

사용자 패턴의 자동추출을 통한 TV-Anytime 기반 사용자 선호정보 관리 시스템

신사임, 이종설, 이석필
전자부품연구원
{miror,leejs,lsplib}@keti.re.kr

Metadata based digital broadcasting with TV-Anytime specification

Saim Shin, Jong-Soel Lee, Soek-Pil Lee
Digital Media Research Center, Korea Electronic Technology Institute

요 약

디지털 방송에서의 맞춤형 서비스란 사용자가 원하는 방송 프로그램만을 사용자가 원하는 시간에 볼 수 있게 해 두는 서비스를 말한다. 본 연구는 맞춤형 방송 서비스를 위한 사용자 별 선호정보를 자동으로 추출하여 관리하는 사용자 선호정보 관리 시스템을 개발하였다. 이 시스템은 맞춤형 방송 서비스를 위한 표준인 TV-Anytime에서 제안하고 있는 메타데이터를 기반으로 사용자 선호정보와 사용자 히스토리 정보를 사용하여 맞춤형 방송 서비스를 위한 선호 프로그램과 채널을 예상하여 사용자에게 제시한다. 이 과정에서 사용하는 사용자 선호정보는 사용자 별로 청취 과정에서 채널, 시간대, 방송에 따른 반응 행동기록을 토대로 자동으로 추출하고 업데이트하는 방법을 제안한다.

1. 서 론

디지털 방송 환경이 조성되고 디지털 방송 서비스가 확대되면서, 사용자는 다채널 다매체에서 원하는 정보나 방송 프로그램만을 보기 원하게 된다. 이러한 사용자를 위한 방송서비스가 맞춤형 서비스이다. 맞춤형 서비스를 위해 방송 프로그램의 정보를 알 수 있도록 하는 메타데이터(Metadata)를 검색하여 사용자에게 제공한다. 이 밖에도 사용자의 취향정보를 위한 메타데이터인 사용자 선호정보(User Preferences)도 필요하다.

본 논문은 맞춤형 방송 서비스를 위한 사용자 선호정보 관리 시스템의 구축에 대한 연구이다. 본 시스템은 사용자 별 사용기록(Usage History)을 분석하여 사용자가 선호하는 시간대와 방송 프로그램 시청 시 어떤 행동 패턴을 보이는지를 자동으로 학습한다. 이렇게 학습하여 추출한 사용자의 선호패턴과 비선호패턴을 기반으로 사용자의 선호정보를 직접 입력 받지 않아도 추출 및 업데이트가 가능하게 해 준다. 새롭게 입력 받은 전자프로그램 가이드(Electronic Program Guide)와 프로그램 별 메타데이터 정보들을 사용자 선호정보를 기준으로 검색하여 사용자에게 제공하여 원하는 방송 프로그램의 선택을 도와 준다.

사용자 선호정보 관리 시스템은 디지털방송 관련 국제표준인 MPEG-2와 TV-Anytime 표준안을 기반으로 개발하여 호환성을 고려하였다. 또한, 웹 정보검색의 이론과 방법론을 적용하여 멀티미디어 검색을 위한 특징추출 및

검색방법을 제안하여 검색속도를 향상시키고 웹 콘텐츠로 의 시스템 확장이 향후연구로 용이하도록 개발하였다.

2. 관련 연구

2.1 MPEG과 TV-Anytime

MPEG은 국제표준화 기구인 ISO/IEC 산하의 공동기술위원회(JTC1)에 소속된 작업그룹(WG : Working Group)으로서, 공식명칭은 ISO/IEC JTC1 SC29 WG11이며 그 별칭인 Moving Picture Experts Group(동영상전문가그룹)의 약자이다. MPEG은 1980년대 후반에 활동을 시작하여 그동안 동영상 및 오디오의 압축 전송 표준을 개발하여 왔으며, 현재는 다양한 멀티미디어 응용과 디지털 콘텐츠의 전자상거래를 위한 표준화 작업을 하고 있다.

MPEG 표준에는 표준화가 완료된 MPEG-1, MPEG-2, MPEG-4, MPEG-7과 현재 표준화가 진행 중인 MPEG-21이 있다. 이 중 고품질 통신 및 방송에의 응용을 목표로 한 MPEG2는 국제표준을 완성하여 비디오 CD 및 인터넷 파일 전송, DVD 및 디지털 방송 등에 이미 활발하게 활용되고 있다. 본 연구에서 시스템 개발의 기준으로 삼고 있는 TV-Anytime 국제표준안은 MPEG-2를 기반으로 하여 제정되었다.

