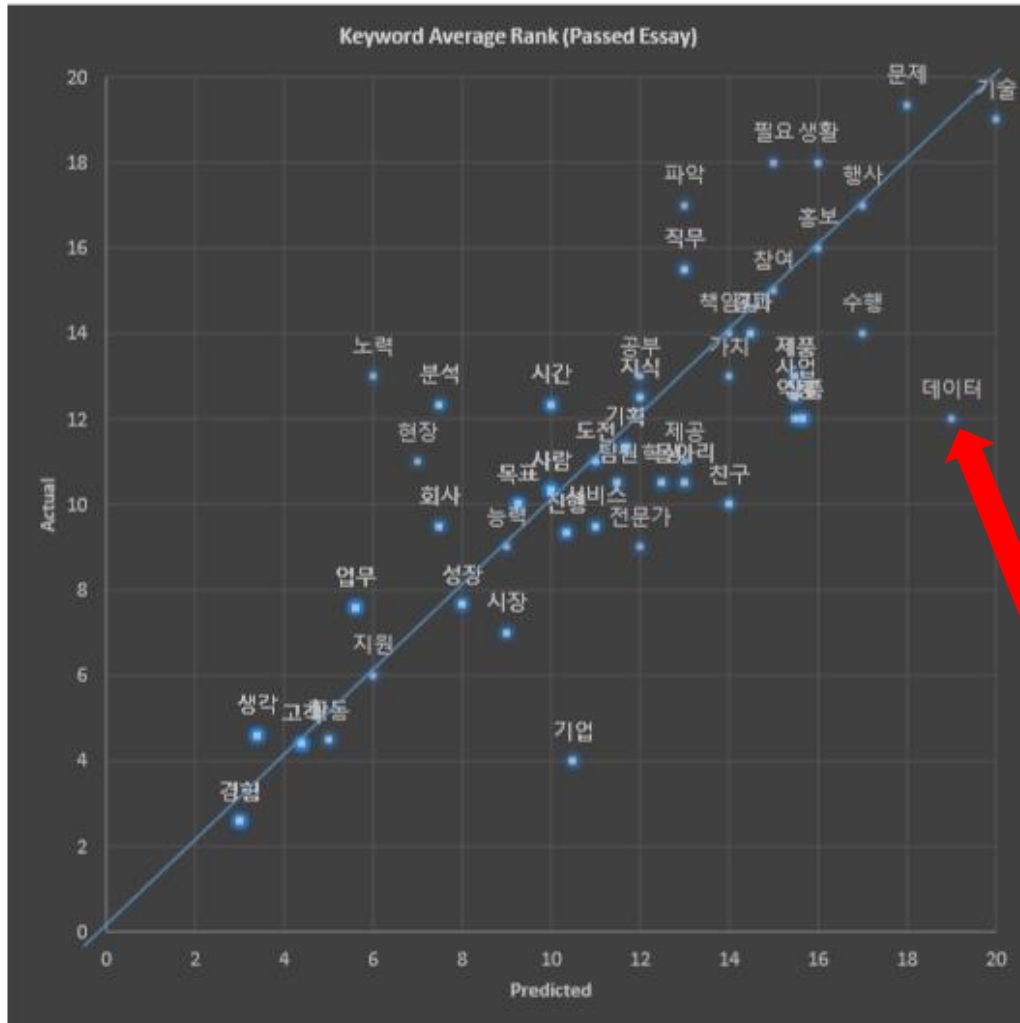


뭔가 중헌디!
어떤 질문이
중헌냐고!



1. 딥러닝으로 우수한 자기소개서를 분류할 수 있을까?

