

# 연관상품 추천을 위한 회귀분석모형 기반 연관 규칙 척도 결합기법

이동원

한성대학교 경영학부  
(dongwonlee@hansung.ac.kr)

인터넷과 모바일 관련 기술의 발전과 기기의 보급은 물리적 공간의 제약을 극복하게 하고, 다양한 상품과 서비스를 소비자에게 제공함으로써, 소비자에게 선택의 폭을 넓히는 기회를 제공하는 반면, 많은 시간과 노력을 기울이고도 소비자가 자신의 기호에 적합한 상품을 선택하기 힘들어지는 부작용을 낳았다. 이에 따라, 기업은 추천 시스템을 활용하여 소비자가 원하는 상품을 더 쉽게 찾는 수단을 제공하고 있다. 상품 간의 연관성을 통계적으로 분석하는 연관 규칙 마이닝 기법은 직관적인 형태의 척도를 규칙과 함께 제공함으로써, 이로부터 도출된 규칙에 포함된 품목 간의 관계를 이해하고, 이를 추천에 적용하기 쉽다는 강점을 갖는다. 그러나, 서로 다른 규칙의 척도가 일관되지 어느 한 쪽의 규칙이 더 우위에 있음을 알려주지 못한다면, 수많은 품목 중 추천에 적합한 품목을 적절히 선별해내기 힘든 상황이 발생한다. 본 연구에서는 추천 상품의 순위를 결정할 수 있도록 연관 규칙 마이닝 기법에 회귀분석모형을 보완적으로 적용하는 방안을 제시하고자 수행되었다. 연관 규칙 마이닝에서 보편적으로 사용되고 있는 지지도, 신뢰도, 향상도를 활용하여 모형을 구현함으로써, 직관적으로 이해하기 쉬운 뿐만 아니라, 실무에서도 활용하기 쉬운 방안을 제시하고자 하였다. 국내 최대규모의 온라인 쇼핑물의 주문 데이터를 활용한 실험을 통해, 제안된 모형으로부터 얻어진 추천 점수를 기반으로 추천상품을 결정하고, 이를 추천에 적용함으로써 추천 적중률을 향상시킬 수 있음을 보였다. 특히, 최근 모바일 상거래가 빠르게 확산됨에 따라, 제한된 화면에 한정된 수의 추천 품목을 제시해야 하는 상황에서 적합한 추천 기법임을 확인할 수 있었다.

**주제어** : 추천 시스템, 연관 규칙 마이닝, 회귀분석모형, 온라인 쇼핑, 모형 기반 추천 시스템

논문접수일 : 2017년 3월 6일    논문수정일 : 2017년 3월 6일    게재확정일 : 2017년 3월 13일  
원고유형 : 일반논문    교신저자 : 이동원

## 1. 서론

인터넷이 보편화되면서 급속도로 성장한 전자상거래는 최근 스마트폰, 태블릿PC와 같은 모바일 기기의 보급이 확산됨에 따라 또 다른 성장 기회를 맞고 있다. 물리적인 공간의 제약을 받는 오프라인 상점과는 달리 온라인 상점은 인터넷이라는 가상 공간에 무한에 가까운 상품을 보다

적은 비용으로 진열하고 판매할 수 있다는 장점을 갖는다. 이는 주력 상품의 판매에 의존하던 기존과는 달리 다양한 상품의 판매를 통해 매출을 크게 늘릴 수 있는 기회로 작용하고 있다.

그러나, 또 다른 한편으로 이렇게 많아진 상품은 소비자가 자신이 원하는 상품을 찾기 힘든 부작용을 야기시키고 있다. 한정된 상품만을 보고 이로부터 구매결정을 내리던 과거와는 달리, 최

근 소비자들은 다양한 정보를 습득하고 이로부터 자신의 취향에 맞는 상품을 찾기 위해 노력하고 있다. 아마존을 비롯한 많은 온라인 쇼핑몰에서는 이런 소비자의 요구에 부응하고 매출을 증대하기 위한 목적으로 추천상품을 제시하려는 노력을 기울이고 있다.

연관상품추천은 이러한 필요를 충족시킬 수 있는 방법으로서 많은 기업들이 추천 시스템 구축에 활용하고 있는 방안이다. 이는 소비자가 방문한 상품 페이지의 일부 영역에 추천상품을 제시함으로써 원하는 상품에 빠르고 쉽게 접근할 수 있도록 한다. 이런 방식을 통해 소비자는 자신이 방문한 페이지의 상품뿐만 아니라 추천상품까지도 자신의 구매대안으로 함께 고려할 수 있게 된다. 즉, 소비자는 자신이 더 선호하는 상품을 고를 수 있는 기회를 얻게 될 수 있고, 미처 인지하지 못하던 상품에 대한 니즈를 알게 되는 기회를 갖게 된다.

이런 연관상품을 추천하기 위해 널리 활용되고 있는 방법이 연관 규칙 마이닝인데(Agrawal, 1993), 이는 상품 간의 연관성을 지지도, 신뢰도, 향상도라는 세 가지 척도로 평가하고 이를 연관 규칙이라는 형태로 표현하는 기법이다. 연관 규칙 마이닝에서 제시되는 규칙의 형태는 아래의 예와 같이 표현된다.

