

사용자 선호도 자동 학습 방법을 이용한 개인용 전자 프로그램 가이드 어플리케이션 개발

임정연*, 정현*, 강상길**, 김문철*, 강경옥***
한국정보통신대학교(ICU)*, 수원대학교**, 한국전자통신연구원(ETRI)***
{jylim, hjeong, mkim}@icu.ac.kr, sgkang@suwon.ac.kr, kokang@etri.re.kr

Personalized EPG Application using Automatic User Preference Learning Method

J Lim*, H Jeong*, M Kim*, S Kang**, K Kang***
Information and Communications University (ICU)*, Suwon University**,
Electronics and Telecommunications Research Institute***

요약

디지털 방송의 시작과 함께, 지상파, 위성, 케이블과 같은 다양한 매체를 통한 다채널 방송 시청 환경의 도래는 사용자에게 많은 방송 프로그램 시청 정보를 전달하게 되었다. 이와 더불어, 방송 단말에 전송된 다양한 방송 프로그램 정보를 탐색하고 선호 방송 프로그램을 선별하기 위해서는 사용자에게 많은 노력이 요구된다. 이러한 요구에 따라, 본 논문에서는 다채널 방송 시청 환경 하에서 사용자의 방송 프로그램 시청 히스토리를 분석하고, 특정 시간에 따른 사용자의 방송 프로그램 시청 패턴을 추출하여 방송 프로그램 장르에 대한 사용자 선호도를 자동으로 계산하는 알고리즘을 제안하고, MPEG-7 MDS 구조에 따른 사용자 선호도 서술과 사용자의 선호도에 따라 방송 프로그램을 자동적으로 추천하는 TV 프로그램 추천 어플리케이션을 소개한다. 본 실험을 위해 실제 연령대별, 성별, 시간대별로 사용자의 TV 시청 자료를 사용하였으며, 실험결과를 통해 본 논문에 제안된 베이지안 네트워크 기반 사용자 자동 학습 알고리즘이 효과적으로 사용자 선호도를 학습할 수 있음을 확인하였다.

1. 서론

디지털 방송 환경의 도래는 TV 시청자에게 더 많은 채널을 통한 다양한 프로그램 시청 및 관련 프로그램 정보와 같은 다양한 부가 서비스를 제공을 가능하게 하였다. 이러한 방송 환경의 시작은 시청자에게 더 많은 프로그램과 자세한 정보를 제공하지만, 사용자 구미에 맞는 적절한 프로그램을 선별해야 하는 노력이 요구된다. 따라서 다양하고 복잡한 방송 프로그램 정보들 중에서, 사용자가 원하는 정보 및 방송 프로그램을 탐색하고 자동적으로 선택해야 하며 정해진 방송사의 스케줄에 맞춘 수동적 환경에서 벗어나 사용자가 원하는 콘텐츠를 언제 어디서나 자동적으로 제공할 수 있는 개인화 서비스가 요구된다.

본 논문에서는 사용자 선호도 및 콘텐츠 정보를 서술하기 위하여 MPEG-7 사용자 선호도 서술 구조 (UserPreference DS)와 사용자 시청 히스토리를 저장하기 위한 MPEG-7 사용 히스토리 프로그램 서술 구조(Usage History DS)를 이용한다. 프로그램 자동 추천을 위한 장르에 대한 사용자 선호도를 계산하기 위하여 베이지안 네트워크 기반 가중치 지도 학습 방법(Weighted supervised learning method)을 이용한 사용자 선호도 자동 학습 알고리즘을 제안한다. 마지막으로 제안된 사용자 선호도 학습 알고리즘을 개인화된 전자 프로그램 가이드(Personalized EPG)와 같은 TV 프로그램 추천 시스템에 적용하고 그 결과를 분석한다[1][2][3].

논문의 2장에서는 MPEG-7 멀티미디어 서술 구조에서 정의 하고 있는 사용자 선호도 서술 구조 및 콘텐츠 사용에 대한 기록 모델인 사용 히스토리 서술 구조에 대해 설명하고 3장에서는 베이지안 네트워크 기반의 사용자 선호도 학습 알고리즘에 대해 설명한다. 4장에서는 실험에서 사용될 MPEG-7 사용자 선호도 서술 구조 및 사용 히스토리 서술 구조에 기반한 사용 히스토리와 사용자 선호도 서술 데이터를 제시하고, 제안된 사용자 선호도 학습 알고리즘을 통한 사용자 선호도 예측 정확도 및 제안된 알고리즘에 기반한 TV 프로그램 추천 시스템 적용에 대한 실험 결과를 제시한다. 마지막으로 결론을 제시한다.

