

협업 필터링 및 하이브리드 필터링을 이용한 동종 브랜드 판매 매장간 취급 SKU 추천 시스템*

조용민

한국과학기술원 경영대학 경영공학부
(moneyjoe@kaist.ac.kr)

남기환

한국과학기술원 경영대학 경영공학부
(namkh@kaist.ac.kr)

최근 인터넷 기반의 웹 및 모바일 기기를 통한 소비 패턴의 다양화와 개성화가 급진전됨에 따라 전통적 유통 채널인 오프라인 매장의 효율적 운영이 더욱 중요해졌다. 매장의 매출과 수익 모두를 제고하기 위해 매장은 소비자에게 가장 매력적인 상품을 적시에 공급·판매 해야 하는데 많은 상품들 중에서 어떤 SKU를 취급하는 것이 판매 확률을 높이고 재고 비용을 낮출 수 있는지에 대한 연구가 부족한 실정이다. 특히, 여러 지역에 걸쳐 다수의 오프라인 매장을 통해 상품을 판매하는 기업의 경우 고객에게 매력적인 적절한 SKU를 추천 받아 취급할 수 있다면 매장의 매출 및 수익을 제고에 도움이 될 것이다.

본 연구에서는 개인화 추천에 이용되어 왔던 협업 필터링과 하이브리드 필터링 등의 추천 시스템 (Recommender System)을 국가별, 지역별로 복수의 판매 매장을 통해 동종 브랜드를 취급하는 유통 기업의 매장 단위 취급 SKU 추천 방식을 제안하였다. 각 매장의 취급 품목별 구매 데이터를 활용하여 각 매장 별 유사성 (Similarity)을 계산하고 각 매장의 SKU별 판매 이력에 따라 협업 필터링을 하여 최종적으로 매장에 개별 SKU를 추천하였다.

또한 매장 프로파일 데이터를 활용하여 주변수 분석 (PCA : Principal Component Analysis) 및 군집 분석 (Clustering)을 통하여 매장을 4개의 군집으로 분류한 뒤 각 군집 내에서 협업 필터링을 적용한 하이브리드 필터링 방식으로 추천 시스템을 구현하고 실제 판매 데이터를 바탕으로 두 방식의 성능을 측정하였다.

현존하는 대부분의 추천 시스템은 사용자에게 영화, 음악 등의 아이템을 추천하는 방식으로 연구가 진행되어 왔고 실제로 산업계에서의 적용 또한 개인화 추천 시스템이 주류를 이루고 있다. 그 동안 개인화 서비스 영역에서 주로 다루어져 왔던 이러한 추천 시스템을 동종 브랜드를 취급하는 유통 기업의 매장 단위에 적용하여 각 매장의 취급 SKU를 추천하는 방식에 대한 연구는 거의 이루어지지 않고 있는 실정이다.

기존 추천 방법론의 추천 적용 대상이 '개인의 영역이었다면 본 연구에서는 국가별, 지역별로 복수의 판매 매장을 통해 개인의 영역을 넘어 매장의 영역으로 확대하여 동종 브랜드를 취급하는 유통 기업의 매장 단위 취급 SKU 추천 방식을 제안하고 있다. 또한 기존의 추천시스템은 온라인에 한정되었다면 이를 오프라인으로 활용 범위를 넓히고, 기존 개인을 기반으로 분석을 하는 것보다 매장영역으로 확대 적용하기에 적합한 알고리즘을 개발하기 위해 데이터마이닝 기법을 적용하여 추천 방법을 제안한다.

본 연구의 결과가 갖는 의의는 개인화 추천 알고리즘을 동일 브랜드를 취급하는 복수의 판매 매장에 적용하여 의미 있는 결과를 도출하고 실제 기업을 대상으로 시스템으로 구축하여 활용할 수 있는 구체적 방법론을 제시했다는 데에 있다. 개인화 영역을 위주로 이루어졌던 기존의 추천 시스템과 관련한 학계의 연구 영역을 동종 브랜드를 취급하는 기업의 판매 매장으로 확장시킨 첫 시도라는 데에도 의미가 있다.

2014년 03주차 ~ 05주차 소 매장 판매 수량 실적 Top 100개 SKU로 추천의 대상을 한정하여 협업 필터링과 하이브리드 필터링 방식으로 52개 매장 별로 취급 SKU를 추천하고, 추천 받은 SKU에 대한 2014년 06주차 매장별 판매 실적을 집계하여 두 추천 방식의 성과를 비교하였다. 두 추천 방식을 비교한 이유는 본 연구의 추천

* 본 연구는 한국과학기술원의 미래선도형 특성화연구사업 중, IoT기반 초연결 사회를 위한 미래 비전에 관한 연구의 일환으로 이루어졌음.

방법이 기존 추천 방식 보다 높은 성과를 입증하기 위해 단순히 오프라인에 협업필터링을 적용한 것을 기준 모델로 정의하였다. 이 기준 모델에 오프라인 매장 관점의 특성을 잘 반영한 본 연구 모델인 하이브리드 필터링 방법과 비교 함으로써 성과를 입증한다.

연구에서 제안한 방식은 기존 추천 방식보다 높은 성과를 나타냈으며, 이는 국내 대기업 의류업체의 실제 판매데이터를 활용하여 입증하였다. 본 연구는 개인 수준의 추천시스템을 그룹수준으로 확장하여 효율적으로 접근하는 방법을 이론적인 프레임 워크를 만들었을 뿐 아니라 실제 데이터를 기반으로 분석하여 봄으로써 실제 기업들이 적용해 볼 수 있다는 점에서 연구의 가치가 크다.

주제어 : Store SKU Recommender System, Collaborative filtering, Hybrid filtering, Retail marketing, Apparel industry

논문접수일 : 2017년 7월 21일 논문수정일 : 2017년 10월 11일 게재확정일 : 2017년 11월 20일
 원고유형 : 일반논문 교신저자 : 남기환

1. 연구의 배경 및 목표

최근 인터넷 기반의 웹 및 모바일 기기를 통한 소비 패턴의 다양화와 개성화가 급진전됨에 따라 전통적 유통채널인 오프라인 매장의 효율적 운영이 더욱 중요해졌다. 오프라인 매장의 매출 극대화를 위한 경쟁이 치열해지면서 다양한 판매 전략이 전개되고 있는데, 시각에 소구하는 상품 정책인 비주얼 머천다이징(V.M.D : Visual Merchandising)의 활용을 통한 매장 디스플레이 전략이 일반적이다(이홍림, 2002). 한 매장에서 취급하는 SKU(Stock Keeping Unit)가 다양하여 매장의 공간적 제약으로 인해 모든 SKU를 디스플레이 할 수 없기에 소비자의 구매 욕구를 최대한으로 이끌어낼 수 있는 상품을 진열하는 것이 효율적인데, 이를 위해 소비자의 시각, 행동에 따른 구매 패턴에 관한 VMD 관점에서의 다양한 연구가 진행되어 왔다(이홍림, 2002). 또한, 매장 운영 시 매출 관점에서는 취급하는 브랜드의 모든 제품을 재고로 보유하면서 공간의 제약 없이 모든 제품을 진열하면 노출 빈도를 높여 판매 확률을 높일 수 있겠지만, 비용과 수익 관점에서는

보유 재고를 최소화 하고 한정된 공간 안에서 매출을 극대화하여 비즈니스의 수익률을 제고하기 위한 노력이 필요하다. 매장의 매출과 수익 모두를 제고하기 위해 매장은 소비자에게 가장 매력적인 상품을 적시에 공급-판매 해야 하는데 많은 상품들 중에서 어떤 SKU를 취급하는 것이 판매 확률을 높이고 재고 비용을 낮출 수 있는지에 대한 연구가 부족한 실정이다. 특히, 여러 지역에 걸쳐 다수의 오프라인 매장을 통해 상품을 판매하는 기업의 경우 고객에게 매력적인 적절한 SKU를 추천 받아 취급할 수 있다면 매장의 매출 및 수익률 제고에 도움이 될 것이다.

추천(Recommendation)이란 본래 목표 고객과 유사한 집단의 선호도에 대한 의견을 이용하여 고객이 다양한 선택가능 상황에서 관심을 가지고 있는 내용들을 효과적으로 파악할 수 있도록 도와주는 과정이라고 정의된다(Resnick and Varian, 1997). 개인화 서비스 중에서 추천시스템(Recommender system)은 목표고객에게 좋아할 것 같은 서비스나 제품을 추천해주는 서비스로써 아마존 등 온라인 쇼핑몰에서 많이 사용되고 있다(Lee and Park, 2007). 기업이 정보 과다 환경

에서 소비자를 도와줄 수 있는 방법 중 하나가 소비자가 필요로 하는 정보를 걸러주는 필터 역할을 하는 것이다. 정보 과다 상황에서 소비자 개인은 자신 나름의 휴리스틱을 통해서 정보 처리 방법을 찾는다. 그렇지만 정보가 빠르게 늘어나면 소비자 개인이 이를 처리하는 데에는 한계가 있다. 만일 각 소비자의 정보 처리를 대신해주는 시스템이 있으면 소비자에게는 매우 큰 도움이 된다. 추천은 다음과 같은 이유로 중요하다 할 수 있다. 만일 매장에서 잘 팔리지 않는 상품을 보유하고 진열해 놓으면 점주는 그 상품으로 인한 보유 비용뿐 아니라 다른 상품을 취급하여 팔렸을 것을 감안한 기회 비용까지 감수해야 한다. 또한, 최종 고객은 그 상품을 보면서 시간적, 감정적인 비용을 지불하는 낭비를 하게 된다. 이런 경험이 반복되면 해당 브랜드에 대한 고객 충성도가 떨어지게 될 것이고 결국 고객의 이탈로 이어질 수 있다. 따라서, 고객이 좋아하는 상품을 정확히 추천해 주는 것은 복수 매장을 거느린 유통 비즈니스의 성패를 좌우할 수도 있다. 추천 시스템은 다양한 기법을 통해 구현될 수 있는데 최근 전자상거래 분야에서 쓰이는 기법 중 대표적인 것이 협업 필터링(Collaborative Filtering)이다. 협업 필터링을 기반으로 한 추천 시스템은 아이템과 사용자의 관계를 이용하여 사용자의 선호에 맞는 아이템을 추천한다(Bell and Koren, 2007). 협업 필터링은 고객들의 프로파일 정보를 활용하여 목표 고객이 높게 평가할 것으로 예상되는 서비스나 아이템을 추천하는 기법으로 다음과 같은 과정으로 아이템을 추천한다. 먼저, 아이템들에 대한 고객의 평가치를 직간접적으로 수집하여 고객별 프로파일을 생성한다. 생성된 프로파일을 기반으로 목표고객과 유사한 성향을 보이는 고객들로 최근접 이웃을 구성한 후 최근

접 이웃의 평가치를 이용하여 목표 고객이 평가하지 않은 아이템의 평가치를 예측한다. 이렇게 예측된 평가치를 기반으로 목표 고객이 높게 평가할 것이라고 예상되는 서비스나 아이템을 추천한다. 최근의 연구는 협업 필터링에 유전자 알고리즘(Gao and Li, 2007), 퍼지 알고리즘(Shamri and Bharadwaj, 2008), 신경망 분석(Christakou and Stafylopatis, 2007), 베이지안 네트워크(Campos et al., 2010), 클러스터링(Shinde and Kulkarni, 2012) 등을 접목한 하이브리드 필터링 방법에 초점이 맞추어져 있다. 이 중에서도 추천 시스템의 군집분석방법의 적용은 아이템의 추가 삭제, 수정으로 인한 추천시스템의 성능저하를 해결하는 방법(Wen and Zhou, 2012), 사용자 군집과 아이템 군집을 결합하는 방법(Gong, 2010), 사용자들의 비 적극성에 기인하는 희박성 문제를 해결하는 방법(Simon et al., 2013) 등 다양한 방법으로 활발하게 연구되어 지고 있다.

현존하는 대부분의 추천 시스템은 이렇듯 사용자에게 영화, 음악 등의 아이템을 추천하는 방식으로 연구가 진행되어 왔고 실제로 산업계에서의 적용 또한 개인화 추천 시스템이 주류를 이루고 있다. 그 동안 개인화 서비스 영역에서 주로 다루어져 왔던 이러한 추천 시스템을 동종 브랜드를 취급하는 유통 기업의 매장 단위에 적용하여 각 매장의 취급 SKU를 추천하는 방식에 대한 연구는 거의 이루어지지 않고 있는 실정이다. 추천의 유용성과 필요성은 많은 기업들이 인지하고 있으나 정교한 추천이 이루어지지 않을 경우, 오히려 고객의 불만 사유가 되어 영업 기회를 잃을 수 있기 때문에 보다 정교하게 추천시스템을 개발하여야 한다(Kim and Kim, 2014).

