

TV 관련 논문

Time-dependent User Profiling for TV Recommendation

INTRODUCTION

TV 추천 시스템 검증 기술 연구 시 문제점

- 시청자 로그를 쉽게 사용할 수 없음
- 모든 개인에 대한 Watch log 수집 불가능

연구 방법

- 하루의 시청로그를 시간대 기준으로 분석
예) 4-5PM은 부모가 일하는 시간이므로 아이들의 선호도 반영
반대로 자정에는 아이들이 자므로 부모의 선호도 반영
- Time-dependent Profiling 사용

PROFILING

Typical Profiling(Content-based)

- Item(TV프로그램) <Item ID, Description Vector>
- Watch log <User ID, Item ID>
- Watch log → Profile <User ID, Preference Vector>
- Preference Vector는 사용자가 보고있는 프로그램의 Description Vector를 집계하여 생성
- 사용자와 보지않은 프로그램 간의 선호도는 사용자의 Preference Vector와 프로그램의 Description Vector의 내적으로 계산

PROFILING

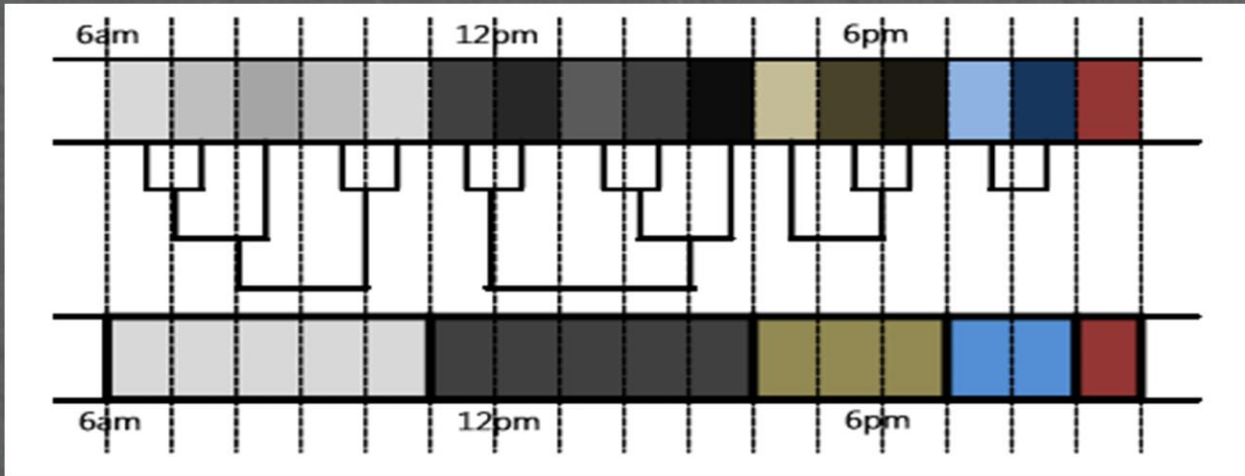
Time Dependent Profiling

- 시간 기반의 프로필을 생성하는 새로운 TV 추천방법
- 방송시간이 자주 바뀌지 않기 때문에 효과적
 - > 누군가 TV프로그램을 좋아한다면 방송 시간에 보고싶어 함
- 사람들도 대부분 고정된 일과를 따름
- 주 단위로 Watch log 분석, 30분 간격으로 분할(총 336개의 타임슬롯)
- 각 타임슬롯에 대해 Item Vector를 집계함으로써 Preference Vector 생성

PROFILING

Profile Smoothing(Clustering)

- 시청자의 일상을 프로그램에 고정시켰기 때문에 오버피팅이 일어남
- 응집 클러스터링을 이용하여 연속적인 프로파일 유사할 경우 이를 병합



PROFILING

Profile Smoothing(Clustering)