규칙1: 햄 → 계란 (지지도 2%, 신뢰도 30%)

규칙2: 햄 → 식빵 (지지도 1%, 신뢰도 40%)

위의 예시에 나타난 두 개의 규칙은 햄이라는 상품에 대해 연관성이 높은 계란과 빵을 통계적인 방법으로 발견하고, 이를 지지도, 신뢰도라는 척도와 함께 정형화된 형태로 제시하고 있다. 이런 규칙은 계란과 빵을 추천하면 햄만을 구매한

소비자로부터 추가 구매를 유발할 수 있다는 것을 암시적으로 표현하기 때문에 연관상품 추천시에 활용될 수 있다. 그러나, 제한된 화면에 한정된 수의 추천상품을 제시해야 하는 온라인 쇼핑 상황에서, 연관 규칙 마이닝 기법은 우선적으로 배치해야 할 추천상품을 어떻게 선별할 것인지에 대한 적절한 해답을 제시하기 위한 일관적이고 보편적인 방법을 제시하지 못한다.

위의 예에서, 연관 규칙의 척도는 햄이라는 상품 페이지를 방문한 소비자에게 계란과 식빵 중 어느 것이 더 추천에 적합할 것인지에 대한 해답을 제공하지 못한다. 지지도가 더 높다는 것은 해당 규칙을 적용할 기회가 상대적으로 더 많음을 암시한다. 위의 규칙에서 햄과 계란을 함께 구매한 경우가 전체 장바구니의 2%, 햄과 식빵이 1%에서 발견되었음을 보임으로써 앞으로도 햄은 식빵보다 계란과 함께 더 많이 구매될 것임을 암시한다. 이는 앞으로 햄을 구매하는 고객에게 식빵보다 계란을 추천하는 것이 더 유리한 것처럼 보일 수 있다. 그러나, 햄을 구매한 고객 중 30%가 계란을 구매한 데 반해, 40%가 식빵을 구매했다는 것을 보여주는 신뢰도는, 지지도와는 반대로 계란보다 식빵이 더 유리할 수 있다는 것을 암시한다. 이와 같이 연관 규칙 마이닝 기법의 두 척도는 상충되는 결과를 보일 수 있다.

이와 같이 서로 다른 규칙 간에 서로 절대적 우위를 판단하는 것이 불가능한 경우, 어느 규칙이 더 우선적으로 적용되어야 할 지 결정하기 힘든 상황이 발생한다. 온라인 상점에서 거래되는 상품의 수가 수십만 개에 이른다는 점을 감안하면, 다양한 상품 간에 수 많은 연관 규칙이 존재할 수 있으므로, 어떤 상품을 우선적으로 추천해야 할 것인지를 결정하는 것은 쉬운 일이 아니다. 더구나, 모바일 기기와 같이 한정된 크기의

화면에 제한된 수의 상품을 배치하려면, 소비자가 선호할 만한 상품을 최우선적으로 제시하는 것이 매우 중요하다. 따라서, 추천상품의 우선순위를 결정할 수 있는 일관된 척도를 개발하는 것이 필요하다.

이러한 통합 척도는 서로 상충되는 추천결과를 중재하는 역할을 해야 함은 물론이고, 우수한 추천성적을 보장할 수 있어야 할 것이다. 또한, 기존의 척도와 마찬가지로 직관적일 이해될 수 있고, 실무에 적용이 용이해야 한다. 따라서, 본 연구는 이런 요건을 갖춘 척도를 개발하되 기존 척도를 효과적으로 결합하여 통합된 척도를 만드는 방법을 제시하려는 목적으로 수행되었다.

## 2. 이론적 배경

### 2.1 추천 시스템

본 연구는 추천 시스템에 관한 기존연구들을 이론적 배경으로 한다. 추천 시스템은 기존에 사용자가 표현한 선호도나 상품 간 유사도를 기반으로 상품이나 서비스를 추천하는 일종의 의사결정지원 시스템으로서의 기능을 수행한다. 이는 사용자로 하여금 구매와 관련된 상품이나 서비스의 검색 노력을 줄이고, 기업으로 하여금 고객 충성도와 매출을 증대하는 효과를 가져온다 (Ansari et al., 2000). 효과적인 추천 기법을 개발하거나 (Balabanovic and Shoham, 1997; Ansari et al., 2000; Adomavicius and Tuzhilin, 2011; Choi et al., 2016), 추천 시스템의 성과를 측정하거나 (Bodapati, 2008; Fleder and Hosanagar, 2009), 다양한 분야에 활용하기 위한 목적으로 (Choi et al., 2015; Kim and Lee, 2013; Kim et al., 2010),

추천 시스템과 관련한 다양한 연구들이 수행되었다.

추천 시스템의 개발과 관련한 다양한 기법이 널리 연구되고 있으며, 그 중에서도 내용기반 필터링(content-based filtering)과 협업 필터링(collaborative filtering)이 주류를 이루고 있다. 내용기반 필터링 기법은 상품 또는 서비스의 유사성에 기반을 두고 추천 상품이나 서비스를 찾는 반면, 협업 필터링은 사용자 간의 유사성을 활용하는 방법이다(Konstan et al., 1997; Ansari et al., 2000). 내용기반 필터링은 상품 간의 유사성을 찾기 위해 의사결정나무, 최근접 이웃 기법 등 다양한 분류 기법을 활용한다. 사용자가 기존에 선호한 상품과 유사한 특성을 갖는 상품 중 아직 경험하지 않은 상품을 추천하는 방식이다. 반면에, 협업 필터링은 사용자가 상품이나 서비스에 대해 표현한 선호도를 기반으로 다른 사용자와의 유사도를 측정한다. 사용자와 유사한 다른 사용자가 선호한 상품 중 아직 사용자가 경험하지 않은 상품이 추천 대상으로 선정된다.

