ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2020 June: 26(2): 27~42 http://dx.doi.org/10.13088/jiis.2020.26.2.027

# 카테고리 연관 규칙 마이닝을 활용한 추천 정확도 향상 기법

이동원
한성대학교 사회과학부
(*dongwonlee@hansung.ac.kr*)

인터넷이라는 가상 공간을 활용함으로써 물리적 공간의 제약을 갖는 오프라인 쇼핑의 한계를 넘어선 온라인 쇼핑은 다양한 기호를 가진 소비자를 만족시킬 수 있는 수많은 상품을 진열할 수 있게 되었다. 그러나, 이는 역설적으로 소비자가 구매의사결정 과정에서 너무 많은 대안을 비교 평가해야 하는 어려움을 겪게 함으로써 오히려 상품 선택을 방해하는 원인이 되기도 한다. 이런 부작용을 해소하기 위한 노력으로서, 연관 상품 추천은 수많은 상품을 다루는 온라인 상거래에서 소비자의 구매의사결정 과정 중 정보탐색 및 대안평가에 소요되는 시간과 노력을 줄여주고 이탈을 방지하며 판매자의 매출 증대에 기여할 수 있다. 연관 상품 추천에 사용되는 연관 규칙 마이닝 기법은 통계적 방법을 통해 주문과 같은 거래 데이터로부터 서로 연관성 높은 상품을 효과적으로 발견할 수 있다. 하지만, 이 기법은 거래 건수를 기반으로 하므로, 잠재적으로 판매 가능성이 높을지라도 충분한 거래 건수가 확보되지 못한 상품은 추천 목록에서 누락될 수 있다. 이렇게 추천 시 제외된 상품은 소비자에게 구매될 수 있는 충분한 기회를 확보하지 못할 수 있으며, 또 다시 다른 상품에 비해 상대적으로 낮은 추천 기회를 얻는 악순환을 겪을 수도 있다. 본 연구는 구매의사결정이 결국 상품이 지닌 속성에 대한 사용자의 평가를 기반으로 한다는 점에 착안하여, 추천 시 상품의 속성을 반영하면 소비자가 특정 상품을 선택할 확률을 좀 더 정확하게 예측할 수 있다는 점을 추천 시스템에 반영하기 위한 목적으로 수행되었다. 즉, 어떤 상품 페이지를 방문한 소비자는 그 상품이 지닌 속성들에 어느 정도 관심을 보인 것이며 추천 시스템은 이런 속성들을 기반으로 연관성을 지닌 상품을 더 정교하게 찾을 수 있다는 것이다. 상품의 주요 속성의 하나로서, 카테고리는 두 상품 간에 아직 드러나지 않은 잠재적인 연관성을 찾기에 적합한 대상이 될 수 있다고 판단하였다. 본 연구는 연관 상품 추천에 상품 간의 연관성뿐만 아니라 카테고리 간의 연관성을 추가로 반영함으로써 추천의 정확도를 높일 수 있는 예측모형을 개발하였고, 온라인 쇼핑몰로부터 수집된 주문 데이터를 활용하여 이루어진 실험은 기존 모형에 비해 추천 성능이 개선됨을 보였다. 실무적인 관점에서 볼 때, 본 연구는 소비자의 구매 만족도를 향상시키고 판매자의 매출을 증가시키는 데에 기여할 수 있을 것으로 기대된다.

**주제어** : 추천 시스템, 추천 정확도 향상, 연관 규칙 마이닝, 카테고리 연관 규칙 마이닝

논문접수일 : 2020년 4월 4일 논문수정일 : 2020년 5월 11일 게재확정일 : 2020년 5월 26일
원고유형 : 일반논문 교신저자 : 이동원

## 1. 서론

오프라인 매장은 물리적인 공간을 확보하기 위해 많은 비용이 발생하는 반면, 온라인 매장은 웹 페이지라는 가상의 공간을 활용함으로써 상품 당 전시에 필요한 추가비용이 거의 소요되지 않아 무한에 가까운 상품을 진열하는 것이 가능하게 되었다. 이는 구매자에게는 다양한 상품을

* 본 연구는 한성대학교 교내학술연구비 지원과제 임(This research was financially supported by Hansung University.)

구매할 수 있게 됨으로써 자신의 기호에 좀 더 적합한 상품을 구매할 수 있는 기회를 제공한다. 그러나, 오히려 이는 역설적으로 원하는 상품을 더 찾기 힘들어지는 결과를 야기하는데, 이는 너무 많은 대안이 오히려 구매의사결정에 장애가 될 수 있기 때문이다(Chernev et al., 2015; Scheibehenne et al., 2010). 이런 이유로, 온라인 판매자는 구매자에게 검색, 추천 등의 편의 기능을 제공함으로써 구매자가 고려해야 할 대안을 찾고 비교하는 노력을 줄여주는 기능을 제공하고 있다.

이와 같이 추천 시스템은 온라인 상거래에서 구매자가 선택할 수 있는 상품의 대안을 줄여주는 역할을 수행하게 되는데(Aljukhadar et al., 2012), 그 중에서 연관 상품 추천 기능은 이미 구매된 상품 간의 연관성을 기반으로 하여 구매자에게 선택될 가능성이 높은 상품을 제안하는 기능을 수행한다. 이런 추천 방식은 상품의 상세 페이지 내에 관련 상품의 목록을 함께 보여줌으로써 구매자가 상품의 대안을 쉽게 비교할 수 있게 해줄 뿐만 아니라 미처 인지하지 못했던 니즈를 만족시켜줄 수 있는 상품을 발견할 수도 있게 해준다. 추천 상품 목록에서 관심상품이 적절히 제시되면 해당 상품 페이지의 방문 및 구매가 증가되기 때문에 판매자의 입장에서는 제한된 수의 목록에 좀 더 구매자의 관심을 끌어 선택될 수 있는 상품을 선별하여 배치하는 것이 매우 중요하다.

추천 목록에 배치될 연관 상품의 선정을 위해서 널리 활용되고 있는 방법은 연관 규칙 마이닝(Agrawal, 1993)으로서, 지지도, 신뢰도, 향상도라는 세 가지 척도를 상품 간의 연관성을 평가하는 방법으로 제시한다. 이는 연관 규칙이라는 형태로서 아래의 예와 같이 표현된다.

연관규칙1: 바지1 → 바지2
(지지도 2.0%, 신뢰도 35.5%)
연관규칙2: 바지1 → 바지3
(지지도 1.5%, 신뢰도 36.5%)

위의 예시에 표시된 두 개의 연관규칙은 '바지1'이라는 상품에 대해 연관성이 높은 상품 중 '바지2'와 '바지3'이 있음을 보여주고 있으며, 그 연관성의 크기를 지지도와 신뢰도라는 척도로 표시하고 있다. 지지도는 전체 구매자 중 '바지1'과 '바지2'를 모두 구매하는 비율을 의미하며, 신뢰도는 '바지1'을 구매한 고객 중 '바지2'를 구매하는 고객의 비율로 계산된다. 여기서, 지지도는 두 상품 위의 규칙을 적용할 대상의 규모를, 신뢰도는 '바지1'에 관심을 보이는 고객 중 '바지2'에 대한 판매 기회의 크기를 보여준다는 측면에서 모두 중요한 척도로 간주되며, 추천 상품 선정 시에 중요한 지표로 활용된다.

