맞춤형 추천을 위한 다중센서기반 사용자 인지 지능형 TV 플랫폼

김은솔[1], 김지섭[1], 이대근[1], 장병탁[1]
eskim@bi.snu.ac.kr, jkim@bi.snu.ac.kr, elnn@elnn.kr, btzhang@bi.snu.ac.kr
[1]서울대학교 컴퓨터공학부

Intelligent TV That Uses Multi-sensors to Recognize User Behavior for Customized Recommendation

요 약

IPTV가 보급되면서 사용자에게 공급되는 컨텐츠의 양이 기하급수적으로 늘어나고 있는 가운데, 사용자가 좋아할 만한 컨텐츠를 추천하는 시스템의 필요성이 대두되고 있다. 컨텐츠 추천에 주로 사용되는 알고리즘에는 크게 협력적 여과와 내용기반 여과 방법이 많이 사용되며, 이 두 가지 방법을 혼합하여 사용하기도 한다. 하지만 이 방법들은 사용자의 컨텐츠에 대한 선호도를 바탕으로 비슷한 성향의 사용자들을 그룹지어 컨텐츠를 추천하기 때문에, 사용자의 성향과 관심도가 급격히 변화하는 텔레비전 환경에서는 신뢰도가 떨어진다. 우리는 이를 보완할 수 있는 방법으로 컨텐츠 모델링과 사용자의 센서 데이터를 이용한 사용자 모델링을 기반으로 하는 실시간 맞춤형 추천 알고리즘을 제시한다. 본 논문에서는 이 알고리즘을 위한 실험 플랫폼을 제시하고 실제 실험 플랫폼에서 얻은 데이터에 대해 논의한다.

1. 서 론

UCSD 대학의 Bohn 교수팀이 2011년에 출간한 미국 국민의 정보 소비 성향에 관한 레포트에 따르면, 미국인들은 하루에 12시간동안 정보를 얻고 그 중에서 5시간은 텔레비전을 통해서 얻는다고 한다[1]. 이는 인터넷, 라디오, 전화, 책과 같은 다른 정보 매체와 비교했을 때 월등히 높은 수치이다. 2000년대 들어, 컴퓨터와 인터넷이 보급되면서 텔레비전을 통한 정보 소비 시간은 점차 줄어들었지만 최근 IPTV, smart TV 등이 보급되면서 다시 늘어나고 있는 추세이며, 컴퓨터와 비교했을 때 2.5배 정도 높은 수치를 보인다. IPTV나 smart TV 환경에서 장시간 정보 소비가 가능한 이유는 일방적으로 정보가 제공되는 것이 아니라 사용자가 자신의 요구에 따라 정보를 선택할 수 있기 때문이다. 이 때문에 사용자의 요구를 정확히 예측하여 알맞은 컨텐츠를 추천하는 기술이 매우 중요하다.

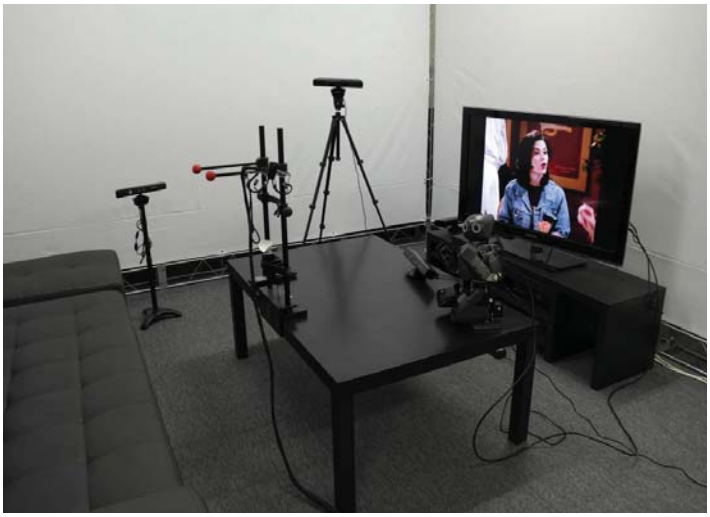
현재 컨텐츠 추천에 주로 사용되는 협력적 여과(collaborative filtering) 방법은 상품 추천에 주로 사용되는 방법으로 사용자의 과거 구매이력에 근거하여 새로운 상품을 추천한다[2][3]. 하지만 이 방법은 사용자의 성향과 관심도가 급격히 변화하는 텔레비전 환경에서는 적합하지 않다. 이에 본 논문에서는 사용자의 상태를 반영한 맞춤형 컨텐츠 추천 방법을 제시한다. 제시하는 방법은 실시간으로 측정한 사용자의 생체 정보를 바탕으로 사용자의 상태를 모델링하고, 동시에 컨텐츠의 내용을 분석하여 사용자의 상태와 컨텐츠의 상관관계를 학습한다. 이를 바탕으로 사용자의 상태를 기반한 컨텐츠의 추천이 가능하다.

