

국내 TV 시청률 자료를 이용한 시청률 예측 모델 구축 및
시청률 최대화 전략 연구
TV Audience Forecast Modeling and Study for Its Strategic Use
Using Korean TV Market Data

정규환, 이재욱, 이대원, 김보현
(우편번호 : 790-784) 경상북도 포항시 남구 효자동 산 31번지
포항공과대학교 산업경영공학과
E-mail : onlyou7@postech.ac.kr
Phone : (054)279-8258

Abstract

시청률은 각 방송사의 주요 수입원인 광고비 책정의 중요한 변수이다. 따라서 프로그램 편성자(Program Planner)는 경쟁 방송사의 편성을 고려한 시청률 최대화 전략을 수립하려 하며, 시청률 예측 모델을 사용하여 그 목적을 달성할 수 있다. 나아가 시청률 예측 모델이 정확할수록 현재의 임의적인 광고비 책정 방식보다 정확한 광고비의 가격 책정이 가능해져 보다 합리적인 광고 및 미디어 시장 환경을 성립할 수 있다. 본 연구에서는 최근의 국내 TV 시청률 자료를 이용하여 국내 공중파 방송에 대한 시청률 예측 모델을 성립하고, 이를 이용한 시청률 최대화 전략을 제시한다.

1. 연구 배경

현재의 TV market의 규모는 나날이 증가하고 있으며, 케이블TV, 위성방송, DMB 등의 새로운 전송 기술의 발달로 인해 채널 수도 꾸준히 증가하고 있는 실정이다. 공익 및 사설 방송국들은 이러한 미디어 환경의 변화에 따라 미디어의 가치를 대표적으로 나타내 주는 시청률을 두고 보다 치열한 경쟁을 하고 있다.

실제로 시청률은 그 프로그램의 상업적 가치를 나타내는 중요한 수단이며, 높은 가치를 갖는 프로그램의 경우 방영 전후의 광고방송(CF)에서 각 방송사의 주요 수입원인 광고비를 높게 책정할 수 있다.

방송국의 프로그램 편성은 3~5개월을 주기로 변경되며, 이때 프로그램 편성자는 시청률을 최대화 시키는 전략을 이용하려 한다. 이러한 전략을 세우기 위해서는 과거의 시청률과, 프로그램과의 상관관계를 분석하는 작업이 선행이 되어야 하며, 또한 이를 통해 현재의 프로그램이 어떠한 시청률을 갖게 될지 또한 예측할 수 있어야 한다. 이 때, 시청률은 단순히 자사의 편성뿐 아니라 타 경쟁사의 편성에 의해서도 영향을 받게 되므로 반드시 이를 고려하여 예측 모델을 세워야 한다.[1]

본 연구에서는 국내 TV market에서 주어진 최근 2년 여간의 시청률 및 프로그램 자료를 이용하여, 앞으로의 시청률을 예측하는 모델을 제시하고, 이를 바탕으로 시청률을 최대화 할 수 있는 전략을 제시하는 것을 목적으로 한다. 시청률을 예측하는 것은 어려운 문제이고, 이는 시청자 개인의 선호도, 취미, 그리고 프로그램의 질에 따라 다르며, 또한 프로그

램의 질은 프로그램의 내용과 다양한 요소들이 복합적으로 영향을 주기 때문이다. 따라서 본 연구에서는 시청자의 행동 특성을 파악하기 위한 예비 데이터 분석, 시청률을 예측하기 위한 모델 작성, 모델 분석 및 각 모델간의 비교를 통해 프로그램 편성자가 시청률을 최대화 시킬 수 있는 전략 성립에 도움이 되는 방법을 제시한다.

2장에서는 사용할 데이터를 원하는 형식으로 변환하는 과정을 설명하였고, 3장에서는 사전 분석을 통한 전반적인 정보 도출에 대해 언급하였다. 4장에서 모델을 만들고 각 모델을 비교 분석한 후, 적절한 시청률 전략을 제시하였으며, 5장에서 결론을 도출하였다.

