Design exam
20180789 이해준

Observation

SongTube의 사용자 수가 증가함에 따라 비디오 송출을 위해 사용하는 메모리의 크기의 한계로 인해 Thrashing이 빈번하게 발생하였다. 현재 SongTube의 비디오 송출 알고리즘은 유저의 동영상 리퀘스트를 받으면 해당 동영상 전체를 메모리에 로드하는 것이다.

이 문제상황을 해결하기 위해 인터넷 스트리밍 서비스를 이용하는 사용자들의 경향성을 분석했다. 해당 분석 결과는 다음과 같다.

1. 쇼츠 시청률 > 긴 비디오 시청률
긴 길이의 비디오보다 1분 미만의 짧은 쇼츠 형태의 비디오가 시청률이 월등히 높았다. 시청량의 90%는 짧은 쇼츠인 것으로 나타났다.

2. 시청 도중 나가기
인터넷 스트리밍 동영상의 경우 대부분 동영상의 초반 부분의 시청률에 비해 후반 부분의 시청률은 낮다. 이는 동영상을 보던 중 많은 사람들이 동영상 재생을 그만둔다는 의미이다.

3. 시청 수 > 업로드 수
동영상 시청 리퀘스트보다 동영상 업로드 리퀘스트의 수가 압도적으로 적었다. 이는 대부분의 사용자가 동영상의 시청을 위해 플랫폼을 이용함을 의미한다.

위 경향성으로 볼 때 현재의 알고리즘은 다음과 같은 문제가 있다.
1. 큰 메모리 소모
전체 동영상을 처음 시청할 때에 로드하므로, 메모리의 소모가 크다. 특히, 긴 동영상의 시청 시에는, 많은 메모리를 소모하여 이전에 로드된 다른 데이터를 swap out하므로 이로 인해 많은 page fault가 발생한다.

2. 시청하지 않는 영상 로드
동영상 재생 시작시에 전체 동영상을 로드하기 때문에 후반부에 시청률이 낮은 경우의 영상은 시청률에 비해 소모하는 메모리의 공간이 크다. 시청하지 않는 동영상 부분의 메모리 로드로 다른 데이터가 swap out되어 비효율이 발생한다.

3. 인기 동영상의 잦은 swap in
사용자의 수가 많아짐에 따라, 대부분의 동영상이 지속적으로 메모리에 있지 못하고, 디스크로 swap out된다. 이에 따라 인기가 많은 동영상의 경우 메모리에 지속적으로 swap in과 swap out되어 비효율이 발생한다.

Idea

이러한 현재 SongTube의 디자인을 해결하기 위해 다음과 같은 아이디어를 제시한다.

1. 일부 로드

유저가 동영상 재생을 시작할 때, 전체 동영상을 한 번에 메모리에 로드하는 것이 아닌, 현재 시점으로부터 일정 시간 후의 영상만을 메모리에 로드한다. 단, 유저가 동영상을 배속한 경우, 동영상 길이를 기준으로가 아닌, 송출 시간을 기준으로 동영상을 로드한다. 예를 들어, 10초만 로드하는 경우, 유저가 동영상을 1.5배속으로 시청하는 경우, 동영상 길이 기준으로 15초 후의 영상까지 로드한다.

2. 인기 급상승 쇼츠 영상 메모리 고정

인기 영상이 시청 수의 대부분을 담당한다. 따라서, 최근 자주 로드되는 쇼츠의 경우, 메모리에서 swap out하지 않고, 전체를 메모리 내에 로드한 상태로 유지한다. 이로 인해 시청이 잦은 인기 쇼츠를 디스크에 swap-in / out에 사용하는 시간을 줄인다. 일정 시간동안 인기 쇼츠의 전체 동영상 중 시청 비율을 고려하여, 쇼츠의 데이터를 항상 로드해 둘 메모리의 공간 크기를 결정한다.

3. 추천 알고리즘 개선

인기 급상승 쇼츠 영상을 추천 영상으로 제공하는 비율을 늘린다. 유저가 인기 급상승중인 어떤 쇼츠 영상을 보지 않았을 경우 이를 추천해주는 비율을 늘린다. 인기 영상은 유저 반응이 좋은 경향성을 보이므로 이를 통해 메모리의 사용성을 향상시킬 수 있을 뿐 아니라, UX또한 향상시킬 수 있다.