합격 자기소개서 키워드 비교



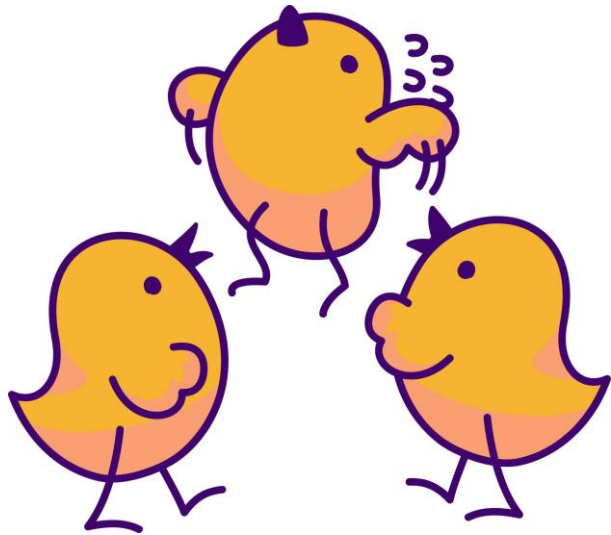
- 자기소개서를 Komoran 형태소 분석기로 명사 추출 ->
-> 1글자 제거 -> TextRank로 키워드 추출 ->
추출된 키워드 빈도수로 순위 부여 ->
딥러닝의 예측 순위 / 실제 순위를 비교한 결과를 2차원으로 표현
- 정확도가 100% 딥러닝 모델은 키워드가
모두 $y = x$ 직선에 위에 놓인다.
- $y = x$ 직선에서 멀어질수록 실제와 예측간 순위 차이가 크다.
- $y = x$ 직선보다 아래쪽에 있는 키워드는
실제 순위가 예측 순위보다 높다.
- [데이터] 키워드는 채용 담당자는 합격으로 많이 분류했지만,
딥러닝 모델은 그보다 적게 분류했음을 알 수 있다.

- **혼동행렬을 기준으로 각 영역별 키워드 표현**
 - **많이 등장한 키워드일수록 크고, 중앙에 표현**
 - **[생각], [업무] 등 무난한 의미를**
- 지닌 단어들은 불용어 처리**



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

[illegible]



AI가 자기소개서
쓰는걸 도와주면
얼마나 편할까?



2. 트랜스포머를 활용한 자기소개서 문장 생성

기술 키워드 : 딥러닝, Transformer, NER, TF-IDF 알고리즘

2. 트랜스포머를 활용한 자기소개서 문장 생성

- *자기소개서의 문장을 입력하면

다음 문장을 추천해주자!

Ex)

Encoder : '저는 00동아리의 회장으로 1년동안
활동했습니다.'

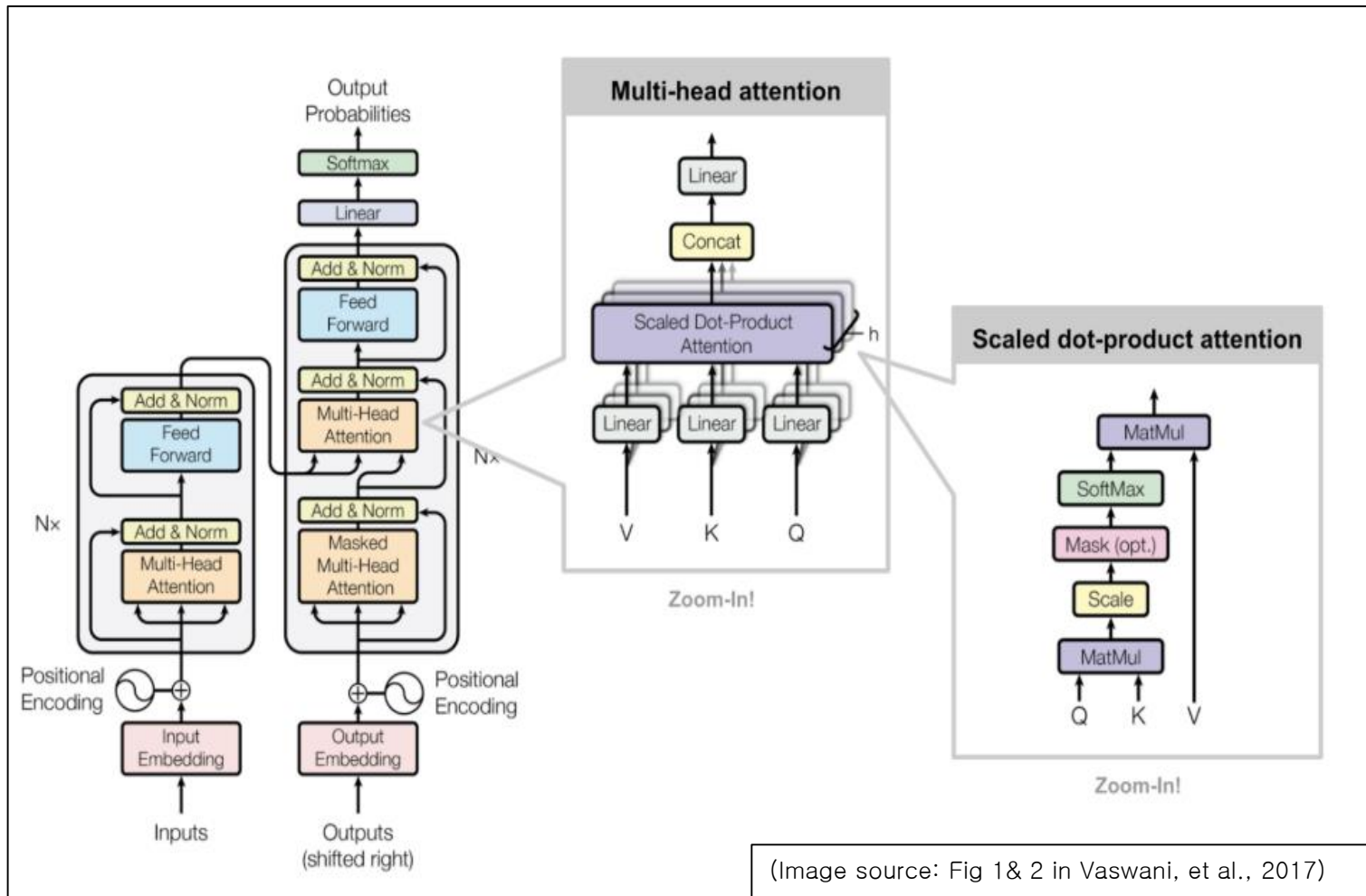
Decoder : '당시 동아리 출석률이 저조한 문제가
있었습니다.'



