
순차 패턴 마이닝 기법을 이용한 개인 맞춤형 (IP)TV 프로그램 스케줄 자동 추천

프로그램 시청 시간의 정량적 정보를 고려한 패턴 추출 및 개인 선호도 정보 추출을 통한 스케줄 추천 시스템

Automatic Recommendation on (IP)TV Program schedules in a personalized way using sequential pattern mining

표신지, Shinjee Pyo*, 김은희, Eunhui Kim**, 김문철, Munchurl Kim**

요약 기존의 TV 시청 환경은 제한된 채널과 콘텐츠들 중에서 사용자가 스스로 채널 이동을 통해 자신이 원하는 콘텐츠를 선택하여 시청하는 형태였다. 그러나 IPTV 의 도입과 더불어 다채널/다매체 시청 환경에서는 사용자가 수많은 콘텐츠에 접근이 가능하며 그로 인해 자신이 선호하고 자신에게 친숙한 콘텐츠를 찾는 것이 많은 부담이 되고 있다. 따라서 본 논문에서는 이러한 TV 시청 환경에서 사용자에게 적합한 프로그램들을 시간적 순서에 따라 추천해 주는 개인화된 (IP)TV 프로그램 자동 스케줄러를 제안한다. 개인 맞춤형 프로그램 스케줄 추천 시스템은 사용자의 누적된 시청 기록을 기반으로 사용자의 시청 패턴을 추론하고 사용자 개인의 선호도 정보를 적용하여 순차적인 프로그램을 추천한다.

Abstract Conventional TV viewing environment had provided limited numbers of channels and contents so that accessibility of contents was made user's manual change of TV channels and by manual selection of TV program contents. However, with advent of IPTV and various contents and channels available to users' terminals, excessive numbers of TV contents become available to users' terminals, thus leading to totally different TV viewing environments. In this TV environment, users are required to make much effort to choose their preferred TV channels or program contents, which becomes much cumbersome to the users. Therefore, in this paper, we will propose TV contents schedule recommendation by making reasoning on users' TV viewing patterns from TV viewing history data using sequential pattern mining so that so that it increases accessibility of users to many TV program contents which may be or may not be aware of the users.

핵심어: *Sequential pattern mining, Data mining, Recommendation, IPTV*

본 연구는 지식경제부 및 정보통신연구진흥원의 IT 핵심기술개발사업의 일환으로 수행하였음. [A1100-0801-3015, Development of Open-IPTV Technologies for Wired and Wireless Networks]

*주저자 : 한국정보통신대학교 공학부 석사과정 e-mail: vy311@icu.ac.kr

**공동저자 : 한국정보통신대학교 공학부 석사과정 e-mail: ehkim@icu.ac.kr

**공동저자 : 한국정보통신대학교 공학부 교수: e-mail: mkim@icu.ac.kr

1. 서론

기존의 (IP)TV 시청환경은 사용자가 자신이 원하는 프로그램을 채널 검색 등을 통하여 선택하고 시청하는 형태였다. 하지만 채널 및 콘텐츠의 다양화와 인터넷의 발달로 사용자는 매우 다양한 콘텐츠를 접할 수 있게 되었다. 이러한 시청 환경에서 사용자는 이전보다 다양해진 프로그램 중에서 자신이 선호하고 자신에게 친숙한 콘텐츠를 찾는 것에 대해 부담을 느낄 수 있다. 따라서 사용자에게 적합한 콘텐츠와 시청 스케줄을 추천해 줄 수 있는 추천 시스템이 요구된다.

이러한 추천 시스템은 사용자의 TV 시청 기록을 기반으로 시청 패턴을 분석하고 사용자 선호도 정보를 추론하는 과정을 통해 이루어 질 수 있다. 시청 패턴 분석을 위해 데이터 마이닝의 한 분야인 순차 패턴 마이닝을 적용하여 사용자의 의미 있는 패턴을 추출하고 개인 사용자의 선호도를 고려하여 개인 맞춤형 프로그램 스케줄을 추천한다.

