

웹 사이트 이용 고객의 행동 정보를 기반으로 한 고객 선호지수 산출 방법*

서동렬

㈜넷스루 연구소
(dyseo@nethru.co.kr)

김두진

㈜넷스루 연구소
(doojin@nethru.co.kr)

윤정기

㈜넷스루 연구소
(jkyun@nethru.co.kr)

김재훈

㈜넷스루 연구소
(jhoon1980@nethru.co.kr)

문강식

㈜넷스루 연구소
(ksmoon@nethru.co.kr)

오재훈

㈜넷스루 연구소
(jaehoon@nethru.co.kr)

최근 웹 서비스의 발달과 함께 웹 콘텐츠를 다양하게 활용함으로써, 사용자의 경험을 기반으로 한 개인화 분석이 주목 받고 있다. 기존의 개인화 분석은 주로 데이터베이스의 데이터를 활용한 규칙 및 통계 모형을 기준으로 수행되고 있다. 이에 시장조사 소요기간에 따른 적시성을 반영하는데 어려움이 있었으며, 데이터베이스 적체 데이터가 고객 행동에 대한 결과였기 때문에 고객의 이용 특성을 반영하는데 한계가 지적되어 왔다. 그러나, 최근 고객의 사이트 방문에서부터 방문을 종료할 때까지의 모든 행동을 추적하고 분석하여 개인화된 서비스를 제공하기 위한 많은 연구와 상용화된 기술 개발이 진행되었다. 본 연구에서는 온라인상에서의 고객 행동을 웹 로그 분석을 이용하여 분석함으로써 고객의 행동정보를 U-Score(Usage Score, 이용지수), P-Score(Preference Score, 선호지수), M-Score(Mania Score, 마니아지수) 등 다양한 고객 선호지수를 도출하였다. 이러한 고객의 선호지수를 통해 웹 콘텐츠에 대한 고객의 선호정보를 파악함으로써, 고객에 대한 심도 있는 리포팅과 고객관계관리 가능하며 개인화 추천 서비스에 유용하게 사용할 수 있다.

※ 주제어: 고객관계관리, 선호도 산출, 개인화, 추천시스템

논문접수일: 2011년 2월 10일

논문수정일: 2011년 3월 11일

게재확정일: 2011년 3월 15일

* 본 연구는 서울시 산학연 협력사업(JP100080)의 연구비 지원으로 이루어 졌음

I. 서론

인터넷과 IT 기술의 발전은 e-Business 시장의 성장과 함께 이용 가능한 정보 및 콘텐츠들의 폭발적인 증가로 이어졌다. 이에 사용자들은 수많은 정보 및 콘텐츠를 검색하고 선별해야 하는 부담감을 가지게 되었다. 그러나 이러한 문제점들을 해결하기 위하여 사용자들이 원하는 정보 및 콘텐츠를 추천하기 위한 연구가 진행되어 왔다. 특히 늘어나는 웹 콘텐츠를 다양하게 활용함으로써 사용자들의 경험을 강화하고자 하는 움직임이 웹 콘텐츠 관리(Web Contents Management) 분야뿐 아니라 CRM 분야에서도 활발하게 이루어져 왔다. 이 가운데 가장 주목 받고 있는 핵심 트렌드로는 웹 분석을 기반으로 한 개인화(Personalization) 서비스이다. 개인화 서비스는 이용자의 요구에 적합한 정보 및 콘텐츠를 추출하여 제공하는 방법으로, 온라인 비즈니스에 있어서 중요한 성공요인으로 인식되고 있다[7].

그러나 기존의 개인화 서비스는 시장 조사(Market Search)나 기업이 보유하고 있는 고객의 데이터베이스의 데이터를 활용한 규칙 및 통계 모형을 기준으로 수행되었다. 이에 따라 시장조사 소요 기간에 따른 적시성을 반영하는데 어려움이 있었고, 데이터베이스 적재 데이터가 고객 행동에 대한 결과 정보였기 때문에 고객의 다양한 행동 특성을 반영하는데 어려움이 있었다. 예를 들어, 단순히 구매 기록만을 분석할 경우, 고객이 상품을 구매하는 과정에서 참조했던 다른 상품들에 대한 관심 정보는 반영되지 않는다. 뿐만 아니라 종래에 있어서는 온라인상의 고객 행동이 복잡해져 감에 따라 동적으로 변화하는 고객의 전체적인 행동정보를 수집하지 못하

고 있으며, 고객의 선호에 영향을 미칠 수 있는 고객의 성향 정보가 적절하게 반영되지 못하는 실정이다.

본 연구에서는 온라인 웹 사이트에서 발생하는 Click-stream 데이터를 이용하여, 고객의 다양한 클릭 행동들이 반영된 고객-아이템 선호도를 분석하고자 한다. 이를 위하여 웹 사이트의 URL 패턴을 이용하여 고객 행동을 정의하고, 고객이 각 아이টে 대해 어떤 행동을 했는지를 분석하며, 고객이 각 아이টে 대하여 표현한 선호도를 정량적으로 측정하는 방법을 제시한다. 또한 고객의 아이টে 대한 선호도를 아이টে간 또는 고객간 특성 비교가 가능하도록 표현하기 위하여 U-Score(Usage Score, 이용지수), P-Score(Preference Score, 선호지수), M-Score(Mania Score, 마니아지수) 등 다양한 고객 지수를 개발하였다. 그리고 도출된 지수들을 실제 적용 함으로써 개발된 선호지수들의 유용성을 실증적으로 검토하였다.

II. 선행연구

2.1 웹 로그 분석

웹 로그(web log)란 웹 사이트 방문자들이 제품이나 서비스를 구매하는 과정을 통해 발생하는 데이터이다. 정보획득이나 구매를 목적으로 인터넷 사이트를 방문하는 방문자들은 로그의 형태로 사이트 내에 흔적을 남기는데 이러한 데이터를 기반으로 해서 다양한 정보를 추출해 내는 것이 웹 로그 분석이다[11].