### 2.2 연관 규칙 마이닝

본 연구의 또 다른 이론적 배경은 연관 규칙 마이닝이다. 이는 Agrawal et al. (1993)이 제시한 기법으로, 품목(상품 또는 서비스)이 주문과 같은 거래에서 동시 출현한 패턴을 연관 규칙의 형태로 표현한 것이다. 연관 규칙 마이닝 기법은, 두 품목이 동시에 같은 거래에 나타난 빈도를 측정함으로써, 한 품목의 판매가 다른 품목의 판매에 얼마나 강하게 영향을 미치는지를 표현한다는 점에서 마케팅 프로모션과 교차판매 등에 적용될 수 있다 (Anand, 1998; Chen et al., 2006; Kim and Street, 2004; Lee et al., 2013; Kim and

Kim, 2005).

연관 규칙은 선행 품목(antecedent item,  $A$ )과 후행 품목(consequent item,  $C$ )을 지지도(support,  $sup$ ), 신뢰도(confidence,  $conf$ )와 함께 아래와 같이 정형화된 형태로 표시된다.

$$A \rightarrow C (sup\%, conf\%)$$

지지도와 신뢰도는 연관 규칙의 강도를 표현하기 위해 사용되는데, 지지도는 전체 거래에서 선행 품목과 후행 품목이 동시에 나타난 거래의 비율로 측정되며, 신뢰도는 선행 품목이 나타난 거래 중 후행 품목이 포함된 거래의 비율로 계산된다. 선행 품목을 구매한 소비자에게 여러 후행 품목 중 신뢰도가 더 높은 품목이 추천되는 경우, 이를 소비자가 받아들일 확률이 더 높을 수 있음을 암시하는 것으로 해석 가능하다는 점에서 중요한 의미를 갖는다. 그러나, 이런 거래가 사전에 충분히 발생하지 않았다면 이런 추천에 대한 확신이 부족할 것이라는 측면에서 지지도도 함께 고려되어야 한다. 즉, 신뢰도와 지지도가 상대적으로 높은 규칙이 그렇지 않은 규칙보다 더 우선적으로 적용되는 것이 타당할 것이다.

그러나, 동일한 선행 품목에 대해 서로 다른 후행 품목을 갖는 두 개 이상의 연관 규칙이 존재하는 경우, 이들 중 어느 규칙도 다른 규칙에 비해 지지도와 신뢰도가 모두 우세하지 않다면, 어떤 규칙을 적용하는 것이 더 바람직할 것인지에 대한 판단에 어려움이 발생할 수 있다. 즉, 신뢰도가 더 높지만 지지도가 낮은 품목과 반대로 신뢰도는 낮지만 지지도가 더 높은 품목 중 어느 것이 추천에 적합할 것인지를 결정해야 하는 고민에 빠지게 된다.

### 3. 연구 모형

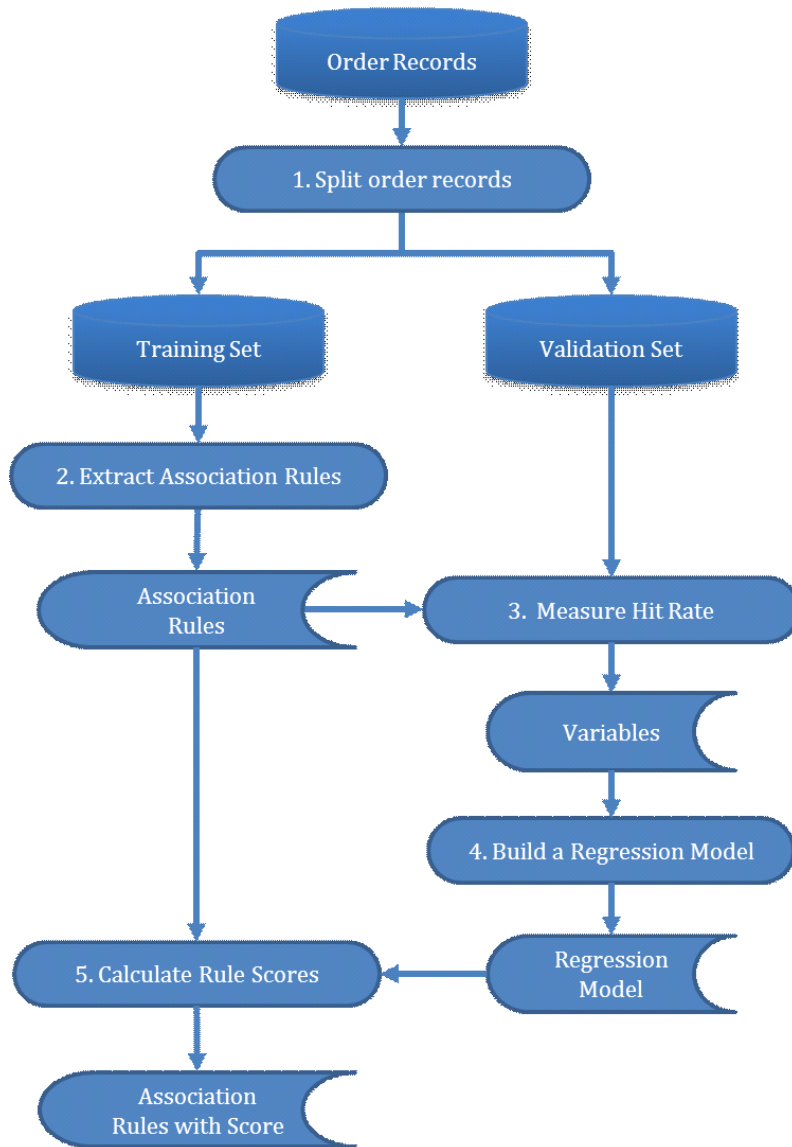
주어진 기준 상품에 대해 어떤 상품을 우선적으로 추천하는 것이 적합한지를 결정하기 위해 추천 상품의 순위를 결정할 수 있는 점수의 계산이 필요하다. 연관상품의 추천은 상품 간의 연관성에 기반하여 특정 상품에 대해 추천에 적합한 연관상품을 찾는 방식으로 이루어진다. 온라인 쇼핑에서는 일반적으로 개별 상품 페이지의 특정 영역에 대략 10개 정도의 상품들이 노출되는데, 소비자의 선택 가능성을 높이기 추천상품의 선정과 배치가 매우 중요하다. 본 연구에서는 추천상품의 노출 순서를 결정하기 위한 방안으로서 연관 규칙 마이닝 기법과 회귀분석모형을 활용하여 개별 상품에 대한 추천상품의 선호도를 점수로 계산하는 방안을 제시한다.

