2018-1학기 캡스톤 디자인 칠색조 최종발표

KT영화 데이터 분석을 통한 올레tv 서비스 차별화 방안 제안

01. 연구 배경

배 경

- 빅데이터 시대에 따라 미디어 산업에서도 이용자 개인의 취향이나 성향에 맞는 콘텐츠를 추천해주는 **'개인화 추천 서비스'**가 주목 받고 있다.
- VOD 이용자는 본인이 원하는 콘텐츠를 탐색하거나 검색해서 이용하는 것이 쉽지 않다.
- 이에 따라 IPTV는 **고객의 선호**(preference)를 관찰하고 분석하여 더 나은 **고객맞춤 추천 시스템**을 구축할 필요가 있다.

이 슈

- 주요 경쟁사인 U+tv와 SK Btv와 제품, 가격 면에서 거의 유사하여 KT만의 서비스 차별화가 요구된다.

연 구 목 적

- 장르, 국적, 시청 등급 세 가지 독립 변수의 교호 작용이 구매율에 미치는 영향을 파악한다.
- 시놉시스 속 특정 단어 포함 여부와 장르의 연관성이 영화 매출에 어떠한 영향을 미치는지 파악한다.
- 이를 기반으로 올레tv만의 차별적 고객맞춤 서비스를 구축하여 올레tv의 **market share**와 **brand loyalty**를 높이고자 한다.

02. 의사결정나무

연구 개요 및 목적

장르, 국적, 시청 등급 세 가지 독립 변수의 교호 작용이 영화 이용 횟수에 미치는 영향을 살펴보고 이를 의사결정나무로 시각화한다. 이를 통해 매출 극대화를 위한 방안을 제시한다.

연구 수행 과정

1> 영화 장르를 세 가지로 분류한다.

데이터 분석의 용이함을 위해 장르를 Active, Nonactive, Comedy 세 가지로 분류한다.
Active: 공포/스릴러, 서부, 무협, SF, 전쟁, 판타지, 범죄/추리
Nonactive: 다큐멘터리, 로맨스, 드라마, 뮤지컬, 애니메이션
Comedy: 코미디

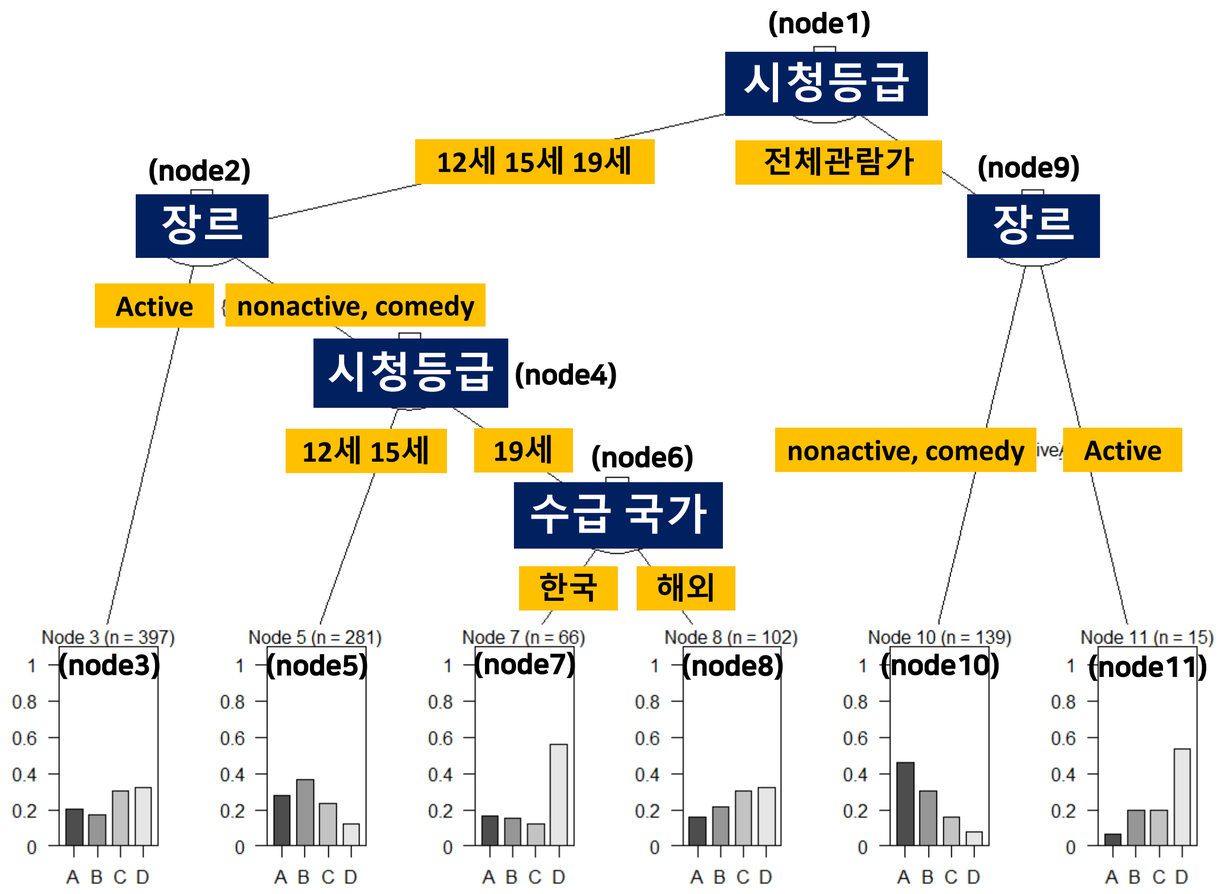
2> 종속변수인 이용횟수를 4가지 범주로 나눈다.