초기의 웹 로그분석은 웹 사이트 상에서 고객의

행동을 분석하고, 이를 웹 사이트 디자인에 반영하기 위하여 사용되었다. 그러나 e-Business 성장과 함께 온라인 거래가 활성화되면서 웹 로그 분석은 고객의 행동패턴을 분석하고 구매를 유도하기 위한 마케팅 수단으로 이용되고 있다[1].

이재신(2003)은 웹 사이트 방문자들의 이동경로를 관찰하고 예측함으로써, 웹 사이트 방문자들에게 대한 마케팅 프로모션을 수립하였다. 장운경(2002)은 온라인 상에서 고객의 구매행동을 분류하여 상품구입 관계를 분석하였다[1][9][10]. 김의환 외(2008)은 사용자들의 검색어와 웹 사이트에서의 행동 패턴을 근거로 의미 있는 정보가 담긴 웹 페이지를 선별 후 자동 태깅을 통해 검색 효율을 높이는 방법을 제안하였다[4].

이와 같이 웹 로그분석은 웹 사이트 방문자들이 많이 본 페이지를 방문페이지로 변경하거나, 웹 사이트에서의 방문경로를 파악하여 상품을 배치하는 등 고객의 행동을 파악하고 마케팅을 수행하기 위한 방법으로 이용되고 있다.

2.2 e-CRM

CRM(Customer Relationship Management)은 1990대 이후, 기업의 가치를 향상시키기 위한 전략이 '상품중심'에서 '고객중심'으로 변화하면서 주요 경영전략으로 부각되었다[6]. 특히 IT 기술의 발전은 좀 더 세밀하고 전문화된 고객 중심 마케팅을 야기시켰다. 더욱이 인터넷의 활성화를 통하여 데이터베이스와 연계한 다양한 고객정보 분석기법이 개발되어 1:1 마케팅, 관계 마케팅 등으로 더욱 발전함으로써 e-CRM의 등장을 이끌게 되었다[2]. e-CRM은 인터넷을 활용한 온라인 상에서 고객관리를 통해 고객의 충성도를 높이고, 고객으

로부터 수익을 최대화하려는 일련의 활동이라고 할 수 있다[1]. 이러한 e-CRM 분결 결과를 온라인 서비스에 적용한 대표적인 성공사례로써 Amazon, CD Now, YouTube 등 해외 사이트 뿐 만 아니라 교보문고, 삼성 몰 등 국내 쇼핑 사이트들도 적용하였거나 적용하기 위한 노력을 기울이고 있다.

2.3 개인화 추천

추천 기술은 e-Business에서 개인화 마케팅을 제공하기 위한 유용한 도구중의 하나로써, 고객의 인구통계학적 정보 및 웹 로그분석, 구매이력 등을 활용하여 특정 고객에 대해 적합한 검색결과 및 상품을 추천하기 위하여 이용된다[7]. 이러한 추천 기술의 목적은 기본적으로 서비스 이용자의 과거 정보를 기반으로 하여 추천 정보를 제공함에 있다[3]. 즉, 서비스 이용자의 특성을 대상으로 하여 이와 관련된 다양한 통계와 분석 및 비교를 통해 개별화 할 수 있는 지식 및 규칙을 찾아내는 과정을 통하여 도출된 결과이다[8].

사용자의 아이템에 대한 확률적 선호도에 기반한 개인화 추천 기법과 사용자의 선호도 모델에 관한 선행 연구들이 있었으나[12][13], 실제 웹 사이트 이용 고객의 Click-stream 데이터로부터 선호도를 구하는 방법을 제시하지는 못하였다.

과거의 개인화 추천기술은 시장 조사(Market Search)나 기업이 보유하고 있는 고객의 이력 및 구매 데이터를 활용한 규칙 및 통계 모형을 기준으로 수행되어왔다. 이에 따라 시장조사 소요 기간에 따른 적시성을 반영하는데 어려움이 있었고, 데이터베이스 적재 데이터가 고객 행동에 대한 결과 정보였기 때문에 고객의 특성을 반영하는데 어려움이 지적 되었다.

최근의 온라인 추천 시스템은 대부분 고객의 구매이력자료를 바탕으로 추천하고 있으며, 구매이력이 없는 경우 추천 자체가 불가능한 문제점이 지적되어 왔다[5]. 또한, 고객의 선호도를 표현하는 방법에 있어서 이용률을 나타내는 절대적인 수치 외에 다양한 정보를 제공하지 못하는 문제점이 지적되고 있다.

III. 웹 사이트 이용 고객 선호지수 산출

3.1 웹 사이트 이용 고객 선호지수의 특징

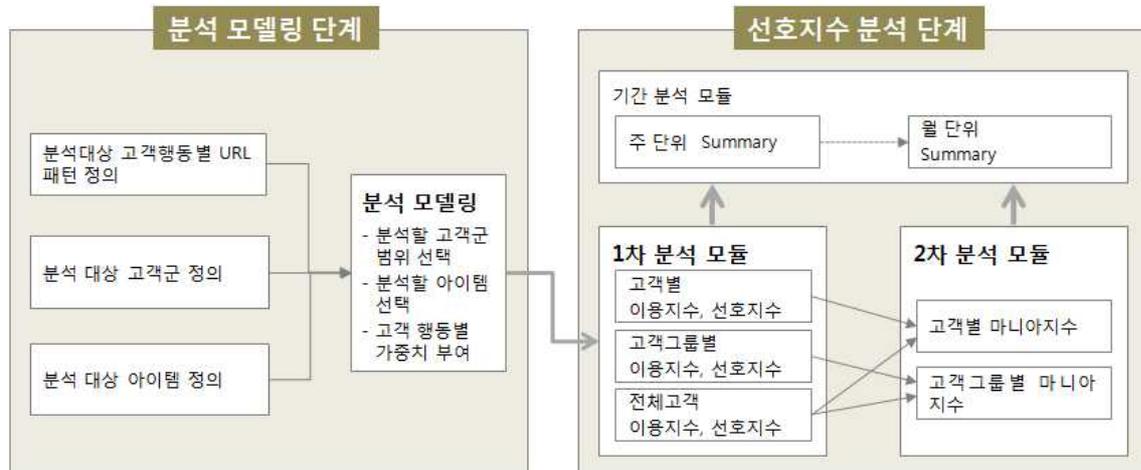
본 연구에서는 선행연구 및 현재 활용되고 있는 개인화 추천기술이 가지는 제한점을 해결할 수 있는 선호지수 추출 방법을 제안한다. 본 연구에서 제안하는 선호지수의 특징은 다음과 같다.