2. 사용자 선호도 모델

MPEG-7 MDS(Multimedia Description Scheme: ISO/IEC 15938-5)는 멀티미디어 콘텐츠의 내용, 시공간 및 콘텐츠 종합 정보 등을 구조적으로 표현 할 수 있는 스키마 규격에 대한 표준을

정의 하고 있다. 사용자의 콘텐츠에 대한 선호도 정보 및 콘텐츠 사용 기록 정보 등을 서술할 수 있도록 사용자 취향에 맞는 방송 프로그램을 필터링 하거나, 시청자가 방송 단말에서 접근할 수 있는 모든 방송 프로그램 채널을 실시간 검색하고 선택된 프로그램을 브라우징할 수 있는 사용자 선호도 서술 구조를 제공하고, 시청자가 소비한 콘텐츠의 사용 히스토리를 서술하기 위한 구조화된 사용 히스토리 서술 구조를 정의한다.

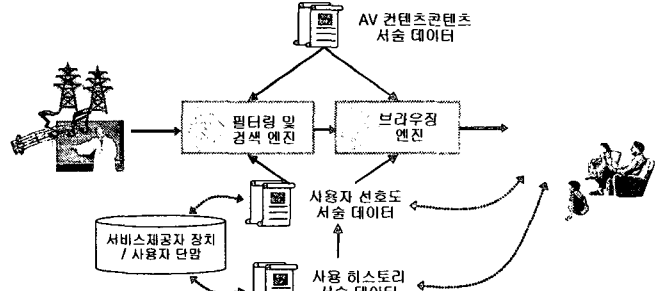


그림 1 필터링, 검색 및 브라우징을 이용한 응용 프로토타입

사용자가 콘텐츠를 소비할 경우 콘텐츠 사용 히스토리 메타데이터는 PC나 TV 셋톱 박스와 같은 TV 단말에 저장된다[2][3][4]. 그림 1은 사용자 선호도 서술 데이터와 사용 히스토리 기술 정보를 이용하여 콘텐츠 필터링, 검색, 브라우징 어플리케이션을 보인다. 사용자 선호도 정보는 맞춤형 방송 서비스로 응용 가능하며, 사용 히스토리 정보는 시간에 따라 변할 수 있는 사용자 선호도 정보를 자동 계산하는데 입력 데이터로 사용될 수 있다[2][3][4].

2.1 사용자 선호도 서술 구조

사용자 선호도 서술 데이터는 소비자 선호도에 따른 TV 프로그램 콘텐츠에 접근하고 이를 소비하는데 쓰일 수 있다. 그림 2는 MPEG-7 사용자 선호도 서술 구조의 계층적 구조를 나타낸다. MPEG-7 사용자 선호도 서술 구조는 그림 2에서와 같이 필터링 및 검색 서술 구조(Filtering and Search Preferences DS), 브라우징 선호도 서술 구조(Browsing Preferences DS) 그리고 사용자 선호도 서술 데이터를 위해 사용자를 구별하기 위한 사용자 식별자 (User Identifier)를 포함한다.

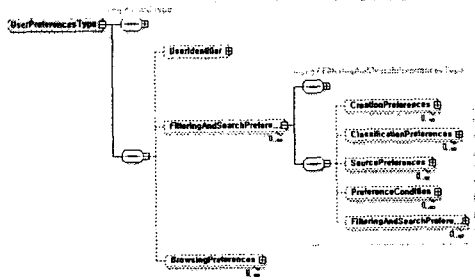


그림 2 MPEG-7 사용자 선호도 서술의 계층적 구조

필터링 및 검색 서술구조는 CreationPreference, ClassificationPreference 그리고 SourcePreference를 포함한다. 브라우징 선호도 서술구조는 SummaryPreferenceType과 PreferenceCondition을 포함하며 SummaryPreferenceType은 SummaryTheme과 SummaryType을 기술한다[5].

2.2 사용 히스토리 서술 구조

MPEG-7 사용 히스토리 서술 구조는 특정한 시간에 소비자의 수행한 콘텐츠 소비형태가 반영된 것이다.

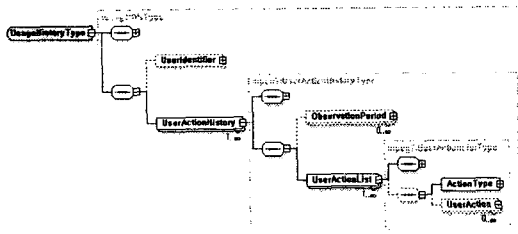


그림 3 MPEG-7 사용 히스토리 서술의 계층적 구조

사용 히스토리 서술 구조의 계층적 구조 MPEG-7 사용 히스토리 서술 구조는 다양한 콘텐츠의 사용 내역을 서술하기 위한 선택스와 시멘틱스를 기술하는데, 여기서 사용 내역이라 함은 사용자의 사용 행동 혹은 행동 타입, 즉, 저장, 실행, 정지, 빨리 찾기, 역으로 찾기 등을 포함한다. 그림 3은 MPEG-7 사용 히스토리 서술 구조의 계층적 구조를 보인다[5]

