

대 흥행 드라마의 조건

소개

1.

시청 메커니즘

분석부족으로

방송사

수익성 저하

2.

경쟁력을 위한

시청 메커니즘

분석이

매우 중요

3.

각 요인들이

상호작용하며

시청률에

영향을 미침

4.

시청률을

결정하는

변수들의

독립 및 결합

효과를 분석

시청률 결정 요소

1. 장르의 친숙성

친숙함이 주는 감정적 편안함, 의사결정 상의 편리함, 인지적 구조, 사회적 상호측면

2. 주연배우의 지명도

주연배우의 연기력으로 드라마를 평가, 의사결정 상의 편리함

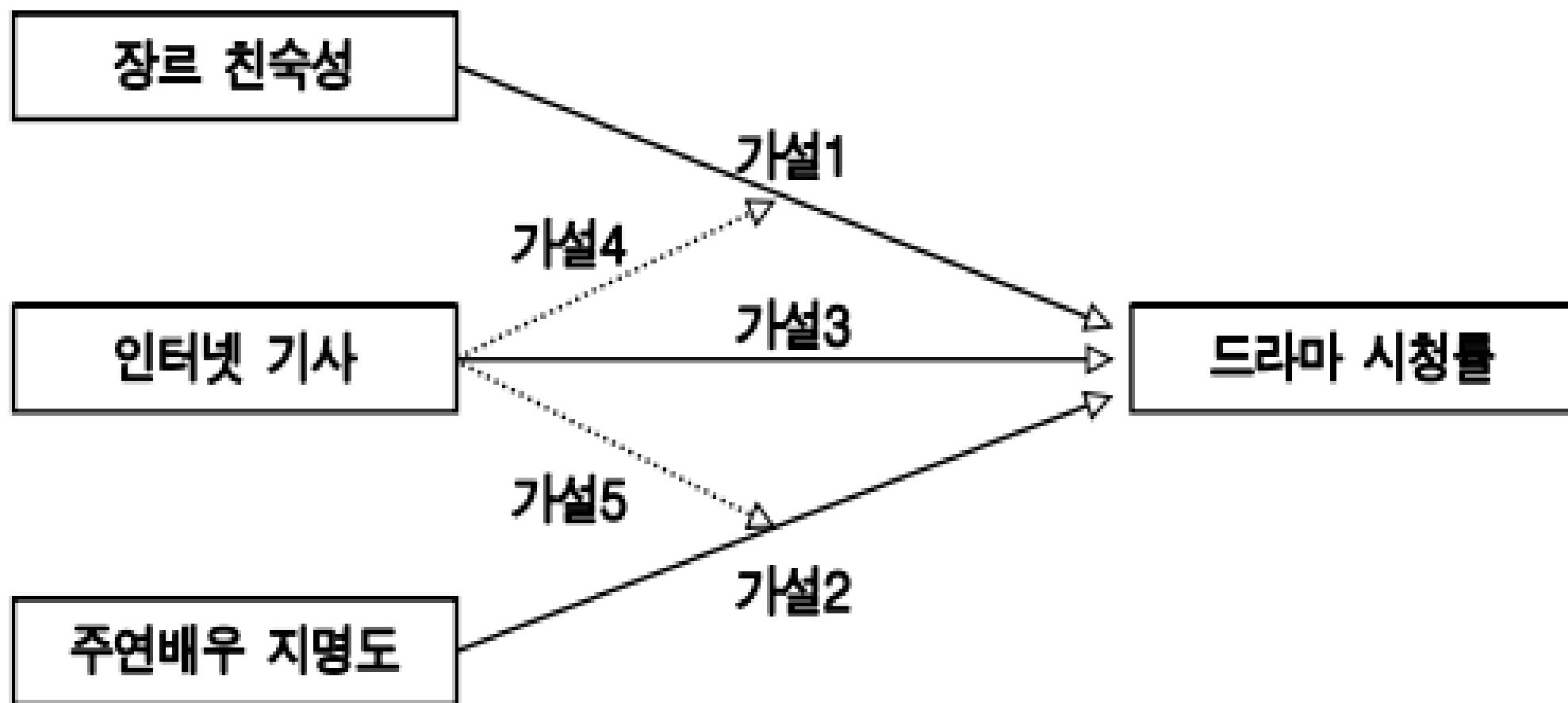
3. 인터넷 기사

전문 비평가 및 일반인들의 평가로 시청여부를 결정함

가설

1. 드라마의 장르 친숙성은 드라마 시청률에 영향을 끼칠 것이다.
2. 주연배우의 지명도는 드라마 시청률에 영향을 끼칠 것이다.
3. 드라마 인터넷 기사 수는 드라마 시청률에 영향을 끼칠 것이다.
4. 드라마 인터넷 기사 수가 많으면 장르 친숙성은 시청률에 더 큰 영향을 끼칠 것이다.
5. 드라마 인터넷 기사 수가 많으면 배우 지명도는 시청률에 더 큰 영향을 끼칠 것이다.

가설



<그림 1> 드라마 시청률에 영향을 미치는 요인에 관한 연구 모형

연구 방법

- 연구 표본

2000년 1월부터 2008년 6월까지 방영된 것

- 종속 변수

Nielsen, TNS 등에서 발표한 시청률 최고치의 평균

- 독립 변수

장르의 친숙성-직전 2개년동안 상영한 드라마들의 장르 비율

주연배우 지명도-대상 드라마 출연 이전 10년동안 시청률 30% 이상 기록한 드라마에 주연을 맡은 회수

인터넷 기사 – 다음, 네이버 – 각 연도 드라마 기사 수 * 각 연도별 인터넷 이용률 역수(조정계수)

<표 1> 각 연도별 인터넷 이용자, 이용률, 그리고 조정 계수

	2000	2001	2002	2003	2004	2005	2006	2007	2008
이용자 수(만명)	1904	2438	2627	2922	3158	3301	3412	3482	3536
