

다운로드 기반의 주문형 비디오 서비스에서 다중 지수를 고려한 동영상 프리페칭 기법

Multi-index Prefetching Mechanism for Download-based Video on Demand Services

문양찬* · 임민규*
(YangChan Moon · Mingyu Lim)

Abstract - In video content watching service, when a user requests video content, the content server has to transmit the entire video to the client for watching. This transmission delay increases as the size of video content increases. In order to solve the transmission delay problem, a prefetching technique can be used in which a video content to be watched by a user is predicted and transmitted to a client before the user requests it. In this paper, we propose a prefetching system considering multiple indices for video content. In the proposed method, video content to be prefetched is selected by comprehensively analyzing the order relation index indicating the order of viewing the videos of the users, the similarity index between the video contents, and the popularity index reflecting the viewing frequency of the video content. Experimental results show that the maximum accuracy is achieved when prefetching uses only the order relation index for movie contents.

Key Words : Prefetching System, Video Content, Recommendation, Transmission Delay

1. 서 론

인터넷이 발전하고 네트워크 속도가 향상되면서 빠른 속도로 많은 데이터를 주고받을 수 있기 때문에 사용자들은 동영상 콘텐츠를 공유할 수 있게 되었다. 동영상 서비스는 유튜브 같은 동영상 콘텐츠 공유 사이트를 통하여 다양한 동영상 콘텐츠를 스트리밍 방식으로 시청하거나, IPTV 같이 선택한 영상 전체를 다운로드 받아 시청하는 방식이 있다. 다운로드 방식의 동영상 서비스인 경우, 사용자가 원하는 동영상 콘텐츠를 요청한 순간부터 실제로 동영상을 시청할 때 까지 전송 지연을 매번 요청할 때마다 경험하는 문제가 있다.

기존에 연구되었던 추천시스템은 콘텐츠와 사용자에 대한 정보를 통하여 수많은 동영상 콘텐츠 중에서 사용자가 선호하는 동영상을 찾을 수 있도록 하여 접근 지연을 줄일 수 있었다. 하지만 이러한 추천시스템은 사용자의 요청 이후 동영상 콘텐츠가 실제로 존재하는 서버에서 사용자의 클라이언트에 전송 받은 뒤에 화면에 출력해야하기 때문에 전송 지연을 줄일 수는 없다.

반면에 전송 지연 문제를 해결하기 위해 미리 클라이언트 디바이스로 동영상 콘텐츠를 다운로드하는 동영상 프리페칭 시스템 [1]이 연구되었지만, 이 연구에서의 프리페칭 시스템은 실시간 스트리밍 중인 동영상 콘텐츠를 대상으로 하기 때문에 사용자가

동영상 콘텐츠를 프리페칭 받으려면, 다른 사용자가 그 동영상 콘텐츠를 미리 스트리밍 중이어야 한다는 전제조건이 필요하다. 이러한 방법은 다른 사용자가 스트리밍 중이라면 무조건 프리페칭을 수행하기 때문에 대상 콘텐츠를 예측할 필요는 없지만, 다른 사용자가 보는 동영상 콘텐츠가 아니라면 프리페칭을 시도할 수 없다. 다른 사용자들이 보고 있지 않은 동영상 콘텐츠를 보는 사람은 프리페칭을 전혀 받지 못하는 것이다.

사용자가 다음에 시청할 동영상을 추측하여 요청 전에 미리 다운로드하기 때문에, 프리페칭에서 가장 중요한 것은 정확도이다. 정확도가 높다는 것은 서버에서 클라이언트에 프리페칭한 동영상 콘텐츠를 사용자가 실제 시청할 확률이 높다는 의미이다. 정확도가 낮으면 프리페칭에 사용되었던 자원이 낭비되는 결과를 초래한다. 따라서 정확도를 높이기 위해서 프리페칭의 대상이 되는 동영상 콘텐츠를 선정하는 방법이 중요하다. 사용자가 선호할 동영상 콘텐츠를 선정한다는 점에서 프리페칭 시스템은 추천 시스템과 동일하다. 따라서 기존에 연구된 추천 시스템의 방법을 적용할 수 있다.

본 논문에서는 동영상 콘텐츠 간의 순서관계와 유사도, 인기도를 고려하여 사용자가 다음에 볼 동영상 콘텐츠를 예측하고 그 동영상 콘텐츠를 클라이언트에 프리페칭 함으로써 전송 지연 없이 동영상 콘텐츠를 시청할 수 있는 프리페칭 시스템을 제안한다. 먼저 순서관계는 특정한 두 동영상 콘텐츠를 연달아 본 사용자의 수를 의미한다. 많은 사용자가 일정한 순서로 동영상 콘텐츠를 시청했다면, 새로운 사용자도 그 순서를 따라 시청할 것이라는 추측에 기반 한다. 유사도는 두 동영상 콘텐츠의 유사한 정도를 의미한다. 어떤 동영상 콘텐츠를 본 사용자가 그것과 유사한 동영상 콘텐츠를 시청할 것이라는 추측에 기반 한다. 마지막

† Corresponding Author : Dept. of Smart ICT Convergence, Konkuk University, Korea.
E-mail: mlim@konkuk.ac.kr

* Dept. of Internet & Multimedia Engineering, Konkuk University, Korea.

Received : March 8, 2017; Accepted : July 9, 2017

으로 인기도는 해당 동영상 콘텐츠의 시청 횟수를 의미한다. 더 많은 사용자가 시청한 동영상 콘텐츠이면 다른 사용자도 그 동영상 콘텐츠를 보려고 할 것이라는 추측에 기반 한다.

본 논문에서 제안하는 다중 지수를 고려한 동영상 콘텐츠 프리페칭 시스템의 정확도를 측정하기 위해, 실제 운용되는 프리페칭 시스템의 구조를 수정하여 데이터 셋을 이용한 정확도 실험을 수행하였다. 순서 관계 지수, 유사도 지수, 인기도 지수를 개별적으로 적용하여 정확도를 실험한 후에, 세 가지 지수를 다양한 가중치로 결합하여 정확도를 실험하였다. 실험 결과, 순서 관계 지수가 세 가지 지수중에 가장 높은 정확도를 보였고 세 지수를 결합하였을 때에도 순서 관계 지수에 대한 가중치가 높을 때 정확도가 높음을 확인하였다.