2. 데이터 변환

분석에 사용한 데이터는 세계 시장 조사 기관인 AGB닐슨의 국내 시청률 조사 자료이다[2]. 본 데이터는 2003년 5월부터 2005년 6월까지 2년 2개월 치의 국내 3개 공중파 방송사 4개 채널(KBS1, KBS2, MBC, SBS)에 대한 정보를 가지고 있으며, 모두 3만 2천 여 개의 sample 데이터로 이루어져 있다. 각각은 전국에서 임의로 추출한 5000여 가구에 대한 평균적 시청률 자료이고, 주요 내용은 방영일자, 채널, 프로그램 명, 방영 시간, 시청률, 방송의 유형 코드 등이다.

일반적인 이러한 형태의 시청률 조사 자료는 각 프로그램의 시작 시간과 종료 시간의 임의성 때문에 시간대별 분석이 용이하지 않다. 시청자는 특정 시간에 하나의 방송을 선택적으로 시청하고 있으므로 시간대별 시청률로의 변환과정이 필요하다. 변환 과정에 대한 설명 및 가정은 다음과 같다.

1) 시간대별 분석을 위하여 방영 시간을 기준으로 정렬하였다. 분석의 용이성 및

목적에 부합하도록 시간대를 방송사의 황금 시간대(오후 8시~11시)로 축소하고 30분 단위로 분석을 실시 하였다.

- 2) 방영 날짜보다 방영 요일이 중요할 것이라는 가정하에 방영일을 요일로 변환하였다.
- 3) 원본 데이터의 시청률은 국내 모든 채널을 고려한 시청률이지만, 본 분석의 목적에 부합하도록 공중 파 주요 4개 채널 간의 상대적 시청률로 변환하였다.
- 4) 유형 코드에 의한 프로그램의 종류는 모두 400여 가지이지만, 분석의 용이성 및 분석 결과의 유용성을 위하여 6가지로 크게 분류하였다. 즉, 각각 최대한의 객관성을 유지하며 Info, News, Show, Drama, Sports, Film으로 분류되었다.
- 5) 특정 모델의 경우, 목적 변수인 시청률이 연속형인 경우보다 명목형인 경우 분석이 용이 하므로, 시청률을 각각 매우 높음, 높음, 보통, 낮음, 매우 낮음으로 분류 하였다. 분류는 보통을 기준으로 가우시안 분포의 형태를 따르도록 결정하였다.

이러한 가정들과 규칙을 바탕으로 최종적인 분석용 데이터의 변수들은 다음과 같고 sample의 수는 2735개이다.

<표 2.1> 분석용 데이터 변수

그룹	변수 설명	그룹	변수	설명
Day	Mon 월요일	KBS2_Pro	Show	쇼
Day	Tue 화요일	KBS2_Pro	Drama	드라마
Day	Wed 수요일	KBS2_Pro	Sports	스포츠
Day	Thu 목요일	KBS2_Pro	Film	영화
Day	Fri 금요일	MBC_Pro	Info	정보
Day	Sat 토요일	MBC_Pro	News	뉴스
Day	Sun 일요일	MBC_Pro	Show	쇼
Time_In	2000 시작시간 20시00분	MBC_Pro	Drama	드라마
Time_In	2030 시작시간 20시30분	MBC_Pro	Sports	스포츠

Time_In	2100	시작시간 21시00분	MBC_Pro	Film	영화
Time_In	2130	시작시간 21시30분	SBS_Pro	Info	정보
Time_In	2200	시작시간 22시00분	SBS_Pro	News	뉴스
Time_In	2230	시작시간 22시30분	SBS_Pro	Show	쇼
KBS1_Pro	Info	정보	SBS_Pro	Drama	드라마
KBS1_Pro	News	뉴스	SBS_Pro	Sports	스포츠
KBS1_Pro	Show	쇼	SBS_Pro	Film	영화
KBS1_Pro	Drama	드라마	Audience	KBS1_Aud	KBS1 시청률
KBS1_Pro	Sports	스포츠	Audience	KBS2_Aud	KBS2 시청률
KBS1_Pro	Film	영화	Audience	MBC_Aud	MBC 시청률
KBS2_Pro	Info	정보	Audience	SBS_Aud	SBS 시청률
KBS2_Pro	News	뉴스			

