

# 온라인 쇼핑 콘텐츠의 접근 패턴 분석

유정민, 이문영, 권태경

서울대학교 컴퓨터 공학부

{jmyoo, mylee}@mmlab.snu.ac.kr, tkkwon@snu.ac.kr

## Analysis of Content Access Patterns for Online Shopping Contents

Jungmin Yoo, Munyoung Lee, Ted "Taekyong" Kwon

School of Computer Science and Engineering, Seoul National University

### 요약

인터넷의 발달과 함께 등장하게 된 온라인 쇼핑은 우리 생활에 중요한 부분을 차지하고 있다. 온라인 쇼핑 사용자와 매출의 규모는 크게 증가하였지만, 사용자들이 온라인 쇼핑 콘텐츠를 어떤 식으로 접근하고 어떻게 소비하는지는 다른 종류의 콘텐츠에 비해 상대적으로 적은 관심을 받아왔다. 본 논문에서는 온라인 쇼핑 사이트의 접속 데이터 로그를 분석하여, 사용자들의 온라인 쇼핑 콘텐츠의 접근 패턴 역시 Zipf 분포를 따라 소수의 콘텐츠에 대해 편중되어 있으며 동일한 콘텐츠를 반복해서 소비하는 특징이 존재하는 것을 확인하였다. 이러한 사용자들의 온라인 쇼핑 콘텐츠에 대한 접근 패턴은 온라인 쇼핑 사이트의 성능 개선에 이용되거나 특정 상품들에 대한 온라인 마케팅에 활용되어 경제적 효과를 얻을 수 있을 것으로 기대된다.

### I. 서론

인터넷의 발달과 함께 등장하게 된 온라인 쇼핑은 직접 매장에 방문하지 않고도 원하는 물건을 구매할 수 있는 편리함 때문에 매년 꾸준히 성장하여 점차 많은 사용자들이 온라인 쇼핑을 이용하고 있고, 오늘날에 이르러서는 우리 생활에 중요한 부분을 차지하게 되었다. 최근 조사결과에 따르면 2013년 한 해에만 온라인 쇼핑이 차지하는 전체 매출 규모가 전 세계적으로 1조 2330억 달러를 차지하였고, 2015년에는 1조 9220억에 다다를 것으로 예상되고 있다 [1].

이처럼 온라인 쇼핑의 규모는 급격히 성장하였고, 온라인 쇼핑을 이용하는 사용자의 수도 증가하였지만 사용자들이 온라인 쇼핑 콘텐츠를 어떤 식으로 접근하고, 어떻게 소비하는지에 대한 연구는 YouTube와 같은 멀티미디어 콘텐츠 제공 서비스나 트위터, Facebook 같은 소셜 네트워크 사이트 서비스와 비교하여 상대적으로 적은 관심을 받아왔다. 기존 연구에 따르면 YouTube 사이트의 경우, 사용자들은 일부 인기 있는 콘텐츠를 집중해서 이용하기 때문에 약 상위 10%의 비디오 콘텐츠가 전체 요청의 약 80%를 차지한다고 한다[2].

본 논문에서는 실제 사용자들이 온라인 쇼핑 사이트에 접속하여 소비활동을 한 데이터 로그를 이용하여 온라인 쇼핑 사용자들이 어떻게 온라인 쇼핑 콘텐츠에 접근하고, 소비(구매)하는지를 파악하고자 한다. 또한, 하루 동안 사용자들이 온라인 쇼핑 콘텐츠에 대한 접근이 어떤 특징을 보이는지를 분석하고자 한다.

### II. 본론

#### 2.1 데이터 셋 설명

본 논문에서 분석한 데이터 셋은 third-party 업체가 온라인 쇼핑 사이트로부터 수집한 로그이다. 이 데이터 셋은 사용자가 HTTP request를 통해 사이트에 접근할 때마다 기록된 정보 (third-party 쿠키, 식별 주소, request time)와 사용자가 탐색하거나 구매한 상품들과 관련한 정보를 포함한다. 표 1. 은 이 데이터 셋 형식의 일부를 설명한 표이다. 사용자 식별 주소는 privacy 보호를 익명화 처리되어 수집되었다. 페이지 정보 (page type) 정보는 상품이 탐색된 것인지 (Item), 구매가 완료된 것인지

(Purchase)를 나타내며, 본 논문에서는 페이지 정보를 이용하여 온라인 쇼핑 콘텐츠를 일반 콘텐츠 (Item)와 유료 콘텐츠 (Purchase)로 분류하였다.