2. 관련 연구

2.1 추천 시스템

사용자가 다양한 콘텐츠 중에서 원하는 콘텐츠를 찾을 때 생기는 접근 지연을 줄이기 위해 추천 시스템이 연구되었다. 이러한 추천 시스템은 다수의 콘텐츠 중에서 사용자가 선호할만한 콘텐츠를 선별한다는 점에서 프리페칭 시스템과 유사하다. 추천 시스템에서 사용되는 대표적인 선별 기준으로는 아이템 기반 방법과 사용자 기반 방법, 협동 필터링 방법이 있다.

아이템 기반 방법[2]은 해당하는 아이템, 즉 콘텐츠에 대한 정보를 기반으로 선별하여 추천하는 방법이다. 대표적으로 인기도를 기준으로 추천하는 방법이 있는데, 이 방법은 아이템들 중에 조회수나 사용자들이 평가한 점수를 사용하여 각 아이템들이 사용자들에게 얼마나 선호되는지를 판단하여 아이템을 선별한다.

사용자 기반 방법[3]은 사용자에 대한 정보를 기반으로 선별하여 추천하는 방법이다. 대표적으로 해당 사용자와 친구 관계이거나 구독 관계에 있는 사용자들이 선호하는 아이템을 해당 사용자도 선호할 것이라 판단하여 추천하는 것이다. 이 방법은 특정 관계에 있는 사용자들끼리 선호하는 아이템이 비슷할 것이라는 가정이 필요하다.

또한 아이템 기반 방법과 사용자 기반 방법을 결합시킨 협동 필터링 방법[4]이 연구되었다. 이 방법은 아이템에 대한 정보와 사용자에 대한 정보를 사용하여 행렬을 만들고 어떤 사용자가 어떤 아이템을 선호할지 예측하는 방법이다. 협동 필터링은 아이템 기반 방법이나 사용자 기반 방법 중 하나만 사용하는 것보다 더 높은 정확도를 보였기 때문에 다양한 콘텐츠를 대상으로 연구되고 사용되었다.

최근에는 아이템과 사용자에 기반한 정보만을 사용했던 방법들과 다르게, 시간이나 장소, 다른 사용자와의 관계 등을 고려한 상황 인식 기반 방법[5]이 연구되었다. 상황(Context) 정보는 대상이 되는 콘텐츠 또는 추천 시스템의 목적에 따라 달라지기 때문에 일반적으로 적용되는 시스템을 만들기 어렵지만, 원하는 대상이나 목적에 특화되어있기 때문에 더 적합한 시스템으로 만들 수 있다. 이러한 추천시스템은 사용자가 선호할만한 콘텐츠를 검색하는 시간을 줄여줄 수는 있지만, 실제로 사용자가 콘텐츠를

시청하기 위해 겪어야하는 전송 지연 문제는 여전히 존재한다.

2.2 콘텐츠 프리페칭 시스템

프리페칭 시스템은 어떤 데이터를 전송 받은 이후에 처리할 때, 전송 지연을 없애기 위해 미리 데이터를 전송하는 방법으로 처리 장치의 대기 시간을 줄여주는 방법으로 웹 콘텐츠, 프록시 서버, 동영상 스트리밍 시스템 등에서 사용 되었다.

웹 콘텐츠 프리페칭 시스템[6]은 사용자가 웹 페이지를 요청했을 때, 해당하는 HTML 파일에 속해있는 콘텐츠 또는 링크를 사용자가 요청할 것을 예측하여 프리페칭 하는 시스템이다. 일반적인 웹 캐시는 사용자가 이미 본 콘텐츠를 저장하는 것이지만, 웹 콘텐츠 프리페칭은 사용자가 아직 보지 않은 콘텐츠를 미리 다운로드 놓는 점에서 차이가 있다.

프록시 서버의 프리페칭 시스템[7]은 낮은 대역폭의 클라이언트가 웹 페이지를 요청할 때, 먼 거리에 있거나 마찬가지로 낮은 대역폭의 서버에서 전송 받을 경우에 전송 지연이 문제가 되기 때문에 프록시 서버를 사용하여 더 빠르게 전송 받도록 하는 시스템이다.

기존에 연구되었던 프리페칭 시스템은 사용자가 요청한 웹페이지 자체 또는 그 웹페이지에 포함된 링크의 콘텐츠만을 대상으로 하였다. 동영상 콘텐츠를 대상으로 하는 프리페칭 시스템은 상대적으로 네트워크 대역폭이나 저장에 필요한 크기 등 소모되는 자원에 비해 전송 지연을 줄일 수 있는 경우가 확실하게 보장되지 않았다. 본 논문에서 제안하는 프리페칭 시스템은 사용자가 시청할 동영상 콘텐츠를 예측하고 정확도를 측정함으로써 전송 지연을 줄일 수 있는 경우를 수치화하여 볼 수 있다. 또한 자원 소모를 최소화하고 전송 지연을 최대한 줄일 수 있도록 일정 수준 이상의 정확도를 보일 때만 제한적으로 프리페칭 하도록 하였다.

3. 프리페칭 지수

프리페칭을 위해서 가장 중요한 작업은 프리페칭 대상 콘텐츠를 선정하는 것이다. 이번 장에서는 프리페칭 대상 콘텐츠 선정을 위해 본 논문에서 적용한 세 가지 기준에 대해 기술한다.

3.1 순서 관계

한 명의 사용자가 여러 콘텐츠를 시청할 때, 그 콘텐츠들 사이에는 순서 관계[1]가 있을 수 있다. 특히, 많은 사용자들이 같은 순서로 여러 개의 콘텐츠를 시청했다면, 새로운 사용자 또는 그 순서의 콘텐츠를 처음 접하는 사용자 또한 그 순서대로 콘텐츠를 시청할 확률이 높다. 이러한 점을 활용한다면 어떤 사용자가 콘텐츠 A를 시청하였을 때, 다른 많은 사용자들이 콘텐츠 A를 시청한 이후에 콘텐츠 B를 시청하였다면, 그 사용자 또한 콘텐츠 A를 시청한 이후에 콘텐츠 B를 시청할 확률이 높다는 것을 예측할 예를 들면, 어떤 가수의 지난 무대 영상을 보고 최신 무

대 영상을 본 사용자가 많다면, 지난 무대 영상을 본 사용자 또한 최신 무대 영상을 볼 것이라고 예측할 수 있는 것이다. 수 있다. 이 때 콘텐츠 B를 프리페칭 대상 콘텐츠로 지정하는 것이다.