- 프로파일 간 유사성은 프로파일에 속하는 항목의 유사성의 합
- 프로그램 간의 유사도는 코사인 유사도를 사용

$$S(p_i, p_j) = \frac{1}{|p_i||p_j|} \sum_{I_t \in p_i} \sum_{I_k \in p_j} S(I_t, I_k)$$

$$S(I_t, I_k) = \frac{I_t \cdot I_k}{|I_t||I_k|}$$

SETTING

Watch log

- 607가구의 감시로그를 1년 동안 기록(2008명, 58206개)
- <Family ID, Member ID, Channel, Starting Time, Ending Time>
- 9개월 (Train Data), 3개월 (Test Data)

EPG and Meta Dataset

- 콘텐츠 기반 필터링 사용(Baseline)
- Daum1에서 크롤링
- EPG <Program ID, Start Time, End Time, Channel>
- 메타 데이터 – 디렉터, 타이틀, 장르 , 요약, 배우 등

SETTING

Profile Variants

- Watch log에 두 개의 ID가 있으므로 프로필 2개 생성가능
- Member ID 기반 프로필, Family ID 기반 프로필

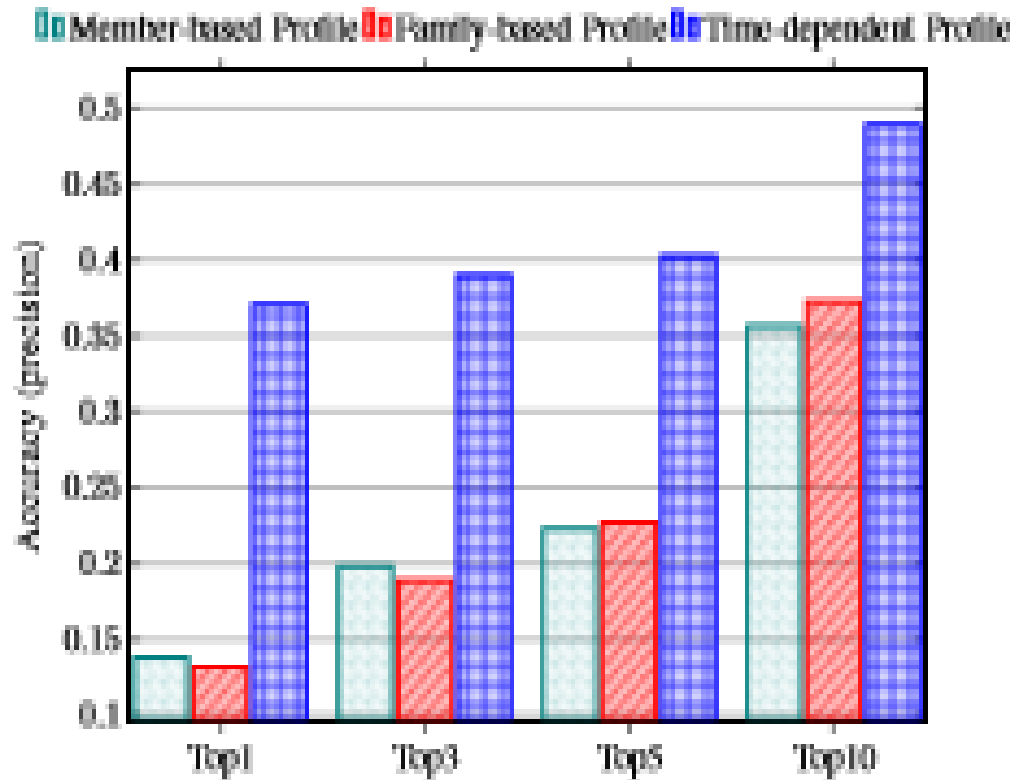
Accuracy measurement

- Precision 사용
- 한 번에 하나의 프로그램을 추천하기 때문에 의미가 있음

$$Precision = \begin{cases} 1, & \text{Watched Item} \in \text{Recommended Item set} \\ 0, & \text{Watched Item} \notin \text{Recommended Item set} \end{cases}$$

