

IPTV 주문형 비디오 시청 순위 함수의 모수에 영향을 끼치는 요인 분석

*윤협상 이승재 정상현

KT 중앙연구소

*yoonhs@kt.com

Analysis on the parameters of IPTV VOD access rank function

*Yoon, Hyoup-Sang Lee, Sueng-Jae ***Joeng, Sang-Hyoun

Central R&D Lab., KT

요약

최근 널리 보급되고 있는 IPTV의 주요 서비스인 주문형 비디오 서비스의 시청 순위는 Zipf의 법칙을 따른다고 알려져 있으며 이를 기반으로 계층형 네트워크 스토리지 구조로 서비스를 제공하여 네트워크 비용과 스토리지 비용을 절감하고 있다. 그러나, 최근 연구에 의하면 미디어 시청 순위는 stretched exponential 함수에 더 근사함이 알려졌다. 본 연구에서는 국내에서 현재 상용 서비스 중인 대규모 주문형 비디오 서비스의 시청 순위-회수 데이터를 분석하여, 시청 순위로부터 stretched exponential 함수의 모수를 추정하여 시청 순위의 분포함수를 파악하고자 한다. 특히, 콘텐츠 파일 크기, 콘텐츠 파일 전달 방식, 콘텐츠 장르 등의 요인과 시청 순위 분포함수의 모수 사이에 상관관계가 있는지 분석하고자 한다.

1. 서론

현재 국내에서 IPTV 서비스를 제공하고 있는 사업자들은 주문형 비디오 서비스와 채널 서비스, 양방향 서비스를 주요 서비스로 제공하고 있다. 특히 주문형 비디오 서비스는 IPTV 서비스 중에서 가장 최초로 시작되었기 때문에 대중적인 인지도가 높고 널리 확산되었다. 예를 들어, KT의 QOOK TV는 100만 이상의 가입자에게 6만 편 이상의 콘텐츠를 서비스하고 있다.

주문형 비디오 서비스를 제공하기 위해서는 콘텐츠를 저장할 수 있는 스토리지 장비와 콘텐츠를 전송할 수 있는 네트워크 회선에 대한 투자가 필수적이다. 가입자와 콘텐츠의 수가 증가할수록 더 많은 투자가 수반되어야 하기 때문에 서비스 비용을 줄일 수 있는 주문형 비디오 서비스 제공 구조에 대한 연구가 계속 진행되어 왔다[1].

현재 국내 IPTV 서비스 제공 업체의 주문형 비디오 서비스는 시청이 빈번한 인기 콘텐츠를 사용자 근처에 위치한 스토리지에 미리 저장해 놓는 계층형 네트워크 스토리지 구조[2]를 사용하여 서비스되고 있다. 이 구조는 소수의 일부 콘텐츠에 대부분의 시청이 집중된다는 사실을 근거로 하고 있다. 이러한 특성을 Zipf의 법칙 [3, 4]으로 근사화하려는 연구가 진행되었다. Zipf의 법칙은 특정 콘텐츠의 시청회수 순위 $i(i=1, 2, \dots, N)$ 와 시청회수 y_i 사이에 $y_i \sim 1/i^\alpha$ (단, $0 < \alpha < 1$)인 관계가 존재함을 의미한다[3].

최근의 연구 결과에 의하면 주문형 비디오 시청 회수의 순위 함수는 Zipf의 법칙보다는 stretched exponential 함수에 더 근사하다고 밝혀졌다[5, 6]. Stretched exponential 함수는 'fat tails' 현상을 설명하기 위해 제안되었으며[5], 인터넷 미디어 서비스 시청 순위 함수를 통계적으로 근사하게 설명함이 입증되었다[6]. Stretched exponential 함수의 형태를 결정하는 모수는 미디어 파일의 크기에 영향을 받는다는 결과도 제시되었다[6].

본 논문에서는 QOOK TV의 시청 순위를 분석하여 국내의 대규

모 상용 주문형 비디오 시청 순위가 stretched exponential 함수로 적합하게 근사화 되는지 분석하고 모수와 상관관계가 있는 요인을 실증적으로 규명하고자 한다.

