

# KCI 문헌 유사도 검사 결과 확인서

## \* 유의사항

KCI 문헌 유사도 검사에서 나타나는 유사도 수치는 단순한 자동검사 결과이므로,  
문헌 간 유사여부 판단을 위해서는 반드시 해당 분야 전문가의 직접 검사가 필요함을 알려드립니다.

확 인

유사율		3%	
발급번호	00011176156	발급일자	2024.01.19 14:22
검사일자	2024.01.19 14:22		
검사명	글로벌경영_20240119_제출논문검사		
검사문서	글로벌경영학회지-송지호.hwp		
비고			

검사설정	유사율 기준 [5이절], 인용문장 [포함], 출처표시문장 [포함], 목차/참고문헌 [제외]
비교범위	[KCI 논문] [학술대회 논문]

## 유사 분석 정보(상세)

문서유사율	전체문장	동일문장	유사의심문장	인용포함문장	출처표시문장
3%	259	0	14	1	13

## 비교 문서 정보

번호	유사율	출처정보	비고
1	1%	[KCI 논문] <a href="http://www.jitam.or.kr">www.jitam.or.kr</a> - 제목 : 양방향 인제매칭을 위한 BERT 기반의 전이학습 모델 - 저자 : 오소진(한남대학교) - 발행년 : 2021.08	
2	1%	[KCI 논문] KCI 논문 - 제목 : 워드 임베딩을 이용한 관광여행지 선호도 비교에 대한 실증 연구 : 강릉 관광여행지 SNS 이용자 중심으로 - 저자 : 장원중(가톨릭관동대학교) - 발행년 : 2020.12	
3	1%	[KCI 논문] KCI 논문 - 제목 : 직무몰입이 지식은폐행동에 미치는 영향: 지식영역성의 매개 효과와 서번트 리더십의 조절효과 - 저자 : 김현우(울산대학교) - 발행년 : 2021.04	
4	1%	[KCI 논문] KCI 논문 - 제목 : 온라인 쇼핑 추천서비스 성공 요인: 한국 중국 비교 연구 - 저자 : 짜오권(경북대학교) - 발행년 : 2023.08	
5	1%	[KCI 논문] KCI 논문 - 제목 : 경영권 집중화 수준이 원가의 하방경직성에 미치는 영향 - 경영자 유형에 따른 차이를 중심으로 - 저자 : 이승현(경기대학교) - 발행년 : 2023.08	
6	1%	[KCI 논문] KCI 논문 - 제목 : 내부시장지향성이 시장지향성에 미치는 영향 - 저자 : 김동윤(충실대학교) - 발행년 : 2021.04	
7	1%	[KCI 논문] KCI 논문 - 제목 : 유비쿼터스 환경의 고객맞춤형 서비스를 위한 빅데이터 분석 기법 - 저자 : 김재수(서울과학기술대학교) - 발행년 : 2012.12	
8	1%	[KCI 논문] KCI 논문 - 제목 : 소셜데이터 분석방법을 이용한 제주관광 트렌드 서비스 설계 및 구현 - 저자 : 김후찬(제주한라대학교) - 발행년 : 2014.06	
9	1%	[KCI 논문] KCI 논문 - 제목 : 텍스트마이닝 기반의 재난현장 응급의료시설 대상지선정 프로세스 연구 - 저자 : 서상욱(가천대학교) - 발행년 : 2018.06	
10	1%	[KCI 논문] KCI 논문 - 제목 : KOSPI200 주가지수선물의 잔존기간에 따른 개인, 기관 및 외국인 투자자 거래비중 연구 - 저자 : 김봉수(상지대학교) - 발행년 : 2020.10	
11	1%	[KCI 논문] KCI 논문 - 제목 : 음수 값을 처리할 수 있는 DEA 모형의 변별력 평가 - 저자 : 박만희(부산가톨릭대학교) - 발행년 : 2020.08	
12	1%	[KCI 논문] KCI 논문 - 제목 : 설명가능한 인공지능(XAI) 방법론의 산업별 적용가능성에 관한 연구 - 저자 : 배재권(계명대학교) - 발행년 : 2023.04	

13	1%	[KCI 논문] KCI 논문 - 제목 : 영농조합법인 제주보름왓의 인적자원관리에 내재된 인간중심주의(Humanocracy)에 관한 연구 - 저자 : 박현철(성균관대학교 미래인문학 소셜앙트레프레너십 융합전공 박사과정) - 발행년 : 2023.04
14	1%	[KCI 논문] KCI 논문 - 제목 : 스마트팩토리 구축에 따른 생산성 효과 분석 - 저자 : 김태수(부산대학교 경영대학원) - 발행년 : 2023.09
15	1%	[KCI 논문] KCI 논문 - 제목 : 대규모기업집단 가족기업 여부가 기업신용등급 계열지원가능성에 미치는 영향 - 저자 : 강동창(동양미래대학교) - 발행년 : 2019.12
16	1%	[KCI 논문] KCI 논문 - 제목 : 커피전문점 모바일 어플리케이션 이용자의 지속사용의도에 관한 연구 - 저자 : 한수민(송실대학교) - 발행년 : 2021.08
17	1%	[KCI 논문] KCI 논문 - 제목 : 토픽 모델링을 이용한 고객 상품평 분석 - 저자 : 박만희(부산가톨릭대학교) - 발행년 : 2021.08
18	1%	[KCI 논문] KCI 논문 - 제목 : 차량 외부 디자인 변경에 따른 풍절음 예측을 위한 설명 가능한 딥러닝 모델 - 저자 : 김태연(고려대학교 산업경영공학부) - 발행년 : 2022.10
19	1%	[KCI 논문] KCI 논문 - 제목 : 차량 DTC 고장 예측을 위한 딥러닝 적용 사례 연구 - 저자 : 김한숨(홍익대학교) - 발행년 : 2023.09

## 검사대상문서

문장유사율: 20%

글로벌 경영학회지 DOI : <https://doi.org/10.38115/asgba.2020.17.0>. 글로벌 경영학회지 제XX권 제X호 제XX권 제X호 20XX년 X월 pp.1~1 글로벌 경영학회 트랜스포머 모델을 기반으로 한 직무추천에 관한 연구 전정우\*·송지호\*\*·안태호\*\*\* †본고는 전정우의 2023년도 박사학위 논문 「딥러닝 기반 텍스트 임베딩을 활용한 직무 추천 모델 연구」의 일부를 발췌하여 재작성한 것임을 밝힙니다.

## 비교대상문서

KCI 논문 | 제목: 설명가능한 인공지능... | 저자: 배재권(계... | 발행년: 2023.04

글로벌 경영학회지 DOI : <https://doi.org/10.38115/asgba.2020.17.195>. 글로벌 경영학회지 제20권 제2호 제20권 제2호 2023년 4월 pp.195~208 글로벌 경영학회 설명가능한 인공지능(XAI) 방법론의 산업별 적용가능성에 관한 연구 배재권\* \* 계명대학교 디지털 경영학부 경영정보학전공 교수 (\*\*\*\*\*@\*\*\*\*\*) 최초접수일 2023년 04월 19일 심사수정일 2023년 04월 27일 게재확정일 2023년 04월 29일 <요약> 기계학습(machine learning) 및 딥러닝(deep learning) 등 인공지능 기술의 학습 성능은 인간에 근접해 있거나 또는 앞서고 있으며, 인간은 인공지능이 찾아낸 숨겨진 패턴과 규칙을 통해 새로운 통찰력을 얻고 있으나 인공지능이 제시한 결과물에 대한 전달력과 설명력은 부족한 상황이다.

KCI 논문 | 제목: 스마트팩토리 구축... | 저자: 김태수(부... | 발행년: 2023.09

[9] 주용한, 조인수 (2020), 텍스트 마이닝을 이용한 중소제조업 스마트 팩토리 성공요인 분석, 글로벌 경영학회지, 17(5), 115~131, doi: <https://doi.org/10.38115/asgba.2020.17.5.115>

KCI 논문 | 제목: 워드 임베딩을 이... | 저자: 장원중(가... | 발행년: 2020.12

DOI : <https://doi.org/10.38115/asgba.2020.17.6.54> 54 글로벌 경영학회지 제17권 제6호 2020년 12월 pp.54~71 글로벌 경영학회 워드 임베딩을 이용한 관광여행지 선호도 비교에 대한 실증연구 : 강릉 관광여행지 SNS 이용자 중심으로 장원중 \* <요약> 본 연구는 워드임베딩의 단어표현기법을 이용하여 관광여행지 선호도 분석을 위한 실증 연구를 수행하고, 강릉 관광여행지 방문자의 네이버블로그와 인스타그램 이용자를 대상으로 비교 분석하여 살펴보고자함.

문장유사율: 0%

\* 공동주저자, 숭실대학교 일반대학원 프로젝트 경영학과 박사 (certimaster@gmail.com) \*\* 공동주저자, 숭실대학교 일반대학원 프로젝트 경영학과 박사과정 (songjiho@gmail.com) \*\*\* 교신저자, 숭실대학교 경영학과 교수 (ahnt@ssu.ac.kr) 최초접수일 : 201X년 0X월 2X일 심사수정일 : 201X년 0X월 0X일 게재확정일 : 201X년 0X월 1X일 <요약> 본 연구는 구직자의 이력서를 기반으로 한 직무 추천 모델들을 작성하고, 모델을 성능을 확인하며, 추천된 직무의 활용도를 높이는 방안을 제안한다.

문장유사율: 0%

이를 위하여 2차례의 실험을 하였으며 1차 실험은 Transformer를 적용한 직무 추천 모델에 이력서 데이터를 52개의 특징(Feature)으로 구성한 학습 데이터셋과 하이퍼 파라미터 설정을 사용하여 성능을 평가하였다.

문장유사율: 0%

2차 실험은 1차 실험에서 사용한 52개 특징에서 결측치가 30% 이상인 특징을 삭제한 34개 특징으로 구성된 학습 데이터셋과 파인튜닝 기법을 적용한 딥러닝 모델의 성능을 평가하였다.

문장유사율: 0%

Transformer 모델의 Epoch 값의 상승은 큰 의미가 있다. 본 연구를 통해 학습 데이터의 최적화 및 파인튜닝을 적용한 딥러닝 모델 설계는 직무 추천 모델의 성능 향상에 큰 영향을 미치는 것으로 확인되었으며, Epoch 값의 상승률을 보았을때 Transformer 모델에 양질의 이력서 데이터를 적용하여 추가적인 모델 학습을 진행한다면, 직무 추천 모델의 성능이 더욱 향상될 것으로 기대한다.

문장유사율: 0%

주제어 : 이력서 기반 직무 추천, 딥러닝, 트랜스포머 모델, I. 서론 본 연구는 취업자의 전공과 취업자의 실제 직무와의 미스매칭과 관련하여 경력자의 이력서 데이터의 학습을 통한 직무 추천 모델의 성능평가에 대한 연구를 진행한다.