사용자가 콘텐츠 c를 시청 중일 때, 콘텐츠 c와 다른 콘텐츠 t에 대한 순서 관계 지수(Order-relation Index, OI_{ct})는 식 (1)과 같이 계산한다. 여기서 O_c 는 콘텐츠 c와 콘텐츠 t의 순서 관계 수, 즉 콘텐츠 c를 시청한 이후 콘텐츠 t를 시청한 사용자의 수이고, O_c 는 콘텐츠 c의 모든 순서 관계 수, 즉 콘텐츠 c를 시청한 이후 다른 콘텐츠를 시청한 모든 사용자의 수이다.

$$OI_{ct} = O_{ct} / O_c \quad (1)$$

3.2 유사도

콘텐츠에는 그 콘텐츠에 대한 정보를 나타내는 메타데이터가 있다. 영화인 경우 제목, 감독, 작가, 배우, 상영시간, 개봉일, 줄거리 등 많은 정보가 존재하게 되는데 이러한 메타데이터를 활용하여 두 개의 콘텐츠가 얼마나 유사한 지를 알아볼 수 있다. 메타데이터 중 많은 부분이 동일하다면, 그 두 개의 콘텐츠는 유사하다고 할 수 있는 것이다. 어떤 사용자가 한 콘텐츠를 시청했다면, 그 콘텐츠와 유사한 다른 콘텐츠로 시청할 가능성이 있다. 이 때 그 사용자가 시청 중인 콘텐츠와 유사한 콘텐츠를 프리페칭 타겟 콘텐츠로 지정하는 것이다.

표 1 유사도 지수 계산 기준

Table 1 Calculation criteria of similarity index

	Number of same items	Point
Genre	equal to or more than 2	30
	1 and the total number of items is 1	30
	1 and the total number of items is more than 1	10
	0	0
Director	equal to or more than 2	40
	1 and the total number of items is 1	40
	1 and the total number of items is more than 1	20
	0	0
Writer	equal to or more than 2	20
	1 and the total number of items is 1	20
	1 and the total number of items is more than 1	10
	0	0
Actor	equal to or more than 3	10
	2	8
	1	4
	0	0

표 1은 두 콘텐츠의 유사도 지수(Similarity Index, SI_{ct})를 계산하기 위한 각 항목별 점수를 나타낸 표이다. 영화 콘텐츠의 메타데이터 중에서 두 콘텐츠 사이의 유사성을 짐작할 수 있는 장르, 감독, 작가, 배우 항목이 몇 개나 일치하는지를 각각의 점수로 나타낸 다음 합산하는 것이 두 콘텐츠의 유사도 지수가 된다.

3.3 인기도

시스템에 여러 명의 사용자가 접속하여 콘텐츠를 시청하게 되

면, 어떤 콘텐츠는 많은 사용자가 시청하고 어떤 콘텐츠는 적은 사용자가 시청하게 된다. 즉 시청횟수에 따라 콘텐츠의 인기도를 가늠할 수 있는데 많은 사용자가 본 콘텐츠의 경우 다른 사용자 또한 그 콘텐츠를 시청할 가능성이 있다. 따라서 시청횟수가 높은 콘텐츠를 프리페칭 타겟 콘텐츠로 지정하는 것이다.

콘텐츠 c에 대한 인기도 지수(Popularity Index, PI_c)는 식 (2)와 같이 계산한다. 여기서 P_c 는 콘텐츠 c의 시청횟수이고, P_{all} 은 각 콘텐츠의 시청횟수를 모두 합산한 것이다. 즉 그 콘텐츠의 인기도 지수는 모든 콘텐츠의 전체 시청횟수 중에서 그 콘텐츠가 차지하고 있는 시청횟수의 비율과 같다.

$$PI_c = P_c / P_{all} \quad (2)$$

3.4 세 지수의 결합

순서 관계 지수와 유사도 지수, 인기도 지수를 각각 단독으로 프리페칭 대상 콘텐츠 선별 기준으로 사용하는 것보다 세 가지를 동시에 고려하는 것이 프리페칭의 정확도를 높일 가능성이 있다. 예를 들어, 어떤 새로운 콘텐츠가 급격하게 인기도가 상승하고 있어서 많은 사용자들이 그 콘텐츠를 볼 것이라고 예상하지만 순서 관계 지수만 가지고 프리페칭 대상 콘텐츠를 선별하고 있다면 인기도가 급상승하는 콘텐츠를 구별해 낼 수 없는 것이다. 반대로 순서 관계가 충분한 두 콘텐츠가 있어도 그 콘텐츠들의 인기도가 낮은 상황이라면 인기도 지수만 가지고 프리페칭 대상 콘텐츠를 선별하는 시스템에서는 그 콘텐츠를 프리페칭 시켜서 적중할 확률이 낮은 것이다. 순서 관계 지수, 유사도 지수, 인기도 지수를 동시에 고려하여 프리페칭 하기 위해서는 세 가지 지수를 하나의 지수로 결합하는 방법이 필요하다.

세 지수는 각각의 계산방법이 다르기 때문에 하나의 지수로 결합하기 위해서는 표준화 과정을 거쳐야 한다. 세 지수를 각각 표준화 하는 방법은 식 (3), (4), (5)와 같다.

$$NOI_{ct} = (OI_{ct} - \mu_o) / \sigma_o \quad (3)$$

$$NSI_{ct} = (SI_{ct} - \mu_s) / \sigma_s \quad (4)$$

$$NPI_c = (PI_c - \mu_p) / \sigma_p \quad (5)$$

이 때, μ 는 각 지수의 평균이고, σ 는 표준 편차이다. 위 식으로 표준화된 순서 관계 지수, 유사도 지수, 인기도 지수는 -1과 1 사이의 값을 갖게 되는데, 1에 가까울수록 해당하는 콘텐츠의 지수가 다른 콘텐츠의 지수에 비해 높은 수치를 가진다는 의미이다.

