

Word2Vec 모델 기반의 유사도를 이용한 상품기획 모델

안영휘^o, 박구락^{*}

^o국립공주대학교 컴퓨터공학과,

^{*}국립공주대학교 컴퓨터공학과

e-mail: yhahn02@naver.com^o, ecgrpark@kongju.ac.kr^{*}

Product Planning using Similarity Analysis Technique Based on Word2Vec Model

Yeong-Hwi Ahn^o, Koo-Rack Park^{*}

^oDept. of Computer Engineering, Kongju National University,

^{*}Dept of Computer Engineering, Kongju National University

● 요약 ●

소비자가 남긴 댓글이나 상품평은 상품기획의 주요 정보가 될 수 있다. 본 논문에서는 버티컬 무소음 마우스 7,300개에 대한 온라인 댓글을 딥러닝 기술인 Word2Vec을 이용하여 유사도 분석을 시행하였다. 유사도 분석결과 클릭 키워드에 대한 장점으로 소리(.975), 버튼(.972), 무게(.971)가 분석되었으며 단점은 가법다(.959)이었다. 이는 구매 상품에 대한 소비자의 의견, 태도, 성향 및 서비스에 대한 포괄적인 의견들을 데이터화 하여 상품의 특징을 분석할 수 있는 의미있는 과정 이라고 볼 수 있다. 상품기획 프로세스에 딥러닝 기술을 통한 소비자의 감성분석자료 포함시키는 전략을 적용한다면 상품기획의 시간과 비용투자의 경제성을 높일 수 있고 나아가 빠르게 변화하는 소비자의 요구사항을 적기에 반영할 수 있을 것으로 생각된다.

키워드: 상품(Product), 유사도(Similarity), 기법(Technique), 텍스트 전처리(Text Preprocessing), 워드투벡터(Word2Vec)

I. Introduction

온라인 환경 속에서 이루어지는 여러 형태의 상품 구매는 구매의 경로와 현상, 소비자의 요구를 반영한 상품 출시의 중요한 정보가 될 수 있다[1]. 특히, 소비자가 남긴 댓글이나 상품평은 자신의 경험을 다른 소비자나 제조사와 상호작용한 결과로서 의미 있는 정보가 된다. 댓글이나 상품평에 텍스트 마이닝을 적용하면 텍스트에 나타난 소비자의 의견, 태도, 성향 뿐 아니라 서비스에 대한 소비자들의 의견들을 데이터화 하여 트렌드를 분석할 수 있다. 본 연구에서는 Word2Vec Model를 적용하여 기업 상품기획의 핵심요소인 소비자의 질적 데이터를 체계적으로 분석하여 최적의 상품기획, 신상품 출시에 적용 가능한 개선, 대안방안을 모색할 수 있는 자료를 제시하고자 한다.

II. Preliminaries

1. Related works

1.1 Word2Vec

Word2Vec은 구조가 간단하면서 성능이 우수하여 임베딩에 가장 널리 사용된다. 이 모델은 SG(Skip-gram)와 CBOW(Continuous Bag of Word) 두 가지 학습 아키텍처를 제안한다. CBOW에서는 주변 단어를 통해 중심단어를 예측하며, SG은 중심 단어에서 주변단어를 예측 한다[2].

1.2 Similarity

유사도는 벡터, 점 등으로 표시되는 일정한 기준이 없는 두 지점 간의 거리를 측정하여 유사점을 분석한다. 유사도를 측정 때 사용되는 알고리즘은 피어슨 상관계수, 코사인 벡터, 유클리드 거리가 있다[3].

III. The Proposed Scheme

2.1 Model Configuration

다음의 Fig. 1.은 제안 모델의 구성도이다.

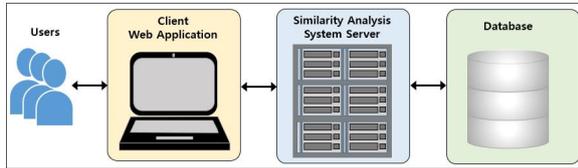


Fig. 1. Model Configuration

2.2 System Architecture

제안시스템의 아키텍처는 세부부분으로 이루어져 있다. 첫째, 웹크롤링을 통해 빅데이터를 수집할 특정 상품을 선택하여 대량의 데이터를 수집한다. 둘째, 소비자 댓글 유사도분석 시스템은 웹 크롤링을 통해 수집된 비정형데이터를 정형데이터로 변환한다. 자연어 처리된 데이터는 Word2Vec 임베딩 학습을 통해 단어 유사도를 분석한다. 셋째, 데이터베이스 파트로서 유사도분석 시스템 파트에서 생성된 분석 데이터 내역을 저장하여 웹 애플리케이션을 통하여 분석내용을 시각화한다.

2.3 System Process

다음은 5단계로 구성된 제안모델의 소비자 댓글 유사도 분석 프로세스이다. 첫째, 빅데이터를 수집하기 위해 사이트, 대상 상품, 기간을 선택한다. 둘째, 실시간으로 빅데이터를 수집한다. 셋째, 실시간으로 수집된 비정형데이터를 자연어 전처리 과정을 통해 정형데이터로 변환한다. 넷째, Word2Vec 임베딩 학습을 통해 단어 유사도를 분석한다. 다섯째, 장단점을 TSNE로 시각화 한다.

IV. Test And Results

Table1, 2은 버티컬 무소음 마우스에 대한 온라인 소비자의 댓글 7,300개를 예비 적용한 개발환경 및 학습결과이다.

Table 1. Development Environment

구분	내용
데이터	온라인 소비자 댓글 7,300개
대상 제품	버티컬 유/무선 무소음 마우스
개발 언어	Python
워드 임베딩	Word2Vec

Table 2. Learning Result

구분	학습 결과			
	장점		단점	
	내용	유사도	내용	유사도
클릭	소리	.975	가볍다	.959
	버튼	.972		
	무게	.971		

V. Conclusions

본 연구에서는 온라인 소비자 댓글 일부를 샘플링하여 유사도 분석을 진행 하였다. 유사도 분석결과 클릭 키워드의 장점으로 소리(.975), 버튼(.972), 무게(.971) 이라는 분석 되었으며, 단점으로는 가볍다(.959)가 분석 되었다. 본 결과는 매우 중요한 소비자 피드백 내용으로 기업에서는 분석된 내용을 토대로 상품기획 프로세스에 보완 함으로써 긍정적인 영향을 미칠 것으로 기대한다. 향후 연구에서는 소비자 댓글을 25,000개 수집하여 유사도분석 및 소비자의 감성분석 정확성을 강화한 기법을 반영한 상품기획프로세스 구현에 대한 연구가 계속 되어야 할 것이다.

REFERENCES

- [1] Lee, Mi-Jin, and Jeong-Won Han, "Analysis of Interior Design Elements in Food and Beverage Applying the Psychological Process of Purchasing (AIDMA) Model," Korean Institute of Interior Design Journal 21, Vol. 21, No. 2, pp.112-122, April 2012.
- [2] Kang, Hyungsuc, and Janghoon Yang, "Optimization of Word2vec Models for Korean Word Embeddings," Journal of Digital Contents Society, Vol. 20, No. 4, pp.825-833, April 2019.
- [3] Seungsoo Lim, and Hayon Lee, "Prediction of New Drug-Side Effect Relation using Word2Vec Model-based Word Similarity," Journal of KIIT. Vol. 18, No. 11, pp. 25-33, Nov 2020.