

개인화 추천을 위한 태그 가치 측정 알고리즘

정광재*, 박건우*, 이상훈**

*국방대학교 국방정보체계학과

**국방대학교 국방정보체계학과 교수

e-mail:jamhang@naver.com

Tag Value Measurement Algorithm for Personalized Recommendation

Kwang-Jae Jeong*, Gun-Woo Park*, Sang-Hoon Lee**

Dept. Defense Information System, Korea National Defense University

요 약

웹 2.0의 영향으로 인터넷 상에 범람하는 콘텐츠를 이용함에 있어 태깅 시스템은 매우 유연하고 효과적인 분류를 가능케 한다. 대부분의 웹 2.0 사이트에서는 검색된 정보에 해당하는 태그와 연관성이 있는 태그를 나타냄으로써 또 다른 관련 콘텐츠를 이용할 수 있는 서비스를 제공한다. 콘텐츠 사용자에 의해 생성되는 태그는 개인 성향에 따라 동일 콘텐츠에 다양하게 적용될 수 있으며 이로 인해 태그를 이용한 검색은 낮은 정확도를 나타낼 수 있다. 본 논문에서는 태그 선택에 있어 인간 상호작용의 특성을 파악하여 개인이 선호하고, 필요로 하는 콘텐츠에 대한 태그를 추천할 수 있는 태그 가치 측정 알고리즘을 제안한다. 콘텐츠 선택에 있어 의사결정에 영향을 미치는 요인을 식별하고 선호영화 추천 서비스인 MovieLens 사이트의 데이터 셋을 적용하여 태그 추천의 예측 정확도를 비교 평가함으로써 향상된 태그 가치 산정 결과를 제시한다.

1. 서론

폭발적인 인터넷 사용자의 증가와 콘텐츠의 다양화로 웹서비스 환경은 진보를 거듭해 왔으며 참여와 공유로 대변되는 인터넷 환경 변화의 흐름을 잘 반영하고 있는 것이 웹 2.0이다. 웹 2.0에서의 제작자와 구독자간 경계는 없다. 사용자 모두가 정보생산자가 될 수 있으며 웹 환경은 수 많은 경로를 거쳐 정보 자원의 공유를 가능케 한다. 태그는 일종의 메타데이터로서 사용자들이 생산한 문서, 영상, 음악 등의 웹 리소스에 검색 및 브라우징을 용이하게 하는 웹 2.0의 기술 중의 하나이다. 대부분의 웹 2.0 사이트에서는 검색된 정보에 해당하는 태그와 연관성이 있는 태그(연관태그)를 나타냄으로써 또 다른 관련 정보자원을 이용할 수 있는 서비스를 제공한다. 연관 태그를 통해 검색자가 의도했던 태그 외에 해당 정보를 대표할 수 있는 더 나은(개인적 선호에 맞는) 태그를 찾을 수 있다.

폭소노미(folksonomy)는 협업 태깅(collaborative tagging) 또는 소셜 태깅(social tagging)으로도 설명할 수 있다. 위키피디아의 정의를 찾아보면 “폭소노미란 태그를 협력적으로 생성하고 관리하는 행위 및 방법”을 의미한다[1]. 소셜 태깅을 주로 사용하는 대표적인 인터넷 서비스로 소셜 북마킹 사이트인 mar.gar.in, delicio.us, 사진 공유 사이트인 Flickr 등이 있다. mar.gar.in은 나의 관심 태그를 등록하면 등록된 태그를 포함하고 있는 다른 콘텐츠를 제공한다. 검색된 콘텐츠들 중 필요한 것을 내 북마크에 태그를 이용하여 저장할 수 있으며 여기서 다른 이들이 사용한 여러 태그목록을 통해 다른 콘텐츠를 검색할 수 있다[2]. 영화 추천 사이트인 MovieLens는 영화의 장르, 제작자, 배우, 제작사, 배경 등을 평가자들이 태그로 지정하며 연관성이 높다고 판단되는 태그에 대해 Thumb up/thumb

down 방식으로 평가하여 순위를 지정할 수 있다[3]. 어떠한 정보자원에 대해서 자신이 연상했던 태그와 다른 이들이 사용한 태그를 비교하여 해당 정보자원과 가장 연관성이 있는 태그를 선별하여 저장할 수 있으며 이러한 과정이 반복적으로 수행되면 해당 콘텐츠에 가장 많이 사용된 몇 개의 태그가 대표성을 갖게 된다.