Sports	0.5%	1.1%	1.2%	1.5%
Film	0.2%	0.3%	0.2%	0.5%

<표 1.2> 각 채널 프로그램 종류별 시청률(평균)

	KBS1	KBS2	MBC	SBS
Info	14.4%	21.4%	22.3%	19.3%
News	34.8%	11.0%	22.0%	19.0%
Show	20.2%	22.2%	26.7%	18.0%
Drama	34.8%	37.9%	30.2%	27.6%
Sports	26.7%	24.3%	32.1%	32.4%
Film	5.0%	13.0%	29.5%	30.9%

	A	B	C	D	E	F
1	Day	Time_In	KBS1_Pro	KBS2_Pro	MBC_Pro	SBS_Pro
2	Thu	2000	Info	News	Show	News
3	Thu	2030	Drama	News	Drama	News
4	Thu	2100	News	Info	News	Show
5	Thu	2130	News	Show	News	Drama
6	Thu	2200	Info	Drama	Drama	Drama
7	Thu	2230	Info	Drama	Drama	Drama
8	Fri	2000	Info	News	Drama	News
9	Fri	2030	Drama	News	Drama	Info
10	Fri	2100	News	Info	News	Show

<그림 2.1> raw data

3. 사전 분석

본격적인 모델을 세우기 전에 사전 분석을 통해 데이터의 특성을 살펴 보았다. 전체적인 데이터 특성의 파악을 통해 모델을 세울 때의 변수 선택과 모델 설명력 향상에 도움을 줄 수 있다. 사전 분석을 통하여 각 채널의 프로그램 종류별 빈도, 각 채널의 프로그램 종류별 시청률, 각 채널의 요일별 시청률을 알아 보았으며 각각에 대한 결과를 요약한 표는 다음과 같다.

<표 1.1> 각 채널의 프로그램 종류별 빈도 (%)

	KBS1	KBS2	MBC	SBS
Info	40.8%	32.2%	6.9%	15.6%
News	31.8%	18.7%	34.9%	36.5%
Show	8.9%	30.8%	28.3%	18.0%
Drama	17.8%	16.9%	28.5%	27.9%

<표 1.3> 각 채널 요일별 시청률(상대적 평균)

	KBS1	KBS2	MBC	SBS
월	29.9%	18.7%	30.8%	20.6%
화	30.2%	18.8%	31.6%	19.4%
수	28.2%	23.1%	30.2%	18.5%
목	26.1%	22.8%	29.3%	21.8%
금	26.1%	24.9%	29.2%	19.8%
토	17.9%	30.9%	20.6%	30.1%
일	19.3%	33.3%	17.8%	29.6%

<표 1.1>에서 관찰할 수 있듯이 각 방송사는 각 프로그램의 편성 빈도에 특색을 가지고 있다. KBS1은 공공방송의 특성을 띠어 정보와 뉴스에 치중한 편성을 보이며, KBS2는 KBS1과의 차별성을 염두 해두어 쇼의 비중이 높다. MBC와 SBS는 상대적으로 드라마의 빈도가 높으며 MBC의 경우 쇼 프로그램의 편성 비율이 4개 방송사 중 가장 높다. 전체적으로 뉴스의 편성 빈도가 KBS2를 제외한 모든 방송사에서 높은데, 이는 황금 시간대에 항상 정규 뉴스가 편성되기 때문이다.

