

# ShipMate: 딥러닝을 이용한 해상물류 전문상담 챗봇

유현수<sup>1</sup>, 남서연<sup>1</sup>, 백주영<sup>2</sup>, 안소영<sup>2</sup>, 황세진<sup>3</sup>, 이규영<sup>4</sup>

<sup>1</sup>이화여자대학교 통계학과

<sup>2</sup>숙명여자대학교 소프트웨어학부

<sup>3</sup>서울시립대학교 환경공학부

<sup>4</sup>한국과학기술원 전산학부 정보보호대학원

hsu607@ewhain.net, systats11@ewhain.net, juyeong@sookmyung.ac.kr, 2011635@sookmyung.ac.kr,  
s09j10@uos.ac.kr, leeahn1223@kaist.ac.kr

## ShipMate: Marine Logistics Specialist Consultation Chatbot using Deep Learning

Hyun-Su Yu<sup>1</sup>, Seo-Yeon Nam<sup>1</sup>, Joo-Yeong Baek<sup>2</sup>, So-Yeong Ahn<sup>2</sup>, Se-Jin Hwang<sup>3</sup>, Gyu-Young Lee<sup>4</sup>

<sup>1</sup>Dept. of Statistics, Ewha Womans University

<sup>2</sup>Dept. of SoftWare, Sookmyung Women's University

<sup>3</sup>Dept. of Environmental Engineering, University Of Seoul

<sup>4</sup>Graduate School of Information Security, KAIST

### 요 약

본 논문에서는 한국무역협회(KITA)의 오픈상담 자료들을 바탕으로, 딥러닝 기술을 이용하여 구현한 해상물류 대화형 챗봇 ShipMate를 제안한다. 챗봇 ShipMate는 KoGPT2를 활용한 답변과 Doc2Vec 기반의 유사 상담사례 추천이 가능하고, 무역상담을 시간 제약 없이 진행할 수 있기 때문에, 기존 해상물류 서비스의 접근성을 한층 더 높일 수 있으며 이를 실험을 통해 입증하였다.

### 1. 서론

‘대화형 인공지능(ChatBot)’은 인간과 자연스러운 방식으로 대화할 수 있도록 하는 애플리케이션이다. 자연어 처리(NLP) 기술을 사용하여, 다양한 데이터를 통해 사용자가 웹 서비스 및 앱 등과 상호 작용할 수 있도록 돕는다.[1] 최근에는 해운 항만 분야에서도 챗봇을 비롯한 빅데이터 기반 서비스를 구축하려는 노력이 이어지고 있다. 한국해양수산개발원에서 발표한 ‘항만부문 정책 의사결정 지원 빅데이터 플랫폼 구축 방안 연구’에 따르면 국내 해운 항만 분야에서 빅데이터를 활용하는 방안이 수립되어야 한다고 밝히고 있다.[2]

따라서 업무 효율을 증대시키고 편리성과 사용자 접근성을 향상시키기 위해서는, 해운항만 관련 사이트에 빅데이터 기반의 챗봇 기술을 도입하는 것이 필요하다.

본 논문에서는 한국무역협회(KITA) TradePro 플랫폼의 오픈상담 데이터로 학습한 ShipMate 챗봇 모델을 제안하고 실험을 통해 그 유용성을 입증하고자 한다.

### 2. 이론적 배경

#### 1) GPT-2

GPT-2는 주어진 텍스트의 다음 단어를 예측할 수 있도록 학습한 언어모델로, 문장 생성에 최적화되어 있다. 이는 트랜스포머(transformer)의 디코더(decoder) 스택만 사용하여 예측 성능을 높였다. KoGPT2는 GPT-2의 부족한 한국어 처리 성능을 극복하기 위해 40GB 이상의 한국어 텍스트로 학습한 한국어 디코더 언어모델이다.

#### 2) Doc2Vec

Doc2Vec은 Word2Vec을 문장 단위로 확장시킨 문서 임베딩 기법이며, 문서 임베딩은 해당 문서의 주제 정보를 함축하고 있다. PV-DM<sup>1</sup>과 PV-DBOW<sup>2</sup>의 2가지 모델이 존재하는데, PV-DM 모델은 단어가 등장하는 순서를 고려하기 때문에 일반적으로 더 좋은 성능을 보인다.

### 3. 한국무역협회 해상물류 챗봇

#### 1) ShipMate 개요

<sup>1</sup> the Distributed Memory Model of Paragraph Vector

<sup>2</sup> the Distributed Bag of Words version of Paragraph Vector

본 논문에서 개발한 챗봇 ShipMate는 KoGPT2 기반의 문장생성 모델에 한국무역협회(KITA) 웹사이트에서 크롤링한 무역상담 데이터를 학습시켰다. 그리고 기본적인 대화기능에 덧붙여 연관된 상담사례 추천, 1:1 전문가 상담연결 등의 다양한 기능을 제공한다.

2) 데이터 수집

본 논문은 KITA 웹사이트의 오픈상담 목록을 크롤링하여 만든 무역상담 데이터셋을 사용하였다.

category	title	question	answer	url
관세/통관	중국에 화장품 수출	중국에 화장품 수출	안녕하세요, 반갑습니다.	<a href="https://www.kita.net/e">https://www.kita.net/e</a>
관세/통관	수출 허가 처리 관련	수출품 허가로 인한 글	안녕하세요 제조업종	<a href="https://www.kita.net/e">https://www.kita.net/e</a>
무역거래/결제	특송을 이용한 아마존	특송을 이용한 아마존	안녕하세요 아마존의	<a href="https://www.kita.net/e">https://www.kita.net/e</a>

(그림 1) 무역상담 데이터셋 (일부 발췌)

무역상담 데이터셋의 수량은 3904 개이며 5 개 컬럼으로 구성되어 있다. 모델을 학습시킬 때는 'question'을 질문 텍스트, 'answer'를 답변 텍스트, 'category'를 분류 레이블로 사용하고, 'title'과 'url'은 사용자의 질문 내용과 유사한 상담 사례를 제공할 때 적용하였다.