3. 사용자 선호도 학습 모델

사용자 선호도 모델은 사용자가 시청한 프로그램의 사용 히스토리를 분석함으로써 사용자의 선호도를 예측하고 이를 기반으로 사용자 선호 방송 콘텐츠를 자동 추천하는데 사용된다. 사용자 선호도의 자동 계산을 위해, 임의의 변수들 간에 관계를 표현하는 베이지안 네트워크(Bayesian Network)를 이용하여 사용 히스토리에서 결과에 영향을 미칠 수 있는 변수들(노드) 간 인과관계를 표현(예지)하고, 이들 임의의 변수에 보다 중요도가 높은 베이지안 네트워크 상의 노드의 특성 간에 가중치를 주기 위한 인과관계에 의한 가중치 구조(weighting scheme)를 적용하며, 각 특성에 할당된 각각의 적절한 가중치 값을 알아내기 위해 지도 학습 방법(supervised learning method)을 사용한다.

3.1 가중치 구조(weighting scheme)을 이용한 베이지안 네트워크(Bayesian Network)

베이지안 네트워크는 변수의 인과관계를 나타내기 위한 비주기의 시각적 표현 방법이며, 그림 4과 같이 변수들과 직접적으로 연관된 변수의 집합으로부터 임의의 변수에 조건부 확률 분포를 나타낸다. 베이지안 네트워크는 $B = \langle G, X \rangle$ 로 나타내지며, G 는 각 노드를 연결하는 그래프이고 X 는 B 에 임의의 변수의 집합이다[6].

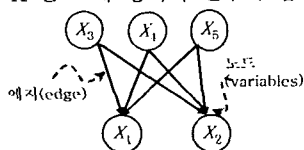


그림 4 베이지안 네트워크(Bayesian Network)

임의의 벡터 $X = [X_1, X_2, \dots, X_M]$ 는 n_i 개의 가능한 특성 $X_i = [x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in_i}]$ 을 가지며, 변수 X_i 가 또 다른 변수

X_j 와 연결되어 있으면, X_i 는 X_j 의 부모노드 혹은 X_j 는 X_i 의 자식노드라 한다. X_i 의 부모 노드는 $X_p \subseteq X \setminus \{X_i\}$ 로 표현될 수 있으며, 베이지안 네트워크 B 와 에 X_p 기반한 변수 X_i 의 상태 x_{ij} 에 조건부 확률 $p(X_i = x_{ij} | X_p, B)$ 은 아래와 같이 표현될 수 있다.

$$p(X_i = x_{ij} | X_p, B) = \frac{n_{ij}}{N_i} \equiv \theta_{x_{ij}|X_p} \quad (1)$$

여기서 N_i 는 변수 X_i 에 샘플의 총 수이고 n_{ij} 는 한 상태 x_{ij} 의 샘플 수를 가리킨다. $\theta_{x_{ij}|X_p}$ 는 조건부 확률 파라미터이다. 통계적 예측 모델인 베이지안 네트워크 상에서 조건부 확률 분포는 콘텐츠의 사용 히스토리로부터 예측될 수 있다. 따라서 제안된 TV 프로그램 장르에 대한 사용자 선호도 계산 방식에서는 요일과 시간에 따른 장르의 선호도를 결정한다. 즉, 요일과 시간은 장르의 부모 노드가 된다. 장르의 선호도를 결정하기 위해, 식 (1)에서와 같이, 조건부 확률은 모든 콘텐츠에 변수 X_i 가 가진 모든 상태에서 어떤 상태 x_{ij} 의 상대적 분포로 계산된다. 하지만 모든 특성에 동등하게 고려되는 이 방법은 사용자의 선호도가 시간이 지남에 따른 변화에 능동적이지 못하다. 일반적으로 사용자는 특정 시간과 요일에 따라 이전에 시청한 프로그램을 이어서 보는 경향이 크며, 이러한 사용자의 경향이 최근의 과거 데이터에서 바뀌기 시작할 경우, 조건부 확률을 이용하여 예측된 결과는 단지 사용 히스토리 내역 중 가장 많은 분포를 차지하는 특성에 치우치게 된다. 또한 사용자의 사용 히스토리 정보의 양에 따라 사용자의 의도와 다르게 사용자 선호도가 예측될 수 있다. 따라서 본 논문에는 최근의 사용 히스토리를 기준으로 하는 가중치 구조를 적용한 사용자 선호도 예측 방식을 제안한다.

