

TV Anytime 응용을 위한 사용자 선호도 추출 및 갱신 알고리즘

배빛나라¹, 류지웅², 김문철³, 남제호⁴, 강경옥⁵, 노용만⁶

한국정보통신대학원대학교 공학부^{1,2,3,6}
한국전자통신연구원 방송미디어연구부^{4,5}

User Preference Extraction and Update Algorithm for TV Anytime Applications

Correspondence³: (e-mail) mkim@icu.ac.kr, (phone) 042-866-6137

요 약

사용자에게 적합한 프로그램을 추천하거나 필터링을 수행하는 지능형 방송 단말 응용 소프트웨어 에이전트에서 필수적으로 사용되는 사용자 선호도 (User Preference)를 추출하는 알고리즘을 연구하였다. 시청자 선호도 추출 알고리즘으로는 시청자의 프로그램 장르나 출연 배우 등에 대한 선호도, 프로그램 시청 시간 등에 대한 선호도나 시청 프로그램에 대한 사용자의 인터랙션 습성 분석에 의한 프로그램 선호도 등을 probabilistic framework과 rule-based framework을 근간으로 추출 하는 알고리즘 연구에 대한 결과를 본 논문에서 제시한다.

1. 서 론

최근 TV환경의 변화는 채널 수의 급격한 증가 뿐만 아니라 프로그램 수의 급격한 증가와 콘텐츠의 다양화도 예상되고 있다. [1] 기존의 TV와 같이 단순한 시청 형태가 아닌 사용자의 action을 많이 요구하는 형태로 변화하였다. 한편 대부분의 사용자들은 자신의 기호에 맞는 TV 프로그램을 시청하는 경향이 있다. 이런 사용자의 프로그램 선호도에 대한 기술(description)은 MPEG-7 MDS(Multimedia Description Schemes) 규격과 TV-Anytime 메타데이터 규격에서 지원되고 있다 [2, 3].

MPEG-7 MDS와 TV Anytime 메타데이터 규격의 User preference DS는 다양한 값들로 사용자 선호도를 기술(description) 할 수 있도록 정의하고 있으나,

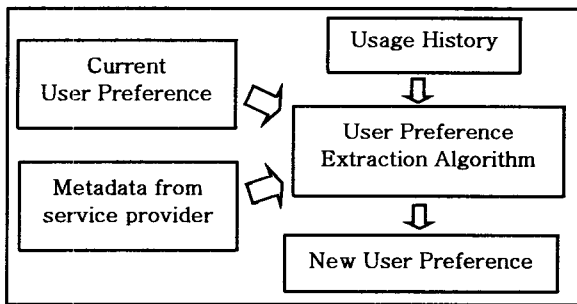
이러한 선호도 값을 추출하는 방법은 표준으로 정의하지 않고 있다. 즉 사용자 선호도에 대한 선택스와 시맨틱스만 표준으로 정의하고 있다. 선호도 값을 추출하는 방법은 표준 규격을 응용한 시스템 개발자의 쪽의 몫으로 남아 있다.

방송 프로그램 (또는 콘텐츠)에 대한 사용자 선호도를 이용한 기존의 방법은 방송 단말에서 사용자가 직접 제한된 사용자 선호도 정보를 직접 설정하고 이를 이용한 프로그램 추천 기능 등이 지원되는 정도에 불과하였다.