문장유사율: 0%

직무 정보가 포함된 경력자의 이력서를 학습 데이터로 구성하고, 파인튜닝(Fine-Tuning)을 통한 모델 설계와 Transformer를 임베딩 기법으로 적용한 직무 추천 모델의 성능평가에 대한 연구를 진행한다.

## 출처 표시 문장

문장유사율: 21%

취업포털사람인에서는 1,124개 기업을 대상으로 '1년 이내 조기 퇴사 현황'에 대한 설문 조사를 한 결과, 기업 10곳 중 8곳(84.7%)이 '1년 이내 조기 퇴사자가 있다'라고 응답했으며, 조기 퇴사 사유 중에서도 '직무가 적성에 안맞아서'가 45.9%로 가장 많았다(사람인, 2022).

## 출처 표시 문장

문장유사율: 0%

고용노동부 2023년 총예산 30.3조 중 직업훈련(9.2%, 2.8조)과 고용서비스(5.9%, 1.8조)에 배정된 예산은 4.6조(15.1%)로, 2020년 총예산 33.6조 중 3.6조(10.9%)로 증가하였다(고용노동부, 2023).

KCI 논문 | 제목: 영농조합법인 제주... | 저자: 박현철(성... | 발행년: 2023.04

여기에는 '정시 퇴근', '생산성 위주의 회의', '효율적 보고', '유연한 근무' 등이 수록되어 있는데, 이는 세대를 불문하고 장시간 근무 관행, 일하는 방식, 나아가 일하는 1) 커리어 테크 플랫폼인 사람인(www.saramin.co.kr)이 2022년 1,124개 기업을 대상으로 '1년 이내 조기 퇴사 현황'에 대해 조사한 결과, 84.7%가 조기 퇴사한 직원이 있다고 답했다.

문장유사율: 0%

특히 증가한 예산의 대부분은 직원훈련과 고용서비스에 배정하여 일자리 창출뿐만 아니라 일자리 미스매치 문제를 해결하기 위하여 노력하고 있다.

문장유사율: 0%

정책과 예산을 지원하는 정부기관, 일자리 매칭을 위한 이력서를 보유하고 있는 취업 전문기관, 구직자의 학습·경험을 함양시키는 교육기관은 각자의 분야에서 최선의 노력을 하고 있다.

문장유사율: 0%

하지만, 미스매치 문제 해결을 위한 노력과 책임은 최종적으로 구직자에게 있다.

문장유사율: 0%

미스매치 문제 해결을 위해서 일자리 미스매치(수요-기업)와 직무 미스매치(공급-구직자)로 구분하여 문제의 원인 분석이 필요하다.

문장유사율: 0%

일자리 미스매치는 구직자의 능력과 기술이 일자리 요구사항과 맞지 않는 현상으로 구직자의 경험 부족, 역량 부족, 교육 및 훈련의 미비, 구인기업 현황 등이 복합적 원인이라 할 수 있으며, 교육 및 훈련 강화, 일자리 요구사항 파악을 위한 정보제공, 취업 지원을 위한 정보공유 플랫폼 제공을 통하여 많은 부분 해소되고 있다고 볼 수 있다.

문장유사율: 0%

특히 저출산·고령화로 인한 노동 인구 감소는 생산 인력의 감소로 이어져, 구직자에게 다양한 취업 기회를 제공할 것이다.

문장유사율: 0%

직무 미스매치는 직무수행과 관련된 역량과 기술이 해당 직무에서 요구하는 것과 맞지 않는 현상으로 특정직무에 대한 이해부족, 관련 경험 부족, 기술적인 역량의 부족, 조직의 불명확한 역할 수행 등을 원인으로 볼 수 있으며, 구직자가 스스로 직무분석과 역할 수행을 위한 학습경험을 사전 탐색하기에는 역부족인 상황이다.

출처 표시 문장

문장유사율: 0%

이러한 문제점을 해소하기 위해서는 일자리 예산 중 미스매치 해소를 위한 예산을 확대해 나가고, 구인업체와 훈련기관이 함께 참여하는 인력양성 프로그램 개발이 필요하며 구직자 스스로 직무를 탐색하고 학습 계획을 수립할 수 있는지원 시스템이 필요하다(박재홍, 2020).

문장유사율: 0%

본 연구의 목적은 대규모 데이터에서 복잡한 패턴을 학습하는데 우수한 성능이 검증된 딥러닝 모델중 Transformer를 사용하여, 이력서 데이터를 기반으로 구직자에게 적합한 직무를 추천하는 모델의 성능을 확인하는 것이 목적이다.

문장유사율: 0%

또한, 독립 변수인 이력서 정보와 종속 변수인 추천 직무 간의 관계를 학습하여 새로운 이력서에 대한 적합한 직무를 추천하는 것에 있다.

문장유사율: 32%

II. 선행연구 검토 및 연구가설 설정 2.1 딥러닝(Deep Learning) 딥러닝(Deep Learning)은 기계학습 기법의 일종으로 심층 신경망(Deep NeuralNetwork, DNN)이라고도 한다.

KCI 논문 | 제목 : 직무몰입이 지식은... | 저자 : 김현우(울... | 발행년 : 2021.04

II. 선행연구 검토 및 연구가설 설정 2.1 직무몰입과 지식영역성 2.2.1 직무몰입의 양면성 직무몰입은 조직행동 분야에서 많은 연구가 이루어진 주제이므로 연구자들에 따라 다양하게 정의된다.

KCI 논문 | 제목 : 경영권 집중화 수... | 저자 : 이승현(경... | 발행년 : 2023.08

마지막 V장에서는 요약 및 결론을 제시한다. III. 선행연구 검토 및 연구가설 설정 2.1 경영권 집중화 수준과 관련된 선행연구 Finkelstein(1992)는 최고경영자가 갖는 파워를 네 가지 측면(dimension)으로 분리하여 제시하였다.

KCI 논문 | 제목 : 온라인 쇼핑 추천... | 저자 : 짜오권(경... | 발행년 : 2023.08

II. 선행연구 검토 및 연구가설 설정 2.1 온라인 쇼핑 추천 서비스 추천 서비스란 사용자별 정보나이용기록 등을 바탕으로 추천하는 것을 말하며(김민주·김민균, 2018), 온라인 쇼핑에서의 추천서비스는 해당 고객의 검색 또는 구매 정보에 기반하거나 사이트내 유사 고객들이 구매한 제품들을 추천하는 것을 의미한다(박병자·최선형, 2018).

문장유사율: 0%

심층 신경망은 인공 신경망(Artificial NeuralNetwork, ANN)의 일종이지만 은닉층을 추가하여 비선형적인 문제에 대한 해결방법을 제시하여 주목받게 되었다.

문장유사율: 0%

딥러닝은 뉴런 네트워크가 뇌의 동작을 모방하여이를 알고리즘으로 만든 것으로 네트워크를 구성 하기 위해 층(Layer)으로 연결되는 많은 뉴런이라는 노드를 포함한다.

## 출처 표시 문장

문장유사율: 0%

최소한의 층으로 입력층과 출력층을 가져야 하며, 더 많은 정보를 얻기 위해서는 여러 은닉층을 두어 심층 신경망으로 구성할 수 있다(우영준 외, 2019).

문장유사율: 0%

신경망의 학습을 위한 알고리즘의 필수적인 요소는 활성화 함수, 손실 함수, 최적화 기법이다.

문장유사율: 0%

신경망은 입력값을 가중치와 선형대수 연산을 하고 결과의 값을 출력 형태로 변환하는데 가중 합산한 값을 활성화 함수에 입력하여 출력으로 변환한다.

## 출처 표시 문장

문장유사율: 0%

활성화 함수는 수학적으로 선형 결합한 입력값들을 다양한 형태의 비선형 또는 선형 결합으로 변환하는 역할을 한다(이창기 외, 2014).

문장유사율: 0%

최종 출력층이 아닌 은닉층에서 학습의 효율 및 성능을 위해 사용하는 활성화 함수로는 소프트맥스(Softmax) 함수도 사용하지만 하이퍼볼릭 탄젠트(Hyperbolic Tangent, tanh), 렐루(Rectified Linear Unit, ReLU) 등 그 특성에 따라 다양한 함수가 있다.

## 출처 표시 문장

문장유사율: 0%

또한, 학습하고자 하는 목표의 최종 정답 유형에 따라 사용하는 마지막 출력층의 분류를 위한 활성화 함수로는 이진 분류를 통한 예측의 경우 시그모이드(Sigmoid) 함수, 다중 분류를 위해서는 통상적으로 소프트맥스 함수를 사용한다(조윤환 외, 2016).

문장유사율: 0%

신경망 학습은 최종 출력층의 예측값과 정답값과의 차이를 줄이기 위해 순전파(Forward Propagation)와 역전파(Back Propagation)를 연결하고, 손실 함수의 그래디언트(Gradient) 계산을 통하여 가중치를 업데이트하는 과정을 반복하는 것이며, 이때 예측값과 정답값의 차이를 수치화 해주는 함수를 손실 함수(Loss Function)라고 한다.

문장유사율: 0%

역전파는 순전파 단계에서 계산된 출력값과 실제 값의 오차를 역방향으로 전파하여 각 노드의 그래디언트를 계산하고 가중치를 업데이트하는 과정을 반복하므로써 신경망의 예측 오차가 최소화될 수 있도록 한다.

문장유사율: 0%

즉, 손실 함수를 통하여 계산된 차이값을 줄여가는 역전파 학습은 최적화 방법(Optimizer)의 적용에 따라 신경망 모델의 성능에 영향을 미친다.

## 출처 표시 문장

문장유사율: 0%

주로 사용하는 최적화 방법은 기울기의 반대 방향으로 일정 크기만큼 이동하며 연산을 반복하여 손실 함수의 값을 최소화하는 가중치(W)와 편향(b)을 찾는 기법인 경사 하강법 계열의 방법과 Momentum 최적화 기법, RMSProp 기법, Adam 기법 등이 있다(Yazan & Talu, 2017).

문장유사율: 0%

트랜스포머 모델을 기반으로 한 직무추천에 관한 연구 2.2 텍스트 임베딩(Text Embedding) 텍스트 임베딩(Text Embedding)은 텍스트 마이닝(Text Mining)이나 자연어를 처리하는 분야에서 활용되고 있는 기법으로 하나의 단어 또는 문장을 인공 신경망을 적용하여 숫자 형태의 벡터로 변환하는 작업을 의미한다.

문장유사율: 0%

기존에는 단어 자체를 아스키코드(ASCII)나 유니코드(Unicode)로 처리해서 사용했지만, 이러한 숫자 코드만을 가지고는 단어의 실제적인 의미를 추론하기는 어렵다.

## 출처 표시 문장

문장유사율: 0%

예를 들어 '왕'과 '여왕'이 관련이 있는 단어이고 '왕'의 성별은 남성이고 '여왕'의 성별은 여성이라는 사실을 기존의 숫자 코드만을 가지고는 알아내기 어렵다(Li et al., 2015).