첫째, 선호지수 추출 시 고객 행동별 가중치를 부여하여, 사이트의 특성 및 특정 행동의 중요성을 반영하였다. 둘째, 콘텐츠 추천 시 추천 대상이 되는 고객을 개별고객, 고객그룹, 전체고객으로 구분함으로써, 마케팅 대상 고객(그룹)에 특화된 콘텐츠 추천이 가능하다. 마지막으로 산출하고자 하는 선호지수를 양적인 선호를 나타내는 U-Score, 질적인 선호도를 나타내는 P-Score, 고객의 특이성 향에 따른 선호도를 나타내는 M-Score 등 다양하게 표현하여, 고객 및 콘텐츠의 선호정보를 다각도로 분석 및 이용할 수 있도록 하였다.

3.2 웹 사이트 이용고객 선호지수의 산출방법

〈그림 1〉은 본 연구에서 제안하는 웹 사이트 이용고객 선호 지수 산출을 위한 시스템 구조이다. 시스템은 분석 모델링 모듈, 선호지수 분석 모듈로 구성된다.

〈그림 1〉 시스템 구조도



3.2.1 분석 모델링

분석 모델링 모듈은 선호지수 분석을 위하여 분석 대상의 정보를 계획하고 등록하는 역할을 수행하며, 본 논문에서 제안하는 선호도 분석 모델은 고객행동-고객군-아이템으로 구성된다. 이를 위하여 먼저 URL 표현식을 이용하여 분석 대상 URL을 패턴화하고 추출할 코드를 정의한다.

이때 URL 표현식은 웹로그 분석 대상이 되는 URL 패턴 정규표현식을 의미한다. 웹 로그에 기록되는 URL은 고객이 클릭한 웹 페이지를 의미하며, 이 URL 정보를 해석하면 고객이 어떤 아이템에 대하여 어떤 행동을 했는지를 알 수 있다. 또한 각각의 URL 패턴에 대해서 행동 가중치를 부여함으로써 각 고객 행동의 중요성을 선호도에 반영한다. 예를 들어, 상품을 상세보기한 행동에 비하여 상품을 장바구니에 담은 행동은 고객이 해당 상품에 대해 더 큰 선호도를 표현했다고 볼 수 있다.

분석 대상 고객군 정의 단계에서는 웹 사이트를 이용하는 고객의 그룹을 설정한다. 즉, 개별고객 및 전체고객뿐 아니라 고객의 데모그래픽 정보, 행동정보, 선호정보 등 다양한 정보를 조합하여 분석하고자 하는 특정 고객그룹을 설정할 수 있다.

분석 대상 고객군을 정의한 후에는 분석할 아이템의 범위를 정의한다. 예를 들어, 상품에 대한 선호도를 구할 것인지, 상품이 속한 카테고리에 대한 선호도를 구할 것인지, 어떤 상품들에 대해서 선호도를 구할 것인지를 결정한다.

3.2.2 선호지수 분석

선호지수 분석 모듈은 분석 모델링 모듈에서 전달된 정보를 해석하고 분석하는 역할을 수행한다. 이를 위하여 분석 전처리 단계 및 선호지수 분석 단계를 수행한다. 분석 전처리 단계에서는 분석 모

델링 모듈에서 설정된 정보를 읽어와 URL 표현식을 해석하여 사용자가 어떤 행동을 했는지를 인식하고, URL 파라미터에 포함된 아이템 코드 정보를 추출하여 웹 로그와 외부 데이터를 분석에 사용할 수 있는 형태로 변환한다. 전 처리된 분석 대상의 웹 로그정보 및 외부 데이터(상품 코드 데이터 및 고객 데이터 등)는 다음 단계의 선호지수 분석 단계로 전달된다. 선호지수 산출단계에서는 본 연구에서 제안한 선호지수를 분석하여 추출한다. 이러한 선호지수 산출 단계는 개별고객, 고객그룹, 전체고객의 U-Score, P-Score를 추출하는 1차 분석 모듈과, M-Score를 추출하는 2차 분석 모듈, 그리고 상위기간에 대한 지수를 산출하는 기간 분석 모듈로 구성된다.

① U-Score, P-Score

고객의 아이템 이용현황을 바탕으로 고객의 콘텐츠에 대한 이용 정도를 파악할 수 있는 U-Score와 고객이 이용한 다른 콘텐츠 대비 해당 콘텐츠의 상대적인 선호 정도를 파악할 수 있는 P-Score를 산출한다. 이러한 1차 분석모듈은 개별고객/고객그룹/전체그룹 산출 모듈로 구분되며, 각 모듈을 통하여 개별고객/고객그룹/ 전체고객의 U-Score 및 P-Score를 산출한다.

먼저 개별고객의 U-Score 계산식은 수식(1)과 같이 고객이 접근한 아이템에 대한 행동별 이용횟수를 계산하고, 마케터가 생각하는 각 고객행동의 중요성을 가중치로 반영한 Weighted-sum으로 정의한다. 따라서 U-Score는 사용자가 특정 아이템을 많이 이용할수록, 또한 가중치가 높은 행동을 많이 할수록 높아진다.

