

그룹 추천에서 사용자 선호도의 편차를 고려한 그룹 모델링 전략

(A Group Modeling Strategy Considering Deviation of the
User's Preference in Group Recommendation)

김 형 진 ^{*} 서 영 덕 ^{*} 백 두 권 ^{**}
(HyungJin Kim) (Young-Duk Seo) (Doo-Kwon Baik)

요 약 그룹 추천은 개인이 아닌 그룹의 특성 및 성향을 분석하여 구성원들에게 적합한 정보를 제공하는 추천 방식이다. 기존의 그룹 추천 방식은 평균 선호도나 선호 횟수에 기반한 그룹 모델링 전략을 사용한다. 하지만 평균이 높고 선호 횟수가 많은 관심사더라도 선호도의 편차가 크다면, 그룹 내 구성원 모두를 만족시키는 추천 결과를 제공하기가 어렵다. 본 논문에서는 이를 개선하고자 관심사에 대한 평균 선호도에 선호도 편차를 가중치로 하는 그룹 모델링 전략을 제안한다. 제안하는 방법은 평균 선호도가 높으면서 선호도 편차가 작은 관심사들을 추천 결과로 제공해줌으로써 기존의 그룹 모델링 전략보다 더 많은 그룹 내 구성원들을 만족시키는 정보를 제공하는 것이 가능하다. 실험을 통해 제안하는 그룹 모델링 전략이 기존의 방식에 비해 높은 성능을 보였고, 소규모의 사용자뿐만 아니라 많은 수의 사용자가 형성하는 그룹에서도 높은 성능을 가짐을 확인하였다.

키워드: 추천 시스템, 그룹 추천, 그룹 모델링, 그룹 모델링 통합 전략, 선호도 편차

Abstract Group recommendation analyzes the characteristics and tendency of a group rather than an individual and provides relevant information for the members of the group. Existing group recommendation methods merely consider the average and frequency of a preference. However, if the users' preferences have large deviations, it is difficult to provide satisfactory results for all users in the group, although the average and frequency values are high. To solve these problems, we propose a method that considers not only the average of a preference but also the deviation. The proposed method provides recommendations with high average values and low deviations for the preference, so it reflects the tendency of all group members better than existing group recommendation methods. Through a comparative experiment, we prove that the proposed method has better performance than existing methods, and verify that it has high performance in groups with a large number of members as well as in small groups.

Keywords: recommender system, group recommendation, group modeling, aggregation strategy of the group modeling, deviation of the preference

· 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대 정보·컴퓨팅기술개발사업의 지원을 받아 수행된 연구임(No.2012M3C4A7033346)

^{*} 학생회원 : 고려대학교 컴퓨터학과
flslzk@korea.ac.kr
seoyoungd@korea.ac.kr

^{**} 종신회원 : 고려대학교 컴퓨터학과 교수(Korea Univ.)
baikdk@korea.ac.kr
(Corresponding author임)

논문접수 : 2016년 6월 23일
(Received 23 June 2016)
논문수정 : 2016년 7월 13일
(Revised 13 July 2016)
심사완료 : 2016년 7월 14일
(Accepted 14 July 2016)

Copyright©2016 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제43권 제10호(2016. 10)

1. 서론

추천 시스템에 대한 연구는 사용자 개인의 선호 정보를 기반으로 유사한 성향을 갖는 사용자들을 통해 적합한 정보를 제공하는 개인화 추천(Personalized Recommendation)이 주로 연구되고 있다[1]. 하지만 데이터가 나날이 급증하는 온라인 환경에서 사용자 개개인의 모든 성향을 파악하는 것은 다량의 오버헤드(Overhead)를 발생시키므로, 이를 실제 시스템이나 서비스에 적용하기엔 많은 비용(Cost)이 요구된다는 문제점이 있다. 또한 영화, 음악, 여행 등 개인의 관심사를 공유하기 위해 온라인 커뮤니티와 같은 그룹 활동이 꾸준히 증가하는 추세를 보이고 있어, 개인보다는 그룹을 대상으로 한 추천 연구가 필요하다[2,3].

그룹 추천(Group Recommendation)은 개인이 아닌 그룹의 특성 및 성향을 분석하여 구성원들에게 적합한 정보를 제공하는 추천 방식으로 Poly-Lens[3], MusicFx[4], Travel Decision[5] 등의 다양한 시나리오를 기반으로 활발히 연구가 진행되고 있다[2]. 이는 그룹 탐색(Group Detection), 그룹 모델링(Group Modeling), 그룹 선호도 추정(Group Preference Estimation)의 세 과정을 통해 이루어진다[1,6]. 그룹 탐색은 온라인 상에서 직접적인 소셜 관계를 갖거나 유사한 선호 정보를 지니는 사용자들을 추천의 대상인 그룹으로 형성하는 과정이다. 그룹 모델링은 그룹 탐색을 통해 형성된 그룹 내 구성원들의 선호도를 통합하는 단계로, 그룹에 적합한 정보를 찾기 위한 전략을 수립한다. 마지막으로 그룹 선호도 추정에서는 앞선 과정을 통해 형성된 그룹과 그룹 모델링 전략에 의해 생성된 정보에 기반하여 정규화된 그룹의 선호도를 측정한다.

그룹 추천을 위해서는 그룹 모델링을 위한 전략을 수립하는 것이 가장 중요하며, 기존의 그룹 모델링 연구에서는 관심사에 대한 구성원들의 선호 정보를 통합하는 방식을 주로 사용한다. 이를 통합 전략(Aggregation Strategy)라 하며, 통합 전략으로는 선호도를 모두 합하는 방식, 평균 선호도를 측정하는 방식, 선호 횟수를 세는 방식, 최소 선호도를 찾는 방식 등이 제안되고 있다[2-5,7-9]. 이러한 통합 전략들은 주로 평균 선호도가 높거나 선호 횟수가 많은 관심사를 찾아, 이를 적합한 정보로 그룹에 제공한다. 하지만 추천된 결과들이 높은 평균 선호도나 선호 횟수를 갖더라도 선호도의 편차가 크다면, 이는 그룹 내 구성원 모두를 만족시킬 수 있는 구성원 개개인의 관심사를 제대로 반영한 그룹 모델링 전략이라 할 수 없다.

선호도와 선호 횟수에 의존하는 기존의 전략에서 벗어나기 위해 관심사들에 대한 선호 순위를 점수화하여

모두 합하는 방식과 특정 임계값(Threshold) 이상의 선호 횟수를 세는 방식, 평균 선호도를 측정하되 구성원들에게 선호되지 않는 관심사를 제외하는 방식이 제안되었다[4,8,9]. 하지만 여전히 선호도와 선호 횟수에 대한 의존도가 높거나 극단적인 경우의 편차만 고려하는 방식이기 때문에 그룹 내 구성원 전체를 고려하는 전략이라 할 수 없다. 따라서 모든 구성원들의 관심사를 반영하기 어렵고 많은 수의 사용자들로 구성된 그룹에서는 추천의 성능이 떨어진다.