이용률(%)	44.7	56.6	59.4	65.5	70.2	72.8	74.8	76.3	77.1
조정 계수	2.23	1.76	1.68	1.52	1.42	1.37	1.33	1.31	1.29

연구 방법

- 통제변수

드라마 유형(방송의 요일)-주말극, 주중극, 일일극으로 구분 1과 0 값 부여

드라마 방송 채널-MBC드라마 시청률 높음, MBC 1, SBS,KBS 0

외주제작여부-외주제작 1, 자체제작 0

방송 기간 통제-드라마 시작부터 끝까지 걸린 날짜의 수

가설 검정

<표 2> 기초 통계 및 상관관계^a

변수	평균	표준 편차	1	2	3	4	5	6	7	8	9	10	11
1. MBC	.37	.48											
2. 일일극	.13	.34	.04										
3. 주말극	.29	.45	-.11	-.24									
4. 주중극	.53	.50	.08	-.40	-.67								
5. 방송 기간(일)	134.73	107.76	-.09	.33	.25	-.45							
6. 외주 제작	.55	.49	.25	-.05	-.04	.06	-.16						
7. 작가 지명도	1.79	2.40	-.01	-.05	.16	-.06	.12	-.16					
8. 연출가 지명도	1.03	1.24	-.02	-.03	.09	.02	.21	-.02	.39				
9. 장르 친숙성	32.00	16.33	-.14	.01	.11	-.08	-.08	-.10	.18	.10			
10. 주연배우 지명도	9.35	7.20	-.06	.27	.15	-.26	.46	-.08	.28	.34	.07		
11. 인터넷 기사	651.9	893.9	.01	-.17	-.11	.20	.03	.17	.18	.23	.01	-.00	
12. 시청률(%)	24.16	12.59	-.10	.09	.04	-.01	.034	-.09	.44	.58	.25	.51	.30

N=225

^a 상관 계수의 절대 값이 0.11 이상은 $p < 0.05$ 임.