해석의 용이함을 위해 데이터를 이용횟수 rank에 따라 4가지로 분류한다.
Group A: 이용횟수 201회 이하(25%)
Group B: 이용횟수 201회 초과 1148회 이하(25%)
Group C: 이용횟수 1148회 초과 3931회 이하(25%)
Group D: 이용횟수 3931회 초과(25%)

3> 결과물을 **Decision Tree**로 시각화한다.

분석한 결과물을 의사결정나무(Decision Tree)로 시각화하여 장르, 국적, 시청 등급 세 가지 독립 변수의 교호 작용이 영화 이용 횟수에 어떤 영향을 미치는지 살펴본다.

연구 수행 과정



종속변수: 이용횟수 -> **인기의 척도로 해석**

10-fold validation으로 확인한 accuracy : 37.2%

의사결정나무 해석

Node11 : 전체관람가의 **Active**장르의 영화의 이용횟수가 높음
몇 안되는 표본(n=15)임에도 불구하고 이용횟수가 높은 이유는 소장용 어린이 선호영상(ex.파워레인저)등을 한 번 구매 후 지속적으로 반복 이용하기 때문으로 보인다.

Node10 : 전체관람가의 비 **Active**장르의 영화의 이용횟수가 높음
반면, 대부분의 사람이 지루하다고 생각하는 다큐멘터리와 서정적인 휴먼/가족 영화들은 이용횟수가 낮게 집계되었다.

Node3,5 : **Active**장르가 비**Active**장르에 비해 이용횟수가 높음
액션, 모험, 스릴러, 공포 등 상대적으로 Active한 장르의 영화가 그렇지 않은 장르의 영화보다 상대적으로 선호됨을 알 수 있었다.

Node7,8 : 관람등급 **19**세 이상의 영화들은 국산영화 선호
19세 이상의 영화 중 한국 영화의 선호도가 높은 이유로는 몇몇 자극적인 한국 성인영화가 견인차 역할을 하기 때문으로 보인다. 이는 뒤에서 나올 시놉시스 분석 결과와 같은 맥락을 보인다.

의사결정나무 : 제안 및 기대효과

이용 금액 및 플랜 설정

- 어린이들은 똑같은 작품을 여러 번 반복 시청하는 경향이 크다는 점을 감안하여 Active 장르인 전체 관람가 영화의 경우 가격 설정 시 **이용 금액을 인상하거나 소장용 영화를 따로 런칭**하여 매출을 극대화한다.

개인별 영화 추천 서비스

- 전체 관람가 작품을 주로 시청하는 고객에게는 Active 장르의 전체 관람가 작품을 추천하고 19세 이상 작품을 즐겨 보는 고객에게는 19세 이상 작품들 중에서도 국내 작품들을 위주로 추천하는 등의 개인별 영화 추천 서비스를 도입하여 타 경쟁사와의 차별화를 도모한다.