본 연구에서 제안하는 기법을 적용하기 위한 프로세스는 <Figure 1>과 같이 5단계로 이루어진다.

첫째, 주문 데이터를 학습용 집합과 검증용 집합으로 구분한다. 주문 데이터는 임의추출 방식에 의해 두 개의 집합으로 구분하되, 동일한 소비자의 주문은 동일한 집합에 포함되도록 한다.

둘째, 학습용 집합으로부터 연관 규칙을 추출한다. 동일한 소비자가 구매한 상품끼리 같은 장바구니에 포함된 것으로 간주하고, 연관 규칙 마이닝 기법을 활용하여 상품의 구매과정에서 서로 연관성을 갖는 상품을 찾아 연관 규칙을 작성한다. 이 과정에서 각 규칙 별로 지지도, 신뢰도, 향상도를 측정한다.

셋째, 검증용 집합에 규칙을 적용하여, 각 규칙 별 추천 성공률을 구한다. 검증용 집합의 소비자별로 기준상품(선행 품목)이 포함된 경우 추천상품도 함께 포함되었는지를 확인하는 방식으



〈Figure 1〉 Research Model

로 추천의 성공여부를 측정한다. 각 규칙 별로 전체 추천회수에 대해 적중된 수를 비율로 계산하여 0에서 1사이의 값을 갖는 추천 성공률을 구한다.

넷째, 각 규칙의 추천성공률과 연관 규칙 척도를 이용하여 회귀모형을 작성한다. 앞서 구한 추천 성공률(HitRate)을 종속변수, 규칙의 지지도(sup)뿐만 아니라 기준상품의 지지도(sup<sub>A</sub>)와 추

천상품의 지지도( $sup_B$ ), 규칙의 신뢰도( $conf$ )와 향상도( $lift$ )를 변수에 포함하여 아래의 식과 같이 회귀모형을 구축한다.

$$HitRate = \beta_0 + \beta_1sup_A + \beta_2sup_B + \beta_3sup_Asup_B + \beta_4sup + \beta_5conf + \beta_6lift + \varepsilon \quad (1)$$

다섯째, 회귀모형의 계수를 이용하여 연관 규칙의 추천점수를 계산한다. 추천점수는 연관 규칙모형에서 등호의 우측에 놓인 수식으로부터 오차항( $\varepsilon$ )을 제외하고 계산하여 얻을 수 있다. 즉, 학습용 집합으로부터 생성된 연관 규칙의 지지도, 신뢰도, 향상도와 함께 기준상품과 추천상품의 지지도를 함께 식(1)에 포함하여 적중률을 계산하고 이를 추천점수로 사용한다. 이로부터 하나의 기준상품에 대해 추천대상 상품의 목록을 정할 수 있는데, 추천상품을 그 추천점수가 높은 순서대로 노출시킬 수만큼 선정한다.

## 4. 실험

### 4.1 실험 목적

실험은 본 연구에서 제안한 기법의 유효성을 검증하기 위한 목적으로 실시되었다. 즉, 회귀모형을 활용하여 선택된 추천상품이 기존의 방식에 의해 선택된 추천상품보다 더 높은 추천 적중률을 보이는지를 평가하기 위한 것이다. 본 연구에서 사용된 기존 추천방식은 신뢰도를 기반으로 추천상품을 선택하는 방식을 취하였다.

본 실험에 사용된 데이터는 국내 최대 규모의 쇼핑몰 중 하나인 11번가의 주문 데이터로서, 웹과 모바일을 통해 주문한 내역이다. 온라인에서

는 상품의 수명주기가 다소 짧은 편이므로, 주문기간의 간격이 너무 긴 경우 상품 간의 연관성이 낮아진다고 보여, 2010년 6월부터 2011년 5월까지 1년 간의 거래 데이터만을 활용하였다.

