Journal of the Korea Academia-Industrial cooperation Society
Vol. 24, No. 9 pp. 769-775, 2023

https://doi.org/10.5762/KAIS.2023.24.9.769
ISSN 1975-4701 / eISSN 2288-4688

BERT 기반 국내 스타트업-해외 바이어간 B2B 매칭 모델

최정석[1], 문남미[2*]
[1]호서대학교 융합공학과, [2]호서대학교 컴퓨터학부 교수

BERT-based B2B matching model for K-Startups and Overseas Buyers

Jungsuk Choi[1], Nammee Moon[2*]
[1]Dept. of Convergence Engineering, Hoseo Graduate School of Venture
[2]Dept. of Computer Engineering, Hoseo University

요 약 국내 스타트업이 자사의 시장 확대를 위해 글로벌 시장으로 진출하기 위해서는 적합한 해외 바이어를 효과적으로 발굴하는 작업이 필수적이다. 이에 본 논문은 자연어처리 사전학습 모델인 BERT 기반의 국내 스타트업-해외 바이어간 B2B 매칭 모델을 제안한다. 제안 모델은 국내 스타트업 및 바이어 홈페이지, 상담일지 등에서 수집한 비정형 텍스트 데이터를 입력 데이터로 사용했고, 출력층에서 적합/부적합 파트너를 이진분류하도록 설계했다. 기업정보는 국내 스타트업 2,860개사, 해외 바이어 2,316개사의 데이터를 수집했고, 매칭 정보는 30,456건의 상담데이터 중 긍정 케이스 6,458건(21.2%), 부정 케이스 23,998건(78.8%)의 데이터를 활용했다. 제안모델의 성능평가를 위한 기준모델은 Word2Vec을 양방향 LSTM으로 학습시킨 콘텐츠기반 필터링 모델이다. 실험결과 제안모델은 기준모델에 비해 정확도(1.7%p), 정밀도 (3.1%p), 재현율(5.2%p), f1-score(4.1%p) 등에서 고르게 더 높은 성능을 보여주었다. 국내 스타트업이 이 모델을 활용해 적합한 바이어를 보다 효과적으로 발굴할 수 있을 것이다.

Abstract For K-Startups to advance into the global market for their business expansion, it is critical to find suitable overseas buyers effectively. This paper proposes a business-to-business (B2B) matching model between K-Startups and overseas buyers based on BERT, a pre-trained NLP model. The proposed model was designed to classify binary classification for right or non-right business partners in the output layer using unstructured text data collected from home pages of K-Startups and overseas buyers and business meeting records. For corporate information, data from 2,860 startups and 2,316 overseas buyers were collected, and for matching information, out of 30,456 business meeting records data, 6,458 positive cases (21.2%), and 23,998 negative cases (78.8%) were used. The base model for evaluating the proposed model is a content-based filtering model trained by bidirectional LSTM with Word2Vec. Based on the experimental results, the proposed model demonstrated better performance compared to the comparative model in accuracy (1.7%p), precision (3.1%p), recall (5.2%p), and f1-score (4.1%p). Using this model, K-Startups can find overseas buyers more effectively.

Keywords : B2B Matching, BERT, Recommender Systems, K-Startups, NLP

*Corresponding Author : Nammee Moon(Hoseo Univ.)
email: nammee.moon@gmail.com
Received August 1, 2023 Revised August 31, 2023
Accepted September 1, 2023 Published September 30, 2023

1. 서론

1.1 연구배경 및 필요성

국내 스타트업이 작은 국내시장에서의 과밀경쟁을 넘어 글로벌시장 진출을 통해 스케일업하기 위해서는 이들의 제품 및 서비스에 적합한 해외 바이어를 효과적으로 발굴하는 작업이 필수적이다. 그러나, 해외 네트워크가 부족한 국내 스타트업은 바이어 발굴에 애로를 겪고 있으며, 국내 서비스 스타트업의 해외진출 시 가장 큰 애로사항이 '바이어 발굴'로 조사된 바 있다[1]. 이 해외 바이어 발굴작업에 인공지능 추천시스템(Recommender Systems)을 활용해 B2B 매칭시스템을 개발함으로써 국내 스타트업이 해외 바이어 발굴 작업에 투입하는 시간과 비용을 줄여 업무효율성을 높이는데 일조하고자 한다.