이처럼 태깅 시스템은 매우 유연하고 역동적인 분류체계를 제공하지만 이로 인한 문제점 또한 배재할 수 없다.

첫째, 시스템을 사용하는 특정 사용자의 성향을 전체 사용자와 동일시 할 수 없다[4]. 예를 들어, Java라는 용어를 접했을 때, 프로그래머가 가장 먼저 떠올리는 것과 커피애호가나 여행전문가가 떠올리는 것은 분명 다르다. 성별, 나이, 배경지식, 거주지, 직업, 취미, 전공 등 개인별 특성에 따라 콘텐츠에 적용되는 태그 용어를 결정하는 다양한 Human Factor가 존재한다.

둘째, 태그를 이용한 검색은 종종 낮은 정확도를 보인다. 태그는 어떤 정보를 넓은 범주에 위치시키기는 용이하지만, 콘텐츠에 ‘악의적인 태거(malicious tagger)’에 의해 생산된 부정확한 태그들이 많이 존재하기 때문이다[5].

본 논문에서는 태그 선택에 있어 인간 상호작용의 특성을 파악하여 개인에게 가치 있는 태그를 추천할 수 있는 태그 가치 측정 알고리즘을 제안한다. 콘텐츠 선택에 있어 의사결정에 영향을 미치는 요인을 식별하고 선호 영화 추천 서비스를 제공하는 MovieLens 사이트의 데이터 셋을 적용하여 태그 추천의 예측 정확도를 비교 평가함으로써 향상된 태그 가치 산정 결과를 제시한다.

2. 관련연구

2.1 개인화 검색

웹(Web)이라는 개념의 빠른 확산과 광대한 인프라 구

축을 통해 웹 검색은 가장 손쉬운 정보 검색 방법 중 하나가 되었다. 하지만 최근까지의 웹 검색은 구글(Google)의 페이지랭크(PageRank) 등의 기법을 이용해 많은 사용자가 원하는 결과를 사용자에게 동일하게 제공하는 것이었다. 그러나 차세대의 검색은 기존의 대다수 사용자에게 대안 일률적인 결과 제공이 아닌, 각 개별 사용자가 원하는 결과를 제공하는 것을 목표로 하고 있다. 이러한 ‘개인화 검색’ 기술을 이용한 서비스는 아직 미미한 수준이지만 각종 검색 포털이나 온라인 마켓에서 경쟁적으로 사용자에게 제공되고 있다. 예를 들어 Rollyo.com의 경우 개인이 원하는 검색 엔진을 선택하여 해당 검색 엔진에서 검색된 자료를 제공해주는 서비스를 제공하고, 하단의 Epinions.com은 개인별 프로파일과 신뢰목록(Web of Trust), 상품 후기를 이용하여 개인의 기존 선호도에 맞는 상품을 추천하는 서비스를 제공한다.

개인화 검색은 개인의 행동과 요구, 선호도에 기반을 둔 맞춤 콘텐츠와 서비스를 제공하여 사용자에게 개인이 원하는 정보를 보다 신속하게 제공하기 때문에 개인 만족도가 높다. 이는 개인이 정보를 찾아가는 것이 아니라 정보가 개인을 찾아오는 것이 중요함을 의미한다[6].

2.2 협업 필터링(Collaboration Filtering)

협업 필터링은 많은 사용자들로부터 얻은 기호정보에 따라 사용자들의 관심사를 자동적으로 예측하게 해주는 방법이다. 협업 필터링 접근법의 근본적인 가정은 사용자들의 과거 경향이 미래에도 그대로 유지될 것이라는 전제에 있다. 예를 들어 음악에 관한 협업 필터링 추천 시스템은 사용자들의 기호에 대한 부분적인 목록을 이용하여 그 사용자의 음악에 대한 기호를 예측하게 된다. 이 시스템은 많은 사용자들로부터 이와 같은 기호를 수집하여 사용자의 선호도와 관심 표현을 바탕으로 비슷한 선호 패턴을 가진 사용자를 식별해낸다.