2. 적용 환경 및 데이터 구성

순차 패턴 마이닝 기법을 적용하여 사용자의 순차적인 시청 패턴을 추론하기 위해서 본 연구에서는 적용할 TV 시청 환경에 적합하도록 기존의 순차 패턴 마이닝 기법 중 하나인 PrefixSpan 알고리즘을 (IP)TV 프로그램 시청에 적용되도록 확장하였다. 기존의 순차 패턴 마이닝은 주로 물건 구매 환경에 적용되어 순차적인 구매 활동을 추출하는 데에 사용되었다[1]. 따라서 기본적인 트랜잭션의 경우, 물건 구매 환경에서는 물건 구매의 행위가 되고, TV 시청환경에서는 프로그램 시청이 된다. 또한 구매 항목(아이템) 시퀀스의 구성에 있어서 기존의 적용환경에서는 일정 기간 동안 구매한 물건들로 구성이 되고, TV 시청 환경에서는 하루 동안의 시간 순서에 따른 시청 기록이 된다. 아이템의 경우, TV 시청환경에서는 시청 프로그램이고, 빈도의 경우에는 단순한 시청 횟수뿐만 아니라 시청 시간도 빈도의 개념에 포함된다.

프로그램 시청 패턴 추출을 위해서 본 논문에서 사용된 데이터는 사용자의 id, 성별, 나이, 시청 프로그램, 프로그램 방영 시간, 시청 시작 시간, 시청 종료 시간, 프로그램 장르 정보, 프로그램 방영 횟수로 구성된 2002년 12월부터 2003년 5월까지 6개월간 6개의 채널(KBS1, KBS2, MBC, SBS, EBS, iTV)에 대해서 사용자 2000여명의 시청 기록이다. TV 프로그램의 경우, 대개 일주일 단위로 프로그램이 반복되는 형태이므로 시청 기록 데이터를 요일 별로 나누어서 패턴 추출을 사용할 수 있다.

3. 순차 패턴 마이닝 기법을 이용한 개인 프로그램 스케줄 추천 시스템

본 논문에서 제안하는 개인 맞춤형 프로그램 스케줄 추천 시스템은 다음과 같이 구성된다.

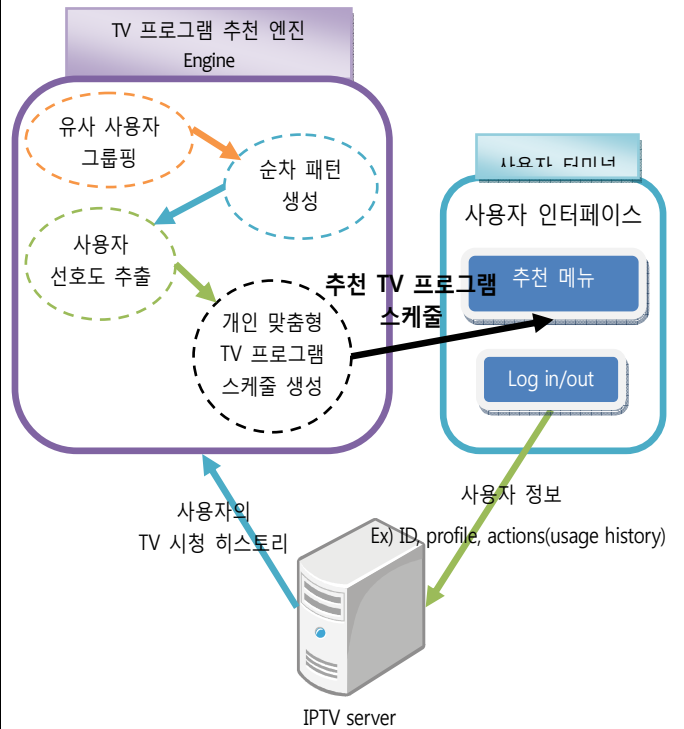


그림 1. 개인 프로그램 스케줄 추천 시스템

다음 그림과 같이 사용자의 시청 기록을 이용하여 유사한 시청 습관을 갖는 유사 사용자 그룹을 형성하고, 형성된 사용자 그룹 내에서 순차적인 시청 패턴을 추출 한다. 또한 사용자 개인별 선호도를 추출하여, 그룹 내에서 형성된 순차 패턴에 적용하여 개인별 프로그램 스케줄을 추천할 수 있도록 시스템을 구성하였다. 다음 절에서 각 단계마다의 세부기술에 대해 설명하겠다.