표준화된 세 지수를 더하여 하나의 숫자로 표현한 것을 프리페칭 지수라 정의하면 콘텐츠 c에 대한 콘텐츠 T의 프리페칭 지수는 식 (6)과 같이 계산할 수 있다.

$$PI_{ct} = NOI_{ct} + NSI_{ct} + NPI_c \quad (6)$$

위 식은 순서 관계 지수, 유사도 지수, 인기도 지수를 각각

표준화 한 뒤에, 모두 더하여 프리페칭 지수를 계산한 것으로 이 프리페칭 지수를 사용하여 프리페칭 할 콘텐츠를 정한 다는 것은 곧 세 지수를 동시에 고려하여 프리페칭 하는 것을 의미한다. 하지만 세 지수를 단순히 더하는 것은 동일한 정도로 고려한 다는 것이고 이것은 사용자의 특성이나 콘텐츠의 특성, 또는 특정 기간에 따라 달라지는 경향을 반영하지 못할 수 있다. 따라서 더하기 전에 가중치를 곱하여 상황에 따라 다르게 프리페칭 지수를 계산하도록 만들 수 있다. 가중치가 반영된 프리페칭 지수는 식 (7)과 같이 계산할 수 있다.

$$WPI_{ct} = W_{oi} * NOI_{ct} + W_{si} * NSI_{ct} + W_{pi} * NPI_c \quad (7)$$

이 때, W_{oi} 는 순서 관계 지수에 대한 가중치, W_{si} 는 유사도 지수에 대한 가중치, W_{pi} 는 인기도 지수에 대한 가중치를 나타낸다.

$$t = \underset{t}{\operatorname{argmax}} WPI_{ct} = \{t | WPI_{cs} \leq WPI_{ct}, \forall s \subset C, s \neq t \neq c\} \quad (8)$$

식 (8)은 프리페칭 대상 목록을 생성하기 위해 전체 콘텐츠 집합 C 중에서 사용자가 현재 시청 중인 콘텐츠 c에 대한 프리페칭 지수를 최대로 만드는 콘텐츠 t를 의미한다. 이 콘텐츠 t가 프리페칭 대상 콘텐츠이며, 프리페칭 대상 목록에 추가시킴으로써 프리페칭 대상 선정을 수행하는 것이다. 또한 프리페칭 대상 목록의 개수에 따라 식 (8)을 반복하여 수행할 수 있다.

4. 프리페칭 시스템 설계

동영상 콘텐츠 프리페칭 시스템에는 그림 1과 같이 사용자가 사용하는 클라이언트와 동영상 콘텐츠를 보유하고 서비스하는 서버가 있다. 클라이언트와 서버에는 공통적으로 통신 모듈과 프리페칭 모듈이 존재한다. 통신 모듈은 클라이언트와 서버의 연결 및 데이터 전송을 관리하고, 프리페칭 모듈은 본 논문에서 제안하는 프리페칭 기법을 적용하여 프리페칭을 수행한다.

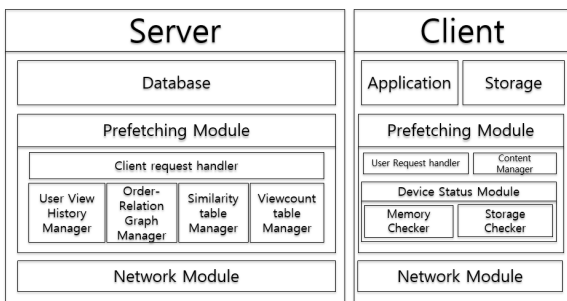


그림 1 동영상 콘텐츠 프리페칭 시스템 구조
Fig. 1 Video content prefetching system structure

4.1 서버의 프리페칭 모듈

서버의 프리페칭 모듈에는 사용자 시청 기록 관리 모듈, 순서 관계 그래프 관리 모듈, 유사도 테이블 관리 모듈, 시청 횟수 테이블 관리 모듈이 있다. 사용자 시청 기록 관리 모듈은 이 시스템에 접속한 사용자가 동영상 콘텐츠를 시청한 기록을 저장하고 관리한다. 클라이언트에서 콘텐츠 시청 메시지가 오면 그 콘텐츠와 사용자를 기록한다. 이 때, 이 사용자가 이전에 다른 콘텐츠를 시청한 기록이 있다면 순서 관계 그래프 관리 모듈에 해당하는 순서 관계를 추가하도록 요청한다. 그것을 위해 사용자 시청 기록 관리 모듈은 사용자별로 콘텐츠를 요청한 기록을 보유하고 있어야 한다. 순서 관계 그래프 관리 모듈, 유사도 테이블 관리 모듈, 시청 횟수 테이블 관리 모듈은 각각 순서 관계 지수, 유사도 지수, 인기도 지수를 계산하기 위한 모듈이다.