따라서 협업 필터링은 비슷한 취향을 가진 사용자들에게 서로 아직 접하지 못한 아이템을 교차 추천하거나 분류된 사용자의 취향에 따라 관련 아이템을 추천하는 형태의 서비스를 제공하기 위해 사용된다[7].

3. 가치측정 알고리즘

3.1 유사도

협업 필터링 알고리즘에서의 사용자 기반 유사 가중치는 추천 시스템에서 가장 널리 사용되며, 성능이 입증된 공분산 기반의 피어슨 상관계수를 이용하여 예측 정확도가 높은 계산법이다[8]. 사용자 x 와 y 의 유사도 계산을 위한 피어슨 상관계수 $T_{x,y}$ 는 수식1과 같다[6].

$$T_{xy} = \frac{\sum_{t \in S_{xy}} (r_{x,t} - \bar{r}_x)(r_{y,t} - \bar{r}_y)}{\sqrt{\sum_{t \in S_{xy}} (r_{x,t} - \bar{r}_x)^2} \sqrt{\sum_{t \in S_{xy}} (r_{y,t} - \bar{r}_y)^2}} \quad [수식 1]$$

- $r_{x,t}$: 사용자 x 가 태그 t 에 부여한 평점
- \bar{r}_x : 사용자 x 가 태그에 부여한 평점 평균
- $r_{y,t}$: 사용자 y 가 태그 t 에 부여한 평점
- \bar{r}_y : 사용자 y 가 태그에 부여한 평점 평균

유사도는 특정 사용자와 동일한 태그를 평가한 사용자들의 평점평균인 \bar{U}_t 합하여 최종적으로 예측 평점 \bar{U}_t 를 구하게 되며, [수식1]과 같다.

$$U_t = \bar{U}_t + \frac{\sum_{j \in raters} (J_t - \bar{J}) T_{uj}}{\sum_j |T_{uj}|} \quad [수식 2]$$

- U_t : 사용자 u 의 태그 t 에 대한 예측 평점
- \bar{U}_t : 태그 t 를 평가한 사용자들의 평점 평균
- J_t : 태그 t 에 대한 사용자의 평점
- \bar{J}_t : 사용자 u 의 태그 평점 평균
- $T_{u,j}$: 사용자 u 와 u 의 유사도

3.2 알고리즘 구성요소

3.2.1 태그 선호도

어떠한 정보자원을 검색하고자 할 때 개인의 배경지식과 필요분야에 대한 정보를 바탕으로 특정 키워드를 선택하여 검색을 수행한다. 여기서 단일 검색 활동에 대한 과정은 큰 의미가 없을지 모르나, 장기간 동안 여러 자료에 대한 검색 패턴은 큰 의미를 가질 수 있다. 검색자가 특정 태그를 포함하는 콘텐츠를 자주 선택한다면 그 검색자는 특정 태그를 포함하는 콘텐츠에 대해 선호도가 높다고 표현할 수 있다.

태그가 적용되어 있는 영화와 상호작용에 기반한 사용자의 태그 선호도 알고리즘을 계산한다. 이 때, 사용자에 따른 태그와 영화와의 상대적인 연관성을 고려해야 한다. 영화별 태그 적용 빈도를 산출하기 위해 $tag_quality$ 를 정의한다. 「타이타닉」이라는 영화에 적용된 50개의 태그 중 「재난」이라는 태그가 8개(최다)가 있다면 이 「재난」이라는 태그의 $tag_quality$ 는 $8/50 = 0.16$ 이 된다. 적용 태그 개수별 변수의 차이와 동일 태그에 대한 평가자들의 상대성을 보완하기 위하여 오일러 수를 이용하여 가중치를 산정한다[9].