하지만, 위의 두 척도는 두 상품의 연관성을 서로 다른 관점으로 통계적으로 표현하는 반면, 소비자로부터 얼마나 많은 선택을 받을 것인가를 평가할 수 있는 일관된 척도로 사용되기에는 한계가 있다. 즉, 지지도가 상대적으로 높고 신뢰도가 상대적으로 낮은 경우(연관규칙1)와 지지도가 상대적으로 낮고 신뢰도가 상대적으로 높은 경우(연관규칙2) 중 어느 것이 추천 시 더 많은 소비자의 선택을 받을 것인지를 직접 비교하는 것이 불가하다. 이런 이유로 두 척도를 적용한 예측모형을 활용함으로써 추천의 성공확률을 높이는 방안이 연구된 바 있다(Lee, 2017b). 본 연구에서는, 이와 같은 통계적 방법을 적용하되 그 정확도를 높일 수 있는 추가적인 방안을 제시하고자 수행되었다.

연관규칙은 통계를 기반으로 이미 발생한 상

품 간의 연관성을 잘 표현할 수 있는 반면, 잠재적인 연관성을 더 찾아내는 데에는 한계를 갖는다. 앞서 언급된 두 개의 연관규칙과 함께 아래의 연관규칙을 추가로 고려해보도록 한다.

연관규칙3: 바지1 → 바지4
(지지도 1.5%, 신뢰도 35.5%)

위의 연관규칙은 연관규칙1, 연관규칙2에 비해 각각 낮은 지지도, 낮은 신뢰도를 갖는다는 점으로 인해 두 규칙보다 중요도가 낮게 평가될 수 있다. 연관규칙1, 2의 추천대상 상품, 즉, '바지2'와 '바지3'이 면바지인 반면, '바지1'과 '바지4'는 청바지였다고 가정해보자. 그리고, 청바지와 면바지에 대해 상품 카테고리 수준에서 발견된 연관규칙이 다음과 같다고 하자.

연관규칙4: 청바지 → 청바지
(지지도 5%, 신뢰도 50%)
연관규칙5: 면바지 → 청바지
(지지도 3%, 신뢰도 30%)

즉, 연관규칙으로 나타난 결과는 청바지와 청바지 간의 연관성이 면바지와 청바지 간의 연관성보다 높다는 것을 보여준다. 즉, 면바지에 관심을 보인 소비자보다 청바지에 관심을 보인 소비자가 청바지에 관심을 보이는 확률이 더 높다는 것으로 해석될 수 있다. 이는 청바지나 면바지에 속한 개별 상품인 바지들 간의 연관성에 이미 반영되었을 수 있다는 측면에서, 카테고리 간의 연관성보다 더 강한 개별 상품 간의 연관성으로 인해 발생한 예로 해석될 수도 있다.

그러나, 노출기회를 많이 얻지 못했거나, 이로 인해 구매 건수가 많이 발생하지 않은 상품의 경우에는 이런 척도의 값이 실제 판매가능성을 충분히 표현하지 못할 수도 있다. 예를 들어, '바지2'와 '바지3'이 '바지4'에 비해 출시가 빨라 좀 더 많이 거래된 경우, '바지4'의 누적 구매 건수가 다른 두 상품에 비해 상대적으로 낮아 지지도와 신뢰도 또한 낮게 평가되었을 수 있다. 본 연구에서는 이렇게 아직은 발현되지 않은 상품 간의 잠재적인 연관성을 찾아내고 이를 이미 발현된 연관성에 추가로 고려함으로써 추천의 적중률을 높일 수 있는 방안을 제시하고자 수행되었다.

## 2. 이론적 배경

### 2.1 추천 시스템

추천 시스템은, 여러 사용자로부터 개별 상품에 대한 선호도를 수집하고, 이를 기반으로 서로 다른 상품 간의 유사도나 서로 다른 사용자 간의 유사도를 계산하여, 이를 기반으로 특정 사용자가 특정 상품에 대해 갖게 될 선호도를 예측하는 기능을 수행한다. 구매의사결정의 관점에서 보면 추천 시스템은 사용자가 원하는 상품을 찾기 위한 탐색 노력을 줄여주는 역할을 수행한다. 또한, 기업은 고객의 이탈을 방지하고 충성도를 높여 고객과의 관계를 유지하고 매출을 증대하는 효과를 얻을 수 있다(Ansari et al., 2000). 추천 기법의 성능을 높이기 위한 추천 기법을 개발하거나(Balabanovic and Shoham, 1997; Ansari et al., 2000; Adomavicius and Tuzhilin, 2011; Choi et al., 2016), 추천 시스템의 성과를 측정하며(Bodapati, 2008; Fleder and Hosanagar, 2009), 상거래 이외의 다양한 분야에 적용하기(Choi et al.,

2015; Kim and Lee, 2013; Kim et al., 2010) 위한 목적으로 다양한 연구들이 활발히 수행되고 있다.

추천 성능의 향상을 위한 추천 기법에 관한 연구는 내용기반 필터링(Content-based Filtering)과 협업 필터링(Collaborative Filtering)으로 크게 구분된다. 내용기반 필터링 기법은 상품 간의 유사성을 기반으로 적합한 추천 상품을 찾기 위해 수행되며, 의사결정나무, 최근접 이웃 기법, 사회연결망 분석 등 다양한 분류 기법을 활용한다(Konstan et al., 1997; Ansari et al., 2000; Kim and Kim, 2014; Kim and Kim, 2016; Lee, 2017b; Shin et al., 2012). 반면, 협업 필터링은 사용자 간의 유사성을 기반으로 추천 상품을 찾는 방식으로서, 추천 대상이 되는 사용자와 유사한 사용자가 선호한 상품 중에서 선호도가 높을 것으로 기대되는 상품을 찾는 기법이다(Konstan et al., 1997; Ansari et al., 2000).