그러나 콘텐츠 정보를 풍부히 그리고 상세하게 기술할 수 있는 리치(rich) 메타데이터 규격의 등장과 소프트웨어 에이전트 기술의 등장으로 메타데이터와 소프트웨어 에이전트 기술을 접목하여 방송 환경에 응용한 연구가 시도되었다 [4]. 디지털 방송의 등장과 함께 다매체/다채널 방송 환경의 등장으로 시청자에게 제공되는 방송 채널 및 프로그램의 수가 폭발적으로 증가하게 되었다. 이렇게 증가된 방송 프로그램을 시청자가 일일이 채널을 돌려 가며 원하는 프로그램을 선택 하는 것이 매우 번거롭고 어려운 일이 되었다. 이러한 새로운 정보 선택형 방송 환경에서 리치 방송 메타데이터 기술과 소프트웨어 에이전트 기술을 이용하여 사용자가 원하는 프로그램을 선별/검색하고 추천해 줄 수 있는 사용자 방송 단말 기술의 개발에 대한 필요성이 대두되었다. 본 논문에서는 시청자의 프로그램 장르나 출연 배우 등에 대한 선호도, 프로그램 시청 시간 등에 대한 선호도나 시청 프로그램에 대한 사용자의 인터랙션 (저장, 빨리감기, 되감기, 멈춤, 요약보기 등) 습성 분석에 의한 프로그램 선호도 등에 대한 선호도 값을 자동 추출하는 알고리즘과 사용자의 콘텐츠 소비 성향에 대한 분석을 통해 변화하는 사용자 선호도를 적응적으로 반영할 수 있는 사용자 선호도 갱신 알고리즘에 대한 연구

를 수행 하였다.

선호도 정보의 추출에는 <그림1>과 같이 현재의 사용자 선호도 정보와 사용자의 TV 시스템 이용정보가 사용된다. 사용자 취향 정보 추출에는 statistical framework와 rule-based framework이 사용된다. Statistical framework는 discrete probability density function과 conditional probability를 기초로 한다. 그리고 사용자 선호도는 다양 형태로 나타나기 때문에, 특정한 규칙을 만들어서 사용자 취향 정보를 추출해야 한다. 이런 이유로 statistical framework과 rule-based framework을 이용하였다.



<그림 1> User Preference 값 추출 과정

2. MPEG-7 과 TV-Anytime의 User Preference DS(Description Scheme)

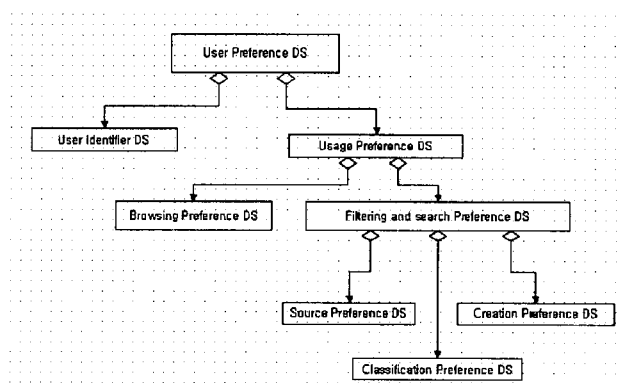
MPEG-7과 TV-Anytime에서 정의된 사용자 선호도 정보의 정의는 동일하다. MPEG-7의 사용자 선호도 기술 구조(User Preference DS)는 사용자를 식별할 수 있는 사용자 식별 기술 구조(User Identifier DS)와 사용자의 선호도 정보를 수록하는 이용 선호도 기술 구조(Usage Preference DS)로 구성된다.

MPEG-7 기술 정의 언어 (DDL: description definition language)로 표현한 사용자 선호도 기술 구조에 대한 스키마 구조는 다음과 같다.

```

<complexType name="UserPreferencesType">
  <complexContent>
    <extension base="mpeg7:DSType">
      <sequence>
        <element name="UserIdentifier"
          type="mpeg7:UserIdentifierType" minOccurs="0"/>
        <element name="UsagePreferences"
          type="mpeg7:UsagePreferencesType"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>
  
```

<그림 2>는 사용자 선호도 기술구조를 UML(Unified Modeling Language) 다이어그램으로 나타내었다.



<그림2> User Preference DS의 구성

2.1. Usage Preference의 구성

Usage Preference는 사용자의 browsing과 관련된 preference의 정의와 Filtering과 search와 관련된 preference의 정의로 구분된다.

