

OTT 서비스 Cold-Start 해결 방안

지도교수 : 이강호

연구자 : 윤선아

< 목 차 >

1. 서론

- 1.1 OTT 서비스
- 1.2 추천 알고리즘이란?
 - 1.2.1 혼합형 필터링

- 2.1.3 New User
- 2.2 Cold-Start 해결 방법
 - 2.2.1 머신러닝 Active Learning
 - 2.2.2 하이브리드 추천시스템

2. 본론

- 2.1 Cold-Start
 - 2.1.1 New Community
 - 2.1.2 New item

3. 결론

요 약

모바일 중심의 콘텐츠 소비가 지속되면서 영상 콘텐츠에 대한 수요가 꾸준히 높아지고 OTT(Over-The-Top) 서비스는 미디어 시장에서 가장 주목받고 있다. OTT 서비스는 기존에 TV 시청을 위해 사용했던 셋톱박스를 넘어 인터넷을 기반으로 언제 어디서나 쉽게 콘텐츠를 제공하는 온라인 콘텐츠 스트리밍 서비스를 말한다.

국내 방송시장에서 OTT 서비스 소비자의 로그 데이터는 방대해지고 이용은 빠르게 증가하는 요즘 콘텐츠 제공자와 중개업자는 많은 조회 수를 얻기 위한 맞춤 추천시스템에 대한 요구가 높아지고 있다.

본 논문에서는 OTT 서비스의 추천시스템에서 자주 발생하는 Cold-Start 문제의 원인을 머신러닝과 하이브리드 추천시스템을 통해 해결하고 영상 추천시스템을 제안한다.

주요어 : OTT 서비스, 추천시스템, 혼합형 필터링, Cold-Start

1. 서론

1.1 OTT 서비스

최근 국내 방송시장에서 OTT(Over-The-Top 이하 OTT)가 주목할 만한 성장세를 보이고 있다. OTT 서비스는 인터넷을 통해 방송 프로그램·영화·교육 등 각종 미디어 콘텐츠를 제공하는 서비스를 말한다. Over-the-X는 '기존의 범위를 넘어서'라는 뜻이 있다. Top은 TV 셋톱박스 같은 단말기를 의미한다. 따라서 직역하면 셋톱박스를 통해서, 넘어서의 의미를 가진다. 이처럼 OTT 서비스는 초기에 단말기를 통해 영화, TV 프로그램 등 프리미엄 콘텐츠를 VOD 방식으로 제공하는 서비스를 지칭했다.



[사진 1] OTT 서비스

글로벌 OTT 시장의 경우에는 연평균 성장률이 39.4%로 가파른 상승세를 보이며, 가입자 수 역시 5억 9,500만 명으로 지속적인 성장세를 보인다. OTT는 TV를 포함한 스마트폰이나 태블릿PC 등 다양한 스마트미디어를 통해 이용 가능하다는 특성을 바탕으로 국내 방송시장에서 새로운 동영상 콘텐츠 서비스 강자로 급성장하고 있다. 이후 인터넷 기술변화에 따라 콘텐츠 유통이 모바일까지 포함하면서 OTT의 의미가 확대됐다. 2020년 2월 26일 발표된 과학기술정보통신부의 '2019 인터넷 이용실태조사 결과' 보고서에 따르면 국민의 81.2%가 인터넷을 통해 동영상 서비스를 이용하며, 73.7%는 매일 동영상 서비스를 이용하는 것으로 나타났다. 동영상 사용량이 늘자 온라인 OTT 서비스의 경쟁은 점차 과열되는 모습을 보이지만, 국내 동영상 플랫폼 사용량 1위인 유튜브와 '구독자 수의 꾸준한 증가' '동영상 플랫폼 만족도 1위'라는 타이틀을 가진 넷플릭스를 주목할 수 있을 것이다. 유튜브나 넷플릭스에서 나의 취향에 맞는 또는 나를 간파한 듯 영상을 추천하게 되는데 이것이 OTT 서비스에서 제공하는 인터넷 추천시스템이다.

인터넷 추천시스템은 사용자가 선호하는 항목 정보에 기반을 두어 사용자가 전에는 몰랐던 새롭고 좋아할 만한 것을 예측하여 추천해주는 시스템을 뜻한다. 이러한 추천시스템은 인터넷의 발전이 가져다준 정보의 홍수 속에서 사용자는 자신의 취향에 맞는 제품 혹은 콘텐츠를 사용자 스스로 찾아야 하는 불편함이나 문제점을 해결해줄 수 있다. 추천

시스템을 통한 장점은 4가지로 분류할 수 있다.