### 4.2 실험 방법

본 실험에서는 온라인 쇼핑몰의 주문정보를 활용하여 일부의 데이터로부터 추천규칙을 추출하고, 나머지 데이터를 검증에 활용함으로써 추천 적중률을 확인하는 방식으로 진행되었다. 새로 제안된 추천방식과 기존의 방식에 대해 동일한 과정을 수행하여 그 결과를 비교함으로써, 새로 제시된 방안이 기존의 방식에 비해 얼마나 개선된 결과를 보여주는지를 확인하였다.

본 실험은 3장에서 소개된 프로세스에 따라 진행되었으며, 본 연구에서 제안한 방식의 개선도를 보여주기 위한 평가과정을 추가로 실시하였다. 구체적인 방법은 다음과 같다.

우선, 임의추출 방식으로 주문 데이터를 3개의 집합으로 나누어, 각각을 학습용, 검증용, 평가용으로 활용하였다. 주문 데이터는 온라인 쇼핑몰의 1년치 데이터를 사용하였다. 다음으로, 학습용 집합에서 연관상품규칙을 추출하였다.

두 경쟁 모형에 대해 동일하게, 각 기준상품당 100개의 추천상품을 확보하고, 각 모형에서의 추천상품 순위선정 방식에 따라 추천상품에 대해 1부터 100까지의 순위를 부여하였다. 기존 모형에 대해서는 신뢰도가 높은 순서에 따라 추천상품의 순위를 결정하였고, 제안 모형에서는 회귀모형의 추천 적중률이 높은 순서로 추천상품의 순위를 결정하였다.

두 모형으로부터 추출된 규칙의 추천성과를 비교하기 위해 각 모형에서 추출된 규칙을 평가

집합에 적용하여 추천성공여부를 확인하고 그 성과를 상호 비교하였다. 이 과정에서 평가에 적용된 고객의 수는 10,004명이며 추천대상이 된 기준상품의 수는 36,550개로서, 이를 고객별로 누적한 결과는 129,975개이며 이는 고객 당 평균 13개의 상품에 대해, 상품 당 평균 3.56명의 고객에게 추천이 이루어진 것에 해당한다.

### 4.3 실험 결과

<Table 1>과 <Table 2>는 각각 실험에 사용된 변수의 기술통계량과 변수 간 상관계수를 보여준다. 변수에는 학습용 집합으로부터 추출된 연관 규칙을 검증용 집합에 적용하여 측정된 결과로부터 얻어진 적중률(hit rate)과 연관 규칙의 지지도, 신뢰도, 향상도, 선행 품목의 지지도(sup1) 후행품목의 지지도(sup2)를 포함하고 있다. 변수

들 간의 강한 상관관계가 발견되지 않아 그대로 분석에 사용되었다.

<Table 3>은 연관 규칙에서 얻어진 설명변수들로부터 종속변수인 적중률을 추정하는 회귀분석을 실시한 결과이다. 연관 규칙의 기본 척도인 지지도, 신뢰도, 향상도만을 설명변수로 포함한 모형(Model 1)과 이 척도들에 선행 품목, 후행 품목의 지지도를 추가로 포함시킨 모형(Model 2)에 대해 분석을 실시한 결과 2번 모형이 더 설명력이 우수한 것으로 판단되어 이를 본 연구에서 제안하는 최종모형으로 선정하고 이후의 실험과정을 진행하였다.

<Table 4>는 제안 모형과 경쟁 모형의 성과를 비교한 결과로서, 추천순위 1위부터 20위까지의 상품이 보인 추천결과를 보여준다. 경쟁모형에 대해서는, 보편적으로 추천 기준으로 많이 사용

<Table 1> Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
hit_rate	160,674	0.1599701	0.2266645	0.003876	1
sup	160,674	0.000184	0.0001436	0.0001325	0.0055666
conf	160,674	0.1640605	0.2302552	0.0043478	1
lift	160,674	219.1239	792.6603	0.4851154	7545
sup1	160,674	0.0040229	0.0050981	0.0001325	0.0304838
sup2	160,674	0.0040229	0.0050981	0.0001325	0.0304838
sup1sup2	160,674	0.0000124	0.0000222	1.76E-08	0.0006141

<Table 2> Correlation Coefficients

	sup	conf	lift	sup1	sup2	sup1sup2
sup	1					
conf	-0.0153	1				
lift	-0.0675	0.5497	1			
sup1	0.2311	-0.4055	-0.1775	1		
sup2	0.2311	0.0966	-0.1775	-0.1467	1	
sup1sup2	0.5623	-0.2191	-0.1436	0.4458	0.4458	1

〈Table 3〉 Regression Analysis of Experiment Data

Variables	hit_rate Coefficient (standard error)	
	Model 1	Model 2
Constant	0.0543619*** (0.0006333)	0.063724*** (0.0007833)
sup	-97.28487*** (2.479409)	-80.76567*** (3.036262)
conf	0.7324337*** (0.0018465)	0.6844672*** (0.0020559)
lift	0.0000153*** (0.000000537)	0.0000215*** (0.00000056)
sup1		-3.428722*** (0.0895096)
sup2		2.130937*** (0.0901961)
sup1sup2		-54.34277** (25.7462)
Observations	160674	160674
Adjusted R-squared	0.6058	0.6135