TV-Anytime은 디지털 방송의 맞춤형 서비스를 위한 국제 표준 규격이다. 이 규격은 맞춤형 서비스에서 필요한 다양한 메타데이터 규격을 웹 서비스 규격을 기반으로 하여 제정하고 제공한다. 이 중 메타데이터를 이용해

사용자가 원하는 프로그램만을 선택하기 위한 검색 및 탐색 서버에 대한 규격을 구현하는데 본 논문에서 제안하는 사용자 선호정보 관리 방법으로 선호정보 메타데이터의 자동 관리를 가능하게 하였다. TV-Anytime은 개개인의 사용자 선호정보를 바탕으로 한 Targeting 서비스에 대한 표준을 제정하였다. Targeting 서비스는 사용자 선호 정보에 추가적인 세부 정보 - 기분, 기후, 개인 소비정보 등 - 를 바탕으로 개개인의 현재 상황에 적합한 콘텐츠를 제공하는 것을 목적으로 한다. 이와 같은 서비스를 통해 TV-프로그램, 광고, 정보, 위치 등에 대한 targeting 서비스가 가능하다.

2.2 맞춤형 방송의 선호정보 관리

[1]은 시청자의 행동을 모니터링해 주는 사용자 행동 정보를 사용자 히스토리 (User History) 메타데이터를 바탕으로 선호도를 자동으로 유추하는 시스템을 구축하였다. 해당 프로그램에 대한 선호도가 높은 경우와 낮은 경우의 사용자 행동들을 규칙화한다. 사용자 히스토리에서 추출한 행동 패턴과 규칙과의 비교를 통하여 사용자 별 세부 선호정보들을 추출하고 있다. 이 연구는 TV-Anytime 발표 전 MPEG-7의 사용자 히스토리와 사용자 선호정보 메타데이터 규격을 바탕으로 시스템을 기준하였다. 기본적인 사용자 선호정보의 자동 추출에 대한 개념을 정리하였지만, 사용자의 행동패턴의 분석을 임의로 선정한 간단한 몇 가지 규칙들을 토대로 구현하는 데 머물렀다.

[2]는 사용자 히스토리 정보를 통계적으로 분석하여 순차적으로 적용하면서 선호 정보를 걸러내는 방식으로 선호정보를 추출 및 관리하고 있다. 또한, 세부정보 간의 계층관계를 적용하여 정교한 가중치 및 확률계산을 적용하였다. [2]는 다량의 사용자 행동정보를 통계적인 방법을 도입하여 정형화시켜 적용하였지만, 적용 순서, 행동의 조합, 가중치 및 일반화 방법은 임의로 규칙화하여 단순화 한 한계를 가지고 있다.

[3]은 Mutual Information과 Bayesian Network을 기반으로 하는 사용자 선호정보 예측 알고리즘을 제안하였다. 또한, 만족도 평가를 위한 평가 방법을 제안하였다. 시스템의 추천과 실제 사용자의 선택 프로그램과의 차이의 평균 error rate를 계산하여 성능평가 방법을 제안하였다. 기존의 방법론들과 비교하여 feature들을 자동으로 순위화 및 체계화하여 추론 네트워크를 구성하는 베이지안 네트워크 방법을 사용하여 자동화 하였고, 자동화를 위하여 본격적으로 통계적 학습 방법을 도입하였다는 점에서 이 연구의 큰 의의를 둘 수 있다. 그러나, 자동 통계추론 방식은 베이지안 방법을 사용한다면, 더 많은 feature들의 자동 추출 및 적용이 가능할 텐데, 사용한 feature가 단순하여 통계적 추론방식의 장점을 최대한 살리지 못하였다. 또한, [3]은 통계적인 평가방법을 최초로 제안하여 시스템의 객관적인 평가를 시도하였다.

