

오프라인 쇼핑몰에서 고객의 과거 구매 패턴을 활용한 아이템 기반 협업필터링 성능 개선에 관한 연구

정 석 봉*

Improvement of Item-Based Collaborative Filtering by Applying Each Customer's Purchase Patterns in Offline Shopping Malls

Seok Bong Jeong*

Abstract

Item-based collaborative filtering (IBCF) is an important technology that is widely used in recommender system of online shopping malls. It uses historical information to compute item-item similarity and make predictions. However, in offline shopping each customer's purchasing pattern can be occurred continuously and repeatedly due to time and space constraints contrast to online shopping. Those facts can make IBCF to have limitations from being applied to offline shopping malls directly. In order to improve the quality of recommendations made by IBCF in offline shopping mall, we propose an ensemble approach that considers both item-item similarity of IBCF and each customer's purchasing patterns which are modeled by item networks. Our experimental results show that this approach produces recommendation results superior to those of existing works such as pure IBCF or bestseller approaches.

Keywords : Recommender System, Collaborative Filtering, Offline Shopping, Item Based Collaborative Filtering

Received : 2017. 04. 06. Revised : 2017. 06. 28. Final Acceptance : 2017. 07. 11

※ This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A8012533).

* Professor, Department of Railway, Kyungil University, Kyungil University, 50, Gamsil-gil, Hayang-eup, Gyeongsan-si, Gyeongsangbuk-do, 38428, Rep. of Korea, Tel : +82-53-600-5071, e-mail : sbjung@kiu.ac.kr

1. 서 론

1990년대 이래로 ‘개인 맞춤형 추천’에 대한 학술적·상업적 연구는 활발히 진행되고 있으며, 이를 기반으로 한 추천 시스템은 소셜 네트워크, 전자상거래 등 다양한 영역에 폭넓게 적용되고 있다[Nicolaus Mote, 2004]. 이러한 추천 시스템에서는 개별 고객에게 얼마나 정확한 추천을 수행하는지가 그 품질을 결정한다[Liu et al., 2009].

특히, 온라인 쇼핑 시장에서 협업필터링(Collaborative Filtering)은 그 추천 정확도 측면에서 가장 성공적인 추천기법으로 알려져 있다[Park et al., 2003, 2009]. 협업필터링은 각 고객을 위한 추천 아이템을 도출하기 위해 동일 관심사나 구매패턴을 가진 주변 고객(이웃)들을 찾고 그들이 선택한 아이템을 추천 아이템으로 활용하는 방식이다.

이러한 협업필터링은 사용자 기반 협업필터링(user based collaborative filtering; UBCF)과 아이템 기반 협업필터링(item based collaborative filtering; IBCF)으로 나눌 수 있다. UBCF는 앞서 설명한 바와 같이 추천 대상 고객과 유사한 사용자 그룹을 찾고 이 그룹이 공통적으로 선호하는 아이템을 추천하는 방식으로, 정확한 추천이 가능하다는 장점이 있다. 반면 IBCF는 고객들의 선호도를 바탕으로 아이템간의 유사도를 계산하고, 특정 사용자가 어떤 아이템을 구매하거나 좋다고 평가하면 그와 비슷한 아이템을 추천하는 방식이다.

협업필터링에 관한 기존 연구들을 살펴보면, IBCF 방식이 더 효과적이고 정확하다는 일부 사례연구도 있으며[Xu, 2013], 그와 상반된 주장의 연구도 있다[Im, 2016]. 그러나 처리해야 하는 고객 및 아이템 수가 증가할수록 계산속도의 이점 때문에 IBCF를 사용하는 것이 효과적이라는 주장에는 이견이 없으며[Xu, 2013; Im, 2016], 따라서

아마존과 같은 대형 온라인 쇼핑몰에서는 IBCF를 활용하는 것으로 알려져 있다[Linden et al., 2003].

한편, 오프라인 쇼핑 시장의 경우 우편물 발송(DM), 텔레마케팅(TM) 등의 전통적인 추천 활동을 수행하고는 있지만, 온라인 대비 가용한 고객 정보 및 정보전달 매체의 한계로 인해 그 활동이 제한적이었다[Kim and Jeong, 2016; Choi, 2011]. 이러한 이유로 오프라인 쇼핑 시장에서 상품 추천과 관련된 연구는 주로 상품 간의 연관성 파악에 의한 진열 및 배치, 세트 상품 구성 등에 국한되었으며[Kim, 2012; Radeder and Chawla, 2009], 고객맞춤형 추천에 관한 연구는 상대적으로 빈약한 것으로 파악된다.

이에 본 연구에서는 오프라인 쇼핑몰을 대상으로 IBCF를 응용한 실시간 맞춤형 브랜드 추천 기법을 구현하고, 그 성능을 향상시키는 방안을 모색하고자 한다. 특히, 실내측위 및 IoT 기술의 발전으로 인해 오프라인 쇼핑 시장에서도 고객의 구매 행위를 파악하는 것이 가능해지고 있으며[Choi, 2016], 스마트폰 등을 통한 실시간 쿠폰 전송 등의 마케팅 활동이 가능해진 상황에서[Lee, 2014], 마케팅 활동의 수단으로 추천 시스템의 역할은 더욱 중요해질 것으로 예상된다.