사용 히스토리 데이터에 가중치 구조를 적용하기 위해 입력된 데이터는 이전 선호도 영역, 현재 선호도 영역 그리고 미래 선호도 영역으로 구분한다. 각 영역의 변수로, $X_{i,t}$ 는 이전 선호도 영역의 i 번째 노드의 변수를 의미하며, $X_{i,c}$ 는 현재 선호도 영역의 i 번째 노드의 변수를 가리킨다. 여기서 이전 선호도 영역은 시간 간격 L 상에서, 겹치지 않는 M 개의 서브 윈도우 $w_1(s), w_2(s), \dots, w_M(s)$ 로 나뉜다. 여기서 s 는 현재 선호도 영역의 위치를 의미한다. 일반적으로 최근의 윈도우에 특성은 미래의 선호도를 예측하는데 가장 유력한 정보가 될 수 있다. 따라서 제안한 사용자 선호도 학습 알고리즘은 현재 선호도 영역의 사용 히스토리 데이터를 목표 데이터로 두고 이전 선호도 영역의 사용 히스토리 데이터를 훈련 데이터로 사용된다. 미래 선호도 영역에 데이터는 테스트 데이터로 이용된다. 이전 선호도 영역의 각 윈도우 w_1, w_2, \dots, w_M 에서 콘텐츠의 빈도는 가중치 w_1, w_2, \dots, w_M 를 갖게 된다. 가중치에 의한 식 (1)의 표현은 아래와 같다.

$$\hat{\theta}_{x_{ij}|X_p} = p(X_i = x_{ij} | X_p, B, w) = \frac{\sum_{m=1}^M w_m n_{ij,m}}{\sum_{m=1}^M w_m N_i} \quad (2)$$

최적의 가중치는 상호정보량(mutual information: MI)이 최대가 되는 관점에서 얻어질 수 있다[7].

3.2 사용자 선호도 계산을 위한 지도 학습 알고리즘

제안된 알고리즘은 최적성 측도의 최대화 관점에서 가중치의 최적화된 집합을 찾는 것이다. 그림 5는 데이터 가중치 구조를 갖는 사용자 선호도 지도 학습 방법(supervised user preference learning method)을 나타낸다.

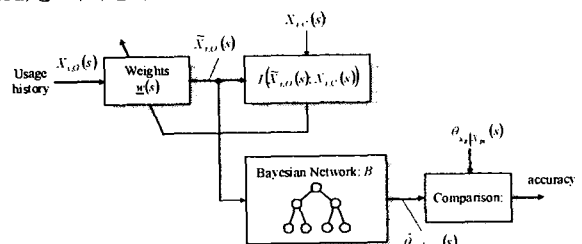


그림 5 가중치 기반 사용자 선호도 지도 학습 방법

최적의 가중치를 얻기 위한 이전 선호도 데이터와 현재 선호도 데이터 간의 상호정보량은 다음과 같이 정의된다.

$$I(\tilde{X}_{i,o}(s), X_{i,c}(s)) = \log \left(\frac{p(X_{i,c}(s) | \tilde{X}_{i,o}(s))}{p(X_{i,c}(s))} \right) \\ = \log(p(X_{i,c}(s) | \tilde{X}_{i,o}(s))) - \log(p(X_{i,c}(s))) \quad (3)$$

최적성의 측도로 쓰이는 $X_{i,c}(s)$ 과 $\tilde{X}_{i,o}(s)$ 간의 상호정보량 $I(\tilde{X}_{i,o}(s), X_{i,c}(s))$ 는 $X_{i,c}(s)$ 과 $\tilde{X}_{i,o}(s)$ 의 상관관계로 표현된다. 우리의 제안된 학습 방법은 상호정보량(MI)이 최대화되는 점에서 이전 선택도 영역의 윈도우 가중치에 최적 값을 결정하도록 한다. 윈도우 가중치의 적용 구조는 그림 5에서와 같다. 윈도우가 움직임에 따라, 기울기 증가(gradient ascent) 알고리즘을 기반으로 가중치 갱신(weight update)이 이루어지며, 가중치 갱신은 상호정보량이 최대치가 될 때까지 지속된다. 상호정보량에 조건부 확률 $p(X_{i,c}(s) | \tilde{X}_{i,o}(s))$ 은 식 (4)에서와 같이 이전 선택도 영역의 분할된 데이터 집합의 빈도에 가중치를 부과함으로써 계산된다. 또한 한 노드의 각 특성에 대한 조건부 확률은 서로 독립이므로, 조건부 확률 $p(X_{i,c}(s) | \tilde{X}_{i,o}(s))$ 은 $X_{i,c}(s)$ 에 모든 요소의 조건부 확률이 곱이 된다. 여기에서 $n_x(s)$ 는 노드 $X_{i,c}(s)$ 에 특성 $x_y(s)$ 의 개수가 된다.

$$p(X_{i,c}(s) | \tilde{X}_{i,o}(s)) = \prod_j p(x_{y_j}(s) | \tilde{X}_{i,o}(s))^{n_{y_j}(s)} \\ = \prod_j \left(\frac{\sum_m w_m(s) n_{j,m}(s)}{\sum_m w_m(s) N_{j,m}(s)} \right)^{n_{y_j}(s)} \quad (4)$$