4. 업로드 메모리 제한

UI와 UX에 대한 기대감은 일반적으로 업로드시보다, 시청 시에 높다. 일반적으로 동영상 시청 시에는 유저가 화면을 지속적으로 시청하며 끊기지 않는 동영상을 기대하지만, 크리에이터가 업로드할 시에는 User interaction이 크게 중요하지 않다. 따라서, 업로드에 사용되는 메모리의 크기를 제한하여 유저 경험을 개선한다. 업로드 / 시청 비율을 지속적으로 계산하여, 업로드에 사용하는 메모리의 크기를 이 비율의 절반으로 고정한다.

Design
위에서 살펴 본 아이디어에 대해 각각 구체화한 디자인은 다음과 같다.

1. 10 Sec Load
해당 디자인은 유저가 동영상을 재생할 때, 송출 시간 기준으로 현재 시점으로부터 짧은 시간만큼의 영상 데이터만을 메모리에 로드한다. 짧은 시간을 로드한다는 것을 효과적으로 알 수 있도록 해당 메모리를 로드하는 디자인의 이름을 “10 Sec Load”라고 명명하였다.

동영상을 로드할 길이는 동영상의 인기에 따라 5초~15초 사이로 다르게 지정한다. 인기가 많은 동영상의 경우 15초까지 미리 데이터를 로드하고, 인기가 거의 없는 동영상은 5초까지만 데이터를 로드한다. 이는 인기가 많은 영상의 시청 수가 많기 때문에, 더 긴 길이만큼을 로딩함으로써 동영상 끊김을 경험하는 유저의 수를 감소시켜 전체적인 유저 경험을 향상시키기 위함이다.

동영상의 인기를 파악하기 위해 각 동영상에 popularity라는 variable을 정의하였고, 인기는 시간에 따라 달라질 수 있으므로, 이를 시간에 따라 지속적으로 변화시키고자 한다. 단, 모든 동영상의 popularity를 계산하여 지속적으로 반영하는 것은 Overhead를 증가시켜 성능을 하락시킬 수 있다. 따라서, 어떤 동영상을 재생할 때 이의 popularity를 전체 글로벌 변수를 빼서 계산하기 위해 전체 글로벌 변수를 지속적으로 증가시키고, 최근 재생된 동영상의 popularity만을 변화시킨다.

이 전체 글로벌 변수를 global_time이라고 하자. 이는 0부터 1시간에 한 번씩 1씩 증가하고, 10000이 될 경우 0으로 초기화된다. 이 global_time이 업데이트 될 때, 최근 재생된 동영상 중 많이 시청된 상위 20% 동영상의 popularity를 2 증가시킨다. 어떤 동영상이 재생될 때, 해당 동영상의 popularity - global_time이 0 이하일 경우 동영상의 로드 길이를 5초만 로드하고, 1씩 증가할 때마다 로드 길이를 1초씩 늘려, 10 이상일 경우 15초를 미리 로드한다.

예를 들어, global_time이 0일 때, 어떤 동영상의 popularity가 1일 경우 이 동영상은 6초씩 미리 로드되고, 해당 동영상이 1시간동안 시청된 순위가 상위 20%아래일 경우, 다음 번 업데이트 이후에 계산된 global_time - popularity의 값은 0이 되어, 5초만 로드된다.

또한, 1시간동안 시청된 동영상을 파악하기 위해 watched_list를 정의한다. 이 리스트에 최근 한시간동안 시청된 동영상의 ID정보가 들어간다. 또, 동영상별로 prev_count라는 변수를 추가하여, 한 시간 이전의 조회수를 저장한다. 이 변수들은 global_time이 업데이트 될 때 같이 초기화 또는 업데이트된다. global_time이 초기화될 때, 모든 영상의 popularity는 0으로 초기화된다.

2. Hot clip fix
해당 디자인은 인기가 많은 쇼츠 영상을 메모리에서 swap out하지 않고 메모리에 지속적으로 남아있도록 한다. 이렇게 메모리에서 지속적으로 남아 있는 쇼츠 영상을 인기 급상승 쇼츠라고 명명한다. 인기 급상승 쇼츠를 메모리에서 swap-out하지 않고 고정시킨다는 것을 지속적으로 알 수 있도록 해당 디자인의 이름을 “Hot clip fix”라고 명명하였다.

쇼츠 영상이 아닌 긴 길이의 영상은 메모리에 고정시키지 않는다. 이는 길이가 긴 영상의 인기가 일반적으로 쇼츠 영상에 비해 떨어진다는 이유도 있지만, 사용자의 UX에 대한

기대감이 쇼츠 영상에서 더 높은 측면도 있다. 긴 영상 중에는 “음악”, “요리 레시피” 등처럼 다른 일을 하면서 시청하는 영상의 비율이 높은 등 쇼츠 영상에 비해 유저 상호작용의 중요성이 떨어진다. 이 뿐 아니라, 쇼츠를 더 자주 시청하는 사용자 풀의 UX에 대한 기대감이 더 높은 것으로 나타났다.