먼저 순서 관계 그래프 관리 모듈은 순서 관계 그래프의 생성 및 갱신을 수행한다. 프리페칭 시스템에는 하나의 순서 관계 그래프가 존재하며, 그래프의 각 노드는 콘텐츠를 의미하고, 두 노드를 잇는 연결선은 사용자가 해당하는 두 콘텐츠를 이어서 시청했다는 것이다. 또한 연결선의 가중치는 그 순서로 시청한 사용자의 수를 의미한다. 사용자의 시청 기록을 토대로 순서 관계를 발견하여 갱신하는 것은 곧 해당 연결선의 가중치를 증가시키는 것이다. 이 모듈에서는 한 콘텐츠가 가진 순서 관계 중, 상위의 콘텐츠 목록을 찾을 수 있고 이것이 순서 관계 지수를 기준으로 프리페칭 대상 콘텐츠를 정하는 것이다. 유사도 테이블 관리 모듈은 유사도 테이블의 생성 및 갱신을 수행한다. 하지만 유사도 테이블은 순서 관계 그래프와는 다르게 시스템 구동 직후, 사용자의 시청기록을 수신하기 이전에 만들어져야 한다. 유사도 지수는 사용자가 아닌 콘텐츠들 간의 메타데이터를 이용하기 때문에 미리 계산하여 완성시킬 수 있다. 그리고 미리 만들어놓고 유사도 지수를 조회하기만 한다면 프리페칭 할 때 서버의 계산 부담을 줄일 수 있다는 점도 있다. 이 모듈에서는 두 콘텐츠 간의 유사도를 알 수 있으므로, 한 콘텐츠가 가진 유사도 중, 상위의 콘텐츠 목록을 찾을 수 있고 이것이 유사도 지수를 기준으로 프리페칭 대상 콘텐츠를 정하는 것이다. 시청 횟수 테이블 관리 모듈은 시청 횟수 테이블의 생성 및 갱신을 수행한다. 모든 콘텐츠의 시청 횟수를 저장하는 것 뿐 아니라, 실시간으로 사용자의 시청 기록에 따라서 해당 콘텐츠의 시청 기록을 증가시키는 작업을 수행한다. 이 모듈에서는 한 콘텐츠의 시청 횟수를 알 수 있으므로, 현재 시스템에서 시청 횟수가 높은 상위의 콘텐츠 목록을 찾을 수 있고 이것이 인기도 지수를 기준으로 프리페칭 대상 콘텐츠를 정하는 것이다.

4.2 클라이언트의 프리페칭 모듈

클라이언트의 프리페칭 모듈에는 디바이스 상황판단 모듈, 프리페칭 콘텐츠 관리 모듈, 사용자/서버 요청 처리 모듈이 있다. 디바이스 상황판단 모듈은 사용자/서버 요청 처리 모듈에서 요청을 하면 클라이언트가 프리페칭을 할 수 있는 상태인지 아닌지를 검사하고 판단하여 가능한지 불가능한지 응답한다. 프리페칭 콘

텐츠 관리 모듈은 클라이언트의 콘텐츠 저장소에 저장되는 프리페칭 콘텐츠를 관리한다. 프리페칭 콘텐츠란 현재 사용자가 시청하고 있는 동영상 콘텐츠가 아닌, 서버에서 전송되어 콘텐츠 저장소에 저장된 동영상 콘텐츠를 의미한다. 사용자/서버 요청 처리 모듈은 사용자가 하는 요청과 서버가 하는 요청을 처리하는 역할을 한다.

4.3 프리페칭 동작 순서

그림 2는 동영상 콘텐츠 프리페칭 시스템의 동작 순서를 나타낸 것이다. 이 동작은 프리페칭 개수를 10개로 정했음을 가정하여 설명한다. 처음 시스템을 시작하게 되면 대기 상태가 되며, 사용자의 콘텐츠 시청을 기다린다. 사용자가 어떤 콘텐츠를 시청하면 그 기록을 시청기록 테이블에 저장하고, 순서 관계 그래프를 갱신한다. 시청 기록에서 순서 관계가 존재한다는 것은 그 사용자의 최신 시청기록과 그 이전의 시청기록이 연속된 시청이라고 생각될 만큼 짧은 시간에 기록되었다는 것이다. 그 후 사용자의 클라이언트가 프리페칭 가능한 상황을 확인한다. 만약 클라이언트가 프리페칭이 가능한 상황이라면 시스템은 프리페칭 동작을 시작한다. 3가지의 지수를 고려하기 위해서 순서 관계 그래프에서 해당 콘텐츠의 순서 관계중 상위 10개, 유사도 테이블에서 해당 콘텐츠의 유사도중 상위 10개, 시청횟수 테이블에서 상위 10개를 모아서 프리페칭 대상 목록을 만든다. 동일한 콘텐츠가 목록에 존재할 수 있기 때문에 프리페칭 대상 목록의 콘텐츠 숫자는 최소 10개에서 최대 30이다. 프리페칭 대상 목록에 있는 모든 콘텐츠의 3가지 지수를 모두 계산한 후에 식 (7)을 사용하여 프리페칭 지수를 계산한다. 최종적으로 프리페칭 대상 목록의 프리페칭 지수 상위 10개의 콘텐츠를 클라이언트에 프리페칭 한다.

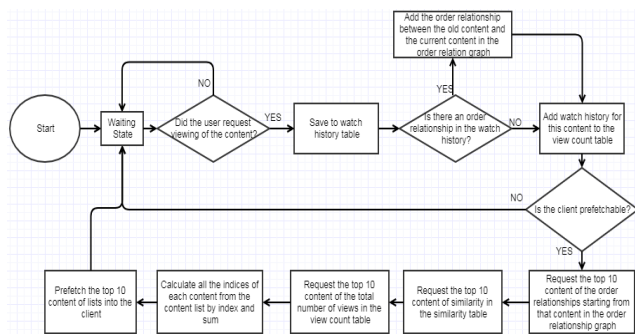


그림 2 동영상 콘텐츠 프리페칭 시스템 순서도
 Fig. 2 Video content prefetching system flowchart

단 그림 2는 순서 관계 지수, 유사도 지수, 인기도 지수를 모두 고려했을 때를 나타낸다. 지수 중 하나를 선택하여 실험 할 때는 다른 지수와 관련된 동작은 수행하지 않고 넘어간다. 예를 들어 순서 관계 지수를 고려한 프리페칭을 할 때에는 순서 관계 그래프에서 프리페칭 대상 콘텐츠 목록을 받은 이후에 그 목록을 클라이언트에 프리페칭 하면 된다. 다른 지수에 관한 목록을 요청하거나, 지수를 표준화하고 가중치를 적용하여 합산할 필요가

없다. 다만 2개 이상의 지수를 고려한다면 지수 계산 방식의 차이 때문에 표준화하고 합산하는 동작을 반드시 수행해야 될 필요가 있다.