2. 제안 추천 방법론

온라인에 비해 오프라인 쇼핑 환경에서 고객은 특정 쇼핑몰을 영업시간 내에 방문해야 한다는 시간적·공간적 제약을 갖게 된다. 쇼핑을 위해 요구되는 이러한 제약들은 고객이 신규 브랜드를 탐색하기 보다는 과거 구매 경험이 있는 브랜드를 선호하는데 영향을 줄 것이라 가정할 수 있다. 즉, 브랜드의 전환비용(switching cost)이 온라인에 비해 크다고 가정할 수 있으며, 이는 온라인 쇼핑에 주로 활용되는 협업필터링 기법

을 오프라인 쇼핑에 직접적으로 적용하는데 한계가 있을 수 있음을 시사한다.

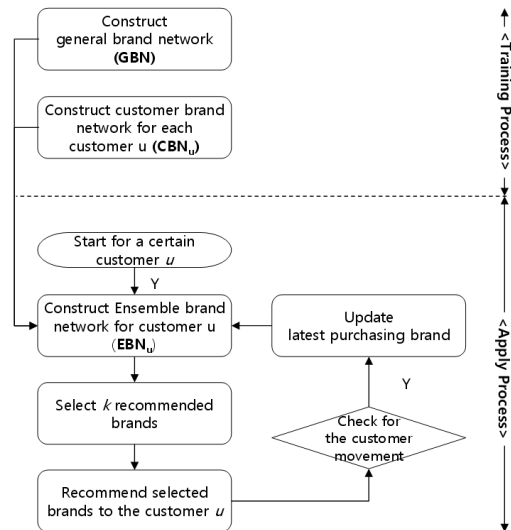
본 연구에서는 이러한 가정에 근거하여 협업 필터링을 오프라인 쇼핑의 특성을 고려하여 효과적으로 적용할 수 있도록, 고객의 과거 구매 패턴을 기존 IBCF 방식과 앙상블 하여 추천 시스템을 구현하는 방안을 제안한다.

앙상블 기법은 다수의 알고리즘을 조합하여 결과를 도출하는 방법으로, 단일 알고리즘에 비해 그 추천 정확도가 높다는 것이 많은 실증 연구를 통해 밝혀졌다[Kim and Cho, 2015; Piotte and Chabbert, 2009; Bar et al., 2013; Lee and Kim, 2013].

한편, 본 연구에서는 IBCF의 대상 아이템을 오프라인 쇼핑물(백화점) 내에 있는 각 브랜드로 간주하였다. 이는 백화점의 확보 가능한 거래내역 데이터가 개별 상품 수준에서는 파악이 불가능하며, 주로 POS 번호를 기준으로 한 브랜드 수준에서 파악이 가능하였기 때문이다. 그러나 제안된 알고리즘은 IBCF를 기반으로 설계된 기법이기 때문에 많은 상품을 취급하는 일반적인 오프라인 쇼핑물의 상품수준에서도 적용 가능할 것으로 예상할 수 있다.

본 연구에서 제안하는 기법은 크게 두 부분으로 구성되는데, <Figure 1>과 같이 학습과정(Training Process)과 적용과정(Apply Process)으로 이루어진다. 학습과정은 쇼핑물의 과거 일정 기간 동안 각 고객별/일자별 브랜드 구매이력을 기반으로 이루어지며, 일반적인 IBCF에서 생성되는 general brand network(GBN)와 각 개별 고객 u 의 과거 구매패턴의 특성을 반영하는 customer brand network(CBN_u)를 생성하는 역할을 한다.

학습과정이 완료되면 적용 과정에서는 당일 방문 고객에 대한 실시간 맞춤형 추천 브랜드를 도출한다. 제안된 기법에서는 고객 u 의 직전 방



<Figure 1> The Whole Framework for the Proposed Scheme

문 브랜드에 기반하여 GBN 및 CBN_u 를 앙상블 하여 Ensemble Brand Network(EBN_u)를 생성하고, 이를 활용하여 k 개의 브랜드를 추천하게 된다. EBN_u 는 고객 u 를 위한 추천 브랜드 도출에 기반이 되는 네트워크로, 본 연구에서는 이를 기반 네트워크라고 칭한다.

이후 쇼핑물의 마케팅 총괄실은 스마트폰 앱이나 SMS, 고객 근처의 DID(Digital Information Display) 등을 통해 해당 고객에게 추천 브랜드의 쿠폰을 제공하거나, 광고를 수행하는 등의 실시간 맞춤형 마케팅 활동을 수행할 수 있다.

각 과정에 대한 상세 설명은 다음절부터 제시되어 있다.