가설 검정

Hierarchical regression analysis

<표 3> 시청률에 대한 회귀분석 결과^a

변 수	모형1	모형2	모형3	모형4	모형5
1단계: 통제변수					
MBC	-0.07	-0.04	-0.06	-0.05	-0.07
일일극	0.15	0.11	0.13	0.11	0.13
주말극	0.09	0.07	0.08	0.05	0.06
주중극	0.24*	0.21	0.17	0.17	0.13
방영 기간	0.24***	0.17***	0.09	0.15	0.07
외주 제작	-0.02	-0.4	-0.03	-0.03	-0.03
작가 지명도	0.25***	0.15***	0.13*	0.15***	0.13*
연출가 지명도	0.41**	0.31***	0.25***	0.30***	0.24***
2단계: 독립변수					
장르 친숙성		0.18***	0.12*	0.19***	0.13*
배우 지명도		0.28***	0.35***	0.21***	0.27***
인터넷 기사		0.19***	0.17**	0.08	0.06
3단계: 상호작용					
장르 친숙성 X 인터넷 기사			0.22***		0.20***
배우 지명도 X 인터넷 기사				0.18*	0.19*
상수	9.99	10.01	5.65	4.75	7.50
ΔR^2	0.45	0.11	0.04	0.01	0.05
ΔF	19.48***	15.82***	14.94***	4.57*	9.95***
수정 후 R^2	0.43	0.54	0.53	0.55	0.54
전체 모형의 F	19.48***	21.87***	17.10***	20.81***	16.50***

* 가 많을수록 의미있는 값

- 부정적인 영향

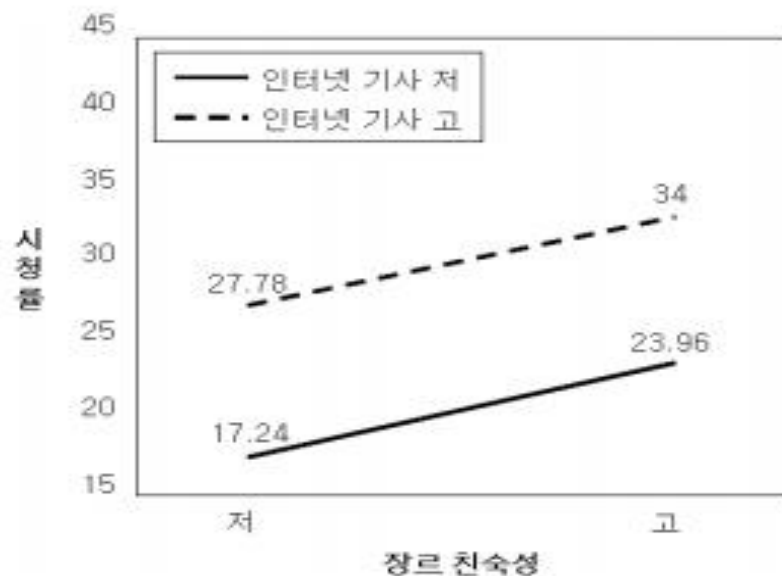
+ 긍정적인 영향

가설 1,2,3 검증

^a N=225. 입력치(Entries)는 표준화된 베타(Beta) 값

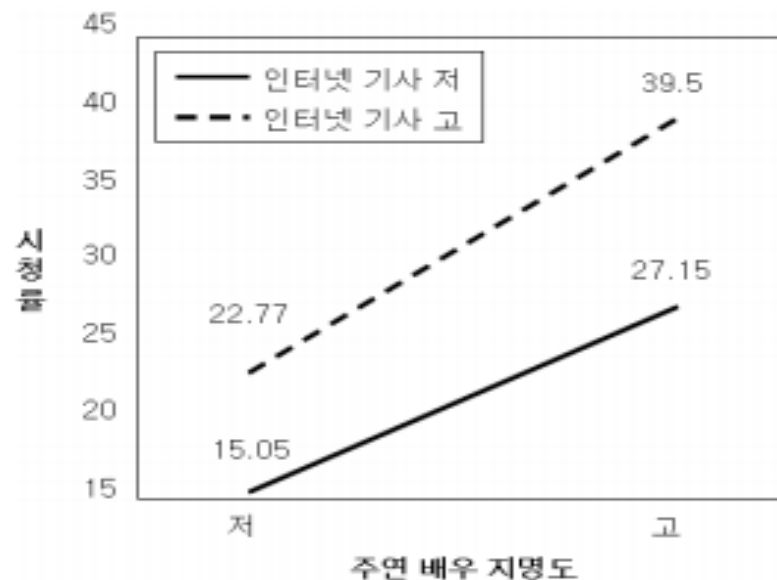
* $p < .05$, ** $p < .01$, *** $p < .001$

가설 검증



드라마 인터넷 기사 수와 장르 친숙성 고저(高低)는 평균을 기준으로 하여 분류했으며, 그래프의 값은 시청률 수준(단위, %)을 표시함

<그림 2> 드라마 인터넷 기사 수와 장르 친숙성 상호작용 항이 시청률에 미치는 영향



드라마 인터넷 기사 수와 주연배우 지명도 고저(高低)는 평균을 기준으로 하여 분류했으며, 그래프의 값은 시청률 수준(단위, %)을 표시함.