본 논문에서는 기존의 그룹 모델링 과정에서 사용되는 통합 전략들이 고른 분포로 높은 선호도를 갖는 관심사를 고려하지 않는 문제점을 개선하고자, 관심사에 대한 평균 선호도에 선호도 편차를 가중치로 하는 전략을 제안한다. 제안하는 전략은 그룹 내에서 구성원들의 관심사에 따라 변하는 선호도의 편차를 고려하여 평균 선호도가 높으면서 선호도 편차가 작은 관심사들을 추천 결과로 제공하기 때문에, 기존 그룹 모델링 방식에 비해 그룹 내 구성원 개개인의 성향을 반영하여 보다 적합한 정보를 제공하는 것이 가능하며, 많은 수의 사용자가 형성하는 그룹에서도 높은 성능을 갖는다.

이 논문의 구성은 다음과 같다. 제 2장에서는 그룹 추천에 대한 간략한 설명 및 일반적인 그룹 추천 프로세스에 대한 소개를 하며, 특히 그룹 모델링에 대한 기존의 전략들을 소개하는데 초점을 맞춘다. 제 3장에서는 그룹 추천에서 자주 사용되는 Movie-Lens 데이터에 대한 분석과 분석된 결과를 통해 본 논문의 연구배경에 대해서 언급한다. 제 4장에서는 본 논문이 제안하는 그룹 모델링 전략을 적용한 그룹 추천 방법을 기술하고, 제 5장에서 이를 기존의 모델링 전략들과 비교하여 실험하고 평가한 내용을 언급한다. 마지막으로 제 6장에서는 본 논문의 결론과 향후 연구를 제시한다.

2. 관련 연구

그룹 추천은 그룹의 특성 및 성향을 분석하여 구성원들에게 적합한 정보를 제공하는 추천 방식으로 웹/뉴스[10], 여행[5,11], 음악[4,12], TV 및 영화[3,13] 등의 다양한 분야에서 연구가 진행되고 있다[2]. 이는 추천의 대상인 그룹을 찾는 그룹 탐색, 그룹에 적합한 정보를 제공하기 위한 전략을 세우는 그룹 모델링, 마지막으로 앞선 과정을 통해 형성된 그룹과 그룹 모델링 전략에 의해 생성된 적합한 정보에 기반하여 정규화된 그룹 선호도를 측정하는 그룹 선호도 추정의 세 과정으로 이루어진다.

2.1 그룹 탐색

그룹 탐색은 추천의 목적에 따라 온라인 상에서 직접적인 소셜 관계를 갖거나 유사한 선호 정보를 지니는

사용자들을 KNN(K-Nearest Neighbor) 혹은 K-Means와 같은 클러스터링 기법을 통해 분류하여 찾는 과정이다.

[14]는 Movie-Lens에서 임의의 사용자들을 선택해 그룹으로 연결하였으며, [15]는 Sina¹⁾에서 4개 토픽(Topic)에 대한 500개의 블로그(Blog)를 기반으로 172개의 소셜 관계로 연결된 21명의 사용자들을 그룹으로 보았다. [6]과 [16]은 Movie-Lens에서 선호 정보의 유사성을 기반으로 그룹을 탐색했는데 [6]은 K-Means를 활용하여 유사성이 높은 사용자들을 그룹으로 형성하며, [16]는 Pearson Correlation Coefficient를 활용하여 높은 유사성을 갖는 상위 33%의 사용자 쌍들을 사용해 그룹을 형성했다.

2.2 그룹 모델링

그룹 모델링은 그룹에게 제공하기 위한 적합한 정보를 생성하는 과정으로 개인화 추천 방식에 기반하여 구성원 개개인의 예측 선호 정보를 측정하고 이를 통합하는 방식[3,6], 하나의 단일 그룹을 기반으로 그룹 예측 선호 정보를 측정하는 방식[5], 구성원들의 선호 정보를 통합해 그룹의 예측 선호 정보를 측정하는 방식[4,16]의 세가지 접근법이 제안되고 있다[2,7]. 하지만 개인화 추천 방식에 기반하여 측정된 구성원의 예측 선호 정보를 통합하는 방식은 그룹에 대한 연구라기보다 개인화 추천을 확장한 방식으로 볼 수 있다. 또한 이는 개인화 추천과 같이 개개인의 예측 선호도를 각각 측정해야 하므로 다량의 오버헤드가 발생하며, 추천의 결과 또한 좋은 성능을 보이지 않는다[2]. 단일 그룹 기반의 그룹 추천 방식은 그룹 내 서브 그룹(Sub Group)들의 선호 정보를 통합해 그룹의 예측 선호 정보로 측정하지만, 추천 성능이 좋은 효율을 갖지는 않는다[2,7]. 그룹 추천에서는 일반적으로 구성원들의 선호 정보를 통합해 그룹의 예측 선호 정보를 측정하는 접근법에 기반하며, 이는 통합 전략이라는 기법들을 통해 이루어진다.

통합 전략으로는 선호도를 모두 합하는 Additive Utilitarian[5], 선호도를 모두 곱하는 Multiplicative Utilitarian[8], 평균 선호도를 측정하는 Average[8], 선호 횟수를 세는 Approval Voting[9], 최소 선호도를 찾는 Least Misery[3], 최대 선호도를 찾는 Most Pleasure[8], 관심사들에 대한 선호 순위를 점수화하여 모두 합하는 Borda Count[8]과 단순히 선호 횟수를 세는 것이 아니라 특정 임계값 이상의 선호 횟수를 세는 Approval Voting with Threshold[9], 평균 선호도를 측정하되 구성원들에게 선호되지 않는 관심사를 제외하는 Average without Misery[4] 등이 사용된다. 다음 표 1은 1점에서 5점 사이의 선호도(선호도가 없는 경우 \bullet 로 표시)로 표현된

표 1 그룹 모델링을 위한 통합 전략의 예

Table 1 Example of aggregation strategy for group modeling

User \ Item	i_1	i_2	i_3	i_4	i_5
u_1	2	3	5	4	\bullet
u_2	3	2	5	4	3
u_3	2	\bullet	1	3	4
Additive Utilitarian	7	5	11	11	7
Multiplicative Utilitarian	12	6	25	48	12
Average	2.33	2.5	3.67	3.67	3.5
Approval Voting	3	2	3	3	2
Approval Voting (Threshold 4)	0	0	2	2	1
Least Misery	2	2	1	3	3
Most Pleasure	3	3	5	4	4
Borda Count	2.5	1	6	6	4.5
Average without Misery				3.67	3.5

구성원 u_1, u_2, u_3 의 관심사 i_1, i_2, i_3, i_4, i_5 에 대한 통합 전략들의 예를 보인다.