2.1 학습과정 : GBN 과 CBN_u 의 구성

제안된 기법의 학습과정은 쇼핑물의 과거 브랜드 구매이력을 기반으로 GBN 및 CBN_u 를 구성하는 과정이다.

먼저, 쇼핑물에 입점한 브랜드의 전체 집합을 B , 해당 쇼핑물에서 구매 경험이 있는 고객의 전체 집합을 U 라고 하자. 또한, <Table 1>과 같이

각 고객이 특정 일자에 구매한 브랜드 리스트를 구매 트랜잭션이라 하고, 학습기간 동안 발생한 구매 트랜잭션의 전체 집합을 T 라고 정의하자. 이는 동일 고객이라 하더라도 서로 다른 일자에 발생한 브랜드 구매 이력은 서로 다른 트랜잭션임을 의미한다.

<Table 1>의 트랜잭션 $t1$ 은 특정 일자에 특정 고객이 브랜드 $b1$, $b4$ 및 $b5$ 에서 구매를 수행했음을 나타낸다.

(Table 1) Purchasing Transaction

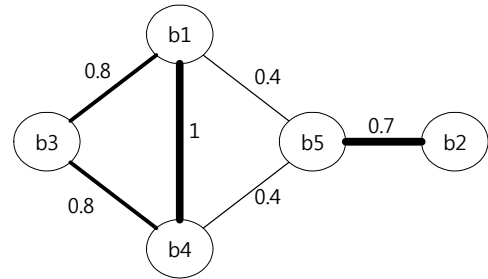
	$b1$	$b2$	$b3$	$b4$	$b5$
$t1$	1	0	0	1	1
$t2$	1	0	1	1	0
$t3$	0	1	0	0	1
$t4$	1	0	1	1	0

위와 같은 구매 트랜잭션 T 를 기반으로 각 브랜드 간 유사도를 나타내는 GBN 을 구성할 수 있다. 이때 브랜드 간 유사도를 측정하는 방법으로는 피어슨 상관관계수, 코사인 계수, 자카드 계수 등이 있으나, 본 연구에서는 IBCF에서 그 성능이 우수하다고 알려진 코사인 계수를 사용한다[Kim and Jeong, 2016].

두 브랜드 b_i 와 b_j 간의 코사인 계수는 다음 식 (1)과 같이 계산할 수 있다. 이는 T 상에서 두 브랜드 벡터가 갖는 내적공간의 각도에 대한 코사인 값을 이용하여 벡터 간의 유사한 정도를 측정하는 방법이다[Linden et al., 2003].

$$\cos(b_i, b_j) = \frac{\vec{b}_i \cdot \vec{b}_j}{\|\vec{b}_i\| \times \|\vec{b}_j\|} \quad (1)$$

예를 들어, <Table 1>의 트랜잭션 데이터를 코사인 계수를 통해 GBN 으로 변환한 결과는 <Figure 2>과 같다. 코사인 계수는 항상 0과 1 사이의 값을 갖는다.



(a) Graph

	$b1$	$b2$	$b3$	$b4$	$b5$
$b1$	0	0	0.8	1	0.4
$b2$	0	0	0	0	0.7
$b3$	0.8	0	0	0.8	0
$b4$	1	0	0.8	0	0.4
$b5$	0.4	0.7	0	0.4	0

(b) Matrix

(Figure 2) General Brand Network(GBN)

일반적으로 IBCF 추천 기법은 고객이 선택한 아이템과 유사도가 가장 높은 아이템을 추천하는 방식으로 작동한다. <Figure 2>의 예에서 만일 특정 고객이 브랜드 $b5$ 에서 구매를 수행할 경우, $b5$ 와 유사도가 높은 $b2$ 를 추천하는 방식이다.

한편, 오프라인 쇼핑몰에서는 고객의 과거 구매 패턴을 활용하여 추천 정확도를 보다 향상시킬 수 있다. 앞서 언급한바와 같이 오프라인 쇼핑 고객의 경우 시간적·공간적 제약으로 인해 기존 구매패턴이 반복되는 경향이 있을 수 있다. 즉, 협업필터링이 활발히 적용되고 있는 영화나 도서, 온라인 쇼핑시장 등과는 달리, 각 고객 별로 브랜드 구매 패턴이 지속적이고 반복적으로 나타나는 경향이 존재할 수 있으며, 따라서 고객의 과거 구매패턴을 추천에 활용할 경우 추천의 정확도가 향상될 가능성이 있다.

본 연구에서는 이러한 고객 별 구매 패턴을 CBN_u 라는 고객 u 에게 특화된 브랜드 네트워크를 통해 구현한다. 특정 고객 u 의 구매 패턴을

파악하기 위한 CBN_u 의 구성 방법은 GBN 과 동일하지만, 학습 대상 트랜잭션으로 T 가 아닌 해당 고객이 발생시킨 트랜잭션만을 활용한다. 즉, 고객 u 의 학습기간 동안의 구매 트랜잭션을 T_u 라고 하면($T_u \subset T$), T_u 를 기반으로 CBN_u 를 구성하며, 이때 브랜드간 유사도는 GBN 의 구성 때와 동일하게 코사인 계수를 사용한다.