3.1 유사 시청 사용자 그룹 형성 단계

유사 시청 사용자 그룹은 사용자들 간의 유사성을 벡터 공간 모델(vector space model)을 이용하여 계산하여 구성한다. 벡터 공간 모델은 두 벡터 간의 유사성을 계산하는 방법으로 다음과 같은 수식을 사용한다.

$$\cos \theta = \frac{\mathbf{v}_{u_a} \cdot \mathbf{v}_{u_i}}{\|\mathbf{v}_{u_a}\| \|\mathbf{v}_{u_i}\|}, \mathbf{v}_{u_a} = [w_{1,u_a}, w_{2,u_a}, \dots, w_{10,u_a}] \quad (1)$$

벡터 공간 모델은 두 벡터간의 유사성을 비교하여 유사 그룹을 형성하는 방법이므로, 기준이 되는 사용자 한 명을 선택하여, 이 사용자의 최상위 10 개의 프로그램에 대한 선호도 값으로 벡터를 구성한다. 그 외의 사용자들에 대해서도 선택된 사용자의 최상위 10 개의 프로그램과 동일한 프로그램을 시청했을 시, 그 프로그램에 대한 선호도 값으로 벡터를 구성하여 10 차원의 벡터를 구성한다. 벡터를 구성하는 값은 86 명의 50 대 여성의 2002년 12월~2003년 2월까지의 3개월간의 시청 기록을 기반으로 프로그램 별 시청 시간에 대해 TF-IDF(Term

Frequency Inverse Document Frequency)을 계산하여 구성한다. 각 사용자 별 벡터 값으로 벡터공간모델을 이용하여 그 값이 1 과 가까운 사용자들로 30 명의 유사 사용자들로 사용자 그룹을 형성한다.

3.2 순차 패턴 형성 단계

3.1 절에서 형성된 그룹내의 사용자들의 3 개월 간의 시청 기록을 요일 별로 나누어 요일 별 빈발 시청 프로그램을 선별한다. 빈발 시청 프로그램은 사용자의 시청 시간을 고려한 정량적인 순차 패턴 마이닝 기법을 적용하여 빈발 시청 프로그램을 선별한다. 기존의 순차 패턴 마이닝의 경우, 아이템의 구매 횟수 만을 고려하여 빈발 아이템을 선별하였다. 하지만 TV 프로그램 시청 환경의 경우에는 시청 빈도뿐만 아니라 시청 시간 또한 고려해야 한다. 시퀀스는 각 사용자의 각 날짜 별 시청 프로그램들로 구성하며, 프로그램 제목은 인덱스 번호를 부여하여 표현하였다. 사용자의 프로그램 별 시청 시간 정보는 프로그램의 방영 시간대비 사용자의 시청 시간으로 다음과 같은 수식으로 표현하였다.

$$q_{j,u_i} = \frac{wt_{j,u_i}}{d_j} \quad (2)$$

q_{j,u_i} 는 사용자 u_i 의 프로그램 p_j 에 대한 시청 시간의 정량적인 정보이고 d_j 는 프로그램 p_j 의 방영 시간, wt_{j,u_i} 는 사용자 u_i 의 프로그램 p_j 의 시청 시간이다. 이러한 q_{j,u_i} 값으로 표현한 시퀀스는 다음과 같다.

사용자ID /요일/날짜	정량적인 정보를 포함한 시퀀스
u1/일/2002.12.1	$\langle P_{134,0.5}, P_{56,0.2} \dots \rangle$
u2/일/2002.12.8	$\langle P_{578,1}, P_{134,0.6} \dots \rangle$
u3/일/2002.12.1	$\langle P_{134,0.4}, P_{87,0.4} \rangle$
u4/일/2002.12.15	$\langle P_{134,0.6}, P_{912,0.7} \dots \rangle$

표 1 시청시간에 대한 정량 정보를 포함한 시퀀스

$$Q_{p_i} = \frac{\sum_{k=1}^N q_{i,u_k}}{B_{p_i}} \quad (3)$$

Q_{p_i} = p_i 의 정량적 값;

N = 유사 사용자그룹의 총 명수;

q_{i,u_k} = 사용자 u_k 의 p_i 에 대한 시청 비율;

B_{p_i} = p_i 의 방영횟수

TV 프로그램 시청 환경에 적합한 형태로 제한한 각 프로그램 별 정량적인 시청 정보를 포함한 시퀀스를 기반으로 각 요일 별 빈발 시청 프로그램을 선별한다. 각 프로그램에 대한 q_{j,u_i} 값을 30 명의 모든 사용자의 시퀀스 내에서 모두 더하여 다음과 같은 프로그램 정량 정보를 계산한다.