영화 m 에 적용된 태그 t 에 대한 $tag_quality$ 에 대한 가중치와 태그 평가값에 대한 가중치를 합산하여 태그 선호 연관가중치 $w(m,t)$ 를 산출한다.

$$w(m,t) = \frac{1}{e^{-tag_quality}} + \frac{1}{e^{-rating}} \quad [수식 3]$$

$r_{j,m}$ 을 영화 m 에 대한 사용자 u 의 평가 값, tag가 적용된 영화 전체를 M_t 라 하면, 태그 선호도 $F_{m,t}$ 는 다음과 같다.

$$F_{m,t} = \frac{\sum_{m \in M_t} w(m,t) \cdot r_{u,m}}{\sum_{m \in M_t} w(m,t)} \quad [수식 4]$$

각 합계에서 사용자가 평가하지 않은 영화 항목의 값은 무시한다.

3.2.2 태그 관심도

관심도란 사용자의 기존 태그 선택 패턴 분석을 통하여 새로운 아이템에 대한 관심 정도를 예측한 값이다. 개인의 정보 선택에 따른 기존 지식으로 인해 온라인 사용자는 자신의 직업, 취미활동 등 자신과 연관성 있는 분야에 대해 자주 검색하는 경향이 있다[4]. 물론 자신의 특성과 연관 없는 분야에 대한 검색행위도 있겠지만 지속성이나 빈도를 고려할 때 주 관심 분야의 추천은 그만큼 사용자에게 유용할 확률이 높다. 이런 의미에서 관심도는 추천 시스템의 중요한 요인이 될 수 있다. 사용자가 평소에 관심을 가지고 보았던 분야는 그렇지 않은 분야보다 좀 더 높

은 가중치를 가질 수 있는 것이다. 사용자는 여러 분야에 관심을 가지고 있으나, 각 분야마다 관심 정도가 다르다. 여기서 관심도 I_t 는 전체 평가 태그에 대한 관심 태그의 비율이다. 태그 평점범위 1~5점 중 4점 이상이 만족을 나타내므로 $I_{u,h}$ 는 평점 4점 이상이 적용된 태그를 적용한다.

$$I = \frac{I_{u,h}}{I_{u,tot}} \quad \text{[수식 5]}$$

- $I_{u,h}$: 사용자 u 의 4점 이상의 평점 태그 수
- $I_{u,tot}$: 사용자 u 의 총 평가 태그 수

3.2.3 태그 인기도

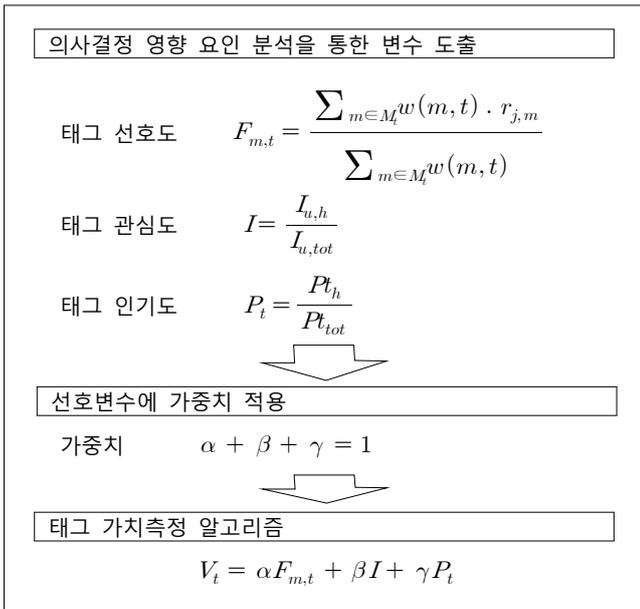
인기도는 많은 사람이 선호하는 정도를 의미한다. 인기가 높다는 것은 많은 사용자들에게 인정을 받았다는 의미이며, 그만큼 새로운 사용자도 높은 관심을 가지고 신뢰할 확률이 높다. 결국 인기도는 온라인 준거집단의 영향력을 반영하는 것으로, 아이টে에 대한 정보가 부족한 새로운 사용자는 인기가 높은 아이টে에 선택한 준거집단의 영향을 받아 그 아이টে에 더욱 큰 관심을 가지고 신뢰할 확률이 높다는 것이다.