문장유사율: 0%

텍스트 임베딩의 연구는 텍스트 마이닝의 문제점을 개선하기 위하여 진행된 것이라 할 수 있으며, 텍스트 마이닝은 비정형 데이터인 자연어를 분석하여 의미 있는 정보를 추출하는 방법이다.



문장유사율: 89%

응용 분야로는 문서 분류(Document Classification), 문서 군집(Document Clustering), 정보 추출(Information Extraction), 문서 요약(Document Summarization) 등이 있다.

KCI 논문 | 제목: 유비쿼터스 환경의... | 저자: 김재수(서... | 발행년: 2012.12

주요 응용 분야로 문서 분류(Document Classification), 문서 군집(Document Clustering), 정보 추출(Information Extraction), 문서 요약(Document Summarization) 등이 있다.

KCI 논문 | 제목: 소셜데이터 분석방... | 저자: 김휴찬(제... | 발행년: 2014.06

주요 응용 분야로 문서 분류(Document Classification), 문서 군집(Document Clustering), 정보 추출(Information Extraction), 문서 요약(Document Summarization) 등이 있다.

KCI 논문 | 제목: 텍스트마이닝 기반... | 저자: 서상욱(가... | 발행년: 2018.06

주요 응용 분야로 문서 분류(Document Classification), 문서 군집(Document Clustering), 정보 추출(Information Extraction), 문서 요약(Document Summarization) 등이 있다.

출처 표시 문장

문장유사율: 0%

하지만, 대용량 텍스트의 경우 차원이 높고 희소성 문제가 있어서 단어의 의미를 잘 파악하지 못하고 문맥을 고려하지 못하는 한계가 있다(주용한 외, 2020).

문장유사율: 0%

텍스트 임베딩은 이러한 텍스트 마이닝의 한계를 극복하기 위해 개발된 기술이며, 단어나 문장을 저차원의 실수 벡터로 표현하여 의미와 문맥을 더욱 잘 이해할 수 있도록 한다.

문장유사율: 0%

또한, 사전 훈련된 임베딩 모델을 사용하여 단어의 의미적 유사성과 관련성을 반영하는 벡터 표현을 생성하고, 이를 활용하여 텍스트 마이닝에 적용하여 성능을 향상할 수 있다.

문장유사율: 42%

텍스트 임베딩은 NPLM을 비롯한 다양한 알고리즘과 모델을 통해 발전해 왔으며, Word2Vec은 NPLM의 개선된 형태로, 단어 간 유사도를 반영하여 단어를 벡터화할 수 있는 대표적인 방법이다.

KCI 논문 | 제목: 워드 임베딩을 이... | 저자: 장원중(가... | 발행년: 2020.12

워드투벡터는 단어 간 유사도를 반영하여 단어를 벡터화할 수 있는 대표적인 방법이다.

출처 표시 문장

문장유사율: 90%

Word2vec은 인공지능망 알고리즘이고, 이를 구현하는 방법에는 CBOW(Continuous Bag of Words)와 Skip-Gram 두 가지 방식이 있다(장원중, 2020).

KCI 논문 | 제목: 워드 임베딩을 이... | 저자: 장원중(가... | 발행년: 2020.12

Word2vec은 인공지능망 알고리즘이고, 이를 구현하는 방법에는 CBOW(Continuous Bag of Words)와 Skip-Gram 두 가지 방식이 있다(Mikolov et al, 2013a; Mikolov et al, 2013b; Mikolov et al, 2010).

문장유사율: 0%

2.3 트랜스포머(Transformer) 트랜스포머(Transformer)는 입력과 출력 시퀀스 사이의 양방향 정보교환을 가능하게 하여, 문장의 의미를 파악하고, 번역이나 요약과 같은 작업을 수행한다.

문장유사율: 0%

Transformer의 주요 구성 요소로는 어텐션(Attention), 셀프 어텐션(Self Attention), 인코더(Encoder), 디코더(Decoder)가 있다.

문장유사율: 0%

어텐션은 입력 시퀀스의 각 위치에서 다른 위치의 정보에 가중치를 부여하여 중요한 정보를 집중하도록 하는 메커니즘이며 셀프 어텐션은 입력 시퀀스 내 다른 위치의 정보 간의 관련성을 계산하는데 사용하여 문장내의 단어 간의 의미적인 연결을 파악한다.

문장유사율: 0%

인코더는 입력 시퀀스를 처리하는 부분으로, 다수의 층으로 구성되어 있다. 각 층은 어텐션과 피드포워드 신경망(Fully Connected Neural Network)으로 구성되어 입력 시퀀스의 특징을 추출한다.

문장유사율: 0%

디코더는 출력 시퀀스를 생성하는 부분으로서 인코더와 비슷한 구조를 가지고 있지만, 더불어 인코더의 출력에 셀프 어텐션을 추가로 활용하여 입력과 출력 간의 상호 의존성을 고려한다.

문장유사율: 0%

Transformer는 병렬 처리가 용이하며, 긴 시퀀스 처리에도 효과적이며, 층이 쌓이는 구조로 인해 깊은 네트워크를 구축할 수 있고, 복잡하고 풍부한 문맥 정보를 다룰 수 있다.

문장유사율: 0%

특히, 번역, 요약, 질문 응답 등의 자연어 처리 분야에서 혁신적인 모델 주목받고 있으며, 다양한 언어 처리 작업에서 우수한 성능을 보여준다.

문장유사율: 0%

[그림 2-1] 트랜스포머 모델 구조 [그림 2-1]는 트랜스포머 모델 구조를 보여주며, 트랜스포머 모델은 기본적으로 인코더-디코더 구조로 되어 있으며, 큰 특징중 하나는 셀프 어텐션을 수행한다는 것이다.

출처 표시 문장

문장유사율: 0%

트랜스포머 모델은 기계번역 시스템에서 악명 높은 문제인 상호 참조를 셀프 어텐션을 통해 해결 하였다(오소진, 2022).

문장유사율: 0%

2.4 어텐션 매커니즘(Attention Mechanism) Transformer는 시퀀스 데이터를 처리하고 문맥을 이해하는 능력을 향상시키기 위해 어텐션 매커니즘(Attention Mechanism)을 활용한다.

문장유사율: 0%

어텐션은 각 시퀀스 요소에 가중치를 할당하여 관련 있는 정보에 더 많은 영향력을 부여한다.

문장유사율: 70%

어텐션 매커니즘은 자연어 기계 번역을 위한 Seq2Seq(Sequence-To-Sequence) 모델에 처음 도입되어, 디코딩 과정에서 현재 스텝에서 가장 관련된 입력 부분에 집중할 수 있도록 해줌으로써 기계 번역의 품질을 크게 향상시켰다.

KCI 논문 | 제목 : 양방향 인제매칭을... | 저자 : 오소진(한... | 발행년 : 2021.08

어텐션 매커니즘은 seq2seq모델이 디코딩 과정에서 현재 스텝에서 가장 관련된 입력 부분에 집중할 수 있도록 해줌으로써 기계 번역의 품질을 크게 향상시켰다.

KCI 논문 | 제목 : 차량 DTC 고장... | 저자 : 김한숨(홍... | 발행년 : 2023.09

구에서 사용된 모델은 1D-CNN-Dropout-ooling-LSTM로 구성하였으며, 모델에 대한 Table 4와 같이 정리하였으며, 전체 구조 . 2와 같다. 4.2.3 Anomaly Transformer 마지막으로 비교적 최근에 알려진 Anomaly Transformer 딥러닝 모델도 본 문제에 응용해 보았다. Transformer는 Vaswani et al. (2017)에 의해 처음 제안된 Seq2Seq(Sequence to Sequence) 모델에 attention 매커니즘을 적용한 구조이다.

출처 표시 문장

문장유사율: 48%

이 매커니즘은 다양한 작업에서 입력과 출력의 원소들 사이의 거리와 무관하게 의존성을 학습할 수 있게 되었다(Bahdanau et al., 2015; Kim et al., 2019).

문장유사율: 83%

Vaswani, A., et al.(2017)의 연구에서는 기존 순환 신경망에서 발생하는 문제들을 해결하고 셀프 어텐션(Self-Attention)을 이용하는 Transformer를 제안하였는데, 이 연구에서는 반복(Recurrence)을 제거함으로써 오직 어텐션 매커니즘만을 이용해서 입력과 출력 사이의 전역의존성(Global Dependency)을 학습하게 된다.

KCI 논문 | 제목 : 양방향 인제매칭을... | 저자 : 오소진(한... | 발행년 : 2021.08

또한 다양한 작업에서 강력한 시퀀스 모델 및 변환 모델의 필수적인 부분이 되었으며, 입력과 출력의 원소들 사이의 거리와 무관하게 의존성을 학습할 수 있게 되었다

KCI 논문 | 제목 : 양방향 인제매칭을... | 저자 : 오소진(한... | 발행년 : 2021.08

2017 연구에서는 기존 순환 신경망에서 발생하는 문제들을 해결하고 Self-attention을 이용하는 트랜스포머(Transformer를 제안하였는데 이 연구에서는 반복(Recurrence)을 제거함으로써 오직 어텐션 매커니즘만을 이용해서 입력과 출력 사이의 전역의존성(Global dependency)을 학습하게 된다.

KCI 논문 | 제목 : 차량 외부 디자인... | 저자 : 김태연(고... | 발행년 : 2022.10

Vaswani et al.(2017)은 특정 영역의 입력 값을 활용하여 학습에 중요한 부분에 큰 가중치를 주어서 학습을 진행하는 셀프 어텐션(self-attention)을 활용하여 자연어 처리 학습 과정에서 중요한 역할을 한 부분을 탐지하였다.

문장유사율: 0%

반복 작업의 제거로 더 손쉽게 병렬 처리가 가능하고 훈련 시간을 단축하였다.

출처 표시 문장

문장유사율: 0%

셀프 어텐션은 독해, 추상적 요약, 텍스트 수반, 학습과제 독립적인 문장 표현을 포함한 다양한 과제에서 성공적으로 사용되었다(Cheng et al., 2016; Parikh et al., 2016; Paulus et al., 2017; Lin et al., 2017).

문장유사율: 35%

III. 연구방법 3.1 연구모형 및 변수측정 본 연구에서는 구직자에 맞춤형 직무 추천시스템을 구축하기 위해 텍스트 임베딩을 적용하여 자연어로 구성된 이력서 데이터를 Embedding Vector로 추출하여 분석하였다.