1.2 연구목표

본 논문은 이 과제를 수행하기 위해 자연어처리(Natural Language Processing, 이하 NLP) 기반의 추천시스템을 활용한 B2B 매칭 모델을 제안하고자 한다. 특히, 자연어처리 분야에서 높은 성능을 보여주고 있는 BERT (Bidirectional Encoder Representations from Transformers)[2]를 사용해 매칭 모델을 구축한다. 매칭 모델의 설계 및 구축을 위해 국내 스타트업과 해외 바이어의 상품 및 서비스 설명서, 그리고 상담일지 등에서 수집한 비정형 텍스트 데이터셋을 사용하고, BERT 사전학습 모델을 미세조정을 거쳐 전이학습(Transfer Learning)시킨 B2B 매칭모델을 제안한다.

이 BERT기반 B2B 매칭 모델이 유용성을 입증하기 위해 텍스트 기반 추천시스템의 협업필터링(Collaborative Filtering)과 컨텐츠기반 필터링(Content-based Filtering) 분야에서 고르게 사용되는 모델과 성능을 비교하고자 한다[3]. 한편, BERT가 양방향 사전학습 모델이므로 보다 의미있는 모델간 성능비교를 위해 Word2Vec을 양방향 LSTM[4](Bidirectional Long Short-Term Memory, 이하 BiLSTM)으로 학습시킨 결합모델 Word2Vec_BiLSTM과 성능을 비교한다.

본 논문의 구성은 다음과 같다. 2장에서는 B2B 매칭을 위한 추천시스템, BERT 등에 관한 관련연구를 정리했고, 3장에서는 BERT기반의 B2B 매칭모델을 제안한다. 4장에서는 제안모델의 성능평가를 위해 실험용 데이터셋과 성능평가 방법 및 결과를 분석한다. 5장에서는 본 연구의 결론과 향후 과제에 대해 정리했다.

2. 관련연구

2.1 자연어처리 기반 추천시스템

자연어처리 기반 추천시스템은 협업필터링과 컨텐츠기반 필터링 분야에서 다양하게 시도되었다. 협업필터링 분야에서 Oren Barken, et al.[5]은 자연어처리 알고리즘인 Word2Vec을 추천시스템에 접목시킨 Item2Vec을 제안했다. 이 모델은 유저 데이터 없이 아이템의 암묵적 데이터(Implicit data)와 부가정보 데이터만으로 추천시스템의 성능을 향상시켜 협업필터링 모델의 단점으로 지적되는 데이터 희소성(sparsity), 콜드 스타트(cold start) 등의 문제를 극복하는 의미있는 성과를 보여주었다. 그리고, W. Kang, et al.[6]은 트랜스포머의 셀프어텐션 메커니즘을 이용해 유저의 아이템 상호작용 시퀀스를 모델링한 SASRec을 발표해 추천시스템의 성능향상을 이끌었다.

한편, 컨텐츠기반 추천시스템 분야에서는 D. M. Blei, et al.[7]가 자연어처리를 접목한 토픽 모델링 LDA(Latent Dirichlet Allocation) 기법을 활용해 아이템 설명이나 사용자 리뷰 등의 토픽 분포를 구하고, 토픽 선호도를 기반으로 추천하는 새로운 방법을 제시했다. T. Mikolove, et al.[8]가 발표한 Word2Vec 확장모델은 단어의 의미를 벡터화해 주변 단어들과의 유사도를 바탕으로 영화, 음악 등 다양한 분야에서 유사한 아이템을 추천하는데 사용되었다. 이후 Q. Le, et al.[9]은 Word2Vec을 확장한 Doc2Vec 모델을 제안해 주어진 문장에서 단어뿐 아니라 문장 전체를 이해하려는 시도를 통해 컨텐츠기반 추천의 성능향상을 꾀했다. 또, J. Pennington, et al.[10]이 발표한 Glove는 데이터셋에서 단어들의 동시출현 통계인 공분산 정보를 활용해 단어간 관계와 유사도를 학습시킴으로써 단어의 의미를 더 잘 표현할 수 있게 했다. 자연어처리 기반의 컨텐츠 필터링 추천시스템은 A. Vaswani, et al.[11]가 트랜스포머를 발표하며 또 한번의 큰 발전을 이루었다. 트랜스포머의 인코더 모델인 BERT가 자연어 이해(Natural Language Understanding) 분야에서 뛰어난 성능을 보이자, BERT를 활용한 추천모델에 관한 연구가 다수 진행되었다. S. Oh[12]는 BERT를 추천시스템에 적용한 양방향 인재매칭 모델을 발표해 BERT 기반 추천시스템의 유용성을 입증했다.