첫째, 제공자와 사용자 모두 이득이 된다. 추천시스템을 통하여 사용자는 원하는 항목을 찾기 위한 시간과 비용을 감소할 수 있으며, 잠재적 요구에 합당하는 항목을 발견할 기회를 제공받는다.

둘째, 일반 방문자의 구매 혹은 사용자로 전환 가능성이 크다 일반 방문자의 성향에 맞는 항목 정보를 받음으로써 상품의 구매 혹은 콘텐츠 이용자로서 전환된다.

셋째, 관련 항목 판매 사용이 증가한다. 사용자가 선호할 제품이나 관련 콘텐츠를 지속적으로 제공하여 구매 혹은 선택했던 제품의 관련된 항목의 판매나 사용이 증가한다.

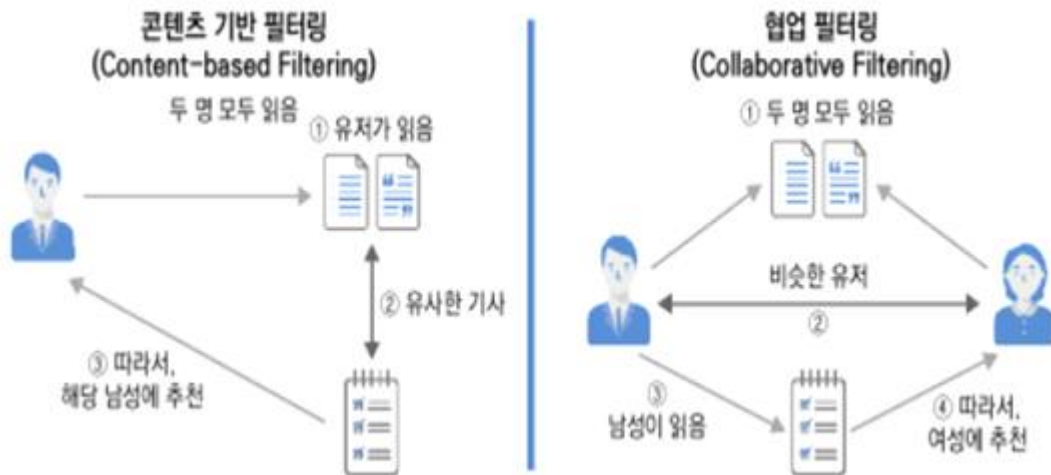
그리고 넷째는 지속적 관계가 유지된다. 추천시스템으로 인해 만족도를 느낀 사용자는 해당 제품 및 콘텐츠 제공자에 대한 신뢰도가 상승하여 타 사이트로의 전환이 감소한다.

1.2 추천 알고리즘이란?

알고리즘이란 어떠한 문제를 해결하기 위해 정한 절차 또는 규칙의 모음이라고 볼 수 있다. 조금 더 정확한 의미를 따져보자면 알고리즘은 어떠한 행동을 하기 위해서 만들어진 명령어들의 유한 집합(finite set)이다. 컴퓨터 프로그램은 정교한 알고리즘들의 집합이라고 간주할 수 있다. 수학이나 컴퓨터 과학에서 말하는 알고리즘은 보통 반복되는 문제를 풀기 위한 작은 진행 절차를 의미한다. 컴퓨터 시대 이후로는 알고리즘이라고 하면 컴퓨터를 통해 실행되는 것이라고 여겨지는 경향이 있으나 사실 알고리즘 자체는 컴퓨터가 등장하기 이전부터도 존재했다. 즉 사람이 수동으로 종이를 사용해 일정한 절차로 문제를 풀더라도 알고리즘에 해당한다.

다만, 컴퓨터의 등장과 함께 알고리즘 역시 급속도로 발전하게 된 것은 사실이다. 포털 검색 결과를 사용자가 원하는 방향으로 제공하거나 저작권 침해 콘텐츠 혹은 해로운 콘텐츠를 골라내는 등 다양한 방면으로 알고리즘이 활용되고 있다. 그중 우리의 일상에서 가장 익숙한 것은 추천 알고리즘이다. 이와 더불어 추천 알고리즘은 방대한 데이터 중에서 일정한 규칙에 따라 사용자가 좋아할 만한 콘텐츠를 추천해 사용자의 만족도를 높여준다. 그리고 기업에는 사용자의 플랫폼 체류 시간 및 콘텐츠 운영의 효율성을 높여 구독자 이탈을 막는 데 효과적인 전략이다.

