

E-commerce 환경에서 실시간 사용자 구매 패턴 분석을 통한 사용자 상품 추천 시스템 연구

김범중¹, 허지혜, 이협진, 김영운
¹한국폴리텍대학교 서울강서캠퍼스 빅데이터과
 qjawnd7939@gmail.com, floramiss@naver.com,
 luckkim@kopo.ac.kr, hglee67@kopo.ac.kr

A Study on the Real-time user purchase pattern analysis User Product Recommendation System in E-Commerce Environment

Beom Jung Kim¹, Ji Hye Huh, Hyeopgeon Lee, Young Woon Kim
¹Department of Big Data, Seoul Gangseo Campus of Korea Polytechnics College.

요 약

IT 기술의 발달로 E-Commerce 분야는 실시간으로 발생하는 데이터양이 증가하고 있으며, 발생한 데이터는 개인화 맞춤 서비스에 많이 활용되고 있다. 그러나 신생 E-commerce 기업은 신규 상품 및 기존 상품에 대한 정보와 고객 간의 상호 작용 데이터가 존재하지 않아 콜드 스타트 문제가 발생한다. 이에 본 논문에서는 E-commerce 환경에서 실시간 사용자 구매패턴 분석을 통한 사용자 상품 추천 시스템을 제안한다. 제안하는 시스템은 Kafka와 Spark를 사용해 실시간 스트림을 데이터를 처리한다. 주요 기능은 ALS 알고리즘과, FP-Growth 알고리즘을 적용해 콜드 스타트 문제를 해결하며, 사용자 구매 패턴 분석을 통한 분석 결과에 맞는 상품을 사용자에게 추천한다.

1. 서론

IT 기술의 발달로 E-Commerce 분야에서는 실시간으로 발생하는 데이터양이 증가하고 있으며 이러한 스트리밍 데이터는 개인화 추천 서비스에 많이 활용되고 있다. E-Commerce Data 연구 논문[1]에 따르면, Amazon 매출의 35%, BestBuy의 23.7%가 추천 시스템에 의해 이뤄지며, Netflix는 대여되는 영화의 최대 75%까지 추천을 경유한다.

그러나 신생 E-Commerce 기업은 신규 상품 및 기존 상품에 대한 정보와 고객 간의 상호 작용 데이터가 존재하지 않는다. 이로 인해 기업은 구매 패턴 분석을 진행하기에 충분한 데이터를 확보하지 못하여 콜드 스타트 문제가 발생한다.

이에 본 논문에서는 E-commerce 환경에서 실시간 사용자 구매 패턴 분석을 통한 사용자 상품 추천 시스템을 제안한다.

제안하는 시스템은 분산 시스템인 Kafka를 사용하여 실시간으로 발생하는 구매 데이터를 Spark로 전달하고 ALS 알고리즘과, FP-Growth 알고리즘을 적용해 사용자 구매 패턴 분석을 통해 분석 결과에 맞는 상품을 사용자에게 추천한다.

본 논문의 구성은 다음과 같다. 2장에서는 ALS 알고리즘과 FP-Growth 알고리즘을 알아보고 3장에서는 E-commerce 환경에서 실시간 사용자 구매 패턴 분석을 통한 사용자 상품 추천 시스템을 제안한다. 마지막 4장에서는 제안하는 시스템에 대한 결론을 제시한다.

2. 관련 연구

본 장에서는 ALS 알고리즘과 FP-Growth 알고리즘에 대한 연구를 살펴본다.

2.1 ALS 알고리즘

ALS 알고리즘[2]은 협업 필터링 기반 알고리즘이다. 실제 사용자 평가 점수와 추천된 상품 점수 차이를 최소화하는 방향으로 학습하며, 최적화를 통해 구한 최적의 x , y 벡터는 추후 새로운 사용자의 상품 평가 점수를 예측한다.

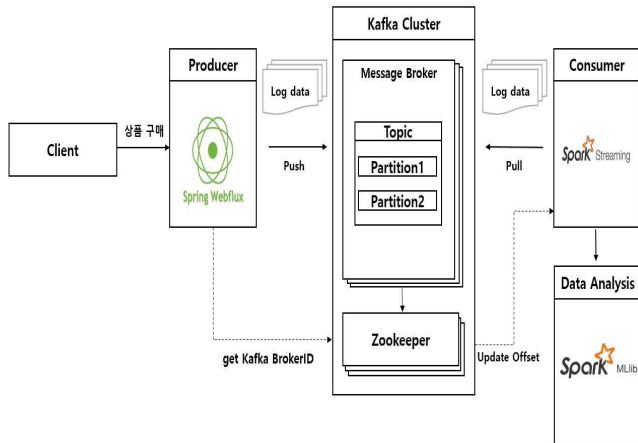
2.2 FP-Growth 알고리즘

FP-Growth 알고리즘[3]은 항목 집합에 대한 모든 조합을 생성하지 않고, 조건 패턴 베이스를 사용하여 재귀적으로 찾아낸다.