본 연구에서는 국가별, 지역별로 복수의 판매 매장을 통해 동종 브랜드를 취급하는 유통 기업

의 매장 단위 취급 SKU 추천 방식의 제안 및 구현을 목표로 하고 있다. 이에 따라, 의류 및 패션 잡화를 제조 유통하는 국내 기업 ‘E社’의 ‘K’라는 브랜드 제품을 판매하는 대한민국 내 52개 매장을 대상으로 추천 시스템을 개발하였다. 본 연구를 통하여 그 동안 개인화 추천 영역에 국한되어 왔던 추천 시스템(Recommender System) 연구 영역이 복수의 판매 매장을 보유한 기업 대상으로까지 확장되어 기업들에 실질적인 도움이 되는 활발한 연구가 이루어지기는 계기가 되었으면 한다.

2. 관련 연구

2.1 추천 시스템

추천 시스템이란 정보를 여과하는 기술의 일종으로, 특정 사용자가 관심을 가질만한 정보(영화, 음악, 책, 뉴스, 웹 페이지 등)를 추천하는 것이다(Bobadilla et al., 1961). 추천 시스템 연구

의 주된 관심사는 고객과 아이템에 대한 이용 가능한 정보를 분석하여 고객들이 관심을 가질 아이템이 무엇인지를 파악하는 것이다. 추천 시스템은 보통 아이템의 속성들을 기반으로 목표 고객이 관심을 가진 아이템과 비슷한 속성의 아이템을 추천하거나, 고객들이 아이템들을 경험하고 부여한 평가 결과를 기반으로 목표 고객이 평가하지 않은 아이템들 중에서 높게 평가할 것이라 예상되는 아이템을 추천한다. 평가 결과는 고객들에게 경험한 아이템에 대한 평가를 묻는 방법을 통해 얻어진다(Lekakos and Giaglis, 1972). 이러한 추천 시스템은 보통 개인화 영역에서의 추천 서비스에 대한 연구가 주를 이루고 있는데 이는 고객 개개인에 대한 충성도(Loyalty)를 높여줄 뿐 아니라 타겟 마케팅(Target marketing)을 가능하게 해준다는 점에서 의미가 크다.

개인화 추천 서비스에 대한 연구는 1990년대 중반 Hill, Rensnick, Shardanand 등이 학술적으로 발표하기 시작했던 것이 효시가 되었는데, 데이터 마이닝 분야의 다양한 방법을 이용한 추천 방법들이 제안되어 왔으며 기업에서 실제 추천 시

〈Table 1〉 Recommendation Technology Classification

분류	설명	입력 값	추천 과정(Process)
협업 (Collaborative)	유사한 정보를 갖는 이웃들의 선호에 기반한 추천	추천 대상 아이템들에 대한 사용자들의 평가	사용자의 선호도와 관심 표현의 패턴을 분석, 비슷한 패턴을 가진 고객을 선별하여 아이템 또는 서비스 추천
내용 기반 (Content-based)	정보 검색에 뿌리를 두며, 유사한 속성의 아이템을 추천	추천 대상 아이템들에 대한 사용자들의 평가	사용자의 아이템에 대한 평가 데이터를 통하여 규칙을 형성, 그 규칙에 따라 추천
인구 통계 기반 (Demographic)	개인 특성 기반 이용자 분류가 목적, 인구 통계 분류 기반 추천	사용자에 대한 인구 통계학적 정보	인구 통계학적으로 유사한 사용자를 구분, 아이템에 대한 그들의 평가로부터 추정하여 추천
효용 기반 (Utility-based)	개별 아이템에 대한 이용자 별 효용성에 기반하여 추천	사용자가 선호를 나타내는 아이템에 대한 효용 함수	아이템에 대한 효용 함수를 적용하고 순위를 결정하여 추천
지식 기반 (Knowledge-based)	이용자의 필요와 선호에 대한 추론에 기반하여 추천	사용자의 필요 또는 흥미에 대한 정보	아이템과 사용자의 필요 또는 선호 사이의 부합 여부를 추론하여 추천

시스템 구현에 이러한 방법론들이 적용되어 왔다. 대표적인 추천 방법은 추천 과정(Process)에 따라 <Table 1>과 같이 다섯 가지로 분류할 수 있으며 (Burke, 2002) 이중 가장 대중적으로 쓰이는 방법은 협업 필터링 방식이다.

2.2 추천 방법론 비교

모든 추천 방법들은 각각 저마다의 장점과 단점을 갖고 있는데, <Table 1> 에서 언급한 추천 방법론에 대한 장단점은 아래 <Table 2>와 같다 (Burke, 2002).

추천시스템에 관련된 기존연구들은 상당수가 사용자의 연관성을 기반으로 상품을 추천하는 협업필터링(Collaborative Filtering)을 기반으로 진행되었다(Billsus and Pazzani, 1998). 협업 필터링(Collaborative filtering, CF) 추천시스템은 추천 대상이 되는 고객과 유사한 패턴을 가지고 있는 고객들을 확인하여, 이들이 과거 선호했던 상품들 중 대상 고객이 아직 경험하지 않은 상품들

추천하는 시스템이다(Jeon and Ahn, 2015). 상품과 상품 사이의 유사성을 기반으로 추천하는 내용 기반 추천 알고리즘 보다, 협업 필터링이 상대적으로 더 높은 추천 정확도를 나타내는 것으로 알려져 있다(Kim and Ahn, 2009; Kim and Ahn, 2011). 협업 필터링을 사용한 초기 시스템은 1990년 초 제록스 팔로 알토 리서치 센터(Xerox PARC) 에서 개발한 태피스트리(Tapestry) 프로젝트인데, 이때 협업 필터링이라는 용어가 처음 사용되었다. 이 시스템은 전자우편 등에 사용자의 평가 또는 주석 등을 붙임으로써 사용자가 문서를 검색할 때 이를 참고하도록 하였다 (Goldberg et al., 1992).

협업 필터링은 각 사용자 별로 그 사용자와 가장 선호 정보가 비슷한 사용자들을 이웃(neighbor)으로 찾은 다음, 이웃들이 선호하는 상품을 그 사용자에게 추천하는 방식이다. 비슷한 사용자인 이웃을 찾을 때는 Pearson correlation, cosine, constraint Pearson correlation and mean squared differences등이 활용된다(Bobadilla et al.,

<Table 2> Comparison of Advantages and Disadvantages of Recommendation Method

분류	장점	단점
협업 필터링 (Collaborative Filtering)	다양한 형태의 정보에 적용 가능함 도메인 지식이 필요치 않음 시간이 지날수록 추천 결과에 대한 품질이 향상됨	충분한 트랜잭션 데이터 필요함 (희박성) 사용자 및 아이템 규모가 클수록 많은 연산량이 요구됨 (확장성)
내용 기반 (Content-based)	추천 대상의 속성 및 이용자의 성향 반영이 가능함 시간이 지날수록 추천 결과에 대한 품질이 향상됨	신규 이용자 문제 계산의 복잡성 충분한 트랜잭션 데이터 필요함
인구 통계 기반 (Demographic)	다양한 형태의 정보에 적용 가능함 구축이 용이함	충분한 트랜잭션 데이터 필요함 인구 통계학적 정보 반드시 필요
효용 기반 (Utility-based)	선호도 변화에 민감하게 반응함 데이터 희박성 문제로부터 자유로움 아이템 자체 외의 정보 포함 가능(제품 보증 기간, 운송 스케줄 등)	어떤 항목에 얼마 만큼의 가중치로 효용을 느끼는지에 대한 사용자 별 정보가 필수로 요구됨
지식 기반 (Knowledge-based)	가벼운 탐색(Casual exploration) 시 유용함	지식 공학 필요

2013). 이러한 측정 방식을 활용하여 각 사용자와 성향이 유사한 사용자를 찾아낸 다음, 각 사용자가 접하지 못한 아이템을 추천하게 된다. 각각의 유사성 측정 방식에 따라 이웃(neighbor)이 다르게 형성되고 이에 따라 선호도 점수가 상이하게 나타나므로, 이웃 형성 과정이 협업 필터링에 매우 중요한 요소라 할 수 있다. 2000년대 중반까지 이루어진 협업 필터링에 관한 연구는 주로 사용자가 접한 뉴스 또는 음악의 선호도를 사용자가 직접 평가한 데이터(explicit data) 또는 상품의 구매여부(binary data)와 같은 데이터를 바탕으로 협업 필터링에 대한 연구가 이루어졌다. 그러나 2000년대 중 후반부터는 전자상거래 사이트에서 사용자의 행동 및 탐색 패턴을 고려한 협업 필터링 기법이 개발되었다. 즉, 사용자가 선호도를 직접 입력하지 않더라도, 사용자의 이용 패턴을 통해 추천된 선호도를 활용함으로써 자동화된 협업 필터링 기법을 개발한 것이다(김용수, 2012). 그러나, 협업 필터링 추천 방법은 많은 계산량을 요구하지 않는 중소규모의 환경에서 실시간 추천을 위해서 고안되었기 때문에 개인화된 추천을 위한 계산은 복잡하지 않고 속도도 빠르나 대용량 환경에서는 사용자 간의 유사도 및 아이템 선정에 위해 많은 계산량이 요구된다. 이를 보완하기 위해 차원감소 기법을 협업적 필터링에 적용한 연구도 수행되었다(Sarwar et al., 2000). 차원감소기법은 사용자가 평가하지 않은 (또는 클릭하지 않은) 아이템의 초기 선호도 값을 평균값 등으로 모두 채운 다음 추정 작업을 진행하므로 모든 아이템에 대해 선호도를 예측할 수 있다는 장점이 있다. 따라서, 데이터의 밀도가 높지 않은 경우에는 차원감소 기법이 전통적인 협업 필터링 기법의 대안이 될 수 있다.

3. 매장 취급 SKU 추천 방법

3.1 추천 방법의 개요

앞에서 살펴본 기존 추천 방법론의 추천 적용 대상이 ‘개인의 영역이었다면 본 연구에서는 국가별, 지역별로 복수의 판매 매장을 통해 개인의 영역을 넘어 매장의 영역으로 확대하여 동종 브랜드를 취급하는 유통 기업의 매장 단위 취급 SKU 추천 방식을 제안하고 있다. 또한 기존의 추천시스템은 온라인에 한정되었다면 이를 오프라인으로 활용 범위를 넓히고, 기존 개인을 기반으로 분석을 하는 것보다 매장영역으로 확대 적용하기에 적합한 알고리즘을 개발하기 위해 데이터마이닝 기법을 적용하여 추천 방법을 제안한다. 각 매장의 취급 품목별 구매 데이터를 활용하여 각 매장 별 유사성(Similarity)을 계산하고 각 매장의 SKU별 판매 이력에 따라 협업 필터링을 하여 최종적으로 각 매장에 개별적으로 SKU를 추천하는 방식과 매장 프로파일 데이터를 활용하여 주변수 분석(PCA: Principal Component Analysis) 및 군집 분석(Clustering)을 통하여 매장을 N개의 군집으로 분류한 뒤 각 군집 내에서 협업 필터링을 적용한 하이브리드 필터링 방식의 추천 시스템을 제안하고 있다.

각 매장이 최근접 이웃(들)으로부터 추천 받은 SKU 목록 중 일부는 어떤 매장에서는 일회성으로 판매된 상품인 경우도 있고 베스트셀러 상품인 경우도 있는데 매장의 공간은 한정적이라는 제약을 고려하여 모든 SKU를 추천할 수 없는 바, 추천 대상의 우선순위를 정하기 위하여 소매장의 3주간 판매 실적을 합하여 판매 수량의 합이 높은 순으로 100개를 한정하여 추천 대상 목록을 만들었다. 추천 대상의 우선순위 기준은

수량뿐만 아니라 매출의 크기, 판매 확률 등 다양한 기준의 조합이 제시될 수 있으나 본 연구에서는 현업 전문가의 의견을 반영하여 판매 수량만을 이용하여 상위 100개의 추천 대상 SKU를 정하였다.

3.1.1 협업 필터링을 이용한 매장 별 취급 SKU 추천

매장의 최근접 이웃을 구성하고 SKU를 추천하는 방식 다음과 같다.

1) 매장 x 품목 매트릭스 준비

p개의 매장에서 q개의 품목을 판매한 이력을 정리하여 <Table 3>과 같이 p x q의 매장 x 품목 매트릭스를 만든다. 이때 판매 이력은 판매 수량이며, 반품의 경우 마이너스가 될 수 있다. 품목은 동일한 속성의 SKU들의 집합이며 ‘코트’, ‘가방’, ‘바지’ 등의 품목 속에 생산연도, 계절, 색깔 등의 분류 기준에 의해 나뉘어진 복수의 SKU들이 존재하게 된다. <Table 3>에서 QA1은 매장 A에서 품목1에 속하는 SKU들이 팔린 수량의 합을 뜻한다. ∅은 해당 기간 동안 고객이 해당 품목에 해당하는 SKU를 하나도 구매하지 않았거나 판매 수량과 반품 수량의 합이 0임을 뜻한다.