CONCLUSION

Accuracy Comparison



- 추천 횟수가 동일 할 때 Time-dependent Profile이 가장 성능이 좋음
- 추천 수가 줄어들수록 다른 프로파일과 큰 차이가 남
- 현실에서는 추천 수가 매우 적기 때문에 Time-dependent Profile이 제일 효과적

시청시간패턴을 활용한 TV 프로그램 추천 시스템

소개

- 기존 추천시스템 연구들은 정적인 데이터(특성, 평점)만 사용
 - > 동적 데이터가 변해도 선호도 점수에 영향을 주지 않음
- 본 논문에서는 시청시간패턴 기반 협업 필터링 추천 시스템 사용
 - > 시청비율, 종료시간과의 관계, 최근시청이력 등

배경 지식

협업 필터링 추천 시스템

- 사용자 취향과 유사한 취향을 갖는 사용자의 취향을 바탕으로 아이템 추천
- 사용자와 유사 취향 사용자 간의 유사성을 계산해 평가되어지지 않은 부분을 예측

$$corr_{a,b} = \frac{\frac{1}{m} \sum_{i=1}^m (r_{a,i} - \bar{r}_a)(r_{b,i} - \bar{r}_b)}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2 \sum_{i=1}^m (r_{b,i} - \bar{r}_b)^2}}$$

피어슨 상관 계수

$$\cos_{\vec{a}, \vec{b}} = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

코사인 유사도

배경 지식

협업 필터링 추천 시스템

- 예측 선호도

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

\bar{r}_a : 사용자 a 의 평점 평균

\bar{r}_u : 사용자 u 의 평점 평균

$w_{a,u}$: 사용자 a 와 다른 사용자 u 의 유사도

배경 지식

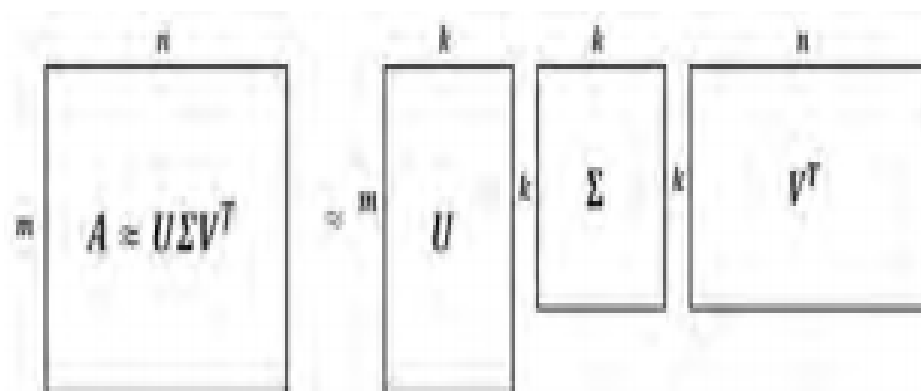
SVD (Singular Value Decomposition)

$$A = U \Sigma V^T$$

$$A^T A = V \Lambda_1 V^T, \text{ } n \times n \text{ symmetric matrix}$$

$$A A^T = U \Lambda_2 U^T, \text{ } m \times m \text{ symmetric matrix}$$

$$\Sigma = \sqrt{\Lambda}, \text{ singular value of } A$$



제안방법

시청비율

- 시청자 u_i 가 TV 프로그램 p_j 의 k 번 째 시청했을 때 시청비율
- 더 많은 시청을 할수록 선호도가 높다는 가정

$$ratio(i, j, k) = \frac{Vet_{i,j,k} - Vst_{i,j,k}}{Pet_{j,k} - Pst_{j,k}}$$

Pst : TV 프로그램 방영 시작시간
(Program start time)