한편, 학습과정에서 구성된 GBN 및 CBN_u (for all $u \in U$)는 적용과정에서 각 고객의 추천 브랜드를 도출하는데 활용된다.

2.2 적용과정 : EBN_u 구성 및 추천 브랜드 도출

적용 과정은 당일 방문 고객 u 에 대한 추천을 위해 기반 네트워크인 EBN_u 를 구성하고, 고객 u 의 가장 최근 브랜드 구매정보를 기반으로 추천 브랜드를 실시간 도출하는 과정이다.

이를 위해 먼저 학습과정에서 도출한 GBN 과 CBN_u 의 결과를 앙상블 하여 기반 네트워크인 EBN_u 를 구성한다. 본 연구에서 제안하는 앙상블 방법은 GBN 과 CBN_u 의 브랜드 간 개별 유사도를 결합하는 방법으로, 김민정과 조윤호[2015]의 UBCF 관련 연구에서 고객 간 유사도를 앙상블한 방법과 유사하다.

본 연구에서는 브랜드 간 유사도를 앙상블하기 위하여 단순평균 및 가중평균 방식을 활용한다. 단순평균 방법은 GBN 과 CBN_u 에서 각 브랜드 간의 유사도를 산술평균하여 EBN_u 의 브랜드 간 유사도로 사용하는 방식이며, 가중평균 방식은 GBN 과 CBN_u 의 유사도를 평균할 때 가중치로 각각의 네트워크의 추천 성능을 나타내는 F1값을 사용한다¹⁾. 즉, GBN 과 CBN_u 각각을 추천을 위한 기반 네트워크로 간주하고 추천 브랜드

를 도출했을 때의 F1값을 $F1_{GBN}$ 및 $F1_{CBN_u}$ 라고 하면, 가중평균 방식으로 앙상블한 EBN_u 는 다음 식 (2)와 같이 구할수 있다.

$$EBN_u = F1_{GBN} \times GBN + F1_{CBN_u} \times CBN_u \quad (2)$$

EBN_u 가 구성된 이후에는 고객 u 의 최근 구매 브랜드 b_t 을 기준으로 추천 브랜드를 도출한다. 즉, 추천 브랜드의 수 k 를 미리 설정하고 EBN_u 에서 b_t 의 이웃 노드 중에 b_t 과의 유사도가 높은 순으로 k 개의 브랜드를 추출한다. 물론 이웃 노드의 수가 k 보다 작을 경우에는 이웃 노드의 수만큼의 추천브랜드를 생성한다.

이렇게 선택된 추천 브랜드는 고객의 스마트폰의 앱, SMS 및 DID 등의 다양한 광고 매체를 통하여 실시간으로 고객에게 전달된다.

한편, 해당 고객이 다른 브랜드로 이동하게 될 경우 고객의 최근 구매 브랜드 b_t 을 업데이트하고 업데이트된 b_t 을 기준으로 다시 추천 브랜드를 생성한다. 이러한 과정은 고객이 당일 쇼핑을 마치고 쇼핑물을 이탈할 때까지 반복된다.

3. 실험

3.1 실험 데이터

본 연구에서는 제안된 기법의 성능 및 적용가능성을 검증하기 위하여 국내의 유명 백화점 중 하나인 A백화점의 실제 구매 데이터를 대상으로 실험을 진행하였다. A백화점은 년 매출 약 5천억 원에 12만 여 명의 회원을 보유한 국내 중형의 백화점이다.

실험에 포함된 고객은 전체 고객 중 구매 이력의 추적이 가능한 회원고객으로 한정하였으며, 구매이력의 파악을 위해 실험 기간인 1년 동안 최소 3회 이상 방문한 고객을 대상으로 하였다.

1) F1 값에 대한 자세한 설명은 제 3.2절을 참조.

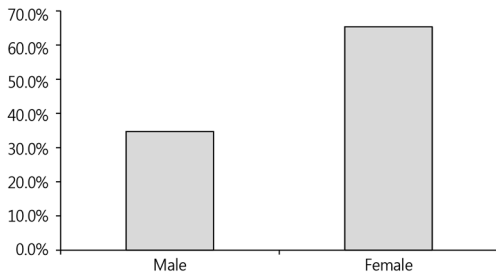
또한, 실험에 앞서 브랜드 구분이 애매한 행사 및 푸드코트 관련 브랜드와 실험 기간 동안에 신규 입점 혹은 폐점한 브랜드를 제외하는 등의 전처리 작업을 거쳤다.

실험에 활용한 구매 내역 데이터의 특성은 <Table 2>에 정리되어 있다.