2018-1학기 캡스톤 디자인 칠색조 최종발표

KT영화 데이터 분석을 통한 올레tv 서비스 차별화 방안 제안

03. 시놉시스 분석

연구 개요 및 목적

시놉시스 분석을 통해 특정단어들과 시놉시스와의 관계를 파악하고 소비자들의 구매를 유도하는 시놉시스 구성을 제안한다.

연구수행 과정

1> 시놉시스의 단어 분석

KT자료와 외부자료인 CGV자료를 통해 시놉시스에서 많이 등장하는 **단어들을 추출한다**.

2> 특정단어의 사용유무에 따른 관객수 파악

시놉시스에서 특정단어를 사용했을 때와 사용하지 않았을 때 누적관객수와 이용횟수의 차이를 파악한다.
장르에 따라 특정단어 사용여부에 따른 누적 관객수와 이용횟수의 차이를 파악한다.

3> 시놉시스의 특징 분석

특정단어 사용에 따른 누적관객수와 이용횟수의 차이가 큰 장르를 파악하고 **특정단어가 들어가는 시놉시스들을 비교하여 분석**한다.

'위험' 단어 사용에 따른 분석 : 멜로/로맨스

CGV 데이터 분석 : '위험'

전체 데이터	
위험	누적관객수
단어 없음	346,996
단어 있음	415,021

멜로/로맨스	
위험	누적관객수
단어 없음	209,850
단어 있음	908,215

전체 데이터	
위험	이용횟수
단어 없음	8680
단어 있음	7218

로맨스	
위험	이용횟수
단어 없음	8210
단어 있음	20556

KT

- 선생님이 가르쳐 주신 사랑, 장모님께 보답하겠습니다! 내추럴 본 욕정남의 격정적인 파격 멜로! 열아홉 살의 진구는 미모의 과외 선생님과 통제 불능 로맨스를 즐긴다. 어느덧 성인이 된 진구는 데릴사위가 되어 아이, **장모와 기묘하고도 위험한 동거를 시작한다.**

CGV

1959년 모스크바. 평생 조국을 위해 살아온 소련 정부 관료 '사샤'. 부모의 죽음을 목격하고 소련 체제를 반대하며 스파이로 성장한 '카티야'. 남몰래 사랑하고 있던 '카티야'를 친구인 '사샤'에게 의도적으로 접근시킨 스파이 '미샤'. 하지만 '사샤'와 '카티야'는 진실한 사랑에 빠지게 된다. 결국 사랑하는 연인 '카티야'와 함께하기 위해 조국을 버리고 미국으로의 도피 망명을 준비하는 과정에서 어느 날 '카티야'가 갑자기 사라지는데… 친구도 연인도 믿을 수 없었던 **잔혹한 시대에 시작된 위험한** 사랑 그리고 사라진 연인… 영원한 기다림이 시작된다. 그리고 1992년 뉴욕, 그 사랑의 진실이 밝혀진다…

'로맨스' 장르에서 '위험'이란 단어가 등장할 때 이용횟수가 훨씬 높게 나왔으며, 주로 '치명적, 금지된, 뜨거운'과 같은 자극적인 단어와 함께 등장하였다. 한편 CGV 데이터보다 KT 데이터에 자극적인 단어가 더 많이 포함되어 있었는데, 이는 영화관에서 상영되지 못하는 자극적인 로맨스 영화들이 KT Olleh TV에서 상영을 하기 때문인 것으로 보인다.

'위험' 단어 사용에 따른 분석 : 액션

CGV 데이터 분석 : '위험'

전체 데이터	
위험	누적관객수
단어 없음	346,996
단어 있음	415,021

액션	
위험	누적관객수
단어 없음	1,391,901
단어 있음	379,638

KT 데이터 분석 : '위험'

전체 데이터	
위험	이용횟수
단어 없음	8680
단어 있음	7218

액션	
위험	이용횟수
단어 없음	18602
단어 있음	1400

KT

멈출 수 없는 살인 본능, 타고난 킬러로 불린 살인청부업자 빅터! **그와의 위험한 인터뷰가** 시작된다! 세르게이 밑에서 살인 청부업자가 되기 위한 교육을 받으며 킬러로 키워진 빅터는 우연히 만난 베네스다와 사랑에 빠지고 그녀가 임신한 사실을 알고 혼란스러워 한다.