각 프로그램에 대한 이러한 Q_{p_i} 값을 빈발 시청 프로그램을 선별하는 기준으로 사용한다. 빈발 시청 프로그램으로 선별 될 수 있는 Q_{p_i} 의 최소 지지도 값은 2 로 정하여 빈발 시청 프로그램을 선별하였다. 임의로 정하는 최소 지지도 값은 패턴 형성에 밀접한 영향을 주므로 최소 지지도 값에 따른 패턴 추천 성능에 대해 5 절 실험 결과에서 논의하도록 한다.

선별된 빈발 시청 프로그램으로 패턴을 생성하기 위해 PrefixSpan 기법을 적용하였다. 요일 별로 선별된 빈발 시청 프로그램을 포함하는 시퀀스들을 골라 낸 후, 이 시퀀스를 빈발 시청 프로그램으로 projection 하는 과정을 거쳐 각 빈발 시청 프로그램 별 projected database 를 형성한다. projected database 내에서 앞서 논의한 빈발 시청 프로그램 선별 과정을 동일한 방법으로 거쳐 projected database 내의 빈발 시청 프로그램을 선별한다. 이러한 경우, '요일 별 빈발 시청 프로그램 - 빈발 시청 프로그램의 projected database 내의 빈발 시청 프로그램'의 형태로 길이-2 인 순차 패턴이 형성된다. 다음 표는 projection 과정을 거쳐 생성된 길이-2 의 요일 별 순차 패턴들을 나타낸다.

요일	길이-2의 순차 패턴들
일	$\langle P_{528} - P_{617} \rangle, \langle P_{528} - P_{366} \rangle$... 263개의 순차 패턴이 발견됨
월	$\langle P_{520} - P_{1083} \rangle, \langle P_{520} - P_{465} \rangle$... 25개의 순차 패턴이 발견됨
화	$\langle P_{1083} - P_{371} \rangle, \langle P_{501} - P_{465} \rangle$... 26개의 순차 패턴이 발견됨
수	$\langle P_{1081} - P_{324} \rangle, \langle P_{501} - P_{342} \rangle$... 39 개의 순차 패턴이 발견됨
목	$\langle P_{1081} - P_{299} \rangle, \langle P_{459} - P_{501} \rangle$... 46개의 순차 패턴이 발견됨
금	$\langle P_{445} - P_{281} \rangle, \langle P_{445} - P_{372} \rangle$... 34개의 순차 패턴이 발견됨
토	$\langle P_{54} - P_{593} \rangle, \langle P_{528} - P_{54} \rangle$... 160개의 순차 패턴이 발견됨
총 593개의 길이-2의 순차 패턴이 발견됨	

표 2 생성된 길이-2 의 순차 패턴들

선별된 projected database 내의 빈발 시청 프로그램들에 대해서 두 번째 projection 을 시행하여, 길이-3 의 순차 패턴을 생성할 수 있다. projection 을 통한 projected database 생성과 projected database 내에서의 빈발 시청 프로그램 선별 과정을 반복적으로 시행함으로써 길이가 긴 순차 패턴들을 생성할 수 있다. 본 연구에서는 길이-3 인 순차 패턴까지 생성하였다. 다음 표는 생성된 길이-3 인 순차 패턴이다.