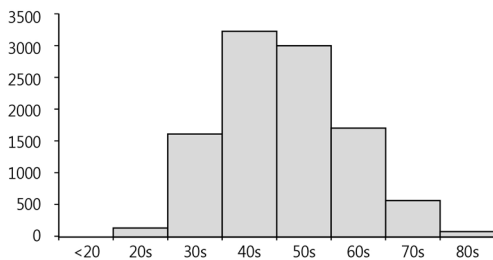
<Table 2> Customers' Purchasing Data

period	2015. 1~2015. 12
no. of transactions	106,593
no. of customers	10,393
no. of brands	459

전처리 이후 실험 고객의 인구통계학적 특성은 <Figure 3> 및 <Figure 4>에 표시되어 있다. 여성이 65.3%, 40대와 50대 전체 고객이 60.1%로 다수를 차지하고 있음을 알 수 있다.



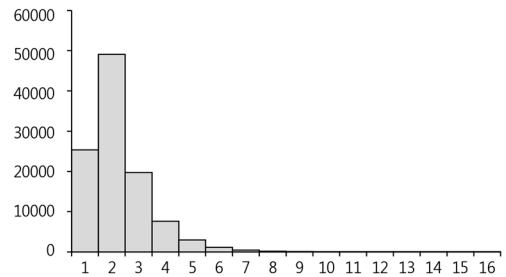
<Figure 3> Gender Ratio



<Figure 4> Age Distribution

<Figure 5>는 하나의 트랜잭션에 포함된 브랜드 개수의 분포를 보여주고 있다. <Figure 5>

에 따르면 전체 트랜잭션의 46.0%가 두개의 브랜드만을 포함하고 있어, 고객 중 약 절반은 일회 방문 시 두 개의 브랜드에서 상품을 구매하는 것으로 파악되었다. 각 트랜잭션에 포함된 브랜드 개수의 평균은 2.26개이며, 3개 이상의 브랜드를 포함하는 트랜잭션의 비중은 30.2%이다.



<Figure 5> Transaction Distribution According to the Number of Brands

한편, 일반적인 머신러닝 기법의 성능평가 방법에 따라 제안된 기법의 성능을 검증하기 위하여 전체 트랜잭션 데이터를 <Table 3>과 같이 학습집단(training set)과 테스트집단(test set)으로 구분하였다.

<Table 3> Training set vs. Test set

	training set	test set
period	2015. 1~2015. 11 (11 months)	2015. 12 (1 months)
no. of transactions	90,950 (85.3%)	15,643 (14.7%)

<Table 3>에서 보듯이, 2015년 1월부터 11월까지 11개월의 트랜잭션 데이터(약 85.3%)를 제안된 기법의 학습을 위해 사용하였으며, 이 결과를 2015년 12월 트랜잭션 데이터(약 14.7%)에 적용함으로써 제안된 기법의 성능을 평가하였다.

이러한 실험 환경은 사후적인 고객 구매 데이터를 통해 그 성능을 확인한다는 점에서 실제 추천

시스템이 운영되는 환경과는 다소 상이하다는 한계가 존재한다. 하지만 실제 추천 시스템의 구축 및 운영을 통한 데이터 수집이 어려운 현실을 감안할 때, 이전의 다수의 연구[Park et al., 2009; Xu et al., 2013; Im, 2016; Kim and Jeong, 2016, Lee and Kim]에서와 마찬가지로 본 연구에서도 유사한 방식으로 실험을 진행한다.

3.2 분석 방법

기존의 많은 연구들이 추천 시스템의 성능을 평가하기 위하여 정확도(Precision), 재현율(Recall) 및 F1 값을 사용해 왔다. 정확도는 추천 시스템이 추천한 아이템 중 실제 고객이 구매하는 아이템의 비율이며, 재현율은 고객이 구매한 아이템 중에서 추천 시스템의 추천이 적중시킨 비율이다.

한편, 정확도와 재현율은 시스템이 추천하는 아이템의 개수 k 에 영향을 받는다. 즉, k 가 많아지면 재현율은 올라가고 정확도는 떨어지게 된다. 따라서 최근의 추천시스템 연구에서는 주로 식 (3)과 같이 정확도와 재현율을 가중 평균한 F1 값을 추천 정확도의 평가방법으로 사용한다 [Im, 2016; Kim and Jeong, 2016; Lee and Kim, 2013].

$$F1 = \frac{Recall \times Precision}{(Recall + Precision)/2} \quad (3)$$

본 실험에서도 제안된 기법을 대상으로 정확도, 재현율 및 F1 값을 산출하여 그 성능을 살펴볼 것이며, 또한 성능의 비교평가를 위하여 기존의 IBCF 및 베스트셀러 기반 추천기법과 그 값을 비교해 본다.