개별 고객의 특정 아이템에 대한 P-Score는 해당 고객이 각기 다른 콘텐츠에 대해 각각 어느 정

도의 선호도를 가지는가를 측정하는 지표이다. 고객이 접근한 모든 콘텐츠의 이용 정도에 대비하여 특정 콘텐츠에 대한 상대적 선호도를 나타내며, 그 계산식은 수식(2)와 같이 해당 고객이 이용한 모든 아이тем들의 U-Score 평균값 대비 해당 아이тем의 U-Score 비중으로 정의한다.

어느 고객의 특정 아이тем에 대한 P-Score가 1보다 작을 경우, 이 고객은 다른 아이тем들에 비하여 해당 아이тем을 덜 선호한다고 해석할 수 있으며, 반대로 1보다 크면 클수록 다른 아이тем에 비하여 해당 아이тем을 상대적으로 더 선호한다고 해석할 수 있다.

$$S_u^i = \sum w_a * h_a \quad [1]$$

S_u^i : 사용자 u의 콘텐츠 i에 대한 U-Score
 w_a = 행동(Action)a에 대한 선호도 가중치
 h_a = Action a에 대한 사용자 u의 히트(Hit)수

$$P_u^i = \frac{S_u^i}{average_{count(I_u)}(S_u)} \quad [2]$$

P_u^i : 사용자 u의 콘텐츠 i에 대한 P-Score
 S_u : 사용자 u의 Usage가 있는 콘텐츠들의 U-Score
 h_a = Action a에 대한 사용자 u의 히트(Hit)수

도출된 고객별 U-Score 및 P-Score 결과는 고객그룹 산출 모듈 및 전체고객 산출 모듈로 전달되어 고객그룹과 전체고객의 U-Score 및 P-Score 산출에 적용된다. 고객그룹의 콘텐츠 별 U-Score 계산식은 수식(3)과 같이 개별 고객의 해당 콘텐츠에 대한 U-Score의 합으로 산출된다. 고객그룹의 P-Score는 고객그룹이 접근한 모든 콘텐츠의 이용정도에 대비하여 해당 콘텐츠에 대한 상대적 선호도를 나타내며, 수식(4)와 같이 해당 고객그룹이 이용한 모든 아이тем들의 U-Score 평균값 대비 해당 고객 그룹의 U-Score 비중으로 정의한다.

$$S_{seg}^i = \sum_{u \in seg} S_u^i \quad [3]$$

S_{seg}^i : 해당 고객그룹의 콘텐츠 i에 대한 U-Score

$$P_{seg}^i = \frac{S_{seg}^i}{average_{count(I_{seg})}(S_{seg})} \quad [4]$$

P_{seg}^i : 해당 고객그룹의 콘텐츠 i에 대한 P-Score
 S_{seg} : 해당 고객그룹의 Usege가 있는 콘텐츠들의 U-Score
 $count(I_{seg})$: 해당 고객그룹의 Usege가 있는 모든 콘텐츠 수

마지막으로 전체고객의 콘텐츠 별 U-Score는 수식(5)와 같이 개별 고객의 해당 콘텐츠에 대한 U-Score의 합으로 산출된다. 전체 고객의 콘텐츠 별 P-Score는 다음의 수식(6)과 같이 콘텐츠별 U-Score와 모든 콘텐츠의 이용지수 평균을 근거로 산출된다.

$$S_U^i = \sum_{u \in U} S_u^i \quad [5]$$

S_U^i : 전체 고객의 콘텐츠 i에 대한 U-Score

$$P_U^i = \frac{S_U^i}{average_{count(I_U)}(S_U)} \quad [6]$$

P_U^i : 전체 고객의 콘텐츠 i에 대한 P-Score
 S_U : Usege가 있는 콘텐츠에 대한 모든 사용자들의 U-Score
 $count(I_U)$: Usage가 있는 모든 콘텐츠 수

② M-Score

마니아 지수는 해당 아이тем에 대하여 특정 고객 또는 고객 그룹이 다른 고객 또는 고객 그룹에 대비하여 상대적으로 어느 정도의 선호도를 나타내는지를 표현한다. 즉, 타 고객 또는 타 고객그룹에 대비하여 해당 콘텐츠를 많이 이용할수록 M-Score는 높게 도출된다. 개별고객 M-Score는 수식(7)과 같이 앞서 도출된 개별고객의 U-Score와 전체고객의 U-Score를 근거로 산출되며, 특정 아이тем을 이용한 모든 고객들의 U-Score 평균값 대비 해당 고객의 U-Score 비중으로 정의한다. 고객

롭의 M-Score는 수식(8)과 같이 해당 아이템을 이용한 모든 고객들의 U-Score 평균값 대비 해당 고객 그룹에서 해당 아이템을 이용한 고객들의 U-Score 평균값의 비중으로 정의한다.

$$M_u^i = \frac{S_u^i}{\text{average}_{count(u \in U^i)}(S_U^i)} \quad [7]$$

M_u^i : 사용자 u의 콘텐츠 i에 대한 M-Score
 S_u^i : 콘텐츠 i에 대한 Usage가 있는 모든 고객의 U-Score
 $count(u \in U^i)$: 콘텐츠 i에 대한 Usage가 있는 모든 고객수

$$M_{seg}^i = \frac{\text{average}_{count(u \in seg^i)}(S_{seg}^i)}{\text{average}_{count(u \in U^i)}(S_U^i)} \quad [8]$$

M_{seg}^i : 해당 고객그룹의 콘텐츠 i에 대한 M-Score
 $count(u \in seg^i)$: 해당 seg에서 콘텐츠 i에 대한 Usage가 있는 고객수

어느 고객의 특정 아이템에 대한 M-Score가 1보다 작을 경우, 이 고객은 다른 고객들에 비하여 해당 아이템을 덜 선호한다고 해석할 수 있으며, 반대로 1보다 크면 클수록 다른 고객들에 비하여 해당 아이템을 상대적으로 더 선호한다고 해석할 수 있다.

