creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

碩士學位 論文

음악 목록의 유사성 분석을 통한 협업 필터링 추천 방법

A Method of Collaborative Filtering by Analyzing Similarity between Music lists

2009 年 12月

崇實大學校 大學院

미디어學科

최 창 환

碩士學位 論文

음악 목록의 유사성 분석을 통한 협업 필터링 추천 방법

A Method of Collaborative Filtering by Analyzing Similarity between Music lists

2009 年 12月

崇實大學校 大學院

미디어學科

최 창 환

碩士學位 論文

음악 목록의 유사성 분석을 통한 협업 필터링 추천 방법

指導敎授 고 일 주

이 論文을 碩士學位 論文으로 제출함

2009 年 12月

崇實大學校 大學院

미디어學科

최 창 환

최 창 환의 碩士 學位 論文을 認准함

審 査 委 員 長 _____ 印

審 査 委 員 _____ 印

審 査 委 員 _____ 印

2009 年 12月

崇實大學校 大學院

목 차

國文抄錄 ····· iv
英文抄錄 ····· v

제 1 장 서론 ····· 1
1.1 연구 배경 ····· 1
1.1.1 음악 시장 현황 ····· 1
1.1.2 음악 추천 서비스 ····· 2
1.2 연구 목적 ····· 7
1.2.1 협업 필터링의 문제점 ····· 7
1.2.2 협업 필터링 개선 방법 제안 ····· 9
1.3 논문의 구성 ····· 11

제 2 장 관련 연구 ····· 12
2.1 특징 연구 ····· 12
2.1.1 사용자 정보 ····· 12
2.1.2 사용자 이력 ····· 12
2.1.3 음악 콘텐츠 메타 데이터 ····· 13
2.1.4 내용 기반 특징 ····· 13
2.1.5 특징 선택 ····· 13
2.2 군집화 기법 ····· 15
2.2.1 유클리디안 거리 ····· 15
2.2.2 K-Means 알고리즘 ····· 16

2.2.3 SOM ····· 16
2.3 협업 필터링 ····· 21
2.3.1 내용 기반 필터링 ····· 21
2.3.2 협업 필터링 ····· 22

제 3 장 음악 목록의 유사성 분석을 통한 협업 필터링 26
3.1 특징 선택 ····· 26
3.2 특징의 정규화 ····· 27
3.3 음악 목록 유사성 분석 ····· 28
3.4 음악 추천 ····· 30
3.5 시스템 구조 ····· 34

제 4 장 실험 및 결과 ····· 36
4.1 실험 데이터 ····· 36
4.2 평가 방법 ····· 37
4.3 실험 결과 ····· 38
4.3.1 사용자 음악 목록 유사성 분석 ····· 38
4.3.2 클래스를 이용한 SOM 학습 ····· 41
4.3.3 추천 결과 ····· 44

제 5 장 결론 및 향후 연구 ····· 48

참고문헌 ····· 50

그 림 목 차

[그림 1-1] 인기순위 100곡 서비스 ····· 3
[그림 1-2] 공개 앨범 서비스 ····· 4
[그림 1-3] 애플 iTunes Genius ····· 5
[그림 1-4] Last.fm ····· 6
[그림 1-5] 제안된 협업 필터링 흐름도 ····· 10
[그림 2-1] SOM의 구성 형태 ····· 17
[그림 2-2] SOM 학습 시 이웃 반경의 감소 ····· 18
[그림 2-3] 가우시안 함수 ····· 19
[그림 3-1] 사용자 음악 목록 유사 분석 시 클래스 예시 ····· 29
[그림 3-2] 추천 클래스 선택 ····· 31
[그림 3-3] 곡 추천 ····· 33
[그림 3-4] 시스템 구조도 ····· 34
[그림 4-1] 사용자별 보유 곡 개수 ····· 38
[그림 4-2] 클래스별 보유 곡 개수 ····· 39
[그림 4-3] 사용자별 클래스 분포 평균 ····· 40
[그림 4-4] SOM 초기화 ····· 41
[그림 4-5] SOM 학습 종료 ····· 42
[그림 4-6] SOM 학습결과에 따른 클래스 분포 ····· 43
[그림 4-7] 클래스별 추천 결과 ····· 44
[그림 4-8] 추천 테스트 페이지 ····· 46
[그림 4-9] 클래스별 추천 MAE 결과 ····· 46