## 2.2 연관 규칙 마이닝

연관 규칙 마이닝 기법(Agrawal et al., 1993)은 주문과 같은 거래 데이터에서 반복적으로 동시에 출현하는 상품을 찾아 이들 간의 패턴을 연관규칙이라는 형태로 표현하는 방법이다. 이런 규칙은 그 안에 포함된 상품 간의 빈도를 기반으로 계산된 연관성 척도를 함께 표현한다. 함께 주문되는 빈도가 높은 상품들이 있을 때, 하나의 상품이 구매될 가능성이 높은 상황에서 다른 상품의 구매 가능성도 높을 것으로 예측할 수 있다는 점에서 상품 판매를 촉진하기 위해 온라인 거래가 이루어지는 사이트에서 널리 활용되고 있다(Anand, 1998; Chen et al., 2006; Kim and Street, 2004; Lee et al., 2013; Kim and Kim, 2005). 연관규칙은 아래와 같이 정형화된 형태를 갖는데, 여기서 A와 C는 각각 선행 상품(Antecedent Item)과 후행 상품(Consequent Item)이며, 이들의 연관성을 보여주는 척도인 지지도(Support)와 신뢰도(Confidence)는 각각 sup와 conf로 표현된다.

$$A \rightarrow C\ (\mathrm{sup}\%, \mathrm{conf}\%)$$

여기서, 지지도는 선행 상품과 후행 상품이 동시에 나타난 거래 건수가 전체 거래의 건수 중 차지하는 비율로 계산되며, 신뢰도는 선행 상품과 후행 상품이 함께 포함된 거래 건수가 선행 상품이 나타난 거래 건수 중 차지하는 비율로 계산된다. 높은 신뢰도는 선행 상품을 구매했거나 이에 관심을 보인 소비자가 후행 상품에 대해서도 흥미를 가질 가능성이 높다는 것을, 높은 지지도는 두 상품이 함께 거래되는 경우가 우연히 발생하지 않았고 두 상품에 모두 관심을 보이는 소비자가 많다는 것을 암시한다고 할 수 있다. 따라서, 현업에서는 이런 연관 규칙을 찾아 선행 상품의 상세 페이지에 후행 상품을 추천 상품으로 노출함으로써 매출을 높이기 위해 노력하고 있다.

선행 상품과 후행 상품이 함께 포함된 거래가 충분히 발생하지 않아 두 상품 간의 지지도가 낮은 경우 두 상품 간의 연관성은 낮게 평가될 수 있다. 즉, 상품이 판매되기 시작한 시점에서는 노출기회가 적음으로 인해 충분한 거래 기록이 쌓이지 않고 이는 낮은 지지도로 이어져 다시 노출기회가 적어지는 악순환에 빠질 수 있다. 현업에서는 이런 상품들의 판매 기회를 높이기 위해 신규 상품을 소비자에게 적극적으로 노출시키는 등의 방법을 통해 노력하고 있다. 그러나, 담당자의 직관에 의해 신규 상품을 임의로 선택하는

방법을 취하는 경우 잠재적인 판매 가능성에 따른 차별적 마케팅이 효율적으로 이루어지기 쉽지 않다.

## 3. 연구 모형

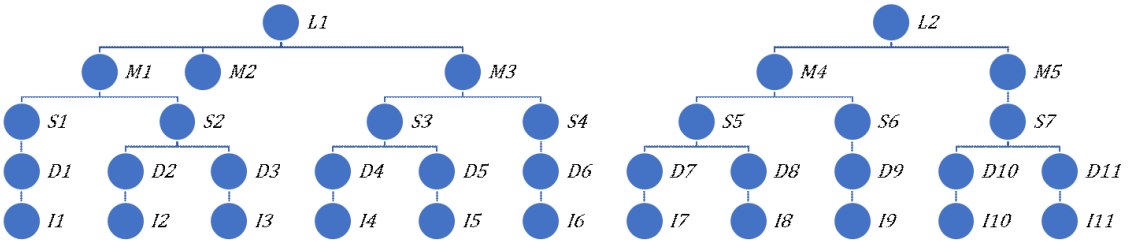
연관상품의 추천목록은 상품 페이지의 특정 추천공간에 배치되며 제한된 수의 상품만을 보여줄 수 있으므로, 이에 포함될 상품의 선정과 나열순서는 추천의 적중률을 결정하는 데에 매우 중요한 역할을 한다. 일반적으로 선행상품 하나 당 10개 안팎의 후행상품이 목록에 포함되는데, 본 연구에서는 성능을 충분히 검토하기 위해 후행상품 20개를 추출하기로 한다. 이들은 추천에 대한 적중률이 높을 것으로 예측되는 순서대로 나열되는데, 적중률을 예측하기 위한 방안으로 회귀모형을 적용하기로 한다. 즉, 연관규칙 마이닝 기법에서 전통적으로 사용되는 지지도와 신뢰도를 독립변수로, 이에 따른 적중률을 종속변수로 삼아 통계분석을 실시하고, 이를 기반으로 적중률을 예측하는 모형을 개발한다.

본 연구에서는 상품 수준의 연관성에 이들이 속한 카테고리 수준의 연관성을 결합함으로써 추천의 적중률을 높이기 위한 목적으로 수행된다. 따라서, 추천 적중률의 예측모형에는 상품 수준에서 계산된 지지도와 신뢰도뿐만 아니라, 카테고리 수준의 지지도와 신뢰도를 함께 독립변수에 포함시킨다. <Figure 1>은 개별 상품의 카테고리 구조와 이에 포함된 상품의 예를 보여준다. 여기서, 카테고리는 계층형 구조를 갖는데, 각 상품에 대해 대분류, 중분류, 소분류, 세분류의 4 계층으로 구성되어 있다.

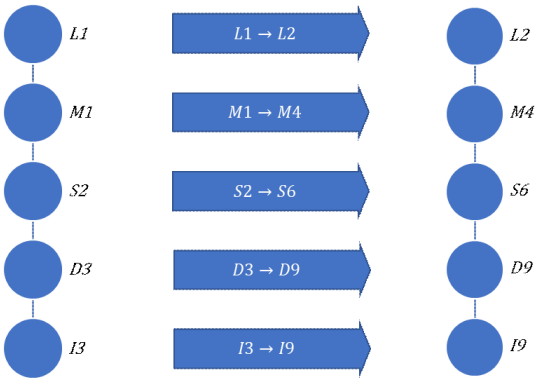
본 연구에 사용된 주문 데이터를 기반으로 상품 수준을 포함하여 4개의 카테고리 계층에 대해 연관규칙을 생성하고 지지도와 신뢰도를 계산하도록 한다(<Figure 2>). 예를 들어, 상품 I3와 I9에 대한 연관규칙은 $I3 \rightarrow I9$과 같이 표현되며, 이 두 상품이 속한 카테고리의 각 계층으로부터 $L1 \rightarrow L2$, $M1 \rightarrow M4$, $S2 \rightarrow S6$, $D3 \rightarrow D9$과 같은 연관규칙을 얻을 수 있다. 이를 기반으로 작성된 회귀모형은 아래의 식과 같다.

$$\begin{aligned} HitRate = {} & \beta_0 + \beta_1 Conf_I + \beta_2 Conf_L + \beta_3 Conf_M \\ & + \beta_4 Conf_S + \beta_5 Conf_D + \beta_6 Sup_I + \beta_7 Sup_L \\ & + \beta_8 Sup_M + \beta_9 Sup_S + \beta_{10} Sup_D + \beta_{11} Sup_{I1} \\ & + \beta_{12} Sup_{L(A)} + \beta_{13} Sup_{M(A)} + \beta_{14} Sup_{S(A)} \\ & + \beta_{15} Sup_{D(A)} + \varepsilon \end{aligned} \quad (1)$$