<Table 3> Store Item Matrix in Collaborative Filtering

매장 \ 품목	품목1	품목2	품목3	품목4	품목5
매장A	QA1	QA2	QA3	QA4	QA5
매장B	QB1	QB2	QB3	QB4	QB5
매장C	∅	QC2	∅	QC4	∅

2) 최근접 이웃의 구성

매장 별 품목들의 판매 수량 이력을 이용하여 매장들 간의 유사도(Similarity)를 계산한 후 최근접 이웃을 구성한다. 유사도를 계산하기 위한 측정지수로는 일반적으로 Euclidean distance, Pearson Correlation Coefficient, Cosine 등이 사용된다.

유클리디안 거리 d(A,B)는 두 매장 A, B에서 판매된 각 품목에 해당하는 SKU들의 판매 수량의 합을 이용하여 식 (1)과 같이 계산한다.

$$d(A,B) = \sqrt{(Q_{A1} - Q_{B1})^2 + (Q_{A2} - Q_{B2})^2 + (Q_{A3} - Q_{B3})^2 + \dots + (Q_{Aq} - Q_{Bq})^2} \quad (1)$$

Pearson Correlation Coefficient r(A,B)는 두 매장 A, B에서 판매된 각 품목에 해당하는 SKU들의 판매 수량의 합을 이용하여 식 (2)과 같이 계산한다. [15]

$$r(A,B) = \frac{\sum_{i=1}^q (Q_{Ai} - \overline{Q_A})(Q_{Bi} - \overline{Q_B})}{\sqrt{\sum_{i=1}^q (Q_{Ai} - \overline{Q_A})^2 \sum_{i=1}^q (Q_{Bi} - \overline{Q_B})^2}} \quad (2)$$

여기서 Q_{Ai} 와 Q_{Bi} 는 A, B 두 매장에서 판매된 각 품목에 해당하는 SKU들의 판매 수량을 뜻한다. 그리고 $\overline{Q_A}$ 와 $\overline{Q_B}$ 는 매장 A, B에서 판매된 모든 SKU들의 평균 판매 수량을 뜻한다.

Cosine $\cos(\vec{A}, \vec{B})$ 는 A, B 두 매장에서 판매된 품목에 해당하는 SKU들의 판매 수량을 q차원 공간에 벡터화 한 후 두 벡터 사이 각의 Cosine 값을 식(3)과 같이 계산한다.

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \times \|\vec{B}\|} \quad (3)$$

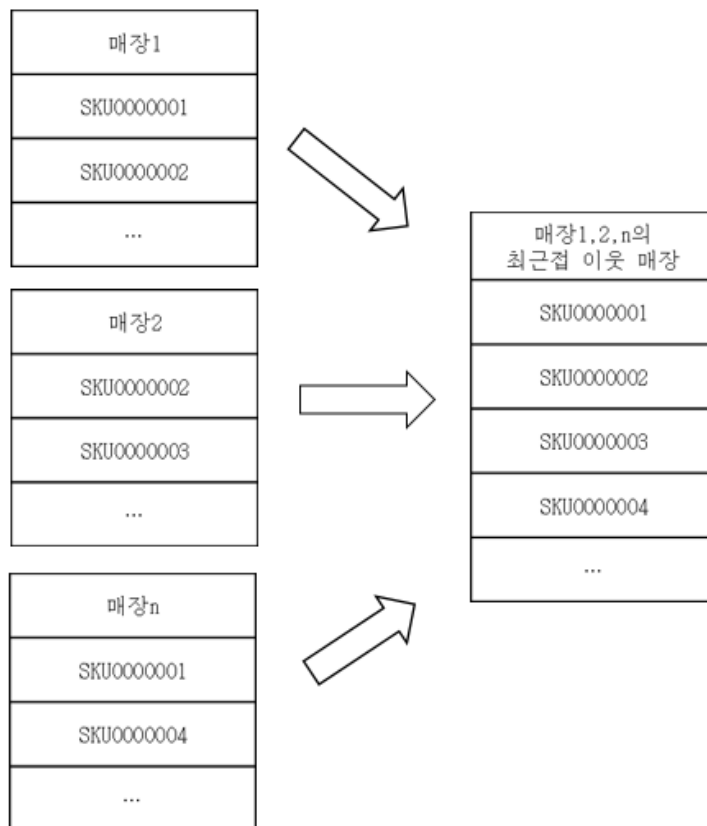
Herlocker et. al은 유사도 계산 시, Pearson Correlation Coefficient를 사용하는 것이 Cosine을 사용하는 것보다 높은 추천 성과를 보인다고 (Herlocker et al., 1999) 발표하는 등 기존에 유사도 계산 방식에 관한 연구가 있었다. 본 연구에서는 최근접 이웃 선정에 대한 직관적 이해가 쉬워 시스템 구현 시 현업에서 채택한 유클리디안 거리 방식으로 최근접 이웃을 선정하였다.

3) 최근접 이웃 매장(들)에 당 매장에서 판매한 SKU 추천

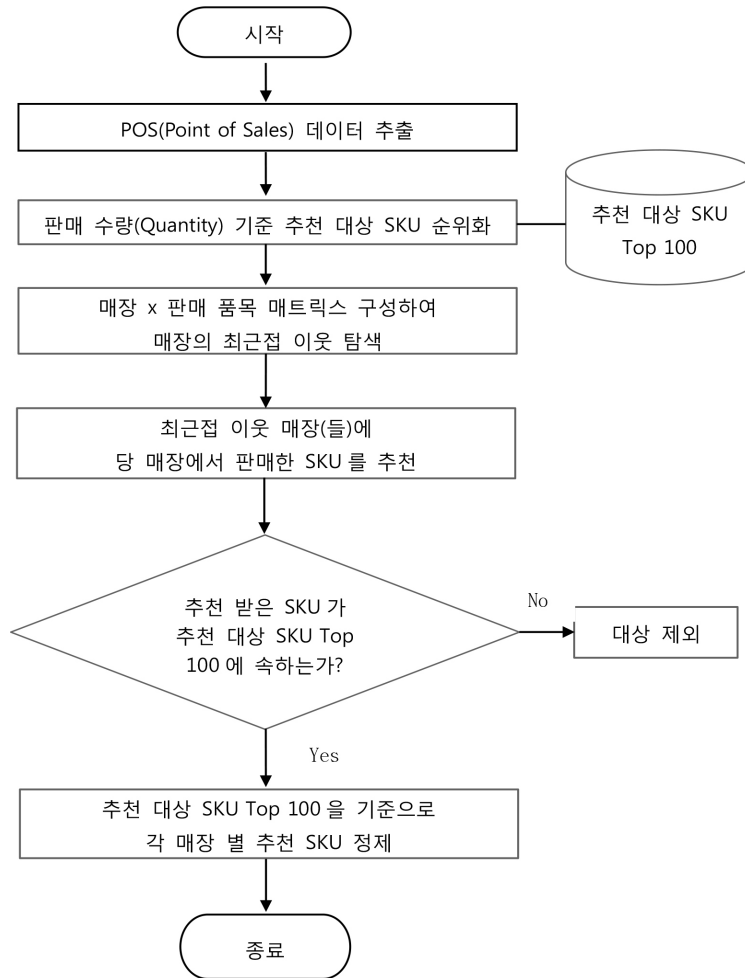
최근접 이웃을 선정 한 후에는 해당 매장에서

팔린 SKU를 선정된 최근접 이웃 매장에 추천하며 반대로도 수행하여 최근접 이웃 매장끼리 서로의 판매 SKU를 추천한다. 최근접 매장(들)로부터 추천 받은 SKU들을 중복을 제거하여 각 매장의 추천 받은 SKU를 정리한다. 간략히 도식화하면 <Figure 1>과 같다.

마지막으로, 전체 매장에서 가장 많이 판매되었던 Top 100 SKU로 추천 SKU의 수를 한정하여 최종적으로 매장 별 취급 추천 SKU를 선정한다. 위에서 설명한 협업 필터링 기법으로 매장 별 취급 SKU를 추천하는 과정은 <Figure 2>와 같이 도식화 할 수 있다.



<Figure 1> SKU Recommendation to the Nearest Neighbor Store



(Figure 2) Recommended SKU Selection Flowchart through Collaboration Filtering Recommendation Method

3.1.2 매장 군집화를 통한 하이브리드 방식의 매장 별 취급 SKU 추천

하이브리드 방식의 취급 SKU 추천 방식은 기본적으로 협업 필터링과 동일한 과정을 따르는데, 한가지 차이점은 매장의 규모, 매출, 고객 수, 고객 단가 등의 정보를 바탕으로 군집화(Clustering) 한 후 군집 내의 매장들 끼리 최근접

이웃 매장을 탐색하여 취급 SKU를 추천한다는 점이다. 매장 프로파일 기반 군집화 과정은 다음과 같다.

- 1) 주성분 분석(Principal Component Analysis)
주성분 분석법은 서로 연관되어 있는 데이터들을 표현하는데 매우 효율적인 통계적 방법으

로 알려져 있다. 주성분 분석법의 기본적인 아이디어는 다차원의 데이터 셋을 원래의 데이터의 다양함을 최대한 유지하면서 저 차원으로 줄이는 것이다. 이 분석은 동일한 척도로 측정되고 서로 상관관계가 높은 항목들로 구성된 하위 집합을 가질 경우에 특히 유용하다. 이 경우에 주성분 분석은 전체 원시데이터 집합에서 설명력을 가지고 있는 원시 변수들을 가중치를 이용하여 선형 조합한 소수의 변수들을 생성시키는데 이렇게 차원 축소된 변수들을 가지고 원시 데이터를 표현하여 직관적으로 데이터의 분포를 파악할 수 있다.

주성분 분석의 절차를 공식화하면 다음과 같다. p 개의 원시변수들을 X_1, X_2, \dots, X_p 로 표기한다고 하면, 주성분 분석을 통해 새로운 변수의 집합으로서 원시변수로부터 이 변수의 평균값을 차감한 값들의 가중평균값인 Z_1, Z_2, \dots, Z_p 를 찾는다.

$$Z_i = a_{i,1}(X_1 - \bar{X}_1) + a_{i,2}(X_2 - \bar{X}_2) + \dots + a_{i,p}(X_p - \bar{X}_p) \quad i = 1, \dots, p \quad (4)$$

여기서, Z 의 각 쌍은 상관계수가 0이다. Z 의 분산의 크기에 따라서, 즉 Z_1 은 가장 큰 분산을 가지고, Z_p 는 가장 작은 분산을 가지는 순서로 Z 값의 순위를 매긴 후 가중치 $a_{i,j}$ 를 계산한 다음, 주성분 점수를 계산한다.

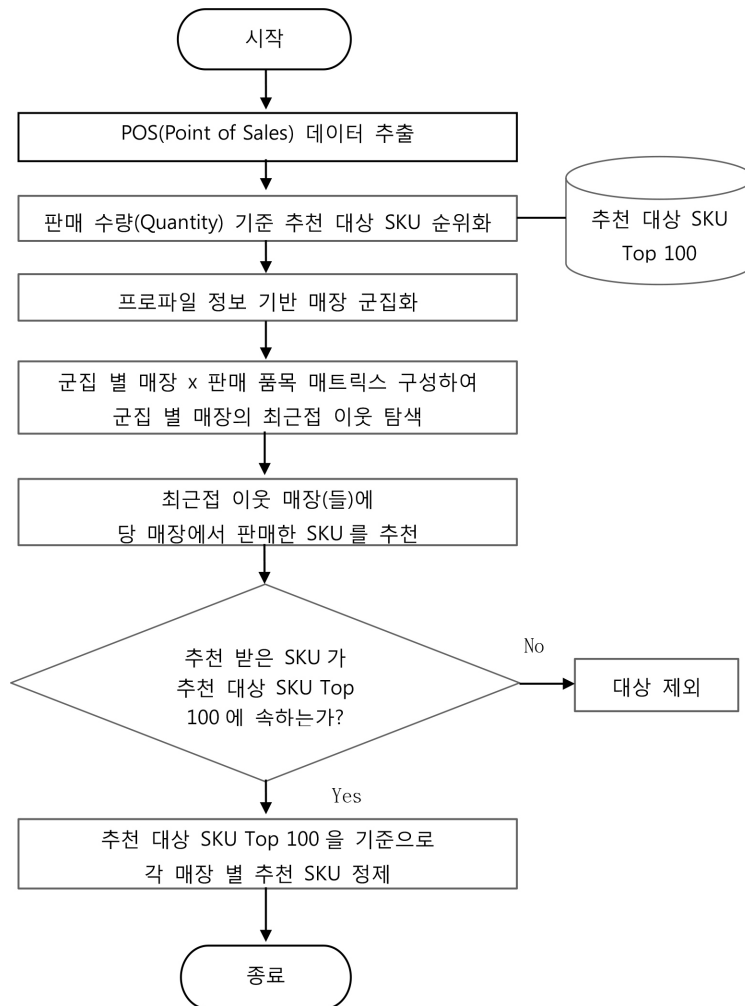
2) K-Means 군집화(Clustering)

K-Means 군집화 알고리즘은 군집화를 위한 자율학습(Unsupervised Learning) 알고리즘 중 하나이다. 사전에 정해진 어떤 수의 군집을 통해서 주어진 데이터 집합을 분류하는 간단하고 쉬운

방법이다. K-Means 군집화 알고리즘은 사례들을 K 개의 초기 군집으로 나누는 것으로 시작한다. 그 다음 단계는 군집의 중심에서 각각의 관찰치까지의 거리 합을 최소화하기 위해 군집을 수정한다. 군집의 수정은 각 레코드를 이전 군집의 중심점에서 가장 가까운 거리의 다른 중심점에 속하는 군집에 재할당하는 것이다. 군집의 수정을 통해 이전보다 더 작은 거리 합을 제시하는 새로운 군집이 만들어진다. 새로운 군집들의 평균값들이 다시 계산되고, 그 다음에는 거리의 감소가 매우 작아질 때까지 이러한 절차가 반복된다. 데이터 이외에 군집의 수 K 를 입력 값으로 받으며, 이 때 K 를 시드 포인트(seed point)라고 한다. 시드 포인트는 임의로 선택되며 바람직한 군집 구조에 관한 해당 영역의 지식들이 시드 포인트를 선택하는데 사용될 수 있다. K-Means 군집화 알고리즘 수행 과정은 다음과 같다.