<그림 3> 드라마 인터넷 기사 수와 주연배우 지명도 상호작용 항이 시청률에 미치는 영향

가설 검정

<표 4> 대 흥행 드라마에 대한 로지스틱 회귀분석 결과^a

변 수	모형1	모형2	모형3	모형4
(통제변수)				
MBC	-1.55(0.78*)	-1.43(0.79)	-1.59(0.83)	-1.60(0.87)
일일극	38.13(8503)	41.48(8503)	27.27(9038)	27.33(9214)
주말극	36.95(8503)	40.28(8503)	25.91(9038)	25.95(9214)
주중극	36.77(8503)	40.03(8503)	25.44(9038)	25.49(9214)
방영 기간	-0.00(0.00)	-0.00(0.00)	-0.00(0.01)	-0.00(0.01)
외주 제작	-1.61(0.84*)	-1.73(0.87*)	-2.05(0.95*)	-2.05(0.96*)
작가 지명도	0.15(0.15)	0.13(0.15)	0.17(0.15)	0.17(0.15)
연출가 지명도	0.96(0.28***)	0.97(0.29***)	1.04(0.30***)	1.04(0.30***)
(독립변수)				
장르 친숙성	0.03(0.25)	0.01(0.04)	0.03(0.03)	0.03(0.04)
배우 지명도	0.24(0.07***)	0.24(0.07***)	0.16(0.08*)	0.16(0.08)
인터넷 기사	0.00(0.00***)	0.00(0.00+)	0.00(0.00)	0.00(0.00)
(상호작용 변수)				
장르 친숙성 X 인터넷 기사		0.00(0.00)		0.00(0.00)
배우 지명도 X 인터넷 기사			0.00(0.00*)	0.00(0.00+)
상수	-43.10(8503)	-45.77(8503)	-31.24(9038)	-31.29(9214)
x2(df)	126.22(11)	126.65(12)	130.25(12)	130.25(13)
-2 log likelihood	71.05	70.62	67.02	67.02

^a N=149. 제시된 수치는 회귀계수(Beta)이며 괄호 안은 표준오차.

+p < .10, * p < .05, ** p < .01, *** p < .001

시청률 30%를 기준으로 비 흥행 드라마와 대 흥행 드라마로 구분

연구 결과

- 시청자들은 새로운 장르의 드라마보다는 기존에 시청해 왔던 장르의 드라마에 대하여 더 큰 즐거움과 애착을 가진다.
- 의사결정의 편리성은 드라마 시청 여부를 결정하는 데 있어서 중요한 요인이다.
- 드라마의 주된 시청자는 주부, 주부들의 일상적인 생활과 밀접하게 연계되는 장르의 드라마가 시청률이 높다.
- 지명도 높은 배우의 출연을 알고 인터넷 기사를 많이 접할수록 시청할 확률이 커진다.

개인 단위 시청률 빅데이터를 이용한
시청률 분석방법 연구

연구 목적

1.

효율적인 광고비

집행을 위한

정확한 시청률

예측

2.

다채널 시대의 복

잡한 경쟁 환경을

고려한 시청률

예측 분석모델

제안

3.