Encoder : '당시 동아리 출석률이 저조한 문제가
있었습니다.'

Decoder : '저는 이러한 문제를 해결하기 위해
00아이디어를 모두에게 실천해보자고 제안했습니다.'

- (문장 생성 시연)



*참고 논문

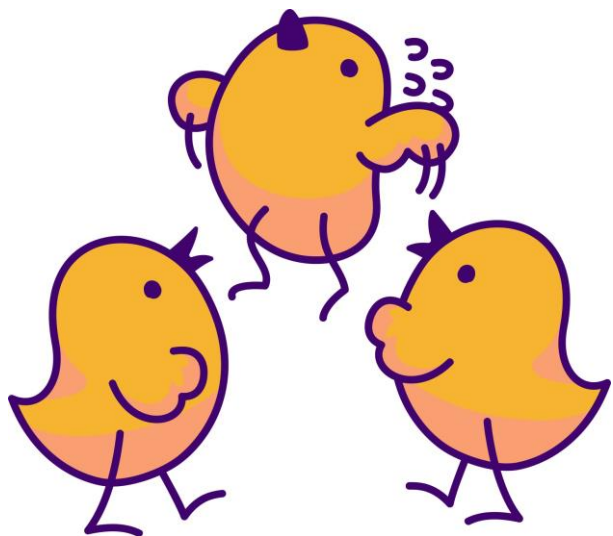
[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000–6010.

2. 트랜스포머를 활용한 자기소개서 문장 생성

- TF-IDF, 정규표현식을 사용하여 같은 비정상 자기소개서 제외
- 좋은 문장을 생성하기 위해 자기소개서 중 우수한 자기소개서 추출
- 우수한 자기소개서를 문장별로 분리하여 연속된 문장을 Pairing
- 답변 별로 모델을 분리
(지원동기 / 성장과정 / 사회활동 / 직무경험 / 10년계획)별 5개의 트랜스포머 모델 생성
- 회사명 사전을 구축하여 데이터내 회사명은
'[회사명]'으로 마스킹 처리(NER)
- 파이썬의 sent tokenize 모듈이 간혹 문장을 잘못
분리하는 경우를 방지하기 위해 자기소개서 [,] 등 제거
- 문장 분리 후 문장길이가 50이하인 것은 제거
ex) 14.7 ~ 15.7까지 동아리 부회장으로 활동했습니다.
-> 날짜에 찍인 .을 구두점으로 오인식하는 경우 방지

1. 'ooo...ooo'
2. 'as:fjasdajsdd'
3. 'ㅇㄱㄴ제다쓰냥'
간혹 외계어로
제출하신 지원자가
있습니다!