표 목 차

[표 1-1] 디지털 음악 산업 발전 협의회 2007 ····· 2
[표 1-2] 포털 사이트 검색 결과 ····· 8
[표 3-1] mRMR 결과 ····· 26
[표 3-2] 가수 형태 정규화 ····· 28
[표 4-1] 사용자별 보유 곡 개수 ····· 38
[표 4-2] 클래스별 Coverage Ratio 결과 ····· 45
[표 4-3] 클래스별 MAE 결과 ····· 47

국문초록

음악 목록의 유사성 분석을 통한 협업 필터링 추천 방법

미디어 학과 최창환

指 導 敎 授 고일주

네트워크의 속도 향상은 일반 텍스트 콘텐츠부터 미디어 콘텐츠까지 개인용 컴퓨터 또는 네트워크가 가능한 디바이스를 통해서 각종 정보들에 대해 쉽게 접근할 수 있는 통로를 마련해 주었다.

특히, 음악 시장은 오프라인 시장에서 온라인 시장으로 이동하면서 이용 가능한 콘텐츠 규모가 커지게 되었으며, 음악 검색 엔진도 초창기 정보 검색 엔진의 발전 과정과 같이 단순 검색에서 멜로디 검색, 내용 기반 검색 그리고 추천의 형태로 진화하고 있다.

추천의 방법에는 사용자들의 콘텐츠에 대한 선호도 점수를 이용하여 유사한 음악목록을 가진 사용자끼리 추천을 하는 협업 필터링 방식이 주로 이용되고 있다. 그러나 음악 추천에서는 이미 알고 있는 곡이 아닌 모르는 곡을 추천 받는 것에 대한 요구가 높기 때문에 추천이 요구되는 곡은 사용자의 재생 또는 구매이력에 포함 될 확률이 낮다.

본 논문에서는 이와 같은 문제를 해결하기 위해서 사용자의 음악 목록을 여러 개의 선호 그룹으로 나눈 후, 추천에 대한 요구가 높을 것으로 예측되는 보유 목록이 적은 그룹을 대상으로 사용자간 유사도를 측정하여 음악을 추천하는 방법을 제안한다.

실험 결과는 추천 요구가 높을 것으로 예상되는 그룹이 다른 그룹보다 더 나은 추천 결과를 보였다.

ABSTRACT

A Method of Collaborative Filtering by Analyzing Similarity between Music lists

CHOI, CHANG-HWAN
Department of Media
Graduate School
Soongsil University

An increase in network speed made a passage to approach each kind of information easily through devices for personal computer or network from general text contents to media contents.

Especially, in the music market, a large of usable contents became bigger by moving to online market from offline market. Music search engine is being developed from simple search to melody search, contents-based search and form of recommendation with the developmental process of early search engine.

In a method of recommendation, a cooperative filtering method, recommending among users with similar music lists with users' preferential scores for contents, is mainly used.

However, in recommendation of music, music for recommendation has the probability to be included in users' playback or purchase records since users have a high demand for being recommended unknown music, not already known music.

The study proposes how to recommend music by measuring similarity among users with groups with a few lists in hand that it is expected that users will have a high demand for recommendation after dividing users' music lists into diverse preferential groups to settle the above problem.

As a result of experiments, a group expected that users would have a high demand for recommendation had better results of recommendation than other groups.

제 1 장 서 론

1.1 연구배경

1.1.1 음악 시장 현황

2000년대에 접어들어 음반 시장은 온라인과 오프라인으로 양분화 되었다. 그 후, 디지털 음반 시장 급속히 발달하여 오프라인 음반 시장을 넘어서게 넘어 설 정도로 크게 성장하게 되었는데, 세계적인 추세도 2003년 국제 음반 산업 연맹(IFPI)의 통계에 따르면 온라인에서 합법적으로 판매된 음악이 오프라인 싱글 음반 판매량에 비해 3배나 많은 것으로 나타났다(한국 소프트웨어 진흥원, 2008).