2.2 BERT 사전학습 모델

Word2Vec과 같이 단어를 공간에 임베딩해 사전학습

시킨 자연어처리 모델은 기존의 단어 수 카운팅에 기초해 통계기법으로 분석하는 TF-IDF 등의 통계적 언어모델에 비해 우수한 성능을 보여주었다. 이후 K. Cho, et al.[13]은 RNN(Recurrent Neural Network)기반의 인코더와 디코더 구조를 도입해 자연어처리 태스크에 광범위하게 적용되었다. 그러나, Word2Vec은 단어의 문맥(context)에 따른 의미가 무시되고, RNN 모델은 문장이 길어지면 문장의 앞부분 단어를 잘 고려하지 못 해 성능이 떨어지는 현상이 나타났다. 이런 문제점을 극복하기 위해 중요한 단어에 집중하는 어텐션 메커니즘(Attention Mechanism)이 등장했다. 어텐션 메커니즘을 활용한 BERT는 구글이 발표한 사전학습 모델로 양방향으로 문맥을 고려해 텍스트를 이해한다. 또, 다양한 문장의 길이에 유연하게 대응할 수 있어 모델 구조를 크게 변경할 필요없이, 미세조정(fine-tuning)만으로도 높은 성능을 보인다. 그리고, 100개 이상의 언어를 지원해 다국가의 언어를 원문 그대로 학습시킬 수 있는 장점이 있다.

2.3 B2B 매칭

일반적으로 B2B 매칭은 특정 제품이나 서비스의 공급자와 수요자를 연결하는 것을 뜻한다. 최근에 인공지능(AI: Artificial Intelligence, 이하 AI) 기술이 이런 B2B 매칭의 프로세스를 자동화하고 개선하는 데 활용되고 있다. F. Prieto, et al.[14]은 B2B 판매를 위해 AI를 사용하는 다국적기업 Danfoss의 사례조사를 통해 AI 활용이 비즈니스의 프로파일 검색 및 분석 시간을 줄이고, B2B 기업의 판매활동에 미치는 영향을 평가해 AI 활용의 중요성을 강조했다. Y. L. Yoon, et al.[15]은 온라인 마켓플레이스에서 공급자-구매자간 매칭을 촉진하기 위한 방법을 연구해, 적절하게 설계된 추천시스템이 구매자의 정보 과부하를 줄여 매칭률을 높일 수 있다는 점을 밝혔다.

3. B2B 매칭모델 설계

3.1 BERT기반 B2B 매칭 모델

본 연구는 국내 스타트업과 해외 바이어간 제품 및 서비스의 유사성을 바탕으로 AI 추천시스템을 활용해 매칭하는 B2B 매칭모델을 제안한다.

이를 위해 기업의 상품 및 서비스 설명서와 국내 스타트업-해외 바이어간 상담일지의 비정형 텍스트를 데이터셋으로 활용하고, 문맥의 의미를 이해할 수 있는 BERT 사전학습 모델을 이용하면, Word2Vec과 같이 단순히 단어를 임베딩해 단어 유사도 기반으로 추천하는 모델보다 우수한 성능을 보일 것이라는 가설에 기초해 BERT 기반 B2B 매칭 모델을 제안한다.

제안하는 BERT 기반 B2B 매칭 모델의 기본구조는 Fig. 1에서 보는 바와 같이 국내 스타트업의 아이템, 투자유치 이력, 마케팅 활동 등의 항목으로 구성된 스타트업 프로파일과 해외 바이어의 아이템, 국가, 잠재 오더량 등의 항목으로 구성된 해외 바이어 프로파일을 만들어 입력 데이터로 사용하고, 상담일지의 상담내용 텍스트를 토큰화(tokenize)해 모델에 입력한다. 데이터셋이 한글과 영어, 2가지로 작성되어 BERT 다중언어 모델을 적용한다. 다음 단계로 토큰 임베딩(token embedding), 세그먼트 임베딩(segment embedding)과 포지션 임베딩