잡았다 요놈!



3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

기술 키워드 : 딥러닝, CNN, NER

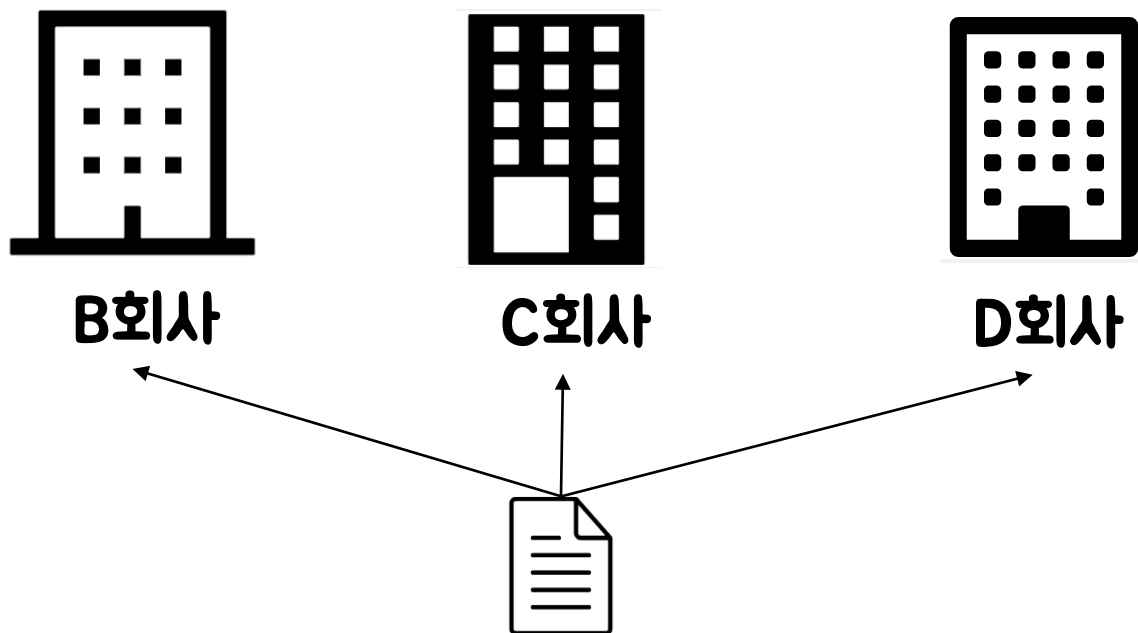
3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?

- 자기소개서 빅데이터를 활용하여 표절여부를 판단하고 있습니다.
- 본 세션에서는 표절에 관한 기술은 제외하고, 딥러닝을 활용하여 타 회사에 지원한 자기소개서를 그대로 복사, 붙여넣기 했을때 어떻게 찾아낼 수 있는지 소개해드리겠습니다.

A회사에 지원할때 썼던,
자기소개서를
B회사에 그대로 복사한다면
딥러닝은 찾을 수 있을까?



3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?



A회사에 제출한 자기소개서



대세는
원소스멀티유즈
(One source Multi use)
아니겠어!

[1명의 지원자가 제출한 자기소개서 일부 예시]

Oracle의 창업주인 래리 엘리슨은 73세의 나이에도 기술 고문을 하고 있습니다. 그의 이런 경력은 끊임없는 창조와 혁신을 통한 자기개발이 있었기 때문에 가능한 일이라 생각합니다. 저도 "self-challenging"을 즐기는 개발자가 되고 싶습니다.

이러한 꿈을 롯데정보통신에서 시작하고 싶습니다. 그 이유는 롯데정보통신 또한 실력을 키우기 위해 끊임없이 노력하는 개발자를 인재상으로 결코 있습니다. 또한, IoT, 빅데이터, 모바일, 클라우드, 핀테크 등 변화하는 ICT분야에서 시장을 선점하기 위해 음 채널, 스마트팩토리, 물류 IT, 클라우드, L페이 등 새로운 사업에 도전하고 있어 세계에서 인정받는 개발자로 거듭날 수 있는 가능성이 크다고 생각했기 때문입니다.

끊임없이 노력하고 변화에 도전하는 self-challenging 엔지니어가 되어, 롯데정보통신에 반드시 필요한 인재가 되겠습니다.