P-Score가 특정 고객의 아이템간 선호도 관계를 나타내는 반면, M-Score는 특정 아이템에 대하여 고객들 간의 선호도 관계를 나타낸다. 두 Score 모두 1보다 큰 값을 가질 때 평균 대비 선호 우위를 가진다고 해석할 수 있으며, 1보다 크면 클수록 상대적인 선호도 편중 성향이 크다고 볼 수 있다.

③ 기간 분석

기간분석은 개별고객의 일별 U-Score를 바탕으로 주별 U/P/M-Score와 월별 U/P/M-Score를 산출한다. 주별, 월별 선호지수 산출방법은 수식 [1]~[8]과 동일하며, 산출 기간에 따라 Usage 기록이 있는 집계 대상 고객 및 아이템 집합이 달

라진다. 주별 또는 월별 U-Score는 해당 기간 동안의 고객-아이템 U-Score의 합으로 산출되며, 이렇게 얻은 U-Score 값을 이용하여 해당 기간 동안의 P/M-Score를 계산한다.

IV. 웹 사이트 이용고객 선호지수 분석

4.1 선호지수

본 연구에서는 앞서 도출된 선호지수 및 산출과정에 따른 실증분석을 실시하였다. 분석 결과의 직관성을 위하여 <표 1>과 같이 10명의 웹 사이트 고객에 대해 10개 아이템 이용현황을 활용하였다. 또한 고객그룹별 지수산출을 위하여 고객의 성별을 활용하였다.

<표 1>의 각 Data Cell은 수식(1)을 활용하여 특정 고객이 특정 아이템을 이용한 횟수에 이용 행동별 가중치를 적용한 값(Weighted Sum)으로서, 해당 고객의 해당 아이템에 대한 U-Score를 나타낸다. <표 2>에서는 수식(2)를 활용하여 ID가 11인 고객이 이용한 각 아이템들에 대하여 P-Score를 산출한 결과를 예시하였다.

U-Score는 개별고객이 해당 콘텐츠를 얼마나 이용하는지에 대한 양적인 선호도를 나타내는 절대지수이다. 또한 P-Score는 고객이 사용한 모든 콘텐츠의 평균 이용값 대비 해당 콘텐츠의 U-Score를 바탕으로 도출되는 결과로써, P-Score를 기준으로 해당 콘텐츠에 대한 선호정도를 파악할 수 있다.

이러한 개별고객의 U-Score 및 P-Score는 고객그룹과 전체고객의 U-Score 및 P-Score 산출에 적용된다. 수식(3)(4)가 적용된 고객그룹별 U-Score

〈표 1〉 분석 데이터

User			Item별 Weighted Sum														
ID	Sex	Age	I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅
11	M	10	11	4	4	3	3	3									
12	M	10	8	1		1			4	2							
13	F	10	6	2							4	1					
14	F	10	6										3				
15	M	40	5					1			11			1			
16	M	70	5	3							1						
17	X	X		10											2		
18	M	40		9												6	
19	M	30		10				2								1	4
20	F	50	4	7					1							1	

〈표 2〉 개별고객의 P-Score

User		Usage		Item별 P-score														
ID	Sum	Avg	I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅	
11	28	4.67	2.36	0.86	0.86	0.64	0.64	0.64										

〈표 3〉 고객그룹별 U-Score 및 P-Score

Seg.	고객수	Item별 U-Score															
		I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅	
남자	6	11+8+5+5 =29	4+1+3+9+10 =27	4	3+1 =4	3	3+1+2 =6	4	2	11+1 =12				1		6+1 =7	4
여자	3	6+6+4 =16	2+7 =9						1		4	1	3			1	
기타	1		10												2		

Seg.	U-Score		Item별 P-Score														
	Sum	Avg.	I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅
남자	103	8.58	29 /8.58 =3.38	27 /8.58 =3.15	4 /8.58 =0.47	4 /8.58 =0.47	3 /8.58 =0.35	6 /8.58 =0.69	4 /8.58 =0.47	2 /8.58 =0.23	12 /8.58 =1.40			1 /8.58 =0.12		7 /8.58 =0.82	4 /8.58 =0.47
여자	35	5	16 /5 =3.2	9 /5 =1.8					1 /5 =0.2		4 /5 =0.8	1 /5 =0.2	3 /5 =0.6			1 /5 =0.2	
기타	12	6		10 /6 =1.67											2 /6 =0.33		

및 P-Score 결과는 〈표 3〉과 같다. 이러한 성별 분류를 통한 지수 산출과정은 하나의 예로써, 고객의 데모그리픽 정보, RFM 정보, 방문행동특성, 선호도 등을 이용한 다양한 조합들에 대한 고객그룹 정의가 가능하다. 이와 같이 도출된 고객그룹별

U-Score 및 P-Score는 특정 고객그룹의 콘텐츠 이용량 및 선호도를 파악할 수 있는 자료로 활용 가능하다.

고객그룹별 U-Score 및 P-Score 산출 후 수식 (5)(6)을 통하여 〈표 4〉와 같이 전체 고객의 U-

〈표 4〉 전체고객의 U-Score 및 P-Score

Item별 U-Score														
I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅
11	4													
+8	+1													
+6	+2													
+6	+3													
+5	+10													
+5	+9							4					6	
+4	+10		3		+1	4		+11					+1	
+4	+7		+1		+2	+1		+1					+1	
=45	=46	4	=4	3	=6	=5	2	=16	1	3	1	2	=8	4
Avg. = (45+46+4+4+3+6+5+2+16+1+3+1+2+8+4)/15 = 150/15=10														
Item별 P-Score														
I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅
45/10	46/10	4/10	4/10	3/10	6/10	5/10	2/10	16/10	1/10	3/10	1/10	2/10	8/10	4/10
=4.5	=4.6	=0.4	=0.4	=0.3	=0.6	=0.5	=0.2	=1.6	=0.1	=0.3	=0.1	=0.2	=0.8	=0.4

Score 및 P-Score를 도출하게 된다. 전체 고객의 U-Score 및 P-Score는 마케터에게 웹 사이트에서 제공되는 콘텐츠의 전체적인 이용량 및 전체적인 콘텐츠 선호도를 제공하며, 이는 고객 마케팅에 있어 중요한 판단 기준으로 활용 가능하다.