본 논문에서는 이에 대한 초기 연구 결과로서, 실험을 위한 통합 플랫폼을 제안하고, 이 플랫폼에서 얻은 실험 데이터에 대해서 논의한다.

2. 통합 플랫폼의 개요

본 논문에서 제시하는 통합 플랫폼은 크게 사용자 모델링 부분과 컨텐츠 모델링 부분으로 나눌 수 있다. 사용자 모델링 부분은 사용자로부터 측정되는 다양한 센서 데이터를 바탕으로 사용자의 감정과 집중도를 분석한다. 한편 컨텐츠 모델링은 비디오의 이미지 및 사운드 스트림에서 등장인물의 얼굴을 인식하고 화자를 인식하여 내용을 분석한다[그림 1]. 이렇게 분석된 사용자 및 컨텐츠 정보는 상관관계 학습 과정을 거쳐 맞춤형 추천 서비스에 사용될 수 있다.

[그림 1] 통합 플랫폼의 개요. 통합 플랫폼은 사용자 모델링과 컨텐츠 모델링 부분으로 나뉘어진다.



[그림 2] 실제로 구축한 통합 플랫폼의 모습. 최대한 거실환경과 비슷하게 설계되어 있다.

2.1 실험 스튜디오 설계

우리는 실험 스튜디오를 설치하여 제안하는 플랫폼을 구축하였다. 스튜디오는 최대한 거실 환경과 같은 분위기로 설계되었고 실제 설치한 스튜디오는 [그림 2]와 같다.

2.2 가상 TV 프로그램

사용자가 실제 TV를 볼 경우, 하드웨어의 문제로 채널의 변화를 기록하기가 어렵다. 이러한 이유로 본 플랫폼은 가상 TV 프로그램을 만들어 실험에 사용하였다. 실제로 방영되는 채널 4개 (KBS1, KBS2, SBS, MBC)의 컨텐츠를 3시간 분량씩 확보하였고 별도로 미국 드라마 5편 (CSI, ER, Friends, Bigbang theory, Grey's anatomy) 또한 3시간 분량씩 확보하여 5개의 드라마 채널을 구성하였다. 즉, 가상 TV은 3시간 분량의 9개 채널을 보유하고 있으며, 3시간동안의 채널 변화는 모두 기록된다.

구현한 가상 TV는 채널 변경 신호가 들어오면 재생하던 동영상 파일을 멈추고 원하는 동영상 파일을 특정 시점부터 재생하는 방식으로 동작한다. 추후 분석을 위해 채널 변경 요청이 들어올 때마다 채널 정보와 시간을 파일로 기록하였다. Windows 환경에서 메시지 후킹을 통해 키보드 이벤트를 받았으며, 동영상 재생은 ffplay 미디어 플레이어를 사용하였고, 코드는 python 스크립트 언어로 작성하였다.

3. 세부 모듈

3.1 사용자 모델링

사용자 모델링을 위한 센서 데이터는 4가지가 사용되었는데(그림 1), 모든 센서 장비는 비부착형이기 때문에 사용자가 크게 불편을 느끼지 않고 실험에 참가할 수 있다. 또한 eye-tracker를 제외한 나머지 장비는 시중에서 쉽게 구할 수 있는 저가형 장비이기로 구성하여 실제 가정의 거실 환경에서 활용될 수 있도록 설계하였다.