Additive Utilitarian 전략은 각 관심사들에 대한 구성원의 선호도를 모두 합하는 것으로, i_3 과 i_4 를 예로 들면 각각 $5+5+1=11$ 과 $4+4+3=11$ 이 된다. Multiplicative Utilitarian 전략은 모든 선호도를 곱하는 것으로, i_1 과 i_5 는 선호 횟수는 다르지만 $2 \times 3 \times 2 = 12$ 와 $3 \times 4 = 12$ 의 동일한 선호도로 측정된다. Average 전략은 선호도의 평균을 측정하는 것으로, i_2 를 예로 들자면 $(3+2)/2=2.5$ 가 된다. Approval Voting 전략은 아이টে에 대한 구성원의 선호 횟수를 측정하는 것으로, 임계값이 존재하지 않는 경우 i_2 와 i_3 은 선호 횟수에 따라 각각 2와 3의 선호도로 측정되며, 임계값이 4인 경우 i_2 는 4점을 넘는 구성원이 없기 때문에 0이 되고 i_3 은 4점을 넘는 구성원이 2명이기 때문에 2가 된다. Least Misery 전략과 Most Pleasure 전략은 각각 최소 선호도와 최대 선호도를 그룹의 선호도로 갖는다. i_1 의 선호도 2, 3, 2 중 Least Misery 전략의 그룹 선호도는 2가 되며, Most Pleasure 전략의 그룹 선호도는 3으로 측정된다. Borda Count 전략은 구성원 개개인의 관심사에 대한 선호 순위를 수치화하여 측정하는 것으로, 가장 낮은 선호도를 갖는 관심사를 0점으로하고 순위가 높아질수록 1점을 더한다. Borda Count 전략에 따라 u_1 의 선호도를 변환하면 0, 1, 3, 2, \bullet 이 되고, u_2 는 1.5, 0, 3, 2, 1.5로, u_3 은 1, \bullet , 0, 2, 3으로 변환되며, 이를 통해 i_1 은 2.5의 선호도로 측정된다. Average without Misery는 구성원의 부정적인 선호도 값을 갖는 관심사를 제외하는 방식으로, 1점과 2점이 존재하는 i_1, i_2, i_3 를 제외한 i_4, i_5 의 평균 선호도만을 그룹 선호도로 측정한다.

1) <http://www.sina.com>

통합 전략은 그 속성에 따라 오직 선호도만을 고려하는 전략, 선호 횟수가 중요시되는 전략, 앞선 두 전략을 개선하기 위한 전략으로 나눌 수 있다. 오직 선호도만을 고려하는 전략은 Average, Least Misery, Most Pleasure가 있고, 선호 횟수가 가장 중요시되는 전략은 Additive Utilitarian, Multiplicative Utilitarian, Approval Voting이 있다. 하지만 평균 선호도가 높거나 선호 횟수가 많은 관심사라 하더라도 선호도의 분포나 편차가 큰 경우에는 그룹 내 모든 구성원들을 만족시키는 추천 결과를 제공해주지는 못한다. 따라서 이러한 전략들은 구성원 개개인의 성향을 제대로 반영하는 그룹 추천 방식이라 하기 어렵다.

앞선 전략들을 개선하기 위해 Borda Count와 Approval Voting with Threshold, Average without Misery가 등장하였지만, Borda Count과 Approval Voting with Threshold는 여전히 선호 횟수에 대한 의존도가 높으며, Average without Misery는 극단적인 경우의 편차만 고려하기 때문에 선호도 편차의 고른 정도는 측정할 수 없다.

본 논문에서는 기존의 그룹 모델링 과정에서 사용되는 통합 전략들이 그룹 내 구성원 개개인의 관심사에 대한 선호도의 분포와 편차를 고려하지 않는 문제점을 개선하고자, 관심사에 대한 평균 선호도와 선호도 편차를 모두 고려한 전략을 제안한다.

3. 그룹 추천에서 선호도 편차의 중요성

Movie-Lens는 Group Lens에서 제공하는 그룹 추천에서 자주 사용되는 데이터 집합으로 영화에 대한 사용자들의 선호도가 기록되어있다. 선호도는 1점에서 5점 사이의 값으로 표현되며 5점과 4점은 긍정, 3점은 보통, 2점과 1점은 부정이라 할 수 있다[17]. 개인화 추천에서는 각 영화나 상품에 대한 개별 사용자들의 평점으로 긍정과 부정을 나눌 수 있지만, 그룹 추천에서는 그룹 내 모든 구성원들의 평점을 기반으로 해당 영화나 상품

표 2 그룹 내 구성원들의 영화에 대한 평균 선호도와 표준 편차

Table 2 The average and the standard deviation of the preference for the movie by group members

User \ Movie	i_1	i_2	i_3	i_4	i_5
u_1	2	3	5	4	•
u_2	3	2	5	4	3
u_3	2	•	1	3	4
Average	2.33	2.5	3.67	3.67	3.5
Standard Deviation	0.58	0.71	2.31	0.57	0.71

에 대한 긍정과 부정을 판단해야 한다.

영화나 관심사에 대한 구성원들의 선호도는 구성원의 성향에 따라 달라질 수 있기 때문에, 그룹 내에서 평균적으로 선호도가 높은 영화나 관심사더라도 그 영화나 관심사를 부정적으로 생각하는 구성원들이 존재할 수 있다. 표 2는 실제 Movie-Lens 데이터 상에서 u_1 , u_2 , u_3 을 사용자 집합으로 하는 그룹 G_1 의 5개의 영화에 대한 평균과 표준 편차를 측정하여, 이와 같은 상황이 발생할 수 있음을 보인다. 표 2를 살펴보면 영화 i_3 과 i_4 는 가장 높은 평균 선호도를 갖는다. 하지만 i_3 의 경우 사용자 선호도의 분포가 i_4 에 비해 넓기 때문에 표준 편차가 더 크다. 그룹 추천에서는 그룹 내의 모든 구성원들에게 만족할만한 결과를 제공하는 것이 더 중요하기 때문에, G_1 에서는 i_3 보다는 i_4 를 더 높은 순위의 추천 결과로 제공해야 한다.