알고리즘이 추천할 콘텐츠를 선택하는 방법은 대표적으로 혼합형 필터링이 있는데 사진2와 같이 방법은 크게 ‘콘텐츠 기반 필터링’과 ‘협업 필터링’의 두 갈래로 나누어진다.



[사진 2] 혼합형 필터링

1.2.1 혼합형 필터링

혼합형 필터링은 두 가지 이상 단일 필터링 기법을 혼합하여 사용하는 방법을 말한다. 콘텐츠 기반 필터링은 콘텐츠 정보를 기반으로 다른 콘텐츠를 추천하는 방식이다. 영화 콘텐츠의 경우라면 영화의 줄거리와 등장 배우 그리고 장르 등을 데이터화하고 상품이라면 상품의 상세정보를 분석하여 추천하는 방법이다. 콘텐츠 기반 필터링의 장점은 콘텐츠 자체를 분석하는 것이기 때문에 초기에 사용자의 행동 데이터가 적더라도 추천할 수 있다는 것이다. 하지만 콘텐츠 기반 필터링 역시 한계점을 갖는다. 콘텐츠의 정보를 모두 함축하는 것에 어려움이 따른다는 점이다. 예를 들어 K-POP 그룹 EXO의 멤버들 중 세훈의 팬은 멤버 세훈만이 등장하는 추천 콘텐츠를 바라지만 겹치는 콘텐츠 정보가 많아 원하지 않는 그룹 전체 혹은 그룹 내 다른 멤버의 활동 모습이 담긴 콘텐츠를 추천받기도 한다. 이렇듯 알고리즘의 입장에서 이용자의 성향을 세부적으로 파악하기가 어렵다는 문제가 발생한다.

협업 필터링은 많은 사용자로부터 얻은 기호 정보에 따라 사용자들의 관심사를 자동으로 예측하게 해주는 방법이다. 같은 행동을 한 사람들을 하나의 프로파일링 그룹으로 묶어서 그룹 내의 사람들이 공통으로 봤던 콘텐츠를 추천하는 방식이다. 예를 들어 쇼핑몰에서 상품을 구매하면 해당 상품을 구매한 사람들이 구매한 다른 상품들을 추천 상품으로 보여주거나 SNS에서 나와 친구를 맺고 있는 사람들의 친구들을 자동으로 추천해주는 개념이다. 하지만 협업 필터링의 경우, 기존 데이터가 없는 신규 사용자의 경우 추천이 힘들다는 점, 사용자가 많아질수록 추천에 계산시간이 오래 걸린다는 점 그리고 다수의 사용자가 관심을 보이는 소수의 콘텐츠가 전체 추천 콘텐츠로 보이는 비율이 높아져 소외되는 콘텐츠가 생긴다는 점 등의 한계를 갖는다. 이렇듯 협업 필터링과 콘텐츠 기반 필터링에서 발생하는 단점을 극복하기 위하여 두 가지 방법 이상을 혼합하기도 하고 또한 협업 필터링에서 사용자 기반 협업 필터링과 콘텐츠 기반 협업 필터링을 결합한 필터링도 혼합형 필터링의 방법 중 하나라고 볼 수 있다. 혼합형 필터링은 현재 대부분 추천시스템에서 단일 필터링 추천 기법의 단점을 보완하여 사용되고 있다.

2. 본론

2.1 Cold-Start

Cold-Start란 추천시스템이 새로운 또는 어떤 사용자들에 대한 충분한 정보가 수집된 상태가 아니라서 해당 사용자들에게 적절한 영상을 추천해주지 못하는 문제를 말한다. 앞서 말한 OTT 서비스의 핵심 알고리즘 기술인 추천시스템에서 사용되는 필터링은 사용자의 정보를 활용하여 적합한 정보를 제공하지만 새로운 사용자에게는 적합한 정보를 제공하지 못하는 Cold-Start 문제가 발생한다. 물론 기업마다 적용하고 있는 추천시스템에 따라 다르겠지만 사용자들의 주로 영상들에 대한 평점, 시청 이력, 좋아요 이력 등과 같은 일종의 상호작용 정보들을 이용한다. 이 문제가 발생하는 이유는 크게 3가지가 존재한다.