여기서, $Conf_X$는 각각 X 수준에서의 신뢰도를 뜻한다. 즉, $Conf_I$는 상품(Item), $Conf_L$은 대분류(Large Group), $Conf_M$는 중분류(Medium Group),




〈Figure 1〉 Hierarchy of Items




〈Figure 2〉 Association Rules between Items and their Hierarchical Categories

$Conf_S$는 소분류(Small Group), $Conf_D$는 세분류(Detail Group)에 대한 신뢰도이다. 마찬가지로, $Sup_X$도 각각 X 수준에서의 지지도를 의미하며, 신뢰도와 마찬가지로 대분류, 중분류, 소분류, 세분류에 대해 계산된 값이 적용된다. 지지도는 전체 거래에 대한 비율(=두 상품의 동시 구매 건수/전체 거래 건수)로 정의되나, 이는 모든 연관규칙에 대해 상대적인 차이를 발생시키지 않으므로 편의상 두 상품의 동시 구매 건수로 계산하기로 한다. $Sup_{X(A)}$(X=I, L, M, S, D)은 연관규칙의 지지도와 신뢰도에 영향을 미치는 통제변수로서 모형에 추가되는데, 연관규칙 상의 선행 상품(또는 카테고리)의 지지도를 의미한다. 전체 거래건수에 대한 해당 상품(또는 카테고리)이 포함된 거래건수의 비율로 계산되나, 모든 연관규칙에 동일하게 적용되므로 본 모형에서는 편의상 분모를 생략하고 계산하기로 한다.

아래는 제안모형의 성능을 비교하기 위한 기존모형의 회귀식으로서, 제안모형과는 달리 상품(Item) 수준의 지지도와 신뢰도만으로 독립변수가 구성된다.

$$HitRate = \beta_0 + \beta_1 Conf_I + \beta_2 Sup_I + \beta_3 Sup_{I(A)} + \varepsilon \quad (2)$$

## 4. 실험

본 연구에서 제안한 모형이 기존의 모형에 비해 높은 성능을 보이는지 검증하기 위한 목적으로 실험을 실시한다. 앞서 개발한 두 개의 회귀모형을 예측모형으로 활용하여 성능을 서로 비교하는 방식으로 진행한다. 전체 과정은 다음과 같다. 먼저, 각 모형 별로 학습용 데이터로부터 지지도와 신뢰도를 포함한 연관규칙을 추출한 후, 각 연관규칙에 대해 예상 추천 적중률을 계산한다. 이를 기반으로, 각 선행상품에 대해 예상 적중률이 높은 순서대로 N (1≤N≤20) 개의 후행상품을 선정하는 방식으로 추천 목록을 작성한다.

### 4.1 데이터

본 연구에서는 국내의 온라인 상거래 기업으로부터 2015년 5월부터 2016년 4월까지 12개월 동안 수집된 130만 건의 주문 거래 데이터를 분석한다. 고객을 기준으로 모형의 학습을 위한 데이터와 검증을 위한 데이터로 80:20의 비율로 구분하고, 학습용 데이터 중 50%는 독립변수의 생성을 위해, 나머지 50%는 종속변수의 생성을 위해 사용한다. 즉, 전체 데이터의 40%로부터 지지도와 신뢰도를 계산하여 연관규칙을 생성하고, 다른 40%의 데이터에서 적중률을 계산하여 회귀모형을 작성한다. 각각의 데이터는 동일 고객에 대해 선후행 상품을 연관짓는 방법으로 생성되며, 학습용 283만건과 검증용 234만건으로 구성된다. 이렇게 작성된 회귀모형을 예측모형으

로 활용하여 나머지 20%에 적용함으로써 모형의 성능을 평가하도록 한다. 성능의 비교를 위해서 제안모형과 경쟁모형 각각에 대해 위의 과정을 실시하도록 한다.

실험에 사용하기 위한 두 모형의 학습에 사용된 변수의 기술통계량은 <Table 1>에 나타나 있다.

〈Table 1〉 Descriptive Statistics

| Variable | Obs. | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| HitRate | 283,585 | 0.006915 | 0.057416 | 0.000000 | 1.000000 |
| $Conf_I$ | 283,585 | 0.071890 | 0.169246 | 0.000259 | 1.000000 |
| $Conf_L$ | 283,585 | 0.067744 | 0.041437 | 0.000873 | 0.195244 |
| $Conf_M$ | 283,585 | 0.033994 | 0.033332 | 0.000024 | 0.458763 |
| $Conf_S$ | 283,585 | 0.019389 | 0.033143 | 0.000029 | 1.000000 |
| $Conf_D$ | 283,585 | 0.013393 | 0.032955 | 0.000031 | 1.000000 |
| $Sup_I$ | 283,585 | 1.211224 | 1.449460 | 1 | 227 |
| $Sup_L$ | 283,585 | 3743.185000 | 3552.565000 | 5 | 12,813 |
| $Sup_M$ | 283,585 | 697.276700 | 874.095300 | 1 | 3,807 |
| $Sup_S$ | 283,585 | 170.235700 | 417.931000 | 1 | 3,387 |
| $Sup_D$ | 283,585 | 64.586020 | 169.990700 | 1 | 2,415 |
| $Sup_{I(A)}$ | 283,585 | 471.153800 | 724.214900 | 1 | 3,865 |
| $Sup_{L(A)}$ | 283,585 | 55070.770000 | 31459.040000 | 799 | 100,660 |
| $Sup_{M(A)}$ | 283,585 | 21420.460000 | 14319.980000 | 6 | 40,912 |
| $Sup_{S(A)}$ | 283,585 | 9748.760000 | 9925.264000 | 1 | 34,344 |