전체적인 Movie-Lens 데이터 상에서도 이러한 경향이 나타나는지 확인하기 위해, 그림 1과 같이 영화에 대한 평균 선호도를 측정하여 긍정(3점 초과), 보통(3점), 부정(3점 미만)으로 나누고, 전체 사용자 중 몇 % 비율의 사용자가 평균 선호도와 다른 선호도를 갖는지를 확인한다. 이를 위해 Movie-Lens 100K와 1M의 사용자 중 80%는 각 영화들의 평균 선호도를 측정하기 위한

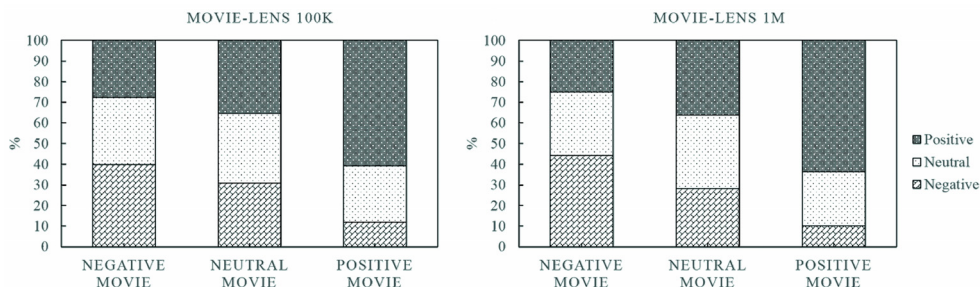


그림 1 영화들의 평균 선호도에 따른 사용자들의 실제 선호도 비율
Fig. 1 Users' preference according to the average preference for movies

훈련 집합(Training Set)으로, 20%는 각 영화에 대한 사용자들의 실제 선호도 비율을 측정하기 위한 실험 집합(Test Set)으로 나누었다. 또한 통계적으로 유의한 측정 결과를 위해 실험 집합과 훈련 집합을 달리하여 5회 반복해 측정하였다.

그림 1의 결과를 통해 평균적으로 긍정적인 평가를 받은 영화라 하더라도 약 11%의 사용자는 해당 영화를 선호하지 않음을 알 수 있다. 또한 평균적으로 부정적인 평가를 받은 영화의 경우도 약 26%의 사용자는 해당 영화를 긍정적으로 평가한다. 표 2와 그림 1은 그룹 내에서 평균적으로 높은 선호도를 갖는 관심사(영화)라 할 지라도 선호도의 편차가 크다면 구성원 모두를 만족시키는 관심사라 하기 어렵다는 것을 의미하며, 그룹 추천에서 구성원의 선호도 분포와 편차를 고려하는 것은 매우 중요한 요소 중 하나라 할 수 있다.

4. 그룹 추천에서 선호도 편차의 중요성

그룹 추천은 그룹 탐색, 그룹 모델링, 그룹 선호 추정 의 세 과정을 통해 이루어진다[6]. 그 중 그룹 모델링의 과정에서 그룹 선호도를 측정하기 위한 전략을 수립하는 것이 가장 중요하며, 본 논문에서도 새로운 그룹 모델링 전략을 제안하는 것에 초점을 맞춘다. 제안하는 그룹 모델링 전략은 그룹 내에서 구성원들의 관심사에 따라 변하는 선호도의 편차를 고려하여 평균 선호도가 높으면서 선호도 편차가 작은 관심사들을 추천의 결과로 제공하는 것을 목표로 한다. 기존의 전략들은 단순히 평균 선호도가 높거나 선호 횟수가 많은 관심사를 추천 결과로서 그룹에 제공하는 반면, 제안하는 전략은 평균 선호도와 선호도 편차를 함께 고려하여 고른 분포로 높은 선호도를 갖는 관심사를 그룹에 제공한다. 이는 기존의 전략에 비해 그룹 내 구성원 개개인의 성향을 반영하는 것이 가능하여 보다 많은 수의 그룹 내 구성원들에게 적합한 정보를 제공할 수 있다.

4.1 그룹 탐색

그룹 추천을 위해서는 우선적으로 유사한 성향을 갖는 사용자 집합을 통해 그룹을 형성하는 과정이 필요하다[6,14-16]. 추천의 대상이 되는 그룹을 탐색하기 위해 사용자들 간의 유사도를 측정하며, 측정된 유사도를 통해 높은 유사성을 갖는 사용자들을 하나의 그룹으로 형성한다. 본 논문에서는 그룹 형성을 위해 사용자들 간의 유사성을 측정하는 수단으로 추천 시스템에서 가장 많이 사용되는 Pearson Correlation Coefficient를 사용한다. I_a 와 I_u 를 사용자 a 와 u 가 선호하는 관심사들의 집합이라 하고, $r_{a,i}$ 와 $r_{u,i}$ 는 관심사 i 에 대한 사용자 a 와 u 의 선호도, \bar{r}_a 와 \bar{r}_u 는 사용자 a 와 u 가 선호한 모든

관심사에 대한 선호도의 평균이라고 할 때, 두 사용자 a 와 u 간의 유사도 $w_{a,u}$ 는 식 (1)과 같이 계산된다.

$$w_{a,u} = \frac{\sum_{i \in I_a \cap I_u} (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I_a \cap I_u} (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i \in I_a \cap I_u} (r_{u,i} - \bar{r}_u)^2}} \quad (1)$$

다음 알고리즘은 식 (1)을 기반으로 사용자들 사이의 유사성을 측정하고, 상위 유사도인 N 명을 통해 그룹을 형성하는 본 논문의 그룹 탐색 과정을 나타낸다. u 를 그룹의 중심이 되는 사용자라 하고, U 는 사용자 집합을 의미하고 a 는 집합 U 상의 사용자라 할 때, 그룹 G 는 u 와 유사한 상위 $N-1$ 명의 a 를 통해 N 명의 구성원을 형성한다.