- (a) 임의로 K 개의 데이터를 선택하여 최초 중심을 정한다. (사용자가 K 를 선택함)
- (b) 매 단계에서, 각 레코드는 ‘가장 가까운’ 중심을 가진 군집에 할당된다.
- (c) 레코드가 빠지거나 추가되면 군집의 중심을 다시 계산하고, 단계 (b)를 반복한다.
- (d) 새롭게 구한 군집의 중심을 바탕으로 각각의 레코드에 대하여 가장 가까이 있는 중심을 찾는다. 군집들 사이에 더 이상 레코드를 이동하는 것이 군집의 분산을 증가시킨다면, 군집화를 멈춘다.

군집화를 수행한 후 군집 별로 이루어지는 최 근접 이웃 매장 선정 및 취급 SKU 추천 과정은 3.1.1에서 기술한 협업 필터링 과정과 동일하다. 이 모든 과정을 포함한 매장 군집화를 통한 하이브리드 방식의 매장 별 취급 SKU 추천 과정은 <Figure 3>과 같이 도식화 할 수 있다.



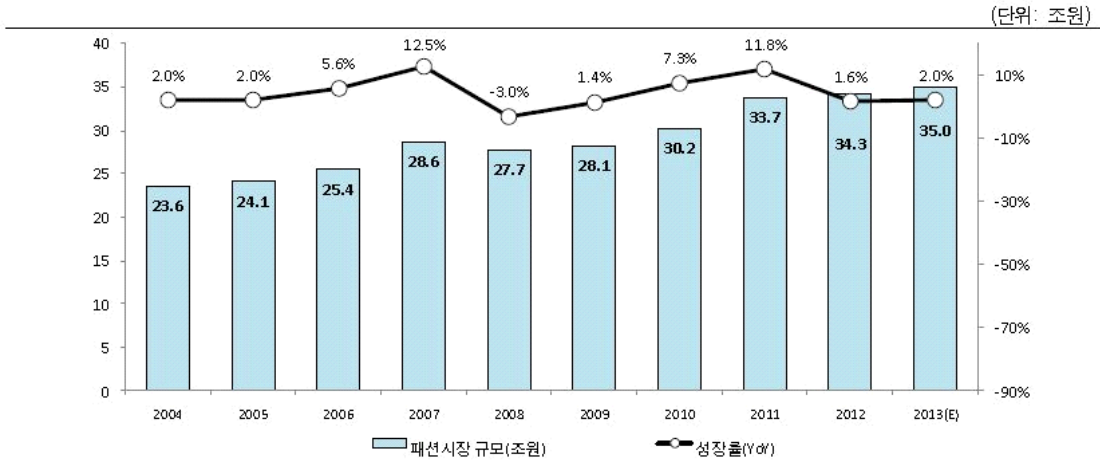
〈Figure 3〉 Recommended SKU selection flowchart through hybrid method

3.2 매장 별 취급 SKU 추천 방법의 구현

3.2.1 국내 패션 시장과 'E社'의 여성 의류 브랜드 'K'의 개요

국내 패션 시장은 보세 매장, 동/남대문 시장, 온라인 개별 쇼핑몰 등에서 취급하는 브랜드 인지도가 없는 제품 등이 거래되는 시장을 일컫는

‘그레이 마켓(Gray Market)’을 포함하여 2013년 기준 2012년 대비 약 2.0% 성장한 약 35조원 (출처: 삼성패션연구소)으로 추정된다. 이는 아래 Figure 4에서 확인 할 수 있다. 패션 사업은 과거 가두 대리점 위주의 고급 기성복 업체가 시장을 지배하였으나, 최근 고급 백화점 및 대형 할인점 등이 성장함에 따라 유통 산업의 패션시장 영향



(출처: 삼성패션연구소)

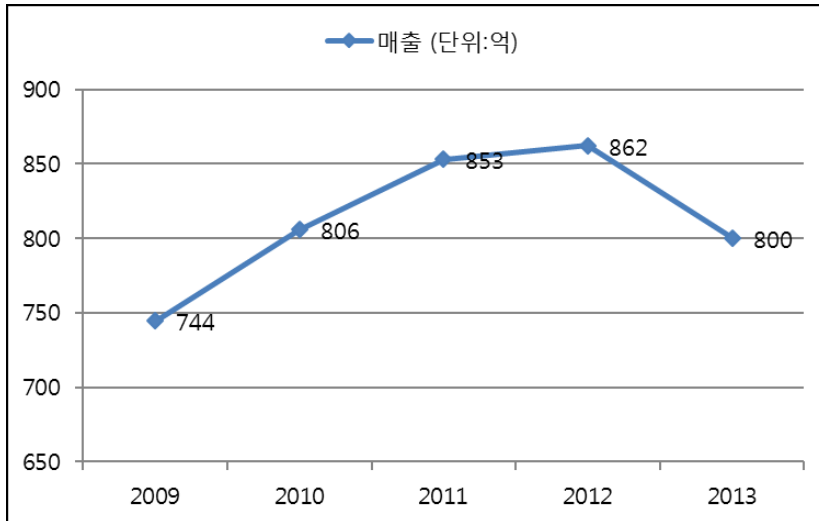
〈Figure 4〉 Domestic Fashion Market Size and Market Growth Rate

력이 커지고 있다. 이에 대응하여 패션 업체는 핵심 가두점 확보, 브랜드파워 강화 등 마케팅을 강화하고 있으며, 단순하며 안정적인 의류 사업에서부터 문화, 라이프스타일, 엔터테인먼트를 기반으로 한 패션 브랜드 사업으로 급속히 변화하고 있다.

‘E社’는 의류 및 패션 잡화를 제조 유통하는 국내 기업으로, 2013년 12월 한 달에만 1,716억 매출 (출처: 금융감독원 전자공시시스템 DART), 2013년 한 해 동안 약 2조원의 매출을 올린 국내 메이저 기업 중 하나이다. ‘E社’는 자가브랜드 중심의 사업을 근간으로 국내 최고의 위치를 확보하고 있으며 최근 유명브랜드 수입을 통해 사업포트폴리오를 강화하고 있다. 캐주얼 브랜드인 ‘B’는 Sub-Brand의 성공적인 확장을 통해 국내 최고의 캐주얼 브랜드로 자리잡고 있으며, 신사복 브랜드인 ‘G’와 ‘R’ 역시 국내 신사복 브랜드 중 최고의 브랜드가치를 인정받고 있다. 최근에는 시장 성장성이 큰 아웃도어와 SPA(Specialty

retailer of Private label Apparel) 사업에도 진출하였으며, 대표 브랜드를 기반으로 중국시장에도 진출하여 사업을 확대하고 있다. ‘E社’는 복종(服種) 별로 남성복, 여성복, 캐주얼, 스포츠웨어, SPA 및 패션 잡화 등 다양한 포트폴리오를 운영하고 있다.

‘E社’의 브랜드 ‘K’는 20대 후반부터 40대 이후까지 고급스럽고 개성적인 스타일을 추구하는 여성을 타깃으로 2013년에 론칭 10주년을 맞이한 여성 브랜드이다. ‘E社’의 다른 브랜드에 비해 비교적 적은 종류의 SKU, 높은 단가 정책을 고수하면서도 수년간 지속적인 성장을 하며 여성복 커리어 Masstige (대중형 명품) 시장 세그먼트에서 단일 브랜드로 전체 시장 규모의 약10%를 차지하였으나 2013년에는 전년 대비8%에 가까운 역 성장을 하였다. ‘K’ 브랜드의 최근 5년간의 매출 실적은 <Figure 5>와 같으며, 국내 패션 시장 분류 Map상의 ‘K’ 브랜드의 위치는 <Table 4>와 같다.



〈Figure 5〉 'K' brand sales trend

〈Table 4〉 Domestic Fashion Market map and 'k' Brand Position

복종 가격 포지션	남성복		여성복				캐주얼			스포츠웨어			SPA	아동복	내의	패션잡화	
	(5,4조)		(5,7조)				(4,5조)			(4,1조)			(1,9조)	(0,5조)	(1,0조)	(6,1조)	
	정장	캐릭터 캐주얼	어덜트 캐주얼	영 캐주얼	캐릭터 커리어	시니어	TD	베이직 캐주얼	진/유니 캐주얼	아웃 도어	스포츠	골프				제화	잡화
Premium (2,4조)	(0,37조)	(0,22조)	X	X	(0,39조)	(0,34조)	X	X	(0,14조)	X	X	(0,20조)	X	X	X	(0,26조)	(0,52조)
Masstige (7,5조)	(0,52조)	(0,46조)	(0,53조)	(0,32조)	(0,84조)	(0,38조)	(1,00조)	X	(0,45조)	(2,31조)	(0,23조)	(0,28조)	X	(0,16조)	(0,15조)	(1,28조)	(0,87조)
Mass (19,3조)	(1,42조)	(0,85조)	(1,05조)	(1,42조)	(1,07조)	(0,99조)	(0,05조)	(2,3조)	(0,61조)		(0,56조)	(0,48조)	(1,90조)	(0,36조)	(0,84조)	(2,42조)	(0,72조)
Gray (4,5조)	(4,5조)																

(출처: 한국패션브랜드연감, 패션비즈, 삼성패션연구소, 통계청, Expert Interview, Monitor Analysis)

※ 괄호는 2011년 기준 시장 규모, 'X' 박스는 시장이 존재하지 않음을 의미, 'K' 브랜드는 연두색 박스에 위치

3.2.2 분석 대상 데이터 개요 및 전처리 과정

‘K’ 브랜드 제품을 취급 판매하는 국내 매장 중 매장 특성에 따른 군집화의 기준 일관성을 위해 상설점 및 가두점을 제외하고 전국 52개 백화점 매장을 대상으로 2014년 3째주부터 5째주까지의 (2014.01.13(월) ~ 2014.02.02(일)) 3주간의 SKU단위 매출 실적 데이터 11,930건을 추출하였다. 매장의 유사성 (Similarity) 계산이 3주간의 SKU 품목별 판매 실적으로 이루어지므로 매장의 품목별 SKU 팔림세를 실제 그대로 반영하기 위하여 반품 트랜잭션에 대한 데이터는 제거하지 않았다. 다만, ‘E社’의 임직원 구매의 경우 적용 할인율이 일반 고객과 상이하여 구매 패턴이 다를 수 있는 관계로 매장 간 유사성 및 추천되

는 SKU 계산시 결과를 왜곡할 우려가 있어 임직원 할인 구매 데이터는 제거하였다. 각 매장의 판매 이력에 기반한 매장 별 최근접 이웃을 구하기 위하여 POS(Point of Sales) 데이터 11,930건을 24개 품목으로 분류하여 <Table 5>와 같은 표를 구성하였다.

전국 주요 백화점에 입점해 있는 52개 ‘K’ 브랜드 판매 매장을 군집화하기 위해 ‘E社’에서 관리하고 있는 매장 프로파일 데이터는 <Table 6>과 같다. <Table 6>의 모든 데이터는 2013년의 기준으로 2013년 한 해 동안 발생한 실적을 기반으로 구성되었다. 아래 <Table 7>은 <Table 6>의 변수들에 대한 설명을 담고 있다.

<Table 5> Sales Volume by Item of 'K' Brand Store

매장 \ 품목	COAT	자켓	T셔츠	가죽 점퍼	패딩 점퍼	바지	브라우저	스웨터	...
B123	4	5	5	8	5	13	2	6	...
B324	3	7	4	9	-1	17	0	1	...
B353	-1	3	4	-2	6	-1	4	2	...
B369	4	2	5	3	0	-1	1	1	...
B400	12	11	11	1	16	28	11	14	...
B521	5	2	3	0	9	4	2	7	...
B539	6	2	6	1	14	9	0	4	...
B563	8	11	25	9	3	28	3	12	...
B655	11	6	7	1	25	3	1	8	...
B735	6	9	15	3	0	13	3	6	...
...