<표 1.2>에서 각 방송사는 어떤 종류의 프로그램을 편성 하였을 때 시청률이 높은지를 전반적으로 관찰해 볼 수 있다. KBS1은 방

송사의 성격상 뉴스의 시청률이 가장 높으며, KBS2는 드라마의 시청률이 가장 높다. MBC는 전반적으로 높은 시청률 분포를 보이며, SBS는 스포츠와 영화 프로그램의 시청률이 가장 높다. 스포츠는 편성 빈도에 비하여 높은 시청률을 나타내는데, 스포츠의 경우 주로 국가대표 축구 경기의 시청률이 이러한 결과에 가장 큰 영향을 미쳤다는 점을 고려 하여야 한다. 영화 역시 편성 빈도가 낮지만 MBC와 SBS의 경우 시청률이 높는데, 이는 명절 특선 영화 등 비교적 최근 개봉한 영화 방영에 의한 결과로서 이 점은 영화 전문 케이블 채널에 비해 공중파 3사가 가질 수 있는 장점이다. 각각의 시청률은 4개 방송사간의 상대적인 시청률로서, 일반적으로 언급되는 시청률의 단위와는 다름을 염두 해 두어야 한다.

<표 1.3>은 각 요일별 각 방송사의 시청률을 비교 분석한 것이다. KBS1은 주말 시청률이 확연히 낮으며, 월요일과 화요일의 시청률이 높은 형태를 보인다. KBS2는 KBS1과 반대로 주말 시청률이 가장 높으며, 월요일과 화요일의 시청률이 낮다. 이 것은 두 채널이 하나의 방송사에서 관리되고 있음에서 오는 전략적인 결과라고 판단된다. MBC는 전체적으로 높은 시청률을 보이나 주말 시청률이 낮은 편이며, SBS는 주말 시청률이 KBS2와 함께 높아, 두 방송사간의 주말 시청률 경쟁이 관찰된다.

사전분석을 통해 데이터의 특성을 파악하고 모델 구축 단계에 필요한 정보를 얻어 낼 수 있었다.

4. 모델 구축

모델 구축의 목적은 편성할 프로그램의 시청률 예측과 시청률이 높은 프로그램이 가지게 되는 규칙을 밝혀 내는 것이다. 시청률

예측을 위해서는 일반적으로 쓰이는 선형 모델을 사용하였으며 [3] 규칙 발견을 위해서는 의사결정나무를 사용하였다. [4] 데이터의 분할은 편성자가 현재까지의 시청률 자료를 토대로 미래의 시청률을 예상한다는 상황적 가정 하에 2005년 3월 말까지 23개월 치의 데이터 2189개를 training sample로, 4월 초부터 6월 말까지 3개월 치의 데이터를 596개 sample을 test sample로 분할하였다. 선형 모델의 경우 training sample로 모델을 구축한 후, test sample로 시청률을 예측한 후, 실제 시청률과의 비교를 통해 성능을 알아보았다. 의사결정나무의 경우도 training sample로 의사결정나무를 생성한 후 test sample로 오분류율을 측정하여 보았다.

4.1. 선형모델

분석의 목적이 방송사의 미래 시청률의 예측이므로 각 방송사의 시청률을 종속변수로 한 하나씩의 모델이 얻어진다. 선형 분석을 통해 분석해본 결과 다음과 같은 네 개의 모델이 얻어졌다. 변수 선택방법은 단계적 선택법을 선택하였고 기준은 AIC(Akaike Information Criterion) 최대화이며, 유의수준 $\alpha = 0.05$ 이다. 변수의 수가 많으므로 유효한 t-값의 절대값이 큰 순서대로 10개의 변수만 선택하였다.

4.1.1. KBS1 시청률 예측 모델

<표 4.1>KBS1 선형 모델

Parameter		Estimate	t-Value
Day	Tue	5.3805	16.50
Day	Sat	-6.1181	-15.41
Day	Sun	-5.2829	-12.82
KBS2_Pro	News	6.5691	11.55
Day	Mon	3.9323	9.11
KBS2_Pro	Info	3.5447	8.29

KBS1_Pro	Info	-7.4577	-7.37
Time_In	2030	4.8883	7.01
KBS1_Pro	Drama	6.9672	6.74
Time_In	2130	6.2704	6.66

(Adj R-sq: 0.8182, AIC : 7589.0248, MSE : 31.6253)

표에서 관찰할 수 있듯이 KBS1의 경우 무슨 요일인지가 시청률에 가장 큰 영향을 준다. 월요일과 화요일은 시청률에 긍정적 영향을 주나, 주말인 것은 시청률에 부정적인 영향을 준다. 또한 KBS2의 프로그램이 News나 Info일 경우 시청자들은 KBS2보다 KBS1을 선택한다는 점을 관찰할 수 있다.

