

# 전자상거래에서의 협업 추천을 위한 고객 프로필 모델

## A Customer Profile Model for Collaborative Recommendation in e-Commerce

이석기\*, 조 현\*\*, 천성용\*\*\*  
 한양대학교 정보시스템학과\*, 한국과학기술원 경영대학\*\*, 단국대학교 경영학부\*\*\*

Seok Kee Lee(goodsense@hanyang.ac.kr)\*, Hyeon Jo(sineoriz@nate.com)\*\*,  
 Sung Yong Chun(sychun@dankook.ac.kr)\*\*\*

### 요약

협업 추천은 전자상거래 기업들이 고객별로 개인화 된 상품추천을 하기 위하여 널리 활용하는 추천기법이다. 추천 행위는 고객들이 상품에 대해 가지고 있는 선호도를 분석하고 이를 프로파일화 하는 것을 전제로 한다. 전통적인 명시적 평가법은 취급하는 상품이 매우 다양한 전자상거래 기업의 고객들에게 구매활동 외에 추가적 부담을 준다는 점에서 한계를 가진다. 따라서 고객의 개입 없이 간접적으로 선호도를 파악할 수 있는 묵시적 평가법이 보다 바람직하다고 볼 수 있는데, 여기에도 추정된 선호도의 지표화에 주로 사용하는 카디널 척도가 추정 오차를 증가 시킨다는 점에서 문제점이 있다. 따라서 본 연구에서는 이러한 문제의 개선을 위해 웹 마이닝과 사전식 컨센서스 기법에 근간하여 서열 척도 기반의 고객 프로필을 생성, 활용하는 협업 추천 기법을 제안하고자 한다. 실제 온라인 쇼핑물의 거래 데이터를 이용한 실험을 통해 제안된 기법의 우수성을 입증 하였다.

■ 중심어 : | 협업 필터링 | 사전식 컨센서스 | 추천 시스템 |

### Abstract

Collaborative recommendation is one of the most widely used methods of automated product recommendation in e-Commerce. For analyzing the customer's preference, traditional explicit ratings are less desirable than implicit ratings because it may impose an additional burden to the customers of e-commerce companies which deals with a number of products. Cardinal scales generally used for representing the preference intensity also ineffective owing to its increasing estimation errors. In this paper, we propose a new way of constructing the ordinal scale-based customer profile for collaborative recommendation. A Web usage mining technique and lexicographic consensus are employed. An experiment shows that the proposed method performs better than existing CF methodologies.

■ keyword : | Collaborative Filtering | Lexicographic Consensus | Recommender System |

## I. 서 론

협업 추천 시스템 (collaborative recommendation)은

Amazon.com을 시작으로 현재 수많은 전자상거래 기업들이 자사 사이트를 방문하는 고객들에게 개인화 된 상품 추천 서비스를 제공하기 위해 널리 활용하고 있는

기법이다 [1]. 협업 추천 시스템은 고객별 선호도를 분석하고 이 정보를 저장하기 위한 고객 프로필(customer profile)을 필요로 한다. 고객 프로필은 사이트를 방문하는 m 명의 고객들이 그 곳에서 취급되는 n 개의 상품들에 대해 가지는 선호도 정보를 보관하기 위해 논리적으로  $m \times n$ 의 행렬 형태를 갖는다. 이 고객 프로필에 담긴 고객의 선호도 정보 수준에 따라 최종적인 추천의 성과가 결정되기 때문에 고객 프로필을 정확하게 작성하는 것이 협업 추천 시스템에서는 무엇보다 중요하다.

고객프로필의 생성은 먼저 고객의 선호도를 수집, 분석하는 과정과 분석된 정보를 지표화 하는 연속적인 과정으로 이루어진다. 고객의 선호도를 분석함에 있어 기존의 많은 연구들은 고객으로 하여금 직접적으로 상품에 대한 선호도를 입력하게 하는 명시적 방법(explicit ratings)을 주로 활용하고 있다. 비교적 쉽고 정확한 방법이지만 수많은 종류의 상품을 한 곳에서 판매해야 하는 오늘날 대부분의 전자상거래 사이트에서 명시적 평가법을 활용하는 것은 현실적으로 쉽지 않다. 이는 고객으로 하여금 구매활동 이외에 상품의 평가라는 추가적인 활동을 요구함으로써 결국 고객에게 부담을 가중시키는 결과를 초래하기 때문이다. 따라서 전자상거래 사이트에 활용되는 협업추천 시스템의 경우에는 고객의 개입 없이 간접적으로 고객의 선호도를 추정할 수 있는 묵시적 평가법 (implicit ratings)을 사용하여 고객 선호도를 분석하는 것이 보다 바람직하다고 할 수 있다.