2.1.1 New Community

여기서 새로운 커뮤니티란 하나의 회사나 서비스가 설립되고 운영되기 시작한 것이라고 볼 수 있다. 예를 들어, 어떤 서비스가 운영되기 시작했을 때 해당 서비스를 이용하는 사용자들은 단 한 명도 없을 것이다. 물론 서비스에 이미 판매하는 제품들의 카탈로그와 같은 것들이 존재하더라도 이 제품들에 대한 사용자들의 상호작용 정보들이 없으므로 Cold-Start 문제가 필연적으로 발생한다.

2.1.2 New item

신작이 출시된 경우에는 사용하는 추천시스템의 종류에 따라 Cold-Start 문제 발생 경향이 달라지기도 한다. 우선 협업 필터링부터 살펴보자면 협업 필터링은 적절한 영상을 추천해주기 위해 영상에 대한 사용자의 상호작용에 주로 의존하게 된다. 그런데 신작이 출시되었을 때, 신작에 대한 사용자들의 상호작용 정보는 아예 없기 마련이다. 설사 정보가 조금 있더라도 이를 기반으로 협업 필터링 추천시스템을 만들게 된다면 추천의 질이 낮아지게 된다. 따라서 협업 필터링의 추천시스템이면 Cold-Start 문제에 매우 민감하다. 그래서 신작이 나오게 되면 실제로는 정말 좋은 영상이라도 인기 없는 영상들로 치부되어 사람들에게 잘 추천되지 않는 경우가 발생한다.

반대로 콘텐츠 기반 필터링에서는 신제품이 출현했을 때 Cold-Start 문제에 걸리는 경향이 상대적으로 덜하다. 왜냐하면, 콘텐츠 기반 필터링은 영상에 대한 사용자의 상호작용 정보를 기반으로 하지 않고 영상 자체의 특징을 기반으로 추천해주는 시스템이기 때문이다.

그래서 신작이 등장했을 때 콘텐츠 기반 필터링은 제품 자체의 특징을 파악해 기존의 영상들과 유사도를 측정해서 추천해줄 수 있다. 하지만 콘텐츠 기반 필터링도 만약 사용자에서 도출된 특징을 기반으로 구현되었다면 Cold-Start 문제에 빠질 가능성이 크다. 예를 들면 제품에 대한 사용자의 리뷰 텍스트, 태그 텍스트와 같은 특징들로 콘텐츠 기반 필터링을 구현할 경우이다.