4.1.2. KBS2 시청률 예측 모델

<표 4.2>KBS2 선형 모델

Parameter		Estimate	t-Value
KBS2_Pro	Drama	9.9879	11.26
KBS2_Pro	News	-11.1137	-15.41
Day	Tue	-6.0244	-10.49
Day	Sun	5.1757	7.13
KBS1_Pro	News	-10.6227	-5.15
Day	Sat	3.5694	5.11
SBS_Pro	Info	4.7421	5.00
Day	Fri	3.1704	4.99
MBC_Pro	Show	-4.7271	-4.99
MBC_Pro	News	7.7134	4.48

(Adj R-sq: 0.5841, AIC : 10065.4882, MSE : 98.0814)

KBS2의 경우에는 본 채널의 프로그램 종류에 의해 시청률이 가장 크게 영향을 받는다. 또한 주말인 것이 시청률에 긍정적 영향을 주며, KBS1에서 News를 방영할 경우 시청률에 중요한 부정적 영향이 있다는 점을 관찰할 수 있다.

4.1.3. MBC 시청률 예측 모델

<표 4.3>MBC 선형 모델

Parameter		Estimate	t-Value
Day	Sun	-7.3405	-14.83
Day	Tue	5.7688	14.74
Day	Sat	-4.7549	-9.98
SBS_Pro	Drama	3.5903	7.52
Time_In	2000	6.4245	7.01
Day	Mon	2.9985	5.79
KBS2_Pro	Drama	-3.4180	-5.65
MBC_Pro	News	-6.0689	-5.17
Time_In	2200	-3.4257	-4.99
KBS2_Pro	Film	8.1459	4.53

(Adj R-sq : 0.4910, AIC : 8388.7362, MSE : 45.5795)

MBC의 경우 주말인 것이 시청률에 가장 부정적인 영향을 주는 것으로 나타났다. 반면 월요일과 화요일의 경우는 시청률에 긍정적인 영향을 보여 시청률이 무슨 요일인지에 따라 변화가 크다고 할 수 있다.

4.1.4. SBS 시청률 예측 모델

<표 4.4>SBS 선형 모델

Parameter		Estimate	t-Value
Day	Sat	7.3036	13.72
Day	Sun	7.4477	13.47
Day	Tue	-5.1249	-11.72
SBS_Pro	Show	-5.2258	-9.61
KBS2_Pro	Drama	-4.7305	-7.00
Time_In	2030	-6.4477	-6.90
MBC_Pro	Show	4.9584	6.87
Time_In	2000	-6.6878	-6.53
Day	Mon	-3.6701	-6.34
Day	Fri	-2.8510	-5.89

(Adj R-sq : 0.5147, AIC : 8872.9855, MSE : 56.8706)

SBS의 경우도 무슨 요일인지가 가장 중요한 변수로 나타났다. MBC와 반대로 주말의 시청률이 긍정적이며, 월요일과 화요일은 시청

률이 부정적이다. 그 밖에 타 방송사의 편성과 편성 시간에 의해 시청률이 영향 받게 됨을 알 수 있다.