요일	길이-3의 순차 패턴들
일	$\langle P_{528} - P_{617} - P_{590} \rangle, \langle P_{528} - P_{366} - P_{180} \rangle$... 1743개의 순차 패턴이 발견됨
월	$\langle P_{1113} - P_{520} - P_{1083} \rangle, \langle P_{1113} - P_{501} - P_{465} \rangle$... 140개의 순차 패턴이 발견됨

화	$\langle P_{1124} - P_{493} - P_{465} \rangle, \langle P_{1113} - P_{472} - P_{501} \rangle$... 112개의 순차 패턴이 발견됨
수	$\langle P_{501} - P_{629} - P_{342} \rangle, \langle P_{501} - P_{629} - P_{342} \rangle$... 141 개의 순차 패턴이 발견됨
목	$\langle P_{459} - P_{501} - P_{891} \rangle, \langle P_{459} - P_{501} - P_{299} \rangle$... 161개의 순차 패턴이 발견됨
금	$\langle P_{445} - P_{372} - P_{281} \rangle, \langle P_{445} - P_{1033} - P_{281} \rangle$... 116개의 순차 패턴이 발견됨
토	$\langle P_{528} - P_0 - P_{54} \rangle, \langle P_{528} - P_0 - P_{598} \rangle$... 909개의 순차 패턴이 발견됨
총 3181개의 길이-3의 순차 패턴이 발견됨	

표 3 생성된 길이-3의 순차패턴들

3.3 개인별 선호도 추출 단계

유사 사용자 그룹내의 30 명의 사용자 각각의 선호도를 추출하는 단계이다. 선호도는 프로그램, 장르, 채널에 대한 선호도로 추출한다. 프로그램에 대한 선호도는 프로그램에 대한 시청 시간 정보를 기반으로 식(2)로 계산한다. 계산한 값으로 상위 20 개의 프로그램들을 사용자의 선호 프로그램을 선별하여, 각각의 선호도 값을 저장한다. 장르에 대한 선호도 값은 각 장르에 해당하는 프로그램의 시청 시간을 기반으로 전체 시청 시간에 대한 특정 장르에 대한 시청 시간으로 계산하여 선호도 값을 구한다. 채널에 대한 선호도 값도 마찬가지로 각 채널내의 시청 시간을 기반으로 전체 시청 시간에 대한 특정 채널에 대한 시청 시간으로 선호도 값을 구한다.

3.4 개인별 추천 프로그램 스케줄 형성 단계

3.2 절과 3.3 절에서 추출된 유사 그룹내의 길이-2, 길이-3 의 순차 패턴들과 개인별 선호도 정보를 가지고 사용자의 개인별 추천 패턴을 형성한다. 즉, 표 2 와 표 3 에서와 같이, 요일 별로 생성된 패턴들이 매우 다양할 때에 개인 사용자의 선호도에 적합하도록 추천 순위를 매겨 사용자에게 추천할 수 있도록 하는 것이다. 이러한 개인별 추천 순위는 다음 식으로 계산한 값이 큰 순서대로 정렬하여 순위를 정한다.

$$RS_{u_k,i} = (1 + PV_{u_k,i}) * (1 + GV_{u_k,G_i}) * (1 + CV_{u_k,C_i}) * Q_{P_i} \quad (4)$$

다음 식은 생성된 순차 패턴을 구성하는 프로그램에 대한 사용자의 프로그램, 장르, 채널 선호도 값과 그 프로그램 시청시간에 대한 정량적인 값을 이용하여 개인별 패턴에 대한 추천 순위를 계산할 수 있도록 표현한 식이다. 식에 따라 각 패턴에 대하여 각 사용자 별로 계산하여 개인별 추천 순위를 구성하였다.

다음 절에서는 생성된 패턴들의 추천 성능에 대한 실험 결과와 최소 지지도 값에 따른 성능 변화 및 기존의 PrefixSpan 방법을 이용하여 패턴을 추출했을 경우와

제안된 정량정보를 고려한 순차 패턴 마이닝으로 패턴을 추출했을 경우의 성능 비교에 대해 논의할 것이다.

4. 실험 결과

실험은 크게 다음과 같은 형태로 진행하였다. 먼저 유사 그룹 내에서 생성된 길이-1, 길이-2, 길이-3 의 순차 패턴들이 테스트 데이터에 실제로 존재하였는지 그룹대 그룹으로 비교하여 실험하였고, 생성된 길이-2, 길이-3 의 개인별 패턴 추천 순위에 따라 테스트 데이터에서 실제로 시청 되었는지 사용자대 사용자로 비교하여 실험하였다. 또한 최소 지지도의 변화에 따른 성능 변화에 대해 실험하였으며, 기존의 PrefixSpan 방법과 제안된 순차 패턴 마이닝의 추천 성능을 비교하여 실험하였다. 테스트 데이터는 2003 년 3 월부터 5 월까지의 동일한 사용자 30 명의 시청 기록으로 실험하였다.