분석된 선호도 정보의 크기를 지표화 하는 과정에서 기존에 묵시적 평가법들은 5점 척도 혹은 7점 척도와 같은 카디널 척도 (cardinal scale)을 주로 활용하는 경향이 있다. 하지만 불충분한 고객의 주변정보 (예. 행동정보, 구매정보 등)만을 활용하여 추정된 선호도를 카디널 척도 기반으로 지표화 할 경우 실제 값과 예측된 값 사이의 추정 오차 (estimation error)가 오히려 증가할 수 있다는 연구결과들이 존재한다 [2-4]. 결국 많은 고객들을 대상으로 다양한 종류의 상품을 노출하고 판매해야 하는 일반적인 전자상거래 기업들의 특성상, 묵시적 평가법에 의해서 고객의 선호도를 수집하고 또한 수집된 선호도를 카디널 척도보다 추정오차가 적은

서열 척도를 기반으로 표현하는 고객 프로필의 활용이 권장된다.

이에 본 연구에서는 웹 마이닝 기법을 활용하여 사용자의 선호도를 간접적으로 분석하고, 이를 사전식 컨센서스 방법을 활용하여 서열 척도 기반의 고객 프로필로 생성하는 새로운 협업 추천 방법론을 제안하고자 한다.

본 연구의 구성은 다음과 같다. 2장에서는 관련 문헌에 대해 간단히 고찰을 하였고 3장에서는 제안된 방법론을 자세히 설명한다. 4장에서는 실험결과에 대한 분석을 수행하고 5장에서 결론 및 향후 연구에 대해 토론한다.

## II. 관련 문헌 연구

### 1. 협업 필터링

협업 필터링은 가장 효과적인 추천기법으로 널리 알려져 있고 다양한 분야에서 활용되어 왔다 [1][4][5]. 협업 필터링은 실제 추천 과정에서 고객 프로필의 희박성의 문제라든가 새로운 고객 혹은 상품의 등장 시에 곧바로 추천이 되지 않는 점과 같은 잘 알려진 문제점들을 노출한 바 있는데, 이런 문제점들은 수많은 연구들에 의해서 해결책이 제시된 바 있다 [1][5-8]. 이에 반해 비교적 불충분한 고객 주변 정보들을 활용해 간접적으로 선호도를 추정해야 하는 묵시적 평가법의 전제 하에서 보다 덜 민감한 서열 척도를 활용하는 이슈를 제기하고 있는 연구는 거의 없다. 기존 몇몇 연구들에서 서열척도의 활용에 관해 제안한 연구들이 있으나 이들은 대부분 명시적 평가법을 전제로 선호도를 서열척도로 지표화 하는 것에 대한 연구들이거나 비교적 소수의 상품을 취급하는 모바일 환경에서 서열척도를 활용하는 연구 등이다 [4][9][10]. 현재의 대부분 전자상거래 기업들을 방문하는 고객의 수와 취급하는 상품의 종류는 크게 증가하고 있다. 위에서 언급된 연구들은 아예 묵시적 평가법을 전제로 하지 않거나[9][10] 혹은 대량의 상품 및 고객을 대상으로 하는 고객 프로필 생성 작업에 적용했을 때 복잡한 프로필 생성과정으로 인해 응답속도의 저하나 시스템 과부하 등의 문제를 야기할 수 있

는 한계점이 있다 [4]. 한편, 간접적으로 선호도를 추정하는데 활용될 수 있는 기법으로 웹 마이닝을 제안한 연구들이 몇몇 있는데 이들 연구는 카디널 척도를 이용하여 추정된 선호도를 지표화 한 이유로 추정오차의 증가문제를 해결하지 못하고 있다 [11][12]. 이에 본 연구는 위 연구들의 한계점을 개선하기 위한 방안으로 웹 마이닝에 의한 선호도 분석과 사전식 컨센서스 기법을 통해 서열적도 기반으로 선호도 크기를 표현하는 고객 프로필의 생성법을 제안하였다.