5. 실험

본 연구에서 제안하는 다중 지수 기반의 동영상 콘텐츠 프리페칭 시스템의 프리페칭 정확도를 측정하기 위해서 무비렌즈 데이터 셋을 사용하여 실험하였다. 프리페칭 정확도를 실험해보기 위한 가장 적합한 데이터 셋은 실제 사용자들이 동영상 콘텐츠를 시청한 기록들이지만, 개인정보 보호의 이유로 데이터 셋을 준비하는데 어려움이 있다. 따라서 본 논문에서는 실제 사용자가 시청한 기록 대신에 사용자가 콘텐츠에 별점을 준 기록을 시청 기록으로 가정하여 실험하였다.

5.1 무비렌즈 데이터 셋

무비렌즈 데이터 셋[8]은 해당 사이트에 접속한 유저들이 영화 콘텐츠에 0.5개에서 5개의 별점을 준 기록을 모은 데이터 셋으로, 259,137명의 유저가 40,110개의 영화에 준 총 24,404,096개의 별점의 기록을 모은 데이터 셋이다. 데이터 셋의 각 행은 하나의 별점 기록으로, 한 유저가 한 영화에 매긴 별점을 시간과 함께 기록하였다. 따라서 각 별점 기록에는 유저의 ID, 영화의 ID, 별점, 타임스탬프가 존재한다.

엄밀히 말하면 별점 기록이 시청 기록을 대체할 수는 없다. 또한 사용자가 별점을 주는 순간에 그 영화를 시청하고 있다고 말할 수는 없다. 하지만 영화에 별점을 주었다는 것은 해당 영화를 시청했다는 것이고, 최소한 어떤 영화에 별점을 주고 난 다음 시간이 흐른 뒤 다시 접속하여 또 다른 영화에 별점을 주었다는 것은 해당하는 영화들을 일련의 순서로 시청하였다고 가정할 수 있다. 그러한 사용자가 별점을 매긴 순서와 다른 순서로 영화를 시청하였을 확률이 무시할 수 있을 만큼 매우 작다고 예상된다. 따라서 데이터 셋의 사용자가 콘텐츠에 별점을 준 기록인 별점 기록을 사용자가 콘텐츠를 시청한 시청 기록으로 간주하여 실험하였다.

물론 일반적으로 사용자가 한 번에 여러 개의 영화에 별점을 주는 경우가 많기 때문에 1시간 이내에 별점을 매긴 영화들을 하나의 클러스터로 묶어서, 클러스터 단위로 실험하였다. 예를 들어 어떤 사용자가 지난주 10개의 영화에 별점을 주고, 오늘 5개의 영화에 별점을 주었다면 이 사용자의 시청기록에는 각각 10개, 5개의 영화가 있는 클러스터가 2개 존재하는 것이다. 이 때 시간상 이전 클러스터에 있는 10개의 영화들을 사용하여 프리페칭 대상 목록을 만든다. 그 후 시간상 이후 클러스터에 있는 영화들이 이 프리페칭 대상 목록에 있는지, 몇 번째에 위치하는지 확인하는 것이다.

5.2 실험 방법

그림 3은 동영상 콘텐츠 프리페칭 시스템의 실험 순서도를 나

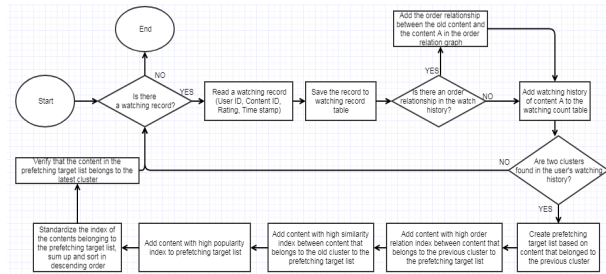


그림 3 동영상 콘텐츠 프리페칭 시스템 실험 순서도

Fig. 3 Video contents prefetching system experiment flow chart

타낸 것이다. 본 절의 실험 방법 설명은 프리페칭 개수를 10개로 했을 때를 가정하여 설명한다. 시스템이 시작되면 무비렌즈 데이터 셋의 시청 기록을 하나씩 읽는다. 시청 기록은 제일 먼저 시청 기록 테이블에 저장된다. 만약 해당 사용자의 시청 기록에서 순서 관계가 존재하면 순서 관계 그래프에 이전 콘텐츠와 콘텐츠 A의 순서 관계를 추가한다. 그리고 시청 횟수 테이블에 있는 콘텐츠A의 시청 횟수를 1만큼 증가시킨다. 그리고 사용자의 시청 기록에서 2개의 클러스터를 발견하면, 프리페칭 동작을 수행한다. 이전 클러스터에 대한 콘텐츠를 기준으로 프리페칭 대상 목록을 생성하는 것이다. 먼저 순서 관계 지수를 적용하기 위해 이전 클러스터에 속한 콘텐츠와 순서 관계 지수가 높은 콘텐츠 상위 10개를 프리페칭 대상 목록에 추가한다. 유사도 지수를 적용하기 위해 이전 클러스터에 속한 콘텐츠와 유사도 지수가 높은 콘텐츠 상위 10개를 프리페칭 대상 목록에 추가한다. 마지막으로 인기도 지수를 적용하기 위해 시청 횟수 테이블에서 인기도 지수가 높은 콘텐츠 상위 10개를 프리페칭 대상 목록에 추가한다. 각 지수 별로 중복된 콘텐츠가 있다면 합쳐지게 되므로 프리페칭 대상 목록에는 최소 10개, 최대 30개의 콘텐츠가 존재하게 된다. 목록에 속한 콘텐츠들의 계산되지 않은 나머지 지수들을 모두 계산한다. 예를 들어 순서 관계 지수가 높아서 추가된 콘텐츠의 경우 유사도 지수와 인기도 지수가 계산되어 있지 않는데 이것을 계산하여 추가하는 것이다. 프리페칭 대상 목록에 있는 모든 콘텐츠들의 지수가 계산되면 표준화 한 뒤에 합산하여 정렬한다. 그 후에 프리페칭 대상 목록에 속한 콘텐츠가 최신 클러스터에 속하여 있는지 확인하고 다시 시청 기록을 읽는 과정으로 넘어간다. 이 과정을 주어진 시청 기록을 전부 읽을 때 까지 반복한다.