국내 온라인 음반 시장은 2003년부터 온라인 기업들이 세계적인 추세에 맞추어 유료화를 실시하면서, 온라인 음악 서비스에 대한 기대 때문에 많은 음악 서비스 업체들이 늘어났다.

2008년에는 온라인 음악 시장으로 진출한 대형음반 업체들 및 유무선 통신회사들의 기존 선점 업체들과의 제휴 및 합병의 진행으로 실제로 온라인 음악 시장도 국내 음반 시장의 큰 영역을 차지하고 있음이 증명되게 되었다.

많은 온라인 음악 서비스의 등장으로 인해 사용자들은 온라인에서의 방대한 음악 데이터베이스에 접속해 손쉽게 세계 각국 및 다른 문화의 다양한 장르의 음악과 실시간으로 빠르게 업데이트 되는 최신의 음악들을 들을 수 있게 되었다.

이러한 손쉬운 접근성으로 인해 사용자들은 계속 늘어나 [표 1-1]과

같이 2006년 온라인 음악 시장의 규모는 3,500억으로 2005년 오프라인 음악 시장의 규모인 1,087억의 2배 이상의 규모로 성장하게 되었다(한국 소프트웨어 진흥원, 2008).

[표 1-1] 디지털 음악 산업 발전 협의회 2007

구분	2000년	2001년	2002년	2003년	2004년	2005년	2006년	2007년(추정)
음반시장	4,104	3,733	2,861	1,833	1,338	1,087	848	600~700
디지털음악시장	-	911	1,328	1,811	2,112	2,621	3,500	3,700
무선(컬러링/벨소리)	-	890	1,290	1,767	1,911	2,251	2,300	2,200
유선	-	21	38	44	201	370	1,200	1,500
총규모	4,104	4,644	4,189	3,644	3,450	3,708	4,384	4,400