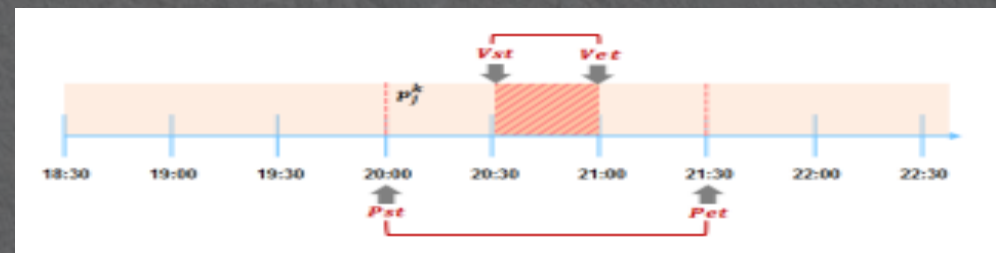
Pet : TV 프로그램 방영 종료시간
(Program end time)

Vst : TV 프로그램 시청 시작시간
(Viewing start time)

Vet : TV 프로그램 시청 종료시간
(Viewing end time)

$$R_{ratio}(i, j) = \frac{1}{N} \sum_k ratio(i, j, k)$$

시청비율 기반 선호도



제안방법

종료시간

- 시청 종료시간과 방영 종료시간의 차이가 클수록 선호도가 낮다는 가정
- 선호도 평준화를 위해 종료시간의 차를 총 방영시간으로 나눔

$$finish(i,j,k) = 1 - \frac{Pet_{i,j,k} - Vet_{i,j,k}}{runningTime(j,k)}$$

$$runningTime(j,k) = Pet_{j,k} - Pst_{j,k} : \text{프로그램 총 방영시간}$$

시청자 u_i 의 TV 프로그램 p_j 의 k 번 째의
종료시간 기반 선호도

$$R_{fin.}(i,j) = \frac{1}{N} \sum_k finish(i,j,k)$$

전체 방영 회차에 대한 종료시간 기반
선호도의 평균

제안방법

최근시청이력

- 시청횟수가 같아도 최근 회차를 많이 볼수록 높은 가중치를 줌
- 한 프로그램의 최근 5회분 중의 시청비율로 이를 계산

$$\omega_{recent}(i,j) = \begin{cases} \frac{recentCnt(i,j)}{5} & (5\text{회 이상 방영}) \\ \frac{recentCnt(i,j)}{totalCnt(i,j)} & (5\text{회 미만 방영}) \end{cases}$$

시청자 u_i 의 TV 프로그램 p_j 에 대한 최근 방영된
5회분 중 시청한 횟수를 $recentCnt(i,j)$
총 방영횟수를 $totalCnt(i,j)$ 라 했을 때

	1	2	3	4	5	6	7	8
user ₁								
user ₂								

그림 3. 최근시청이력 가중치의 예
Fig. 3. Example of $\omega_{recent}(i,j)$

최근시청이력 가중치의 예

제안방법

최종적인 선호도

- 시청비율 기반 선호도와 종료시간 기반 선호도의 합에 최근 시청이력 가중치를 줌
- 시청자 u_i 의 TV 프로그램 p_j 에 대한 최종적인 선호도

$$R(i, j) = \omega_{recent}(i, j) \times (R_{ratio}(i, j) + R_{fin.}(i, j))$$

실험

실험 환경

- TNmS가 수집한 6개월간의 3318명의 시청자 개인 시청기록
- 4개의 지상파 채널의 TV프로그램 방영 데이터
- 전처리 과정으로 10분 미만의 시청기록은 삭제

실험

- 80%를 트레이닝 데이터로 사용하여 선호도 구함
- SVD 이용해 사용자-아이템 행렬의 선호도 점수 예측
- 20%를 테스트 데이터로 사용해 가장 높은 선호도로 비교
- 검증하기위해 시청시간패턴을 고려하지않은 베이스라인 선택
- 평가척도로 Precision 사용

실험

실험 결과

Method	Precision(%)
baseline	34.7
ratio	36.6
finish-time	36.35
recent	39.27
Joint preference	39.4

Thank you