한편, 테스트 수행에서 단지 한 개의 브랜드로 구성된 트랜잭션의 경우 추천의 효과를 측정하지 못하는 한계가 있기 때문에, 이러한 트랜잭션은 실험 대상에서 제외하였다. 또한, <Figure 5>

에서 보듯이 다수의 트랜잭션이 2개 이하의 브랜드만을 포함하고 있어, 본 실험에서는 매장에 방문한 고객에게 처음 방문한 브랜드를 기반으로 k 개의 브랜드를 추천하는 상황을 가정하였다. 즉 테스트 집단 내의 각 트랜잭션을 대상으로, 첫 구매가 발생하는 순간 추천을 수행하고, 추천된 브랜드와 이후 구매가 발생한 브랜드를 비교하여 정확도, 재현율 및 F1 값을 산출한다.

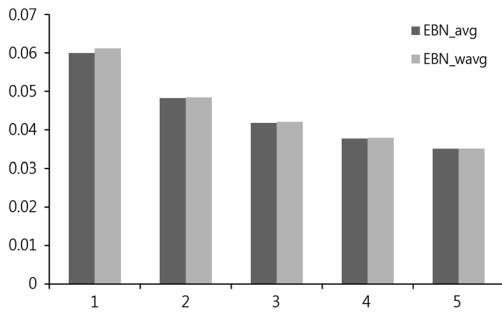
3.3 분석 결과

3.3.1 k 의 변화에 따른 추천 성능 비교

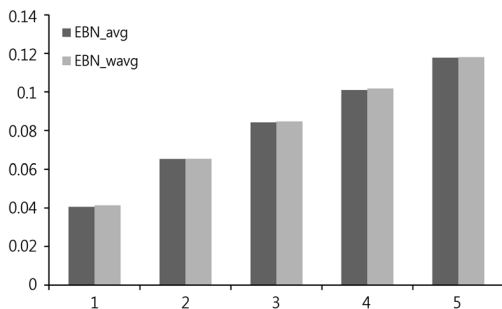
먼저 제안된 기법의 성능을 확인하기 위해 추천 브랜드의 개수 k 를 변화시키면서 성능을 평가하였다. 제안된 알고리즘은 앙상블 방법에 따라 다음과 같이 두 가지로 구분하여 그 성능을 확인하였다.

- EBN_avg : 기반 네트워크인 EBN_u 구성 시, CBN_u 와 GBN 의 유사도를 단순 평균 방식으로 결합
- EBN_wavg : 기반 네트워크인 EBN_u 구성 시, CBN_u 와 GBN 의 유사도를 각 방식의 F1 값을 가중치로 가중 평균하여 결합

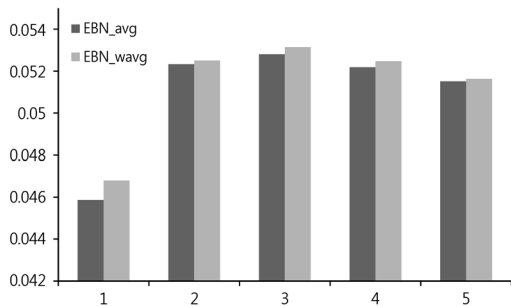
<Figure 6>~<Figure 8>에서 보듯이, 추천 브랜드의 개수 k 를 1에서 5까지 증가시키면서 실험한 결과, 모든 성능평가 지표에서 가중 평균 앙상블 방식인 EBN_wavg가 단순 평균 앙상블 방식인 EBN_avg보다 우수한 것을 확인할 수 있다. 이는 UBCF를 기반으로 앙상블 실험을 시도한 김민정, 조운호[2015]의 결과와 부합하는 것으로, EBN_u 구성 시 CBN_u 및 GBN 의 각 네트워크가 추천 성능에 기여하는 정도를 가중치로 반영하는 방법이 보다 뛰어난 성능을 보임을 의미한다.



<Figure 6> Precision for Various k



<Figure 7> Recall for Various k



<Figure 8> F1 value for Various k

<Figure 8>에서 보듯이, F1 값은 k 가 3일 때 EBN_wavg에서 0.053으로 가장 높음을 알 수 있다. 즉, 쇼핑 중인 고객에게 3개의 브랜드를 추천할 경우 그 추천 성능이 가장 뛰어난 것을 알 수 있다.

한편, 이러한 실험 결과는 본 연구에서 제안된 추천 기법의 성능이 기존의 온라인 쇼핑몰에서 제안된 기법들 보다는 다소 떨어짐을 보여준다. 온라인 쇼핑몰 대상의 김민정, 조윤희[2015]의

연구에서 F1 값은 0.04~0.16수준이며, 영화를 대상으로 진행된 이재식, 박석두[2007]의 연구에서는 본 연구와 유사한 0.05~0.06의 F1 값을 보인다.

이렇듯 제안된 기법의 성능이 온라인 대비 다소 낮게 나타나는 것은 오프라인 쇼핑의 환경적 요인에 기인하는 것으로 보인다. 즉, 온라인 쇼핑의 경우 쇼핑의 매개체인 인터넷 및 스마트폰의 화면이 갖는 제약으로 인해 동시에 제공되는 상품 및 상품 정보의 한계가 존재하며, 이로 인해 고객의 추천에 대한 의존도가 높다고 하겠다. 반면 오프라인 쇼핑에서는 각 상품에 대한 소비자의 직접적인 체험이 구매의사결정에 보다 큰 영향을 미치며, 추천이 미치는 영향이 제한적일 수 있음을 의미한다.