1년 동안의 교내 영어학습동아리 회장으로서 구성원들과 소통하고 협력하여 능동적 참여를 이끌어 낸 경험이 있습니다. 이른 오전 매일 7시 30분에 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 참석률을 목표로 설정했습니다. 이 목표를 이루기 위해 릴레이 모닝콜을 제안했습니다. 학교와 집과의 거리가 먼 사람부터 가까운 사람에게 한명씩 모닝콜을 하는 시스템입니다. 그 결과 참석률을 높이는 데 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.

4년이 지난 지금도 구성원들과 그 당시 추억 이야기를 나누곤 합니다. 이처럼 구성원들과의 무한한 신뢰, 온정과 수용이 깔린 딱딱 같은 불임성으로 공동의 목표를 이루기 위해 나아가겠습니다. 앞으로 롯데정보통신에서 구성원들을 아우르는 결속력을 다지는 힘과 주위를 살피며 작은 변화에도 민감하게 반응하여 발전하는 모습으로 인정받는 신입사원이 되겠습니다.

저의 장점은 개발자로서 끊임없는 창조와 혁신을 통한 "self-challenging"을 즐기는 엔지니어가 되기를 추구한다는 것입니다. 그래서 지속적인 학습과 노력을 통해 자기개발을 하고 있습니다. 3번의 공모전 수상, 1번의 인턴 경험을 통해 저의 이런 장점을 잘 확인 할 수 있었습니다.

특히 최우수상을 수상한 "경상북도 공공데이터 활용 앱 개발 공모전"에서 SNS와 LBS기반의 관광지 정보 앱을 개발하였습니다. 당시 아이디어 고안에서부터 개발까지 혼자서 도전하는 공모전이라 많은 고민을 하였습니다. 중간에 포기하고 싶은 마음도 있었지만, 개발을 진행하면서 하나하나 발전해가는 앱의 모습과 조금씩 성장해 가는 저의 실력에 힘을 내서 개발을 완성 할 수 있었습니다.

디지털 산업진흥원 소속 UNIZEQ사에서 인턴으로서 MEAN스택 기반의 기업용 커뮤니케이션 플랫폼을 개발하는 업무를 담당하였습니다. 당시 저는 최신 기술인 MEAN스택 기술에 대한 지식이 전무하여 주어진 프로젝트를 해낼 수 있을까 생각했지만, 하루 3시간도 자지 않는 강행군 속에 관련서적 공부와 개발을 병행하였으며, 최종일에 다다라서는 30여 시간 동안 잠을 자지 않고 프로젝트를 성공적으로 마무리 해냈습니다.

위의 두 경험을 통해 인터넷, 개발서적없이 완벽한 결과물을 낼 수 없다는 제 약점을 찾을 수 있었습니다. 하지만, 끊임없이 변화하고 매일같이 새로운 기술이 나오는 IT분야에서 최신 최적의 기술을 적용 시키기 위해 인터넷 검색과 관련서적 공부는 필수라고 생각합니다.

입사 후 5년 동안은 현재 롯데정보통신이 추진 중인 음니채널과 엘페이에 기여 하고 싶습니다. 특히 제가 관심을 보여왔고 경험에 있는 모바일, 빅데이터 분야에서 제 역량을 발휘하고 싶습니다. 시스템 사용자의 소비패턴, 행동패턴을 R 언어로 분석하여 필요한 데이터를 수집하고 사용자에게 맞춤형 정보를 제공하게끔 하겠습니다. 또한, 제가 아직 경험해보지 못한 핀테크 기술들을 직· 간접적으로 경험하고 학습하여 해당 기술을 담당하는 사원들과 원활한 의사소통을 할 수 있도록 하겠습니다.

입사 후 5년 후에는 저의 축적된 기술과 노하우를 바탕으로 롯데정보통신의 글로벌 진출을 함께 하고 싶습니다. 특히 고도의 기술 융합체인 IBS분야에 도전하여 롯데정보통신만의 해외고충발딩 시스템을 기획하고 개발하겠습니다. 또한, 지속적인 자기개발을 하여 해당 시스템에 최신 최적의 기술을 적용하겠습니다.

(이하 생략...)

3. [Ctrl+C, Ctrl+V] 자기소개서 어떻게 찾을까?