3.1.1 Eye-tracker

Eye-tracker는 사용자의 시선을 추적하기 위한 장치로서 사용자가 TV 상에서 주시하고 있는 위치를 좌표로 알아낼 수 있다. Eye-tracker는 컨텐츠 모델링 결과와 결합하여 중요한 보를 제공하는데, '사용자가 자막과 영상 중에서 어디를 많이 보는가?', '주로 어떤 등장인물 또는 장면에 집중하는가?'와 같이 동영상과 관련된 중요한 정보를 알아볼 수 있다. 또한 Eye-tracker는 사용자의 집중도 분석에 많이 사용된다. 동공의 크기, 일정 시간 내에 시선의 고정이 일어난 횟수 등을 따져서 사용자의 집중도를 분석할 수 있는데, 주로 정적인 화면에 일정 시간 노출이 되는 실험에서 많이 사용되었다. 우리는 선행연구[4]를 통하여 동적인 비디오 환경에서 집중도를 계산하는 방법을 연구하였고, 이 연구 결과를 본 플랫폼에 적용하여, TV를 연속적으로 시청하는 상황에서 사용자의 집중도의 변화도를 시간에 따라 나타내었다. 본 플랫폼에는 Arrington Research의 BCU902 장비를 이용하여 90Hz로 사용자의 안구 운동을 추적하였다.

3.1.2 Face expression recognition

사용자의 표정은 감정을 나타내는 가장 명확한 지표이다 [5]. 우리는 일반적으로 많이 사용하는 웹캠을 이용하여 사용자의 얼굴을 촬영하였고 이를 오픈소스 프로그램인 CERT를 이용하여 분석하였다. CERT는 사용자의 얼굴을 인식하여 7가지 감정(화남, 경멸, 혐오감, 공포, 행복, 슬픔, 놀람)에 대한 강도를 계산하여 제공한다.

3.1.3 피부 전위 및 심전도 측정

사용자의 감정과 집중도를 알아보기 위한 또 하나의 측정 장비로 Affectiva사의 Q-Sensor를 사용하였다. 이 장비는 사용자의 피부 전위와 온도, 심전도를 측정하여 놀람, 흥분과 같은 기본 감정에 대한 강도를 제공하며, 집중도도 계산하여 제공한다. 이 장비는 손목에 부착할 수 있을 정도로 크기가 작고 블루투스로 데이터를 전송하기 때문에 실험에서 사용하기에 용이하다.

3.1.4 Motion tracking

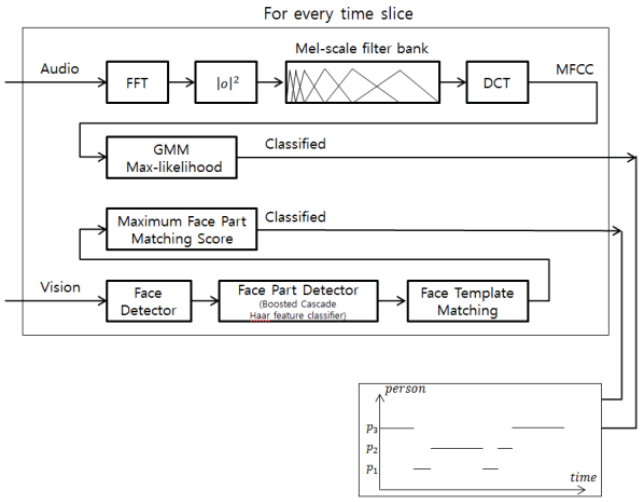
마지막으로 사용자의 자세와 움직임을 추적하기 위해 Kinect를 사용하였다. Kinect는 사용자의 골격을 14개의 관절로 표현하고, 하나의 관절은 6개의 각도 값으로 표현된다. 이 정보는 TV를 시청하고 있는 사용자의 자세를 알아보는데 사용될 뿐만 아니라 심한 머리 움직임으로 eye-tracker의 데이터에 대한 신뢰도가 떨어지는 것을 보정해 줄 수도 있다.