비선형적 관계

까지 고려한

시청률 예측

계량모델 개발

연구 방법

- 빅 데이터의 축적 및 처리
 - 피플미터, 모바일, 온라인
- 대안적 예측모형의 타당성 검증
 - 분석단위의 차이와 선형성 가정을 충족하지 못하는 경우 때문에 대안적 예측모형들이 필요함
- 정성적 정보 활용
 - 기존 모델에는 정성적 정보를 적용하기 힘들
 - Linked Data System을 구축해 정성적 판단의 기본 자료로 사용가능
- 시청률 예측은 드라마 장르로 한정, 최종 예측은 방영 예정인 드라마의 초회분 시청률을 대상

데이터

닐슨의 3년 치 개인 단위 시청률 데이터를 사용 데이터는 HDFS에 저장

	Schooling	패널의 학력수준 정보	College graduate	String
	Occupation	패널의 직업 정보	Professional / Management	String
	Marriage	패널의 결혼 여부	Married	String
	Age	패널의 실제 나이	24 / From 70 Up	String
	Weight	패널의 가중치	10448.4	Float
시청	ID	시청정보가 속한 가구번호	1401077	String
	File Date	시청한 날짜	2013-07-01	String
	Channel	채널코드	SBS ESPN	String
	TV ID	TV 번호	1	Int
	Platform	해당하는 채널에 시청되는 플랫폼	Skylife	String
	Name	시청한 멤버	aa	String
	Time Start	시청 시작 시간	140800	String
	Time End	시청 종료 시간	182559	String
	Time Total	총 시청 시간	5 hour 12 min 59 sec	String
프로그램	Local Code	지역코드	whole country	String
	Channel	채널코드	MBC	String
	End Time	프로그램 종료시간	2010-07-01 20:17:46	Int
	Start Time	프로그램 시작시간	2010-07-01 20:53:11	Int
	Class Code	시급	C	String
	Program Description	프로그램명	일일연속극(황금물고기)	String
	Class Code	광고비	12630	String
	Type Top	대분류	Drama & Movie	String
	Type Middle	중분류	Drama	String
	Type Bottom	소분류	Daily soap opera	String
	Plan Time	프로그램 계획시간	20:15:00	Int

데이터	속성	설명	예제	타입
가구	ID	가구번호	1401077	String
	File Date	날짜	20130701	String
	Area Zone	패널가구의 구별 정보	Jongro-gu	String
	Size	패널가구의 주택평수 정보	20~49 / From 100 Up	String
	Building Form	패널가구의 주택형태 정보	Detached house	String
	Ownership	패널가구의 저가/전세 정보	Own	String
	Income	패널가구의 가구소득 정보	120~199 / From 500 Up	String
	Member Number	패널가구의 가족 수 정보	1~2 / From 5 Up	String
	TV Number	패널가구의 TV대수 정보	1 / From 2 Up	String
	Child under 13	패널가구의 13세 이하 아이 유무	Exist	String
	Cable Form	패널가구의 케이블 상품 정보	Popular(pay)	String
	Sky Life HD	패널가구의 스카이라이프 HD 여부	Non-exist	String
	IPTV	패널가구의 IPTV 여부	Exist	String
	Cable TV	패널가구의 케이블 가입 여부	Narrowband Cable	String
	Satellite	패널가구의 위성수신 여부	Non-exist	String
	Sky Life Reception	패널가구의 스카이라이프 수신여부	Exist	String
	Sky Life Form	패널가구의 스카이라이프 상품 정보	Fundamental(pay)	String
	IP License	패널가구의 IPTV 사업자 정보	KT QOOK LIVE	String
	Market	해당 가구가 속해 있는 마켓	Seoul	String
	Weight	해당 가구의 가중치 정보	8385.8	Float
패널	ID	패널이 속한 가구번호	1401077	String
	File Date	날짜	20130701	String
	Name	구성원 구별자	aa	String
	Sex	패널의 성별 정보	male	String
	Age Scope	패널의 연령대 정보(5세 단위)	10~14 / From 65 Up	String

데이터 전처리

<표 2> 예측모형의 입력 데이터

시청 자ID	시청 일자	드라 마명	채널 코드	진행 률	러닝 타임 (분)	3월~ 5월 방영	6월~ 8월 방영	9월~ 11월 방영	월화 드라 마	수목 드라 마	주말 드라 마	22시 이후 방영	21~2 2시 방영	19~2 1시 방영	KBS1 시청 시간 (분)	KBS2 시청 시간 (분)	MBC 시청 시간 (분)	SBS 시청 시간 (분)
1234 917a a	2010 0701	황금 물고 기	MBC	0.01	36	0	1	0	1	1	0	0	0	1	0	2	25	6
1234 917a a	2010 0701	바람 불어 좋은 날	KBS1	0.01	36	0	1	0	1	1	0	0	0	1	0	1	29	3
1234 917a a	2010 0701	로드 넘버 원	MBC	0.05	71	0	1	0	0	1	0	1	0	0	23	0	0	0
1234 917a a	2010 0701	제왕 왕 김탁 구	KBS2	0.04	73	0	1	0	0	1	0	0	0	1	10	20	3	0
1234 917a a	2010 0701	세자 매	SBS	0.01	37	0	1	0	1	1	0	0	0	1	10	20	3	0
1234 917a a	2010 0701	나쁜 남자	SBS	0.09	72	0	1	0	0	1	0	1	0	0	23	0	0	0