<Table 6> Store Profile Data

매장코드	매장 사이즈 (3.3m ²)	백화점 마진 (%)	구매 고객 수 (중복제거)	등록 고객 매출	객단가	구매 횟수	매장 전체 매출 (원)	전체 수량	백화점 매출 (억 원)
B123	30.0	31	417	1,585,797,000	1,041,917	1,522	3,137,704,000	2,602	2,540
B324	22.0	35	182	1,047,414,200	821,501	1,275	2,062,904,200	1,964	3,466
B353	25.0	34	213	692,797,200	805,578	860	1,370,246,600	1,188	4,361
B369	28.0	34	217	715,260,600	908,844	787	1,447,488,400	1,302	2,968
B400	29.9	36	725	2,001,323,000	1,050,563	1,905	4,144,571,800	3,567	13,614
B521	27.0	34	204	649,218,200	658,436	986	1,285,314,600	1,158	2,951
B539	34.0	36	477	1,528,406,200	1,054,073	1,450	3,022,351,800	2,440	11,367
B563	31.0	34	387	2,067,478,800	1,010,992	2,045	4,113,460,400	3,580	3,486
B655	26.0	32	589	2,166,334,600	986,042	2,197	4,343,189,400	3,741	6,069
B735	21.0	34	246	1,168,205,200	869,848	1,343	2,308,304,400	2,210	6,204
B768	22.0	34	432	962,630,200	822,058	1,171	1,947,604,400	1,549	8,690
B974	29.8	34	293	1,107,615,000	943,454	1,174	2,190,282,000	1,784	5,440
C040	36.0	35	465	1,181,465,400	816,493	1,447	2,357,018,800	2,041	7,019
C058	29.0	36	656	1,598,952,400	976,758	1,637	3,309,057,200	2,882	7,211
C090	26.2	36	642	1,485,360,600	1,000,243	1,485	2,989,633,200	2,547	7,201
C105	37.0	36	187	675,463,400	828,789	815	1,337,108,800	1,235	2,640
...

<Table 7> 'K' Brand Store Profile Variable Details

변수	의미	단위
매장 코드	매장 ID	-
매장 면적	매장 사이즈	평(3.3m ²)
백화점 마진	백화점 수수료	%
구매 고객 수	1회 이상 구매한 등록 고객 수	명
등록 고객 매출	1회 이상 구매한 등록 고객의 구매액	원
객단가	고객 매출 / 고객 구매 횟수	명
고객 구매 횟수	1회 이상 구매한 등록 고객의 총 구매 횟수	회
매장 전체 매출	매장의 전체 매출	원
전체 수량	매장의 전체 판매량	매(개)
백화점 매출	해당 매장이 속한 백화점 전체 매출	억원

3.2.3 유클리디안 거리 측정 방식 최근접 이웃 매장 선정

<Table 5>의 데이터를 기반으로 하여 유클리디안 거리 측정 기법으로 매장 별 거리 즉, 유사

도(Similarity)를 구하면 <Table 8>과 같다. 매장 가로 x 세로 매트릭스의 값이 작을수록 두 매장의 유사도가 높다는 의미이며 B324 매장의 최근접 이웃 매장은 B123 매장이 된다.

이런 방식으로 선정된 최근접 이웃 매장들은 SKU들을 서로에게 추천하게 된다. 최근접 이웃 쌍을 이루어 각각의 매장에서 가장 많이 팔린 매장을 정리하면 <Table 9>와 같다.

<Table 8> Estimation Results of Distance by Store(similarity) by Euclidean Distance Measurement

매장	B123	B324	B353	B369	B400	B521	B539	B563	...
B123	0.00	13.93	20.40	19.00	28.41	16.91	13.93	34.15	...
B324	13.93	0.00	25.50	21.70	32.23	23.32	22.72	33.59	...
B353	20.40	25.50	0.00	12.04	39.99	12.33	17.94	48.12	...
B369	19.00	21.70	12.04	0.00	41.13	14.25	19.62	44.64	...
B400	28.41	32.23	39.99	41.13	0.00	32.73	28.97	28.37	...
B521	16.91	23.32	12.33	14.25	32.73	0.00	12.33	43.45	...
B539	13.93	22.72	17.94	19.62	28.97	12.33	0.00	39.22	...
B563	34.15	33.59	48.12	44.64	28.37	43.45	39.22	0.00	...
B655	26.61	34.47	26.34	28.76	31.64	19.90	17.26	44.09	...
B735	16.40	16.58	23.26	20.05	27.71	20.52	21.05	29.27	...
B768	18.19	23.77	14.39	11.22	36.44	12.85	17.06	41.61	...
B974	16.12	20.98	19.54	16.22	30.87	17.20	17.78	32.34	...
C040	21.31	24.54	30.82	31.73	15.13	23.19	19.90	29.56	...
...

<Table 9> Nearest Neighbors Pair by Store According to Euclidean Distance Measurement

최근접 이웃 ID	대상 매장	최근접 매장	최근접 이웃 ID	대상 매장	최근접 매장
1	B123	D042	53	D042	B123
2	B324	B123	54	B123	B324
3	B353	C896	55	C896	B353
4	B369	D425	56	D425	B369
5	B400	C040	57	C040	B400
6	B521	D140	58	D140	B521
7	B539	C376	59	C376	B539
8	B563	D508	60	D508	B563
9	B655	D760	61	D760	B655
10	B735	C440	62	C440	B735
11	B768	B369	63	B369	B768
12	B974	C467	64	C467	B974
13	C040	C058	65	C058	C040
14	C058	D759	66	D759	C058
...
52	D762	B655	104	B655	D762

3.2.4 매장 군집 별 유클리디안 거리 측정 방식 최근접 이웃 매장 선정

3.1.2에서 설명한 하이브리드 방식의 매장 별 취급 SKU 추천을 하기 위해서 먼저 각 매장의 프로파일 데이터를 기반으로 매장을 군집화 한 후 각 군집 내에서 최근접 매장을 선정해야 한다. 매장을 군집화 하는 방식으로 K-Means 군집화 알고리즘을 채택하였다. 보통 K의 개수를 정하기 위하여 임의의 숫자들로 군집화를 여러 번 수행한 후 각 군집을 이루는 관찰치의 수량이 고르게 분포되면서 군집내 관찰 치끼리는 거리가 가깝고 군집들끼리는 거리가 가장 멀게 되는 군

집의 개수를 정한다. 본 연구에서는, 시각화가 용이하고 직관적인 해석이 가능한 2x2 매트릭스 형태로 매장들을 분류해온 ‘E社’ 현업의 업무 방식에 따라 K를 4개로 정하고 시작하였다.

2x2 매트릭스 상에 매장 군집을 시각화하기 위하여 차원을 축소하고자 <Table 6>의 매장 프로파일 데이터로 주변수 분석(Principal Components Analysis)을 수행하였다. 분석 도구는 엑셀 마이너(XLMiner) 버전 4.0을 이용하였다. 변수의 종류와 단위가 다양하여 데이터의 정규화 과정을 포함하는 Correlation matrix 방식으로 주성분 분석을 수행한 결과는 아래와 같다.

<Table 10> [Inputs]

Variables									
Selected variables	평균	백화점 마진	고객 매출	구매 고객수	객단가/회	구매 횟수	전체 매출	전체 수량	백화점 매출
Parameters/Options									
Fixed # of components					9				
Method					Correlation matrix				
Show data scores					Yes				

<Table 11> [Principal Components]

Components									
Variable	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
백화점매출	0.314	-0.270	0.146	0.603	-0.469	0.473	-0.004	0.007	-0.008
구매횟수	0.411	-0.004	0.052	-0.348	0.024	0.171	0.458	0.675	0.108
전체수량	0.418	0.102	-0.039	-0.199	0.106	0.141	-0.841	0.183	-0.033
고객매출	0.419	0.144	-0.039	-0.150	0.104	0.106	0.236	-0.400	-0.736
전체매출	0.420	0.132	-0.044	-0.140	0.085	0.093	0.140	-0.552	0.666
평균	-0.019	0.333	0.934	0.052	0.112	-0.042	-0.020	-0.012	0.004
백화점마진	0.163	-0.646	0.134	0.219	0.691	-0.112	0.012	-0.004	0.002
객단가/회	0.164	0.575	-0.278	0.617	0.332	-0.154	0.080	0.211	0.031
구매고객수	0.389	-0.142	0.065	0.058	-0.388	-0.818	-0.020	0.023	-0.014
Variance	5.4	1.5	0.9	0.6	0.4	0.1	0.0	0.0	0.0
Variance%	60.0	16.8	10.5	7.0	4.1	1.5	0.1	0.1	0.0
Cum%	60.0	76.8	87.3	94.2	98.4	99.8	99.9	100.0	100.0

누적 분산 76.8%로 전체 데이터의 76.8%를 해석할 수 있는 주성분 1.0과 2.0에서 각각 가장 가중치가 높은 변수인 ‘전체 매출’과 ‘객단가/회’를 선택하여 군집을 4개로 설정하고 K-Means 군집화를 수행한 결과는 <Table 12>, <Table 13>과 같다.

매장의 전체 매출과 객단가 변수를 이용하여 4개의 군집으로 K-Means 군집화를 수행한 결과는 <Table 14>과 같다.

<Table 14>에서 구해진 매장 군집별로 3.2.3에서 수행한 방식과 동일한 방식으로 유클리드 거리 측정을 통해 최근접 이웃 매장을 선정한 결과는 <Table 15>과 같다.

<Table 12> [Inputs2]

Variables		
Selected variables	객단가/회	전체매출
Parameters/Options		
# Clusters	4	
Start Option	Fixed Start	
# Iterations	10	
Show data summary	Yes	
Show distance from each cluster	Yes	

<Table 13> [Cluster Centers]

Cluster	객단가/회	전체매출		
Cluster-1	783,551	1,210,630,546		
Cluster-2	1,155,530	1,917,520,708		
Cluster-3	892,922	2,028,762,748		
Cluster-4	1,003,145	3,706,735,987		
Distance between cluster centers	Cluster-1	Cluster-2	Cluster-3	Cluster-4
Cluster-1	0	706890260.8	818132210.1	2496105451
Cluster-2	706890260.8	0	111242349.9	1789215285
Cluster-3	818132210.1	111242349.9	0	1677973242
Cluster-4	2496105451	1789215285	1677973242	0

<Table 14> 'K' Brand Store Clustering Results

매장	매장 전체매출	객단가	Cluster	매장	매장 전체매출	객단가	Cluster
B123	3,137,704,000	1,041,917	4	C467	1,437,738,400	937,008	3
B324	2,062,904,200	821,501	3	C474	1,385,899,000	816,404	1
B353	1,370,246,600	805,578	1	C506	1,180,954,000	874,953	1
B369	1,447,488,400	908,844	3	C632	1,134,185,200	667,325	1
B400	4,144,571,800	1,050,563	4	C734	1,290,767,000	981,646	3
B521	1,285,314,600	658,436	1	C752	1,383,095,200	882,251	3
B539	3,022,351,800	1,054,073	4	C822	2,810,878,200	1,179,691	2
B563	4,113,460,400	1,010,992	4	C869	1,443,623,200	817,548	1
B655	4,343,189,400	986,042	4	C896	1,044,144,600	851,966	1
B735	2,308,304,400	869,848	3	C958	2,363,305,600	1,251,000	2
B768	1,947,604,400	822,058	3	D042	2,810,717,600	916,611	3
...

<Table 15> Nearest Neighbors Pair by Store According to Euclidean Distance Calculation

최근접 이웃 ID	군집	대상 매장	최근접 매장	최근접 이웃 ID	군집	대상 매장	최근접 매장
1	1	B353	C896	53	1	C896	B353
2	1	B521	D140	54	1	D140	B521
3	1	C105	C869	55	1	C869	C105
4	1	C167	C105	56	1	C105	C167
5	1	C188	D140	57	1	D140	C188
6	1	C376	C188	58	1	C188	C376
7	1	C395	B521	59	1	B521	C395
8	1	C474	B521	60	1	B521	C474
9	1	C506	D140	61	1	D140	C506
10	1	C632	D425	62	1	D425	C632
11	1	C869	B353	63	1	B353	C869
12	1	C896	D221	64	1	D221	C896
13	1	D140	C632	65	1	C632	D140
14	1	D221	C896	66	1	C896	D221
15	1	D425	C632	67	1	C632	D425
16	1	D762	B521	68	1	B521	D762
17	2	C360	D062	69	2	D062	C360
18	2	C822	D062	70	2	D062	C822
19	2	C958	D346	71	2	D346	C958
20	2	D062	C822	72	2	C822	D062
21	2	D236	C360	73	2	C360	D236
22	2	D346	C360	74	2	C360	D346
...
52	4	D757	B400	104	4	B400	D757

3.3 최종 추천 SKU 선정 결과

각 매장의 판매 품목 유사성(Similarity)을 기반으로 추천의 주체와 추천 대상 매장을 결정한 후, 각각의 매장이 판매했던 SKU를 서로 추천할 때에, 추천하고자 하는 SKU의 대상을 우선순위화 하기 위하여 판매 실적에 따라 1위부터 100위까지 순위를 정하고 대상을 한정하였다. SKU의 우선순위는 ‘K’ 브랜드를 취급하는 52개 매장에서 2014년 3째주부터 5째주까지 1개 이상 판매된 404개의 SKU중 현업 전문가 의견에 따라 판매 수량을 제1 정렬 기준, 판매 금액을 제2 정렬 기준으로 하여 <Table 16>와 같이 100개의 SKU를 추천 대상으로 정하였다.