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

[A사에 제출한 자기소개서를 B사에 그대로 제출한 Case]



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

[A사에 인턴으로 일했던 경험을 B사에 제출한 Case]

- 단순히 자기소개서에 타회사명이 언급된다고 문제가 되지 않습니다. 인턴, 공모전 등 타사에서 진행한 사례가 자기소개서에 등장할 수 있기 때문입니다.
- 아래와 같은 방법을 고안했습니다!

- 1) 회사명이 등장한 모든 문장 수집!(NER, 정규표현식 활용)
- 2) 회사명을 학습하지 않게 하기 위해서 문장에서 회사명 제거
(자사뿐만 아니라 다양한 회사에서 사용할 수 있도록)
- 3) 반드시 지원하는 회사명이 있어야 하는 문장인지 혹은
타사명이 있어도 괜찮은 문장인지 문장마다 0 or 1로 라벨링한 train data 생성
- 4) *1D - CNN 모델 학습
- 5) 다음 두 조건을 동시에 만족하면 문장이 있다면 복사한 자기소개서로 판별
* (타사명이 존재) AND (반드시 지원하는 회사명이 와야하는 문장)
⇒ 실제 타사 PoC 결과 **100% 정확도** 보였음.

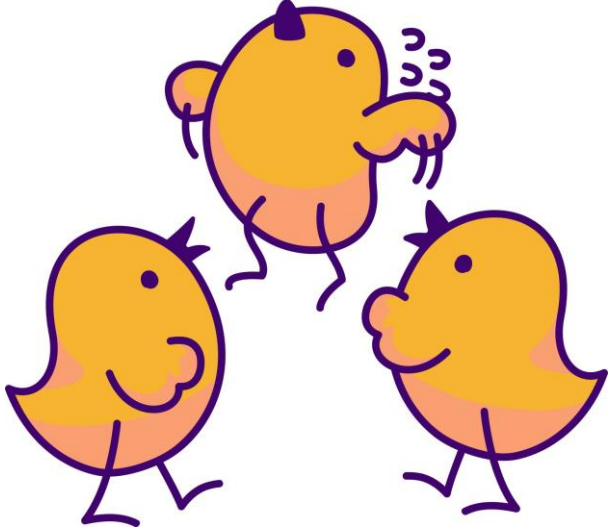
TRAIN DATA 예시>

에서 열심히 일하겠습니다.(0 반드시 지원하는 회사명이 와야함)

에서 주최한 공모전에서 최우수상을 수상했습니다.(1 타 회사명이 와도 괜찮음)

*참고 논문

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



4. 오타 어떻게 찾을까?

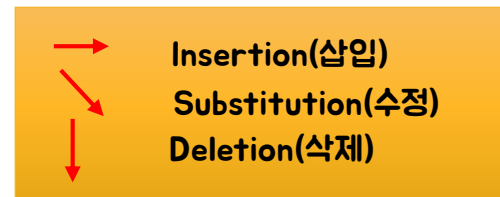
기술 키워드 : Levenshtein distance 알고리즘

4. 오타 어떻게 찾을까?

- Insertion(삽입), Deletion(삭제), Substitution(수정)은 비용을 계산
Ex) [사랑, 사랑해] 는 한 글자가 삽입되어 거리 1
[나무, 나]는 한 글자가 삭제되어 거리 1
[자연어, 자연언]는 연의 초성 수정(0.3). 어의 종성 수정(0.3)되어 거리 0.6

- 그렇다면 [자연어처리짱]과 [자여어철리긱]사이의 거리는?