1. HDFS에 있는 프로그램 테이블과 시청 테이블을 조인
2. 프로그램 테이블의 시작시간과 종료시간을 사용하여 각 Feature 값 계산
3. Hive를 사용하여 데이터의 관계형 테이블을 조인
 - Hive에서 제공하는 질의문을 작성하면, HDFS은 질의문을 자동으로 MapReduce로 변경하여 실행, Map에서 예측변인 값 계산, Reduce에서 출력데이터를 병합

초회분 시청시간 예측

1. 모든 드라마 벡터들을 균등하게 나누어 Map task 실행
2. key는 시청자ID, value는 시청시간과 distance로 설정

ex) (key=141083ad, value={SBS 시청시간=12분, distance=1.62})

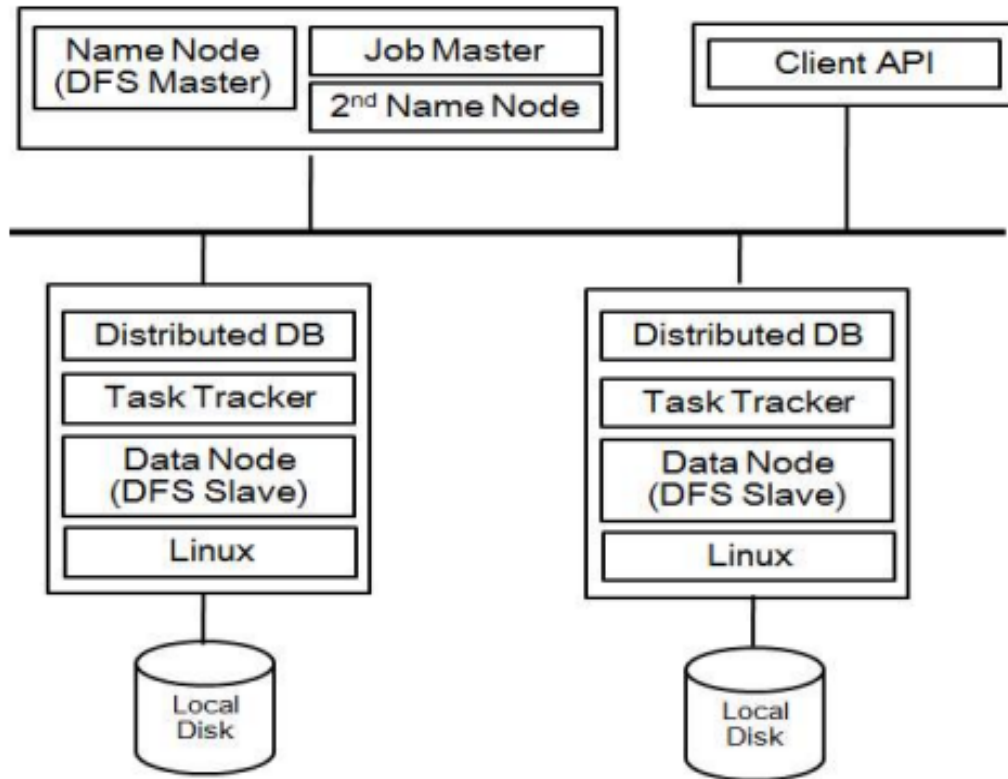
3. Reduce task에서 k-NN 사용해서 SBS 시청시간을 구함

<표 3> k-Nearest Neighbor MapReduce 알고리즘

Procedure k-Nearest Neighbor MapReduce	
Map task	Input: All points Query point p Output: k nearest neighbors (local) Emit the k closest points to p
Reduce task	Input: Key = null Values = local neighbors Query point p Output: k nearest neighbors (global) Emit the k closest points to p among all local neighbors