2. 시청 순위 데이터 수집 및 stretched exponential

KT QOOK TV는 가입자의 네트워크 대역폭에 따라 콘텐츠 전송 방식을 달리 하여, 다운로드 앤드 플레이와 스트리밍 두 가지 방식의 주문형 비디오 서비스를 제공하고 있다. 각 서비스의 가입자는 대략 30만명과 70만명이고, 6만편 가량의 콘텐츠를 서비스하고 있다. 두 방식 모두 공통적으로 중앙의 관리 서버에 요청하여 콘텐츠의 URL을 받는다. 이후 dnp 방식은 셋톱박스의 저장 장치에 콘텐츠를 다운로드 하면서 재생하고, 스트리밍 방식은 셋톱박스에 저장 장치가 장착되어 있지 않으므로 실시간으로 재생하게 된다.

본 연구에 사용된 시청 데이터는 2009년 1월 한 달 동안 1주일 단위로 콘텐츠 전달 방식별로 측정되었다. 1주일 주기로 시청회수의 변화가 관측되고 가장 시청 빈도가 높은 지상파 콘텐츠의 경우 1주일 단위로 신규 콘텐츠가 입수되기 때문에 측정 기간 단위를 1주일로 결정하였다. 측정된 시청 데이터에는 (콘텐츠의 제목, 장르 정보, 시청 회수) 형식의 데이터가 시청 회수의 내림차순 순서로 기록되어 있다. 콘텐츠 전달 방식은 명시적으로 구분하여 측정이 가능하지만, 파일 크기는 장르 정보로부터 유추하여 구분하고자 한다.

3. 모수 추정 및 분석

인터넷 미디어 서비스 시청 순위에 따른 시청회수를 통계적으로 근사화하는 stretched exponential 함수에 대해 살펴보고 각 요인의 수준 차이에 따른 모수의 차이를 분석하여 콘텐츠의 속성이 시청 순위에 따른 시청회수 함수에 어떤 영향을 주는지 알아보하고자 한다.

Stretched exponential 함수는 시청회수 순위 $i(i=1, 2, \dots, N)$ 와

시청회수 y_i 사이에 $y_i^c = -a \ln i + b$ (단, $0 < c < 1$)인 관계를 나타낸다. 총 세 개의 모수가 존재하지만 a 와 b 는 c 로부터 계산되기 때문에 실질적으로는 모수가 c 하나인 함수이다[6].

Stretched exponential 함수의 모수를 추정하기 위해 MLE(maximum likelihood estimation) 방법을 사용하였으며 추정의 정확도를 평가하기 위해 R^2 값을 계산하였다. R^2 는 시청순위와 시청회수 간의 상관관계를 나타내는 값으로 1에 가까울수록 추정함수가 실제 데이터를 정확하게 근사함을 의미한다. 본 연구에서는 시청 순위와 시청 회수를 로그-로그 스케일(zipf 법칙)과 c 승-로그(stretched exponential)에 각각 타점하고 추정된 모수와 R^2 값을 계산하였다. 예를 들어, 그림<1>의 x 축은 시청 순위를 로그스케일로 나타내고 좌측과 우측의 y 축은 각각 c 승 스케일과 로그 스케일을 나타낸다. 모수 c 가 1에 가까울수록 지수함수와 유사하고 로그-로그 스케일에서 휘어짐의 정도가 크게 나타남을 알 수 있다.

가. 콘텐츠 파일 크기

시청 데이터에는 명시적으로 파일 크기에 대한 수치가 포함되어 있지 않기 때문에 정확한 파일 크기를 알 수 없지만, 동일한 방식으로 인코딩된 콘텐츠는 재생시간과 파일 사이즈가 비례한다는 사실을 알 수 있다. 따라서, 재생 시간이 60시간 내외인 지상파 콘텐츠와 5분 내외인 음악 콘텐츠의 파일 크기도 이와 비례함을 가정하였다.

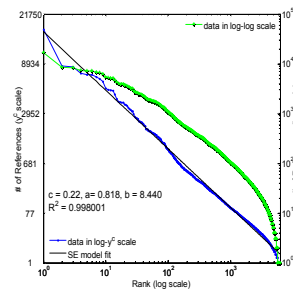


그림 2. MBC 카테고리

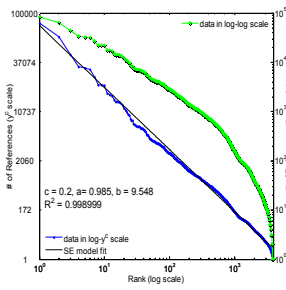


그림 3. 음악 카테고리

그림 1, 2의 분석 결과 파일크기가 서로 10배 이상 차이가 나는 MBC와 음악 카테고리의 추정된 모수 값은 각각 0.22와 0.2로 유의한 차이가 있다고 할 수 없다.