[4]는 TV-Anytime을 바탕으로 맞춤형 방송 시스템을 구축하였다. 사용자의 성향을 반영하는 프로그램 추천 기록에 대한 대표적인 두 가지 패턴을 Usage History에 기록된 사용자의 시청기록에서 카운트하여 간단한 수식을 도출하여 프로그램의 선호 정도를 계산한다. 그러나,

제안하는 선호도 계산 방법이 단순하기 때문에, 빈도수를 통계적 정규화 과정 없이 수식에 그대로 반영하여 애 민감하고 학습에 사용되는 데이터에 의존적인 경향이 강하다. 또한, 사용자 선호정보의 자동 추출에 관한 연구는 진행되지 않은 채, 입력 받은 선호정보를 기반으로 선호 프로그램을 추출하는 부분만을 포함하고 있다.

3. 사용자 선호정보 관리시스템

제안하는 사용자 별 선호정보 관리 시스템 개발의 가정은 다음과 같다.

방송시청 시 선호하는 프로그램에 대한 정보는 사용자의 생활패턴과 취향 등에 따라 다양하게 나타난다. 사용자에게 따라 드라마와 영화를 선호하여 방송을 통하여 이들 장르를 주로 시청하는 사용자가 있는 반면, 스포츠 채널을 선호하기도 한다. 또한, 이러한 사용자의 선호도는 시청 시간에 따라 선호하는 채널이나 장르가 다르게 나타나기도 하는 등, 사용자의 선호패턴은 복잡 다양하게 나타날 수 있다. 그러므로, 이러한 패턴을 자동으로 파악하고 관리하여 선호 프로그램을 우선 사용자에게 제공하여 주는 방식으로 시청만족도를 높이는 시스템을 필요로 한다.

그리고, 사용자 별 선호도는 단기적, 장기적으로 주변환경과 사용자의 관심변화에 따라 꾸준히 변화할 수 있다. 사용자의 선호 프로그램과 채널정보는 일시적으로 바뀌기도 하고, 점진적으로 변하기도 한다. 예를 들어, 월드컵 기간에는 일시적으로 스포츠 채널 및 스포츠 프로그램의 시청빈도가 높아진다거나 하는 일시적인 선호정보의 변화를 들 수 있다. 혹은, 어린 시절에는 만화채널이나 어린이 프로그램 방영시간을 선호하였지만, 성장하면서 선호하는 채널과 장르가 바뀌는 현상도 일반적이다. 사용자 선호정보는 영구적이 아니므로, 꾸준한 업데이트 및 사용자의 시청패턴을 모니터링 해야 만족도가 높은 맞춤형방송 서비스를 기대할 수 있을 것이다.

마지막으로, 선호정보의 변화에 따라 사용자의 시청 시 행동패턴 또한 변화한다. 사용자의 주 시청시간이 밤 시간만 가능하다면, 선호 프로그램을 밤에 즐기 위하여 'Recording' 을 즐겨 한다던가, 스포츠 경기의 주요 장면을 자세히 보기 위하여 'Repeat' 이나 'Zoom' 명령을 자주 사용하게 된다던가 하는 예이다. 즉, 사용자의 시청 시 행동패턴은 선호시간과 장르, 또는 사용자의 맞춤형방송에의 적응 정도에 따라 사용패턴이 확장될 수도 있고, 선호 프로그램과 비선호 프로그램에 대한 대응행동 또한 바뀔 수 있다. 그러므로, 사용자 행동패턴의 수집 또한 이러한 변화에 맞추어 주기적으로 모니터링 하여 최신의 사용자 행동패턴을 수집하여야 최적의 선호정보를 수집할 수 있다.

그림 1은 제안하는 사용자 선호정보 관리 시스템을 보여 준다. 선호정보 관리 시스템은 시청 시 사용자 행동 시그널 정보를 입력 받아 이 정보를 사용자 행동기록을 XML 문서로 생성하는 사용자 행동기록 (Usage History) 생성기와, 생성한 행동기록을 바탕으로 사용자 선호정보를 추출하는 사용자 선호정보 (User Preferences) 관리도들로

나누어진다. 사용자 선호정보 관리모듈에는 사용자 행동 패턴을 분석하는 패턴 수집기 (Pattern collector)와 사용자 별 선호프로그램에 대한 행동패턴 및 비선호 프로그램에 대한 행동패턴을 수집하는 패턴 결정기 (Feature Predictor)를 포함한다. 이러한 패턴 수집기에서 수집된 최적의 선호 / 비선호 패턴들을 기반으로 TV-Anytime 표준안에서 제안하고 있는 사용자 선호정보들을 추출하고 업데이트하는 선호정보 추출모듈 (Preferences predictor)도 있다. 사용자 선호정보 관리모듈에서 추출한 사용자 별 선호정보를 기반으로 추천모듈 (Recommendation Module)에서는 선호정보와 프로그램 별 메타데이터를 기반으로 EPG의 프로그램들을 점수화하고 순위화하여 선호할 만한 프로그램 리스트를 사용자에게 제시하게 된다.