최적의 가중치 집합을 얻기 위해, 우리는 가중치 $w_m(s)$ 의 관점에서 상호정보량에 편미분도함수를 계산한다. 상호정보량(MI)을 얻기 위해 식 (3)을 이용한 상호정보량(MI)의 편미분결과는 아래와 같다.

$$\frac{\partial I(\tilde{X}_{i,o}(s), X_{i,c}(s))}{\partial w_m(s)} = \frac{\partial \log(p(X_{i,c}(s) | \tilde{X}_{i,o}(s)))}{\partial w_m(s)} \\ = \sum_j n_{y_j}(s) \left(\frac{n_{j,m}(s)}{\sum_m w_m(s) n_{j,m}(s)} - \frac{N_{j,m}(s)}{\sum_m w_m(s) N_{j,m}(s)} \right) \quad (5)$$

가중치 갱신 텀인 $\Delta w_m(s)$ 는 학습률 매개변수 η 와 함께 아래와 같이 표현될 수 있다.

$$\Delta w_m(e, s) = \eta \frac{\partial I(\tilde{X}_{i,o}(s), X_{i,c}(s))}{\partial w_m(s)} \quad (5)$$

위와 같이, 가중치 갱신은 상호정보량(MI)를 최대로 하는 기울기 증가 방법을 기반으로 이루어진다. 매 시점 e 에 가중치 갱신치는 다음과 같이 표현될 수 있다.

$$w_m(e+1, s) = w_m(e, s) + \Delta w_m(e, s) \quad (7)$$

e 는 학습반복 수를 나타내며 η 는 기울기상승 과정 동안 가중치 폭의 스텝(searching step) 정도를 정의하는 기울기 증가의 학습률 매개변수이다. 훈련 과정은 상호정보량(MI)가 더 이상 증가하지 않을 때 멈추게 된다[8].

4. 실험 및 결과

본 장에서는, 사용자의 행동에 따른 사용 히스토리와 사용자 선호도 기술 데이터의 표현 방법과 2002년 12월부터 2003년 5월까지 시청된 TV 프로그램 소비 히스토리를 이용한 실험 결과를 보이며, 우리의 지도 학습 방법을 이용한 디지털 TV 프로그램용 장르 추천 시스템 구현에 대해 설명한다. 시청자료는 시청 시간, 프로그램 시간, 세 레벨로 나뉜 프로그램 장르 그리고 사용자의 프로필 등으로 구성되어 있다. 이 중 우리는 10대 여자, 30대 여자 그리고 50대 남자의 데이터를 가지고 실험하였다. 본 TV 프로그램 소비 히스토리 데이터는 AC Nielson Korea로부터 제공되었다.

4.1 사용 히스토리와 사용자 선호도 기술 데이터
사용자 선호도를 예측하기 위한 사용 히스토리는 그림 6에서와 같이 MPEG-7 사용 히스토리 서술 구조의 형태로 표현된다.

```
<Description xsi:type="UserDescriptionType">
  <UsageHistory id="uh-002" allowCollection="true">
    <UserIdentifier protected="true">
      <Name xml:lang="en">Person A</Name></UserIdentifier>
    <UserActionHistory id="uh-001" protected="false">
      <ObservationPeriod>
        <TimePoint>2003-04-12T20:00-21:00</TimePoint>
        <Duration>PT1H</Duration></ObservationPeriod>
      </UserActionHistory>
    </UsageHistory>
  </Description>
```

```
<UserActionList id="ua-list-001" numOfInstances="1">
  <ActionType href="urn:mpeg:mpeg7:cs:ActionTypeCS:2001:1.2">
    <Name>PlayStream</Name></ActionType>
    <UserAction><ActionTime>
      <GeneralTime>
        <TimePoint>2003-04-12T20:16:08+09:00</TimePoint>
        <Duration>PT12M</Duration>
      </GeneralTime></ActionTime>
    </UserAction>
  </UserActionList>
  <ProgramIdentifier>dm-20030412</ProgramIdentifier></ActionDataItem/>
```