4.2. 의사결정나무

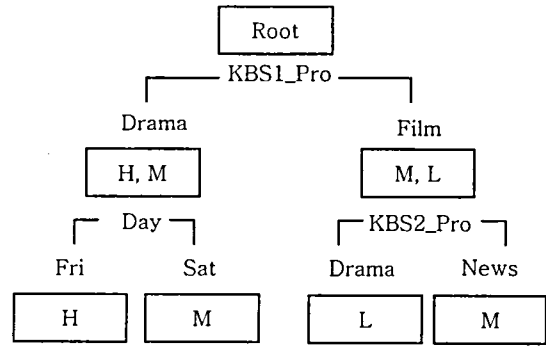
의사결정나무의 최대 장점은 입력 변수에 의해 목표 변수가 결정되게 되는 규칙을 제시해 준다는 점이다. 분석에 사용한 데이터에서 목표 변수인 시청률은 연속형 변수이지만 이러한 의사결정나무의 장점인 시청률 결정 규칙 발견을 위해 명목형 변수로 변환 하였다. 변환한 방법은 아래와 같이 M(Middle)을 기준으로 가우시안 분포의 형태를 따르도록 5단계로 구분하였다.

<표 4.5>시청률 변환 코드(상대 시청률 기준)

시청률 구간	Code
50%이상	HH
35%~50%	H
20%~35%	M
10%~20%	L
10%이하	LL

의사 결정나무의 경우 나무의 깊이 결정이 중요한데, 깊이가 깊어질수록 각 잎의 순수도는 커지지만 모델로 인해 얻을 수 있는 규칙이 복잡해진다. 따라서 본 연구의 취지에 맞는 간단한 규칙 도출을 위해 본 논문에서는 깊이 2의 의사결정나무만 제시한다. 실제 오분류율 측정에 사용된 의사결정나무의 깊이는 3이다. 모델 구축시에 사용한 분리 기준은 Chi-square test이었으며, 유효수준은 0.05이다. 각 채널에 대하여 4개의 의사결정나무 모델을 구하였으며 각각에 대한 결과는 아래와 같다.

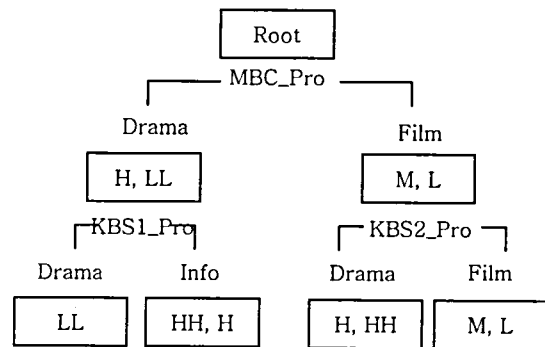
4.2.1. KBS1의 시청률 의사결정나무



<그림 4.1> KBS1 의사결정나무

의사결정나무를 관찰해 보면 KBS1의 시청률은 KBS1이 Drama를 방영하고 금요일이면 시청률이 높고, KBS1에서 영화를 방영하고 KBS2에서 드라마를 방영하면 시청률이 낮다는 규칙을 발견할 수 있다.

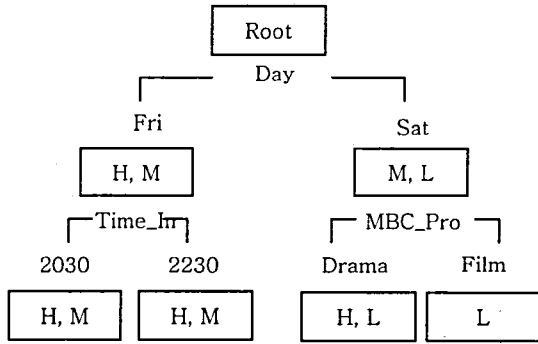
4.2.2. KBS2의 시청률 의사결정나무



<그림 4.2> KBS2 의사결정나무

KBS2의 경우 MBC에서 드라마를 방영하고 KBS1에서 정보 프로그램을 방영하면 시청률이 매우 높거나 높게 된다는 규칙을 발견할 수 있다. 반면 MBC에서 드라마를 방영하나, KBS1에서 드라마를 방영할 경우 KBS2의 시청률은 매우 낮게 된다.