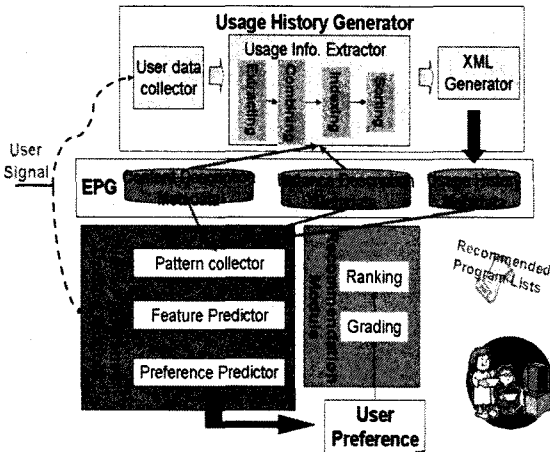


그림 1 사용자 선호정보 관리 시스템

3.1 사용자 행동패턴 관찰

시청 중의 사용자 행동양식을 관찰하기 위하여 패턴 풀 (pattern pool)을 제안한다. 패턴 풀에는 사용자 행동기록에서 추출 가능한 모든 연속적인 사용자 행동패턴들을 수집한다.

패턴 풀에 수집된 행동패턴들은 특성추출 과정에서 평가를 통하여 선호정보 추출에 사용할 선호패턴 및 비선호 패턴 목록을 결정하게 된다. 특성추출과정은 패턴 풀에 수집한 패턴들에 대하여 사용자 행동기록에서 그 빈도를 기반으로 상호 정보값 (Mutual Information)을 계산한다. 특정 패턴 A를 위한 상호 정보값의 계산식은 식 (1)과 같다.

$$MI_A = \log\left(\frac{p(f_A | pos)}{p(f_A)}\right) \quad \text{----- (1)}$$

식 (1)에서 $p(f_A | pos)$ 는 패턴 A가 사용자가 선택하여 시청한 선호 프로그램 시청 시 등장할 빈도를 토대로 하는 확률값을 의미한다. $p(f_A)$ 는 전체 사용자 행동기록에서 패턴 A의 등장확률을 정의한다. 상호 정보값 (MI)는 정보이론에 의하여 사용자의 선호 프로그램과 해당 패턴의 등장빈도 사이의 독립성 정도를 통계적으로 설명한다. 즉, MI가 0보다 큰 패턴은 등장빈도와 사용자의 선호도가

비례한다고 말할 수 있고, MI가 0인 패턴은 사용자의 선호도와 완전히 독립적인 무관한 패턴이다. 또한, MI가 0보다 작은 경우는 사용자 선호도와 패턴의 빈도가 반비례하는 부정적인 패턴이라고 할 수 있다.

본 연구에서는 특성추출 과정에서 사용자의 선호도를 점수화하기 위한 선호패턴과 비선호패턴을 반영하기 위한 비선호 패턴을 식 (1)의 MI값을 기준으로 관리한다. 패턴 풀의 전체 패턴의 평균 상호 정보값인 MI_{avg} 값을 기준으로 MI_{avg} 와 0보다 큰 MI를 가지는 패턴을 선호패턴으로, 0보다 작은 MI의 패턴을 비선호 패턴으로 결정하였다.

이러한 패턴수집과 특성추출 과정은 사용자가 방송을 시청하지 않는 시간에 주기적으로 새로운 패턴을 수집하여 최적의 선호 / 비선호 패턴으로 관리한다. 자동적인 사용자 별 행동패턴 관찰과정은 사용자의 최신의 선호정보를 관리하는데 유리하다.