KCI 논문 | 제목 : 직무물입이 지식은... | 저자 : 김현우(울... | 발행년 : 2021.04

III. 연구방법 3.1 연구모형 및 변수측정 본 연구에서는 직무물입과 지식은폐행동과의 관계에서 지식영역성의 매개효과와 서번트 리더십의 조절효과를 탐색할 것이다.

KCI 논문 | 제목 : 경영권 집중화 수... | 저자 : 이승현(경... | 발행년 : 2023.08

III. 연구방법 3.1 연구모형 및 변수측정 본 연구에서는 경영권 집중화 수준이 판매관리의 하방경직적 원가행태에 미치는 영향에 대해 분석한다.

KCI 논문 | 제목 : 온라인 쇼핑 추천... | 저자 : 짜오쥔(경... | 발행년 : 2023.08

III. 연구방법 3.1 연구모형 및 변수측정 본 연구에서는 2장에서 제시된 연구가설들을 종합하여 [그림 1]과 같은 연구모형을 제시하고, 온라인 쇼핑의 만족도와 충성도 측면에서 추천 서비스 성공 요인을 서로 다른 문화적 차이를 가지고 있는 한국과 중국 표본을 통해 비교 분석하고자 한다.

문장유사율: 0%

이를 통해 직무 간 유사성 및 이를 활용한 딥러닝 기반 직무 추천 모델을 연구하였다.

문장유사율: 0%

본 연구는 이력서의 특징을 딥러닝 기반의 임베딩 기법을 이용하여 독립 변수화한 후 종속 변수인 "JOB\_CODE"를 통해 직무를 추천하는 것이다.



문장유사율: 0%

이를 위하여 이력서의 특징을 52개와 34개의 두 가지 경우로 분류하여 각각의 데이터셋과 데이터 전처리 및 학습 데이터셋을 구축하고 Feature Engineering을 진행한다.

문장유사율: 0%

이후 Transformer 모델을 사용하여 두 가지 경우의 직무 추천 모델을 학습하고 정확도와 정밀도를 비교하였다.

문장유사율: 0%

[그림 3-1]은 본 연구의 직무 추천시스템의 연구모형으로 데이터 분석, 데이터 전처리, 모델 개발, 추천모델, recommendation 순서로 진행되었다.

문장유사율: 0%

[그림 3-1] 직무 추천 시스템 연구 모형 3.2 표본선정 취업포털로부터 연구 목적으로 받은 총 370,824개의 이력서 데이터 중, 이상치 및 결측치를 제거한 구직자 298,465명의 이력서 데이터를 항목별로 전처리 완료한 학습데이터를 기준으로 연구를 진행하였다.

문장유사율: 0%

취업포털이 보유한 이력서 데이터를 익명화 처리하여 사용하였으며, 데이터 항목은 [그림 3-2]와 같이 인적사항, 희망직무 및 업종, 학력사항, 경력사항, 자격 사항 등 5가지로 구성하였다. '이력서 번호, 지원자 코드, 이력서 등록일, 이력서 수정일'

문장유사율: 0%

은 인적사항으로 구분되며, '희망직종, 희망직종 세부, 희망직무, 희망근무 지역, 희망 업종, 희망연봉'은 희망직무로 구분된다. '고등학교, 대학교, 학력구분, 편입여부, 입학연도, 졸업연도, 전공, 부전공, 계열, 평점, 만점, 최종학력'

문장유사율: 0%

은 학력사항으로 구분되며, '회사명, 업종, 근무형태, 부서명, 직위, 담당업무, 근무시작일, 근무종료일, 프로젝트명, 참여기간, 의뢰업체, 내용 및 성과, 전체경력, 최종연봉, 퇴직사유'는 경력 사항으로 구분된다. '언어, 공인시험, 공인시험점수, 자격증명'

문장유사율: 0%

은 자격 사항으로 구분된다. 직무 추천 모델의 종속 변수는 '희망직종', '희망직종 세부직종'이며, 두가지의 데이터를 함께 사용하기 위해 두가지의 데이터를 군집화한 뒤 해당 내용을 Labeling 하여 코드값을 부여하였다.

문장유사율: 0%

이렇게 Labeling 된 데이터를 'JOB\_CODE'라고 지정하여 종속변수로 사용하였다. 즉, '희망직종', '희망직종 세부직종' 컬럼을 제외한 모든 컬럼이 원인변수가 되어 'JOB\_CODE'라는 결과값을 만들어낸다. [그림 3-2] 이력서 데이터 정의 'JOB\_CODE'는 <표 3-1>과 같이 취업포털의 직무체계를 활용하여 총 125개로 Labeling 하였다.

문장유사율: 0%

<표 3-1> JOB\_CODE 생성 희망직종 희망직종 세부직종 Labeling IT·게임 HTML코딩 1 IT·게임 게임 기획 2 IT·게임 게임 디자인 3 ..... 문화·예술 신문·방송 아나운서 ·리포터 ·VJ ·성우 125 3.3 데이터 전처리 데이터 전처리는 1단계 결측치 처리, 2단계 이상치 처리, 3단계 데이터 코드화로 나누어 진행하며 학습 데이터셋 구축을 위한 이전 단계이다.

문장유사율: 0%

<표 3-2>의 1단계 결측치 처리는 Null 값이 데이터에 있을시 딥러닝 모델에 입력이 되었을때 문제가 될 수 있어, 결측치를 모두 제거하거나 값을 넣어주는 방법을 사용하여 처리하였다.

문장유사율: 0%

결측치가 없는 경우 이력서 데이터를 그대로 사용하며, 결측치가 있는 경우에도 데이터의 타입이정수(Integer)인 경우 숫자 0으로 결측치를 입력하였으며, 데이터 타입이 문자열일 경우 Unknown(UNK)으로 결측치를 입력하였다.

문장유사율: 0%

<표 3-2> 결측치 처리 항목 구분 숫자(0) UNK 결측치 처리 항목 등록일, 수정일, 근무형태, 희망연봉, 최종연봉, 근무지역, 최종학력, 편입여부, 입학연도, 졸업연도, 계열, 평점, 업종, 회사명, 근무시작일, 근무종료일, 전체경력, 부서명, 직위, 담당업무, 퇴직사유, 내용 및 성과, 언어, 공인시험, 점수 전공 부전공 자격증 <표 3-3>의 2단계 이상치 처리는 각 데이터의 특성에 따라 ① 불필요한 문자 삭제, ② 데이터 타입 변경, ③ 컬럼 삭제, 총 3가지의 방법을 활용하였다.

문장유사율: 0%

3단계 데이터 코드화는 각 데이터의 특성에 따라 자체코드화, 표준코드화, 수치계산, 기타의 총 4가지 방법을 활용하였다.

문장유사율: 0%

본 연구에서 데이터 코드화는 직무 추천 모델의 경량화를 통한 학습 속도 개선을 위하여 데이터 전처리에 포함하였다.

문장유사율: 0%

〈표 3-3〉 이상치 처리 프로세스 idx 이상치 처리 세부내용 1 불필요한 문자 삭제 영문, 한글, 숫자등 다양한 데이터들이 혼재되어 있는 상태에서 필요한 문자열만 추출하여 원본데이터에 입력한다.

문장유사율: 0%

2 데이터 타입 변경 원본 데이터의 데이터 타입을 필요한 형태로 변경함. 즉, 날짜 데이터인 경우 String 값으로 정의가 되어 있을 때 데이터 타입을 변경하여 DATETIME 타입으로 변경하여 날짜 데이터로 활용한다.

문장유사율: 0%

3 컬럼 삭제 원본 데이터 중 데이터의 모수가 없거나 혹은 불필요한 데이터일 경우 컬럼을 삭제하여 진행한다.

문장유사율: 0%

대부분 명목형 변수로 구성되어 있는 이력서 데이터의 학습 속도를 개선하기 위하여 수치형 변수로 변환하고 자체 코드 매핑과 표준 코드화에 따라 〈표 3-4〉 데이터 전처리 내역에 표기된 데이터 항목에 대하여 전처리를 진행하여 최종적인 학습데이터를 생성하였다.

문장유사율: 0%

〈표 3-4〉 데이터 전처리 내역 데이터 항목명 전처리 전처리 내용이력서 등록일

문장유사율: 0%

수치계산이력서 등록일과 수정일 차이계산이력서 수정일수치계산이력서 등록일과 수정일 차이계산 희망 직종

문장유사율: 0%

자체코드 자체 코드화 진행 희망 세부직종 자체코드 자체 코드화 진행 희망 업종

문장유사율: 0%

자체코드 자체 코드화 진행 전체경력 수치계산 전체 경력기간의 '일 수'를 계산

문장유사율: 0%

근무형태 자체코드 정규직 활용하여 자체 코드 테이블 구성 후 자체 코드화 진행 / 근무형태 값이 복수일 경우 제일 좋은 값(정규직)만 유지희망연봉자체코드

문장유사율: 0%

군집화 후 연봉정보 추출 / 추출된 데이터를 범주형으로 구성된 자체 코드 테이블을 활용

문장유사율: 0%

최종연봉자체코드 근무지역1 자체코드 복수의 값은 근무지역1, 근무지역2, 근무지역3으로 분리 / 자체 코드화 진행

문장유사율: 0%

근무지역2 자체코드 지역에 따른 코드테이블 구성 학력구분 자체코드 자체 코드화 진행

문장유사율: 0%

고등학교 자체코드 표준 코드 테이블 활용 학교명 자체코드 표준 코드 테이블 활용

문장유사율: 0%

편입여부 자체코드 자체 코드화 진행 졸업년도 자체코드 자체 코드화 진행

문장유사율: 0%

졸업기간 자체코드 자체 코드화 진행 입학년도 자체코드 자체 코드화 진행

문장유사율: 0%

전공/부전공 기타 자체 텍스트 처리 알고리즘 사용 계열 자체코드 자체 코드화 진행

문장유사율: 0%

평점/만점 수치계산 백분율 계산 업종 자체코드 자체 코드화 진행 회사명 자체코드

문장유사율: 0%

자체 코드화 진행(값이 있으면1, 없으면 0) 부서명자체코드 자체 코드화 진행(값이 있으면1, 없으면 0)

문장유사율: 0%

직위 자체코드 자체 코드화 진행(값이 있으면1, 없으면 0) 담당업무 자체코드

문장유사율: 0%

자체 코드화 진행(값이 있으면1, 없으면 0) 퇴직사유 자체코드 자체 코드화 진행

문장유사율: 0%

내용 및 성과 자체코드 자체 코드화 진행(값이 있으면1, 없으면 0) 언어 자체코드

문장유사율: 0%

자체 코드화 진행 공인시험 점수 수치계산 점수 표준화 진행 공인시험 자체코드

문장유사율: 0%

자체 코드화 진행 자격증 기타 자체 텍스트 처리 알고리즘 사용 3.2 학습 데이터셋 구축

문장유사율: 0%

학습 데이터셋은 구축자의 적합한 직무뿐만 아니라 부적합한 직무도 포함하여 학습하기 위해서 1:1 Negative Sampling 방식으로 전체 이력서 데이터 370,824개에서 이상치 및 결측치 처리를 완료한 298,465개를 Positive Sample 298,465개, Negative Sample 298,465개를 합쳐서 총 596,930개의 데이터로 구성하였다.

문장유사율: 0%

Positive Sample, Negative Sample을 구분하기 위해 'ANS' 컬럼을 설정하여 이진분류 하였다.