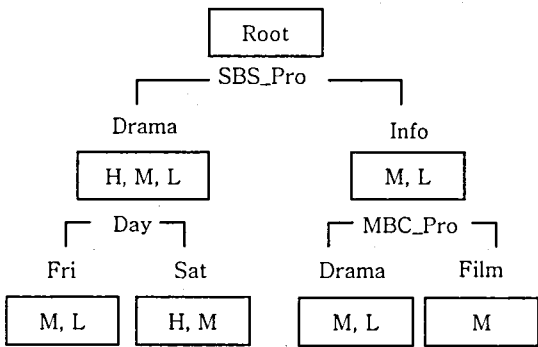
4.2.3. MBC의 시청률 의사결정나무



<그림 4.3> MBC 의사결정나무

MBC의 경우 금요일이면 시간대에 관계 없이 비교적 높은 시청률을 보였다. 하지만 토요일이고 MBC에서 영화를 편성하면 즉, 토요일 영화 프로그램은 시청률이 낮음을 알 수 있다.

4.2.4. SBS의 시청률 의사결정나무



<그림 4.4> SBS 의사결정나무

SBS의 경우 자체의 편성이 드라마 일 경우 요일에 따라 시청률이 달라진다. 주말드라마가 시청률이 높으며 금요 드라마는 시청률이 평균이거나 그 이하이다. SBS자체에서 정보 프로그램을 편성할 경우, MBC의 편성에 따라 시청률이 영향을 받게 된다.

4.2.5. 발견된 규칙의 정리

의사결정나무를 이용하여 밝혀낸 규칙 중 앞 마디의 순수도가 높아(80% 이상) 명확한 것을 정리하면 다음과 같다.

<표 4.6>주요 규칙

번호	규칙
1	KBS1_Pro가 Film-> KBS2_Pro가 News-> KBS1시청률 : 평균적
2	MBC_Pro가 Drama-> KBS1_Pro가 Drama -> KBS2_Pro가 Info -> KBS2 시청률: 매우 낮음
3	금요일-> 시간대가 오후 10시 30분부터-> MBC_Pro가 Info-> MBC시청률 : 평균적
4	토요일 -> MBC_Pro가 Film -> MBC시청률 : 낮음
5	SBS_Pro가 Info-> MBC_Pro가 Drama -> 월요일 -> SBS 시청률 : 낮음
6	SBS_Pro가 Info -> MBC_Pro가 Film -> KBS2_Pro가 News -> SBS시청률 : 평균적

5. 결론

이상으로 국내 TV시청률 조사 자료를 이용한 시청률 예측 모델 및 시청률 결정 규칙을 각각 선형모델과 의사결정나무를 이용하여 구축하였다. 결론적으로 사전 분석에서 관찰한 결과와 대부분 일치하는 결과가 관찰되었으며 시청률에 영향을 주는 각 변수들을 선정할 수 있었다.

실제로 시청률은 본 연구에서 사용된 데이터에 속한 변수들 이외에도 개개인의 성향에 의해 민감하게 영향을 받게 되는 변수이므로 보다 정확한 시청률 예측을 위해서는 각 프로그램의 성격을 보다 세부적으로 정의할 필요가 있다. 드라마의 종류도 장르별로 보다 세분화가 가능하며, 정보 프로그램도 실제로 정보의 내용에 따라 시청률의 차이가 크다. 또한 현재의 미디어 환경을 볼 때 공중과 4개 채널에만 국한 할 것이 아니라 전체 채널을 대상으로 한 분석을 실시함으로써 보다 정확한 시청률 예측 모델을 세울 수 있을 것이다.

<참고문헌>

[1] Paulo Giudici(2003), *Applied Data Mining*,
John Wiley & Sons. pp.323-351

[2] AGB Nielsen Korea :
<http://www.acnielsen.co.kr>

[3] David Hand, Heikki Mannila and Padhraic
Smyth(2001), *Principles of Data Mining*, The
MIT Press. pp.367-397

[4] 배화수, 조대현 et all(2004), *SAS E-miner 를
이용한 데이터 마이닝*, 교우사. pp.147-174