〈Table 2〉 Correlation Coefficients

| | $Conf_I$ | $Conf_L$ | $Conf_M$ | $Conf_S$ | $Conf_D$ | $Sup_I$ | $Sup_L$ | $Sup_M$ | $Sup_S$ | $Sup_D$ | $Sup_{I1}$ | $Sup_{L1}$ | $Sup_{M1}$ | $Sup_{S1}$ | $Sup_{D1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Conf_I$ | 1.0000 | | | | | | | | | | | | | | |
| $Conf_L$ | 0.0618 | 1.0000 | | | | | | | | | | | | | |
| $Conf_M$ | 0.1125 | 0.5950 | 1.0000 | | | | | | | | | | | | |
| $Conf_S$ | 0.1647 | 0.3209 | 0.7001 | 1.0000 | | | | | | | | | | | |
| $Conf_D$ | 0.2107 | 0.2288 | 0.4946 | 0.7027 | 1.0000 | | | | | | | | | | |
| $Sup_I$ | -0.0374 | 0.0684 | 0.0898 | 0.0797 | 0.0673 | 1.0000 | | | | | | | | | |
| $Sup_L$ | -0.0336 | 0.6689 | 0.3590 | 0.0939 | 0.0459 | 0.0594 | 1.0000 | | | | | | | | |
| $Sup_M$ | -0.0205 | 0.4407 | 0.5779 | 0.3154 | 0.1421 | 0.0979 | 0.5266 | 1.0000 | | | | | | | |
| $Sup_S$ | -0.0025 | 0.2279 | 0.3747 | 0.4117 | 0.1675 | 0.0825 | 0.1301 | 0.5693 | 1.0000 | | | | | | |
| $Sup_D$ | -0.0164 | 0.1789 | 0.2901 | 0.3022 | 0.2541 | 0.1298 | 0.1184 | 0.3615 | 0.5415 | 1.0000 | | | | | |
| $Sup_{I(A)}$ | -0.2665 | -0.0584 | -0.0901 | -0.1018 | -0.1025 | 0.1452 | 0.0660 | 0.0744 | 0.0408 | 0.0586 | 1.0000 | | | | |
| $Sup_{L(A)}$ | -0.1059 | 0.0096 | -0.0122 | -0.0915 | -0.0812 | 0.0227 | 0.6392 | 0.2833 | -0.0040 | 0.0109 | 0.1747 | 1.0000 | | | |
| $Sup_{M(A)}$ | -0.1204 | -0.0231 | -0.0647 | -0.0722 | -0.0813 | 0.0421 | 0.3009 | 0.5430 | 0.2518 | 0.1582 | 0.2467 | 0.4912 | 1.0000 | | |
| $Sup_{S(A)}$ | -0.1004 | -0.0330 | -0.0483 | -0.0571 | -0.0809 | 0.0412 | 0.0414 | 0.3147 | 0.4640 | 0.2736 | 0.2293 | 0.0962 | 0.6119 | 1.0000 | |
| $Sup_{D(A)}$ | -0.1238 | -0.0209 | -0.0466 | -0.0730 | -0.1103 | 0.0430 | 0.0374 | 0.1955 | 0.2525 | 0.4175 | 0.2826 | 0.0804 | 0.4071 | 0.6056 | 1.0000 |

### 4.2 예측 모형 개발

성능의 비교를 위해 두 개의 예측 모형을 개발한다. 학습용 데이터로부터 두 개의 회귀모형을 개발한 결과가 <Table 3>과 <Table 4>에 나타나 있다. <Table 3>은 본 연구에서 제안한 바와 같이 상품 간의 연관 규칙뿐만 아니라 그 상품이 포함된 카테고리 간의 연관 규칙으로부터 계산된 지지도와 신뢰도를 독립변수로 추천 적중률을 예측하는 회귀모형(<식1>)을 분석한 결과이며, <Table 4>는 이 중에서 카테고리 간의 연관 규칙은 제외하고 상품 간의 연관 규칙만을 적용한 분석결과이다.

〈Table 3〉 Regression Analysis of Training Data for Proposed Model

| HitRate | Coef. | Std. Err. | t | P>t | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| $Conf_I$ | 0.05871330 | 0.00065120 | 90.16000000 | 0.00000000 | 0.05743700 | 0.05998960 |
| $Conf_L$ | 0.06819340 | 0.00553230 | 12.33000000 | 0.00000000 | 0.05735030 | 0.07903640 |
| $Conf_M$ | 0.00111180 | 0.00662430 | 0.17000000 | 0.86700000 | -0.01187160 | 0.01409520 |
| $Conf_S$ | 0.03523120 | 0.00595320 | 5.92000000 | 0.00000000 | 0.02356310 | 0.04689920 |
| $Conf_D$ | 0.21755440 | 0.00470090 | 46.28000000 | 0.00000000 | 0.20834080 | 0.22676810 |
| $Sup_I$ | 0.00372190 | 0.00007330 | 50.75000000 | 0.00000000 | 0.00357820 | 0.00386560 |
| $Sup_L$ | -0.00000040 | 0.00000008 | -4.97000000 | 0.00000000 | -0.00000056 | -0.00000024 |
| $Sup_M$ | -0.00000036 | 0.00000027 | -1.37000000 | 0.17100000 | -0.00000089 | 0.00000016 |
| $Sup_S$ | -0.00000295 | 0.00000042 | -6.97000000 | 0.00000000 | -0.00000378 | -0.00000212 |
| $Sup_D$ | -0.00000425 | 0.00000083 | -5.12000000 | 0.00000000 | -0.00000587 | -0.00000262 |
| $Sup_{I1}$ | -0.00000161 | 0.00000016 | -10.19000000 | 0.00000000 | -0.00000193 | -0.00000130 |
| $Sup_{L1}$ | 0.00000007 | 0.00000001 | 10.47000000 | 0.00000000 | 0.00000006 | 0.00000009 |
| $Sup_{M1}$ | 0.00000001 | 0.00000001 | 0.58000000 | 0.56300000 | -0.00000002 | 0.00000004 |
| $Sup_{S1}$ | 0.00000014 | 0.00000002 | 7.92000000 | 0.00000000 | 0.00000011 | 0.00000018 |
| $Sup_{D1}$ | 0.00000011 | 0.00000002 | 5.53000000 | 0.00000000 | 0.00000007 | 0.00000014 |
| Constant | -0.01311290 | 0.00043670 | -30.03000000 | 0.00000000 | -0.01396880 | -0.01225710 |

〈Table 4〉 Regression Analysis of Training Data for Competent Model

| HitRate | Coef. | Std. Err. | t | P>t | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| $Conf_I$ | 0.06751060 | 0.00064390 | 104.85000000 | 0.00000000 | 0.06624860 | 0.06877260 |
| $Sup_I$ | 0.00417680 | 0.00007320 | 57.03000000 | 0.00000000 | 0.00403330 | 0.00432040 |
| $Sup_{I1}$ | -0.00000154 | 0.00000015 | -10.11000000 | 0.00000000 | -0.00000183 | -0.00000124 |
| Constant | -0.00227330 | 0.00016160 | -14.06000000 | 0.00000000 | -0.00259010 | -0.00195650 |

