

상품 리뷰 분석을 통한 사용자 맞춤형 추천 시스템

황도연 · 배상중 · 김창수 · 정희경

배재대학교

Customized recommendation system through product review analysis

Doyeun Hwang · Sangjung Bae · Changsoo Kim · Heokyung Jung

PaiChai University

E-mail : hwangdy25@pcu.ac.kr, basaju8242@gamil.com, {ddoja, hkjung}@pcu.ac.kr

요 약

전통적인 방식의 추천 시스템은 사용자가 독립적으로 행동한다는 가정하에 개발된 방식이며, 단순히 상품을 나열하거나 상품의 속성과 사용자의 기호를 연관하는 기능이 부족하여 가독성과 효율성이 떨어지는 문제점이 있다. 이를 해결하기 위해 본 논문에서는 상품 리뷰 데이터를 크롤링을 한 뒤 R을 이용한 텍스트 마이닝 기법을 사용하여 비정형의 리뷰 데이터를 사용자의 구매이력과의 연관 분석을 통해 의미 있는 정보로 가공하여 사용자 맞춤형 정보를 제공하는 시스템을 제안한다. 이를 통해 사용자는 방대한 양의 상품 리뷰 데이터를 분석할 필요 없이 자신에게 필요한 데이터만을 제공받을 수 있게 되어 사용자의 의사결정에 도움 될 것으로 사료된다.

ABSTRACT

The traditional recommendation system is developed on the assumption that users behave independently, and have problem of readability and efficiency are inferior due to simply sort products or lack of function for associate product attributes with user's taste. To solve this problem in this study we propose a system that provides user customized information that the analysis of the unstructured review data with the purchase histories of users processed with meaningful information after crawling product review data using text mining with R. This allows to help user make decisions can be provided only necessary information without analyze massive amounts of products review data.

키워드

상품 리뷰, 상품 속성, 텍스트 마이닝, 추천 시스템

I. 서 론

인터넷에서 발생하는 데이터의 양이 기하급수적으로 늘어나고 있다. 이러한 데이터의 대부분은 비정형 데이터로써 기업 또는 사용자들이 필요로 하는 정보가 되기 위해서는 데이터를 취합하고 가공 및 분석하는 단계를 거쳐야 한다. 또한 개인 맞춤형 서비스가 발달하고 개인의 개성을 중요시 하는 사회가 도래하면서 사용자들의 참여에 의해 발생한 상품 리뷰나 사용 후기 데이터의 양이 증가하고 있다. 이를 위해 상품 리뷰 데이터를 분석하여 결과를 제공하는 시스템과 관련된 연구가 진행되고 있다. 기존의 분석 시스템들은 사용자의 리뷰 중에서 평점을 중심으로 분석하기 때문에 평점 자체의 신뢰성 문제가 제기 될 수 있다[1]. 또한 상품의 분류와 속성을 반영하여 리뷰의 감

정 분석을 통해 분석 결과를 제공한다[2]. 기존의 시스템은 리뷰 데이터 분석에 있어서 상품의 속성 및 상품이 속한 세부 부류만을 반영하기 때문에 사용자 맞춤형 정보를 제공하기에 부족하다.

본 논문에서는 비정형적인 리뷰 데이터를 의미 있는 정보로 가공하고, 가공한 정보를 다시 사용자에게 추천하는 시스템을 제안한다. 상품 리뷰 데이터를 크롤링 한 뒤 R을 이용한 텍스트마이닝 기법을 사용하여 사용자 맞춤형 정보를 제공하는 시스템이다.

II. 시스템 설계

2.1 시스템 구조

리뷰 데이터를 수집 및 분석하는 시스템의 구조는 그림 1과 같다. 시스템은 크롤링, 형태소 분

석, 상품 특징 추출, 리뷰 데이터 구축 순서로 동작한다. 크롤링 단계에서는 국내 최대 규모의 인터넷쇼핑몰 사이트 중 한 곳의 상품 분류, 이름, 리뷰 등을 수집한다. 그 다음 단계에서 수집한 데이터를 한국어 형태소 분석 패키지 KoNLPy를 활용하여 형태소 분석을 수행한다. 각 문장에서 명사와 형용사, 부사를 추출하여 각각의 빈도수를 계산하여 상품의 특징과 각 특징의 점수를 계산하고 이를 리뷰 데이터베이스에 저장한다. 사용자가 구매 의향이 있는 상품을 검색하면 다른 사용자들의 상품 리뷰 분석 결과와 사용자의 구매 이력을 연관 지어 맞춤형 정보를 제공받을 수 있다.