Algorithm Group Detection

```

1: Input :  $U, u$ 
2: Output :  $G$ 
3: for  $a \in U$ 
4:   if  $a \neq u$ 
5:      $u_{sim,a} = w_{a,u}$ 
6:   end
7: end
8:    $G.insert(u)$ 
9: for  $a \in Top-k(u_{sim}, N-1)$ 
10:    $G.insert(u)$ 
11: end

```

4.2 사용자 선호도의 편차를 고려한 그룹 모델링 전략

그룹 모델링을 위한 기존의 통합 전략 중 대표적으로 사용되는 Average 전략은 그룹 내 구성원들의 관심사에 대한 선호도의 평균을 측정하여, 높은 평균을 갖는 관심사를 그룹에 추천 결과로 제공한다[8]. Average 전략은 합리적인 그룹 모델링 전략으로 볼 수 있지만, 앞서 설명하였듯이 높은 평균 선호도를 갖는 관심사라고 해서 항상 그룹 내의 모든 구성원들에게 적합한 정보라 할 수 없다. 따라서 그룹 내 모든 구성원들이 만족할만한 적합한 정보를 추천의 결과로 제공하기 위해서는 고른 분포를 통해 높은 평균 선호도로 측정되는 관심사를 추천해주는 것이 중요하며, 선호도의 평균뿐만 아니라 선호도의 분포 및 편차를 함께 고려할 수 있는 그룹 모델링 전략이 필요하다. 본 논문에서는 선호도의 평균에 선호도의 편차를 가중치로 하는 그룹 모델링 전략을 제안하며, 선호도의 편차를 구하기 위한 방법으로 MSD (Mean Square Deviation)를 활용한다. \bar{r}_i 는 관심사 i 에 대한 그룹 내 모든 구성원들의 평균 선호도를 의미하고, N 이 관심사 i 에 대한 구성원들의 선호 횟수라 할 때, 그룹 G 의

관심사 i 에 대한 편차 MSD_i 는 식 (2)와 같다.

$$MSD_i = \frac{\sum_{u \in G} (\bar{r}_i - r_{u,i})^2}{N} \quad (2)$$

MSD_i 는 데이터 집합의 선호도 범위에 따라 그 값이 너무 커져서 전체적인 결과에 큰 영향을 미치면 정확한 추천 결과를 도출하는데 악영향을 미칠 수 있다. 따라서 MSD_i 의 범위를 0과 1사이로 만들기 위해 사용자의 관심사에 대한 선호도를 0과 1사이의 범위로 정규화 시켜 준다. $maxr$ 과 $minr$ 을 데이터 집합에서 표현 가능한 최대 선호도와 최소 선호도라 할 때, 정규화된 사용자의 선호도 $r'_{u,i}$ 는 다음과 같다.

$$r'_{u,i} = \frac{r_{u,i} - minr}{maxr - minr} \quad (3)$$

MSD_i 는 그 값이 작으면 작을수록 그룹 내에서 선호도 분포가 고른 관심사라는 것을 의미하기 때문에, 제안하는 그룹 모델링 전략에서 $1 - MSD_i$ 를 가중치로 한다. 제안하는 구성원 간의 선호도 평균과 편차를 모두 고려하는 그룹 모델링 전략을 통해 측정되는 그룹 선호도 $r_{G,i}$ 는 식 (4)와 같다.

$$r_{G,i} = \bar{r}_i \cdot (1 - MSD_i) \quad (4)$$

4.3 그룹 선호도 추정

그룹 모델링 전략에 의해 측정된 그룹 선호도는 기존의 추천 시스템 상의 선호도와 그 범위가 다르다. 따라서 그룹 모델링 전략의 성능을 평가하기 위해서는 통합 전략의 성질에 따라 그룹의 선호도를 정규화하는 과정이 필요하다[6]. $max_{u \in G} r_{u,i}$ 는 관심사 i 에 대한 그룹 내 구성원들의 선호도 중 최대 선호도를 의미하고, $max_{j \in I} r_{G,j}$ 는 모든 관심사 I 의 그룹 선호도 중 최대 선호도라 할 때, 제안 전략을 위해 정규화된 그룹 선호도 $p_{G,i}$ 는 식 (5)와 같다.

$$p_{G,i} = \frac{r_{G,i} \cdot \max_{u \in G} r_{u,i}}{\max_{j \in I} r_{G,j}} \quad (5)$$

5. 실험 및 평가

본 논문에서는 제안하는 그룹 추천 방법의 성능을 검증하기 위해 Movie-Lens 100K와 1M 데이터 집합을 기반으로 기존의 Additive Utilitarian[5], Multiplicative Utilitarian[8], Approval Voting[9], Approval Voting (Threshold = 4)[9], Least Misery[3], Most Pleasure[8], Borda Count[8], Average without Misery[4] 전략들과 비교 평가를 수행한다.

5.1 데이터 집합

Movie-Lens는 영화에 대한 사용자들의 선호도가 기

표 3 실험 및 평가를 위한 Movie-Lens 100K와 1M의 구성
Table 3 Dataset of Movie-lens 100K and 1M for experiment and evaluation

Movie-Lens		100K	1M
Contents			
All	User	943	6,040
	Movie	1,682	3,883
	Preference	100,000	1,000,209
Active	User	200	200
	Movie	100	100
	Preference	13,218	17,761

록된 데이터 집합으로 사용자 정보, 영화 정보, 선호 정보로 이루어져있다. 본 논문에서는 Movie-Lens의 영화에 대한 사용자의 선호 정보를 이용하였으며, 보다 정확한 검증을 위해 가장 많이 선호도가 평가된 상위 100개의 영화와 해당 영화들을 많이 선호한 상위 200명의 사용자들을 대상으로 실험하였다. 표 3은 본 논문의 실험 및 평가를 위한 데이터 집합의 명세를 나타낸다.

5.2 교차 검증(Cross Validation)

본 논문에서는 제안하는 그룹 추천 방법의 성능을 검증하기 위해 5-Fold Cross Validation을 수행하였다. 그림 2는 그룹 추천의 정확성 평가를 위한 교차 검증 및 비교 평가를 위한 실험 과정을 나타낸다. 우선, Movie-Lens 100K와 1M 데이터의 관심사(영화) 중 80%를 훈련 집합으로 하여 그룹 탐색을 통해 N 명의 사용자를 그룹으로 형성하고, 훈련 집합을 통해 형성된 그룹을 기반으로 나머지 관심사의 20%를 평가를 위한 실험 집합으로 사용하였다. 실험 집합의 관심사에 그룹 모델링 전략을 적용하고, 그룹 선호도 $p_{G,i}$ 를 측정한다. 마지막으로 평가 기법을 통해 상위 k 개의 관심사에 대한 정확성을 측정한다. 통계적으로 유의한 측정 결과를 얻기 위해 훈련 집합과 실험 집합을 달리하여 5회 반복하여 실험하였고, 그룹을 형성하는 구성원들의 수에 따른 성능을 평가하기 위해 구성원의 수 N 을 4명, 8명, 16명, 32명, 64명인 경우로 나누어 실험을 진행하였다.

5.3 정확성 평가

본 논문에서는 제안하는 그룹 추천 방법의 성능을 검증하기 위해 기존의 모델링 전략들과 비교평가를 수행하였다. 검증을 위한 평가 기법으로는 예측된 그룹 선호도의 정확성을 측정하는 Root Mean Square Error (RMSE)와 Mean Absolute Error(MAE)를 사용하였다. 그룹 추천을 통해 제공되는 상위 k 개의 관심사 집합을 I_k 라 할 때, 관심사 i 에 대한 RMSE와 MAE는 식 (6), (7)과 같다.