태그에 대한 인기도 P_t 는 특정 태그 t 가 얼마나 많은 사용자에게 높은 평가를 받았는지(얼마나 인기 있는지) 나타내는 값으로, 여기서 P_{tot} 는 태그 t 가 평가된 총 횟수이며 P_{t_h} 는 4점 이상의 평점이 적용된 태그 t 의 수를 나타낸다.

$$P_t = \frac{P_{t_h}}{P_{tot}} \quad \text{[수식 6]}$$

3.3 알고리즘 설계

태그 가치측정 알고리즘은 3.2절의 세 가지 요소를 고려하여 작성한다. 태그 가치는 V 로 나타내며, 이 때 각 요소에는 합이 1인 가중치가 각각 적용되어 사용자 u 의 태그 t 에 대한 예측 평점인 V_t 가 계산된다. 설계 프로세스는 다음과 같다.



4. 성능 평가 및 비교 분석

4.1 데이터 셋

본 연구에서는 MovieLens 웹 사이트에서 수집된 데이터를 사용하였다. MovieLens는 사용자가 자신이 평가한 다양한 영화에 대한 평가에 따라 선호 예상 영화를 추천하는 서비스를 제공한다[9]. 실험을 위해 MovieLens를 통해 수집된 사용자 평가 데이터와 E-mail 설문문을 통한 태그 선호도 평가 데이터를 수집하였다.

MovieLens 평가 데이터는 71,556명의 사용자가 10,682개의 영화에 대해 평가한 10,000,054개의 평점과, 95,580회 적용된 태그 목록을 저장하고 있으며, 사용자 일인당 최소 20개의 영화에 대해 평가가 수행되었다. 실제적인 평가는 MovieLens 사이트를 통해 이루어졌다[11]. E-mail 설문에서는 각각의 사용자에게 태그 목록을 제공하고 최소 50개 이상 영화에 대한 태그의 연관성과 태그 선호도를 별점 1에서 5로 평가하도록 요청했다. 불성실한 응답을 제외한 총 530개의 설문결과를 토대로 실험을 수행하였다. 영화에 대한 선호도 평점은 1점 6.1%, 2점 10.8%, 3점 24.4%, 4점 33.3%, 5점 23.7%, 무응답 1.7%이며, 태그에 대한 선호도 평점은 1점 9.8%, 2점 20.3%, 3점 29.7%, 4점 21.1%, 5점 16.4%, 무응답 2.7%로 분포하고 있다.

4.2 평가 방법

본 연구에서는 절대평균 오차(MAE : Mean Absolute Error)를 이용하여 예측 성능을 평가한다. 절대평균 오차는 예측 정확도를 평가하기 위해 추천 시스템 검증에 위한 많은 연구에서 사용되는 방법으로, 사용자의 실제 평점과 예측 평점간의 절대 편차의 평균으로 계산되며, [수식 9]와 같다[7].

$$MAE = \frac{\sum_{x=1}^N |U_x - r_x|}{N} \quad \text{[수식 7]}$$

(단, $0 \leq MAE \leq N$)

- U_x : 예측 평점
- x : 사용자 u 와 유사도가 높은 사용자
- r_x : 실제 평점
- N : 사용자 u 와 유사도가 높은 총 사용자 수

MAE를 통한 예측 정확도는 MAE와 반비례한다. MAE가 낮을수록 예측 정확도가 높은 것이므로 태그 가치측정 알고리즘 적용 결과가 단순 Summation보다 높은 예측 정확도를 가짐을 입증한다.

태그 가치 측정 알고리즘을 수행하기 위한 선결과제는 사용자를 선정하고, 그 사용자와 동일한 태그를 평가한 경험에 있는 사용자들을 추출하는 것이다. 이 때 유사도가 높은 사용자간의 추천이 보다 효율적이기 때문에, 추출된 사용자 중 피어슨 상관계수를 계산하여 유의수준이 0.05 레벨 이상인 유사 사용자들을 선별한다.