2.1.3 New user

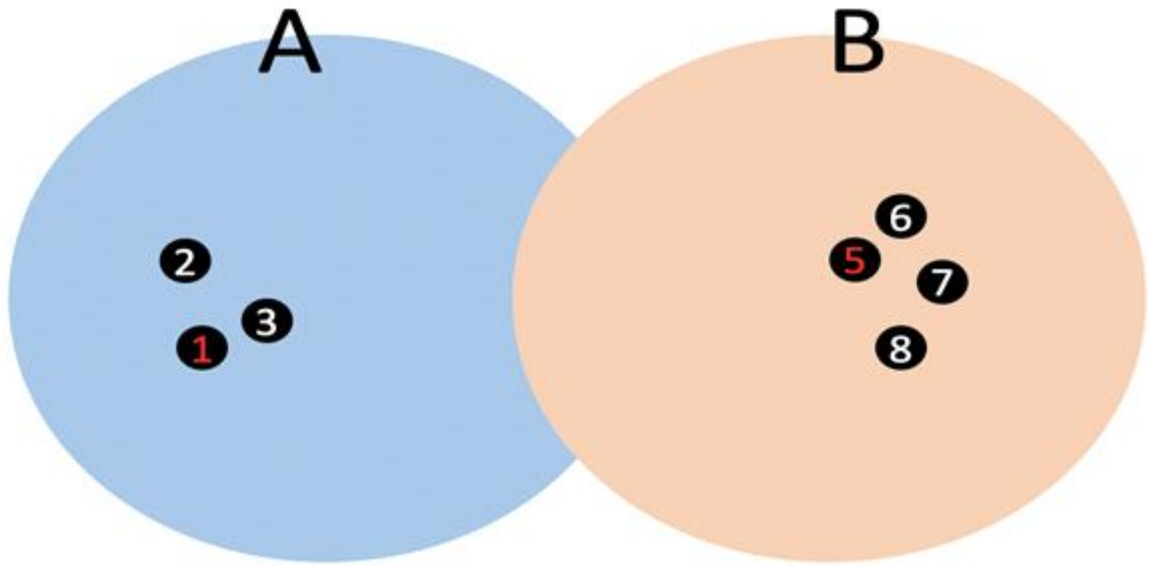
새로운 사용자가 가입했을 때이다. 이 경우 신규 사용자에게 대한 역사가 없으므로 Cold-Start 문제가 주로 발생한다. 보통은 이 문제를 막기 위해 신규 사용자의 초기 프로필을 작성하도록 하면서 그 신규 사용자의 기호가 무엇인지 파악한다. 하지만 이때 초기 프로필 작성 시 시간이 많이 소요된다거나 하는 문제로 인해 사용자들의 귀찮음을 유발하게 되면 오히려 잘못된 정보가 기재되거나 사용자들이 그냥 넘겨버리는 경우가 많다. 나 역시 수많은 OTT 서비스들을 이용하면서 프로필 작성 시 대충하는 경향이 있다. 그래서 새로운 사용자가 들어왔을 때 발생하는 Cold-Start 문제에 대해서도 추천 시스템의 종류에 따라 약간 다르게 대응을 해야 한다. 우선 콘텐츠 기반 필터링 추천 시스템은 신규 사용자가 다른 사용자의 기호들과 얼마나 관련성이 있는지 파악을 해야 한다. 그래서 제품들에 대한 신규 사용자의 기호인 상호작용 정보를 파악해야만 하는데 신규 사용자의 경우 제품들에 대해 평가한 역사가 거의 없는 상태이다. 그래서 Cold-Start 문제에 취약하다. 반면 콘텐츠 기반 필터링 중 사용자 기반인 추천시스템은 상대적으로 Cold-Start 문제에 강건하다. 왜냐하면, 주로 사용자의 특성만을 기반으로 하기 때문이다. 따라서 신규 사용자의 초기 프로필만 어느 정도 잘 작성이 된다면 이 정보를 이용해 User-User 추천시스템을 잘 구현해 Cold-Start 문제를 잘 해결할 수 있다. 그래서 여러 기업들은 초기 프로필을 어떻게 든 완성시키기 위해 신규 사용자에게 대놓고 명시적으로 질문지를 던지거나 사용자의 연동된 SNS 계정의 프로필 정보를 이용하기도 한다.

2.2 Cold-Start 해결 방법

기존의 제품 또는 사용자(이를 Warm items or users라고도 함)들에게는 협업 필터링을, 새로운 영상 또는 사용자(Cold items or users)에는 콘텐츠 기반 필터링을 수행한다. 결국 이 사용자가 서비스에 진입한 지 얼마만큼 되었는지 또는 얼마만큼의 역사를 남겼는지에 따라 추천시스템 종류를 선택적으로 구현하는 것이다. 하지만 이 방법도 단점이 존재한다. 바로 제품이나 사용자의 특성에 대한 전반적인 설명을 파악하기 어려운 상황에서는 콘텐츠 기반 필터링의 추천 질이 매우 낮아진다는 점이다. 그래서 이를 예방하기 위해 만약 신규 사용자의 개인적인 정보들이 입력되어 있지 않거나 영상의 전반적인 설명이 게시되어 있지 않다면 개인화된 추천이 아닌 전체적으로 인기가 많은 영상을 추천해주기도 한다. 이러한 Cold-Start 문제를 해결하는 방법에는 현재 머신러닝과 하이브리드 추천시스템을 이용하는 것이다.

2.2.1 머신러닝 Active Learning

머신러닝 Active Learning을 이용하는 것이다. 즉 사용자가 영상을 평가하도록 하는 것이다. 이때 사용자가 평가할 영상들은 추천시스템의 관점에서 가장 정보를 많이 담고 있을 상위의 영상들만 추출한 것들이다. Active Learning과 추천시스템을 어떻게 사용하는지에 관해서 한 가지 예시를 들어보자면 [사진 3]과 같다.