M-Score는 수식(7)을 활용하여 〈표 5〉와 같이 산출된다. 〈표 5〉를 살펴보면 다른 사용자들에 비하여 특정 콘텐츠에 대한 이용 비중이 높은 마니아 사용자들을 상대적인 지수로 쉽게 파악할 수 있다.

고객그룹별 M-Score는 수식(8)을 통하여 산출되며, 그 결과는 〈표 6〉과 같다. 이를 자세히 살펴보면, 고객그룹별 M-Score는 전체 고객의 해당 아이템 별 U-Score 평균값을 근거로 도출됨을 알 수 있다. 이는 개별고객과 마찬가지로 고객그룹의 콘텐츠 이용 성향을 파악할 수 있는 자료로써 고객관리 및 마케팅 측면에서 유용한 정보로 활용될 수 있다.

4.2 선호지수의 활용 방안

4.2.1 고객관점 선호정보의 제공

본 연구에서 제안하는 웹 사이트 고객의 선호지수는 고객의 웹 사이트 이용정보를 바탕으로 개별 고객, 전체고객, 고객그룹의 콘텐츠에 대한 U-Score와 P-Score를 분석하여 그에 따른 선호정보를 제공한다. 제공되는 U-Score와 P-Score는 고객이 이용한 콘텐츠의 개수와 이용 정도에 따라 다양한 수치로 나타난다. 특히 P-Score의 경우 고객이 이용한 콘텐츠들 사이의 상대적인 선호도를 나타내므로, 특정 고객을 대상으로 콘텐츠를 추천함에 있어 유용하게 사용될 것으로 사료된다.

뿐만 아니라 고객에게 콘텐츠를 추천하는 경우, 고객의 콘텐츠별 P-Score와 함께 고객이 이용한 전체 콘텐츠의 U-Score 평균값을 고려하면 좀 더 의미 있는 추천이 가능하다.

특히 구매기록이 없거나 신뢰성 있는 사용자 평점 정보를 얻을 수 없는 사이트의 경우, 사이트 고객

〈표 5〉 개별고객의 M-Score

Item List		I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅	
전체 고객	Sum	45	46	4	4	3	6	5	2	16	1	3	1	2	8	4	
	고객수	7	8	1	2	1	3	2	1	3	1	1	1	1	3	1	
	Avg	6.43	5.75	4	2	3	2	2.5	2	5.33	1	3	1	2	2.67	4	
개별 고객	11	11/6.43 =1.71	4/5.75 =0.7	4/4 =1	3/2 =1.5	3/3 =1	3/2 =1.5										
	12	8/6.43 =1.24	1/5.75 =0.17		1/2 =0.5			4/2.5 =1.6	2/2 =1								
	13	6/6.43 =0.93	2/5.75 =0.35							4/5.33 =0.75	1/1 =1						
	14	6/6.43 =0.93										3/3 =1					
	15	5/6.43 =0.78					1/2 =0.5			11/5.33 =2.06			1/1 =1				
	16	5/6.43 =0.78	3/5.75 =0.52							1/5.33 =0.19							
	17		10/5.75 =1.74												2/2 =1		
	18		9/5.75 =1.57													6/2.67 =2.25	
	19		10/5.75 =1.74				2/2 =1									1/2.67 =0.37	4/4 =1
	20	4/6.43 =0.62	7/5.75 =1.22						1/2.5 =0.4							1/2.67 =0.37	

〈표 6〉 고객그룹별 M-Score

Item List		I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	I ₉	I ₁₀	I ₁₁	I ₁₂	I ₁₃	I ₁₄	I ₁₅
전체 고객	U-score	45	46	4	4	3	6	5	2	16	1	3	1	2	8	4
	고객수	7	8	1	2	1	3	2	1	3	1	1	1	1	3	1
	평균	6.43	5.75	4	2	3	2	2.5	2	5.33	1	3	1	2	2.67	4
남	U-score	29	27	4	4	3	6	4	2	12			1		7	4
	고객수	4	5	1	2	1	3	1	1	2			1		2	1
	평균	7.25	5.4	4	2	3	2	4	2	6			1		3.5	4
	M-score	7.25/6.43 =1.13	5.4/5.75 =0.94	4/4 =1	2/2 =1	3/3 =1	2/2 =1	4/2.5 =1.6	2/2 =1	6/5.33 =1.13			1/1 =1		3.5/2.67 =1.31	4/4 =1
성 여	U-score	16	9					1		4	1	3			1	
	고객수	3	2					1		1	1	1			1	
	평균	5.33	4.5					1		4	1	3			1	
	M-score	5.33/6.43 =0.83	4.5/5.75 =0.78					1/2.5 =0.4		4/5.33 =0.75	1/1 =1	3/3 =1			1/2.67 =0.37	
기 타	U-score		10											2		
	고객수		1											1		
	평균		10											2		
	M-score		10/5.75 =1.74												2/2 =1	

의 웹로그 정보를 이용하여 Collaborative Filtering 기법을 적용하고자 할 때 Implicit Rating[14] 값으로 유용하게 활용 가능하다.

4.2.2 콘텐츠관점 선호정보의 제공

M-Score는 콘텐츠를 이용한 전체 고객들 사이의 상대적인 선호도를 나타내므로, 특정 콘텐츠에 대한 편중된 선호도를 가지는 고객 및 고객그룹을 추출하여 타겟 마케팅으로 활용 가능하다.