그림 6 사람 A의 방송 콘텐츠 사용 히스토리 서술 데이터

그림 6은 2003년 4월 12일, 오후 8시~9시 사이에 사람 A의 방송 콘텐츠 사용 히스토리 서술 메타데이터를 나타낸다. 한 시간 동안, 사용자 A의 행동 내역은 자신의 아이디 us-list-001이라는 이름으로 사용자 행동 리스트(user action list)에 기록된다. 위의 예는 dm-20030412라는 프로그램의 12분간 시청 내역이 기록되어 있다. 사용자 행동 리스트는 ActionType에 따른 UserActionList와 ActionTime을 포함하고 있는 UserAction을 반복적으로 사용해 기록될 수 있다. 그림 7은 사람 A의 장르에 대한 MPEG-7 사용자 선호도 서술 구조 기반의 사용자 선호도를 나타낸다.

```
<Description xsi:type="UserDescriptionType">
  <UserPreferences>
    <FilteringAndSearchPreferences>
      <ClassificationPreferences>
        <Genre preferenceValue="0"><Name>Education</Name></Genre>
        <Genre preferenceValue="43"><Name>DramaMovie</Name></Genre>
        <Genre preferenceValue="0"><Name>News</Name></Genre>
        <Genre preferenceValue="1"><Name>Sports</Name></Genre>
        <Genre preferenceValue="0"><Name>Children</Name></Genre>
        <Genre preferenceValue="56"><Name>Entertainment</Name></Genre>
        <Genre preferenceValue="0"><Name>Information</Name></Genre>
        <Genre preferenceValue="0"><Name>etc</Name></Genre>
      </ClassificationPreferences>
    </FilteringAndSearchPreferences>
  </UserPreferences>
</Description>
```

그림 7 사람 A의 사용자 선호도 결과

그림 7에서와 같이, 사용자 선호도 서술 메타데이터는 고려된 장르와 그것의 선호도 값으로 구성되어 있다. 오후 8시~9시 사이에 사람 A가 선호하는 장르는 드라마&영화와 오락이며, 각각의 선호도 값은 56과 43을 갖는다. 특정 시간 대에 드라마&영화와 오락 외에 다른 장르의 선호도 값은 해당 프로그램의 히스토리가 거의 없기 때문에 0에 가까움을 알 수 있다. 스포츠의 경우, 사람 A는 이 장르에 해당되는 프로그램을 시청했지만, 사람 A가 이 장르의 프로그램을 최근에 보지 않았기 때문에, 트레이닝 반복의 수가 증가할수록 점점 더 선호도도 떨어져, 장르 시청의 추천율이 매우 작음을 알 수 있다. 다른 장에서는 우리의 지도 학습 방법을 이용한 사용자 선호도 값을 기반으로 사용자가 좋아하는 TV 프로그램의 리스트를 제공하는 추천 시스템을 보인다.

4.2 TV 시청 히스토리를 이용한 사용자 선호도 학습
콘텐츠 장르 기반 사용자 선호 방송 프로그램 자동 추천을 위해 TV 시청 시간 대와 요일을 나누고 실험하였다. 즉, 날짜 변수는 {월, 화, ..., 일}의 상태를 가지며, 시간 변수는 주 시청 대를 고려하여 {6~7, 7~8, ..., 11~12}의 상태를 갖는다. 장르 변수는 {교육, 드라마&영화, 뉴스, 스포츠, 유아, 오락, 정보, 기타}로 나뉜다. 그림 12는 윈도우의 개수가 3인 경우($M=3$, [$w_1(4)$, $w_2(4)$, $w_3(4)$])와 2인 경우($M=2$, [$w_1(5)$, $w_2(5)$]), 가중치 적용 학습 커브를 보인다. 학습률 매개변수 η 는 0.1로 고정시켰으며, 훈련 반복수는 300으로 두었다. 그림 8에서와 같이, $w_1(5)$, $M=2$ 와 $w_2(5)$, $M=2$ 는 각각 $w_2(4)$, $M=3$ 와 $w_1(4)$, $w_3(4)$, $M=3$ 에 비슷한 커브를 갖는다. 트레이닝 반복의 수가 증가할수록 각각의 가중치는 포화된다.

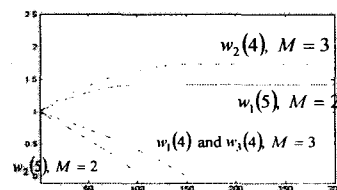


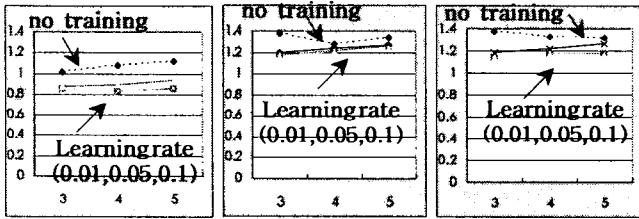
그림 8 반복된 트레이닝에 따른 가중치 적용

4.3 제안된 학습 알고리즘의 정확성 검토
본 장에서는 실제 사용자가 시청한 결과, 즉, 미래 선호도 영역으로 테스트한 결과를 기준으로 가중치 트레이닝 되지 않은 추천 결

과와 우리의 추천 시스템을 통해 예측된 결과를 비교한다. 예측된 결과의 정확성을 확인하기 위해 추천된 장르의 순서와 실제로 많이 본 프로그램의 장르의 순서로 순위를 매겨 각 순위간의 차이를 계산한다. 각 장르의 순위에는 상수 값을 할당하며 순위에 따른 상수 값 간에는 허용오차를 할당하였다. 정확성 오차는 추천된 순위 벡터와 실제 매겨진 벡터간의 거리(Euclidian distance measure)로 아래와 같이 계산될 수 있다.