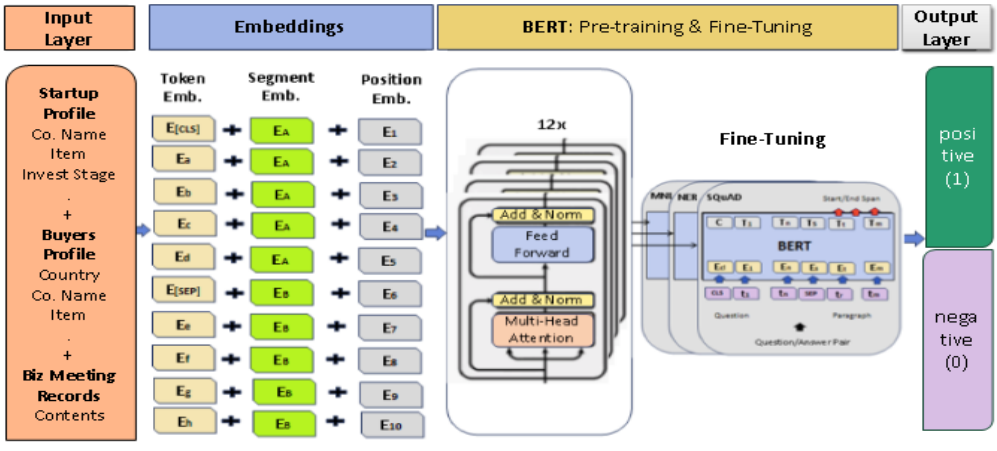

Fig. 1. BERT_B2B matching model structure

(position embedding)의 과정을 거친다.

임베딩된 기업 프로파일과 상담일지 내용을 벡터로 만들고, 상담일지 상의 상담내용을 사전학습 모델에서 학습시킨다. 다음 단계로 미세조정 과정을 거쳐 상담내용과 상담일지 상에 기록된 상담만족도를 학습한다. 마지막 출력층에서 시그모이드(Sigmoid) 함수를 이용해 긍정(1), 부정(0)으로 이진분류하여 실제 상담일지에 기록된 긍정, 부정 라벨링과 비교해 정확도, 정밀도, 재현율, f1-score 등 성능평가 지표를 계산하는 구조다.

한편, 제안모델의 성능평가를 위한 기준모델은 Word2Vec을 BiLSTM으로 학습시킨 모델이다. 국내 스타트업, 해외 바이어의 프로파일과 제품 설명서 및 상담일지의 말뭉치(corpus)를 단어 빈도수 벡터(BoW)로 변환해 Fig. 2에서 보는 바와 같이 Word2Vec 사전학습모델에 입력한다. 이후 BiLSTM 레이어에서 양방향으로 학습시킨 후, 텍스트를 결합해 시그모이드(Sigmoid)함수를 거쳐 최종 출력층에서 적합도 여부를 학습하도록 한다. 단어들은 토큰 임베딩을 통해 고유 ID를 부여받으며, 세그먼트 임베딩과 포지션 임베딩을 이용해 각 단어의 순서와 위치 정보를 설정한다. 이렇게 전처리를 거친 입력 데이터는 BiLSTM에 들어갈 입력 임베딩으로 설정된다. 그리고, Word2Vec과 BiLSTM을 결합하여 기준모델인 Word2Vec_BiLSTM을 만든다.

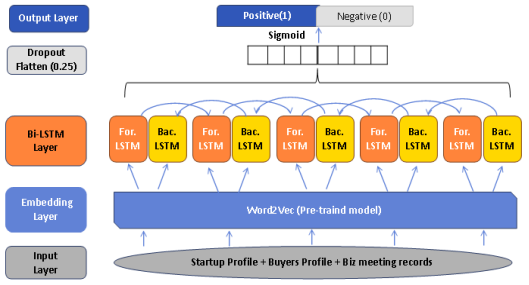

Fig. 2. Word2Vec_BiLSTM matching model structure

4. 실험 및 분석

4.1 데이터셋

모델의 성능실험을 위한 데이터셋은 국내 스타트업 기업정보, 해외 바이어 기업정보, 상담일지 등 3종류의 csv 파일로 작성된 데이터셋을 사용했다. Table 1에서 보는 바와 같이 국내 스타트업 정보는 총 2,860개 사의 기업정보를 기업명, 아이템 설명 텍스트, 투자유치 이력, 마케팅 활동 등 9개의 컬럼으로 구성했다. 해외 바이어 정보는 2,316개의 기업정보를 기업명, 아이템 설명 텍스트, 잠재 주문량 등 6개의 컬럼으로 구성했다. 상담일지는 스타트업 및 바이어의 ID, 상담내용 텍스트, 매칭 만족도 등 10개의 컬럼으로 구성했다. 상담 만족도 기준으로 긍정 케이스 6,458건(21.2%), 부정 케이스 23,998건(78.8%)을 분류해 사용했다. 실험데이터 중 훈련용 데이터는 24,364건(80%), 테스트 데이터는 6,092건(20%)를 임의로 추출해 사용했다.