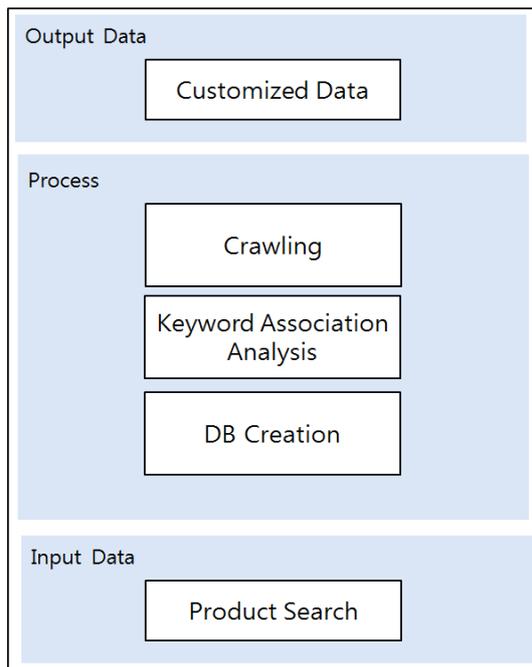


그림 1. 시스템 구조도

2.2 크롤링

온라인 쇼핑 시 이용되는 대표적인 쇼핑물로부터 상품의 분류, 이름, 리뷰 데이터를 수집한다. 그 중 비교적 많은 리뷰데이터가 있는 '유아동'에 해당하는 상품의 데이터만 수집한다. 또한 사이트 내 존재하는 사용자의 구매데이터를 활용하여 '소비금액'과 '상품 카테고리'를 추출하여 수집한다.

2.3 형태소분석

Python 환경에서 한국어 정보처리를 위한 KoNLPy 패키지를 사용하여 형태소 분석을 수행한다. 리뷰 데이터를 하나의 문장으로 분할 한 뒤, 각 문장에서 명사와 형용사, 부사를 추출한다.

2.4 키워드 연관 분석

형태소 분석을 통해 각 문장에서 상품의 분류를 나타내는 명사와 상품의 속성을 나타내는 형용사, 그 형용사를 수식해주는 부사만을 추출한

다. 그 다음 각각의 빈도수를 계산하여 상품의 특징과 각 특징의 점수를 계산하고, 사용자의 구매 이력 데이터에서 상품의 금액과 구매 빈도가 높은 상품 카테고리를 추출하여 연관 분석을 통해 사용자 맞춤형 추천 정보를 생성한다.

III. 결 론

본 연구에서는 비정형의 리뷰데이터와 사용자의 구매이력의 연관 분석을 통해 사용자에게 맞춤형 정보를 제공하는 시스템을 제안한다. 제안한 시스템의 데이터 분석 결과로 상품 속성 중 핵심이 되는 단어들을 제공함으로써 사용자는 한눈에 핵심 속성을 파악할 수 있다 또한, 리뷰에 대한 분석 결과를 사용자가 우선적으로 고려하는 속성별로 상품의 종류가 제공된다. 따라서 기존 시스템에 비해 제안한 시스템을 사용하는 소비자들은 방대한 양의 리뷰 데이터를 일일이 읽을 필요 없이 구매하고자 하는 상품에 대한 정보를 빠르게 제공받을 수 있어 사용자의 의사결정에 도움이 될 것으로 예상된다.

참고문헌

- [1] S. B. Son, J. H. Chun, "Product Feature Extraction and Rating Distribution Using User Reviews" Society for e-Business Studies, vol.22, no.1, pp.65-87 Feb. 2017. (in Korean)
- [2] J. E. C, H. J. R, D. B. Yu, N. R. Kim, Y. H. Kim, "System Design for Analysis and Evaluation of E-commerce Products Using Review Sentiment Word Analysis" KIISE : Transactions on Computing Practices. vol.22 no.5, pp.209-217. May. 2016. (in Korean)