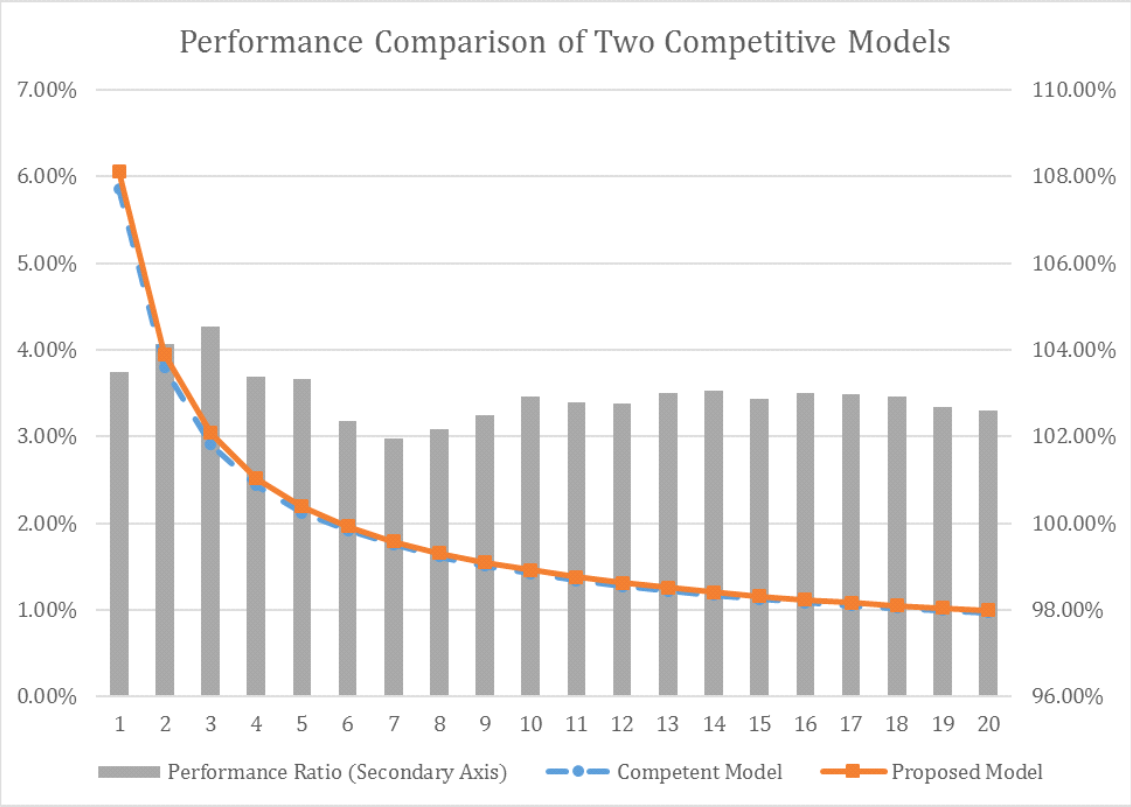
### 4.3 성능 평가

검증 데이터를 사용하여 두 개의 모형에 대한 성능을 비교한 결과는 <Table 5>와 <Figure 3>과 같다. <Table 5>는 제안 모형, 경쟁 모형으로부터 각각의 선행상품에 대해 상위 N ($1 \leq N \leq 20$) 개의 연관규칙을 적용한 결과이다. 즉, 연관 규칙의 선행 상품에 대한 상품 상세 페이지에 1개의 추천 상품이 노출되는 경우 경쟁 모형(5.86%)에 비해 제안 모형(6.06%)이 제시한 추천 상품이 선택될 확률이 3.49%(0.20%p), 10개의 추천 상품이 노출되는 경우 2.92%(0.02%p) 높다는 것을 의미한다.

실험결과에 나타난 바와 같이, 제안 모형을 통해 상품 간 연관성과 함께 상품의 카테고리 간 연관성을 반영하여 추천 적중률을 예측하는 경우, 상품 간 연관성만을 사용하는 경우에 비해 높은 추천 성과를 얻을 수 있다는 점을 확인할 수 있다. 이는 제안 모형을 활용함으로써, 기존 거래를 통해 드러나지 않은 상품 간의 잠재적인 연관성을 효과적으로 찾아 소비자의 기호에 좀 더 적합한 추천 상품을 제시할 수 있음을 시사한다.

〈Table 5〉 Comparison of Performance Using Validation Data for Two Models

| Rank | Proposed Model | | | Competent Model | | | Performance Ratio |
|---|---|---|---|---|---|---|---|
| | Hit Count | Rec. Count | Hit Rate | Hit Count | Rec. Count | Hit Rate | |
| 1 | 5,163 | 85,193 | 6.06% | 4,989 | 85,193 | 5.86% | 103.49% |
| 2 | 6,681 | 169,152 | 3.95% | 6,416 | 169,152 | 3.79% | 104.13% |
| 3 | 7,656 | 252,126 | 3.04% | 7,324 | 252,126 | 2.90% | 104.53% |
| 4 | 8,407 | 334,165 | 2.52% | 8,132 | 334,165 | 2.43% | 103.38% |
| 5 | 9,096 | 415,380 | 2.19% | 8,804 | 415,380 | 2.12% | 103.32% |
| 6 | 9,730 | 495,803 | 1.96% | 9,506 | 495,803 | 1.92% | 102.36% |
| 7 | 10,276 | 575,514 | 1.79% | 10,080 | 575,514 | 1.75% | 101.94% |
| 8 | 10,803 | 654,583 | 1.65% | 10,575 | 654,583 | 1.62% | 102.16% |
| 9 | 11,335 | 733,021 | 1.55% | 11,059 | 733,021 | 1.51% | 102.50% |
| 10 | 11,823 | 810,855 | 1.46% | 11,488 | 810,855 | 1.42% | 102.92% |
| 11 | 12,230 | 888,133 | 1.38% | 11,900 | 888,133 | 1.34% | 102.77% |
| 12 | 12,624 | 964,787 | 1.31% | 12,286 | 964,787 | 1.27% | 102.75% |
| 13 | 13,057 | 1,040,952 | 1.25% | 12,676 | 1,040,952 | 1.22% | 103.01% |
| 14 | 13,426 | 1,116,649 | 1.20% | 13,029 | 1,116,649 | 1.17% | 103.05% |
| 15 | 13,756 | 1,191,931 | 1.15% | 13,373 | 1,191,931 | 1.12% | 102.86% |
| 16 | 14,121 | 1,266,750 | 1.11% | 13,708 | 1,266,750 | 1.08% | 103.01% |
| 17 | 14,466 | 1,341,145 | 1.08% | 14,050 | 1,341,145 | 1.05% | 102.96% |
| 18 | 14,838 | 1,415,107 | 1.05% | 14,417 | 1,415,107 | 1.02% | 102.92% |
| 19 | 15,184 | 1,488,583 | 1.02% | 14,787 | 1,488,583 | 0.99% | 102.68% |
| 20 | 15,532 | 1,561,665 | 0.99% | 15,138 | 1,561,665 | 0.97% | 102.60% |




〈Figure 3〉 Comparison of Performance Using Validation Data for Two Models

## 5. 결론

본 연구는 온라인 쇼핑에서 보편적으로 사용되고 있는 연관 상품 추천 시 상품의 속성을 추가로 활용함으로써 추천의 성공률을 높이기 위한 목적으로 수행되었다. 이런 추천 방식은 주로 상품의 상세 페이지를 방문한 소비자가 관심을 가질 확률이 높은 연관 상품을 함께 노출하는 방식으로 또 다른 상품을 추천한다. 이를 통해, 소비자는 자신이 원하는 상품을 탐색하는 데 소요되는 시간을 줄일 수 있을 뿐만 아니라, 이전에 미처 깨닫지 못했던 새로운 니즈를 인식할 수도 있다. 판매자의 입장에서, 소비자의 만족과 구매 증가는 매출의 증대로 이어지게 되므로 추천의 중요성이 크게 인식되고 있다. 즉, 적절한 상품을 추천 목록에 배치함으로써 소비자의 관심을 끌고 해당 상품의 페이지 방문을 유도하여 이를 구매로 전환할 수 있는 방법이 중요하게 인식되고 있는 것이다.