CGV

대한민국이 위험하다! 역대 최대 규모의 강진에 이어 원자력 폭발 사고까지 예고 없이 찾아온 초유의 재난 앞에 한반도는 일대 혼란에 휩싸이고 믿고 있던 컨트롤 타워마저 사정없이 흔들린다 방사능 유출의 공포는 점차 극에 달하고 최악의 사태를 유발할 2차 폭발의 위험을 막기 위해 발전소 직원인 '재혁'과 그의 동료들은 목숨 건 사투를 시작하는데…!

'액션' 장르에서 단어 '위험'이 등장했을 때 이용횟수에 긍정적인 영향을 미칠 것으로 기대되었지만 실제로는 단어가 등장하지 않았을 경우 이용횟수가 훨씬 증가하였다. 이는 '위험'이란 단어가 액션 장르에서는 다소 진부하게 느껴져 시놉시스에서 궁금증을 유발하지 못하기 때문인 것으로 보인다.

'결혼' 단어 사용에 따른 분석 : 멜로/로맨스

CGV 데이터 분석 : '결혼'

전체 데이터	
위험	누적관객수
단어 없음	362,851
단어 있음	146,189

멜로/로맨스	
위험	누적관객수
단어 없음	263,818
단어 있음	105,371

KT 데이터 분석 : '결혼'

전체 데이터	
위험	이용횟수
단어 없음	8594
단어 있음	9347

로맨스	
위험	이용횟수
단어 없음	9848
단어 있음	2603

KT

신이치와 고토미는 **결혼한 지 10년째인** 부부. 두 사람은 줄곧 불임클리닉에 다니며 노력해 왔지만 좀처럼 아이가 생기지 않는다. 회사 후배 히로유키의 걱정에도 곧 생길 거라며 태평하게 구는 신이치. 그는 사실 2년 전까지 애인이 있었다.

CGV

40cm 아래, 모든 것이 완벽한 이상형이 나타났다?! 능력과 미모를 겸비한 성공한 변호사 디안. **최근 불행한 결혼생활을 정리한** 그녀는 잃어버린 핸드폰을 찾아 준 알렉상드르와의 설레는 만남을 새롭게 시작한다. …(생략)…

'멜로/로맨스' 장르에서 '결혼'이라는 단어가 나올 경우 현실적이고 우울한 결혼 생활에 대해 이야기하거나 행복한 결혼 생활을 엔딩으로 등장하는 경우가 많아 시놉시스에 나오기 어려운 것으로 보았다.

2018-1학기 캡스톤 디자인 칠색조 최종발표

KT영화 데이터 분석을 통한 올레tv 서비스 차별화 방안 제안

‘결혼’ 단어 사용에 따른 분석 : 코미디

CGV 데이터 분석 : ‘결혼’

전체 데이터	
위험	누적관객수
단어 없음	362,851
단어 있음	146,189

코미디	
위험	누적관객수
단어 없음	276,468
단어 있음	1,126,735

KT 데이터 분석 : ‘결혼’

전체 데이터	
위험	이용횟수
단어 없음	8594
단어 있음	9347

코미디	
위험	이용횟수
단어 없음	4066
단어 있음	14237

KT

천재 교수인 필립 브레이너드는 건망증이 보통 심한 게 아니다. **세 번째 결혼식 날, 이번에는 잊지 않고 결혼식에 참석하겠다던** 브레이너드 교수는 오랫동안 진행해온 실험이 성공하기에 이르자 역시 결혼식은 뒷전으로 하고 발명품에 정신을 빼앗긴다.

CGV

‘복’ 터진 줄 알고 시작한 견우의 ‘속’ 터지는 신혼 수난기! …(생략)…‘그녀’! “설마 결혼? 제가 잘못 들은 걸까요?” 꿈인지 생시인지 들어온 복을 얼른 움켜쥐는 반도의 잉여, ‘견우’! 그러나, 밤은 더 살벌하고, 낮은 더 엽기적인데…**예측불가! 새로운 ‘그녀’와 ‘견우’의 상상 못한 엽기적인 결혼!** ‘견우’의 인생수난 여기서 끝날 수 있을까?

➡ 코미디’ 장르에서는 결혼의 힘든 점을 해학적으로 풀어내 인기가 많은 것으로 보인다.