단 그림 2와 마찬가지로 그림 3 또한 순서 관계 지수, 유사도 지수, 인기도 지수를 모두 고려했을 때를 나타낸다. 개별 지수 실험에서는 3.2.3 절의 설명과 마찬가지로 고려하지 않는 지수에 관한 동작과 표준화 및 합산 동작을 제외한다. 즉 총 3번의 개별 지수 실험에서는 각각 순서 관계 지수, 유사도 지수, 인기도 지수를 사용하여 프리페칭 대상 목록을 생성하는 것이다. 결합 지수 실험에서는 세 지수의 가중치를 모두 1로 정하는 단순 결합과 세 지수중 하나의 지수의 가중치를 높이고 다른 두 지수의 가중치를 낮춰서 특정 지수 위주로 결합하여 프리페칭하는 실험을 진행하였다. 이 때 가중치의 변화는 즉, 실험에 쓰이는 가중치를 (순서 관계 지수 가중치, 유사도 지수 가중치, 인기도 지수 가중

치)로 표현한다면 (1, 1, 1), (2, 0.5, 0.5), (0.5, 2, 0.5), (0.5, 0.5, 2)와 같이 4번의 결합 지수 실험을 진행한 것이다.

본 실험에 사용된 실험 환경은 다음 표 2와 같다.

표 2 실험 환경

Table 2 Experiment environment

Item	Detail
CPU	Intel(R) Core(TM) i5-4460 @ 3.20GHz
RAM	4GB DDR3 PC3-10700 * 2EA
HDD	500GB SAMSUNG HD502HJ
OS	Windows 7 64bit
Programming Language	JAVA 1.8

5.3 실험 결과

그림 4는 개별 지수 프리페칭 실험의 결과를 나타낸다. 프리페칭 개수에 따라 순서 관계 지수가 가장 높은 정확도를 보였고, 인기도 지수와 유사도 지수가 낮은 정확도를 보였다. 각 지수별 정확도는 프리페칭 개수가 1개 이상 10개 이하일 때는 큰 차이가 없지만, 10개 보다 많을 때 큰 차이를 보이기 시작했음을 알 수 있다. 다른 지수보다 순서 관계 지수가 유독 높은 정확도를 보이는 것은 본 논문의 실험에서 사용한 무비렌즈 데이터 셋의 특성 때문인 것으로 보인다.

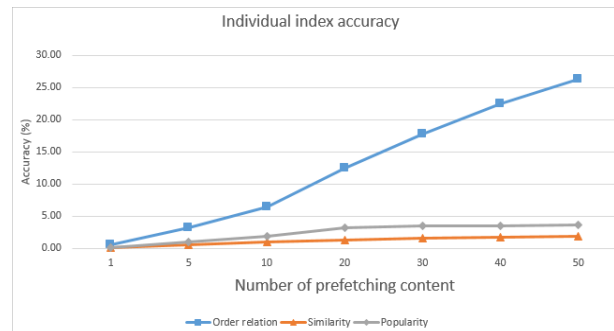


그림 4 개별 지수 프리페칭 실험 결과

Fig. 4 Individual index prefetching experiment results

가장 이상적인 경우, 즉 순서 관계 지수만을 적용하였고 프리페칭 개수가 50개일 때, 프리페칭 정확도는 25%보다 약간 높음을 알 수 있다. 이것은 단순히 50개를 프리페칭 해도 약 13개만이 적중되었다는 의미이다. 그 이하일 경우 그보다 낮은 확률이므로 프리페칭에 사용된 네트워크 대역폭의 4분의 3은 그 순간에는 낭비되었다고 볼 수도 있다. 하지만 프리페칭 된 콘텐츠가 클라이언트에 계속 남아있으면서 향후에 시청될 가능성도 존재한다. 또한 본 논문에서 적용된 세 가지의 지수는 사용자 개인의 기존 시청기록이 적용되지 않은, 즉 개인화 되지 않은 기준으로써 익명의 사용자나 최근 가입한 사용자 등 그 선호나 경향을 알 수 없는 사용자에게 적용되는 최소한의 정확도라고 볼 수도 있다.

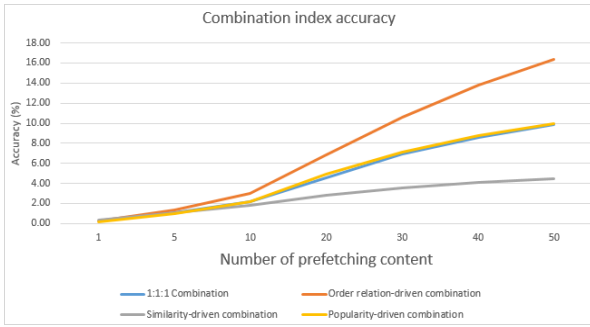


그림 5 결합 지수 프리페칭 실험 결과
 Fig. 5 Combination index prefetching results

그림 5는 결합 지수 프리페칭 실험의 결과를 나타낸다. 결합 지수 정확도에서는 순서 관계 위주의 결합이 가장 높은 정확도를 보였고, 인기도 위주 결합과 1:1:1 결합이 중간 정확도, 유사도 위주 결합이 가장 낮은 정확도를 보였다. 3장의 예상과는 다르게, 세 지수를 결합하여 적용한 실험의 정확도가 하나의 지수만 고려했을 때보다 높아지는 일은 없었다. 이것은 프리페칭 대상 목록의 구성 과정에서, 개별 지수로 계산하였을 때에는 대상 목록에 포함되지 않았던 동영상 콘텐츠가 결합된 지수로 계산하였을 때 추가되면서 정확도를 높이도록 보정하는 역할을 예상하였던 것이었지만 실험의 결과로 개별 지수 중에서 가장 높은 정확도를 가졌던 순서 관계 지수의 정확도를 오히려 떨어뜨리는 것으로 확인되었다. 이것은 개별 지수의 정확도가 더 높거나, 비슷한 경우에 대해서는 확신할 수 없지만 실험에 사용된 유사도 지수나 인기도 지수처럼 개별 지수의 정확도가 확연히 낮은 경우에는 아예 고려하지 않는 것이 더 높은 정확도를 유지할 수 있다는 것을 나타낸다.