문장유사율: 0%

구직자가 선호하는 Positive sample 데이터를 ANS = 1로 선정하고 구직자가 선호하지 않은 Negative sample 데이터를 ANS = 0로 선정하였다.

문장유사율: 0%

즉, 이력서 데이터 전체를 ANS = 1로 설정하고, 복사한 데이터를 ANS = 0로 설정하여 두개의 데이터를 하나의 학습 데이터 셋으로 구성하였다.

문장유사율: 0%

모델의 학습 단계에서는 train\_set, validation\_set, test\_set 6 : 2 : 2로 데이터를 분할하여 사용하였다.

문장유사율: 0%

<표 3-5> 학습 데이터셋 데이터 개수 데이터셋 데이터 수 이력서 데이터 298,465개 학습 데이터 셋 구축 1:1 Negative sampling 전체 : 596,930 긍정 사례 : 298,465개 / 부정사례 : 298,465개 훈련 데이터 477,544개 (80%) 시험용 데이터 119,386개 (20%) 3.2 피처 엔지니어링(Feature Engineering) 학습 데이터셋 구축 후 모델 학습을 하기 위해 변수별로 feature tuning을 진행하였다.

문장유사율: 0%

변수는 소범주형 변수, 다범주형 변수, 자연어 변수로 나뉘어 진행하며, 소범주형 변수는 Category-Encoding, 다범주형 변수는 Embedding, 자연어 변수는 사전학습모델을 이용하였다.

문장유사율: 0%

<표 3-6> 변수별 Feature Tuning 소범주형 변수다범주형 변수 자연어 변수 추론 Category-Encoding Embedding 사전학습모델 용도 범주형 변수를 수치형 변수로 변환 텍스트 데이터를 수치형 데이터로 변환 대규모 데이터를 사전에 학습된 모델 Data Sets (Details) 대학 설립구분, 대학구분, 대학교LOC, 최종학력, 고등학교세부유형, 고등학교 LOC, 대학교LOC, 계열, 근무형태, 업종대분류, 업종소분류, 직위, 퇴직사유, 근무지역 고등학교명, 대학교명, 담당업무, 언어, 공인시험, JOB\_CODE 전공, 자격증 소범주형 변수는 사전에 정의된 라벨(Label)이 30개 이하인 변수이며 학습데이터는 Label Encoding 되어 있는 형태로 입력되며, Category Encoding 층을 거쳐 One-hot Encoding 된 형태로 변환되어 딥러닝 모델에 입력하였다.

문장유사율: 0%

소범주형 변수의 주요 Layer 층은 Category Encoding이며, 해당 Encoding은 정수의 특징을 인코딩하는 전처리 계층이다.

문장유사율: 0%

다범주형 변수는 사전에 정의된 라벨(Label)이 30개 이상인 변수이며 One-hot Encoding으로 처리할 경우 차원의 수가 너무 커져 데이터 희박성(Sparsity) 문제가 발생할 가능성이 높은데이터 형태를 의미한다.

문장유사율: 0%

이러한 다범주형 변수의 경우 모델 내에서 직무 추천을 위한 적절한 임베딩 공간의 탐색이 필요하다.

문장유사율: 0%

특히 본 모델에서 사용되는 다범주형 변수의 경우 범주의 개수가 1,000개 이상인 경우가 다수 존재한다.

문장유사율: 0%

이를 위하여 임베딩 층을 설계하여 다범주형 변수를 적절한 벡터 공간에 표현할 수 있도록 하는 한편, 각 다범주형 변수별로 의미 있는 벡터 공간의 탐색을 위해 직무 추천에 필요한 특성(Feature)만을 추출하기 위한 층을 추가로 설계하여 반영하였다.

문장유사율: 0%

즉, 다범주형 변수를 먼저 큰 벡터 공간에 표현하는 임베딩 층을 통과시켜 벡터화를 진행하고 임베딩 층의 결과를 직무 추천에 적합한 수치로 변환하기 위한 1D 컨브넷(Conv1D) 구조를 반영하였다.

문장유사율: 0%

1D 컨브넷(Conv1D)은 합성곱 연산의 커널(Kernal) 개념을 도입하여 1D 패치(부분 시퀀스)를 추출하여 합성곱 연산을 적용한다.

문장유사율: 0%

그에 따라 시퀀스에 있는 지역 패턴을 인식할 수 있어 각 범주가가지는 시퀀스를 고려하면서 직무 추천에 도움이 되는지역 패턴의 식별이 가능하다.

문장유사율: 0%

이를 통해 원천 다범주형 변수는 임베딩 층을 통과하여 다차원의 벡터 공간에 수치로 표현된다.

문장유사율: 0%

다범주형 변수의 주요 Layer 층은 Embedding, Conv1D로 구성되어 있으며, Conv1D 작업 전에 단어를 밀집 벡터로 만드는 역할을 하는 Embedding 작업이 필요하다.

문장유사율: 0%

Embedding은 정수인코딩이 되는 단어를 임베딩 된 값으로 출력한다. 자연어 처리 알고리즘인 유니버설 센텐스 인코더(Universal sentence encoder)를 통하여 다범주형 변수 처리 방법을 사용하지 못하는 자연어 변수를 처리하고, 신규 전공 또는 전공간의 결합 등의 데이터 변화에도 강건한 모델을 설계하기 위하여 적용하였다.

문장유사율: 0%

특성(Feature) 결합 및 추천을 위한 벡터 공간 탐색 과정은 각 변수 대분류별로 표현된 34개의 수치를 하나로 결합하고 직무 추천을 위한 벡터 공간을 탐색하는 과정이다.

문장유사율: 0%

이를 위해 대분류별 34개의 특징을 결합하고이어 완전 연결층(Fully Connected Layer)을 연결하여 최종 출력값을 추론할 수 있는 벡터 공간을 탐색할 수 있도록 유도한다.

문장유사율: 0%

추천을 위한 벡터 공간 탐색은 모델의 출력값의 형태에 따라 상이하게 설계되며, 최종 출력값의 차원보다 더 큰 벡터공간의 탐색을 하여 정보의 손실을 최소화는 딥러닝 모델을 설계하였다.

문장유사율: 0%

3.3 하이퍼 파라미터(Hyperparameter) 하이퍼 파라미터는 최적의 학습 모델을 구현하기 위해 설정하는 변수로 가중치 초기화, 학습률, 최적화, 학습반복, 학습 배치 크기, 임베딩 차원, 은닉층의 개수 등이 있으며, 하이퍼파라미터의 튜닝 기법은 그리드 탐색, 랜덤 탐색, 베이지안 최적화, 휴리스틱 탐색 등이 있다.

문장유사율: 0%

하이퍼 파라미터는 모델의 매개 변수를 추정하는 프로세스로서 개발자가 임의로 조정할 수 있는 매개변수이며, 하이퍼파라미터의 최적값은 데이터 분석 결과에 의해 결정되는 것이 아니므로 절대적인 최적값은 존재하지 않는다. <표 3-7>는 하이퍼 파라미터 설정 변수이다.

문장유사율: 0%

〈표 3-7〉 하이퍼 파라미터 항목 Idx 하이퍼 파라미터 선정방법 1 가중치 초기화

문장유사율: 0%

Xavier 초기화 기법과 he 초기화 기법을 사용 2 학습률 임의의 학습률 선정 후 반복적 미세조정

문장유사율: 0%

3 최적화 RMSProp, Adam 중 최고 성능 최적화 기법 선정 4 학습반복 검증 손실(Validation loss) 값의 2회 초과 증가시 학습반복 중단(조기종료 조건을 통한 학습 반복 횟수의 선정

문장유사율: 0%

5 학습 배치 크기 16, 64, 128, 256, 512의 각각 다른 사이즈로 반복 학습을 통한 최고 성능의 크기 선정

문장유사율: 0%

6 임베딩 차원 50, 100, 200, 400의 각각 다른 임베딩 차원을 입력 임베딩으로 선정하여 반복학습 후 최고 성능의 임베딩 차원 선정

문장유사율: 0%

7 은닉층 개수 네트워크 구조(은닉층)을 설정하여 가중치를 조절 3.4 딥러닝 모델 학습

문장유사율: 0%

본 연구에서는 그리드 탐색 실험으로 선정된 하이퍼 파라미터를 Transformer 모델에 적용하여 학습하였다.

문장유사율: 0%

활성화 함수는 각 모델의 은닉층에 따라 일부 차이가 있으며, 손실 함수는 이진 분류 Binary Cross Entropy를 이용하였다.

문장유사율: 0%

Transformer 전체 모델 중에서 '언어' 컬럼은 1개의 입력층과 4개의 은닉층, 1개의 출력층으로 구성한다.

문장유사율: 0%

입력층은 정수화한 텍스트 데이터를 입력 받아서 3차원의 임베딩 벡터를 출력하는 임베딩층으로 구성한다.

문장유사율: 0%

은닉층 중 첫 번째 층을 7개의 노드를 가진 TFOLambda 모델로 학습하고 두 번째 층은 7개의 노드를 가진 Transformer 모델중 Multi-Head Attention 모델로 학습한다.

문장유사율: 0%

Flatten 함수로 1차원데이터로 가공한 뒤 텐스층으로 구성한다. 활성화 함수인 Sigmoid 함수를 거쳐 출력층으로 수치가 산출된다.

문장유사율: 10%

[그림 3-3] Transformer 모델 언어 FeatureDense Layers **IV. 실증분석 결과 4.1** 실험 환경 및 도구 본 연구에 사용된 실험 장비의 운영체제는 Ubuntu 20.04.5 LTS, CPU는 Silver 4210, GPU는 Nvidia Quadro A6000 48GB, Memory는 64GB이며, 실험도구는 Python 3.10과 Tensorflow 2.8, Keras를 사용하여 전처리 및 모델 학습을 진행하였다.

문장유사율: 0%

4.2 파인튜닝 선정 실험 본 연구에서는 상위층 파인튜닝과 부분적 파인튜닝의 방법을 사용하여 학습 모델의 성능을 높이하고자 하였으며, 상위층 파인튜닝은 사전 학습된 모델의 상위층을 새로운 작업에 맞게 조정하는 것에 중점을 두었다.

문장유사율: 0%

사전 학습된 모델의 하위층은 일반화된 특성을 학습했기 때문에 그대로 사용하고, 상위층은 새로운 작업과 관련된 특성이 반영되도록 조정하였다.

문장유사율: 0%

이를 통하여 학습 성능 향상에 필요한 세부 정보를 더욱 민감하게 학습하고자 하였다.

KCI 논문 | 제목 : 경영권 집중화 수... | 저자 : 이승현(경... | 발행년 : 2023.08

**IV. 실증분석 결과 4.1** 주요 변수들에 대한 기술통계량 본 연구의 회귀분석에서 사용된 주요 변수들에 대한 기술통계량은 다음의 〈표 1〉에 정리되어 있다.

KCI 논문 | 제목 : 직무몰입이 지식은... | 저자 : 김현우(울... | 발행년 : 2021.04

**IV. 실증분석 결과 4.1** 기초통계 및 상관분석 본 연구의 주요 변인인 직무몰입, 서번트리더십, 지식영역성, 지식은폐행동 사이의 상관관계를 알아보기 위해 상관 분석을 실시하였으며 결과는 아래 〈표 2〉와 같다. 분석 결과를 살펴보면 독립 변인인 직무몰입과 매개변인인 지식영역성의 상관관계는 통계적으로 유의(r = .18, p < .01)하였다.