3.3.1 협업 필터링을 이용한 매장 별 취급 SKU 선정 결과

<Figure 1> 에서 설명한 최근접 이웃 매장에 SKU를 추천하는 과정을 <Table 9> 에서 짚어지진 최근접 이웃 매장들끼리 수행하고 <Table 16>의 2014년 W03 ~ W05 판매 실적에 따른 Top 100 SKU 기준으로 각 매장 별로 추천 받은 SKU를 정제하면 <Table 17>과 같다.

<Table 17>의 결과를 토대로 매장별로 추천 받은 SKU의 개수를 정리하면 <Table 18>와 같다.

<Table 16> Top 100 SKUs by Sales Results in 2014 W03 ~ W05

Rank	SKU	Qty	Sales
1	KF4A3U122R	107	132,032,000
2	KF4A111224	102	68,194,600
3	KF4A3U1221	69	84,096,000
4	KF4A4H1225	60	21,942,900
5	KF4A271213	59	17,045,600
6	KF4A341215	57	75,555,000
7	KF4A3U1215	56	63,897,000
8	KF3D3UU145	54	74,892,000
9	KF4A4H1223	52	18,975,600
10	KF4A6A1133	51	14,721,200
11	KF4A211215	49	18,944,800
12	KF4A5A12AD	47	43,569,900
13	KF4A271214	47	13,708,000
14	KF4A35121H	46	48,168,000
15	KF4A6A1134	45	12,992,800
...
100	KF3D5NZV26	10	3,811,000

<Table 17> Store Recommend SKU based Collaborative Filtering (1: Recommendation)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	0	0	0	1	0	0	0	0	0	0	...
2	KF4A11224	1	0	0	0	0	1	1	0	0	0	...
3	KF4A3U1221	1	0	0	0	0	1	0	0	0	1	...
4	KF4A4H1225	1	0	0	1	0	0	0	0	0	0	...
5	KF4A271213	1	0	0	0	0	1	0	0	0	0	...
6	KF4A341215	0	0	0	0	0	1	0	0	0	0	...
7	KF4A3U1215	1	0	0	0	0	1	0	0	0	0	...
8	KF3D3UU145	1	0	0	0	0	0	0	0	0	0	...
9	KF4A4H1223	1	0	0	0	0	0	0	0	0	0	...
10	KF4A6A1133	0	0	0	1	0	0	0	0	1	0	...
11	KF4A211215	0	0	0	1	0	0	1	0	0	0	...
12	KF4A5A12AD	0	0	0	0	1	1	0	0	1	1	...
13	KF4A271214	1	0	0	1	0	0	0	0	0	0	...
14	KF4A35121H	0	0	0	0	0	1	0	0	0	0	...
15	KF4A6A1134	0	1	0	0	0	0	0	0	0	1	...
16	KF4A11A222	1	0	0	0	0	0	0	0	0	0	...
17	KF4A841115	0	0	0	0	0	1	0	0	0	0	...
18	KF4A5N1225	0	0	0	0	0	0	1	0	0	0	...
19	KF4A4H121R	0	1	0	0	1	1	0	0	1	0	...
...
100	KF3D5NZV26	1	0	0	1	0	0	0	0	1	0	...

<Table 18> Number of SKUs Recommended for Each Store by Collaboration Filtering Method

매장	추천 SKU의 수 (Top 100개 SKU 中)	매장	추천 SKU의 수 (Top 100개 SKU 中)	매장	추천 SKU의 수 (Top 100개 SKU 中)
B123	40	C251	26	D042	37
B324	20	C284	20	D062	40
B353	3	C360	21	D098	20
B369	36	C376	27	D140	52
B400	35	C395	21	D221	11
B521	38	C440	36	D236	37
B539	18	C444	36	D346	49
B563	0	C461	33	D425	44
B655	20	C467	42	D508	32
B735	18	C474	12	D520	51
B768	8	C506	11	D757	22
B974	14	C632	14	D758	16
C040	43	C734	15	D759	42
C058	41	C752	19	D760	46
C090	18	C822	34	D761	27
C105	23	C869	23	D762	27
C167	15	C896	22	합계	1,412
C188	43	C958	14		

3.3.2 매장 군집화를 통한 하이브리드 방식의
매장 별 취급 SKU 추천 결과

<Figure 1> 에서 설명한 최근접 이웃 매장에
SKU를 추천하는 과정을 <Table 15> 에서 짚지

어진 군집별 최근접 이웃 매장들끼리 수행하고
<Table 16>의 2014년 W03 ~ W05 판매 실적에
따른 Top 100 SKU 기준으로 각 매장 별로 추천
받은 SKU를 정제하면 <Table 19>와 같다.

<Table 19> Store Recommend SKU based Hybrid Method (1: Recommendation)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	0	0	1	0	1	1	1	0	1	0	...
2	KF4A111224	0	0	0	0	0	0	1	0	0	1	...
3	KF4A3U1221	1	0	0	0	1	1	0	1	0	0	...
4	KF4A4H1225	0	0	0	0	0	1	0	0	0	1	...
5	KF4A271213	0	0	0	1	0	0	1	0	0	1	...
6	KF4A341215	0	0	0	0	1	0	1	0	1	1	...
7	KF4A3U1215	1	0	1	1	0	1	0	0	1	0	...
8	KF3D3UU145	0	0	0	0	0	1	0	0	1	1	...
9	KF4A4H1223	0	0	0	1	1	1	0	0	1	0	...
10	KF4A6A1133	1	0	0	1	0	0	1	0	0	0	...
11	KF4A211215	0	0	0	0	0	1	0	0	0	0	...
12	KF4A5A12AD	0	1	0	0	0	1	0	0	0	0	...
13	KF4A271214	0	1	1	0	0	1	1	0	1	1	...
14	KF4A35121H	0	1	0	1	0	1	1	0	1	0	...
15	KF4A6A1134	1	1	1	1	1	0	1	1	0	0	...
16	KF4A11A222	0	0	1	1	0	0	0	0	0	1	...
17	KF4A841115	1	0	0	0	0	1	0	0	0	1	...
...
100	KF3D5NZV26	0	0	0	0	0	0	0	0	1	1	...

<Table 20> Number of SKUs Recommended by Each Store in a Hybrid Method

매장	추천 SKU의 수 (Top 100개 中)	매장	추천 SKU의 수 (Top 100개 中)	매장	추천 SKU의 수 (Top 100개 中)
B123	20	C251	29	D042	14
B324	17	C284	18	D062	36
B353	21	C360	39	D098	20
B369	45	C376	15	D140	45
B400	30	C395	14	D221	1
B521	44	C440	21	D236	16
B539	44	C444	30	D346	45
B563	16	C461	33	D425	19
B655	20	C467	41	D508	30
B735	32	C474	12	D520	38
B768	8	C506	11	D757	22
B974	14	C632	15	D758	16
C040	42	C734	11	D759	42
C058	42	C752	28	D760	14
C090	39	C822	21	D761	27
C105	31	C869	28	D762	12
C167	8	C896	19	합계	1,303
C188	34	C958	14		

<Table 19>의 결과를 토대로 매장별로 추천 받은 SKU의 개수를 정리하면 <Table 20>과 같다.

위하여 정확률을 평가 척도로 활용하였다.

정확률(Precision)은 각 매장에 추천된 SKU에 대하여 실제 판매가 이루어진 SKU에 대한 비율로써 수식 (5)와 같이 표현된다.

4. 매장 별 취급 SKU 추천 방법에 대한 평가

$$\text{정확률(Precision)} = \frac{\text{추천 SKU 중에서 실제 판매된 SKU의 수}}{\text{매장에 추천된 SKU의 수}} \quad (5)$$

4.1 추천 시스템 평가 방법

본 연구에서는 기존의 협업 필터링 추천 방법과 매장 군집화 후 군집 내에서 협업 필터링을 하는 하이브리드 추천 방법을 제안하였으며 제안된 추천 방법을 기반으로 추천 시스템을 구현하였다. 두 추천 시스템에 대한 성능을 평가하기

각 매장에서 취급 SKU를 추천 받았지만, 실제로 측정 주차에 해당 SKU를 취급했는지 여부에 따라 판매 여부 측정이 의미가 있으므로 측정 주차의 Top 100 SKU에 대한 각 매장별 기초 재고 보유 여부를 파악할 필요가 있다. <Table 21>은 측정 주차인 2014년 06주차의 ‘K’ 브랜드 취급

<Table 21> 06, 2014 Top 100 SKU base Inventory (1: Inventory)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	1	1	1	1	0	1	0	0	1	1	...
2	KF4A111224	0	1	1	1	1	0	1	1	1	0	...
3	KF4A3U1221	0	1	1	0	0	0	0	0	1	0	...
4	KF4A4H1225	1	0	0	1	1	1	0	1	1	1	...
5	KF4A271213	1	1	0	1	0	0	1	1	1	1	...
6	KF4A341215	1	0	0	0	0	0	1	0	1	1	...
7	KF4A3U1215	1	0	1	1	0	1	0	1	1	0	...
8	KF3D3UU145	1	1	1	1	1	1	1	1	1	1	...
9	KF4A4H1223	1	1	1	0	1	1	1	1	1	1	...
10	KF4A6A1133	0	0	0	0	0	0	0	0	1	0	...
11	KF4A211215	0	0	0	1	1	1	1	1	1	1	...
12	KF4A5A12AD	0	0	0	1	1	0	1	1	1	1	...
13	KF4A271214	0	0	0	1	1	0	1	1	1	0	...
14	KF4A35121H	1	0	0	1	0	0	1	1	1	1	...
15	KF4A6A1134	0	0	0	0	0	0	0	0	1	0	...
16	KF4A11A222	0	0	1	1	1	1	1	1	1	1	...
17	KF4A841115	1	1	1	1	1	1	1	1	1	0	...
18	KF4A5N1225	1	0	1	1	1	1	0	1	1	0	...
19	KF4A4H121R	0	1	0	1	1	1	1	1	1	1	...
20	KF4A3Z1234	1	0	1	1	1	1	1	1	1	1	...
...

<Table 22> 06, 2014 Top 100 SKU Sale (1: Sale)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	0	0	0	0	0	0	0	1	1	0	...
2	KF4A111224	0	1	0	0	0	0	1	1	1	0	...
3	KF4A3U1221	0	0	0	0	1	0	0	0	0	0	...
4	KF4A4H1225	0	0	0	0	0	0	0	1	1	1	...
5	KF4A271213	0	1	0	1	0	0	0	1	1	1	...
6	KF4A341215	0	1	1	0	1	0	0	0	0	0	...
7	KF4A3U1215	0	0	1	0	1	1	0	0	0	0	...
8	KF3D3UU145	1	0	0	0	0	0	1	0	0	1	...
9	KF4A4H1223	0	0	0	0	0	0	0	0	1	1	...
10	KF4A6A1133	1	1	0	0	0	0	0	0	0	0	...
11	KF4A211215	0	1	0	1	1	0	0	0	0	0	...
12	KF4A5A12AD	1	0	0	0	0	0	0	0	1	1	...
13	KF4A271214	0	0	0	1	1	0	0	0	0	0	...
14	KF4A35121H	0	0	0	1	0	0	1	0	1	0	...
15	KF4A6A1134	0	0	0	0	0	0	0	0	0	0	...
16	KF4A11A222	0	0	0	0	1	1	1	1	0	1	...
17	KF4A841115	0	1	0	0	0	0	0	0	1	1	...
18	KF4A5N1225	0	0	0	0	0	0	0	0	0	0	...
19	KF4A4H121R	0	0	0	0	0	0	0	1	1	0	...
20	KF4A3Z1234	0	0	0	0	0	0	0	1	0	0	...
...

매장의 Top 100 SKU에 대한 매장별 기초 재고 보유 여부이다.

<Table 22>은 측정 주차인 2014년 06주차의 'K' 브랜드 취급 매장의 Top 100 SKU에 대한 매장별 판매 실적이다.