표 3 각 실험방법 별 수행시간 (단위 : 초)

Table 3 Execution time by experiment method

(unit: second)

Index	Order relation index	Similarity index	Popularity index	Combination index
Time	1.233	9.958	5.133	12.098

실험방법의 수행시간을 비교하기 위해 프리페칭 대상 목록을 생성할 때 걸리는 시간을 측정하여 비교하였다. 표 3은 각 실험방법의 프리페칭 대상 목록 생성 1000회에 걸리는 수행시간을 5회 측정한 뒤 평균을 계산한 결과이다. 특정 지수 위주의 실험은 결합 지수 실험에서 가중치의 수치만 변경한 것이므로 수행시간에서는 유의미한 차이가 없기 때문에 제외하였다. 또한 각 방법의 구조상 프리페칭 개수에 따른 수행시간 차이가 매우 적었기 때문에 50개를 기준으로 측정하였다. 개별 지수 방법의 수행시간을 보면 유사도 지수가 순서 관계 지수나 인기도 지수에 비해 오래 걸리는 것을 알 수 있는데 이것은 한 영화와 다른 모든 영

화의 유사도를 각각의 파일로 저장해놓고 그것을 불러오는 방식으로 구현했기 때문인 것으로 보인다.

6. 결 론

본 논문에서는 동영상 콘텐츠 간의 순서관계와 유사도, 인기도를 고려하여 사용자가 다음에 볼 동영상 콘텐츠를 예측하고 그 동영상 콘텐츠를 클라이언트에 프리페칭 함으로써 전송 지연 없이 동영상 콘텐츠를 시청할 수 있는 프리페칭 시스템을 제안하였고 세 종류의 지수를 개별적으로 적용시켰을 때와 결합하여 적용시켰을 때의 프리페칭 정확도를 측정하였다. 지수의 계산 방법의 차이 때문에 단순 합산으로 적용이 불가능하여 표준화하여 합산하는 방법을 적용 하였고, 합산 과정에서 가중치를 부여하여 세 지수의 영향력을 변경하여 실험하였다.

결론적으로 세 지수중에서 가장 높은 프리페칭 정확도를 보였던 순서 관계가 결합 실험에서도 가장 영향력이 큰 것으로 나타났으며, 어떤 식으로 결합하던지 가장 높은 정확도를 보인 지수를 단독으로 하는 것이 여전히 가장 높은 정확도를 보이는지를 확인하기 위해선 또 다른 지수를 고려하거나 지수 계산 방법을 변경한 추가적인 실험이 필요할 것으로 보인다.

향후에는 사용자 개인의 영화 선택 경향을 파악하여 프리페칭 지수의 가중치를 조절함으로써 프리페칭 대상 콘텐츠를 선정하는 방법에 대해 연구할 계획이다. 이 방법은 본 논문에서 제안하는 방법과 달리 각 사용자의 기존 시청 기록에 대한 분석을 통해 사용자마다 다른 가중치를 적용하는 것이다. 또한 두 방법을 병행하여 사용할 수 있다면 익명 사용자 및 시청 기록이 없는 사용자 뿐 아니라 시청 기록이 누적되고 있는 사용자까지도 프리페칭의 이득으로 전송 지연 없이 동영상 콘텐츠를 시청할 수 있을 것이다.

실험에 사용한 무비렌즈 데이터 셋은 동영상 콘텐츠 중에서도 영화 콘텐츠를 사용자가 평가한 데이터를 모아놓은 것으로 본 논문의 다중 지수를 고려한 프리페칭 기법을 실험하기 위한 데이터 셋으로 완벽하게 적합하지는 않았다. 하지만 사용자의 시청 기록 데이터 셋은 프라이버시 문제가 있기 때문에 공개된 데이터 셋이 없었고 별도로 구하는 데에 어려움이 있어 무비렌즈 데이터 셋을 각 유저의 기록으로 나누고 클러스터로 묶는 방법을 통해 시청 기록으로 간주하여 실험하였다. 사용자 별 시청 기록 데이터를 구하고 추가로 영화가 아닌 다른 장르의 동영상 콘텐츠 또는 일반 사용자 동영상 콘텐츠에 대한 데이터 셋을 구하여 연구할 계획이다.

감사의 글

이 논문은 2017년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.NRF-2015R1D1A1A01056848)

References

[1] YangChan Moon, "A prefetching system using order relation among video content", Masters thesis, Konkuk University, 2015.

[2] Pazzani, Michael J, and Daniel Billsus, "Content-based recommendation systems", The adaptive web. Springer Berlin Heidelberg, pp.325-341, 2007.

[3] Davidson, James, et al, "The YouTube video recommendation system", In Proceedings of the fourth ACM conference on Recommender systems, ACM, pp. 293-296, 2010.

[4] Schafer, J. Ben, et al, "Collaborative filtering recommender systems", The adaptive web. Springer Berlin Heidelberg, pp. 291-324, 2007.

[5] Adomavicius, Gediminas, and Alexander Tuzhilin. "Context-aware recommender systems", Recommender systems handbook, Springer US, pp.217-253, 2011.

[6] Domnech, Josep, et al, "Web prefetching performance metrics: A survey", Performance Evaluation, pp. 988-1004, 2006.

[7] FAN, Li, et al, "Web prefetching between low-bandwidth clients and proxies potential and performance", ACM SIGMETRICS Performance Evaluation Review, pp.178-187, 1999.

[8] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4, Article 19 (December 2015), 19 pages. DOI=http://dx.doi.org/10.1145/2827872.



임민규(Mingyu Lim)

1998년 KAIST 전산학과 공학사
 2000년 ICU 공학부 공학석사
 2006년 ICU 공학부 공학박사
 2006년~2008년 제네바대학교 MIRALab 선
 임연연구원
 2009년~2016년 건국대학교 인터넷미디어공
 학과 교수
 2017년~현재 건국대학교 스마트ICT융합공학
 과, 소프트웨어학과 교수

저자 소개



문양찬 (YangChan Moon)

2012년 건국대학교 인터넷미디어공학과
 공학사
 2015년 건국대학교 인터넷미디어공학과
 공학석사
 2015년~현재 건국대학교 인터넷미디어
 공학과 박사과정