3.2 컨텐츠 모델링

본 플랫폼에서는 컨텐츠 모델링으로 비디오 등장인물 분

석 기법을 사용하였다. 우리는 선행연구[6]을 통하여, 비디오 스트림에서 등장인물의 얼굴을 인식하는 방법과 사운드를 분석하여 화자를 분석하는 방법에 대해서 연구하였으며, 이 방법을 개선하여 본 플랫폼에 적용하였다.

본 플랫폼에서는 얼굴인식을 위하여 face descriptor를 추출한 후[7], 이를 multi-class SVM을 이용하여 분류하였다. 또한 화자 인식은 사운드에서 MFCC (Mel-Frequency Cepstral Coefficients) 특성값을 추출한 후, GMM (Gaussian Mixture Model)을 최대 우도 방법으로 학습하는 방법을 사용하였다. 학습 결과, 컨텐츠 모델링 모듈은 30ms 주기로 얼굴 인식한 결과와 화자 인식 결과를 multinomial 변수로 나타낸다.



[그림 3] 비디오의 소리 및 이미지를 동시에 분석함으로서 화자와 등장인물을 인식하고, 이를 바탕으로 컨텐츠의 내용을 분석하는 모듈.

4. 결론

본 논문에서 제시하는 플랫폼에서 얻을 수 있는 데이터의 종류와 형식은 [그림 4]과 같다. 이 데이터는 13개 종류의 센서 정보를 30ms 주기로 기록한 시계열 데이터이다. 우리는 본 플랫폼을 설계하고 위와 같은 데이터를 생성하는 작업 자체가 사용자 맞춤형 컨텐츠 추천을 위한 알고리즘 개발에 중요한 역할을 할 수 있을 것이라고 기대하며, 이 데이터를 웹을 통해 공개할 예정이다. 또한 차기 연구로서 위의 데이터를 효과적으로 분석하는 알고리즘과 이를 바탕으로 컨텐츠를 추천하는 알고리즘에 대하여 연구를 진행하고 있다.

감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2012-0005643, Videome), 교육과학기술부의 BK21-IT 프로그램에서 일부 지원되었음.

Source(Device)	Sensor Information	Data type
Eye tracker	Gaze position	(x,y) coordinate
	Attention	Real value, [0,1]
Face expression & EDA	Angry	Real value, [0,1]
	Contempt	Real value, [0,1]
	Disgust	Real value, [0,1]
	Fear	Real value, [0,1]
	Happiness	Real value, [0,1]
	Sadness	Real value, [0,1]
	Surprise	Real value, [0,1]
Kinect	Joint Angles	Real value, [0,180]
TV	Change of Channel	Integer, [1,9]
Video Stream(Image)	Character	Integer, [1,7]
Video Stream(Sound)	Speaker	Integer, [1,7]

[그림 4] 통합 플랫폼에서 기록되는 데이터의 종류와 형식

참고문헌

[1] J. E. Short, R. E. Bohn, C. Baru, How Much Information?, *Enterprise Server Information*, 2011, January.

[2] R. Bambini, P. Cremonesi, R. Turrin, A Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environment, *Recommender Systems Handbook*, 2011, pp.299-331

[3] 김은주, 송원문, 송성렬, 김명원, IPTV서비스를 위한 효율적인 협력적 추천 기법, *정보과학회논문지: 소프트웨어 및 응용*, 제 39권, 제 5호, pp.390-398

[4] 김은솔, 김지섭, 장병탁, Eye-Tracker를 이용한 사용자 집중도 분석, *한국정보과학회 가을학술발표 논문집*, 제38권 2(B), 2011, pp. 353-356.

[5] Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J, and Bartlett M, The Computer Expression Recognition Toolbox (CERT). *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2011, pp. 298-305.

[6] 김지섭, 김병희, 장병탁, 인물 중심 TV 드라마 프로파일링에 의한 소셜 네트워크 분석, *한국정보과학회 가을학술발표 논문집*, 제38권 2(B), 2011, pp. 291-294.

[7] Face Recognition Matlab Code: http://www.robots.ox.ac.uk/~vgg/research/nface/