4.2 협업 필터링 방식 추천 SKU의 판매 정확률

3.3.1에서 협업 필터링 방식으로 각 매장에 추천한 SKU를(<Table 17>) 측정 주차인 2014년 06주차의 기초 재고로 보유하고 있었는지 여부를

<Table 21>과 매핑한 결과는 <Table 23>와 같으며, 해당 SKU들이 2014년 06주차에 판매되었는지 여부를 <Table 22>과 매핑한 결과는 <Table 24>과 같다.

즉, <Table 24>은 2014년 03주차 ~ 05주차의 판매 이력을 바탕으로 협업 필터링 추천 방식으로 2014년 06주차에 각 매장에서 취급하라 추천한 SKU들을 06주차 기초 재고로 보유하고 판매에 성공한 SKU들을 나타낸다.

〈Table 23〉 Collaborative Filtering Method Recommend SKU Inventory (1: Inventory)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	0	0	0	1	0	0	0	0	0	0	...
2	KF4A111224	0	0	0	0	0	0	1	0	0	0	...
3	KF4A3U1221	0	0	0	0	0	0	0	0	0	0	...
4	KF4A4H1225	1	0	0	1	0	0	0	0	0	0	...
5	KF4A271213	1	0	0	0	0	0	0	0	0	0	...
6	KF4A341215	0	0	0	0	0	0	0	0	0	0	...
7	KF4A3U1215	1	0	0	0	0	1	0	0	0	0	...
8	KF3D3UU145	1	0	0	0	0	0	0	0	0	0	...
9	KF4A4H1223	1	0	0	0	0	0	0	0	0	0	...
10	KF4A6A1133	0	0	0	0	0	0	0	0	1	0	...
11	KF4A211215	0	0	0	1	0	0	1	0	0	0	...
12	KF4A5A12AD	0	0	0	0	1	0	0	0	1	1	...
13	KF4A271214	0	0	0	1	0	0	0	0	0	0	...
14	KF4A35121H	0	0	0	0	0	0	0	0	0	0	...
15	KF4A6A1134	0	0	0	0	0	0	0	0	0	0	...
16	KF4A11A222	0	0	0	0	0	0	0	0	0	0	...
17	KF4A841115	0	0	0	0	0	1	0	0	0	0	...
18	KF4A5N1225	0	0	0	0	0	0	0	0	0	0	...
19	KF4A4H121R	0	1	0	0	1	1	0	0	1	0	...
...

〈Table 24〉 Collaborative Filtering Method Recommendation SKU Sales (1: Sales)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	0	0	0	0	0	0	0	0	0	0	...
2	KF4A111224	0	0	0	0	0	0	1	0	0	0	...
3	KF4A3U1221	0	0	0	0	0	0	0	0	0	0	...
4	KF4A4H1225	0	0	0	0	0	0	0	0	0	0	...
5	KF4A271213	0	0	0	0	0	0	0	0	0	0	...
6	KF4A341215	0	0	0	0	0	0	0	0	0	0	...
7	KF4A3U1215	0	0	0	0	0	1	0	0	0	0	...
8	KF3D3UU145	1	0	0	0	0	0	0	0	0	0	...
9	KF4A4H1223	0	0	0	0	0	0	0	0	0	0	...
10	KF4A6A1133	0	0	0	0	0	0	0	0	0	0	...
11	KF4A211215	0	0	0	1	0	0	0	0	0	0	...
12	KF4A5A12AD	0	0	0	0	0	0	0	0	1	1	...
13	KF4A271214	0	0	0	1	0	0	0	0	0	0	...
14	KF4A35121H	0	0	0	0	0	0	0	0	0	0	...
15	KF4A6A1134	0	0	0	0	0	0	0	0	0	0	...
16	KF4A11A222	0	0	0	0	0	0	0	0	0	0	...
17	KF4A841115	0	0	0	0	0	0	0	0	0	0	...
18	KF4A5N1225	0	0	0	0	0	0	0	0	0	0	...
19	KF4A4H121R	0	0	0	0	0	0	0	0	1	0	...
20	KF4A3Z1234	0	0	0	0	0	0	0	0	0	0	...
...

4.3 하이브리드 방식 추천 SKU의 판매 정확률

3.3.2에서 매장 군집화를 통한 하이브리드 방식으로 각 매장에 추천한 SKU를(<Table 19>) 측정 주차인 2014년 06주차의 기초 재고로 보유하고 있었는지 여부를 <Table 21>과 매핑한 결과는 <Table 25>와 같으며, 해당 SKU들이 2014년 06주차에 판매되었는지 여부를 <표 18>과 매핑한 결과는 <Table 26>와 같다.

즉, <Table 26>는 2014년 03주차 ~ 05주차의 판매 이력을 바탕으로 매장 군집화를 통한 하이브리드 추천 방식으로 2014년 06주차에 각 매장에서 취급하라 추천한 SKU들을 06주차 기초 재고로 보유하여 판매에 성공한 SKU들을 나타낸다.

4.4 판매 정확률 기반 성과 비교

2014년 03주차 ~ 05주차 소 매장 판매 수량 실적 Top 100개 SKU로 추천의 대상을 한정하여 협업 필터링과 하이브리드 필터링 방식으로 52개 매장 별로 취급 SKU를 추천하고, 추천 받은 SKU에 대한 2014년 06주차 매장별 판매 실적을 집계하여 두 추천 방식의 성과를 비교하였다. 두 추천 방식을 비교한 이유는 본 연구의 추천 방법이 기존 추천 방식 보다 높은 성과를 입증하기 위해 단순히 오프라인에 협업필터링을 적용한 것을 기준 모델로 정의하였다. 이 기준 모델에 오프라인 매장 관점의 특성을 잘 반영한 본 연구 모델인 하이브리드 필터링 방법과 비교 함으로써 성과를 입증한다. 이때, 동일 매

<Table 25> Hybrid Method Recommend SKU Inventory (1: Inventory)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	0	0	0	1	0	0	0	0	0	0	...
2	KF4A111224	0	0	0	0	0	0	1	0	0	0	...
3	KF4A3U1221	0	0	0	0	0	0	0	0	0	0	...
4	KF4A4H1225	0	0	0	1	0	0	0	0	0	0	...
5	KF4A271213	0	0	0	0	0	0	1	0	1	0	...
6	KF4A341215	0	0	0	0	0	0	1	0	0	0	...
7	KF4A3U1215	1	0	0	0	0	1	0	0	0	0	...
8	KF3D3UU145	1	0	0	0	0	0	0	1	0	1	...
9	KF4A4H1223	1	0	1	0	0	0	0	0	0	0	...
10	KF4A6A1133	0	0	0	0	0	0	0	0	1	0	...
11	KF4A211215	0	0	0	1	0	0	1	0	1	0	...
12	KF4A5A12AD	0	0	0	1	1	0	0	0	0	1	...
13	KF4A271214	0	0	0	1	0	0	0	0	0	0	...
14	KF4A35121H	0	0	0	0	0	0	0	0	0	0	...
15	KF4A6A1134	0	0	0	0	0	0	0	0	1	0	...
16	KF4A11A222	0	0	1	0	0	1	0	0	0	0	...
17	KF4A841115	0	0	0	0	0	1	1	0	0	0	...
18	KF4A5N1225	0	0	0	0	0	0	0	0	0	0	...
19	KF4A4H121R	0	1	0	0	1	1	1	0	0	0	...
...

<Table 26> Hybrid Method Recommendation SKU Sales (1: Sales)

Rank	SKU	매장										
		B123	B324	B353	B369	B400	B521	B539	B563	B655	B735	...
1	KF4A3U122R	0	0	0	0	0	0	0	0	0	0	...
2	KF4A111224	0	0	0	0	0	0	1	0	0	0	...
3	KF4A3U1221	0	0	0	0	0	0	0	0	0	0	...
4	KF4A4H1225	0	0	0	0	0	0	0	0	0	0	...
5	KF4A271213	0	0	0	0	0	0	0	0	1	0	...
6	KF4A341215	0	0	0	0	0	0	0	0	0	0	...
7	KF4A3U1215	0	0	0	0	0	1	0	0	0	0	...
8	KF3D3UU145	1	0	0	0	0	0	0	0	0	1	...
9	KF4A4H1223	0	0	0	0	0	0	0	0	0	0	...
10	KF4A6A1133	0	0	0	0	0	0	0	0	0	0	...
11	KF4A211215	0	0	0	1	0	0	0	0	0	0	...
12	KF4A5A12AD	0	0	0	0	0	0	0	0	0	1	...
13	KF4A271214	0	0	0	1	0	0	0	0	0	0	...
14	KF4A35121H	0	0	0	0	0	0	0	0	0	0	...
15	KF4A6A1134	0	0	0	0	0	0	0	0	0	0	...
16	KF4A11A222	0	0	0	0	0	1	0	0	0	0	...
17	KF4A841115	0	0	0	0	0	0	0	0	0	0	...
18	KF4A5N1225	0	0	0	0	0	0	0	0	0	0	...
19	KF4A4H121R	0	0	0	0	0	0	0	0	0	0	...
20	KF4A3Z1234	0	0	0	0	0	0	0	0	0	0	...
...

장에서 하나의 SKU에 대한 복수 판매 수량은 고려되지 않았으며 판매 수량에 관계없이 개수 1로 계산되었다.

<Table 27>에서 SKU의 개수는 한 SKU의 복수 수량을 의미하는 것이 아니라 SKU의 종류의 개수를 나타낸다. 즉, 각 매장별로 추천, 기초 재

고, 판매 SKU의 개수는 한 SKU의 복수 수량 (Quantity)이 아니라 종류의 개수이며 (1), (2), (3)은 52개 각 매장의 결과를 모두 합한 값이다.

2014년 06주차 'K' 브랜드의 매장별 추천 SKU의 판매 정확률은 정확률 산출 기준 1,2 모두 하이브리드 필터링 추천 방식이 우수하다 할 수 있

<Table 27> Recommendation Method Accuracy Measurement Result (Top 100 SKU)

항목	협업 필터링	하이브리드 필터링
(1) 추천 SKU 개수	1,412	1,303
(2) 기초 재고로 보유한 추천 SKU 개수	912	873
(3) 기초 재고로 보유한 추천 SKU 의 판매 개수	91	95
(4) 정확률2 : (3)/(2)	0.09978	0.10882

다. 정확률2의 기준으로 하이브리드 필터링 추천 방식의 성과는, ‘52개 매장에서 기초 재고로 보유하고 있던 총 873개의 각 매장 별 추천 SKU 중 총 95개 SKU가 전 매장에 걸쳐 판매되었다’는 것으로 설명할 수 있다.

5. 결론

본 연구에서는 개인화 추천에 이용되어 왔던 협업 필터링과 하이브리드 필터링 등의 추천 시스템(Recommender System)을 국가별, 지역별로 복수의 판매 매장을 통해 동종 브랜드를 취급하는 유통 기업의 매장 단위 취급 SKU 추천 방식을 제안하였다. 각 매장의 취급 품목별 구매 데이터를 활용하여 각 매장 별 유사성(Similarity)을 계산하고 각 매장의 SKU별 판매 이력에 따라 협업 필터링을 하여 최종적으로 매장에 개별 SKU를 추천하였다. 또한 매장 프로파일 데이터를 활용하여 주변수 분석(PCA : Principal Component Analysis) 및 군집 분석(Clustering)을 통하여 매장을 4개의 군집으로 분류한 뒤 각 군집 내에서 협업 필터링을 적용한 하이브리드 필터링 방식으로 추천 시스템을 구현하고 실제 판매 데이터를 바탕으로 두 방식의 성능을 측정하였다.

2014년 03주차 ~ 05주차 소 매장 판매 수량 실적 Top 100개 SKU로 추천의 대상을 한정하여 협업 필터링과 하이브리드 필터링 방식으로 매장별 취급 SKU를 추천하고, 추천 받은 SKU에 대한 2014년 06주차 판매 실적을 집계하여 두 추천 방식의 성과를 비교한 결과 기초 재고로 보유하고 있던 추천 SKU 중 두 방식 모두 약 10%의 정확률을 보였다. 즉, 추천한 10개의 SKU중 1개의 비율로 실제로 판매가 이루어졌음을 알 수 있

었다. 그 중 매장 속성으로 매장들을 군집화한 후 협업 필터링을 적용한 하이브리드 방식이 협업 필터링만을 적용한 방식보다 0.9% 우수한 정확률을 보였다. 개인화 추천 연구 영역에서 하이브리드 방식이 우수한 성능을 보여왔던 것과 동일한 결과이다.

본 연구의 결과가 갖는 의의는 개인화 추천 알고리즘을 동일 브랜드를 취급하는 복수의 판매 매장에 적용하여 의미 있는 결과를 도출하고 실제 기업을 대상으로 시스템으로 구축하여 활용할 수 있는 구체적 방법론을 제시했다는 데에 있다. 개인화 영역을 위주로 이루어졌던 기존의 추천 시스템과 관련한 학계의 연구 영역을 동종 브랜드를 취급하는 기업의 판매 매장으로 확장시킨 첫 시도라는 데에도 의미가 있다.