## 2. 컨센서스 기법

웹 마이닝은 기본적으로 고객이 웹상에서 보여준 트랜잭션들 (transactions)을 분석하여 고객의 행동정보 속에 내재되어 있는 선호도를 간접적으로 추정한다. 고객이 한번 방문할 때마다 하나의 트랜잭션이 생성되고 고객별로 일정기간  $n$  개의 트랜잭션에 대해  $n$  개의 부분적 선호도가 추정되게 된다. 이를 하나의 선호도로 종합하여 고객프로필에 저장하기 위해서는 부분적 선호도를 통합할 수 있는 메커니즘이 필요한데 이와 관련하여 다기준 의사 결정 (multi-criteria decision making, MCDM) 분야에서는 선호도 통합을 위한 다양한 컨센서스 기법들이 연구되어 왔다. 그 중에서도 사전식 컨센서스 기법 (lexicographic consensus)은 선호도 차이를 의미하는 지표들을 기준으로 해당 지표의 존재 여부 (existence)와 빈도 (frequency) 등에 의해 대안들 간의 우선 순위를 결정짓는 방법론이다. Condorcet [13], Borda [14], Cook [2] 등에 의해 제안된 방법론들은 보편적인 컨센서스 기법들로 활용되고 있지만 이들은 전통적으로 소수의 한정된 대안(혹은 상품)에 대해서 단일한 선호도 순위를 생성하는데 적합한 경향이 있다. 비교해야 할 상품의 개수가 증가할 경우 동일 순위를 다수 생성하는 등의 문제점을 노출한다. Cook [2]은 자신의 연구에서 선호도의 크기 (degree of preference) 까지 표현할 수 있는 컨센서스 기법을 제안 하였으나, 이 방법은 선호도 크기 계산 과정이 복잡하여서 적은 상품을 대상으로 할 때에도 많은 계산시간과 자원을 필요로 한다. 즉, 많은 종류의 상품과 많은 고객을 대상으로 실시간에 가깝게 추천 목록을 생성해야 하는 전자상

거래 환경 하에서의 추천 시스템에는 이들 컨센서스 기법의 적용은 효과적이지 못하다. 반면에 본 연구에서 활용된 사전식 컨센서스 기법은 대안 상품의 수가 비교적 많은 경우에도 중복이 최소화 된 선호도 순위를 생성할 수가 있어서 고객 선호도를 매우 세부적으로 표현할 수 있다는 장점이 있다. 구현 측면에서도 알고리즘이 매우 간편하여서 빠른 응답 속도를 기반으로 한 고객 프로필의 생성이 가능하다.

## III. 제안된 방법론

본 연구에서 제안하는 협업 추천 방법론은 웹 마이닝 단계, 컨센서스 기법을 활용한 서열 척도 기반의 고객 프로필 생성 단계, 이웃 선정 및 추천 목록의 생성 단계와 같은 연속적인 3개의 과정으로 구성되어 있다. 개별 단계들에 대한 자세한 설명은 다음 장에서부터 차례대로 제시한다.

### 1. 웹 마이닝

첫 번째 단계인 웹 마이닝 단계에서는 고객의 행동정보를 포함하고 있는 웹 데이터 (예. 웹 로그와 구매 데이터 등)를 정제하여 고객의 행동 정보만을 추출하고 여기에 내재된 선호도 정보를 분석하는 과정이다. 본 연구에서도 Kim 외 [11] 와 Lee 외 [4] 등에서 제안한 바와 마찬가지로 고객이 웹상에서 보이는 행동정보는 그 단계별로 내재된 선호도의 차이가 있다는 것을 전제한다. 가령 구매한 (purchase) 상품은 장바구니에 담긴 (basket placement) 상품보다는 선호도가 높다는 가정이 그것이다. 또한 장바구니에 담긴 상품은 단순히 조회만 한 (click-through) 상품보다는 높은 선호도를 갖는다는 것으로 가정 한다. 이 조건에 따라 웹 로그 데이터 내에서 위 세 가지의 고객 행동 기록만을 추출하고 트랜잭션별로 정렬할 수 있다. 본 연구에서는 이 작업을 위해 Lee 외 [4]의 연구에서 고객 행동 패턴의 추출을 위해 제안되었던 매트릭스를 약간 수정하여 활용하였다. [4]의 연구에서는 모바일 환경의 특성을 반영하기 위해 고객의 행동 패턴을 구매, 미리듣기, 클릭하기,

무시하기 등의 4가지로 분류한 반면, 본 논문에서는 전자상거래 환경을 고려하여야래 수식 (1)과 같이 정의되는 고객 행동 집합 (customer action set, CAS), 즉  $V^*$ 를 정의하여 활용 하였다.

$$V^* = (v_{sj}),$$

$$v_{sj} = \begin{cases} P, & \text{상품 } m_j \text{를 구매한 경우} \\ B, & \text{상품 } m_j \text{를 장바구니 담기한 경우} \\ C, & \text{상품 } m_j \text{를 조회한 경우} \end{cases}, \quad (1)$$

수식 (1) 중에서  $s$ 는 트랜잭션을 의미하며  $m_j$ 는 상품 목록 중에서  $j$ 번째 상품을 의미한다.