2.2. User Preference Value

Browsing Preference DS	FilteringAndSearchPreference DS		
	Creation Preference	Classification Preference	Source Preference
◆SummaryType ◆Preferred ◆SummaryType	◆Title ◆Creator ◆Keyword ◆Location ◆DatePeriod	◆Country ◆DatePeriod ◆Language ◆Genre ◆Subject ◆MediaReview ◆Parental Guidance	◆Publication Type ◆Publication Source ◆Oublication Place ◆Publication Date ◆Publisher ◆MediaFormat

표 1. User Preference Value의 종류

- ◆ Browsing Preference DS
사용자가 TV-System의 이용하면서 발생하는 browsing 정보에 대한 사용자의 선호도를 정의한다.
- ◆ Creation Preference DS
Content 및 TV-Program의 제작에 관련된 정보에 대한 선호도를 정의한다. Creation Preference DS에 해당되는 데이터의 수가 무한대에 가깝다는 특징이 있다.
- ◆ Classification Preference DS
Content 및 TV-Program의 분류기준에 대한 선호도를 정의한다. 분류 기준은 MPEG-7의 classification scheme에 의해 정의된다. 즉, 항상 유한한 데이터의 개수를 유지한다.
- ◆ Source Preference DS
Content 및 TV-Program의 publication에 관련된

선호도를 정의한다. 서비스 공급자에 대한 선호도 기술이 대부분이다.

3. User Preference Value Extraction

3.1. Preference value

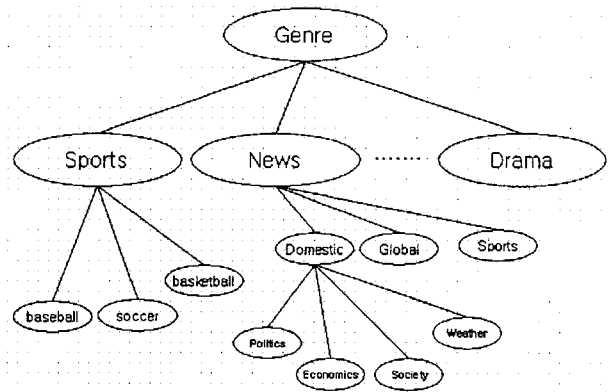
Preference value의 선정은 source preference를 기초로 하였다. Source Preference는 stereo type service에 적용이 용이하다. 통계적 방식의 신뢰성을 확보하기 위하여 데이터의 축적이 요구된다. 데이터의 축적이 충분하지 않은 사용자는 선호도를 직접적으로 입력하도록 요구하거나 default type의 서비스 제공이 필요하다. Default type의 서비스 제공은 현재 시청 가능한 프로그램 또는 content가 된다. 즉, source preference value를 현재 시간을 최대의 선호도로 조절한다.

Classification preference는 classification scheme [2],[3]에 의해 정의되는데, 각 value 응용의 범위가 다양하다. Parental Guidance는 프로그램 혹은 content의 등급별 시청 선호도가 될 수 있겠지만, 사용자의 입력된 연령과 비교하여 사용자의 연령으로 시청할 수 없는 프로그램이나 content를 차단하는 판단기준이 되기도 한다. Media Review는 content의 재 시청의 비율을 정의한다. 각 비율별로 사용자의 선호도가 명확히 표시된다.

Creation preference는 그 데이터의 수가 service provider의 제공과 큰 연관성을 가진다. Creation preference는 program이나 content 별, 또는 program이나 content의 genre별 preference value를 이용하여 간접적으로 추론한다. 데이터의 수의 한계점이 명확하지 않고, 제공되는 방송 정보에 의해 데이터의 변화가 발생하므로 직접적인 preference value의 extraction이 불가능하다. Search와 같은 data와 연관성이 큰 action에 의해서만 preference를 직접적으로 추론할 수 있다.