3.2 정보검색이론 기반 사용자 선호정보 추출

패턴 결정기에서 수집된 선호 / 비선호 패턴들을 기반으로 사용자 선호정보를 점수화하여 수집한다. TV-Anytime 사용자 선호정보 표준안에서 제안하고 있는 각 선호정보들을 패턴의 빈도를 기반으로 점수화한다. 최근 업데이트된 사용자 행동기록을 분석한 후, 각 선호정보 후보목록 별로 선호 / 비선호 패턴의 빈도를 식 (2)와 같이 점수화하여 점수가 높은 선호정보 후보들을 사용자 선호정보로 결정하게 된다.

$$UP_Score_i = \frac{\sum_{m=1}^{N_{pos}} \alpha_m \cdot f_m - \sum_{m=1}^{N_{neg}} \alpha_{m'} \cdot f_{m'}}{N} \quad \text{----- (2)}$$

$$N = N_{pos} + N_{neg}$$

식 (2)는 사용자 선호정보 대상목록 중 i번째 선호정보 후보의 선호도 점수 UP_Score_i 의 계산식이다. N_{pos} 는 선호 패턴의 개수를, N_{neg} 는 비선호 패턴의 개수를 의미한다. α_m 은 각 패턴의 중요도를 나타내는 가중치이고, 본 논문의 평가에서는 패턴 별 가중치를 부여하지 않고 1로 계산하였다. 즉, 가중치를 고려한 선호패턴의 빈도와 비선호 패턴의 빈도의 차를 전체 패턴의 개수로 나눈 값을 각 선호정보 후보의 점수로 결정한다. UP_Score 로 정렬한 선호정보 후보에서 상위의 후보들을 각 사용자 별 선호정보로 결정한다.

제안하는 시스템에서는 최근에 수집한 사용자행동 기록들을 바탕으로 이 장에서 설명한 패턴추출 및 선호정보추출과정을 주기적으로 수행하여 최신의 사용자 선호정보를 관리할 수 있다. 최신의 사용자 선호정보의 확보는 사용자의 만족도가 높은 방송의 추천을 가능하게 하므로, 높은 만족도의 사용자 맞춤형방송 서비스를 실현할 수 있다.

4. 평가

사용자 선호정보 관리를 통한 맞춤형 방송 시스템의 만족도를 평가하기 위하여 본 논문은 정보검색 시스템의 평가방법인 역순위평균 (Mean Reciprocal Rank) 방법을 적

용하였다.

4.1 평가 방법

역순위 평균은 정보검색 시스템에서 정답으로 자동 추출한 웹 문서들의 적합성 정도를 평가하는 통계적인 평가 방법의 하나로, 웹 기반 정보검색 시스템의 성능 평가에 가장 일반적이고 표준이 되는 방법이다. 역순위 평균은 검색 시스템에 요청한 질의의 실제 정답을 포함하는 문서가 시스템이 제시한 검색 결과에서 몇 위에 위치하는지를 파악한다. 각 질의의 정답 순위의 역수를 합하여 그 평균을 계산하여, 제안하는 시스템의 정확도를 평가하는 방법이고, 역순위 평균의 수식은 아래 식 (3)과 같다.

$$MRR = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \quad \text{----- (3)}$$

본 연구에서는 사용자가 실제 선택하여 시청한 프로그램을 정답 데이터로 가정하였다. 사용자가 선택한 프로그램이 맞춤형 방송의 선호정보 기반 추천 목록에서 몇 위에 위치하고 있는지를 파악하여 식 (3)의 방식으로 역순위 평균을 계산하였다.

식 (3)에서 EPGE 등장하는 각각의 콘텐츠를 의미하고, r_i 는 추천 방송 목록에서 i 의 순위를 말한다.

4.2 평가 결과

선호정보 기반 추천 시스템의 평가를 수행한 환경은 다음과 같다. 평가를 위한 데이터는 케이블과 공중파 방송을 포함한 78개 채널의 7일 간의 전자 콘텐츠 가이드 (Electronic Content Guide)를 대상으로 하였다. 평가 대상 전자 콘텐츠 가이드에는 77,882개의 콘텐츠와 각 콘텐츠에 대한 메타데이터 - 채널정보, 제목, 방송시간, 캐스팅 정보 등 - 를 포함하고 있다. 각 콘텐츠의 장르 정보는 10개의 대분류와 78개의 세부장르로 분류하고 있다. 평가 대상 데이터를 대상으로 4명의 시청자의 7일간의 시청기록을 수집하여, 그 중 5일치의 시청기록으로 사용자 별 선호정보를 자동으로 학습하였다. 자동 추출된 사용자 선호정보로 선호정보 추천 시스템이 추천한 사용자 별 추천 프로그램 목록에서 사용자가 나머지 2일 동안 실제로 시청하거나 녹화한 프로그램을 추천 시스템의 정답 콘텐츠로 가정한다면, 추천 목록에서의 정답 콘텐츠의 순위정보를 도출할 수 있다. 이 결과를 가지고 계산한 MRR 기반 평가 결과는 표 1과 같다.