빅데이터 분산처리

2.2GHz hexa core 64bit 워크스테이션 서버
Hadoop 1.2
Cent OS Linux



<그림 1> Hadoop 기반의 빅데이터 처리 시스템

네임 노드 - 파일 분할하여 데이터 노드에 전송,

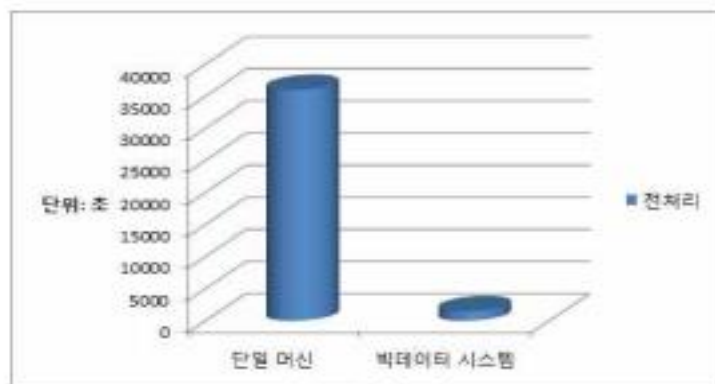
데이터 노드에 대한 메타 정보 저장

데이터 노드 - 약 200MB의 Chunk들을 수백 GB씩 저장,

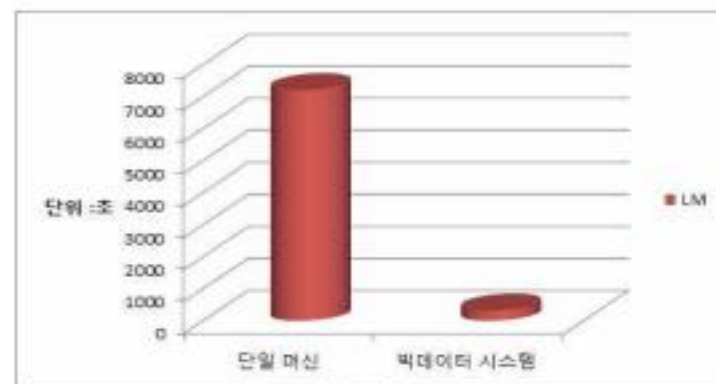
MapReduce 수행, 빠른 검색을 위해

Chunk가 많아지면 병합, 커지면 분할

빅데이터 처리 결과



(a) 전처리



(b) 선형회귀 모형



(c) k-Nearest Neighbour 알고리즘



(d) 전체 성능

<그림 2> 단일 머신과 빅데이터 기술을 이용한 데이터 처리 시간 비교

시청률 예측 모형

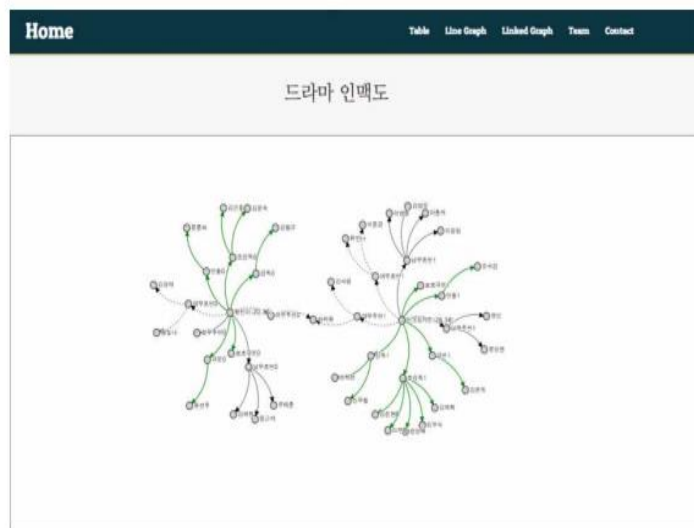
- 가장 관심도가 큰 신작 드라마의 초회분 시청률 예측
- 예측 변수
 - 계절적 요인, 방영시간대, 경쟁채널 시청환경 요인, 인구통계학적 요인이나 시장상황 배제