〈Table 4〉 Performance Comparison of Two Models

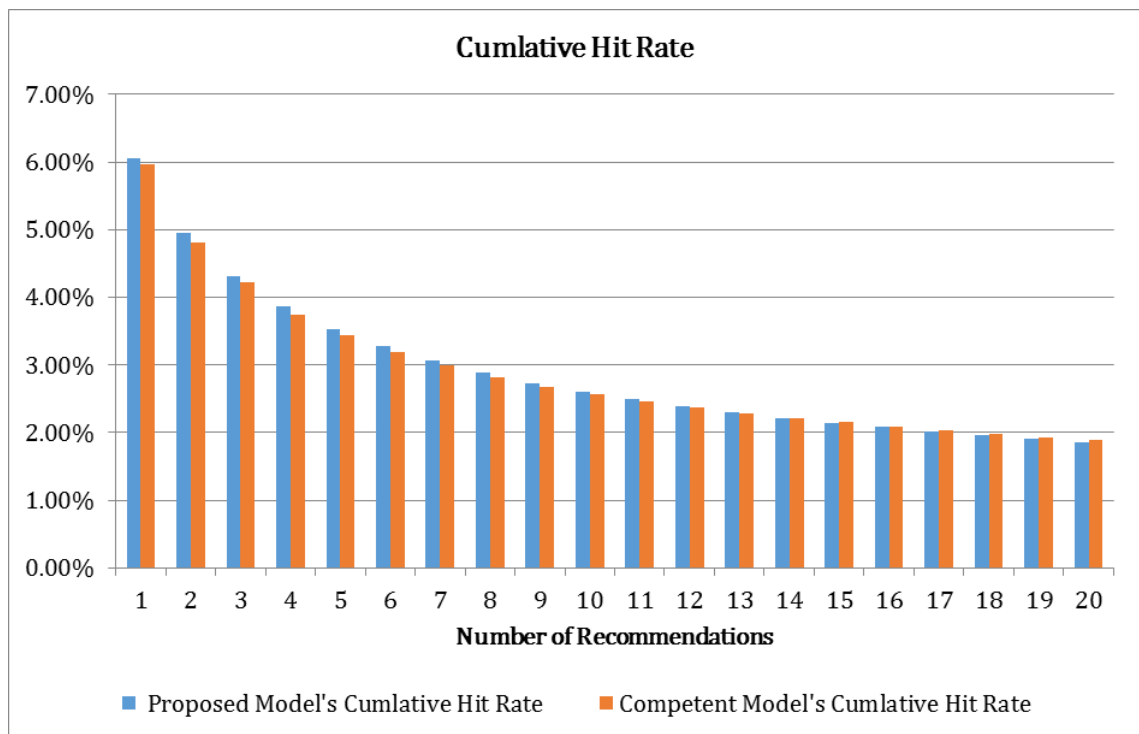
Rankings of Recommendations	Number of Hit Items		Cumulative Number of Hit Items		Cumulative Hit Rate		Cumulative Degree of Hit Rate Improvement
	Proposed Model	Competitive Model	Proposed Model	Competitive Model	Proposed Model	Competitive Model	
1	7,868	7,747	7,868	7,747	6.05%	5.96%	1.6%
2	5,018	4,754	12,886	12,501	4.96%	4.81%	3.1%
3	3,939	3,969	16,825	16,470	4.31%	4.22%	2.2%
4	3,283	2,998	20,108	19,468	3.87%	3.74%	3.3%
5	2,799	2,834	22,907	22,302	3.52%	3.43%	2.7%
6	2,618	2,585	25,525	24,887	3.27%	3.19%	2.6%
7	2,329	2,298	27,854	27,185	3.06%	2.99%	2.5%
8	2,198	2,133	30,052	29,318	2.89%	2.82%	2.5%
9	1,942	2,039	31,994	31,357	2.74%	2.68%	2.0%
10	1,839	1,977	33,833	33,334	2.60%	2.56%	1.5%
11	1,756	1,872	35,589	35,206	2.49%	2.46%	1.1%
12	1,679	1,813	37,268	37,019	2.39%	2.37%	0.7%
13	1,602	1,695	38,870	38,714	2.30%	2.29%	0.4%
14	1,519	1,665	40,389	40,379	2.22%	2.22%	0.0%
15	1,495	1,561	41,884	41,940	2.15%	2.15%	-0.1%
16	1,403	1,446	43,287	43,386	2.08%	2.09%	-0.2%
17	1,305	1,594	44,592	44,980	2.02%	2.04%	-0.9%
18	1,259	1,390	45,851	46,370	1.96%	1.98%	-1.1%
19	1,226	1,410	47,077	47,780	1.91%	1.93%	-1.5%
20	1,224	1,364	48,301	49,144	1.86%	1.89%	-1.7%