Table 1. Experimental Dataset

Data Set	Data Points	Remarks
K-Startup Info.	2,860	homepages
Overseas Buyers Info.	2,316	homepages
Maching Info.	30,456 positive cases: 6,458 (21.2%) negative cases: 23,998(78.8%)	biz meeting records
Training data	24,364 (80%)	
Test data	6,092 (20%)	

4.2 성능평가 방법

본 연구에서는 매칭 모델의 성능평가 지표로 정확도(accuracy), 정밀도(precision), 재현율(recall), f1-score 및 AUC(Area Under Curve)를 사용한다.

국내 스타트업과 해외 바이어간 B2B 매칭의 적합 여부를 판단하기 위해 상담 만족도를 5등급으로 나누어 매우 낮음(1점), 낮음(2점), 보통(3점), 높음(4점), 매우 높음(5점)으로 분류했으며, 만족도 '낮음'(2점) 이상인 국내 스타트업-해외 바이어 쌍은 긍정(1) 케이스로 분류했다. 상품 및 서비스가 적합한 매칭에도 불구하고, 가격, A/S 조건에 대한 이견 등으로 만족도가 '낮음'으로 분류된 경우가 대부분이어서 '낮음'도 긍정 케이스로 분류했다. 그러나, 만족도가 '매우 낮음'(1점) 이하인 경우는 상품의 매칭이 부적합한 경우가 대부분 포함되어 있어 부정(0) 케이스로 분류했다. 또, 상담 취소나 상담장 노쇼(no show)의 경우 등은 미스매칭으로 간주하여 부정(0) 케이스로 분류했다. 평가지표 중 정밀도는 Eq. (1)과 같이 추천 모형에서 긍정(1)으로 예측된 국내 스타트업-해외 바이어 쌍에서 실제 상담 만족도가 긍정(1)인 쌍의 비율로 계산한다.

$$\text{정밀도(precision)} = \frac{\| P \cap C \|}{\| P \|} \quad (1)$$

$\| P \|$: The count of K-startup and overseas buyer pairs predicted as positive (1) by the model

$\| P \cap C \|$: The count of K-startup and overseas buyer pairs that were actually classified as positive in biz meeting records, among those predicted as positive (1) by the model

재현율은 Eq. (2)과 같이 상담이력이 있는 국내 스타트업-해외 바이어의 쌍 중에서 제안모델이 긍정(1)으로 예측한 쌍의 비율로 계산한다.

$$\text{재현율(recall)} = \frac{\| C \cap P \|}{\| C \|} \quad (2)$$

$\| C \|$: The count of K-startup and overseas buyer pairs that have recorded meeting cases in biz meeting records

$\| C \cap P \|$: The count of K-startup and overseas buyer pairs with recorded meeting cases that were predicted as positive (1) by the model

f1-score 는 정밀도와 재현율의 조화평균으로 계산된다. 또한, AUC(Area Under Curve)는 ROC(Receiver Operating Characteristic) 그래프의 면적으로 계산하여 모델의 성능을 평가한다. 이 지표는 임의의 긍정사례를 임의의 부정사례보다 얼마나 잘 분류하는지에 대한 확률로 해석된다.

4.3 실험 및 결과 분석

본 연구에서는 국내 스타트업과 해외 바이어의 홈페이지와 상담일지 등에서 수집한 비정형 텍스트 데이터를 입력 데이터로 사용해, BERT기반의 B2B 추천 모델을 제안하고, 성능평가를 시행했다.

