



MURERBOT: Massive User REview-based Responsive Bot

Yu-Jin Jo[†], Han-Seong Bae[†], Hyun-A Jo, Eun-Seo Kim, Keejun Han^{*}

School of Computer Engineering, Hansung University

ABSTRACT

Reviews of actual users have a great influence on consumers' judgment when purchasing products. Existing shopping malls tend to focus on positive reviews, and product information and reviews of real users are provided in a form that is difficult to grasp at a glance. Therefore, this paper proposes MURERBot, a shopping mall review analysis chatbot system that provides a summary of not only positive reviews but also negative reviews at a glance using sentiment analysis model. MURERbot converts the user's question into an embedding vector through the SBERT model and the FastText model, and then provides product recommendation, product information, and review summaries through the similarity between the vectors. It also uses a chatbot module that supports interactive queries so that users can easily obtain information.

© 2023 KKITS All rights reserved

KEYWORDS Embedding vectors, Sentiment analysis, Chatbot, Multi-label-classification, Reviews, Cosine similarity, BERT, FastText, Gensim, GRU

1. 서론

대부분의 쇼핑몰들은 상품 판매 촉진을 위해 긍정적인 리뷰를 중점적으로 노출하고 있다. 이에 소비자들은 상품에 대한 종합적인 정보를 파악하기 위해 많은 시간을 할애한다. 따라서 본 논문에서는 상품의 정보 파악에 편의성을 제공하는 쇼핑몰 리

뷰 분석 챗봇 시스템인 MURERBot을 제안한다. MURERBot은 문장 의도를 파악하여 리뷰에 대한 속성 분류와 감성 분류, 상품 추천, 상품 상세정보를 제공한다. 사용자 인터페이스는 소비자가 원하는 상품 정보를 편리하게 얻을 수 있도록 대화 형식인 챗봇을 이용하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는 시스템을 구현하는데 사용된 모델과 주요 알고리즘을 살펴본다. 제 3장에서는 시스템의 구조, 제 4장에서는 결론을 기술한다.

2. 관련 연구

[†]These two authors contributed equally to this work.

^{*}Corresponding author is with the School of Computer Engineering, Hansung University, 116 Samseongyo-ro 16gil, Seoul, 02876

E-mail address: keejun.han@hansung.ac.kr

2.1 사용 모델

2.1.1 SBERT (Sentence-BERT)

SBERT[1]는 주어진 문장에서 의미 있는 임베딩을 출력층에서 반환하는 BERT 네트워크이다. SBERT는 입력 문장이 들어오면 고정된 크기의 벡터를 도출하고, 도출된 벡터는 코사인 유사도, 맨하탄 거리, 유클리드 거리와 같은 유사성 척도를 계산하기 적당한 임베딩을 도출해 준다. 또한 리소스의 효율적인 클러스터링 및 semantic similarity search task를 수행한다.

2.1.2 감성 분석 모델

본 논문에서는 BERT[2] 모델을 이용한 Sequence Classification을 통해 감성 분석 모델을 구현한다. 본 논문에서는 감성 분석 모델 학습에 20,000개의 네이버 쇼핑물 리뷰 데이터와 네이버 영화 리뷰 데이터(NSMC) 5,200개를 이용한다. 모델 학습을 위한 파라미터는 설정은 다음과 같다. 문장 최대 길이는 128, batch size는 16, epoch는 10, learning rate는 $2e-5$ 로 설정한다. BERT multilingual base model, KoBERT, KR FinBert를 기반으로 fine-tuning한 결과는 표 1과 같다. BERT 다국어 모델이 0.80으로 가장 높은 정확도를 보인다. 따라서, 본 논문에서는 BERT 다국어 모델을 이용한 감성 분석 모델을 통해 네이버 쇼핑물 리뷰의 감성 판단을 진행한다.

표 1. 감성 분석 모델 fine-tuning 결과
Table 1. results of fine-tuning a sentiment analysis model

model	accuracy
bert-base-multilingual-cased	0.80
monologg/kobert	0.55
snunlp/KR-FinBert-SC	0.78

2.1.3 Gensim - FastText

본 논문에서는 대규모 말뭉치 유사성 검색을 위해 Gensim의 FastText 모델을 활용하여 유사한 단어를 추출한다. FastText 모델은 단어의 단순성과 독립성을 위해 자음, 모음 단위로 분리하고 벡터화를 진행하여 유사한 단어를 도출해낸다. 본 논문에서는 사용자의 상품 검색 단어가 유사하지만 다른 단어일 경우에 검색이 되지 않는 문제의 해결방안으로 FastText 모델을 사용한다.