이후 단계의 설명 편의를 위해 임의의 대상 고객 A가 있다고 가정하고 고객 A의 CAS를 아래 [표 1]과 같이 구성 하였다.

표 1. 목표고객 A의 고객 행동 집합 (CAS<sub>A</sub>)

	$m_1$	$m_2$	$m_3$	$m_4$
트랜잭션 1	B	B	C	C
트랜잭션 2	B	B		P
트랜잭션 3	B	B	P	
트랜잭션 4	P	C	P	B

[표 1]의 첫 번째 열을 보면 두 개의 B와 C가 포함되어 있는데 이는 고객 A가 첫 번째 트랜잭션에서 상품  $m_1$ 과  $m_2$ 에 대해 장바구니에 담은 행동을 보였으며 상품  $m_3$ 과  $m_4$ 에 대해서는 조회를 했다는 것을 의미한다.

## 2. 서열 척도 기반의 고객 프로필 생성

이 단계는 고객별로 생성된 CAS를 이용하여 해당 고객을 대표하는 통합 선호도를 생성하고 이를 전체 고객에게 반복함으로써 최종적인 고객 프로필을 생성하는 과정이다. 구체적으로 두 단계의 세부과정으로 구성되어 있다.

### 2.1 고객별 통합 선호도의 추출

CAS의 각 행은 일정 기간 동안 개별 고객들이 남긴 전체 트랜잭션의 빈도에 대응하여 증가한다. CAS의 각 행은 해당 트랜잭션 과정에서 고객에게 노출되었던 일

부 상품들에 대한 고객의 선호도, 즉 부분적인 선호도 정보를 내포하고 있다. 따라서 이를 통합하여 개별 고객을 대표하는 선호도로 표현할 필요가 있다. 본 연구에서는 이를 통합 선호도 (unified preference)로 정의하며 가령 고객 b의 통합선호도  $E_b$ 는 아래 수식 (2)에서와 같이 정의될 수 있다.

$$E_b = \{r_{bj} | m_j \in M_b\}, \quad (2)$$

여기서  $r_{bj}$ 는 상품  $j$ 에 대한 고객  $b$ 의 선호도 순위 (rank)를 의미한다.  $E_b$ 는 사전식 컨센서스 기법을 적용함으로써 생성할 수 있다. 알고리즘은 비교적 간단하면서도 효과적이다. 2장에서 잠깐 언급된 바와 같이 선호도의 차이를 결정지을 수 있는 지표 (criteria)들을 활용하여 지표의 존재 여부와 등장 횟수를 비교 기준으로 상품의 그룹을 세분화해 나간다. 사전에 결정된 모든 선호도의 기준 지표들을 활용한 후에는 분류된 그룹 순서대로 내부 상품들에게 순위를 부여하는 원리이다. 우리의 방법론에서는 이미 언급된 세 가지의 고객 행동 패턴을 기준 지표로 활용 하였는데, 가령 [표 1]에 제시된 고객 A의 통합 선호도 ( $E_A$ )가 추출되는 과정은 아래 [그림 1]과 같다.

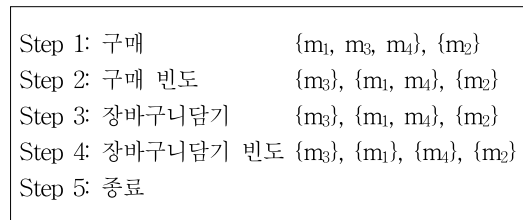


그림 1. CAS<sub>A</sub>에 대한 사전식 컨센서스 기법의 적용

[그림 1]에서 보는 것과 같이 고객 A의 경우 구매 경험 여부에 따라 전체 상품 중에서 구매 경력이 없는  $m_2$  상품과 나머지 3개 상품으로 분류가 되었으며 동일 그룹 내의 3개 상품에 대해서는 구매의 빈도에 따라 다시 세분화 과정을 거쳤다. 이 후 두 번째 지표인 장바구니 담기의 여부와 빈도를 기준으로 사전식 정렬을 수행하여 각 상품별로 차별화된 그룹 생성이 완료된 것을 볼 수 있다. 최종적으로 각 그룹 내의 상품들에 대해서 그

품의 순서별로 서열화된 순위를 부여하면 고객 A의 통합 선호도인  $E_A$ 를 얻게 되는데 이는 다음의 [표 2]와 같다.