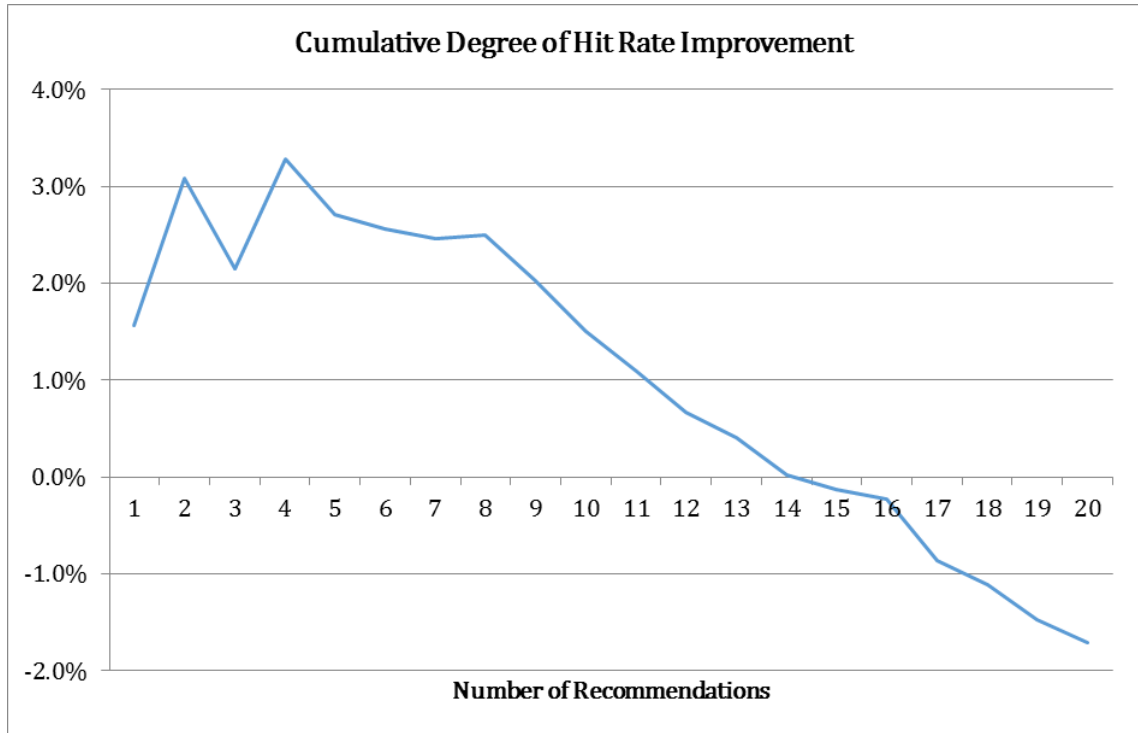
되고 있는 신뢰도를 순위 결정에 적용하였다. 즉, 신뢰도가 가장 높은 상품으로부터 추천 순위가 결정되도록 하였다. 추천에 대해 적중된 상품의 수와 이를 누적한 값을 표기했으며, 이로부터 누적 적중률을 구했고, 두 모형을 비교하여 제안모형의 개선도를 계산한 결과를 포함하고 있다. 각 추천상품의 추천순위가 증가함에 따라 추천성도가 감소하는 경향이 두 모형에서 공통적으로 뚜렷이 나타나는 것으로 보아 두 모형 모두 적중률이 높은 순서대로 추천규칙이 추출되었음을 확인할 수 있다.

두 모형의 성과를 비교하면, <Figure 2>에서 볼 수 있는 것처럼, 두 모형의 1위 추천상품이 각

각 6.05%, 5.96%의 누적 적중률로 제안모형이 더 우세함을 보이다가, 추천상품의 추천순위가 증가함에 따라 그 차이가 점차 감소하는 경향을 확인할 수 있다. <Figure 3>의 누적 개선도에서도 확인할 수 있듯이, 13위의 추천상품까지는 제안모형이 우세하다가 15위 이후부터는 기존모형이 더 우세하게 나타나는 것을 관찰할 수 있었다. 제한된 수의 상품이 추천되어야 하는 온라인 쇼핑 환경에서 일반적으로 10개 정도의 추천상품을 제시한다는 것을 전제로 본다면, 제안모형의 성과가 기존모형에 비해 더 우수하게 나타날 것이라고 기대할 수 있다.



<Figure 2> Cumulative Hit Rate



〈Figure 3〉 Cumulative Degree of Hit Rate Improvement

## 5. 결론

인터넷과 모바일 관련 기술의 발전과 기기의 보급은 물리적 공간의 제약을 극복하게 함으로써, 다양한 상품과 서비스를 소비자에게 제공할 수 있게 되었다. 이런 변화는 소비자에게 선택의 폭을 넓히는 기회라는 측면과 더불어 많은 시간과 노력을 기울이고도 자신의 기호에 적합한 품목을 선택하기 힘들어지는 위기의 측면도 함께 가져왔다. 소비자는 자신이 원하는 상품과 서비스를 더 찾기 쉬운 환경을 원하고 있고, 기업은 이에 부응하기 위하여 추천 시스템을 활용하고 있다.

상품 간의 연관성을 통계적으로 분석하여 규칙의 형태로 제공하는 연관 규칙 마이닝 기법은 최근 많은 온라인 기업들이 활용하고 있는 추천 알고리즘이다. 이 기법은 직관적인 형태의 척도를 규칙과 함께 제공함으로써, 이로부터 도출된 규칙에 포함된 품목 간의 관계를 이해하고, 이를 추천에 적용하기 쉽다는 강점을 갖는다. 서로 다른 규칙의 척도가 일관되지 어느 한 쪽의 규칙이 더 우위에 있음을 알려주지 못한다면, 온라인 상점에서 거래되는 수십만 개의 품목 중 추천에 가장 적합한 소수의 품목을 구별해내기 힘든 상황이 발생한다. 적합하지 못한 추천 품목은 외면당할 뿐만 아니라, 추천 자체에 대한 신뢰를 잃을

수도 있다.