특정 콘텐츠를 선호하는 대상 고객을 추천하는 경우 해당 콘텐츠에 대한 고객별 M-Score와 함께 해당 콘텐츠를 이용한 전체 고객의 U-Score 평균값을 고려하면 좀 더 의미 있는 추천이 가능하다.

높은 U-Score 평균값을 갖는 콘텐츠는 대부분의 고객이 강한 선호를 보인 콘텐츠에 해당되는 것으로, M-Score가 높은 특정 고객이 특별히 선호하기 보다는 일반적으로 전체 고객들에게 인기 있는 콘텐츠로 볼 수 있다. 이러한 콘텐츠는 다양한 고객에게 추천 콘텐츠로 제공할 수 있다.

반면, 낮은 U-Score 평균값을 보이는 콘텐츠는 롱테일(long tail)에 속하는 콘텐츠로서 해당 콘텐츠에 대해 높은 M-Score를 갖는 고객이 다른 고객에 비해 강한 선호를 나타내는 콘텐츠다. U-Score 평균값이 낮은 콘텐츠를 선호하는 대상 고객을 추

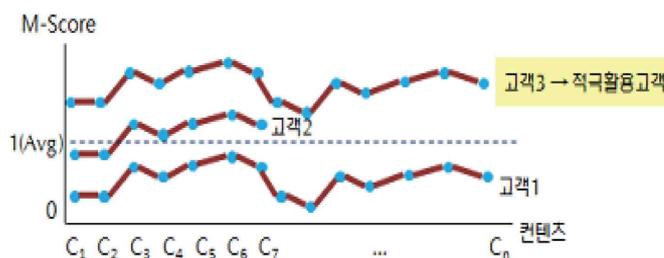
천하고자 할 때는 해당 콘텐츠에 대해 높은 M-Score를 갖는 고객을 선택하도록 하는 것이 바람직 할 것으로 사료된다.

4.2.3 선호지수 간 관계정보의 제공

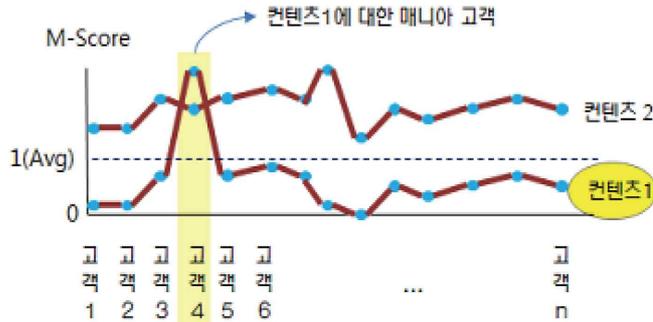
본 연구에서 제안하는 선호지수들을 통하여 고객별, 콘텐츠별 산출결과와 패턴을 분석함으로써 의미 있는 결과를 파악할 수 있다. 예를들어 고객별 콘텐츠에 대한 M-Score 패턴을 분석하고 콘텐츠 별 고객들의 M-Score 분포를 분석하여 그 분석 결과를 근거로, 콘텐츠를 적극적으로 활용하는 고객과 롱테일(long-tail) 타겟 고객을 찾아낼 수 있다.

〈그림 2〉와 같이 고객별 이용 콘텐츠의 M-Score 분포를 해석하면 다수의 콘텐츠에서 평균값 이상의 M-Score를 갖는 고객을 추출할 수 있다. 고객별 M-Score 평균값이 상대적으로 높고, 이용 콘텐츠 중 M-Score가 평균값 이상인 콘텐츠가 많은 경우 해당 고객은 콘텐츠 이용률이 높은 적극 활용 고객으로 유추할 수 있다. 이러한 콘텐츠 적극 활용 고객은 콘텐츠 이용과 평가에 큰 영향을 미치는 중요한 고객으로 고객 가치가 높은 것으로 판단할 수 있다. 따라서, 이러한 콘텐츠 적극 활용 고객은 이벤트나 캠페인 안내 메일, 상품평 참여 유도 쿠폰 발행 등의 방법으로 타겟 마케팅에 적극적으로 활

〈그림 2〉 고객별 이용 콘텐츠에 대한 M-Score 분포



〈그림 3〉 콘텐츠별 고객들의 M-Score 분포도.



용할 것으로 사료된다.

〈그림 3〉은 콘텐츠별 고객들의 M-Score 분포를 나타낸 것이다. 콘텐츠별로 최대 M-Score 값을 제외한 나머지 M-Score들의 평균값을 구했을 때, 전체적인 마니아지수 평균값이 낮은 콘텐츠에 대해서, 최대 M-Score가 높은 값을 보이는 고객들을 해당 콘텐츠의 매니아 고객으로 유추할 수 있다.

즉, 롱테일 콘텐츠 중 다른 고객에 비해 훨씬 높은 선호를 보이는 고객을 찾아 매니아 고객으로 해석할 수 있다. 이러한 매니아 고객은 해당 콘텐츠 혹은 해당 콘텐츠의 연관 콘텐츠에 대한 이용이나 구매 등에 적극적인 고객으로 간주할 수 있다. 매니아 고객의 경우 오프라인 메일이나 아웃바운드 마케팅과 같은 보다 구체적인 타겟 마케팅 방법을 사용하여 그 반응률을 높이는 것이 효과적일 것으로 보인다.

V. 결론

본 연구에서는 온라인상에서의 고객 행동을 웹 로그 분석을 이용하여 분석함으로써, 고객의 행동정보

를 U-Score(Usage Score, 이용지수), P-Score (Preference Score, 선호지수), M-Score (Mania Score, 매니아지수) 등의 다양한 고객 선호지수로 도출하였다. 그리고 도출된 선호지수들을 실제 적용함으로써, 개발된 선호지수들의 유용성을 실증적으로 검토하였다.