3.2. Statistical approach

발생 확률이 높은 데이터는 사용자의 선호도가 높은 데이터이다. 대부분의 preference 정보는 계층적인 구조를 지니고 있다. 예를 들면 Genre의 경우도 depth가 4인 tree의 구조를 지니고 있다. 같은 계층의 노드에는 discrete probability density function의 관계가 성립되고, 계층 간의 관계는 conditional probability가 성립된다.



<그림 3> Example of Genre Tree

각 노드는 usage history를 시청 시간과 방문의 빈도수에 따라서 그 발생 확률을 구할 수 있다.

$$P(\text{baseball} | \text{sports}) = \frac{P(\text{baseball} \cap \text{sports})}{P(\text{sports})}$$

$$P(\text{sports}) = P(\text{baseball}) + P(\text{soccer}) + P(\text{basketball})$$

$$P(\text{baseball}) = \frac{\text{baseball_waste_time}}{\text{total_waste_time}}$$

위의 식들을 일반화 시키면,

$$P(\text{Genre}) = \frac{\text{Genre_waste_time}}{\text{total_waste_time}}$$

Eq 1)

$$P(\text{subGenre} | \text{Genre}) = \frac{P(\text{subGenre})}{P(\text{Genre})}$$

Eq 2)

여기서 Genre_waste_time은 Genre에 해당되는 프로그램을 시청한 총 시간을 나타내며, total_waste_time은 전체 시청 시간을 나타낸다.

시청 시간대별 publication source에 대한 선호도는 시청 시간대별 전체 시청 시간의 분포에 속하게 된다. 즉, 특정 시간대의 publication source과 시청 시간의 분포는 conditional probability, publication source 간의 분포는 discrete probability density function이 성립된다.

사용자의 system 이용정보는 usage history에 축적되어 있다. 사용자 선호도는 그 변화의 정도가 크지 않은 특징이 있다. 즉, 사용자의 선호도가 높은 채널이

나 프로그램, 혹은 그 프로그램의 장르 등은 usage history에서 발생하는 빈도가 높게 된다. 이런 특성을 이용하기 위해 statistical method를 선정하게 되었다.

3.3. Rule-based approach

Preference value의 검출에서 정형화된 과정을 거치게 된다. Preference value에 영향을 미치는 요인이 다양하기 때문에 각 요인들 간의 연관성, 특징을 rule 표현하여 적용하였다.

Action에 따른 가중치의 설정과 preference value 추출에서 statistical method의 적용 순서들을 규칙에 따라 적용하였다.

즉, 각 사용자는 system의 사용에 규칙성을 가지고 있다. 이런 규칙성을 이용하면 사용자의 preference를 정확히 추론할 수 있다. 그리고 선호도 값들의 연관성 또한 rule-based method를 결정한 이유이다.

Rule 1) Statistical method의 적용 순서

순서1) 각 요일 별 시청 시간의 분포를 획득한다.

순서2) 각 요일의 시청 시간대 별 채널(publication source)에 대한 이용 분포를 구한다.

순서3) 채널 정보별 시청 프로그램의 장르의 분포를 구한다.

순서4) 각 장르별 프로그램의 시청 빈도를 측정한다.

채널과 각 채널의 특성을 시간대에 맞추어서 추론할 수 있도록 rule 1 을 설정하였다.

Rule 2) 시청 빈도가 낮거나 시청이 한번도 일어나지 않은 경우에는 MPEG-7에서 정의된 0보다 적은 preference value영역을 사용하여, preference value가 0 보다 적은 경우에는 프로그램 입력과정에서 filtering을 수행하여 프로그램을 차단시킨다.

Rule 3) 사용자의 시스템 이용 정보가 충분하지 않은 경우에는 사용자로부터 직접적인 입력을 받은 preference를 입력 받거나, stereo type의 default preference를 할당한다.

Rule 4) Action의 순서적인 조합 또한 사용자의 선호도를 반영한다. 즉, 하나의 action이 아닌 action의 순서에 의해서 같은 action이라도 그 가중치를 달리 설정하였다.