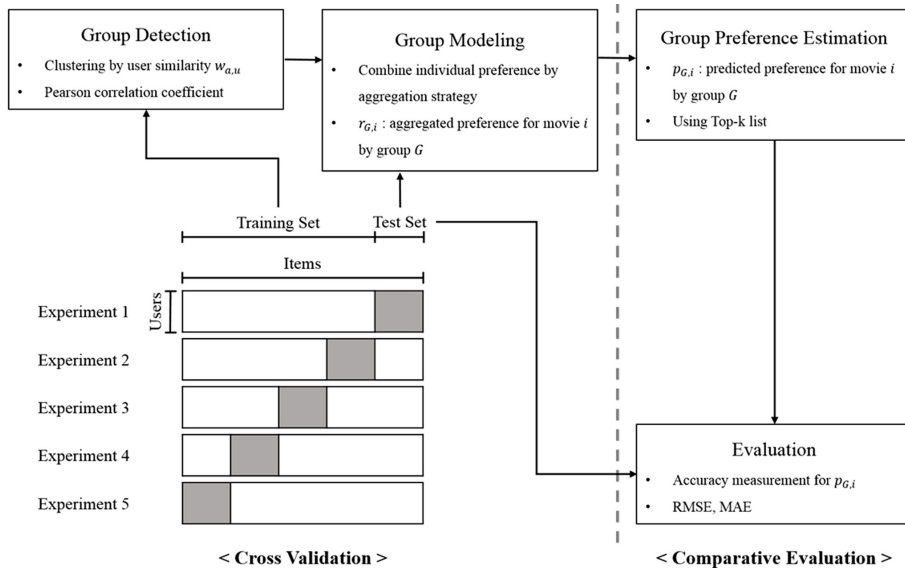


그림 2 교차 검증 및 비교 평가를 위한 실험 과정

Fig. 2 Process of cross validation and comparative evaluation

$$RMSE_i = \sqrt{\frac{\sum_{u \in G} (p_{G,i} - r_{u,i})^2}{N}}, \quad RMSE = \frac{\sum_{i \in I_k} RMSE_i}{k} \quad (6)$$

$$MAE_i = \frac{\sum_{u \in G} |p_{G,i} - r_{u,i}|}{N}, \quad MAE = \frac{\sum_{i \in I_k} MAE_i}{k} \quad (7)$$

그림 3과 그림 4는 Movie-Lens 100K와 1M에서 기존의 그룹 추천 방법과 제안하는 그룹 추천 방법을 정량적으로 비교 평가한 결과를 나타낸다. 평가를 위한 추천 결과의 수 k 는 2, 4, 6, 8, 10으로 하여 $RMSE$ 와 MAE 를 측정하였다. 또한 그룹의 크기와 추천 결과의 수에 따른 제안 전략과 기존의 그룹 모델링 전략들 사이의 비교 평가 결과가 통계적으로 유의한지(p -value < 0.05) 측정하기 위해 t -test를 수행하였다. 제안하는 전략의 성능이 기존의 그룹 모델링 전략에 비해 통계적으로 유의할 경우는 기존 그룹 모델링 전략을 대표하는 마크를 통해 그래프의 x 축 하단에 표시하였다.

실험 결과 Movie-Lens 1M에서 구성원이 64명이고 상위 추천 결과가 8개와 10개인 경우에 대해 Average without Misery 전략이 제안하는 방법보다 더 높은 성능을 보인 것을 제외하고, 모든 경우에서 제안 전략이 높은 성능을 가졌다. 또한 추천 결과의 수가 적을수록 제안하는 방식이 더 높은 정확성을 보였는데, 이는 추천 결과의 수가 많아질수록 각각의 전략들이 유사한 추천 결과를 갖게 되어 결과의 수가 적을 때에 비해 비슷한 결과가 나오기 때문이다.

기존의 그룹 모델링 전략들 중 선호도만을 고려하는

Least Misery와 Most Pleasure 전략은 그룹 내 구성원들의 관심사에 대한 선호도 중 최소 선호도 값과 최대 선호도 값을 그룹 선호도로 하기 때문에, 그룹 내 구성원 대다수를 만족시키는 관심사를 선택한다기보다는 극단적인 선호도를 선택하는 방식이라 할 수 있다. 따라서 타 전략들이 비해 그룹 내 많은 수의 구성원들을 만족시키는 결과를 도출하기가 어렵고, 실제 실험 결과를 통해서도 Multiplicative Utilitarian 전략을 제외한 다른 전략들에 비해 일반적으로 낮은 성능을 보인다. 또한 Least Misery가 Most Pleasure보다 높은 성능을 보이는 데, 최소 관심사가 높다는 것은 일반적으로 그룹 내의 선호도 평균이 높으면서 분포가 작다는 것을 의미하기 때문이다. 제안하는 전략의 $RMSE$ 는 Least Misery와 Most Pleasure 전략에 비해 평균적으로 각각 25.90%와 45.50%의 성능 향상을 보였으며, MAE 는 23.38%와 42.61%의 성능 향상을 보였다.

선호 횟수가 중요시되는 Additive Utilitarian 전략은 선호도만을 고려하는 Least Misery와 Most Pleasure 전략에 비해 비슷하거나 약간 좋은 성능을 보인다. Approval Voting의 경우 Most Pleasure에 비해서는 높은 성능을 보이지만, Least Misery에 비해서는 평균적으로 낮은 성능을 보인다. 하지만 사용자수가 32명 이상인 경우에는 Least Misery에 비해 전반적으로 높은 성능을 갖는다. 또한, Additive Utilitarian가 Approval Voting보다 더 높은 성능을 보이는데, 이는 Approval Voting은 단순히 선호 횟수만을 고려하기 때문에 부정