4.1 선별된 패턴들의 추천 성능

유사 시청 사용자 그룹 내에서 생성된 길이-1, 길이-2, 길이-3 의 순차 패턴들이 실제로 테스트 데이터의 시청 기록 내에서 얼마나 발견되었는지를 계산하여 다음과 같은 그래프로 나타내었다.

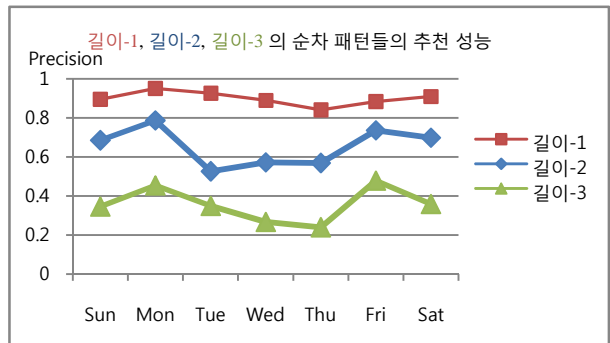


그림 2 최소 지지도 Q=2 일 때, 그룹 내에서의 순차 패턴들의 추천 성능

그래프에서와 같이 길이-1 일 경우에는 패턴의 추천 정확도가 1 에 가까운 것을 확인할 수 있었고 길이가 길어질수록 추천 정확도가 떨어짐을 확인할 수 있었다. 또한 요일 별 추천 성능도 사용자들의 요일 별 시청 정도에 따라 다르게 나타난 것을 알 수 있다. 요일 별로 사용자가 프로그램을 시청하는 시간의 차이로 인해 생성된 패턴들의 개수가 서로 달랐으며, 생성된 패턴에 대한 추천 성능 또한 다르게 나온 것을 확인할 수 있었다.

4.2 개인별 추천 패턴 순위에 따른 추천 성능

길이-2, 길이-3 의 순차 패턴들을 사용자의 선호도를 고려하여 추천 순위를 매겨 사용자에게 추천했을 경우, 사용자가 실제로 추천된 패턴을 시청했는지에 대한 추천 성능을 계산하였다. 그림 3 에서와 같이, 길이-2 인 개인별 추천 패턴을 개개인에게 추천했을 경우, 최고 60%정도의

시청 확률을 보였으며, 길이-3 인 개인별 추천 패턴을 추천해 주었을 때에는, 최고 50%의 시청 확률을 보였음을 확인할 수 있었다. 성능이 예상했던 것 보다 높지 않았던 이유는 사용자의 시청 시퀀스의 부족과 사용자의 시청 패턴의 변화 때문이었다. 순차적 추천 패턴이 실제로 시청되기 위해서는 규칙적/지속적/장기적인 시청 시퀀스가 요구된다. 하지만 테스트 데이터 내에서는 이러한 시퀀스 뿐만 아니라 일시적이고 들쭉 날쭉한 시퀀스 또한 존재한다. 따라서 성능 면에서 다소 떨어지는 것을 확인할 수 있었다. 개인별 추천 패턴의 성능을 높이기 위해서는 시간의 흐름에 따른 사용자의 시청 패턴의 변화를 고려한 패턴 구성이 요구되며, 요일 별 시청 정도에 따른 추천 패턴 구성이 요구된다.