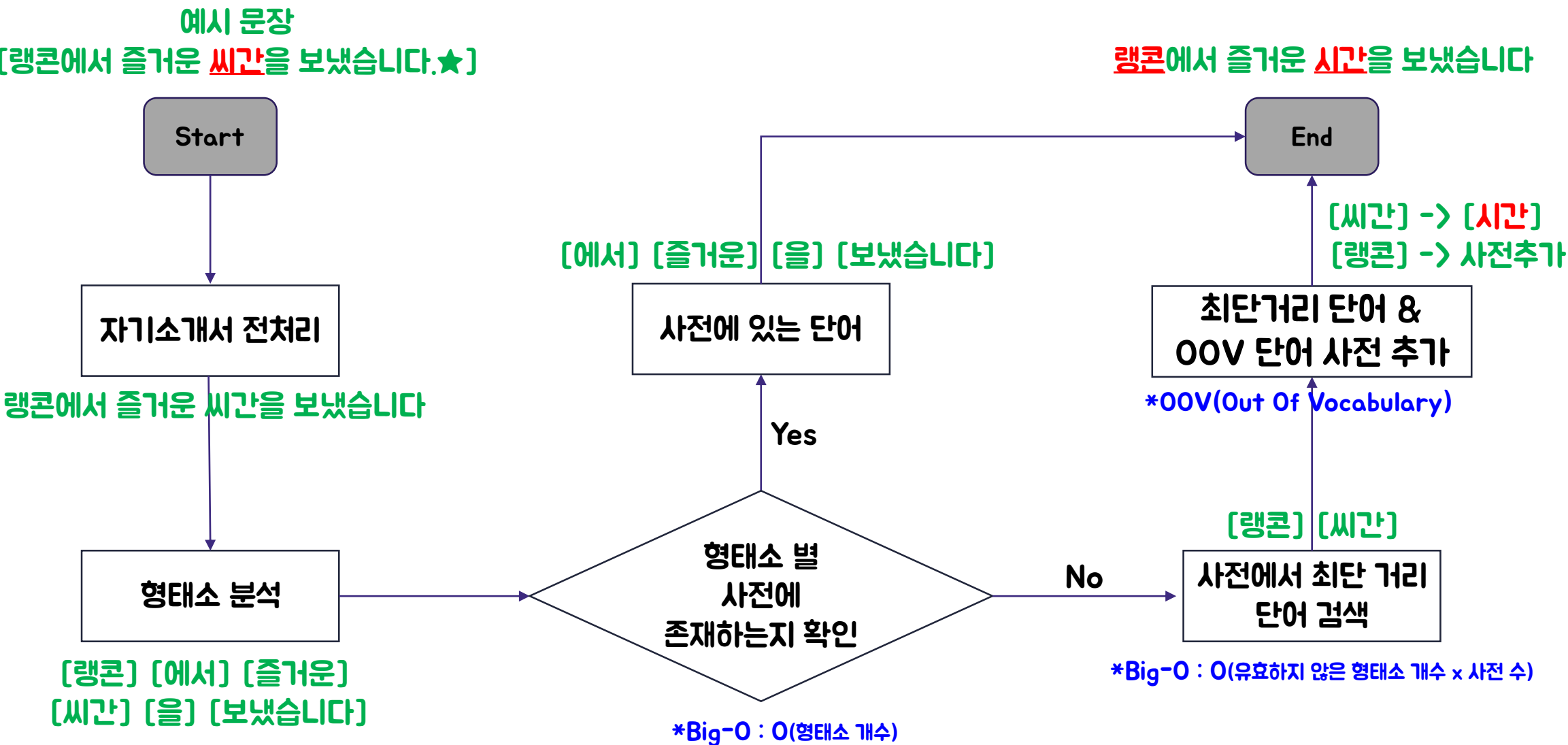
- > [1,1]은 [자, 자] 같기 때문에 0으로 초기화한다.(다르면 1)
- > [1,j]는 j가 1씩 증가함에 따라 한 글자씩 추가되었기 때문에 1씩 더한다.
- > [2,1]은 [자연, 자]가 한 글자가 추가되었기 때문에 1이다.
- > [2,2]은 [자연, 자여]의 종성이 다르기 때문에 0.334(약0.3)이다.
- > [2,3]은 [자연, 자여어]는 [여]와 [연]의 종성이 다르고, 한 글자가 추가되었기 때문에 1.3이다.
- > 위와 같은 방식으로 매트릭스를 계산한다.
- > [6,6]의 값 1.6이 [자연어처리짱]과 [자여어철리긱]의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