[사진 3] 머신러닝 데이터 분류 예시

먼저 A, B 두 집단으로 분류되는 영상 또는 소비자들이 있다고 가정하자. 그리고 무수히 많은 데이터들 중 추천 시스템의 관점에서 보았을 때 정보를 가장 많이 담고 있는 제품 2개를 뽑아서 보여주었다. 그것은 A집단의 1번 영상, B집단의 5번 영상이다. 이때 유저가 1번, 5번 영상에 대해 각각 평가를 한다. 그러면 이제 1번, 5번 영상에 대한 유저의 평가를 기준으로 다른 영상들을 추천해준다. 예를 들어, 유저가 A집단에 속한 1번 데이터에 좋은 평가를 주었다고 해보자. 그렇다면 추천 시스템은 이 평가만을 기반으로 1번과 가까운 거리에 있는 2번, 3번 영상들도 유사한 성격의 영상으로 판단하여 2번, 3번 영상을 유저에게 추천해준다. 설사 동시에 B집단의 5번 영상에도 좋은 평가를 주었다 하더라도 추천 시스템은 좋은 평가를 해준 1번과 비슷한 영상들, 5번과 비슷한 영상들 모두 유저에게 추천해줄 것이다.

2.2.2 하이브리드 추천시스템

하이브리드 추천시스템은 협업 필터링과 콘텐츠 기반 필터링을 조합하여 상호 보완적으로 개발된 알고리즘이다. 신규 콘텐츠는 콘텐츠 기반 필터링 기술로 분석하여 추천하고 충분한 데이터 쌓인 후부터는 협업 필터링으로 추천의 정확성을 높이는 방식이다. 예를 들어 영화 스트리밍 서비스의 이용자 A는 나 홀로 집에 1편의 주인공인 맥컬리 컬킨(Macaulay Culkin) 배우가 나오는 영화를 주로 선호한다. 그래서 나 홀로 집에 1편을 보았다. 영화를 보고 난 후 스트리밍 서비스가 A에게 영화를 추천해준다. 이때 일반적인 콘텐츠 기반 필터링이라면 단순히 나 홀로 집에 2, 3, 4편을 추천해줄 것이다. 하지만 하이브리드 추천시스템이라면 나 홀로 집에 2편과 다른 맥컬리 컬킨(Macaulay Culkin) 배우가 나온 영화들을 추천해줄 것이다. 즉, 하이브리드 추천시스템은 사용자가 중요하다고 생각되는 요소와 영화라고 한다면 출연 배우와 영화감독에 더 가중치를 두어 같은 배우가 출연한 영화와 같은 감독이 만든 영화를 추천해주는 것이다.

3. 결론

OTT 서비스는 영상이나 사용자의 특성에 대한 전반적인 설명을 파악하기 어려운 상황에서는 추천 알고리즘의 질이 매우 낮아진다고거나, 이미 시간이 지나서 유행이 지나버린 콘텐츠를 추천할 가능성이 커진다는 점이다. 본 논문에서는 이러한 OTT 서비스 추천시스템의 Cold-Start 문제 원인과 해결하는 방법들에 대해 알아보았다. 시스템에서 주로 사용되는 혼합형 필터링 기법들은 장단점이 존재하고 잠재적인 문제점으로는 Cold-Start를 가지고 있다. 결론적으로 어떤 하나의 방법을 고수하기보다는 소개된 방법들을 이리저리 결합하여 만든 머신러닝 기법이나 하이브리드 추천시스템 모델이 Cold-Start 문제를 가장 잘 해결한다고 생각한다.

참고문헌

- 1) 김민정, 장성봉, 서영수 and 전수영. (2020). SAMC 알고리즘을 이용한 미디어 패널 개인화 추천 알고리즘. Journal of The Korean Data Analysis Society, 22(5), 1835-1847.
- 2) 과학기술정보통신부 제8기 블로그 기자 윤석인
- 3) 코딩 월드 뉴스(<https://www.codingworldnews.com>)
- 4) <권하영 기자>kwonhy@ddaily.co.kr
- 5) 얹의 공간 영훈YounghunJo 2021. 2. 25
- 6) A Study on Movies Recommendation System of Hybrid Filtering-Based
- 7) 김다산, 박제원, 신용태and 최재현. (2016). OTT 서비스를 위한 영상 추천시스템. 한국IT정책경영학회 논문지,8(1), 87-93.