본 논문에서는 하루 동안의 사용자 접근 패턴에 대해 알아보기 위해 2014년 3월 11일에 온라인 쇼핑 사이트에 접속한 데이터를 분석하였다. 분석이 진행된 날에는 총 4,307,293번의 콘텐츠 접근 및 소비가 있었다. 분석 샘플의 신뢰성을 높이기 위해, 서로 다른 50명 이상의 사용자에게 의해 접근 되고, 100번 이상 접근된 콘텐츠 (상품)들에 대해 필터링하였다.

client_id	익명화된 사용자 식별 주소
request_time	사용자의 HTTP request time
cookie	사용자의 cookie
page_type	페이지 정보 (Item, Purchase)
i	상품의 이름
p	상품의 가격
q	상품의 수량

표 1 온라인 쇼핑물 데이터 셋의 형식

#### 2.2 분석 및 결과

위 데이터 셋에 대해 사용자들이 하루 동안 온라인 쇼핑 콘텐츠에 접근하는 패턴에 어떠한 특징이 있는지 다음과 같이 분석하였다.

##### 2.2.1 일반 콘텐츠 및 유료 콘텐츠에 대한 사용자의 접근 분석

사용자는 온라인 쇼핑 콘텐츠 (상품)를 탐색하기도 하고, 구매를 하기도 한다. 본 절에서는 단순히 탐색해보는 상품과 실제 구매하는 상품에 대한 사용자의 접근 패턴을 각각 알아보았다. 이를 위해, 탐색한 상품은 일반 콘텐츠로, 구매까지 완료한 상품은 유료 콘텐츠로 간주하였다.

먼저 콘텐츠의 인기도에 따라 탐색 (또는 구매)된 빈도수가 관련이 있는지 살펴보기 위해, Zipf 분포와 비교해보았다. Zipf 분포란, 소수의 콘텐츠에 사용 빈도가 편중된 분포이다 [3]. 이 분포는 alpha 변수 값에 의해 그 편중 정도가 결정되며, alpha 값이 클수록 소수의 콘텐츠에 편중되는 정도가 크다. YouTube의 경우에는 약 상위 10%의 비디오 콘텐츠가 전체 요청의 약 80%를 차지하여 사용자들이 소수의 인기 있는 콘텐츠를 집중해서 이용하고 있고, 비디오 콘텐츠에 대한 요청은 alpha 값 1.0을 가지는 Zipf 분포와 유사한 특징을 보인다고 알려져 있다 [2].

사용자들이 온라인 쇼핑 콘텐츠를 사용하는 패턴에 있어서도 소수의 콘텐츠에 집중해서 소비하는지를 확인하기 위해 일반 콘텐츠와 유료 콘텐츠 각각에 대해 인기 있는 상위 100개의 콘텐츠를 추출하여 이 상위 100개의 콘텐츠에 대한 접근이 Zipf 분포를 따르는지 확인해보았다. 그림 1과 그림 2는 각각 탐색 및 구매 횟수로 콘텐츠에 순위를 부여하여 해당 콘텐츠에 대한 접근의 빈도수 분포를 Zipf 분포 그래프와 비교하여 도식화한 것이다. 그림 1과 그림 2에서 확인할 수 있듯이, 탐색 또는 구매된 상위 콘텐츠는 그 빈도수 분포가  $\alpha$ 값 0.4~0.6 정도의 Zipf 분포와 유사한 특징을 보인다. 이는 온라인 쇼핑 콘텐츠에 대한 접근은 순위에 따라 편중된 경향이 크진 않지만, 비디오 콘텐츠와 마찬가지로 Zipf 분포를 따라 소수의 콘텐츠에 대한 접근이 나머지 콘텐츠에 비해 더 많이 분포되어 있음을 알 수 있다.