<표 4> 예측모형의 드라마 시청률 예측 평가

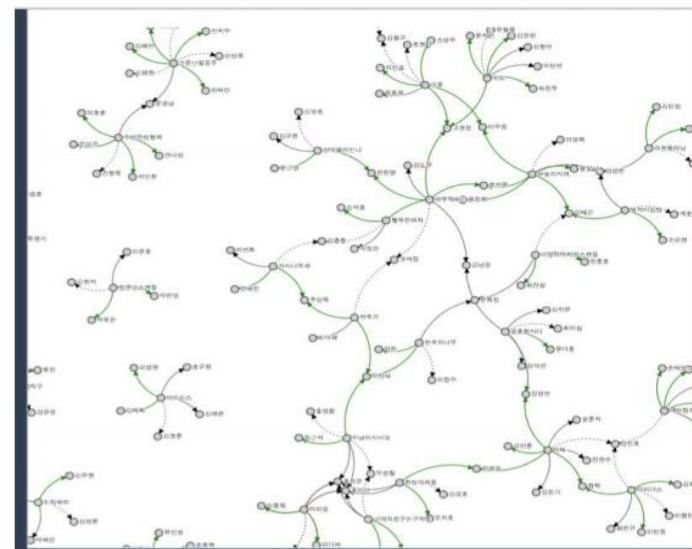
단위: 분수(minutes)						
	방영일	실제시청	K=1	K=2	K=3	평균 (K1,K2,K3)
KBS2 채널						
칼과 꽃-1회†	2013/07/03	6.86	14.73 (2.15)	14.29 (2.08)	13.86 (2.02)	14.29 (2.08)
아이리스2-1회†	2013/02/13	16.87	12.68 (0.75)	12.20 (0.72)	11.65 (0.69)	12.18 (0.72)
MBC 채널						
여왕의교실-1회	2013/06/12	7.76	9.38 (1.21)	8.28 (1.07)	7.59 (0.98)	8.42 (1.09)
남자가사랑할때-1회†	2013/04/03	8.12	12.92 (1.59)	11.99 (1.48)	11.21 (1.38)	12.04 (1.48)
SBS 채널						
황금의제국-1회	2013/07/01	8.67	11.73 (1.35)	10.37 (1.20)	9.52 (1.10)	10.54 (1.22)
야왕-1회	2013/01/14	7.42	7.66 (1.03)	5.93 (0.80)	5.25 (0.71)	6.29 (0.85)

Linked Data 구축

- 정성변수 - 연출자, 작가, 주요배우 등
- 각 요인들의 관련성을 네트워크하여 클러스터링 구축 후 평가
- 3년간 방영되었던 드라마 368편의 32384개 데이터를 SQL데이터로 구성



[주연배우 하지원 인맥 네트워크]



[시청률 15% 이상 인맥 네트워크 클러스터링]

Linked Data 구축

연출가 - 과거 제작한 드라마의 참가 가중치(주연출 1.2, 보조연출 0.8)를
제작 편수에 따른 가중치 평균을 구함

주연배우 - 과거 출연한 작품의 참가 가중치(주연배우 1, 조연배우 0.5)를
출연한 드라마의 평균 시청률에 곱하여 가중치 계산

과거 시청률	친밀도
10% 미만	0.8
10% 이상 15%미만	1
15%이상 20%미만	1.2
20%이상 25%미만	1.5
25%이상 30%미만	1.8
30%이상	2

Linked Data 구축

	연출	작가	주요인물			
인 물	김명옥	김현희	박세영	유건	이해인	박재정
시청률	24.38	16.08	0	9.81	0	30.35
참여률	1.2	1	1	0.5	0.5	1
천명도	1	1	1	1	1	2

개선 될 사항

1. 계절요인과 프로그램 경쟁상황만을 고려

-> 프로그램의 특성에 대한 변수화가 필요

2. kNN은 속도는 빠르나 정확도가 좋지 못함

-> kNN이 아닌 다른 머신러닝 알고리즘의 적용이 필요

3. 시청흐름에 대해 고려하지 못함

4. 가구 내의 다른 구성원들이 시청자에게 주는 영향을 고려하지 못함

-> 구성원 수에 따른 영향 고려가 필요

5. 표본 대표성을 위해 사후 유층화 가중치를 사용해야 함

감사합니다