실험환경은 Table. 2에서 보는 바와 같이 실험 PC는 Intel(R) Core(TM)을 탑재하고 있고, 기본 클럭 속도는 1.20GHz다. RAM 용량은 8.0GB, 하드 드라이브 용량은 460GB이며. 운영체계는 Window 10이다. 서버는 Intel(R) Xeon(R) CPU를 탑재하고 있고, 클럭 속도는 2.30GHz다. RAM 용량은 총 12.0GB, 하드 드라이브 용량은 114GB, 서버 운영체제는 Ubuntu 18.04.5이고, 소프트웨어는 Python 3.7.10, Tensorflow 2.4.1, Keras 2.4.3 및 BERT 모듈이 설치되어 있다.

제안 모델의 하이퍼 파라미터로 옵티마이저는

Table 2. Experimental Setup

Item	Specification
Experimentation PC	CPU : Intel(R) Core(TM) i5 1035G7 CPU @ 1.20GHz, 1.50 Ghz
	RAM : 8.0 GB
	HDD : 460.0 GB
	OS : Window 10 Home
	SW : Google Chrome 89.0.4389.82
Dedicated Server	CPU : Intel(R) Xeon(R) CPU @ 2.30GHz
	RAM : 12.0 GB
	HDD : 114.0 GB
	OS : Ubuntu 18.04.5 LTS
	SW : Python 3.7.10, Tensorflow 2.4.1, Keras 2.4.3, BERT module

RectifiedAdam을 설정했고, 그에 따른 학습률은 1.0e-5, 과대적합을 극복하기 위한 weightdecay는 0.0025를 설정했다. 손실함수는 이진분류 문제에서 가장 성능이 좋은 BinaryCrossEntropy를 사용했고, 훈련 모델은 총 5번(epochs=5)를 진행했다. Fig. 4에서 보는 바와 같이 epochs 10으로 학습시켰을 때, epochs 5 이후부터는 손실값과 정확도가 떨어지지 않았기 때문에 최적의 epochs를 5회로 설정했다. 기타 은닉층 사이즈(hidden_size, 768)와 은닉층 수(num_hidden_layers, 12) 등은 BERT-Base 디폴트값으로 실험했다,

성능비교를 위한 기준모델은 Word2Vec과 BiLSTM을 활용한 Word2Vec_BiLSTM이다. 이 모델은 제품 및 서비스 정보에서 단어 사전을 만드는 BoW 방식을 적용했다. 반면, BERT 모델은 양방향 정보를 활용해 스타트업과 바이어의 제품 및 서비스를 통합하는 전처리 과정을 통해 30,456건의 통합 데이터셋을 구성했다. 이중 임의로 추출한 24,364건(80%)은 훈련 데이터로 사용했고, 6,092건(20%)은 테스트 데이터로 활용했다.

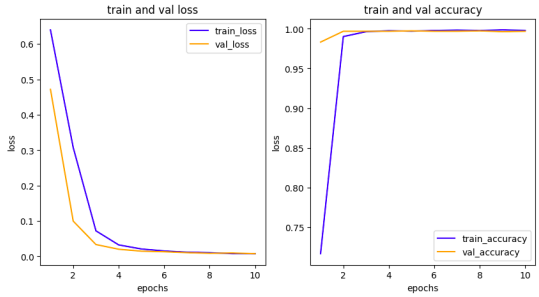

Fig. 3. train and val loss and accuracy

실험결과 Table 3에서 보는 바와 같이 BERT_B2B는 정확도(Accuracy) 99.9%, 정밀도(Precision) 99.8%, 재현율(Recall) 99.9%, f1-score 99.8%의 성능을 보였고, Word2Vec_BiLSTM은 정확도 98.2%, 정밀도 96.7%, 재현율 94.7%, f1-score 95.7%을 보였다. 제안모델이 기준모델에 비해 정확도(1.7%p), 정밀도(3.1%p), 재현율(5.2%p), f1-score (4.1%p) 등에서 고르게 높은 성능을 보였다. 특히, 재현율의 성능차이가 가장 큰 데, 이는 BERT_B2B 모델이 TP(True Positive) 케이스를 더 잘 분류해 적합한 비즈니스 파트너를 추천하는데 유용한 모델임을 입증했다는 점에서 의미가 있다.

Table 3. Experimental Results

Evaluation Metrics	Word2Vec_Bi-LSTM	BERT_B2B
Accuracy	0.9819	0.9993
Precision	0.9668	0.9977
Recall	0.9474	0.9992
f1-score	0.9570	0.9985
ROC_AUC	0.9693	0.9993

두 모델의 분류 능력을 측정한 AUC 값은 Fig. 4와 같이 BERT_B2B 모델 99.9%, Word2Vec_BiLSTM 96.9%를 각각 기록해 제안 모델이 기준모델에 비해 더 유용한 모델임을 보여주었다.