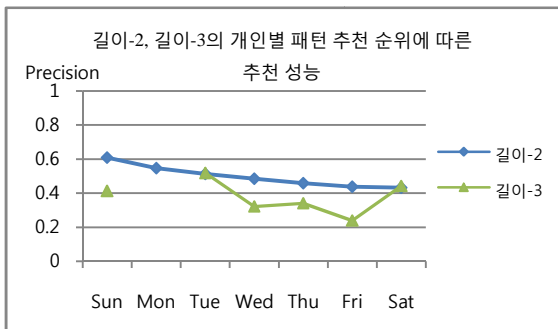


그림 3 최소 지지도 Q=2 일 때, 개인별 패턴 추천 순위에 따른 추천 성능

4.3 최소 지지도에 따른 추천 성능 변화

최소 지지도 Q 값은 빈발 시청 프로그램 선별에 있어서 threshold 값으로 사용된다. 따라서 Q 값이 크면, 기준이 높아지므로 적은 빈발 시청 프로그램들이 선별되며, 반대로 Q 값이 작으면 빈발 시청 프로그램으로 선별될 수 있는 기준이 낮아지므로 많은 빈발 시청 프로그램들이 선별된다. 선별되는 빈발 시청 프로그램에 따라 생성되는 패턴이 달라지고, 이로 인한 패턴의 추천 성능이 달라질 수 있다. 그림 4 와 그림 5 는 길이-2, 길이-3 인 순차 패턴의 추천 성능을 Q 값의 변화에 따라 실험한 결과이다. 그림 4 와 그림 5 에서와 같이 최소 지지도의 Q 값을 1 부터 3.5 까지 변화 시키면서 추천 성능을 보았을 때, Q 값이 3 일때 가장 좋은 성능을 보인 것을 확인할 수 있었다.

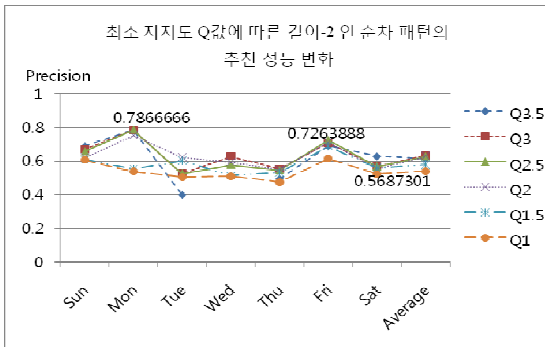


그림 5 최소 지지도 Q 값에 따른 길이-2 인 순차 패턴의 추천 성능 변화

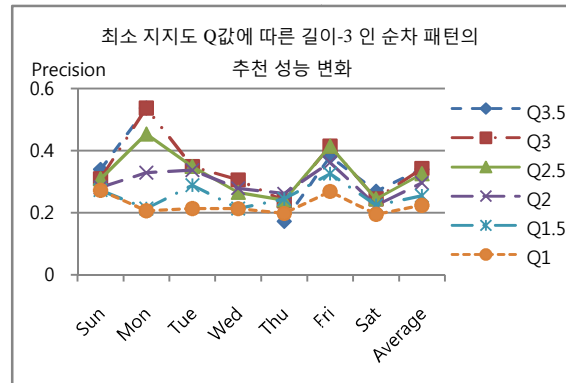


그림 6 최소 지지도 Q 값에 따른 길이-3 인 순차 패턴의 추천 성능 변화

4.4 PrefixSpan 방법과 제안된 정량적 정보를 포함한 순차 패턴 마이닝 방법의 패턴 추천 성능 비교

기존의 순차 패턴 마이닝 방법인 PrefixSpan 방법을 적용하여 생성된 패턴을 추천했을 경우와 사용자의 시청 시간의 정량적 정보를 고려한 순차 패턴 마이닝을 적용하여 생성된 패턴을 추천했을 경우의 성능을 비교하였다. 먼저 길이-1 의 패턴의 추천 성능은 그림 6 와 같다. 그림 6 에서와 같이 제안된 순차 패턴 마이닝 기법을 패턴 추천에 적용했을 때에 보다 높은 성능을 나타낸 것을 확인할 수 있었다.

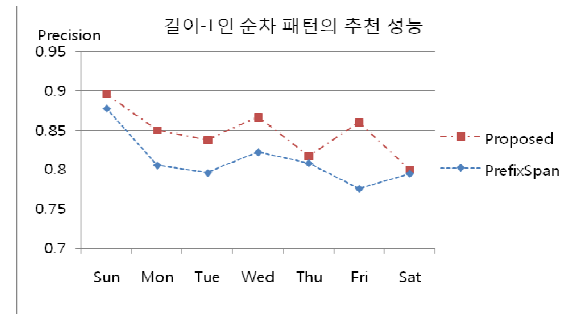


그림 4 길이-1 인 패턴에서 PrefixSpan 방법과 제안된 방법의 추천 성능 비교

그림 7 은 길이-2 인 순차 패턴의 추천 성능을 비교한 그래프이다. 기존의 PrefixSpan 보다 높은 성능을 나타내는 것을 확인할 수 있으며 요일 별 성능이 PrefixSpan 와는 달리 서로 차이가 많이 나타나는 것을 볼 수 있다. 제안한 순차 패턴 마이닝의 경우에는 패턴 추출 시에 시청 시간을 고려하여 추출한다. 따라서 요일 별로 시청 시간이 서로 다르며 시청 습관 또한 다르기 때문에 요일 마다 성능에 차이를 보인 것을 알 수 있었다.