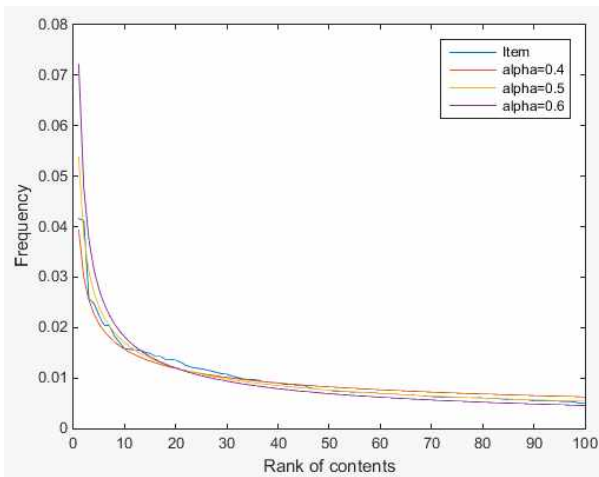


그림 1 일반 콘텐츠에 대한 사용자의 접근 분석

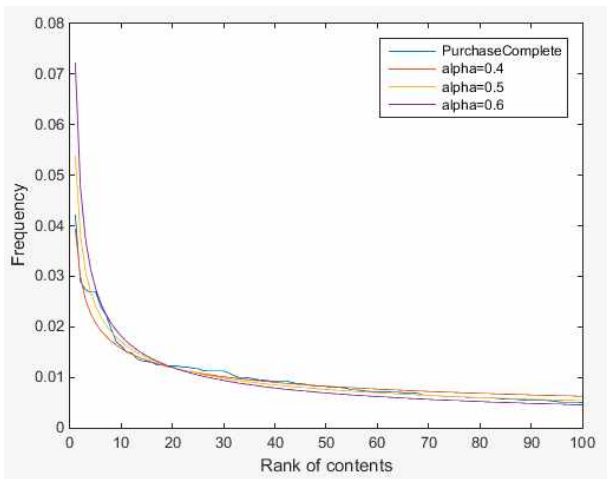


그림 2 유료 콘텐츠에 대한 사용자의 접근 분석

### 2.2.2 반복된 콘텐츠에 대한 사용자의 패턴 분석

기존 연구에 따르면, YouTube와 같은 비디오 콘텐츠 뿐만 아니라 Wikipedia 콘텐츠, Google Maps에서 사용자들이 클릭한 콘텐츠 등 다양한 종류의 콘텐츠들은 반복하여 소비되는 패턴을 보인다 [4]. 온라인 쇼핑 사용자들 역시 구매를 결정하기 전에 혹은 선호하는 상품의 정보를 더 얻거나 재구매하기 위해 동일한 상품을 여러 번 탐색할 수 있다. 본 절에서는 이러한 특성이 온라인 쇼핑에서도 존재하는지 확인하기 위해 온라인 쇼핑 콘텐츠가 검색되었을 때 해당 콘텐츠가 소비된 것으로 간주하여 온라인 쇼핑 콘텐츠 중 반복 소비되는 비율을 측정하였다.

그림 3은 온라인 쇼핑 데이터 셋의 반복 소비된 비율의 분포를 보여준다. X축은 사용자가 콘텐츠를 소비한 횟수 중에서 반복 소비한 횟수가 차지하는 비율이며, Y축은 해당 반복 빈도수를 갖는 전체 사용자의 비율을 뜻한다. 예를 들면, 반복 소비의 비율이 0.5인 사용자가 전체의 9% 정도를 차지하는데, 이는 9%의 소비자가 소비한 콘텐츠 중 50%가 반복되어 소비된 콘텐츠인 것을 의미한다. 이를 통해, 온라인 쇼핑 콘텐츠도 다른 다양한 종류의 콘텐츠들과 마찬가지로 반복 소비되는 현상이 존재함을 확인할 수 있었다.

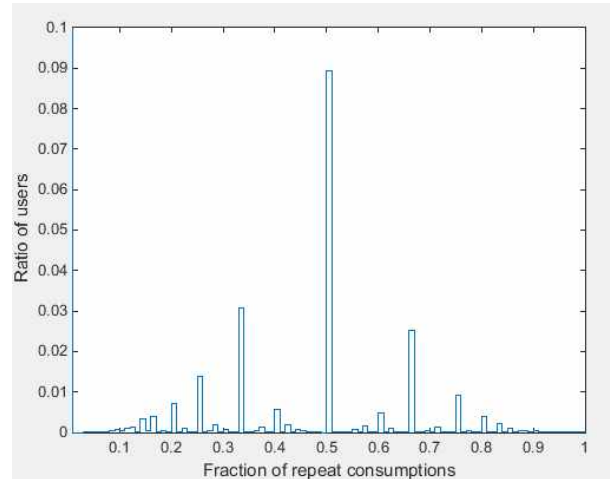


그림 3 반복된 콘텐츠에 대한 사용자의 분포

## III. 결론

지금까지, 온라인 쇼핑 데이터 셋을 분석하여 사용자들이 온라인 쇼핑 콘텐츠에 접근하는 패턴은 Zipf 분포를 따라 소수의 콘텐츠에 대해 편중되어 있으며, 동일한 콘텐츠를 반복하여 소비하는 특징이 있음을 확인하였다. 이러한 결과는 인기 있는 상품들을 효과적으로 프로모션 하거나, 사용자들의 반복 접근하는 특징을 고려하여 최근에 접근했던 상품에 대해 타겟 마케팅을 할 수 있어서 경제적 효과를 얻을 수 있을 것으로 기대한다. 향후 연구로, 이러한 사용자들의 접근 패턴을 모델링하여 추가적인 분석을 진행해 볼 계획이다.

## ACKNOWLEDGMENT

이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구 결과물임을 밝힙니다.  
(No.2014R1A2A2A01016562)

## 참고 문헌

- [1] B2C e-commerce sales worldwide from 2012 to 2018 (in billion U.S. dollars). Available: <http://www.statista.com/statistics/261245/b2c-e-commerce-sales-worldwide/>
- [2] M. Cha et al., "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System", in Proc. of ACM IMC, 2007.
- [3] George K. Zipf (1949) Human Behavior and the Principle of Least Effort. Addison-Wesley.
- [4] Ashton Anderson, Ravi Kumar, Andrew Tomkins, Sergei Vassilvitskii, "The Dynamics of Repeat Consumption", WWW 2014.