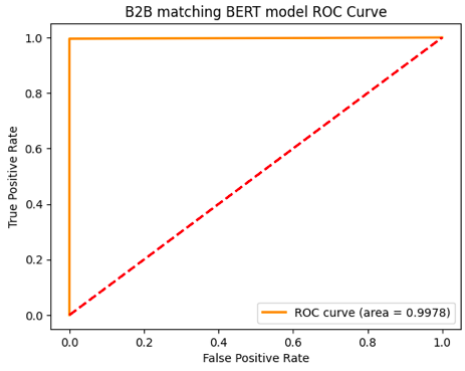

Fig. 4. ROC_AUC of BERT_B2B model

5. 결론 및 향후 과제

5.1 결론

본 연구에서는 국내 스타트업이 추천시스템을 활용해 적합 바이어를 효과적으로 발굴하는 BERT 기반의 B2B 매칭 모델을 제안했다. 제안 모델의 성능 평가를 위해 기준모델 Word2Vec_BiLSTM과 성능 비교를 했고, 제안 모델은 정확도, 정밀도, 재현율, f1-score 등에서 전반적으로 높은 성능을 보여주었다. 특히, 재현율이 높게 나타나 제안 모델이 적합한 비즈니스 파트너를 추천하는데 더 유용한 모델임을 보여주었다.

본 연구가 가지는 의미는 BERT 기반의 양방향 학습 모델을 사용해 문맥적 의미까지 학습시킴으로써 Word2Vec과 같이 단어 임베딩을 바탕으로 단어간 유사도를 바탕으로 추천하는 모델보다 성능을 향상시켰다는 점이다. 이런 특성은 유형의 제품을 수출하는 스타트업뿐 아니라, 무형의 서비스 아이템을 수출하는 서비스 스타트업에게도 유용하게 활용될 수 있다. 본 연구에서 제안한 BERT_B2B 모델이 실무에 적용된다면 국내 스타트업과 해외 바이어간 B2B 매칭의 성능을 보다 향상시킬 수 있을 것이다.

5.2 향후 과제

본 연구에서 제안한 BERT_B2B 매칭 모델은 비정형 텍스트 기반으로 국내 스타트업-해외 바이어간 추천하는 모델이다. 그러나, 복잡한 정밀 제품이나 새로운 제품 및 서비스의 디테일한 설명은 텍스트만으로 커버하기 어려운 한계를 가진다. 따라서, 향후 연구과제로 텍스트-이미지 데이터셋을 활용해 B2B 매칭모델을 구축한다면 매칭의 정확도는 더욱 향상시킬 수 있을 것이다. 또, 국내 공공기관에서 해외 바이어의 제품 및 서비스 수입희망 인콰이어리(Inquiry)를 매일 실시간으로 제공하고 있으므로, 해외 바이어 정보와 인콰이어리를 실시간으로 UD해 국내 스타트업에 추천하는 실시간 추천시스템을 구축한다면 매칭의 성공률을 더 높일 수 있을 것이다. 향후 텍스트-이미지 데이터셋을 활용한 멀티모달(Multi-modal) 추천 모델, 실시간 추천모델을 접목시킨 B2B 매칭모델을 개발해 추천시스템의 성능을 더욱 향상시키고자 한다.

References

[1] H. Shim, H. Lee, Analysis of the Current Status of Overseas Expansion in the Service Industry and Its Challenges, ISSN 2093-3118, KITA, Korea. 2020. Available from : https://www.kita.net/cmmrcInfo/rsrchReprt/rsrchReport/FileDown.do?orig_name=TF%2029%C8%A3.%20%BC

%AD%BA%F1%BD%BA%BE%F7%20%C7%D8%BF%DC%C1%F8%C3%E2%20%C7%F6%C8%B2%B0%FA%20%BE%D6%B7%CE%BF%E4%C0%CE%20%BA%D0%BC%AE.pdf&file_name=iitreporter_2096_11.pdf

[2] J. Devlin, M. W. Chang, K. Lee, K.Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *arXiv preprint arXiv*: 1810.48550, 2018.
DOI: http://doi.org/10.48550/arXiv.1810.04805