표 2. 고객 A의 통합선호도 ( $E_A$ )

	$m_1$	$m_2$	$m_3$	$m_4$
고객 A	2	4	1	3

### 2.2 서열 척도 기반의 고객 프로필 생성

2.1 절에서 생성된 통합 선호도는 개별 고객이 전체 상품에 대해 가지고 있는 선호도 순위를 예측한 정보이다. 따라서 이를 전체 고객에 대해서 동일한 반복 작업을 수행하게 되면 협업 추천에 활용할 수 있는 서열 척도 기반의 완전한 고객 프로필을 얻을 수가 있다. 전체 고객의 선호도 정보를 표현하고 있는 최종적인 서열 척도 기반의 고객 프로필  $U$ 는 다음 수식 (3)과 같이 정의된다.

$$U = \{E_1, \dots, E_Y\}. \quad Y: \text{고객의 수}, \quad (3)$$

### 3. 이웃 선정과 추천 목록의 생성

2 단계에서 생성된 고객 프로필을 이용하여 본 단계에서는 협업 추천의 일반적인 과정인 이웃 선정 (neighborhood formation)과 추천 목록의 생성 (recommendation generation)이 이루어진다.

#### 3.1 이웃의 선정

이웃 선정이란 추천의 대상이 되는 목표 고객과 가장 유사한 선호도를 가진  $k$  명의 고객을 선정하는 것을 말한다. Kamishima [9]의 연구에서 기존의 서열 척도 기반의 고객 프로필을 활용하여 이웃을 선정하는 방법이 제시된 바 있는데 본 연구에서도 이와 비슷한 방식으로 고객 간의 유사도를 계산하고 이웃을 선정 하였다. 가령 목표 고객  $a$ 의 이웃 고객  $b$ 를 선정하기 위해 필요한 고객들 간의 유사도 ( $R_{ab}$ )는 아래 수식 (4)에 의해 계산될 수 있다.

$$R_{ab} = \frac{\sum_{m_j \in M_{ab}} (r_{aj} - \bar{r}_a)(r_{bj} - \bar{r}_b)}{\sqrt{\sum_{m_j \in M_{ab}} (r_{aj} - \bar{r}_a)^2} \sqrt{\sum_{m_j \in M_{ab}} (r_{bj} - \bar{r}_b)^2}}, \quad (4)$$

여기서  $M_{ab}$ 는 고객  $a, b$ 에 의해 모두 경험되어진 상

품의 집합을 말하며  $\bar{r}_a = \frac{\sum_{m_j \in M_{ab}} r_{aj}}{|M_{ab}|}$ 를 뜻한다. 이와 같은 유사도 수식을 이용하여 높은 유사도 ( $R_{ab}$ ) 값을 갖는  $k$  명 (best- $k$ -neighborhood)을 목표고객의 이웃으로 선정할 수 있다.

#### 3.2 추천 목록의 생성

이웃으로 선정된 고객들이 이전에 구매하거나 경험한 상품 정보를 근거로 목표 고객이 선호할 만한  $N$ 개의 상품 목록 (top- $N$  recommendation)을 선정하는 단계이다. 보여지는 상품의 범위를 확장하고 새로운 구매를 유도하기 위하여 목표 고객이 이전에 구매하였던 상품들은 목록 생성에서 제외된다. 목표 고객  $a$ 의 상품  $j$ 에 대한 예상 선호도 순위 ( $\hat{r}_{aj}$ )는 아래의 수식 (5)에 의해 계산된다.

$$\hat{r}_{aj} = \frac{\sum_{b \in M_j} R_{ab} \times (r_{bj} - \bar{r}_b)}{\sum_{b \in M_j} |R_{ab}|}, \quad (5)$$

여기서  $\bar{M}_j = \{b | E_i \in U \text{ s.t. } m_j \in M_b\}$ 을 의미한다.  $\hat{r}_{aj}$  값은 예측값이긴 하나 순위를 의미하므로 후보 상품들 중에서 예측되어진  $\hat{r}_{aj}$  값이 낮은 차례대로  $N$ 개의 상품들이 추천 목록에 포함된다.