2.1.4 GRU(Gated Recurrent Unit)

GRU모델[3]은 LSTM 모델의 장기 의존성 문제를 해결하면서, 매번 은닉상태 업데이트 계산을 줄여 LSTM보다 간단한 구조를 가지는 모델이다. 본 논문에서는 GRU 모델을 사전에 디자인, 무게, 성능, 소음, 사이즈, 만족도, 6가지 속성으로 라벨링한 리뷰 데이터로 학습시켰다. 학습된 모델은 예측하고자 하는 리뷰에 대해 위 6가지 속성 중 해당되는 속성으로 예측, 분류하는데 활용된다.

2.2 주요 알고리즘

2.2.1 TextRank

TextRank[4]는 하이퍼 링크 구조를 가지는 웹 문서의 중요도에 따라 가중치를 부여하는 Google의 PageRank[5]를 활용한 그래프 기반 순위 알고리즘이다. TextRank는 문서 내의 문장 또는 단어를 이용하여 문장의 Ranking을 계산한다. 본 논문에서는 TextRank 알고리즘을 사용하여 중요 문장을 추출하여 상품의 긍정과 부정 리뷰를 요약한다.

2.2.2 Cosine Similarity

Cosine Similarity[6]는 자연어 처리 분야에서 임베딩 벡터 사이의 유사도를 비교할 때 벡터 사이

의 각도를 구해서 얼마나 유사한지 나타낸 수치이다. 본 논문에서는 사용자가 입력한 문장과 사전에 정의된 시나리오 간의 유사도 계산을 통해 의도를 파악하고, 사용자가 추천받고자 하는 상품의 특징과 상품 리뷰의 유사도를 측정해 추천 알고리즘을 구현한다.

3. 시스템 모듈

본 논문이 제시하는 시스템은 챗봇을 이용하여 사용자에게 상품 추천, 상품 리뷰 요약본 및 상품 상세정보를 제공한다. 이를 위해 시스템 내부에는 챗봇 모듈, 추천 모듈, 요약 모듈, 상품 상세정보 제공 모듈이 존재한다. 시스템 내부 모듈 간의 구조는 <그림 1>과 같고, 시스템의 구체적인 구조는 <그림 2>와 같다.

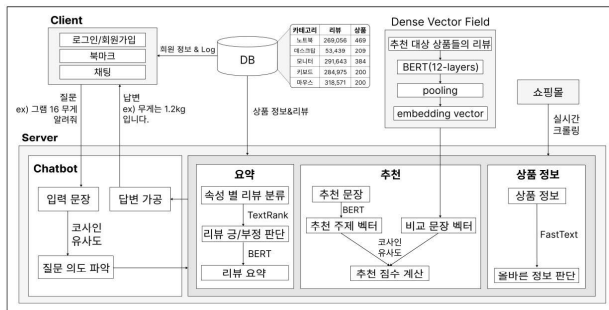


그림 1. 시스템 구조도
Figure 1. System Architecture

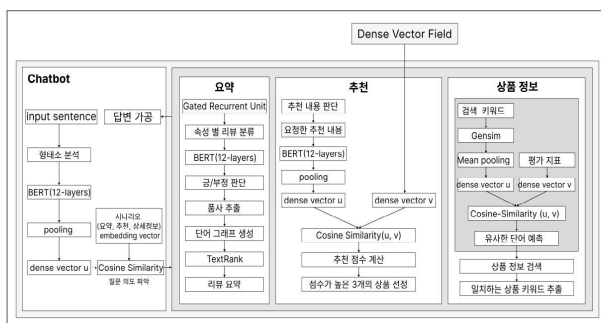


그림 2. 서버 구조도
Figure 2. Server Architecture

3.1 챗봇 모듈

사용자의 질문을 받아 형태소 분석을 통해 제품명과 제품명을 제외한 문장으로 분리한다. 제품명을 제외한 문장에 대한 임베딩 벡터를 SBERT[1]를 통해 추출하여 정해진 시나리오의 문장들과의 코사인 유사도[6]를 구한다. 코사인 유사도 최대값을 갖는 시나리오가 임계치를 넘는 경우 사용자의 질문을 해당 시나리오로 판단하고 판단한 시나리오에 따라 아래와 같은 작업을 수행하여 답변한다.

3.2 추천 모듈

본 논문의 추천 알고리즘은 추천하고자 하는 카테고리의 리뷰들과 사용자의 질문을 [1]을 통해 임베딩 벡터를 도출하고, [6]을 통해 리뷰와 사용자 질문의 유사도를 구한다. A 상품에 임계치 이상의 유사도를 가진 리뷰가 존재한다면 A 상품의 Recommend Score를 증가시킨다. 최종적으로 하나의 추천 카테고리에 대해 모든 상품의 Recommend Score를 계산하여 상위 3개의 상품을 추천 상품으로 결정한다. 추천 모듈을 사용한 시스템 인터페이스는 <그림 3>과 같다.