$$Error(s) = \left\| \bar{R}(\theta_{x_j|Pa_i}, FPR_i(s)) - \hat{R}(\theta_{x_j|Pa_i}, FPR_i(s)) \right\| \quad (8)$$

여기서 $\bar{R}(\theta_{x_j|Pa_i}, FPR_i(s))$ 는 실제 순위의 벡터를 가리키며, $\hat{R}(\theta_{x_j|Pa_i}, FPR_i(s))$ 는 추천된 순위의 벡터를 의미한다. 그림 9는 사람 A의 트레이닝 방법과 트레이닝을 하지 않은 방법간의 정확성 에러 비교이다.



(a) 사람 A (b) 사람 B (c) 사람 C
그림 9 정확성 에러 결과

그림 9의 (a)는 사람 A의 정확성 결과를 비교하며, (b)는 사람 B의 정확성 결과 그리고 그림 9의 (c)는 사람 C의 정확성 결과를 나타낸다. 각각은 훈련 되지 않은 결과와 각 학습률에 따라 훈련된 에러 값이며 윈도우의 개수에 따라 달라진다. 그림 9에서 보는 바와 같이 트레이닝 되지 않은 결과에 비해 트레이닝된 결과의 에러 값이 더 작음을 알 수 있다. 여자 A(a)의 경우 21.4%로 윈도우의 개수가 5인 경우 가장 많은 향상을 보였으며, 여자 B(b)의 경우, 윈도우 개수가 3인 경우 10.1%의 향상을 보였고, 남자 C(c)는 윈도우 개수가 3인 경우에 임의의 샘플에 조건부 확률값인 훈련되지 않은 결과에 비해 10.9%의 정확성 향상을 보였다. 또한 실제 기록된 시청 프로그램 데이터의 크기의 측면에서, 사람 A에 비해 사람 B와 C의 기록된 사용 히스토리가 많다. 이러한 결과를 볼 때, 사람 A에 비해 사람 B와 사람 C가 비교적 일정한 프로그램을 선호하는 경향이 더 큰 것을 알 수 있으며, 사람 A의 경우, 기록된 사용 히스토리가 많지 않고 어떤 프로그램의 선호 경향이 일정하지 않으므로, 윈도우 수 및 정보량이 증가할수록 정확성은 더 향상되게 된다. 즉, 더욱 정확한 결과를 얻기 위해, 사람 A를 위한 선호도 알고리즘은 더 많은 이전 선호도 데이터를 요구하게 된다.

4.4 제안된 알고리즘을 이용한 개인화 전자 프로그램 가이드
본 장에서는 위에서 제안된 자동적 사용자 선호도 학습 알고리즘을 기반으로 구현된 개인화 전자 프로그램 가이드(personalized EPG) 응용을 소개한다. 사용자가 TV를 켰을 때, 개인화 EPG 정보는 미리 계산된 사용자의 장르 선호도에 따라 TV를 켜는 날짜와 시간에 맞게 추천된 프로그램 리스트를 보일 수 있다.

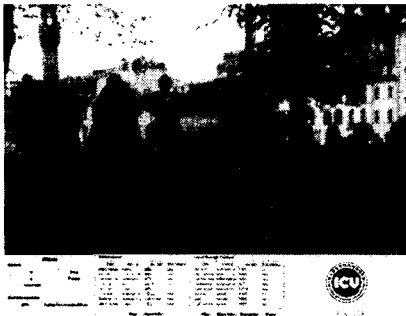


그림 10 개인화 전자 프로그램 가이드 시스템의 GUI

그림 10은 제안된 베이지안 네트워크 기반 사용자 선호도 자동 학습 알고리즘에 의한 개인화 전자 프로그램 가이드의 사용자 인터페이스를 나타낸다. 처음 사용자는 자신의 TV 단말에 저장된 모든 콘텐츠와 현재 방송되고 있는 모든 목록을 볼 수 있으며, 개인화 전자 프로그램 가이드를 선택하게 되면, 추천된 사용자 리스트가 나타난다.