문장유사율: 0%

부분적 파인튜닝은 사전 학습된 모델의 일부 층만을 파인튜닝하는 것이다. 전체 모델을 다시 학습하는 대신, 일부 층만 새로운 작업에 맞게 업데이트 된다.

문장유사율: 0%

이를 통하여 파라미터 수를 줄이고 학습 비용을 절감하고자 하였다. 또한, 학습 모델의 설계 및 실험은 데이터 관점, 분류기 관점, Layers 관점으로 나누어 진행하였다.

문장유사율: 0%

4.2.1 데이터 관점 데이터 관점의 파인튜닝은 사전 학습된 모델에 새로운 작업에 필요한 데이터를 제공하여 모델을 조정하는 것이다.

문장유사율: 0%

직무 추천 모델의 1차 실험을 분석한 결과 조직자의 특징을 충분히 추출하지 못한 것으로 확인되어, 2차 실험에서는 부서명, 담당업무 특징을 추가하여 모델을 조정하였다.

문장유사율: 0%

이를 위하여 해당 데이터에 대해 토큰화(Tokenization)하고, 한국어 형태소 분석기(Ko NLPy)를 사용하여 명사만 추출하였다.

문장유사율: 44%

그 후 **잠재 디리클레 할당(Latent Dirichlet allocation, LDA)** 모델을 사용하여 추출된 명사를 기반으로 키워드를 추출하였다.

KCI 논문 | 제목 : 토픽 모델링을 이... | 저자 : 박만희(부... | 발행년 : 2021.08

토픽 모델링은 대규모 문서 집합에 잠재된 주제를 찾아내기 위한 기법으로 다양한 알고리즘들 중 **잠재 디리클레 할당(Latent Dirichlet allocation LDA)** 알고리즘이 가장 많이 사용되고 있다(Blei, 2012).

문장유사율: 0%

부서명은 1~2개의 키워드로 이루어져 있으므로 이를 추출하여 모델 학습에 사용하였다.

문장유사율: 0%

담당업무는 여러 문장으로 이루어져 있으므로 문장의 영향력 있는 키워드 5개를 추출하였다.

문장유사율: 0%

4.2.2 분류기 관점 본 연구에 사용된 이력서 데이터의 특징(Feature)은 소범주형, 다범주형, 자연어 변수로 구분하였으며, 다범주형 변수인 경우Transformer의 임베딩을 적용하였다.

문장유사율: 0%

자연어 변수의 경우는 Universal-Sentence-Encoder와 XLM(Cross-lingual Language Model) Transformer 기반 사전 언어 학습 모델을 사용하여 모델 학습을 진행하였다.

문장유사율: 0%

4.2.3 Layers 관점 Layers 관점은 새로운 컬럼을 마지막에 Concatenate를 하는 방법과 드롭아웃(Dropout) 기법에 대해 0.5~0.8 사이의 최적값을 찾는 방법으로 딥러닝 모델을 구성하였다.

문장유사율: 0%

Concatenate는 두개 이상의 입력을 결합하는데 사용되는 연산이며, 중요도가 높은 컬럼을 마지막에 Concatenate를 하면 해당 컬럼의 가중치를 최대한 보존하면서 학습이 가능하다.

문장유사율: 0%

이에 따라 마지막에 Concatenate를 하는 경우, 해당 컬럼의 영향력이 높아질 수 있으므로, 해당 내용에 따라 중요도가 높은 컬럼은 마지막에 넣어 가중치 손실을 최소화한다.

문장유사율: 0%

2차 실험에서 추가된 "부서명", "담당업무" 컬럼에 대하여 Concatenate를 마지막에 사용하여 딥러닝 모델을 구성하였다.

문장유사율: 0%

드롭아웃은 훈련 데이터에만 과도하게 학습되어 실제 데이터에 대한 일반화 성능이 떨어지는 오버피팅(Overfitting)을 방지하는 방법으로서, 일반화 성능을 향상시킬 수 있도록 0.5~0.8 사이의 최적값을 찾아 딥러닝 모델을 구성하였다.

문장유사율: 0%

4.3 하이퍼 파라미터 선정 실험 4.3.1 가중치 초기화 기법 Xavier 초기화 혹은 Glorot 초기화 라고도 불리는 가중치 초기화(Weight Initialization) 기법은 각 계층의 출력 분산이 입력 분산과 거의 같도록 가중치를 초기화하여 그레이디언트 소실(Gradient Vanishing) 또는 폭주(Gradient Exploding) 문제를 완화하고, 모델의 학습을 안정화한다.



문장유사율: 0%

본 연구에서는 Xavier 초기화의 Normal, Uniform 두 가지 방법을 활용하여 실험을 진행한다.

문장유사율: 0%

Normal 방법은 초기화할 파라미터 값들의 범위에서 평균은 0으로 유지하고, 분산 값을 루트( $2 / (\text{Input} + \text{Output})$ )으로 조정하는 방법이다.

문장유사율: 0%

Uniform 방법은 하한값과 상한값을 지정하여 그사이 범위에서 파라미터 값을 초기화시키는 것이다.

문장유사율: 0%

본 연구 모델의 실험 결과는 Xavier\_uniform을 사용하여 학습을 진행하였을 때, 0.005 높은 것으로 확인된다.

문장유사율: 0%

〈표 4-1〉 Xavier\_uniform, Xavier\_normal Idx Xavier\_uniform Xavier\_normal 임베딩 차원 딥러닝에서 임베딩 차원은 학습된 임베딩 공간에서 특징을 표현하는데 사용되는 차원수를 나타낸다. 임베딩은 자연어 처리를 위한 신경망, 컴퓨터비전, 추천시스템 등 많은 딥러닝 모델에서 사용된다.

문장유사율: 0%

예를 들어, 자연어 처리를 위한 신경망에서 텍스트 말뭉치의 단어나 토큰은 고차원 공간에서 실수 벡터인 임베딩으로 표현될 수 있다.

문장유사율: 0%

임베딩 벡터의 크기, 즉, 임베딩 차원은 각 단어 또는 토큰을 표현하는데 사용되는 특징의 수를 결정하는 하이퍼파라미터이다.

문장유사율: 0%

임베딩 차원을 선택하는 것은 딥러닝 모델의 성능에 영향을 미칠수 있으므로 중요하다.

문장유사율: 0%

임베딩 차원이 클수록 특징을 더 복잡하게 표현할 수 있지만, 더 많은 학습데이터와 계산 리소스가 필요할 수 있다.

문장유사율: 0%

임베딩 차원이 작을수록 학습 속도가 빨라질 수 있지만, 입력 데이터의 정확도 떨어지거나 표현력이 떨어질 수 있다.

문장유사율: 0%

본 연구에서는 임베딩 차원 선정을 64, 128, 256, 512 총 4가지 차원을 활용하여 실험을 진행하였으며, 512의 차원을 적용한 실험이 가장 높은 것으로 확인된다.

문장유사율: 0%

〈표 4-2〉 임베딩 차원 Idx 학습 배치 크기 배치 크기는 네트워크를 통해 전파될 샘플의 수를 정의한다.

문장유사율: 0%

예를 들어 1,000개의 학습 데이터가 있고 배치 크기를 100으로 설정한다는 가정을 하면 알고리즘은 학습 데이터셋에서 처음 100개의 샘플(1 ~ 100)을 가져와 네트워크를 학습한다.

문장유사율: 0%

그런 다음 두 번째 100개의 샘플(101~ 200)을 가져와 네트워크를 다시 학습한다.

문장유사율: 0%

모든 학습 데이터가 네트워크에 전파될 때까지 이 절차를 계속 수행할 수 있다.

문장유사율: 0%

본 연구 모델의 학습 배치 크기 실험은 총 5가지의 실험을 진행하였으며, 학습 배치 크기 128개를 적용한 실험이 가장 높은 것으로 확인된다.

문장유사율: 0%

〈표 4-3〉 학습 배치 크기 선정 Idx 학습률 최적화도구가 학습 프로세스 중에 모델의 매개 변수를 업데이트하는 단계의 크기를 결정하는 하이퍼파라미터이다.

문장유사율: 0%

학습률은 역전파 중에 계산된 추정 오류 기울기에 따라 모델의 매개 변수가 얼마나 변화하는지 제어한다.

문장유사율: 0%

학습률이 높으면 옵티마이저(Optimizer)가 최적값을 초과하여 학습이 불안정해 지거나 동작이 달라질 수 있으며, 학습률이 낮으면 수렴 속도가 느려지고 학습 시간이 길어질 수 있다.

문장유사율: 0%

따라서 딥러닝 모델을 효과적으로 훈련하려면 적절한 학습률을 설정하는 것이 중요하다.

문장유사율: 0%

본 연구 모델의 학습률 선정은  $1e-2$ ,  $1e-3$ ,  $1e-4$ ,  $1e-5$ ,  $3e-5$  총 5가지의 실험을 진행하였으며, 학습률  $1e-2(0.01)$ 을 적용한 실험이 가장 높은 것으로 확인된다.

문장유사율: 0%

<표 4-4> 학습률 Idx 0.01 ( $1e-2$ ) 0.001 ( $1e-3$ ) 0.0001 ( $1e-4$ ) 0.00001 ( $1e-5$ ) 0.00003 ( $3e-5$ ) 최적화 기법 최적화 기법은 딥러닝 모델의 학습 과정에서 손실 함수를 최소화하기 위해 모델의 파라미터를 조정하는데 사용되는 방법이며, 모델의 출력과 실제 값 사이의 오차를 계산하여 손실 함수를 얻고, 이 손실 함수를 최소화하는 방향으로 모델의 파라미터를 업데이트한다.

문장유사율: 0%

RMSProp(Root Mean Square Propagation)는 경사하강법의 변형인 확률적 경사 하강(Stochastic Gradient Descent, SGD)을 기반으로 기울기의 제곱에 대한 이동 평균을 사용하여 학습률을 조절한다.

문장유사율: 0%

즉, 큰 기울기에는 작은 학습률이 적용되고, 작은 기울기에는 큰 학습률이 적용된다.

문장유사율: 0%

Adam(Adaptive Moment Estimation)은 이전 기울기와 제곱 기울기의 평균을 사용하여 기하급수적으로 감소하는 평균을 유지하고, 이러한 추정치를 사용하여 학습 중에 각 매개 변수에 대한 학습 속도를 조정한다.

문장유사율: 0%

이를 통해 옵티마이저(Optimizer)는 각 파라미터에 대한 학습률을 개별적으로 조정할 수 있으므로 기존 최적화 기법에 비해 더 빠르게 수렴하고 더 나은 성능을 얻을 수 있다.