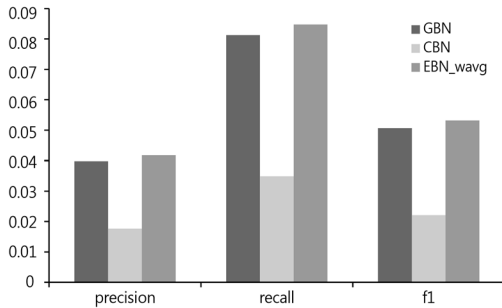
한편, 예상대로 k 가 커질수록 재현율을 키우고 정확도는 감소하는 것을 확인할 수 있다.

3.3.2 다른 기법과의 성능 비교

다음으로는 제안된 기법의 객관적인 성능 비교를 위하여 기존 추천 기법들과의 성능 차이를 확인해 보았다. 성능 비교에 사용된 기존 추천 기법들은 다음과 같다.

- *GBN* 기반 추천 : 일반적인 IBCF 방식으로, 고객 추천을 위한 기반 네트워크로 *GBN*을 사용
- *CBN* 기반 추천 : 대상 고객의 과거 구매 패턴만을 활용하여 추천 브랜드를 도출하는 방식으로, 기반 네트워크로 *CBN_u*를 사용
- Bestseller 기반 추천 : 학습기간 동안 가장 구매 빈도가 높은 브랜드들을 모든 고객에게 동일하게 추천하는 방식

본 비교를 위해 사용된 제안된 기법은 가중 평균 기법인 EBN_wavg이며, 추천 브랜드의 개수는 $k=3$ 을 사용하였고, 다른 기법들에서도 3개의 브랜드를 추출하였다.



<Figure 9> Performance Comparison among Various Recommendation Except for Bestseller Scheme

<Figure 9>에 따르면 제안된 기법의 정확도, 재현율, 그리고 F1 값은 *GBN* 및 *CBN* 기반 추천 방식보다 높음을 알 수 있다. F1 값을 기준으로, 제안된 기법은 *GBN* 기반 추천방식 대비 5.0%, *CBN* 기반 추천방식 대비 140% 우수한 성능을 보이는 것으로 확인되었다. 즉, IBCF 단독의 추천 방식이나 고객의 과거 패턴만을 기반으로 추천하는 방식 보다는 이를 앙상블한 기법이 보다 우수함을 의미한다고 하겠다.

또한, 기존의 IBCF 대비 제안된 기법의 우수성은 온라인 대비 오프라인 쇼핑에서 고객의 구매패턴은 지속적이고 반복적으로 나타날 것이라는 본 연구의 기본 가정이 틀리지 않았음을 보여주는 결과라고 하겠다.

한편, <Figure 9>에서는 가독성의 이유로 표현하지 않았지만, 베스트셀러 기법은 그 성능이 가장 떨어짐을 확인할 수 있었는데, F1 값이 0.00016으로 제안 기법의 0.3% 수준에 그쳐 그 성능의 한계를 확인할 수 있었다.

4. 결 론

본 연구에서는 오프라인 쇼핑물의 추천시스템 구현에 있어 IBCF 기법을 효과적으로 적용하는 방안을 제안하였다. 온라인 대비 오프라인 쇼핑은 시간적·공간적 제약으로 인해 고객의 구매 패턴

이 지속적이고 반복적으로 나타날 수 있다. 본 연구에서는 이러한 가정에 착안하여, 고객의 구매패턴을 모델링하는 방법과 이를 기존의 IBCF 기법에 앙상블 하여 추천에 활용하는 방안을 제안하였다.

제안된 기법은 크게 학습과정과 적용과정으로 구성된다. 학습과정에서는 과거 트랜잭션 데이터를 기반으로 일반적인 브랜드 네트워크(*GBN*) 및 고객의 구매패턴을 반영하는 브랜드 네트워크(*CBN_u*)를 구성하게 된다. 적용과정에서는 학습과정에서 구성된 브랜드 네트워크들을 앙상블하여 고객 맞춤형 브랜드 네트워크(*EBN_u*)를 구성하고, 당일 방문 고객의 브랜드 구매이력을 감안하여 추천 브랜드를 도출한다.

본 연구에서는 제안된 기법의 성능을 평가하기 위하여 국내 한 중형 백화점의 브랜드 구매이력 데이터를 확보하여 실험을 수행하였다. 실험 결과, 제안된 추천기법은 기존의 IBCF 및 베스트셀러기반 추천기법과 비교하여 높은 추천 정확도를 보였다.