## IV. 실험 결과

### 1. 실험 설계

#### 1.1 데이터

제안된 추천 시스템의 성능을 평가하기 위해 실제 H 온라인 쇼핑몰의 웹 로그 데이터가 실험에 활용 되었다. 해당 기업의 IIS 웹 서버들 내에 30 일간 누적된 총

124개, 64,370MB 규모의 웹 로그파일을 수집하였다. 초기 상태의 웹 로그 파일에 대한 전처리 (data pre-processing) 과정을 통해 불필요한 정보를 제거한 결과 총 1,904개의 상품에 대해 66,329명 고객들이 관여되어 있는 2,249,540 건의 트랜잭션을 추출할 수 있었다. 학습 집합 (training set)과 검사 집합 (test set)의 구성은 다양한 사전 검증들 통해 가장 적합한 것으로 판명된 학습: 검사 = 20일:10일의 기준에 따라 데이터를 분리하고 실험에 활용 하였다.

### 1.2 벤치마킹 시스템

제안된 협업 추천 시스템 (이후 실험 결과의 설명을 위해 Lex-CF로 칭함)의 주요 특징은 기존의 카디널 척도 대신 서열 척도를 활용한다는 점과 서열 척도의 생성하는 과정에 사전식 컨센서스 기법을 활용했다는 점이다. 따라서 카디널 척도를 사용하는 시스템과 함께 사전식 컨센서스 기법이 아닌 다른 방법으로 서열 척도를 생성하여 사용하는 시스템과의 성능 비교를 해 보는 것이 보다 객관적으로 제안된 방법론의 성능을 검증하는 방법일 것이다. 따라서 공통적으로 웹 마이닝을 활용한다는 전제 하에서, 카디널 척도로 선호도를 표현하는 추천 시스템 (CS-CF로 칭함)과 서열 척도를 활용하되 단순히 카디널 척도 값의 크기에 따라 우선순위를 부여하는 방식으로 생성된 서열 척도를 활용하는 추천 시스템 (OS-CF로 칭함)과 같은 두 개의 벤치마킹 시스템을 동시에 개발하여 본 연구에서 제안된 시스템과 비교 하였다.

### 1.3 성능 측정 지표

추천 시스템의 성능을 측정하는 데에는 재현율 (recall)과 정확률 (precision), 그리고 F1-metric 이 주로 사용된다. 이 중 재현율과 정확률은 추천의 개수가 증가할수록 재현율은 증가하고 정확률은 감소하는 상반된 결과를 보인다. 따라서 아래 식 (6)과 같이 두 성능 평가 지표를 동일한 가중치로 결합한 F1-metric을 사용하여 시스템간의 성능을 비교 하였다.

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \quad (6)$$

## 2. 실험 결과 및 해석

앞서 전처리된 실제 웹 데이터를 이용하여 비교 대상인 세 개 시스템들의 추천 성능을 측정 하였다. 구체적인 결과는 아래 [그림 2]와 같다. 여기에 제시된 그래프들은 이웃 고객의 수를 변경하면서 측정된 F1-metric 값들이다. 이 중에서 먼저 카디널 척도에 대비하여 서열척도를 활용하는 것이 과연 추천 시스템의 성능 향상에 영향을 주는지를 분석하기 위해 CS-CF와 OS-CF 시스템의 F1-metric을 비교 하였다. 이는 두 시스템의 유일한 차이가 선호도 표현에 사용된 척도의 종류이기 때문이다.

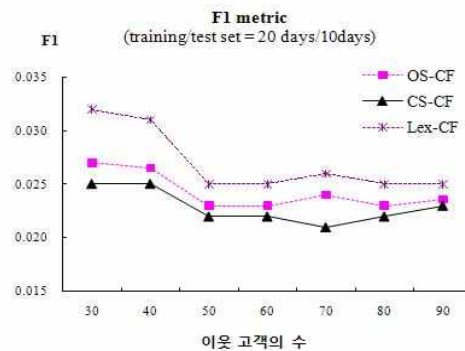


그림 2. 추천 시스템들의 F1-metric

분석의 결과, 그래프의 전 영역에 걸쳐 OS-CF의 F1-metric은 0.023에서 0.026 범위의 값을 가지며 CS-CF의 0.021에서 0.025 범위의 값을 갖는 것으로 나타났다. 이웃 고객의 수에 관계없이 그래프의 전 영역에서 OS-CF의 값이 우수하며 CS-CF에 비해 전체적으로는 8.7% 정도의 성능 향상이 존재 하는 것으로 나타났다. 두 시스템이 갖는 F1-metric 차이의 통계적 검증을 위해 T-test를 수행한 결과, 유의도 1% 수준 이내에서 값의 차이가 유의한 것으로 나타났다.