문장유사율: 0%

본 모델의 실험은 RMSProp 기법과 Adam 기법 두 종류의 최적화 기법을 사용하여 실험하였다.

문장유사율: 0%

<표 4-5> RMSProp, Adam 정확도 비교 Idx RMSProp Adam 딥러닝 모델 실험 본 연구는 Transformer를 사용하여 전체 모델을 1차 실험과 2차 실험으로 구분하여 실험하였다.

문장유사율: 0%

1차 실험은 52개 특징(Feature)에서 추출된 179,589,881개의 파라미터와 하이퍼 파라미터 설정으로 학습하였다.

문장유사율: 0%

2차 실험은 1차 실험에서 사용한 52개 특징(Feature) 중 결측치가 높은 항목을 제거한 34개 특징(Feature)에서 추출된 348,629,913개의 파라미터와 파인튜닝 기법을 활용하여 학습하였다.

문장유사율: 0%

또한, 모든 실험에는 동일한 최적화 기법(RMSProp)과 손실 함수(Binary\_Crossentropy) 및 하이퍼 파라미터 설정 값을 적용하였으며, 데이터셋은 6:2:2로 분할하여 학습, 검증, 시험을 수행하도록 실험하였다.

문장유사율: 0%

<표 4-6>은 딥러닝 모델 실험의 진행에 있어 1차 실험과 2차 실험의 주요 차이점에 정리하였다.

문장유사율: 0%

〈표 4-6〉 딥러닝 모델 실험의 주요 차이점 입력층 은닉층 출력층 매개변수 기타 1차 실험 31 100 1 179,589,881 하이퍼 파라미터 2차 실험 23 135 1 348,629,913 파인튜닝 증감률 -34.78% 25.93% 0.00% 48.49% 입력층은 결측치 항목을 삭제하여 34.78%로 감소하였으며 은닉층은 복잡한 패턴과 상호작용을 학습할 수 있도록 25.95% 증가되었다.

문장유사율: 0%

또한 뉴런(Neuron)의 연결 가중치와 편향을 나타내는 매개 변수(Parameter)는 48.49% 증가되었다.

문장유사율: 0%

하이퍼 파라미터는 1차 실험의 설정 값을 그대로 2차 실험에 적용하였으며, 2차 실험에는 파인튜닝 기법을 추가하여 파인튜닝이 Transformer 모델의 성능에 미치는 영향에 대하여 확인하고자 하였다.

문장유사율: 0%

4.4.1 트랜스포머(Transformer) 〈표 4-7〉은 Transformer의 1차와 2차 실험의 결과를 정리하였다.

문장유사율: 0%

Transformer 기반의 1차 실험은 Accuracy = 0.7898, Loss = 0.4443, epoch = 39의 학습 모델이 최종 모델로 선정 되었다.

문장유사율: 0%

[그림 4-1]에서 Transformer 학습 곡선은 epoch = 39에서 검증손실(val loss)의 증가로 조기종료(Early Stopping)된 것을 나타낸다.

문장유사율: 0%

〈표 4-7〉 Transformer 실험 결과 epoch Loss Accuracy AUC val\_Loss val\_Accuracy val\_AUC 1차 2차 증감률 7% -3.3% -3.0% 4.44% 1.35% -2.2% epoch가 높을수록 모델 학습 성능이 향상되는 것은 명확하지 않지만, 1차 실험에서 동일하게 사용한 입력층, 은닉층, 하이퍼 파라미터가 Transformer 모델에는 적합하지 않을 수 있다는 추정이 가능하다.

문장유사율: 0%

Loss = 0.4443과 val\_Loss = 0.4670의 정확도가 0.0227로 근소한 차이를 보이는 것은 과적합(Overfitting) 되지 않고 일반화(Generalization) 능력이 높다는 것을 나타낸다. Transformer 기반의 2차 실험은 Accuracy = 0.7637, Loss = 0.4779, epoch = 100의 학습 모델이 최종 모델로 선정 되었다.

문장유사율: 0%

[그림 4-1] Transformer 1차 실험 그래프 [그림 4-2] Transformer 2차 실험 그래프 [그림 4-2]의 2차 실험은 설정값인 epoch=100까지 진행되어 종료되었으므로 Epoch의 설정 값을 증가하여 추가 학습을 한다면 성능지표가 더욱 상승할 것으로 예상하며, Accuracy 상승은 동일했으나 train\_loss와 val\_loss가 안정적으로 성능 향상되었다.

문장유사율: 0%

이것은 1차 실험성능보다 2차 실험성능이 개선된 것을 확인할수 있으며 그 원인으로는 파인튜닝과 특징(Feature)을 52개에서 34개로 감소한 것이 원인으로 추정된다.

문장유사율: 0%

4.5 실험 결과 분석 4.5.1 하이퍼파라미터의 영향 분석 딥러닝 모델의 하이퍼 파라미터는 모델 성능에 중요한 영향을 미치므로 하이퍼파라미터의 조정을 통해 모델의 정확도, 학습 속도 등을 개선할 수 있다.

문장유사율: 0%

본 연구실험에는 하이퍼 파라미터 조정의 항목은 가중치 초기화, 임베딩 차원 설정, 학습 배치 크기 선정, 학습률, 최적화 기법을 사용하여 실험한다.

문장유사율: 0%

먼저, 가중치 초기화 기법은 Xavier\_uniform을 사용하여 학습되며, 임베딩 차원 설정은 512차원, 학습 배치 크기 선정 128개, 학습률 1e-3, 최적화 기법, RMSProp를 사용하여 학습한다. 〈표 4-8〉 하이퍼 파라미터 선정 결과 항목 선정 기법 가중치 초기화 기법 Xavier\_uniform 임베딩 차원 설정 512 학습 배치 크기 선정 128 학습률 1e-3 최적화 기법 RMSProp 4.5.2 Transformer 모델의 실험 결과 분석 딥러닝 모델의 성능을 평가하기 위해 정확도, 재현율, 정밀도 등의 지표를 사용한다.

문장유사율: 0%

이러한 지표는 모델이 얼마나 잘 예측하는지에 대한 정보를 제공한다. 실험에서는 Transformer 모델을 사용하였으며, 특징 개수가 52개인 1차 실험과 특징 개수가 34개인 2차 실험을 단계별로 진행하였다.

문장유사율: 0%

Transformer 모델은 2차 실험에서는 학습 초기부터 높은 정확도와 낮은 손실을 보여주며, epoch=100으로 학습을 초기에 종료했음에도 불구하고 높은 성능을 유지하였다.

문장유사율: 0%

Transformer 모델은 epoch=100 이상으로 추가적인 학습을 진행시 높은 성능을 도출할 수 있다고 예상된다.

문장유사율: 0%

모델의 성능평가에서 AUC 지표는 모델의 이진 분류 능력을 평가하는 데 중요한 역할을 했다.

문장유사율: 0%

Transformer 모델은 AUC가 높은 성능을 보였는데, 이는 모델이 Positive와 Negative 클래스를 잘 구분하는 능력이 높다는 것을 의미한다.

문장유사율: 0%

Transformer 모델은 학습 초기부터 안정적인 성능을 보여주어서 조기종료 기법을 적용하더라도 높은 성능을 유지할 수 있다.

문장유사율: 0%

실험 결과를 종합해보면, 1차와 2차 실험에서 특징 개수의 차이가 모델의 성능에 큰 영향을 미쳤다. 특징 개수가 적은 2차 실험에서는 모델의 성능이 상승되었으며, 불필요한 데이터를 처리하고 구직자의 개인적인 Feature를 추가하여 모델 개발 시 특징 개수를 적절히 선택하고 최적의 특징 조합을 고려하는 것이 중요하다.

문장유사율: 0%

또한, 이력서와 같이 텍스트 데이터와 같은 시퀀스 데이터를 다루는 경우에는 셀프 어텐션 메커니즘이 적용된 Transformer 모델을 고려해볼 가치가 있다.

문장유사율: 0%

마지막으로, 최적화 기법으로는 RMSProp을 사용하였는데, 이는 모델의 학습을 효과적으로 수행하기 위해 사용되는 알고리즘이다.

문장유사율: 0%

하지만, 다른 최적화 기법을 사용하거나 하이퍼 파라미터를 조정함으로써 모델의 성능을 더욱 개선할 수 있을 것이다.

문장유사율: 0%

따라서, 최적화 알고리즘과 하이퍼 파라미터 설정에 대한 추가적인 실험과 조정을 수행하여 모델의 성능을 더욱 향상시킬 수 있을 것으로 기대한다.

문장유사율: 0%

4.5.3 추론 모델의 비교분석 모델 학습 후 추론 모델을 만들어 재현율, 정밀도, F1\_Score를 평가지표로 하여 테스트 데이터를 활용하여 성능평가를 진행한다.

문장유사율: 0%

성능평가에 대해 어떤 지표가 가장 중요한지는 모델에 따라 다르므로, 상황에 따라 적절한 지표를 선택하여 모델의 성능을 평가한다.

문장유사율: 0%

본 연구 모델의 직무 추천시스템은 재현율을 사용해서 평가한다. 재현율은 실제 Positive인 데이터 중 모델이 Positive로 예측한 데이터의 비율을 나타낸다. 다시 말해, 모델이 얼마나 실제 Positive 데이터를 빠짐없이 예측할 수 있는지를 나타내는 지표이다.

문장유사율: 0%

Transformer 모델의 재현율은 83%로 이는 실제 정답값을 빠짐없이 찾아내는데 성공했다는 의미이다.

문장유사율: 0%

〈표 4-9〉 Transformer 추론 모델의 성능평가 차수 Precision Recall Transformer 1차 실험 0.29 0.83 0.43 2차 실험 0.26 0.94 0.43 V.

문장유사율: 0%

결론 5.1 연구의 요약 및 결론 본 연구는 구직자들의 전공에 기반을 둔 직무와 산업에서 요구하는 직무 간의 차이로 발생하는 반복적인 미스매칭 중 직무 미스매치에 대한 연구로 직무 미스매치를 취업포털의 이력서 데이터를 활용하여 딥러닝 기반 직무 추천 시스템을 개발하고, 성능평가를 통한 최적의 모델선정과 실제 시스템에 적용할 수 있는 연구를 제안하는 것이다.

문장유사율: 0%

직무 미스매치 문제를 해소하기 위한 딥러닝 기반의 직무 추천 모델은 협업 필터링을 이용하였으며, 이력서 데이터를 학습 데이터로 구성하는데 있어 2가지가설을 설정하였다.

문장유사율: 0%

첫째, 경력자의 이력서 데이터는 구직자의 직무 선택의 기준 지표를 포함하고 있다.

문장유사율: 0%

둘째, 구직자와 유사한 프로파일을 보유하고 있는 경력자의 이력서 데이터를 기반으로 직무를 추천하였을 때 가장 높은 만족도와 직무 추천 성과를 내포한다는 것이다.