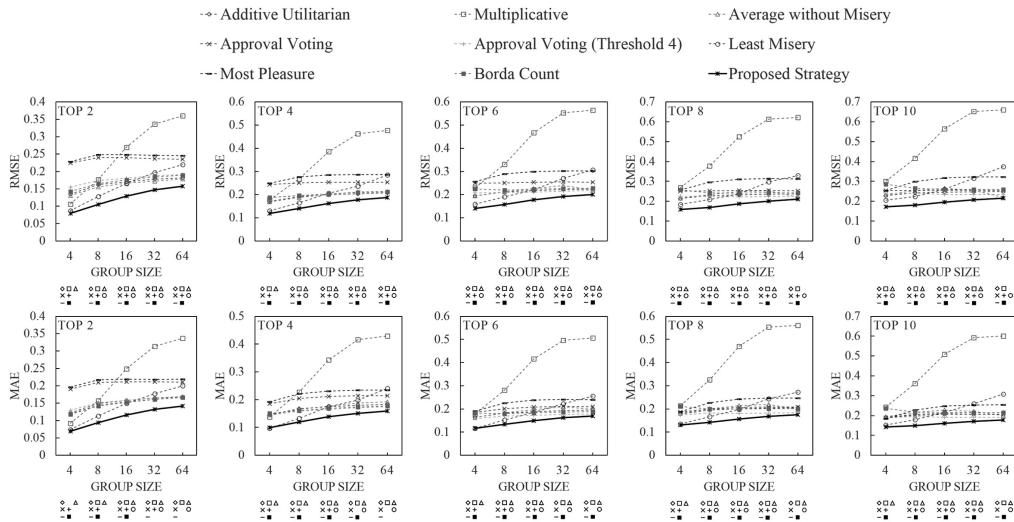


그림 3 Movie-Lens 100K에서 그룹의 크기와 추천 결과의 수에 따른 추천의 정확성 비교
Fig. 3 Accuracy measurement by group size and number of recommendation results in Movie-lens 100k

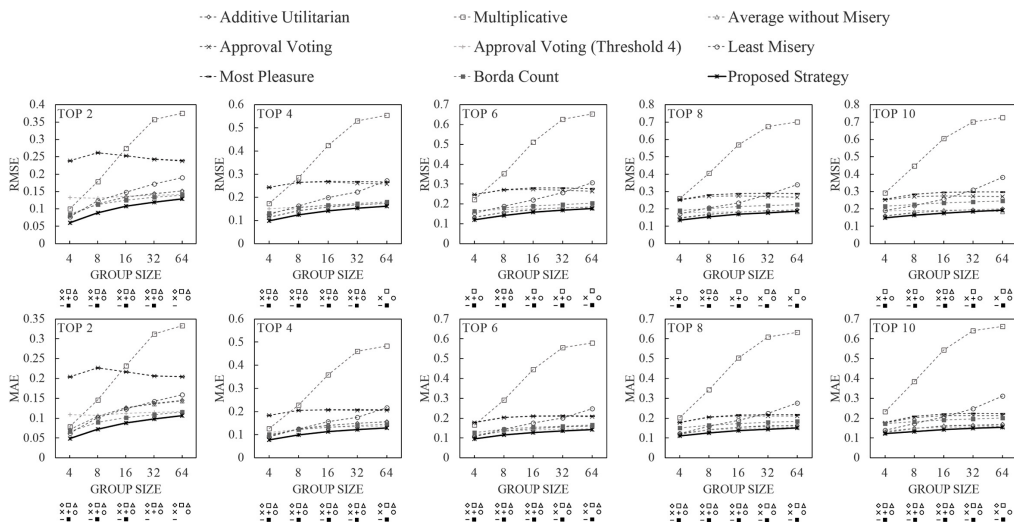


그림 4 Movie-Lens 1M에서 그룹의 크기와 추천 결과의 수에 따른 추천의 정확성 비교
Fig. 4 Accuracy measurement by group size and number of recommendation results in Movie-lens 1M

적인 선호도를 갖는 경우와 긍정적인 선호도를 갖는 경우를 동일하게 취급하기 때문이다. Multiplicative Utilitarian 전략의 경우 선호 횟수가 많아짐에 따라 그룹 선호도 값이 기하 급수적으로 증가하기 때문에 그룹 추천의 성능에 악영향을 미친다. 따라서 다른 전략들에 비해서 가장 낮은 성능을 보이고, 특히 그룹 내 구성원의 수가 많아지면 많아질수록 더 낮은 성능을 보인다. Additive Utilitarian과 Approval Voting 전략에 비해 제안하는 전략의 $RMSE$ 가 평균적으로 16.33%와 41.34%의 성능

이 향상 되었고, MAE 는 20.65%와 39.31%의 성능이 향상되었다. Multiplicative Utilitarian 전략에 비해서는 $RMSE$ 와 MAE 가 각각 57.71%와 41.34%만큼 향상되어 다른 전략들에 비해 큰 향상율을 보였다.

앞선 전략들을 개선하기 위해 제안된 Borda Count, Approval Voting with Threshold, Average without Misery 전략은 Additive Utilitarian을 제외하고 나머지 전략들에 비해서 개선된 성능을 보인다. 하지만 Borda Count 전략의 경우 선호도의 분포와 편차를 고려하지

않기 때문에 Approval Voting with Threshold와 Average without Misery 전략에 비해서 성능이 낮고, 이 두 전략 또한 극단적인 편차만 고려하기 때문에 제안하는 전략에 비해 성능이 떨어진다. 제안 전략은 Borda Count 전략에 비해 $RMSE$ 는 평균적으로 22.42%, MAE 는 평균적으로 21.40%를 개선하였고, Approval Voting with Threshold 전략에 비해서는 18.38%와 16.03%, Average without Misery에 비해서는 15.40%와 20.00%의 성능 향상을 보였다.

제안 전략은 기존의 그룹 모델링 전략들에 비해 높은 성능을 보였으며 통계적으로도 유의함을 확인하였다. 또한 그룹 내의 구성원 수를 늘려가며 실험을 진행하여, 많은 수의 구성원들로 이루어진 그룹에서도 상대적으로 우수한 성능을 가짐을 확인하였다. 이를 통해 관심사에 대한 선호도뿐만 아니라 선호도 분포와 편차를 함께 고려하여 그룹 추천에 적용할 경우, 추천의 성능이 개선되고 더 많은 수의 그룹 내 구성원에게 만족할만한 추천 결과를 제공할 수 있다는 것을 확인하였다.

6. 결론 및 향후 연구

본 논문에서는 고른 분포로 높은 선호도를 갖는 관심사를 그룹 추천의 결과로 제공하기 위해 사용자들의 평균 선호도와 선호도의 편차를 함께 고려한 그룹 모델링 전략을 제안한다. 사용자 선호도의 편차를 고려한 그룹 모델링 전략은 그룹 내에서 구성원들의 관심사에 따라 변하는 선호도의 편차를 고려하기 때문에, 기존의 그룹 추천에 비해 많은 수의 그룹 내 구성원들에게 만족스러운 추천 결과를 제공하는 것이 가능하다. 제안 전략이 실제로 기존의 통합 전략들에 비해 높은 성능을 갖는지를 검증하기 위해 Movie-Lens 100K와 1M에서 4명, 8명, 16명, 32명, 64명의 구성원들로 이루어진 그룹을 통해 기존의 그룹 모델링 전략과 비교 평가를 진행하였다. 실험을 통해 제안하는 방식이 기존의 그룹 추천 방식보다 높은 성능을 보임을 확인하였고, 그룹 추천에서 선호도의 편차가 중요한 요소임을 증명하였다.