3) ShipMate 챗봇모델 구축

3.1 KoGPT2 기반 문장 생성 모델

본 논문에서는 KoGPT2 모델에 한국어 데이터셋을 학습시켜 해상물류 무역상담 챗봇 서비스를 구현하였다. 무역상담 데이터셋을 학습 및 테스트 데이터로 분리한 후 모델 훈련을 진행하였으며, 답변의 품질을 향상시키기 위해 추가적으로 fine-tuning 을 실시하고 세부 파라미터를 수정하여 최종 모델을 완성하였다.

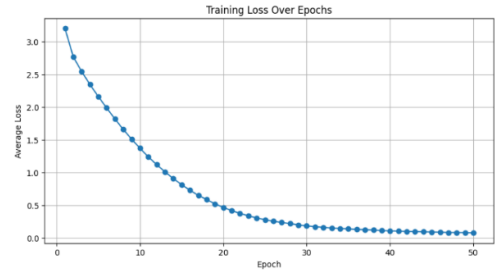
3.2 Doc2Vec 기반 유사 상담사례 추천

본 논문에서는 'question' 데이터에 한국어 전처리와 토큰화를 진행한 후, 파이썬 머신러닝 패키지인 Gensim doc2vec 을 사용하여 문서벡터를 학습시켰다. 새로운 질문을 입력하면 코사인 유사도를 계산하여 기존 상담사례 중에서 입력을 받은 질문과 가장 유사한 상담사례를 추천해 준다.

4. 실험 및 고찰

1) KoGPT2 기반 문장생성 실험결과

본 논문에서 구현한 문장 생성 모델의 임베딩 차원은 768 이고, 입력 가능한 토큰 수는 최대 1024 개이다. 모델에 데이터를 학습시킬 때 Train data 와 Test data 의 비율은 8:2 로 설정하였다. epoch 는 50 이며, 이때의 average loss(평균 손실)이 0.0800 으로 가장 낮았다. 그림 2는 모델 훈련을 진행하면서 지속적으로 감소하는 average loss 를 그래프로 시각화한 것이다.



(그림 2) 모델 훈련 시 average loss 변화 파라미터의 경우, 모델 훈련과 답변 생성 시 표 1의 값을 사용할 때 가장 답변 성능이 좋았다.

<표 1> 모델 훈련 및 답변 생성 시 파라미터 값

파라미터	훈련 batch size	훈련 max length <sup>3</sup>	답변 max length	답변 temperature <sup>4</sup>
값	8	512	512	0.7

2) 유사 상담사례 추천 실험결과

유사한 상담 사례 추천 알고리즘에서 Doc2Vec 모델 학습 시 파라미터는 vector\_size<sup>5</sup> = 50, alpha<sup>6</sup> = 0.025, min\_alpha = 0.00025, min\_count<sup>7</sup> = 5, workers = 8, window<sup>8</sup> = 5, dm = 1<sup>9</sup>으로 설정하였다. 학습시킨 Doc2Vec 모델을 이용하여 유사질문 추천알고리즘을 실행한 결과, 그림 3 과 같이 입력 받은 새로운 질문과 가장 코사인 유사도가 높은 질문을 출력한 것을 확인할 수 있었다.

index	category	title	question	cosine_similarity
0	해외인증	호주 의 표기가 CE 등록	안녕하십니까. 의뢰기기 CE 등록 관련하여 문의드립니다. 호주에서 의뢰기기 CE 등록 시 이에 대한 절차 및 발생하는 비용 등 관련 내용을 알고싶어 문의 드립니다. 감사합니다.	0.9602080583572388

(그림 3) 유사 질문 추천 알고리즘 실험 결과

5. 결론

본 논문에서는 KoGPT2 문장 생성모델에 기반한 해상물류 무역상담 챗봇 ShipMate 를 제안하였다. 사용한 데이터가 상세하고 길어서 일상 대화처럼 자연스러운 답변을 제공하는 데에는 다소 한계가 있었지만, 사용자가 전문적인 답변을 시간의 제약 없이 제공받을 수 있다는 점에서는 그 의의가 크다. 향후 기존의 챗봇 서비스들에 ShipMate 모델을 도입한다면, 사용자 접근성을 더욱 높일 수 있을 것이라 예상된다.

※ 본 논문은 해양수산부 실무형 해상물류 일자리 지원사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

참고문헌

[1] 박동아. 인공지능 기반 대화형 공공 행정 챗봇 서비스에 관한 연구. 멀티미디어학회논문지, 20(8), 1347-1356, 2017  
 [2] 이기열 외 2 인. 항만부문 정책의사결정 지원 빅데이터 플랫폼 구축 방안 연구. 한국해양수산개발원 기본 연구 보고서, 2021-13, 237

3 생성할 최대 텍스트 길이  
 4 생성된 텍스트의 다양성을 조절하는 값  
 5 벡터 차원의 크기  
 6 학습 진행 시 가중치를 얼마나 업데이트할지 결정하는 값

7 학습에 사용할 최소 단어 빈도 수  
 8 모델 학습 시 앞뒤로 보는 단어의 수  
 9 학습방법(1 = PV-DM, 0 = PV-DBOW)