- | |
|--|
| Case 1 : Preview -> Play(high) |
| Case 2 : Play -> Stop(Middle) :: refresh |
| Case 3 : Play -> Preview(Middle) :: refresh |
| Case 4 : Play -> Fast Backward -> Play(high) |
| Case 5 : Play -> Fast Forward -> Play(Middle) |
| Case 6 : Play -> Fast Backward -> Pause -> Play(high) |
| Case 7 : Play -> Pause -> Fast Backward -> Play(high) |
| Case 8 : Play -> Fast Forward -> Pause -> Play(middle) |
| Case 9 : Play -> Pause -> Fast Forward -> Play(middle) |
| Case 10 : Any other action : Low |

표2. Action Type과 action Sequence의 가중치

Action을 8가지로 지정하였고 (Play, Pause, STOP, Fast Forward, Fast Backward, Skip one frame, Back to one frame) 그 순서의 조합에 따른 가중치의 반영은 <표2>와 같다. Action의 순서에 대한 조합을 이용하면 단순히 하나의 action만을 이용하면 선호도를 추출할 때 오류의 가능성이 높아지기 때문에 action의 sequence를 이용한다.

4. 실험 결과 및 검토

4.1. 실험의 과정

사용자의 로그인으로 사용자의 genre preference.xml에 따라서 서비스 인터페이스를 제공하고, 사용자 이용정보를 수록한다. 그리고 시스템의 이용 종료(log-out) 후에는 사용자 이용정보를 이용하여 preference.xml을 재 작성한다.

4.2. 실험의 결과

```

<?xml version="1.0" encoding="UTF-8" ?>
<mpeg7 xmlns="http://www.w3.org/2000/09/XMLSchema-instance" type="complete">
  <userPreferences xmlns="">
    <userIdentifier protected="true">
      <Name>zugoto</Name>
    </userIdentifier>
    <usagePreferences allowAutomaticUpdate="true">
      <FilteringAndSearchPreferences protected="true">
        <ClassificationPreferences>
          <Genre href="um:mpeg:GenreCS" preferenceValue="40">
            <Name>Movies</Name>
          </Genre>
          <Genre href="um:mpeg:GenreCS" preferenceValue="78">
            <Name>Sports</Name>
          </Genre>
          <Genre href="um:mpeg:GenreCS" preferenceValue="70">
            <Name>Drama</Name>
          </Genre>
          <Genre href="um:mpeg:GenreCS" preferenceValue="80">
            <Name>News</Name>
          </Genre>
          <Genre href="um:mpeg:GenreCS" preferenceValue="50">
            <Name>Music</Name>
          </Genre>
          <Genre href="um:mpeg:GenreCS" preferenceValue="12">
            <Name>Education</Name>
          </Genre>
        </ClassificationPreferences>
      </FilteringAndSearchPreferences>
    </usagePreferences>
  </userPreferences>
</mpeg7>
  
```

<그림 3> Before Auto Generation(UserPreference.xml)

현재의 preference는 News Genre에 대한 선호도가 Sports에 대한 선호보다 높은 경우이다. 이때 Sports 중에 하나의 프로그램을 시청하고 시스템을 종료하였다.

```

<?xml version="1.0" encoding="UTF-8" ?>
<mpeg7 xmlns="http://www.itu.int/ITU-T/333/2000/2000_Schema-mpeg7.xsd" type="complete">
  <UserPreferences xmlns="">
    <UserIdentifier protected="true">
      <UserIdentifierName>zugoto</UserIdentifierName>
    </UserIdentifier>
    <UsagePreferences allowAutomaticUpdate="true">
      <FilteringAndSearchPreferences protected="true">
        <ClassificationPreferences>
          <Genre href="urn:mpeg:GenreCS" preferenceValue="36">
            <Name>Movies</Name>
          </Genre>
          <Genre href="urn:mpeg:GenreCS" preferenceValue="80">
            <Name>Sports</Name>
          </Genre>
          <Genre href="urn:mpeg:GenreCS" preferenceValue="63">
            <Name>Drama</Name>
          </Genre>
          <Genre href="urn:mpeg:GenreCS" preferenceValue="79">
            <Name>News</Name>
          </Genre>
          <Genre href="urn:mpeg:GenreCS" preferenceValue="45">
            <Name>Music</Name>
          </Genre>
          <Genre href="urn:mpeg:GenreCS" preferenceValue="10">
            <Name>Education</Name>
          </Genre>
        </ClassificationPreferences>
        <FilteringAndSearchPreferences>
          <UsagePreferences>
            </UserPreferences>
          </UsagePreferences>
        </FilteringAndSearchPreferences>
      </FilteringAndSearchPreferences>
    </UsagePreferences>
  </UserPreferences>
</mpeg7>