그림 8 도 마찬가지로 기존의 PrefixSpan 보다 성능이 좋지만 요일 별 성능에 차이가 많은 것을 볼 수 있다. 이러한 요일 별 성능의 차이를 원만하게 하기 위해서는 사용자의 요일 별 시청 습관에 대해 고려하여 요일 별로

패턴 추출 시 서로 다른 최소 지지도 값을 이용하는 방법 등으로 성능 차이를 줄일 수 있을 것이다.

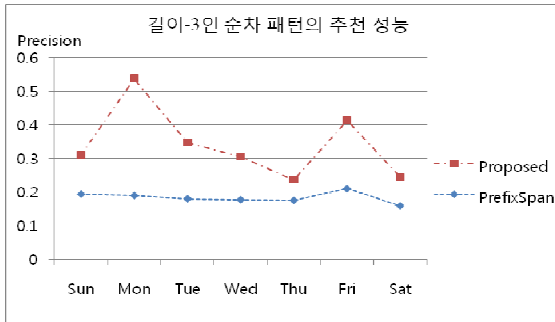


그림 7 길이-3 인 패턴에서 PrefixSpan 방법과 제안된 방법의 추천 성능 비교

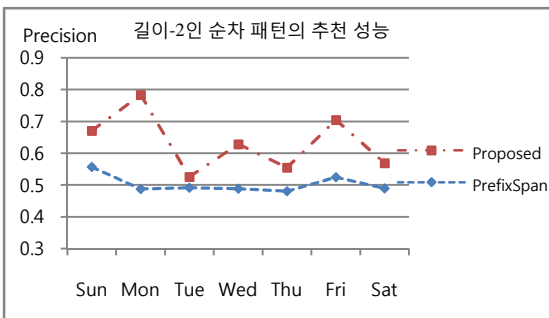


그림 8 길이-2 인 패턴에서 PrefixSpan 방법과 제안된 방법의 추천 성능 비교

5. 결론

본 연구에서는 사용자들의 시청 기록을 기반으로 순차적인 프로그램 시청 패턴을 발견하여 사용자 선호도를 고려한 개인 맞춤형 프로그램 스케줄 추천 시스템을 제안하였다. TV 시청 환경에서의 순차적인 패턴을 발견하기 위해 사용자의 프로그램 시청 시간을 정량적인 정보로

표현하여 패턴을 추출하는 정량정보를 포함한 순차 패턴 마이닝 기법을 제안하였으며, 추출한 사용자의 선호도 정보와 순차 패턴들을 결합하여 개인별 순차 패턴을 구성하기 위해 추천 순위 모델을 제안하였다. 제안된 방법으로 생성된 패턴들의 추천 성능을 실험을 통해 검증하였으며, 기존의 PrefixSpan 기법과의 성능 비교를 통해 제안된 순차 패턴 마이닝 기법이 TV 시청 환경에서 효율적으로 패턴을 추출함을 입증하였다. 향후 과제로는 시간의 흐름에 따른 사용자의 시청 습관의 일관성을 고려한 패턴 추천 연구 및 요일 별 시청 특성을 고려한 순차 패턴 마이닝 연구를 진행 할 계획이다.

참고문헌

- [1] Rakesh Agrawal, Ramakrishnan Srikant, "Mining sequential patterns," 11th International Conference on Data Engineering (ICDE'95), 1995 *IEEE Transactions on Software Engineering*, 1995.
- [2] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Helen Pinto, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," In Proc. 2001 Int. Conf. Data Engineering, Vol. April 2001.
- [3] Baoyao Zhou, Siu Cheung Hui, Kuiyu Chang, "An Intelligent Recommender System using Sequential Web Access Patterns," In Proc. 2004 IEEE Conference on Cybernetics and Intelligent Systems, December 2004.