FP-Tree는 데이터 집합에서 각 아이템의 출현 빈도를 계산하고 최솟값 이상의 단일 품목 집합 아이템들을 순서에 따라 정렬해 생성한다.

3. 제안하는 실시간 사용자 상품 추천 시스템

제안하는 실시간 사용자 상품 추천 시스템은 비동기 분산 시스템을 이용하여 실시간으로 발생하는 사용자 구매 데이터를 처리하고 ALS 알고리즘과 FP-Growth 알고리즘을 통해 콜드 스타트를 해소해 사용자에게 상품을 추천한다. [그림 1]은 제안하는 실시간 사용자 상품 추천 시스템을 나타낸다.



[그림 1] 실시간 사용자 상품 추천 시스템

Client는 상품을 구매하거나 기업에서 제공하는 상품을 확인한다. Producer는 사용자가 선택한 상품에 대한 데이터를 발생시키고 Kafka Streams를 활용해 Kafka Cluster로 데이터를 전달한다. Kafka는 전달받은 데이터를 직렬화, 파티셔닝, 메시지 압축의 과정을 거치고 Broker에 분산 저장하고 Zookeeper에 의해 관리된다. Consumer는 Spark Streaming을 사용해 Kafka Broker로부터 스트리밍 데이터를 가져와 Spark MLlib으로 데이터를 실시간으로 전달한다.

제안하는 시스템은 콜드 스타트 문제를 해소하기 위해 두 가지의 방법으로 상품을 추천한다.

최초 접속한 사용자는 FP-Growth를 통해 상품을 추천한다. FP-Growth를 활용한 상품 추천은 접속한 사용자의 성별과 나이 정보를 전달받아 사용자 정보를 기준으로 구매 데이터 셋에서 빈발 아이템 집합을 찾아 사용자에게 상품을 추천해 준다.

이미 구매 데이터가 존재하는 사용자는 ALS 알고리즘을 적용하여 상품을 추천한다. ALS 활용한 상품 추천은 사용자의 구매 데이터를 분석하여 비슷한

취향을 가진 다른 사용자들과의 연관성을 파악하여 추천한다. ALS를 활용한 상품 추천은 사용자의 구매 데이터가 부족한 경우 상품 평가 점수를 예측하기 어려워 추천의 정확도가 낮을 수 있다.

따라서 일정 규모의 구매 데이터가 축적되기 전까지는 FP-Growth를 통해 상품을 추천해 주고 일정량의 데이터가 쌓이면 ALS 알고리즘을 적용하여 콜드 스타트 문제를 해결하고 개인화 맞춤 추천 서비스 제공할 수 있다.

제안하는 시스템은 비동기 분산 시스템을 활용해 실시간으로 사용자의 구매 데이터를 전달받아 최초 사용자와 기존 서비스를 이용하고 있는 사용자에게 다른 알고리즘을 적용하여 콜드 스타트 문제를 해결하고 구매 데이터를 분석하여 E-Commerce 기반 실시간 상품 추천을 제공하는 시스템을 구축할 수 있다.

4. 결론 및 향후 연구과제

본 논문에서는 E-commerce 환경에서 실시간 사용자 구매 패턴 분석을 통한 사용자 상품 추천 시스템을 제안한다.

제안하는 시스템은 비동기 분산 시스템인 Kafka를 사용하여 실시간으로 발생하는 구매 데이터를 Spark로 전달하고 ALS 알고리즘과, FP-Growth 알고리즘을 적용해 사용자 구매 패턴 분석을 통해 분석 결과에 맞는 상품을 사용자에게 추천한다.

향후, 제안하는 시스템의 검증은 위해 시스템 구현과 시스템에 대한 성능 평가가 필요하다.

참고문헌

[1] Juryon Paik, Methods Comparison: Enhancing Diversity for Personalized Recommendation with Practical E-Commerce Data, Journal of the Korea Society of Computer and Information, 27(9), 2022, pp.59-68.
 [2] Y. Hu, Y. Koren and C. Volinsky, Collaborative Filtering for Implicit Feedback Datasets, 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 263-272.
 [3] Jiawei Han, Jian Pei, and Yiwen Yin, Mining frequent patterns without candidate generation, SIGMOD Rec, 29(2), 2000, pp.1-12.