이러한 실험 결과는 결국 선호도 예측에 필요한 정보의 제공이 불충분한 전자상거래 환경에서 선호도의 크기를 기존의 카디널 척도로 표현할 경우, 앞선 관련 연

구들에서 언급된 바와 같이 실제 선호도와 예측된 선호도 간의 차이를 의미하는 추정 오차가 증가하고 결국 추천 시스템의 성능에 부정적 영향을 미치기 때문으로 해석할 수 있겠다.

한편, 선호도를 서열 척도 기반으로 표현하는 과정에서도 적용되는 컨센서스 기법의 특성에 따라서 추천 시스템의 성능이 영향을 받는지를 분석하기 위하여 OS-CF와 Lex-CF의 F1 metric 값도 추가적으로 비교해 보았다. 그래프의 전 영역에 걸쳐 있는 Lex-CF의 F1-metric 평균값은 약 0.027로 나타났다. 이는 0.025의 평균값을 가진 OS-CF에 비해 약 8% 정도 높은 결과이다. 앞서의 비교에서와 동일한 T-test를 통해 두 시스템의 수치를 비교한 결과 유의도 1% 이내에서 값이 차이가 유의한 것으로 검증 되었다. 비교 대상이 된 두 시스템의 유일한 차이가 선호도 순위 생성에 활용된 컨센서스 기법이기 때문에 어떠한 컨센서스 기법을 적용하느냐에 따라서도 추천 시스템의 성능은 달라질 수 있다는 것으로 해석할 수가 있다. 또한 이는 전통적인 MCDM 분야에서 컨센서스 기법에 따라 통합된 선호도 순위는 조금씩 달라짐을 입증한 기존의 연구 결과들에도 상응하는 결과로 볼 수 있다.

## V. 결론

본 연구에서 제안된 협업 추천을 위한 고객 프로파일 생성 방법론은 전자상거래 사이트들이 대부분 겪고 있는 상품 과부하 (product overload)의 문제를 해소하면서 고객별로 개인화 된 추천 서비스를 가능케 한다. 이전의 방법론들과 비교했을 때 본 연구에서 제안된 방법론은 다음과 같은 차별적 특징을 가지고 있다. 첫째, 웹 마이닝 기법을 적용하여 웹상에서 보이는 고객의 행동 데이터로부터 간접적으로 선호도 정보를 예측한다. 전자상거래 환경은 대량의 상품과 고객이 존재하기 때문에 모든 상품에 대한 선호도를 고객들로부터 직접 획득하는 것은 현실적으로 힘들다. 따라서 웹 마이닝과 같은 목시적 평가법의 사용이 불가피하다. 둘째, 예측된 선호도 정보를 표현하기 위해 기존의 방법과 달리 서열 척도를 활용하였으며 서열화 된 순위를 생성함에 있어

서도 웹 환경에 보다 적합한 사전식 컨센서스 기법을 적용 하였다. 사전식 컨센서스 기법은 MCDM 분야 내에서 연구되어진 다른 컨센서스 기법들과 비교해서 구현 원리가 간단하면서도 선호도 평가 대상들 간에 동일한 순위를 최소화 한다는 장점을 가지고 있다. 따라서 대량의 상품에 대해서 실시간적인 추천 서비스가 요구되는 전자상거래 환경에 보다 적합할 수 있다.

실제 온라인 쇼핑물의 웹 데이터를 이용한 실험을 통해 제안된 방법론을 적용한 추천 시스템의 우수성을 입증 하였다. 또한 전자상거래 사이트에 대해 제안된 방법론을 적용할 경우 고객과 서비스 공급자 모두에게 혜택을 제공할 수 있다. 우선 고객은 그들이 원하는 상품을 보다 찾을 수 있으므로 보다 편리하게 상품 구매를 할 수 있다. 공급자 입장에서도 서비스에 만족하는 고객의 구매 전환율 (purchase conversion rate) 상승을 통해 수익의 증대를 기대할 수 있다.