이를 위해서는 제한된 상품만을 노출시킬 수 있는 추천 목록에 좀 더 적중률이 높은 상품을 위주로 배치해야 할 필요가 있다. 소비자가 이미

방문한 상세 페이지는 다음에 관심을 가질 만한 상품을 찾는 단서가 될 수 있다. 즉, 이미 방문한 페이지의 상품과 연관성이 높은 상품을 찾아내는 연관 규칙 마이닝 기법을 활용하면 적중률이 높은 추천 상품을 찾을 수 있게 되는 것이다. 그러나, 상품 간의 연관성이 거래 데이터에 미처 반영되기 전인 상황에서는 이와 같은 방법이 추천 성공률을 높이는 데에 한계를 드러낼 수 있다. 즉, 상품이 실제로 추천에 성공할 수 있는 잠재성이 있는 경우라도, 거래가 충분히 이루어져 다른 상품과의 연관성이 충분히 드러나기 전까지는 이를 발견하기 힘들다는 것이다. 척도의 계산 자체가 이런 거래 건수를 기반으로 하고 있기 때문에, 낮은 지지도 혹은 신뢰도를 보인 경우 해당 상품은 다시 노출 기회를 잃고 이로 인해 거래 건수가 증가되기 힘든 악순환에 빠질 수도 있다. 신규 상품을 우선적으로 노출시키는 방법을 이를 극복하려는 노력이 이루어지고 있으나, 이는 잠재적으로 거래 건수가 낮은 상품이 함께 노출을 늘리는 결과를 야기함으로써 잠재성이 높은 상품의 판매기회를 뺏을 수 있다는 문제점을 갖는다.

이런 이유로, 본 연구는 해당 상품이 충분한 거래 건수를 확보하지 못한 경우에도 잠재성이 있는 경우 이를 발견할 수 있는 대안으로 상품의 속성을 활용하는 방법을 제안하였다. 상품의 속성에 대한 선호도는 소비자가 구매결정을 내리는 데 있어서 중요하게 고려하는 요소가 된다. 따라서, 본 연구는 선호하는 상품과 유사한 속성을 가진 상품을 추천함으로써 추천확률을 높일 수 있다는 가정을 전제에서 출발하였다. 즉, 상품 간의 연관성은 속성 간의 연관성을 통해 드러나게 될 것이라는 것이다. 여러 중요한 속성 중 카테고리를 고려하였고, 이들 간의 연관성을 추천에 반영하면 추천 정확도가 높아질 수 있다는 점을 검증하고자 하였다. 실험 결과를 통해 알 수 있듯이, 추천 성능은 2~3% 수준으로 향상될 수 있었다.

본 연구에서 제안한 상품의 속성 간 연관성을 활용한 연관 상품 추천 방식의 학문적 기여는 다음과 같다. 먼저, 개별 상품 간의 연관성만을 고려하던 기존의 방법을 확장하여 속성 수준에서의 연관성을 함께 고려한 추천모형을 제시하였다. 상품의 속성은 구매자가 상품을 선택하는 데에 있어 중요한 고려 대상이며 이를 추천모형에 반영함으로써 추천성과를 높일 수 있다는 것을 통계적 모형을 적용한 실험을 통해 보였다. 다음으로, 거래를 통해 연관성이 발견되지 않은 상품에 대해서도 확장할 수 있다는 점이 본 연구에서 제안한 모형의 또 다른 장점이라고 할 수 있다. 즉, 동일 고객에 의한 구매 건수가 전혀 없는 상품 간의 관계라고 하더라도, 이들 간의 상호 연관성을 찾을 수 있는 방법을 제시한 것이다. 즉, 두 상품 간의 속성의 일치를 통해 상품 수준의 연관성의 한계를 극복할 수 있는 것이다.

추천의 정확성은 그 자체로 추천 성공이 늘어나 매출이 증가된다는 장점을 지님과 동시에, 이렇게 발생한 주문 데이터가 다시 상품 간의 연관성을 찾는 데에 기여한다는 점에서 선순환을 기대할 수 있다. 추천 성공률을 향상시킬 뿐만 아니라 잠재된 연관성을 발견하는 데에 있어서 도움을 줄 수 있다는 점에서, 본 연구에서 제안한 모형은 실무적으로도 큰 의미를 갖는다고 할 수 있다. 또한, 카테고리 이외에도 색상이나 디자인과 같은 또 다른 속성 간의 연관성을 추가로 고려해볼 수 있는 가능성을 제시했다. 특히, 의류나 잡화와 같이 트렌드에 민감한 산업에서는 상품의 고유한 특성이 서로 크게 차별화될 것이다.

따라서, 이런 상품의 속성들을 활용하면 유사한 속성을 갖는 상품을 좀 더 쉽게 찾아 추천 성공률을 더 높일 수 있을 것으로 기대된다.

하지만, 본 연구는 이미 발생한 구매 데이터를 기반으로 실험을 수행하였으므로, 그 성능이 실제 환경에서 실무적으로 검증되지 못했다는 한계를 지닌다. 실제 현업에서 두 모형의 추천 적중률과 더 나아가 매출의 의미 있는 차이를 발견할 수 있다면 성능에 대한 객관적 평가가 이루어질 수 있을 것으로 기대된다.

## 참고문헌(References)


Agrawal, R., T. Imielinski, A. Swami. “Mining association rule between sets of items in large databases,” *Proc. 1993 ACM SIGMOD international conference on management of data*, (1993), 207~216.

Adomavicius, G., A. Tuzhilin. “Context-Aware Recommender Systems. *Recommender Systems Handbook*, Springer US, (2011), 217~253.

Aljukhadar, Muhammad, Sylvain Senecal, and Charles-Etienne Daoust. "Using recommendation agents to cope with information overload." *International Journal of Electronic Commerce* Vol.17, No.2(2012), 41~70.

Anand, S.S., A.R. Patrick. “A Data Mining methodology for cross-sales,” *Knowledge-Based Systems*, Vol.10**,** No.7(1998), 449~461.

Ansari, A., S. Essegaier, R. Kohli. “Internet recommender systems,” *Journal of Marketing Research*, Vol.37, No.3(2000), 363~375.

Balabanovic, M., Y. Shoham. “Content-Based, Collaborative Recommendation,” *Communications of the ACM*, Vol.40, No.3(1997), 66~72.

Bodapati, A.V. “Recommender systems with purchase data,” *Journal of Marketing Research,* Vol.45, No.1(2008), 77~93.

Chen, Y.L., J.M. Chen, C.W. Tung. “A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales,” *Decision Support Systems,* Vol.42, No.3(2006), 1503~1520.

Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman. "Choice overload: A conceptual review and meta-analysis." *Journal of Consumer Psychology*, Vol.25, No.2 (2015), 333~358.

Choi, S., Hyun, Y., Kim, N. “Improving Performance of Recommendation Systems Using Topic Modeling,” *Journal of Intelligence and Information Systems*, Vol.21, No.3(2015), 101~116.