나. 콘텐츠 전달 방식

다운로드 앤 플레이 방식과 스트리밍 방식간에는 사용자가 느끼는 사용성에서 차이가 발생한다고 할 수 있다. 그럼에도 불구하고 콘텐츠 전달 방식에 따라서 모수간의 차이가 유의하다고 보기 어렵다. 다운로드 앤 플레이와 스트리밍 방식의 모수는 각각 0.18, 0.205 이다.

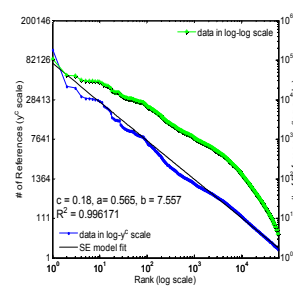


그림 5. 다운로드 앤 플레이

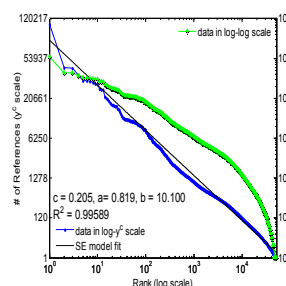


그림 4. 스트리밍

다. 콘텐츠 장르

영화와 지상파(SBS)를 비교한 결과(그림 6, 7 참조) 상당한 차이가 발생함을 알 수 있다. 각각 0.37과 0.17으로 추정되었다. 모수 값이 크다는 것은 시청 회수가 많은 콘텐츠의 빈도가 높다는 것을 의미한다. 영화 콘텐츠는 서비스 기간이 비교적 짧고 사용자의 취향이 다변화되어 나타나는 것으로 예상된다.

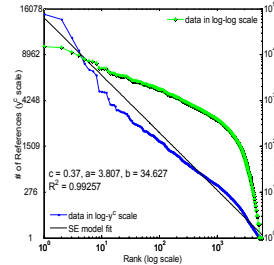


그림 1. 영화 카테고리

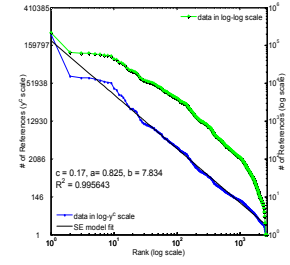


그림 6. SBS 카테고리

4. 결론

본 연구에서는 대규모 상용 주문형 비디오 서비스의 시청 순위와 시청 회수 간의 관계를 분석하여 stretched exponential 함수에 정확하게 근사화 된다는 사실을 실증적인 방법을 사용하여 밝혀내었다. 본 연구의 결과에 의하면 파일 크기와 모수 간의 상관관계는 명확하게 존재하지 않음을 알 수 있다. 또한 콘텐츠 전달 방식에 의한 차이는 유의하지 않았으나 영화와 기타 장르 사이에는 모수에 큰 차이가 발생함을 알 수 있다. 해외의 인터넷 미디어와 같이 국내의 주문형 비디오 시청 순위함수 역시 stretched exponential 함수를 따름이 밝혀졌으므로 스토리지 비용 감소를 위해 P2P 콘텐츠 전달 방식 도입에 대한 필요성이 커지고 있음을 직접 확인할 수 있다.

참고문헌

- [1] F. Thouin and M. Coates, Video-on-demand networks: Design approaches and future challenges, Network, IEEE, vol. 21, no. 2, pp. 42-48, 2007.
- [2] Y.-D. Lin, H.-Z. Lai, and Y.-C. Lai, A hierarchical network storage architecture for video-on-demand services, Proceedings of the 21st IEEE Conference on, 1996, pp. 355-364.
- [3] R. Günther, L. Levitin, B. Schapiro, and P. Wagner, Zipf's law and the effect of ranking on probability distributions, International Journal of Theoretical Physics, vol. 35, no. 2, pp. 395-417, 1996.
- [4] L. A. Adamin and B. A. Huberman, Zipf's law and the internet, Glottometrics, vol. 3, pp. 143-150, 2002.
- [5] J. Laherrère and D. Sornette, Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales, The European Physical Journal B, vol. 2, no. 4, pp. 525-539, 1998.
- [6] L. Guo, E. Tan, S. Chen, Z. Xiao, and X. Zhang, The stretched exponential distribution of internet media access patterns, Proceedings of the 27th ACM symposium on Principles of distributed computing, 2008, pp. 283-294.