한편, 본 연구는 다음과 같은 한계를 갖는다. 먼저, 본 연구에서 제안한 앙상블 기법은 오프라인 쇼핑물의 모든 고객에게 적용할 수 있는 것은 아니다. 신규 고객이나 구매 이력 파악이 불가능한 고객, 또는 구매 빈도가 매우 낮은 고객에게는 기존의 IBCF나 베스트셀러 기법 등의 추천방식이 보다 적합할 것이다. 즉, 동일한 오프라인 쇼핑물이라고 해도 추천 대상 고객의 특성에 맞게 여러 추천 기법들을 선별하여 활용하는 것이 바람직할 것이다.

두 번째, 본 연구에서는 대상 아이템으로 브랜드를 사용하였기 때문에, 상품을 대상으로 할 경우 본 기법의 성능을 확인하지 못하였다는 한계가 존재한다. 그러나 제안 알고리즘은 IBCF에 기반을 두고 있기 때문에 많은 상품과 고객이 존재하는 대형 쇼핑몰 등에서도 효과적으로 적용 가능할 것이라 기대할 수 있다.

References

- [1] Bar, A., Rokach, G., Shani, G., Shapira, B., and Schclar, A., "Improving simple collaborative filtering models using ensemble methods", *Multiple Classifier Systems*, 2013, pp. 1-12.
- [2] Choi, J. S., Kim, D. Y., Han, I. S., and Kim, Y. C., "A Study on Data Analysis in on/off-line shopping mall", *Electronic Commerce Research Association*, Vol. 12, No. 2, 2011, pp. 71-88.
- [3] Choi, S. J., "Have you arrived at Gang-nam Station? On flash sale, 80m in front of you", *Maeil Business Newspaper*, April 21, 2016.
- [4] Im, I., Recommendation system using R, 2nd edition, Chaosbook, 2016.
- [5] Kim, H. K., Kim, J. K., and Chen, Q. Y., "A Product Network Analysis for Extending the Market Basket Analysis", *Expert Systems with Applications*, Vol. 39, 2012, pp. 7403-7410.
- [6] Kim, M. and Y. Cho, "A Multimodal Profile Ensemble Approach to Development of Recommender System Using Big Data", *The Journal of Intelligence and Information Systems*, Vol. 21, No. 4, 2015, pp. 93-100.
- [7] Kim, N. K. and Jeong, S. B., "A Study on Customized Brand Recommendation based on Customer Behavior for Off-line Shopping Malls", *Journal of Information Technology Applications & Management*, Vol. 23, No. 4, 2016, pp. 55-70.
- [8] Lee, M. H., "Lotte Department Store, Strengthening Omni Channel, The main store Smart Beacon Service", *Korea Economic Daily*, December 7, 2014.
- [9] Lee, J. S. and Park, S. D., "Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering", *Journal of Intelligence and Information Systems*, Vol. 13, No. 4, 2007, pp. 65-78.
- [10] Lee, Y. and Kim, K.-J., "Product Recommender Systems using Multi-Model Ensemble Techniques", *Journal of Intelligence and Information Systems*, Vol. 19, No. 2, 2013, pp. 39-54.
- [11] Linden, G., Simth, B., and York, J., "Amazon.com recommendation : Item-to-item collaborative filtering", *IEEE Internet computing*, Vol. 7, No. 1, 2003, pp. 76-80.
- [12] Liu et al., "Overview of the Evaluated Algorithms for the Personal Recommendation Systems", *Complex Systems and Complex Science*, Vol. 6, 2009, pp. 1-3.
- [13] Nicolaus Mote, "The New School of Ontologies[J]", *CSC I 585*, Nov. 30, 2004.
- [14] Park, J. H., Cho, Y. H., and Kim, J.K., "Social Network : A Novel Approach to New Customer Recommendations", *Journal of Intelligent Information Systems*, Vol. 15, No. 1, 2009, pp. 123-140.
- [15] Park, W. B., Cho. Y. S., and Ko, H. H., "Clustering Method of Weighted Preference Using K-means Algorithm and Bayesian Network for Recommender System", *Journal of Information Technology Applications & Management Special Section*, 2013, pp. 219-230.
- [16] Piotte, M. and Chabbert, M., "The Pramatic theory solution to the Netflix grand prize,"

Netflix prize documentation, 2009.

- [17] Radeder, T. and Chawla, N. V., “Modeling a store’s product space as a social network”, *Proceeding of the international conference on advances in social network ana-*

lysis and mining, Athens Greece, 2009, pp. 164-169.

- [18] Xu, Q., Zheng, S., and Cai, M., “IBCF Improved Algorithm Based on the Tag”, *ICCNCE 2013*, 2013, pp. 265-268.

■ 저자소개



Seok Bong Jeong

Seok Bong Jeong is currently a Professor of Dept. of Railway, Kyungil University. He received the B.S. degree in Industrial Management and the M.S. and Ph.D. degree in Industrial Engineering from KAIST. He worked for SAMSUNG SDS and Ernst & Young as a IT consultant and participated in several IT Projects. His research interests are in the areas of Big Data Analysis, Business Intelligence, and Machine Learning.