본 연구에서는 추천 상품의 순위를 결정할 수 있도록 회귀분석모형을 적용하는 방안을 제시하였다. 실무에서 활용하기 쉬운 방안을 제시하는 것이 중요하다는 판단 하에, 연관 규칙 마이닝에서 보편적으로 사용되고 있는 지지도, 신뢰도, 향상도를 활용하여 모형을 구현하였다. 국내 최대규모의 온라인 쇼핑몰인 11번가의 주문 데이터를 활용한 실험은, 모형으로부터 얻어진 추천 점수를 기반으로 추천상품을 결정하고, 이를 적용함으로써 추천 적중률을 높일 수 있음을 보여준다. 특히, 최근 모바일 상거래가 빠르게 확산됨에 따라, 제한된 화면에 한정된 수의 추천 품목을 제시해야 하는 상황에서 적합한 추천 기법임을 확인할 수 있었다.

본 연구는 상품을 취급하는 온라인 쇼핑몰에서 교차판매에 적용되는 것을 전제로 수행되었으나, 본 연구에서 제안한 방법은 연관 규칙을 적용하는 다양한 상황에서 모두 적용할 수 있을 것으로 기대할 수 있다. 또한, 모형의 정교화를 위해 추가적인 변수를 고려해볼 필요가 있을 것이다. 즉, 연관 규칙으로부터 얻은 지지도, 신뢰도, 향상도와 선후행 품목의 지지도 외에도, 가격과 같은 변수를 추가로 고려하는 노력을 통해 더욱 정교한 추천 시스템을 구축하는 것이 가능할 것으로 기대해볼 수 있을 것이다.

## 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.
- 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. Essegai, 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," *J. 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.
- 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, H. K., Choi, I. Y., Ha, K. M., Kim, J. K. “Development of User Based Recommender System using Social Network for u-Healthcare,” *Journal of Intelligence and Information Systems*, Vol.16. No.3(2010), 181~199.
- 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, Y., 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., S. Park, S. Moon. “Utility-based association rule mining: A marketing solution for cross-selling,” *Expert Systems with Applications*. Vol.40, No.7(2013), 2715~25.

Abstract

## A Regression-Model-based Method for Combining Interestingness Measures of Association Rule Mining

Dongwon Lee\*

Advances in Internet technologies and the proliferation of mobile devices enabled consumers to approach a wide range of goods and services, while causing an adverse effect that they have hard time reaching their congenial items even if they devote much time to searching for them. Accordingly, businesses are using the recommender systems to provide tools for consumers to find the desired items more easily. Association Rule Mining (ARM) technology is advantageous to recommender systems in that ARM provides intuitive form of a rule with interestingness measures (support, confidence, and lift) describing the relationship between items. Given an item, its relevant items can be distinguished with the help of the measures that show the strength of relationship between items. Based on the strength, the most pertinent items can be chosen among other items and exposed to a given item's web page. However, the diversity of the measures may confuse which items are more recommendable. Given two rules, for example, one rule's support and confidence may not be concurrently superior to the other rule's. Such discrepancy of the measures in distinguishing one rule's superiority from other rules may cause difficulty in selecting proper items for recommendation. In addition, in an online environment where a web page or mobile screen can provide a limited number of recommendations that attract consumer interest, the prudent selection of items to be included in the list of recommendations is very important. The exposure of items of little interest may lead consumers to ignore the recommendations. Then, such consumers will possibly not pay attention to other forms of marketing activities. Therefore, the measures should be aligned with the probability of consumer's acceptance of recommendations. For this reason, this study proposes a model-based approach to combine those measures into one unified measure that can consistently determine the ranking of recommended items. A regression model was designed to describe how well the measures (independent variables; i.e., support, confidence, and lift) explain consumer's acceptance of

---

\* Corresponding Author: Dongwon Lee  
School of Business Administration, College 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

recommendations (dependent variables, hit rate of recommended items). The model is intuitive to understand and easy to use in that the equation consists of the commonly used measures for ARM and can be used in the estimation of hit rates. The experiment using transaction data from one of the Korea's largest online shopping malls was conducted to show that the proposed model can improve the hit rates of recommendations. From the top of the list to 13th place, recommended items in the higher rankings from the proposed model show the higher hit rates than those from the competitive model's. The result shows that the proposed model's performance is superior to the competitive model's in online recommendation environment. In a web page, consumers are provided around ten recommendations with which the proposed model outperforms. Moreover, a mobile device cannot expose many items simultaneously due to its limited screen size. Therefore, the result shows that the newly devised recommendation technique is suitable for the mobile recommender systems. While this study has been conducted to cover the cross-selling in online shopping malls that handle merchandise, the proposed method can be expected to be applied in various situations under which association rules apply. For example, this model can be applied to medical diagnostic systems that predict candidate diseases from a patient's symptoms. To increase the efficiency of the model, additional variables will need to be considered for the elaboration of the model in future studies. For example, price can be a good candidate for an explanatory variable because it has a major impact on consumer purchase decisions. If the prices of recommended items are much higher than the items in which a consumer is interested, the consumer may hesitate to accept the recommendations.

**Key Words** : Recommender system, association rule mining, regression model, online shopping, model-based recommender system

Received : March 6, 2017 Revised : March 6, 2017 Accepted : March 13, 2017

Publication Type : Regular Paper Corresponding Author : Dongwon Lee

## 저 자 소개



### 이 동 원

LG CNS에서 시스템 엔지니어로 근무하였으며, KAIST 경영대학원에서 MIS 전공으로 석사/박사 학위를 취득하였다. 현재 한성대학교 경영학부 조교수로 재직 중이다. 현재 빅데이터에 기반한 연구를 주로 수행하고 있으며, 주요 관심분야는 고객관계관리, 추천 시스템, 데이터 마이닝 기법의 정교화, 디지털 콘텐츠 마케팅 등이다.