본 논문은 그룹 모델링 전략에 초점을 맞추어 연구를 진행하였다. 향후 연구로는 그룹 탐색에 초점을 맞추어 기존의 그룹 탐색 방식과 차별화된 클러스터링 방식에 제안하는 모델링 전략을 적용한 그룹 추천 시스템을 제안하고자 한다. 또한 제안하는 그룹 모델링 전략이 Movie-Lens뿐만 아니라, Last.fm과 Yelp와 같은 다른 분야의 데이터 집합에서도 높은 성능을 보이는지 실험을 통해 확인하고자 한다.

References

- [1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutierrez, "Recommender systems survey," *Journal of Knowledge-Based Systems*, Vol. 46, pp. 109-132, 2013.
- [2] M. Kompan and M. Bielikova, "Group recommendations: survey and perspectives," *Journal of Computing and Informatics*, Vol. 33, pp. 446-476, 2014.
- [3] M. O'Connor, D. Cosley, J. Konstan, and J. Riedl, "PolyLens: A Recommender System for Groups of Users," *Proc. of European Conference on Computer-Supported Cooperative Work*, pp. 199-218, 2001.
- [4] J. McCarthy, and T. Anagnost, "MusicFX: An arbiter of group preferences for computer supported collaborative workouts," *Proc. of Computer Supported Cooperative Work*, pp. 363-372, 1998. pp. 765-774, 2011.
- [5] L. Ardisson, A. Goy, G. Petrone, M. Segnan, and P. Torasso, "Tailoring the Recommendation of Tourist Information to Heterogeneous User Groups," *Proc. of International Workshops OHS-7, SC-3, and AH-3, LNCS 2266*, pp. 280-295, 2002.
- [6] L. Boratto, S. Carta, and G. Fenu, "Discovery and representation of the preferences of automatically detected groups: Exploiting the link between group modeling and clustering," *Journal of Future Generation Computer Systems*, 2015.
- [7] A. Jameson, and B. Smyth, "Recommendation to groups," *Journal of The adaptive web*, pp. 596-627, 2007.
- [8] J. Masthoff, "Group modeling: Selecting a sequence of television items to suit a group of viewers," *Journal of User Modeling and User-Adapted Interaction*, Vol. 14 pp. 37-85, 2004.
- [9] H. Lieberman, N.W.V. Dyke, and A.S. Vivacqua, "Let's browse: A collaborative web browsing agent," *Proc. of Intelligent User Interfaces*, pp. 65-68, 1999.
- [10] S. Pizzutillo, B. De Carolis, G. Cozzolongo, and F. Ambruso, "Group modeling in a public space: methods, techniques, experiences," *Proc. of World Scientific and Engineering Academy and Society*, pp. 175-180, 2005.
- [11] K. McCarthy, M. Salamó, L. Coyle, L. McGinty, B. Smyth, and P. Nixon, "CATS: A Synchronous Approach to Collaborative Group Recommendation," *Proc. of FLAIRS Conference*, pp. 86-91, 2006.
- [12] A. Crossen, J. Budzik, and K. J. Hammond, "Flytrap: intelligent group music recommendation," *Proc. of Intelligent User Interfaces*, pp. 184-185, 2002.
- [13] Z. Yu, X. Zhou, Y. Hao, and J. Gu, "Tv program recommendation for multiple viewers based on user profile merging," *Journal of User Modeling and User-Adapted Interaction*, Vol. 16, pp. 63-82, 2006.
- [14] Y.-L. Chen, L.-C. Cheng, and C.-N. Chuang, "A group recommendation system with consideration of interactions among group members," *Journal of*

Expert Systems with Applications, Vol. 34, pp. 2082–2090, 2008.

- [15] J. Wang, Z. Liu, and H. Zhao, "Group Recommendation Using Topic Identification in Social Networks," *Proc. of Intelligent Human-Machine Systems and Cybernetics*, pp. 355–358, 2014.
- [16] L. Baltrunas, T. Makcinskas, and F. Ricci, "Group recommendations with rank aggregation and collaborative filtering," *Proc. of Recommender System*, pp. 119–126, 2010.
- [17] J. Bobadilla, F. Serradilla, and J. Bernal, "A new collaborative filtering metric that improves the behavior of recommender systems," *Journal of Knowledge Based Systems*, Vol. 23, pp. 520–528, 2010.



김 형 진

2015년 원광대학교 컴퓨터공학과(학사)
2015년~현재 고려대학교 컴퓨터학과 석사과정. 관심분야는 소셜 네트워크 서비스, 빅 데이터, 추천 시스템, 온톨로지, 자가적응 소프트웨어



서 영 덕

2012년 고려대학교 컴퓨터학과(학사). 2012년~현재 고려대학교 컴퓨터학과 석·박통합과정. 관심분야는 소셜 네트워크 서비스, 빅 데이터, 추천 시스템, 온톨로지, 자가적응 소프트웨어



백 두 권

1974년 고려대학교 수학과 학사. 1977년 고려대학교 산업공학과 석사. 1983년 Wayne State Univ. 전산학과 석사. 1985년 Wayne State Univ. 전산학과 박사. 1986년~현재 고려대학교 정보대학 컴퓨터학과 교수. 1989년~1991년 고려대학교 전산학과 학과장. 1990년~1991년 미국 Arizona 대학교 객원 교수. 1991년~2013년 ISO/IEC JTC1/SC32 전문위원회 위원장. 1993년~1999년 한국과학기술원 객원책임연구원. 1993년~1999년 한국DB진흥센터 표준연구위원. 1996년~1997년 고려대학교 컴퓨터과학기술연구소 초대소장. 1997년~1998년 고려대학교 정보전산원 원장. 1998년~1999년 한국정보과학회 전산교육연구회 운영위원장. 1999년~2001년 정보통신진흥협회 데이터기술위원회 의장. 2002년~2004년 고려대학교 정보통신대학 초대학장. 2002년~2003년 한국시뮬레이션학회 회장. 2003년~2009년 정보통신부 컴퓨터프로그램보호위원회 위원. 2004년~2005년 한국정보처리학회 부회장. 2005년~2009년 한국소프트웨어진흥원 이사. 2009년~2011년 고려대학교 정보통신대학 학장. 관심분야는 소프트웨어공학, 데이터공학, 모델링, 시뮬레이션, 자가적응 SW