Title	Genre	Channel	Recommended
Meteor Garden	Drama-Movie	catchone	yes
Drama 2	Drama-Movie	MBC	yes
Vincent	Entertainment	KBS2	yes
my heart will go on	Entertainment	MTV	yes
Vincent	Entertainment	KBS2	yes
jally mcbcal	Drama-Movie	MBC	yes
my heart will go on	Entertainment	MTV	yes
dark angel	Drama-Movie	OCN	yes

(a) 제안된 선호도 알고리즘에 의한 사람 A의 추천리스트

Title	Genre	Channel	Recommended
MBC News	News	MBC	yes
Meteor Garden	Drama-Movie	catchone	yes
Drama 2	Drama-Movie	MBC	yes
Vincent	Entertainment	KBS2	yes
my heart will go on	Entertainment	MTV	yes
Soccer	Sports	MBC	yes
Vincent	Entertainment	KBS2	yes
jally mcbcal	Drama-Movie	MBC	yes

(b) 제안된 선호도 알고리즘에 의한 사람 C의 추천리스트
그림 11 개인화 전자 프로그램 가이드

그림 11는 구현된 개인화 전자 프로그램 가이드 시스템이며 10대 여자 A와 50대 남자 C가 로그인 했을 경우, 그림 10의 아래와 같은 리스트가 보이게 된다. 왼쪽 리스트는 현재 방송중인 콘텐츠의 목록이며, 오른쪽 리스트는 사용자의 방송 단말 (혹은 셋톱박스)에 저장된 콘텐츠의 목록이다. 사용자가 추천된 프로그램 리스트를 선택하게 되면, 그림 11의 (a), (b)와 같은 제안된 사용자 선호도 알고리즘에 의한 사람 A와 C의 추천 리스트가 보이게 된다. 사람 A의 경우, 토요일 오후 8시에 TV를 켰을 때이며, 사람 A의 추천 장르는 드라마&영화 그리고 오락임을 알 수 있다. 사람 C의 경우, 토요일 10시에 TV를 켰을 때이며, 사람 C의 추천 장르는 드라마&영화, 오락, 스포츠 그리고 뉴스임을 알 수 있다. 사람 A와 C의 제안된 사용자 선호도 알고리즘에 의한 추천 프로그램의 리스트를 비교했을 때 사용자 C가 더 많은 추천 프로그램 리스트 목록을 얻었음을 알 수 있다. 이러한 결과로 볼 때, 사람 C의 경우, 사람 A에 비해 더 풍부한 사용 히스토리를 가지고 있으므로 더 많은 추천 결과를 최근의 사용 히스토리만으로 쉽게 얻을 수 있다.

5. 결론

본 논문은 MPEG-7 사용자 선호도 서술 구조 및 사용 히스토리 서술 구조 표현을 기반으로 TV 개인화를 위한 사용자 선호도 자동 학습 방법과 이를 적용한 개인화 전자 프로그램 가이드 어플리케이션을 제안한다. 사용자 선호도 학습을 위해 우리는 가중치를 부여한 베이지안 네트워크에서 출발하였으며, 최적의 가중치를 찾기 위해 이전 사용자 선호도와 현재 사용자 선호도간의 상호정보량(MI)을 이용하였다. 사용자의 선호도는 시간이 지남에 따라 바뀔 수 있으므로, 특정 시간 간격의 사용 히스토리 데이터에 가중치를 두어, 시간에 따른 사용자 선호도 변화에 유연하게 학습하는 방법을 제안하였다. 실험을 통해 사용자의 최근 콘텐츠 소비 형태에 따른 적응적 사용자 선호도 알고리즘과 사용자 선호도 기반 방송 프로그램 정보를 효과적으로 제시함으로써 다채널 방송 환경에서 지능형 방송 프로그램 가이드의 가능성을 제시하였다.

6. 참고문헌

- [1] 류지웅, 김문철, 남제호, 강경옥, 김진웅, "사용자 선호도 기반 지능형 프로그램 가이드," 한국방송공학회 논문지 2002년 제 7권 제 2호, pp153-167.
- [2] Munchurl Kim, Jeongyeon Lim, Kyeongok Kang and Jinwoong Kim, "Agent-based Intelligent Multimedia Broadcasting within MPEG-21 Multimedia Framework," ETRI journal, Vol. 26, No. 2.
- [3] Seongjoon Pak, Geewoong Ryu and Munchurl Kim, "Agent-based Multimedia Personalcasting," Proceedings of the International Workshop on Advanced Image Technology, pp. 311 ~ pp. 316, Nagasaki, Japan.
- [4] Munchurl Kim, Geewoong Ryu, Beetnara Bae, Jeho Nam, Kyeongok Kang and Jinwoong Kim, "Intelligent program guide for digital broadcasting," Proceedings of the IWAIT2002, pp. 257-263, Hualien, Taiwan, R.O.C.
- [5] ISO/IEC JTC1/SC29/WG11, "ISO/IEC 15938-5 FDIS: Multimedia content Description Interface" MPEG2001/N4242, Sydney, Australia
- [6] F.V. Jensen, Bayesian Networks and Decision Graphs, Springer.
- [7] Y.J. Sung, J.H. Hong, T.S. Kim, "A formal model for user preference", Proceedings of IEEE International Conference on Data Mining, pp. 235-242, Maebashi City, Japan.
- [8] Sanggil Kang, Jeongyeon Lim, Munchurl Kim (2004). "Modeling the user preference on broadcasting contents using Bayesian belief network," proceeding of VCIP, San Jose.