```

<그림 4> After Auto Generation(UserPreference.xml)

4.3. 결과의 분석

Statistical approach는 데이터가 충분히 축적되지 않은 상태에서는 사용자의 특징을 제대로 추론하지 못하는 특징이 있었다. 지나치게 특정 genre나 분야에 preference가 높거나 낮은 경우가 발생하였고, 이런 단점을 보완하기 위하여 일정 시간이상 사용자 정보가 축적되지 않은 경우에는 preference value를 추출하지 않도록 해야 한다.

선호도 값의 변화의 정도는 이전 선호도 값의 반영의 비율에 의해서 차이가 발생한다. 반영의 비율이 크게 설정되었을 때는 사용자의 선호도가 초기의 선호도에서 크게 벗어나지 않고, 반영의 비율이 낮은 경우에는 사용자의 선호도가 최근의 시스템 이용자 정보에 의해 변화가 크게 발생하였다. 선호도 변화의 정도를 조절하기 위하여 많은 실험이 요구되었다.

Statistical approach와 Rule-Based approach는 상호 연관성을 지니고 있다. 사용자 선호도를 추론 및 사용자 선호도를 기반으로 한 서비스의 제공에서는 Rule-Based와 statistical approaches의 복합적인 사용이 필요하다.

5. 결 론

TV 기반에서 사용자 취향 정보를 Statistical Method와 Rule-Based approach를 기반으로 추출하였다. 이런 사용자 취향 정보의 추출은 사용자의 콘텐츠 인터랙션 히스토리를 자동 분석하여 이루어졌으며 사용자 선호도의 변화에 적응적으로 선호도를 자동 갱신할 수 있음을 확인하였다. 사용자의 입력을 요구하지 않고 사용자 취향 정보를 추출해내는 알고리즘은 대체널 방송 환경에서 시청자에게 제공되는 수 많은 방송 콘텐츠에 대해 사용자의 선호도가 높은 프로그램을 추천하고 쉽게 획득할 수 있게 하며, 콘텐츠를 쉽게 탐색할 수 있는 기능을 지원하는 지능형 방송 단말 소프트웨어 에이전트 기술에 중요하게 사용될 수 있을 것으로 기대된다.

Acknowledgement

본 논문은 한국전자통신연구원 “메타데이터 기반 사용자 단말 에이전트 기술 개발”에 관한 위탁연구과제 수행의 일환으로 얻어진 연구결과입니다.

References

- [1] P. Vuorimaa and C. Sancho, "XML based text TV," Web Information Systems Engineering, WISE2000, Hong Kong, June 19-20, 2000.
- [2] ISO/IEC JTC 1/SC29/WG11/N3966, Text of 15938-5 FCD Information Technology - Multimedia Content Description Interface, Singapore, March 2001.
- [3] Metadata, Specification Series of The TV-Anytime forum, Document SP003v1.1, August 17, 2001.
- [4] Ronald Schrooten and Ivo van den Maagdenberg, Agents Technology in Europe, Edited by the ACTS Project Infowin (AC113), 1999, ISBN 3-00-005267-4.