표 1 MRR 기반 평가 결과

사용자	MRR	
	대분류 기준	소분류 기준
30대 남성 (A)	0.431	0.58
30대 남성 (B)	0.297	0.351
30대 여성 (A)	0.673	0.774
30대 여성 (B)	0.355	0.583
평균	0.439	0.571

표 1에서의 MRR 결과는 제안하는 추천 시스템의 추천 결과가 실제 사용자가 선호하여 즐겨 시청하는 대부분의 프로그램들을 상위 1-5위 사이에 포함하고 있음을 통계적으로 나타내고 있다.

5. 결론 및 향후 연구

본 논문은 사용자 맞춤형 방송 서비스를 위한 사용자 선호정보 관리 시스템 개발에 대한 연구이다. 관리 시스템은 사용자의 시청기록을 저장하고 있는 사용자 행동기록을 분석하여, 사용자의 행동패턴을 추출한다. 행동패턴의 분석을 통하여 선호정보 관리 시스템은 사용자가 선호하는 프로그램 과 선호하지 않는 프로그램에 따른 선호패턴 및 비선호 패턴을 기억한다. 추출한 패턴에 의하여 사용자 선호정보를 추출하고 수정하면서, 사용자에게 최적의 선호 프로그램의 추천을 가능하게 한다. 제안하는 선호정보 관리 시스템은 사용자가 번거롭게 자신의 선호정보를 입력하고 수정하지 않아도 최적의 추천결과를 통하여 원하는 프로그램을 추천 받을 수 있다. 또한, 국제 표준인 TV-Anytime 표준을 지원하는 시스템은 그 호환성 및 확장성이 우수하다.

향후 연구로는 제안하는 사용자 선호정보 추출 방식을 정교화하는 작업이 필요하다. 또한, 선호정보 자동학습의 효율성을 높이기 위하여, 학습 시간을 단축하기 위한 알고리즘의 보완도 필요하다. 마지막으로, 맞춤형 방송 시스템의 보다 정확한 성능평가를 위해서는, 웹 정보추출 시스템의 성능평가에서 가져온 MRR 기반 평가 방식의 보완 및 평가 결과를 시스템에 반영하는 피드백 방식의 시스템 보완도 가능한 연구이다.

참고 문헌

[1] 류지웅, 배빛나라, 김문철, 남재호, 강경옥, “ 디지털 방송을 위한 지능형 프로그램 가이드 (Intelligent Program Guide for Digital Broadcasting)”, 한국방송공학회, 2001.

[2] 배빛나라, 류지웅, 김문철, 남재호, 강경옥, 노용만, “ TV Anytime 응용을 위한 사용자 선호도 추출 및 갱신 알고리즘 (User Preference Extraction and Update Algorithm for TV Anytime Applications)”, 한국방송공학회, 2001.

[3] Sanggil Kang, Jeongyeon Lim, Munchurl Kim, “ Modeling the user preference of broadcasting content using Bayesian networks ”, Journal of electronic Imaging, 14 (2), 2006.

[4] 전자부품연구원, “ 맞춤형 서비스 솔루션 기술개발 (최종보고서)”, 산업자원부, 2005. 8.

[5] Fotis G. Kazasis, Nektarios Moumoutzis, Nikos Pappas, Anastatia Karanastasi, Stavros Christodoulakis, “ Designing ubiquitous Personalized TV-Anytime Services ”, Workshop on Ubiquitous Mobile Information and Collaboration Systems, 2003.

[6] 류지웅, 김문철, 남재호, 강경옥, 김진웅, “ 사용자 선호도 기반 지능형 프로그램 가이드 ”, 방송공학회 논문지 7권 2호, 153~167, 2002.

[7] 박경세, “ 퍼스널 TV의 비즈니스 환경 분석 ”, 11. 30, 방송동향과 분석 통권 165호, 2002.