본 연구의 추천 시스템은 과거 판매 제품의 품목 정보를 바탕으로 매장간 유사성을 구하고 가장 유사한 매장 즉, 최근접 매장으로 짝지어진 매장끼리 가장 많이 팔린 SKU 순으로 서로의 취급 SKU를 추천하는 알고리즘을 기본 토대로 하고 있다. 그에 따라 몇 가지 감안해야 할 사항으로, 제품의 특성상 저렴한 SKU 또는 교체 주기가 빈번한 SKU의 경우 전 매장에 걸쳐 많이 팔려 우선순위가 높은 추천 SKU가 될 수 있는데, 이에 따라 단가가 높고 계절성이 있는 제품이 추천 우선순위에서 밀려나는 현실적인 문제가 발생할 수 있다. 예를 들면, 양말과 코트의 경우 위의 알고리즘 대로라면 양말이 추천 우선순위를 가질 확률이 높게 된다. 따라서 SKU별 판매 수량뿐 아니라 매출의 크기를 함께 고려하여 추천 SKU 대상을 선별하는 것이 매장의 매출 신장에 도움이 될 수 있다. 판매 수량과 매출 크기를 어떤 비율의 조합으로 정해야 매출의 크기를 극대화할 수 있는지는 본 연구에서는 다루고 있지 않

으며 향후 최적화 관점에서의 실증 연구가 필요한 부분이다. 비슷한 문제로 판매 확률을 고려하여 추천 SKU 대상을 정하는 부분도 본 연구에서는 다루고 있지 않다. 예를 들어, 제품의 PLC(Product Life Cycle) 정보를 고려하여 EOS(End of Sales) 시점이 다가온 경우 어느 시점에서 추천 대상에서 제외해야 판매 확률을 극대화할 수 있는지도 최적화 관점에서의 실증 연구가 필요한 부분이다.

협업 필터링 방식과 하이브리드 방식 추천 결과에 대한 성능을 실제 판매 데이터를 집계하여 수치화는 하였으나, 실험군과 대조군을 명확히 구분하여 POC(Proof of Concept)를 수행하는 실험 연구는 동일 매장, 동일 시간, 동일 고객 등의 구매 환경을 재현할 수 없는 현실적인 제약으로 진행되지 못하여 추천 SKU의 기초 재고 보유 여부와 판매 여부만으로 두 방식의 결과를 비교하였다.

지금까지의 추천 시스템 연구가 개인화 영역 위주로 이루어져와 본 연구에서와 같은 매장 단위의 취급 SKU 추천에 관한 다양한 연구가 앞으로 활발히 이루어져 전 세계 판매 유통망을 보유한 기업에 유용하게 활용되어야 한다. 특히, 협업 필터링 방식에 연관 판매 추천 방식을 접목한 하이브리드 추천 방식(Ujwala et al, 1961) 제품의 자기잠식(Cannibalization) 효과가 고려된 판매 확률 관점에서의 추천 우선순위 연구, 제품의 부피와 단가, 매장 면적 등을 감안한 SKU별 재고 비용을 고려한 재무적 관점에서의 추천 우선순위 연구 등 현업의 실질적인 요구사항을 고려한 다양한 추천 방식의 연구가 필요하다.

참고문헌(References)

- Al-Shamri, M.Y.H. and K.K. Bharadwaj, "Fuzzy-genetic approach to recommender systems based on a novel hybrid user model," *Expert Systems with Applications*, Vol.35, No.3(2008), 1386~1399.
- Bell, R. and Y. Koren, "Lessons from the Netflix Prize Challenge," *SIGKDD Explorations*, Vol.9, No.2(2007), 71~84.
- Billsus, D. and M. J. Pazzani, "Learning Collaborative Information Filters," *Proceedings of 15th International Conference on Machine Learning*, (1998), 46~45.
- Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A., "Recommender systems survey," *Knowledge-Based Systems*, Vol.46, No.1(2013), 109~132.
- Burke, R. "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User Adapted Interaction*, Vol.12, No.4(2002), 331~370.
- Campos, L.M., J.M. Fernández-Luna, J.F. Huete, and M.A. Rueda-Morales, "Combining content-based and collaborative recommendations: a hybrid approach based on Bayesian Networks," *International Journal of Approximate Reasoning*, Vol.51, No.7(2010), 785~799.
- Christakou, C. and A. Stafylopatis, "A hybrid movie recommender system based on neural networks," *International Conference on Intelligent Systems Design and Applications*, (2005), 500~505.
- Gao, L.Q. and C. Li, "Hybrid personalized recommended model based on genetic algorithm," *International Conference on*

- Wireless Communication, Networks and Mobile Computing*, (2008), 9215~9218.
- Goldberg, D., D. Nichols, B. M. Oki, and D. Terry
“Using collaborative filtering to weave an information tapestry.” *Communications of the ACM*, Vol.35, No.12(1992), 61~70.
- Gong, S., “A collaborative filtering recommendation algorithm based on user clustering and item clustering,” *Journal of Software*, Vol.5, No.7 (2010), 745~752.
- Herlocker, J., J. A. Konstan, R. Borchers, and J. Riedl, “An Algorithmic Framework for Performing Collaborative Filtering,” *Proceedings of the 22th ACM SIGIR Conf. on Research and Development in Information Retrieval*,(1999), 230 ~ 237.
- Jeon, B. and H. Ahn, “A Collaborative Filtering System Combined with Users’ Review Mining: Application to the Recommendation of Smartphone Apps,” *Journal of Intelligence and Information System*, Vol. 21, No. 2(2015), 1~18.
- Kim, K.-j. and H. Ahn, “User-Item Matrix Reduction Technique for Personalized Recommender Systems,” *Journal of Information Technology Applications & Management*, Vol. 16, No. 1(2009), 97~113.
- Kim, K.-j. and H. Ahn, “Collaborative Filtering with a User-Item Matrix Reduction Technique for Recommender Systems,” *International Journal of Electronic Commerce*, Vol. 16, No. 1(2011), 107~128.
- Kim, M. and K. Kim, “Recommender Systems using Structural Hole and Collaborative Filtering,” *Journal of Intelligence and Information System*, Vol. 20, No. 4(2014), 107~120.
- Lee. H. L., “A Study of the V.M.D process for sales promotion in the fashion marketing,” *Design Science Research*, Vol.5, No.2(2002), 75~88.
- Lee, J. and H. S. Park, “Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering,” *Journal of Intelligence and Information System*, Vol. 13, No. 4(2007), 65~78.
- Lekakos, G. and G. M. Giaglis, “Improving the Prediction Accuracy of Recommendation Algorithms: Approaches Anchored on Human Factors”, *Interacting with Computers*, Vol.1, No.18(2006), 410 ~ 431.
- Renaud-Deputter, S., T. Xiong, and S. Wang, “Combining collaborative filtering and clustering for implicit recommender system,” *Proceedings of 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA)*, IEEE, (2013), 748~755.
- Resnick, P. and H. R. Varian, “Recommender Systems,” *Communications of the ACM*, Vol.40, No.1 (1997), 56~58.
- Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, “GroupLens : An Open Architecture for Collaborative Filtering of Netnews,” *Proceedings of the ACM Conf. on Computer Supported Cooperative Work*, (1994), 175 ~ 186.
- Sarwar, B., “Sparsity, Scalability, and Distribution in Recommender Systems,” Ph.D. Diss., Dept. of Computer and Information Sciences, Univ. of Minnesota(2001).
- Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, “Application of dimensionality reduction in recommender system - a case study” *ACM*

- WebKDD Workshop*, Vol.1, No.2(2000), 264-272.
- Shinde, S.K. and U. Kulkarni, "Hybrid personalized recommender system using centering-bunching based clustering algorithm," *Expert Systems with Applications*, Vol.39, No.1(2012), 1381~1387.
- Ujwala H., S. R. Wanaskar, and V. D. Mukhopadhyay, "A Hybrid Web Recommendation System Based on the Improved Association Rule Mining Algorithm," *Journal of Software Engineering and Applications*, Vol.06, No.8(2013), 396~411.
- Wen, J. and W. Zhou, "An improved item-based collaborative filtering algorithm based on clustering method," *Journal of Computational Information Systems*, Vol.8, No.2(2012), 571~ 578.

Abstract

SKU recommender system for retail stores that carry identical brands using collaborative filtering and hybrid filtering

Denis Yongmin Joe* · Kihwan Nam**

Recently, the diversification and individualization of consumption patterns through the web and mobile devices based on the Internet have been rapid. As this happens, the efficient operation of the offline store, which is a traditional distribution channel, has become more important. In order to raise both the sales and profits of stores, stores need to supply and sell the most attractive products to consumers in a timely manner. However, there is a lack of research on which SKUs, out of many products, can increase sales probability and reduce inventory costs. In particular, if a company sells products through multiple in-store stores across multiple locations, it would be helpful to increase sales and profitability of stores if SKUs appealing to customers are recommended.

In this study, the recommender system (recommender system such as collaborative filtering and hybrid filtering), which has been used for personalization recommendation, is suggested by SKU recommendation method of a store unit of a distribution company that handles a homogeneous brand through a plurality of sales stores by country and region. We calculated the similarity of each store by using the purchase data of each store's handling items, filtering the collaboration according to the sales history of each store by each SKU, and finally recommending the individual SKU to the store.

In addition, the store is classified into four clusters through PCA (Principal Component Analysis) and cluster analysis (Clustering) using the store profile data. The recommendation system is implemented by the hybrid filtering method that applies the collaborative filtering in each cluster and measured the performance of both methods based on actual sales data.

Most of the existing recommendation systems have been studied by recommending items such as movies and music to the users. In practice, industrial applications have also become popular. In the

* KAIST College of Business, Korea Advanced Institute of Science and Technology (KAIST)

** Corresponding Author: Kihwan Nam

College of Business at Korea Advanced Institute of Science and Technology (KAIST)

SUPEX 293, 85 Hoegiro Dongdaemoon-gu Seoul, 130-722, South Korea

Tel: +82-10-4930-8317, E-mail: namkh@kaist.ac.kr

meantime, there has been little research on recommending SKUs for each store by applying these recommendation systems, which have been mainly dealt with in the field of personalization services, to the store units of distributors handling similar brands.

If the recommendation method of the existing recommendation methodology was 'the individual field', this study expanded the scope of the store beyond the individual domain through a plurality of sales stores by country and region and dealt with the store unit of the distribution company handling the same brand SKU while suggesting a recommendation method. In addition, if the existing recommendation system is limited to online, it is recommended to apply the data mining technique to develop an algorithm suitable for expanding to the store area rather than expanding the utilization range offline and analyzing based on the existing individual.

The significance of the results of this study is that the personalization recommendation algorithm is applied to a plurality of sales outlets handling the same brand. A meaningful result is derived and a concrete methodology that can be constructed and used as a system for actual companies is proposed. It is also meaningful that this is the first attempt to expand the research area of the academic field related to the existing recommendation system, which was focused on the personalization domain, to a sales store of a company handling the same brand.

From 05 to 03 in 2014, the number of stores' sales volume of the top 100 SKUs are limited to 52 SKUs by collaborative filtering and the hybrid filtering method SKU recommended. We compared the performance of the two recommendation methods by totaling the sales results. The reason for comparing the two recommendation methods is that the recommendation method of this study is defined as the reference model in which offline collaborative filtering is applied to demonstrate higher performance than the existing recommendation method. The results of this model are compared with the Hybrid filtering method, which is a model that reflects the characteristics of the offline store view.

The proposed method showed a higher performance than the existing recommendation method. The proposed method was proved by using actual sales data of large Korean apparel companies. In this study, we propose a method to extend the recommendation system of the individual level to the group level and to efficiently approach it. In addition to the theoretical framework, which is of great value.

Key Words : Store SKU Recommender System, Collaborative filtering, Hybrid filtering, Retail marketing, Apparel industry

Received : July 21, 2017 Revised : October 11, 2017 Accepted : November 20, 2017

Publication Type : Regular Paper Corresponding Author : Kihwan Nam

저 자 소개



조용민

KAIST 경영공학부에서 금융분야 박사학위를 취득하였다. 현재 중앙대학교 경영대학 대우교수, 주식회사 센티언스 파트너로 재직 중이다. 주요 관심분야는 기업재무, 투자를 포함한 재무분야 실증연구와 금융 빅데이터 분석 등이다.



남기환

KAIST 경영공학부에서 MIS 박사학위를 취득하였다. 현재 한양대학교 경영대학 겸임교수, UNIST 경영대학 대우교수로 재직 중이다. 주요 관심분야는 Business Analytics & Business Intelligence, Big Data Analytics, Data Mining, Statistical Analysis, Recommender Systems, Econometrics Models, Machine Learning, Deep Learning 등 이다. IT기업들의 경영 전반에 걸친 문제들을 데이터 기반으로 접근함으로써 보다 효과적인 마케팅 및 경영 전략을 수립하는데 관심을 가지고 있다.