하지만 실험의 결과가 단일한 특정 쇼핑물에 한정된 고객과 상품 정보를 근거로 한다는 점에서 연구의 한계점도 존재한다. 이를 보완하고 연구의 결과를 보다 일반화하기 위해서는 방문 고객의 규모, 매출액, 주력 상품의 종류 등에 따라 다양한 사이트들의 데이터들을 수집하고 이를 활용한 추가적인 성능 분석이 필요하다. 이러한 한계점은 향후의 추가적 연구 과정에서 보완되어야 할 것이다.

## 참 고 문 헌

- [1] F. Liu and H. J. Lee, "Use of social network information to enhance collaborative filtering performance," *Expert Systems with Applications*, Vol.37, pp.4772-4778, 2010.
- [2] W. D. Cook, "Optimal allocation of proposals to reviewers to facilitate effective ranking," *Management Science*, Vol.51, No.4, pp.655-661, 2005.
- [3] C. Wang and W. A. Wulf, "Towards a framework for security measurement," in: Proc.

- of National Information Systems Security Conference pp.522-533, 1997.
- [4] S. K. Lee, S. H. Kim, and Y. H. Cho, "Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations," *Information Sciences*, Vol.180, No.11, pp.2142-2155, 2009.
- [5] 박종학, 조윤희, 김재경, "사회연결망: 신규고객 추천문제의 새로운 접근법", *지능정보연구*, 제15권, 제1호, pp.123-139, 2009.
- [6] 김귀정, 김봉한, 한정수, "복합지식 기반 개인 맞춤형 지능화 추천 시스템", *한국콘텐츠학회논문지*, 제10권, 제8호, pp.26-31, 2010.
- [7] 여운동, 박현우, 권영일, 박영욱, "연구논문 추천 시스템의 전자도서관 적용방안", *한국콘텐츠학회논문지*, 제10권, 제11호, pp.10-19, 2010.
- [8] 정귀임, 박상성, 신영근, 장동식, "역전과 신경망을 이요한 개인 맞춤형 상품 추천 시스템 구축", *한국콘텐츠학회논문지*, 제7권, 제12호, pp.292-302, 2007.
- [9] T. Kamishima, "Nantonac collaborative filtering: recommendation based on order response," in: *Proc. of the Ninth International Conference on Knowledge Discovery and Data Mining*, pp.583-588, 2003.
- [10] T. Joachims, "Optimizing search engine using click through data," in: *Proc. of the Eighth International Conference on Knowledge Discovery and Data Mining*, pp.133-142, 2002.
- [11] C. Y. Kim, J. K. Lee, Y. H. Cho, and D. H. Kim, "VISCORs: a visual-content recommender for the mobile Web," *IEEE Intelligent Systems* Vol.19, No.3, pp.32-38, 2004.
- [12] Y. H. Cho and J. K. Kim, "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, Vol.26, No.2, pp.236-246, 2004.

- [13] M. Condorcet, "Essai sur L'Application de L'Analyse a la Probabilite des Decisions Rendues," a *La Pluralite des Voix*, Paris, 1785.
- [14] J. Borda, "Memoire sur les elections au scrutin," *Histoire de l'academie royale de science*, Paris, 1981.

#### 저 자 소 개

이 석 기(Seok Kee Lee) 정회원



- 2000년 2월 : 고려대학교 컴퓨터학과(이학사)
- 2002년 2월 : 한국과학기술원 경영공학과(공학석사)
- 2009년 8월 : 한국과학기술원 경영공학과(공학박사)
- 2010년 9월 ~ 현재 : 한양대학교 정보시스템학과 조교수
- <관심분야> : 추천 시스템, 데이터 마이닝, CRM

조 현(Hyeon Jo) 정회원



- 2004년 2월 : 한국과학기술원 산업경영학(학사)
- 2006년 2월 : 한국과학기술원 경영공학과(공학석사)
- 2006년 3월 ~ 현재 : 한국과학기술원 경영공학 박사과정
- <관심분야> : 추천 시스템, 소셜 네트워크

친 성 용(Sung Yong Chun) 정회원



- 1999년 2월 : 한국과학기술원 산업경영학(학사)
- 2001년 2월 : 한국과학기술원 경영공학과(공학석사)
- 2007년 8월 : 한국과학기술원 경영공학과(공학박사)
- 2009년 9월 ~ 현재 : 단국대학교 경영학부 조교수
- <관심분야> : 금융 마케팅, 하이테크 마케팅, CRM