‘엄마’ 단어 사용에 따른 분석 : 드라마

CGV 데이터 분석 : ‘엄마’

전체 데이터	
위험	누적관객수
단어 없음	371,136
단어 있음	124,884

드라마	
위험	누적관객수
단어 없음	200,435
단어 있음	8,3092

KT 데이터 분석 : ‘엄마’

전체 데이터	
위험	이용횟수
단어 없음	8,517
단어 있음	11,459

코미디	
위험	이용횟수
단어 없음	3,087
단어 있음	6,644

KT

2014 두근두근 캐스팅! 철없는 아빠 대수와 **당찬 엄마** 미라. 그리고 선천성 조로증으로 16살 나이에 80살의 신체를 가진 아들 아름이. 씩씩하게 살아가는 예쁜 가족에게 다가오는 죽음의 그림자. 이들이 전하는 아주 특별한 감동.

CGV

과거의 영광을 잊지 못한 채 유명 작가를 꿈꾸는 사설탐정 ‘료타’는 태풍이 휘몰아친 날, 헤어졌던 가족과 함께 예기치 못한 하룻밤을 보내게 되는데… 아직 철들지 않은 대기만성형 아빠 ‘료타’ **조금 더 나은 인생을 바라는 엄마** ‘쿄코’ 빠르게 세상을 배워가는 아들 ‘싱고’ 그리고 가족 모두와 행복하고 싶은 할머니 ‘요시코’ 어디서부터 꼬여버렸는지 알 수 없는 ‘료타’의 인생은 태풍이 지나가고 새로운 오늘을 맞이할 수 있을까?

➡ CGV데이터에서 ‘드라마’ 장르에 ‘엄마’가 포함될 경우 누적관개수가 적었던 반면 KT데이터에서는 이용횟수가 더 늘어난 것을 볼 수 있다. 이는 화려하고 웅장한 액션 영화에 비해 드라마 장르의 영화는 영화관에서 잘 보지 않는 경향이 있기 때문이다.

시놉시스 분석 : 제안 및 기대효과

1

- 시놉시스 분석을 통해 같은 단어라도 어떤 장르에서는 높은 누적관객수를 보였고 그 반대의 경우도 존재했다. 예를 들면 '위험'이나 '의문'은 '액션' 장르의 영화 시놉시스에 사용되었을 때 이미 '액션'이라는 장르가 두 단어를 함의하고 있기 때문에 진부하게 느껴지는 경향이 있다. 특정 장르를 생각할 때 연관되어 떠오르지 않는 단어를 사용해서 사용자에게 궁금증을 유발하고 신선함을 느낄 수 있는 시놉시스를 만든다면 영화 이용횟수에 긍정적인 영향을 미칠 것으로 기대된다.

2

- 시놉시스에 포함되어 있을 때 높은 누적관객수를 보이는 단어를 이용해 키워드로 만든다. 올레티비 시청자가 특정 장르의 영화를 검색할 때 그 장르에서 높은 관객수를 보인 키워드를 노출시켜 시청자가 원하는 유형의 영화를 더 효과적으로 추천할 수 있다. 예를 들면 코미디 장르의 영화를 찾고있는 시청자에게는 '성공'이나 '결혼'이라는 키워드를 띄워서 시청자가 수많은 영화 리스트에서 짧은 시간 안에 보고 싶은 영화를 결정할 수 있게 도와줄 수 있다. 이러한 키워드 추천 시스템을 통해 KT Olleh TV는 고객의 만족도를 높일 수 있을 것으로 기대된다.

04. 연구 한계점

- 3가지 변수로 만든 의사결정나무 모형으로 영화 이용 패턴을 모두 추론하기 여렵다는 점
- 시놉시스 단어만 중점적으로 분석하여 시놉시스 문장 고유의 특성 하나하나를 간과하였다는 점
- 시놉시스 단어의 포함유무만 고려하여 단어 출현 횟수에 따른 효과를 고려하지 못하였다는 점
- KT는 한 해의 데이터인데 반에, CGV 실제 시놉시스는 여러 해의 데이터이기 때문에, 동일한 시간적 배열의 데이터를 비교한 것이 아니라는 점
- 시놉시스 단어들 간의 조합을 고려한 분석까지는 하지 못하였다는 점