문장유사율: 0%

이에 따른 단계별 연구 진행을 위하여 1단계로 경력자 중심의 이력서 데이터를 기반으로 52개의 특징과 34개의 특징을 갖는 2종의 학습 데이터셋을 구축하였으며, 2단계로 텍스트 임베딩 기법과 딥러닝 모델을 활용하여 파인튜닝을 통한 직무 추천 모델을 설계하고, 그리드 탐색 실험으로 최적의 하이퍼 파라미터를 선정하였다.

문장유사율: 0%

본 연구는 딥러닝 모델의 성능평가를 위하여, 52개의 특징과 하이퍼 파라미터를 중심으로 하는 1차 실험과 34개 특징과 파인튜닝을 중심으로 하는 2차 실험의 성능을 측정하였다.

문장유사율: 0%

Transformer 모델의 1차 실험의 학습모델 성능평가 결과로는 정확률(Accuracy)이 0.7898, 추론 모델 성능은 모델의 재현률(Recall)이 0.83으로 우수하였다.

문장유사율: 0%

2차 실험의 학습모델 성능평가 결과는 1차 실험과 비교하여 상승이 있었으며, 성능 향상의 주요 요인으로는 특징(Feature)을 52개에서 34개로 축소하여 학습 데이터셋을 최적화한 것과 파인튜닝 과정으로 모델 설계를 진행한 것이라 할 수 있다.

문장유사율: 0%

특히, 파인튜닝을 통해 재설계된 Transformer의 Epoch는 1차 실험에서 epoch=39로 중단되었지만, 2차 실험에서는 epoch=100까지도 달하였다.

문장유사율: 0%

이러한 결과로 파인튜닝을 통한 모델 설계와 양질의 이력서 데이터 확보는 직무 추천 모델의 성능과 직접적인 관계가 있다는 것을 확인하였다.

문장유사율: 0%

5.2 연구의 한계 및 향후 연구 이력서를 기반으로 한 직무 추천 연구는 민감한 정보인 이력서 데이터를 활용하여 구직자에게 적합한 직무를 추천하기 위한 것이다.

문장유사율: 0%

본 연구를 진행하면서 도출된 대표적인 문제점은 이력서 데이터의 수집, 관리, 그리고 컴퓨터 자원의 확보 라고 할 수 있다.

문장유사율: 0%

이력서 데이터를 수집 및 관리하는 관련 기업이 아니라면, 개인정보보호법을 준수하면서 학습 데이터로 활용하기 위한 충분한 데이터의 수집에 한계가 있으며, 수집된 이력서 데이터의 개인정보 유출 등의 관리 위험이 존재하여, 연구에 어려움이 있다.

문장유사율: 0%

또한, 새로운 직무에 대한 정확성과 완결성이 미흡한 학습 데이터는 직무 추천 모델의 성능에 영향을 미치므로 새로운 이력서 데이터를 수집하고 정제하여 데이터 품질을 지속적으로 유지하는것에 어려움이 있다.

문장유사율: 0%

그리고, 구직자의 경험, 기술, 선호도 등은 시간이 지남에 따라 변화하므로 이력서를 기반으로 하는 추천 시스템은 구직자의 동적 변화에 대응하기에는 한계가 있으며, 추가되는 이력서 데이터를 고려한 학습 모델의 학습 주기를 면밀히 고려해야 한다.

## 인용 포함 문장

문장유사율: 33%

향후 연구에서는 Transformer의 Epoch 학습을 지속할 수 있는 환경을 구성하고, 사용성 검증이 필요할 것이며, 직무 추천과 연계된 직무 역량강화를 위한 학습 콘텐츠 추천 모델로의 확장 연구도 고려해볼 수 있을 것이다. "본 게재논문에 대한 윤리적 문제는 전적으로 저자에게 책임이 있음을 확인함"

KCI 논문 | 제목: 온라인 쇼핑 추천... | 저자: 짜오권(경... | 발행년: 2023.08

향후 연구에서는 최신성 및 발전성의 측정 항목을 보완하여야 연구의 구성 타당도가 더 높아질 수 있을 것이다. "본 게재논문에 대한 윤리적 문제는 전적으로 저자에게 책임이 있음을 확인함"

KCI 논문 | 제목: 직무몰입이 지식은... | 저자: 김현우(울... | 발행년: 2021.04

다른 리더십의 효과를 함께 검증하고 통제하여 영향력을 살펴봤을때, 서번트리더십이 중요한 조절효과라는 것을 더 명확히 알 수 있을 것이다. "본 게재논문에 대한 윤리적 문제는 전적으로 저자에게 책임이 있음을 확인함"

KCI 논문 | 제목: 내부시장지향성이 ... | 저자: 김동윤(승... | 발행년: 2021.04

둘째, 내부시장 지향성과 시장지향성의 관계에서 동적역량이 매개효과를 주는 것을 확인 하였는데, 또 다른 매개변수를 찾는다면 내부시장 지향성이 시장지향성 에 미치는 영향을 좀 더 다양하게 설명할 수 있을 것이다. "본 게재논문에 대한 윤리적 문제는 전적으로 저자에게 책임이 있음을 확인함"

## 참고문헌

참고문헌 고용노동부, 2023.01.02., [https://www.moel.go.kr/news/enews/report/newsView.do?news\\_seq=14478](https://www.moel.go.kr/news/enews/report/newsView.do?news_seq=14478).

"구인-구직 간 일자리 미스매치 실태와 개선방안에 관한 연구"박재홍 (2020), , 부산대학교 대학원 석사학위 논문.

사람인, 2022.08.03., [https://www.saramin.co.kr/zf\\_user/hr-magazine/view?hr\\_idx=953](https://www.saramin.co.kr/zf_user/hr-magazine/view?hr_idx=953). 오소진 (2022), , 한남대학교 박사학위 논문.

"BERT 기반의 전이학습 모델을 적용한 양방향 인제매칭 시스템"

우영춘·이성엽·최완·안창원·백옥기 (2019), "디지털 헬스케어 데이터 분석을 위한 머신러닝 기술 활용 동향", 전자통신 동향 분석, 34(1), 98-110.

"딥 러닝을 이용한 한국어의존 구문 분석"이창기·김준석·김정희 (2014), , 제 26 회 한글 및 한국어 정보처리 학술대회, 87-91.

장원중 (2020), "워드 임베딩을 이용한 관광여행지 선호도 비교에 대한 실증연구", 글로벌 경영 학회지, 17(6), 54-71.

조운환·서영덕·박대준·정제창 (2016), , 대한전자공학회, 2016(11), 800-803.  
"DNN에서 효율적인 학습을 위한 활성화 함수에 대한 고찰"

주용한·조인수(2020), "텍스트 마이닝을 이용한 중소기업 스마트 팩토리 성공요인 분석", 글로벌 경영 학회지, 17(5), 115-131.

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). "Neural machine translation by jointly learning to align and translate", In 3rd International Conference on Learning Representations, ICLR 2015.

Cheng, J., Dong, L., & Lapata, M. (2016). , In 2016 Conference on Empirical "Long Short-Term Memory-Networks for Machine Reading"Methods in Natural Language Processing, 551-561.

Association for Computational Linguistics. Kim, D. G., Park, Y. S., Park, L. J., & Chung, T. Y. (2019).

"Developing of new a tensorflow tutorial model on machine learning: focusing on the Kaggle titanic dataset"

, IEMEK Journal of Embedded Systems and Applications, 14(4), 207-218.

Li, Y., Xu, L., Tian, F., Jiang, L., Zhong, X., & Chen, E. (2015). , In Twenty-Fo "Word embedding revisited: A new representation learning and explicit matrix factorization perspective"  
urth International Joint Conference on Artificial Intelligence.

Lin, Z., Feng, M., dos Santos, C., Yu, M., Xiang, B., Zhou, B., & Bengio, Y. (2017). "A structured self-attentive sentence embedding, International Conference . In International Conference on Learning Representations"on Learning Representations, ICLR.

Parikh, A., Täckström, O., Das, D., & Uszkoreit, J. (2016). , In Proceedings of "A Decomposable Attention Model for Natural Language Inference"the 2016 Conference on Empirical Methods in Natural Language Processing, 2249-2255.

Paulus, R., Xiong, C., & Socher, R. (2017). "A deep reinforced model for abstractive summarization", In International Conference on Learning Representations.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I.

(2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. Yazan, E., & Talu, M.



F. (2017)., In 2017 International Artificial Intelligence and Data Processing Symposium "Comparison of the stochastic gradient descent based optimization techniques" posium (IDAP), 1-5.

문장유사율: 0%

IEEE, A Study on Job Recommendation Based on Transformer Model Jeon, Jung Woo\*, Song, Ji Ho\*\*, Ahn, Tae Ho\*\*\* \* 1st author, Ph.D.

문장유사율: 0%

, The Dept. of Business Administration, Soongsil University (certimaster@gmail.com) \*\* 2nd author, Master's Student, The Dept.

문장유사율: 0%

of Business Administration, Soongsil University (songjiho@gmail.com) \*\*\* Corresponding author, Ph.D.

문장유사율: 20%

Professor, The Dept. of Business Administration, Soongsil University (ahnt@ssu.ac.kr) This study proposes job recommendation models based on job seekers' resumes, evaluates the performance of these models, and suggests ways to enhance the utility of the recommended jobs.

KCI 논문 | 제목 : 커피전문점 모바일... | 저자 : 한수민(송... | 발행년 : 2021.08

of Business Administration, Soongsil University (\*\*\*\*\*@\*\*\*\*\* ) \*\* Corresponding author, Ph.D. Professor, The Dept. of Business Administration, Soongsil University (\*\*\*\*\*@\*\*\*\*\* )

KCI 논문 | 제목 : 내부시장지향성이 ... | 저자 : 김동윤(송... | 발행년 : 2021.04

Professor, The Dept. of Business Administration, Soongsil University (\*\*\*\*\*@\*\*\*\*\* ) \*\*\* Co-author, Doctoral Candidate, The Dept. of Business Administration, Soongsil University (\*\*\*\*\*@\*\*\*\*\* )

문장유사율: 0%

Two experiments were conducted for this purpose. The first experiment evaluated the performance of a job recommendation model using the Transformer, applying a training dataset composed of 52 features from resume data and specific hyperparameters settings.

문장유사율: 0%

The second experiment assessed the performance of a deep learning model with fine-tuning techniques using a training dataset that consisted of 34 features, which excluded features with more than 30% missing values from the original 52 features used in the first experiment.

문장유사율: 0%

The increase in the Epoch value of the Transformer model holds significant meaning. This research demonstrates that optimizing training data and applying fine-tuning to the deep learning model design significantly impacts the performance improvement of the job recommendation model.

문장유사율: 0%

It is expected that further enhancing the performance of the job recommendation model is feasible by applying quality resume data to the Transformer model and continuing additional model training, especially considering the rate of increase in the Epoch value. KeyWords : Resume-based job recommendations, deep learning, Transformers models,