쇼츠 영상 중 위에서 계산한 popularity의 값이 높은 상위 75개의 동영상을 인기 급상승 쇼츠로 지정한다. 해당 수는 SongTube의 사용자 수가 증가함에 따라 수정되어야 한다.

3. Watch this first!

해당 디자인은 인기가 많은 쇼츠 영상을 더 자주 추천하도록 추천 알고리즘을 변화시킨다. 따라서 해당 디자인을 인기가 많은 쇼츠를 추천한다는 것을 직관적으로 알 수 있도록 “Watch this first!”라고 명명하였다.

현재 추천 알고리즘은 사용자의 시청 기록과 사용자와 유사한 사용자들의 시청 기록을 토대로 동영상을 추천하는데, 추천을 위해 사용자별로 추천할 영상에 점수를 매겨 점수가 높은 영상을 먼저 추천한다. 이 추천 알고리즘에서는 인기 급상승 쇼츠의 점수를 50% 가중치를 더 줘서 우선순위를 높인다.

이를 통해 사용자가 어느 정도 관심이 있는 주제라면, 해당 주제 내에서 인기가 급상승중인 쇼츠의 추천 우선순위를 더 향상시켜 UX를 개선할 수 있을 뿐 아니라, 메모리에서 swap out하지 않는 인기 급상승 쇼츠의 시청량을 늘려 메모리의 효율성 또한 높일 수 있다.

4. Limit upload

일반적으로 유저 UX 기대치는 업로드 시에 많이 떨어지므로, 전체 요청 중 업로드 요청 수의 비율의 절반에 해당하는 비율만큼만 업로드의 메모리를 제한하고자 한다. 이를 직관적으로 알 수 있게 하기 위해 해당 디자인의 이름을 “Limit upload”라고 명명하였다.

동영상 업로드 시에 일반적으로 화면을 끝까지 보고 있는 사람은 적을 것이다. 따라서, 이 경우 반응성을 조금 감소시키더라도 유저 경험은 크게 악화되지 않는다. 따라서 이에 소모되는 메모리 리소스를 절약하여 시청 시에 사용한다면 메모리 유용성과 UX는 개선될 수 있다.

Discussion

해당 디자인들은 발생한 문제상황인 Thrashing의 빈번한 발생을 효과적으로 감소시킨다.

먼저, 동영상 전체를 로드하지 않고 5~15초정도만 지속적으로 로드하기 때문에, 기존에 전체 동영상을 로드하여 발생했던 메모리 소모에 비해 메모리 소모를 크게 감소시켜, Thrashing을 효과적으로 감소시켰다.

또한, 자주 시청되는 인기 급상승 쇼츠들은 메모리에 고정되어 있기 때문에, 이 영상을 swap-in / out 하는 데에 소모되는 시간이 줄어들어 성능이 향상되고, 이 영역이 100GB에서 차지하는 비율이 크지 않으므로, 메모리에 큰 부담을 주지 않는다. 또한, 이를 추천하도록 알고리즘을 변화시켜, 다른 영상 시청에 소모될 메모리의 리소스를 감소시키고, 이를 유저의 취향에 기반하기 때문에 UX또한 향상시켰다.

마지막으로, 업로드에 사용되는 메모리 리소스를 감소시켜 메모리의 사용성을 더욱 증가시켰다. 업로드에서는 메모리를 많이 할당하지 않더라도 UX가 크게 감소하지 않기

때문에 이 메모리를 절약하여 다른 동영상 시청에 사용한다면 UX와 메모리 유용성을 더욱 향상시킬 수 있을 것이다.

하지만, 해당 디자인들은 SongTube의 사용자 수가 증가한다면 효과적으로 작동하지 않을 수 있다는 단점이 있다. 예를 들어, 10 Sec Load와 Hot Clip Fix, Watch This First에서 사용되는 동영상의 popularity의 경우 현재 상황에서는 유저의 수가 크지 않기 때문에 한 시간동안 시청된 영상의 수가 많지 않지만, 유저의 수가 증가할수록 popularity를 업데이트하는 데에 소모되는 Overhead가 증가할 것이다. 또한, 해당 시청된 동영상을 list형태로 저장하기 때문에, 동영상을 추가하고, 검색하는 데에 많은 시간이 소요될 수 있다.

또한, Workload측면에서만 개선할 방안들을 제시하였기 때문에, 플랫폼 이용자 수가 증가한다면, VM system등을 근본적으로 개선하여 미래 플랫폼이 거대해질 상황을 대비할 필요가 있다.