T_{ij}	U_i	r_j	R_x	$h - j$	$(h - j) * T_{ij}$	$ T_{ij} $	h	$ U_i - r_x $	$\sum U_i - r_x $
0.489	3.150	3.000	3.450	-0.250	-0.122	0.489	3.250		4.328
0.571		4.000	3.050	0.750	0.428	0.571		0.672	
0.372		4.000	3.150	0.750	0.279	0.372	$\sum T_{ij} $		
-0.451		3.000	2.950	-0.250	0.113	0.451	8.938		
0.763		3.000	2.800	-0.250	-0.191	0.763		0.328	
-0.445		4.000	2.600	0.750	-0.334	0.445	$\sum (h - j) * T_{ij}$		
-0.129		3.000	3.200	-0.250	0.032	0.129	1.594		
0.186		3.000	3.050	-0.250	-0.047	0.186			
0.276		3.000	3.000	-0.250	-0.069	0.276	U_i		
0.313		2.000	2.700	-1.250	-0.391	0.313	3.328		
0.551		3.000	3.150	-0.250	-0.138	0.551		0.328	
-0.403		1.000	2.500	-2.250	0.907	0.403			0.328
0.637		3.000	3.600	-0.250	-0.159	0.637			0.328
0.301		2.000	3.100	-1.250	-0.376	0.301			
0.673		3.000	3.700	-0.250	-0.168	0.673			0.328
0.188		5.000	3.500	1.750	0.329	0.188			
-0.207		3.000	3.050	-0.250	0.052	0.207			
-0.434		4.000	2.850	0.750	-0.326	0.434			
0.936		4.000	2.900	0.750	0.702	0.936		0.672	
0.613		5.000	2.800	1.750	1.073	0.613		1.672	

표. 1 유의수준을 고려한 예측 평점 계산

이렇게 선별된 사용자에게 대해 각 분류별로 예측 평점과 MAE를 계산한다.

4.3 알고리즘 비교 분석

4.3.1 알고리즘을 이용한 평점 예측

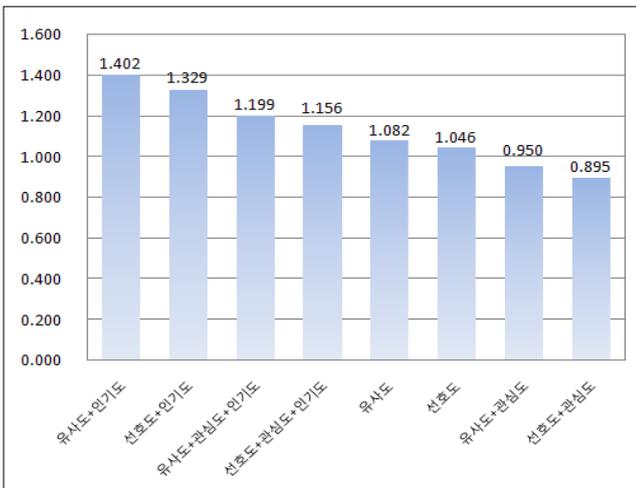
알고리즘을 이용한 평점을 예측하기 전에 태그 선호도, 태그 관심도, 태그 인기도의 영향력을 알아보기 위한 분류를 시행한다.

- ① 유사도
- ② 유사도 + (태그)관심도
- ③ 유사도 + (태그)인기도
- ④ 유사도 + (태그)관심도 + (태그)인기도
- ⑤ (태그)선호도
- ⑥ (태그)선호도 + (태그)관심도
- ⑦ (태그)관심도 + (태그)인기도
- ⑧ (태그)선호도 + (태그)관심도 + (태그)인기도

위 분류에 따라 데이터 셋을 적용하여 예측 평점을 계산하고, 그 결과를 비교함으로써 가치 측정 알고리즘 각 요소의 영향력과 정확도를 분석할 수 있다.

4.3.2 예측 정확도 비교

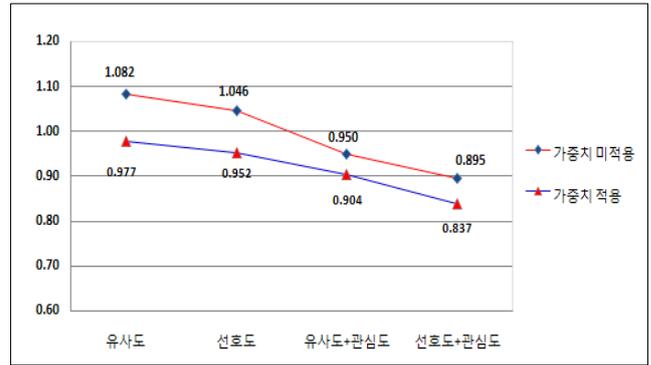
그림 1은 요소별 가중치를 적용하지 않은 요소별 조합에 따른 MAE 값을 보여준다.