그림 3. 추천 모듈 시스템 인터페이스
Figure 3. Recommend Module System Interface

3.3 요약 모듈

사용자가 원하는 상품에 대한 리뷰 요약본을 제공한다. DB로부터 상품의 상세정보와 리뷰 데이터를 가져온다. 리뷰 데이터의 경우 감성 분석 모델 [2]를 통해 사전에 분류하여 DB에 삽입한다. 요약본은 긍정 또는 부정으로 라벨링된 리뷰 데이터만 사용한다. 긍정 리뷰와 부정 리뷰를 각각 TextRank[5]를 통해 2개의 문장으로 요약하여 사용자에게 제공한다. 요약 모듈을 사용한 시스템 인터페이스는 <그림 4>와 같다.



그림 4. 요약 모듈 시스템 인터페이스
Figure 4. Summary Module System Interface

3.4 정보 제공 모듈

사용자가 원하는 상품 정보를 제공한다. 코사인 유사도[6]를 통해 사용자의 입력 문장 내 상품 정보를 포함하는지 파악하고 해당하는 상품 정보를 DB에 검색하여 상품 정보가 존재할 경우 데이터를 가져온다. DB에 상품 정보가 없을 경우 네이버 쇼핑몰 크롤링을 통해 상품 정보를 가져온다. 검색되지 않는 경우 FastText[3]를 사용하여 유사 단어를 추출하고 상품 정보를 검색한다. 검색이 되지 않고 사용자가 원하는 상품 정보에 대한 언어와 DB내에 있는 정보에 대한 언어가 다를 경우 파과고 API를 사용하여 검색한다. 정보 제공 모듈을 사용한 시스템 인터페이스는 <그림 5>와 같다.

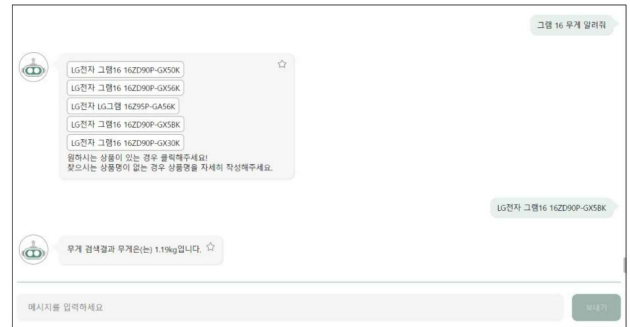


그림 5. 정보 제공 모듈 시스템 인터페이스
Figure 5. Provide Information System Interface

4. 결론

본 논문에서 제시한 MURERbot은 챗봇과의 대화형 질의를 통해 감성 분류를 진행한 리뷰의 전체 비율 및 요약과 속성별 비율 및 요약을 한 페이지로 제공하며, 추천 상품에 대한 리스트와 상품에 대한 상세 정보를 제공하여 사용자가 상품 구매를 결정하는 시간을 절약할 수 있게 한다.

References

- [1] R. Nils, and I. Gurevych, *Sentence-bert: Sentence embeddings using siamese bert-networks*, arXiv preprint arXiv:1908.10084, 2019.
- [2] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, *Bert: Pre-training of deep bidirectional transformers for language understanding*, arXiv preprint arXiv:1810.04805, 2018.
- [3] J. Chung, C. Culcehre, K. Cho, and Y. Bengio, *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*, 2014.
- [4] R. Mihalcea, and P. Tarau, *TextRank: Bringing order into text*, Proceedings of the

2004 conference on empirical methods in natural language processing, 2004.

- [5] S. Brin, and L. Page, *The anatomy of a large-scale hypertextual web search engine*, Computer networks and ISDN systems, Vol. 30, Issues 1-7, pp. 107-117, 1998.
- [6] A. Singhal, *Modern information retrieval: A brief overview*, IEEE Data Eng. Bull. Vol. 24, No. 4, pp. 35-43, 2021.

대화형 상품 리뷰 분석 시스템

조유진, 배한성, 조현아, 김은서
한성대학교 컴퓨터공학과

요 약

실제 사용자들의 리뷰는 소비자들의 상품 구매 판단에 큰 영향을 미칩니다. 그러나 기존의 쇼핑몰은 주로 긍정적인 리뷰만을 강조하고 있으며, 상품 정보와 실제 사용자들의 후기를 한눈에 파악하기 어렵게 제공하고 있습니다. 따라서 이 논문에서는 긍정적 리뷰 뿐만 아니라 부정적 리뷰도 한눈에 볼 수 있는 요약본을 제공하는 쇼핑몰 리뷰 분석 챗봇 시스템인 MURERBot을 제안합니다.

감사의 글

본 논문은 한성대학교의 2023학년도 학술연구조성비를 지원 받음.