Choi, S., Kwahk, K.-Y., Ahn, H. “Enhancing Predictive Accuracy of Collaborative Filtering Algorithms using the Network Analysis of Trust Relationship among Users,” *Journal of Intelligence and Information Systems*, Vol.22, No.3(2016), 113~127.

Fleder, D., K. Hosanagar. “Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity,” *Management Science,* Vol.55, No.5(2009), 697~712.

Kim, B. K., S. Lee, S. Bang, J. Kim, and J. H. Lee, "Personalized Recommendation System Using Social Network," Proceedings of the Conference on Intelligent Information Systems, Vol.20, No.1(2010), 48~49.

Kim, J., Lee, S.-W. “The Ontology Based, the

Movie Contents Recommendation Scheme, Using Relations of Movie Metadata," *Journal of Intelligence and Information Systems*, Vol.19, No.3(2013), 25~44.

Kim, K.-J., Kim, B.-G. "Product Recommender System for Online Shopping Malls using Data Mining Techniques," *Journal of Intelligence and Information Systems*, Vol.11, No.1(2005), 191~205.

Kim, M., and K. J. Kim, "Recommender Systems using Structural Hole and Collaborative Filtering," *Journal of Intelligence and Information Systems*, Vol.20, No.4(2014), 107~120.

Kim, M. G., and K. J. Kim, " Recommender Systems using SVD with Social Network Information," *Journal of Intelligence and Information Systems*, Vol.22, No.4(2016), 1~18.

Kim, S. H., and R. S. Chang, "The Study on the Research Trend of Social Network Analysis and the its Applicability to Information Science," *Journal of the Korean Society for Information Management*, Vol.27, No.4(2010), 71~87.

Kim, Y., and W.N. Street. "An intelligent system for customer targeting: a data mining approach," *Decision Support Systems,* Vol.37, No.2(2004), 215~228.

Konstan, J.A., B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl. "GroupLens: applying collaborative filtering to Usenet news," *Communications of the ACM*, Vol.40, No.3(1997), 77~87.

Lee, D. "A Regression-Model-based Method for Combining Interestingness Measures of Association Rule Mining." *Journal of Intelligence and Information Systems*, Vol.23, No.1(2017), 127~141.

Lee, D. "Extension Method of Association Rules Using Social Network Analysis." *Journal of Intelligence and Information Systems*, Vol.23, No.4 (2017), 111~126.

Lee, D., S. Park, S. Moon. "Utility-based association rule mining: A marketing solution for cross-selling," *Expert Systems with Applications*. Vol.40, No.7(2013), 2715~2725.

Scheibehenne, Benjamin, Rainer Greifeneder, and Peter M. Todd. "Can there ever be too many options? A meta-analytic review of choice overload." *Journal of consumer research* Vol.37, No.3(2010), 409-425.

Shin, C. H., J. W. Lee, H. N. Yang, and I. Y. Choi, "The Research on Recommender for New Customers Using Collaborative Filtering and Social Network Analysis," *Journal of Intelligence and Information Systems*, Vol.18, No.4(2012), 19~42.

Abstract

# A Study on the Improvement of Recommendation Accuracy by Using Category Association Rule Mining

Dongwon Lee*

Traditional companies with offline stores were unable to secure large display space due to the problems of cost. This limitation inevitably allowed limited kinds of products to be displayed on the shelves, which resulted in consumers being deprived of the opportunity to experience various items. Taking advantage of the virtual space called the Internet, online shopping goes beyond the limits of limitations in physical space of offline shopping and is now able to display numerous products on web pages that can satisfy consumers with a variety of needs. Paradoxically, however, this can also cause consumers to experience the difficulty of comparing and evaluating too many alternatives in their purchase decision-making process. As an effort to address this side effect, various kinds of consumer's purchase decision support systems have been studied, such as keyword-based item search service and recommender systems. These systems can reduce search time for items, prevent consumer from leaving while browsing, and contribute to the seller's increased sales. Among those systems, recommender systems based on association rule mining techniques can effectively detect interrelated products from transaction data such as orders. The association between products obtained by statistical analysis provides clues to predicting how interested consumers will be in another product. However, since its algorithm is based on the number of transactions, products not sold enough so far in the early days of launch may not be included in the list of recommendations even though they are highly likely to be sold. Such missing items may not have sufficient opportunities to be exposed to consumers to record sufficient sales, and then fall into a vicious cycle of a vicious cycle of declining sales and omission in the recommendation list. This situation is an inevitable outcome in situations in which recommendations are made based on past transaction histories, rather than on determining potential future sales possibilities. This study started with the idea that reflecting the means by which this potential possibility can be identified indirectly would help to select highly recommended products. In the light of the fact that the attributes of a product affect the consumer's

* Corresponding Author: Dongwon Lee
Division of Social Sciences, Hansung University
116 Samseongyoro-16gil, Seongbuk-gu, Seoul 02876, Korea
Tel: +82-2-760-4250, Fax: +82-2-760-4482, E-mail: dongwonlee@hansung.ac.kr

purchasing decisions, this study was conducted to reflect them in the recommender systems. In other words, consumers who visit a product page have shown interest in the attributes of the product and would be also interested in other products with the same attributes. On such assumption, based on these attributes, the recommender system can select recommended products that can show a higher acceptance rate. Given that a category is one of the main attributes of a product, it can be a good indicator of not only direct associations between two items but also potential associations that have yet to be revealed. Based on this idea, the study devised a recommender system that reflects not only associations between products but also categories. Through regression analysis, two kinds of associations were combined to form a model that could predict the hit rate of recommendation. To evaluate the performance of the proposed model, another regression model was also developed based only on associations between products. Comparative experiments were designed to be similar to the environment in which products are actually recommended in online shopping malls. First, the association rules for all possible combinations of antecedent and consequent items were generated from the order data. Then, hit rates for each of the associated rules were predicted from the support and confidence that are calculated by each of the models. The comparative experiments using order data collected from an online shopping mall show that the recommendation accuracy can be improved by further reflecting not only the association between products but also categories in the recommendation of related products. The proposed model showed a 2 to 3 percent improvement in hit rates compared to the existing model. From a practical point of view, it is expected to have a positive effect on improving consumers' purchasing satisfaction and increasing sellers' sales.

**Key Words** : Recommender System, Recommendation Accuracy Improvement, Recommendation Acceptance Rate Improvement, Association Rule Mining, Category Association Rule Mining

Received : April 4, 2020 Revised : May 11, 2020 Accepted : May 26, 2020
Publication Type : Regular Paper Corresponding Author : Dongwon Lee

# 저 자 소 개

**이 동 원**

한양대학교에서 재료공학사, 카이스트 경영대학원에서 경영정보학석사 학위와 경영공학박사 학위를 취득했다. LG CNS에서 시스템 엔지니어로 근무한 바 있고, 한성대학교에 교수로 재직 중이다. 주요연구분야는 데이터마이닝, 딥러닝, 소비자 행동, 추천 시스템이다.