[3] T. Mikolov, K. Chen, G. Corrado, J. Dean, "Efficient estimation of word representations in vector space", *arXiv preprint arXiv*:1301.3781, 2013.
DOI: http://doi.org/10.48550/arXiv.1301.3781

[4] S. Horhreiter, J. Schmiduber, "Long Short-Term Memory", *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.
DOI: http://doi.org/10.1162/neco.1997.9.8.1735

[5] O. Barken, N. Koenigstein, "ITEM2VEC: Neural Item Embedding for Collaborative Filtering", *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing*, pp. 1-6, 2016.
DOI: http://doi.org/10.1109/MLSP.2016.7738886

[6] W, Kang, J. McAuley, "Self-Attentional Sequential Recommendation" *2018 IEEE International Conference on Data Mining*, pp. 197-206, 2018.
DOI: https://doi.org/10.48550/arXiv.1808.09781

[7] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent dirichlet allocation", *The Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003
DOI: https://dl.acm.org/doi/10.5555/944919.944937

[8] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, "Distributed representations of words and phrases and their compositionality", *Advances in neural information processing systems*, pp. 3111-3119, 2013.
DOI: http://doi.org/10.48550/arXiv.1310.4546

[9] Q. V. Le, T. Mikolov, "Distributed Representations of Sentences and Documents", *International Conference on Machine Learning*, pp. 1188-1196, 2014.
DOI: http://doi.org/10.48550/arXiv.1405.4053

[10] J. Pennington, R. Socher, C. D. Manning, "Glove: Global vectors for word representation", *2014 Conference of Empirical Methods in Natural Language Processing*, pp. 1532-1543, 2014. Available from: https://nlp.stanford.edu/pubs/glove.pdf

[11] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is All You Need", *Advances in neural information processing systems*, 2017.
DOI: http://doi.org/10.48550/arXiv.1706.03762

[12] S. Oh, M. Jang, H. S. Song, "A BERT-based Transfer Learning Model for Bidirectional HR Matching", *Journal of Information Technology Applications & Management*, pp. 33-43, 2021.
DOI: http://doi.org/10.21219/jitam.2021.28.4.033

[13] K. Cho, C. Gulcehre, D. Bahdanau, H. Schwen, Y. Bengio, "Learning Phase Representation using RNN Encoder-Decoder for Statistical Machine Translation", *arXiv preprint arXiv*:1406.1078, 2014.
DOI: http://doi.org/10.48550/arXiv.1406.1078

[14] F. Prieto, H. F. T. Braga, "Contribution of Artificial Intelligence in B2B Sales: A Danfoss Case Study", *International Journal of Advanced Engineering Research and Science*, vol. 8, no. 4, pp. 225-234, 2021.
DOI: http://doi.org/10.22161/ijaers.84.27

[15] Y. L. Yoon, Y. Yoon, H. Nam, J. Choi, "Buyer-supplier matching in online B2B marketplace: An empirical study of small and medium sized enterprises(SMEs)", *Industrial Marketing Management*, vol. 93, pp. 90-100, 2021.
DOI: http://doi.org/10.1016/j.indmarman.2020.12.010

최 정 석(Jungsuk Choi) [정회원]

- 2018년 6월 : 호서대학교 글로벌 창업대학원 창업컨설팅학과 (석사)
- 2020년 2월 ~ 현재 : 호서대학교 벤처대학원 융합공학과 박사과정
- 1992년 1월 ~ 현재 : 대한무역투자진흥공사 글로벌인재 센터

〈관심분야〉
추천시스템, NLP, 컴퓨터 비전, 딥러닝

문 남 미(Nammee Moon) [정회원]

- 1985년 2월 : 이화여자대학교 컴퓨터공학 (석사)
- 1998년 2월 : 이화여자대학교 컴퓨터공학 (박사)
- 1999년 9월 ~ 2000년 2월 : 아주대학교 디지털미디어학과 조교수
- 2000년 3월 ~ 2003년 2월 : 이화여자대학교 조교수
- 2003년 3월 ~ 2008년 2월 : 서울벤처정보대학원대학교 디지털미디어학과 부교수
- 2008년 3월 ~ 현재 : 호서대학교 컴퓨터학부 교수

〈관심분야〉
빅데이터 처리 및 분석, 딥러닝, HCI, Social Learning