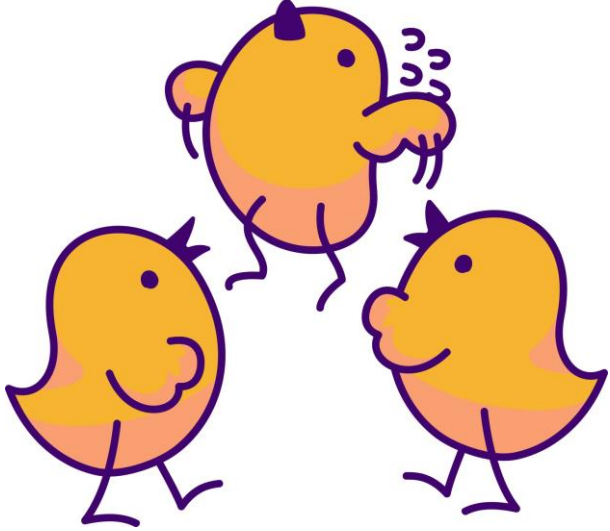


	자	여	어	철	리	긱
자	0	1	2	3	4	5
연	1	0.3	1.3	2.3	3.3	4.3
어	2	1.3	0.3	1.3	2.3	3.3
처	3	2.3	1.3	0.6	1.6	2.6
리	4	3.3	2.3	1.6	0.6	1.6
짱	5	4.3	3.3	2.6	1.6	1.6

4. 오타 어떻게 찾을까?

• 오타 검정 순서도(simple version)





5. 롯데그룹 채용 안내

5. 롯데그룹 채용 안내

함께 가는 친구, 롯데

2019 LOTTE RECRUITMENT

로그인 | 회원가입 | 나의지원정보 | 인사지원가이드

CONTENTS FAMILY 채용공고 입사지원 나의 지원정보

채용정보
발령과 책임감을 갖춘 글로벌 인재를 찾습니다.

입사지원
당신의 활동과 열정이 미래의 롯데를 완성합니다.

롯데HR
롯데만의 인사노하우를 공개합니다.

롯데소개
여제, 오늘 그리고 내일의 롯데가 있습니다.

롯데 채용 브로슈어
롯데 채용 직무 FAQ

좋아하는 것보다
좋은 스펙은 없어!
롯데 인턴·SPE태클
2019.10.24~11.06 Job.lotte.co.kr
별개 인턴십 / SPE태클전형
Job-Cafe: 11.01 (금) 밤세라미스 전대역점
사전신청: 10.28(월)~30(수), 3일간

[롯데푸드 채용 꿀팁] 롯데푸드 생...
L-TALK
롯데푸드 지원패널
생산/품질 관리 직무 Q&A

채용소식
롯데푸드 채용 꿀팁 롯데푸드 생...
2019년 하반기 롯데 인턴/SPEC태... 2019.10.11
2019년 하반기 캠퍼스 리크루팅 안내 2019.08.30
2019년 하계 인턴십/상반기 SPEC... 2019.06.27
2019년 하계 인턴십 서류전형/상반... 2019.06.05
2019년 상반기 신입 사원/전역장교/... 2019.06.04

직무소개
내가 꿈꾸는 커리어, 롯데인의 하루
부문선택 식음료
회사선택 롯데제과
바로가기

롯데 채용 일반 FAQ
회원가입 | 입사지원 | 지원서작성/수정 | 채용결과
회원가입은 어떻게 하나요? (14세 이상 내국인, 이주...
주민등록번호 등 개인정보 수집은 어떻게 되나요?
해외거주 외국인 회원가입은 어떻게 하나요?

롯데 채용 Q&A
입사지원에 대한 문의사항
입사지원에 대한 궁금사항을
신속, 친절하게 답변해 드립니다.
바로가기

관계사

• 신입 채용

- 2020년 상반기 공채 예정
- 스펙태클 채용 예정

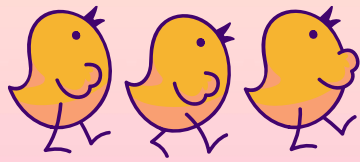
: SPE태클(스펙태클) 채용은 2015년에 시작된 롯데 고유의 블라인드 채용 전형입니다. 학벌이나 스펙 중심의 서류 전형에서 벗어나 지원자 직무 수행 능력과 역량 중심으로 인재를 선발합니다. 화려한 볼거리(Spectacle)처럼 다양한 지원자의 이력에 집중하고, 무분별한 스펙 쌓기에 태클을 건다(spec-tackle)는 중의적인 의미를 담고 있습니다.

• 경력 채용

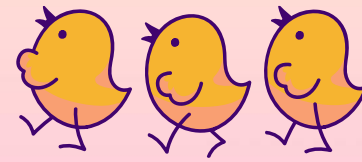
- 롯데 채용 사이트내 수시 공고

*롯데그룹 채용 URL

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



LangCon 2020



QnA
감사합니다.
😊

김영준

datajuny@gmail.com