검토결과 고객 및 고객그룹별 콘텐츠에 대한 상대적인 선호도를 파악하고, 특정 콘텐츠를 선호하는 고객 및 고객그룹을 추출할 수 있었다. 뿐만 아니라 고객별, 콘텐츠 별 콘텐츠에 대한 결과를 비교함으로써 콘텐츠를 적극적으로 이용하는 고객과 롱테일(long-tail) 타겟 고객을 찾아낼 수 있으며, 특정 콘텐츠를 적극적으로 활용하는 중요 고객을 파악할 수 있었다.

이와 같이 본 연구에서 제안하는 선호지수는 고객의 콘텐츠 이용 현황 및 성향을 다양한 지표로 파악할 수 있는 수단을 제공한다. 또한, 웹 사이트를 이용하는 고객의 행동정보를 이용하여 콘텐츠별 다양한 선호 지수를 생성함에 있어서, 정성적인 고객의 성향을 정량적인 지수를 근거로 도출하여 마케팅의 의사결정을 지원하고, 사이트의 특성 및 콘텐츠의 중요성에 따른 가중치를 반영함으로써, 고객 관리의 효율성이 향상된 차별화된 마케팅이 가

능할 것으로 사료된다.

본 연구의 실증 결과는 제한된 고객 집합의 이용 내역을 바탕으로 한 것으로서, 보다 광범위한 데이터 집합을 통한 실증 연구가 필요하다.

또한 본 연구에서 제시한 다양한 고객-컨텐츠 선호도 지표를 활용한 개인화 추천 시스템 및 추천 기법에 대한 추가적인 연구가 필요할 것으로 보이며, Collaborative Filtering 방식의 추천 시스템을 위한 입력 데이터로 활용될 수 있을 것이다.

참고문헌

- [1] 고재문, 서준용, 김운식, “수주생산기업 B2B에서 e-CRM을 위한 웹 로그 분석,” **산업공학**, 제18권, 제2호, pp. 205-220, 2005.
- [2] 김미지, 김강, “e-CRM 구성요인이 e-쇼핑몰 고객만족 및 브랜드 충성도에 미치는 영향에 관한 연구,” **한국컴퓨터 정보학회**, 제16권, 제1호, pp.211-218, 2011.
- [3] 김용, 김문석, 김윤범, 박재홍, “이용자 이용행위 및 컨텐츠 위치정보에 기반한 개인화 추천방법에 관한 연구,” **한국정보관리학회**, 제26권 제1호, 2009, pp. 81-105, 2009.
- [4] 김의환, 이승훈, 윤태복, 이지형, “사용자 행동과 웹 페이지 분석을 통한 자동 태깅 시스템,” **한국지능시스템학회 추계학술대회**, 제18권, 제2호, 2008.
- [5] 박기남, 강낙중, “고객자본에 기초한 e-CRM기반의 상품추천 시스템개발에 관한 연구,” **인터넷 전자상거래 연구**, 제10권, 제1호, pp125-144, 2010.
- [6] 박소영, 배미현, 이춘수, “호텔기업의 e-CRM 비교를 통한 고객만족 전략에 관한 사례연구-FGI방법으로,” **인터넷 전자상거래 연구**, 제9권, 제1호, pp. 175-195, 2009.
- [7] 박수환, 김종우, 이홍주, 조남재, “전자상거래 개인화 추천을 위한 상품 카테고리 중립적 사용자 프로파일링,” **경영정보학연구**, 제16권, 제3호, pp.159~176, 2006.
- [8] 송창우, 김종훈, 정경용, 류중경, 이정현, “시맨틱 웹에서 개인화 플로파일을 이용한 컨텐츠 추천 검색 시스템,” **한국컨텐츠학회**, 제8권, 제1호, pp. 318~327, 2008.
- [9] 이재신, “데이터마이닝을 활용한 동적인 고객분석에 따른 고객관계관리 기법,” **한국지능정보시스템학회**, 제9권, 제3호, pp.23-47, 2003.
- [10] 장윤경, “로그분석을 통한 고객행동과 고객 가치와의 관계에 관한 연구,” 아주대학교 석사학위논문, 2002.
- [11] Cooley, R., Mobasher, B., & Srivastava, J., “Data preparation for mining world wide web browsing patterns,” *Knowledge and Information System*, Vol.1, No.1, pp.123-132, 1999.
- [12] Breese, J.S. et al., “Emperical analysis of predictive algorithms for collaborative filtering,” *Proc. Fourth Conf. Uncertainty in Artificial Intelligence*, 1998.
- [13] Jung, S.Y. et al., “A statistical model for user preference,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No.6, pp.834-843, 2005.
- [14] Herlocker, J.L., “Evaluating Collaborative Filtering Recommender Systems,” *ACM Trans. on Information Systems*, Vol. 22, No.1, pp.5-53, 2004.

Method for Preference Score Based on User Behavior

Seo, DongYal* · Kim, DooJin* · Yun, JeongKi* · Kim, JaeHoon* ·
Moon, KangSik* · Oh, JaeHoon*

Abstract

Recently with the development of Web services by utilizing a variety of web content, the studies on user experience and personalization based on web usage has attracted much attention. Majority of personalized analysis are have been carried out based on existing data, primarily using the database and statistical models. These approaches are difficult to reflect in a timely mannern, and are limited to reflect the true behavioral characteristics because the data itself was just a result of customers' behaviors. However, recent studies and commercial products on web analytics try to track and analyze all of the actions from landing to exit to provide personalized service. In this study, by analyzing the customer's click-stream behaviors, we define U-Score(Usage Score), P-Score (Preference Score), M-Score(Mania Score) to indicate variety of customer preferences. With the devised three indicators, we can identify the customer's preferences more precisely, provide in-depth customer reports and customer relationship management, and utilize personalized recommender services.

※ Key Words: Customer Relationship Management, Preference Scoring, Personalization, Recommender System

* Nethru Data Mining Lab