(그림 1) 평가 요소 조합에 따른 MAE 값

인기도 항목이 적용된 알고리즘의 MAE가 상대적으로 높음을 알 수 있다. 이는 태그 평가에 있어서 인기도의 영향이 높지 않음을 나타낸다. 일반적으로 준거집단의 영향을 받는 온라인 마켓에서의 판매물품에 있어서의 인기도는 아이템 추천의 고려 요소가 되나 콘텐츠의 작성자 또는 관리자에 의해 생성되는 태그는 개인적 성향이 반영되기 때문에 인기도 보다는 선호분야의 영향력이 높음을 알 수 있다. SPSS Clementine 10.1을 이용 각 요소별 가중치 계산 결과 α, β 의 값은 각각 0.891, 0.109이며 γ 의 값은 10^{-5} 이하의 극히 적은 값이 산출되었다. 유사도+관심도의 조합에 있어 가중치는 각각 0.82, 0.18이다. 이에 따라 제안 알고리즘에서 인기도를 배제한 유사도, 관심도, 선호도의 조합을 통해 예측 정확도를 계산한다.

그림 2는 인기도를 제외한 요소들의 가중치 적용 전과 적용 후의 MAE 값을 나타낸 것이다. 요소별 가중치를 적용 후 예측 정확도에 대한 MAE 값은 전체적으로 감소하였다.



(그림 2) 가중치를 적용한 요소별 MAE 값

선호도 및 관심도 조합의 예측 정확도에 대한 MAE는 0.837로 단일 요소의 예측정확도보다 높음을 알 수 있다. 즉, 두 요소를 모두 고려한다면 보다 정확한 개인화된 태그 추출이 가능할 것으로 판단된다.

5. 결론 및 향후 발전방향

본 연구에서는 콘텐츠 선택에 있어서의 인간의 행위특성을 고려한 가치 측정 알고리즘을 설계하였다. 선호도, 관심도, 유사도에 대한 고려요소에 MovieLens 데이터 셋과 태그 평점 설문 데이터를 이용하여 각 요인에 대한 영향력을 알아보았다. 태그의 적용에 있어서의 실험 결과 선호도와 관심도의 조합이 가장 높은 예측 정확도를 나타낼 수 있었다. 본 논문에서 사용한 데이터 셋에서는 "animation", "animation_action", "animation_pixar"와 같이 동일 단어가 중복된 태그를 단일 태그로 계산하였다. 향후에는 단어가 중복 적용된 태그에 대해서는 유사정도에 따라 태그들을 군집하는 방법에 대한 연구가 필요하며, 또한 태그 선택에 있어서의 추가적인 영향력 요인에 대한 발굴 및 본 논문에서 다룬 영화 이외에 다양한 콘텐츠에 적용된 태그에 대한 연구를 통해 각 아이템별 태그 추천에 대한 알고리즘 구현이 요구된다.

참고문헌

- [1] www.wikipedia.org
- [2] mar.gar.in
- [3] www.movielens.org
- [4] Jamie Teevan, Meredith R. Morris, Steve Bush "Discovering and Using Groups to Improve Personalized Search", Microsoft Research, 2009
- [5] G. Koutrika, F. Effendi, Z. Gyöngyi, P. Heymann, and H. Garcia-Molina. "Combating spam in tagging system." In Proceedings of the 3rd international workshop on Adversarial information retrieval on the web, pages 57-64. ACM Press New York, NY, USA, 2007.
- [6] 피터 모빌, "검색 2.0 발견의 진화", Yuna 역, 한빛미디어, 2006
- [7] 조성훈, "웹 환경에서 신규 사용자와 추천 성능 향상을 위한 상품 품질 기반의 가중치 추천 기법", 한남대 박사 논문, 2008.
- [8] 이종석, 권준범, 전치혁, "고객-제품 구매여부 데이터를 이용한 협동적 필터링에서의 유사성 척도의 사용", 한국경영과학회 춘계 학술발표대회, SB7-9, 2004.
- [9] <http://www.grouplens.org/node/73#attachments>