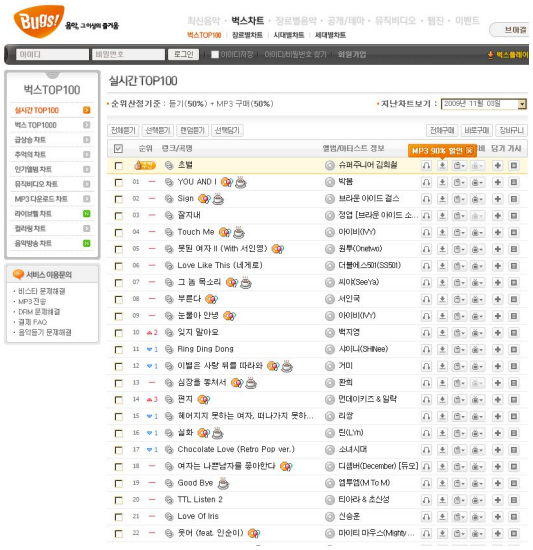
출처 : 디지털 콘텐츠 산업백서

1.1.2 음악 추천 서비스

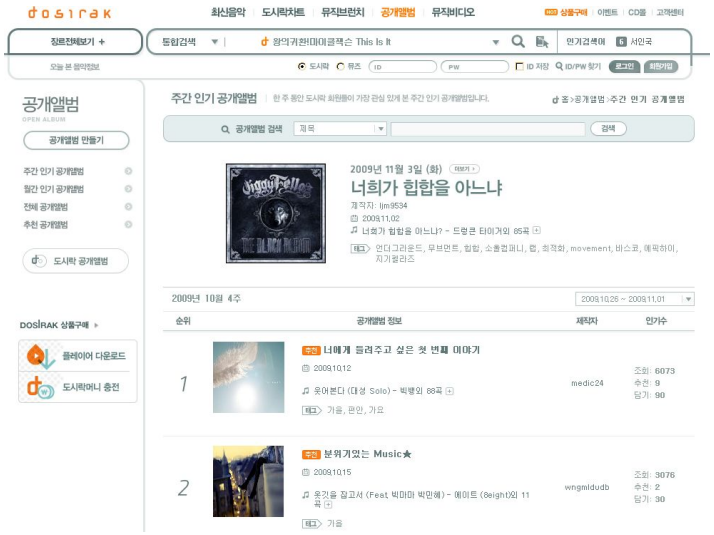
온라인 음악 시장에서 간편한 접근성으로 인한 사용자들의 증가는 인터넷 서비스에 대한 기술의 발전과 새로운 요구를 가져오게 되었다. 각종 온라인 서비스 업체들은 검색 서비스를 기본으로 제공하며, 사용자에게 좀 더 편하게 원하는 곡을 찾을 수 있는 환경을 마련함으로써 매출 신장에 대한 기대를 하였고, 검색으로는 장르별 검색, 연대별 검색 등을 제공 하였고, 여기에 조금 더 나아가 유사 아티스트 검색, 자연어 검색(지식 묻고 답하기), 허밍 질의 검색 등을 지원하게 되었다. 하지만, 하루에도 수백의 신규 음악이 발표되는 음악 시장에서 방대한 음악 데이터베이스를 구축하여 제공하는 것은 쉽지 않은 작업이며, 더욱이 곡에 대한 각종 메타 데이터를 입력하여 제공하는 것은 많은 전문가들과 작업자들과 시간을 소모해야 하는 일이다. 그 예로, 마이크로 소프트의 MSN Music의 경우 장르별 검색을 위해 연 100,000곡을 약 30명의 음악 전문

가들이 장르 분류를 하는 작업을 하고 있다(Nicolas Scaringella & Giorgio Zoia & Daniel Mlynek, 2006). 이러한 검색을 통한 서비스 모델은 보편화되어 서비스 회사들은 모바일과 같은 디바이스를 통한 서비스의 확대, 각종 이벤트 및 프로모션을 통한 신곡 및 신보 추천을 통해서 사용자들의 구매를 유도하였다. 이렇게 고객 확보를 위한 경쟁은 음악 추천이라는 영역을 거치지 않을 수 없었는데, 이로 인해 새로운 비즈니스 모델의 탄생과 서비스 질의 향상도 더불어 이루어졌다고 할 수 있다.

음악 추천의 초기에 각종 온라인 음악 업체들은 전통적인 오프라인 시장의 상위 100곡에 대한 목록을 제공하는 서비스와 유사한 "TOP 100" 서비스를 제공하는데, 기존에 음반 판매량과 라디오 방송 횟수로 결정했던 오프라인 시장에서의 집계 방식을 음악 다운로드 횟수, AOD 청취 횟수, 음반 판매량의 집계로 확장하여 서비스 하였으며([그림 1-1]), 여기에 사용자들이 직접 참여하여 제작하는 컴필레이션 앨범 형태인 공개 앨범이나 테마 앨범 같은 서비스도 같이 제공 하고 있다([그림 1-2)].



[그림 1-1] 인기순위 100곡 서비스



[그림 1-2] 공개 앨범 서비스

실질적인 음악 추천은 국내보다는 해외에서 먼저 시작하였다.

“뮤직 게놈 프로젝트” 인 Pandora.com 은 인터넷 라디오 서비스로 뮤지션과 음악 전공자등 음악 전문가들을 고용하여 장르와 가수, 앨범과 같은 일반적인 메타 정보를 포함하여 음악이 가지고 있는 속성을 약 400개의 항목으로 확장 분류하였다. 이렇게 입력된 속성에 따라 만들어진 가수나 곡의 ″DNA″를 통하여 사용자들이 요구하는 항목이나 유사한 음악을 재생리스트로 만들어 주는 추천을 제공하고 있다. 하지만, Pandora.com 또한 수동으로 데이터베이스를 구축하여야 하기 때문에 아직 전 세계 음악을 모두 포함하기엔 역부족이다.

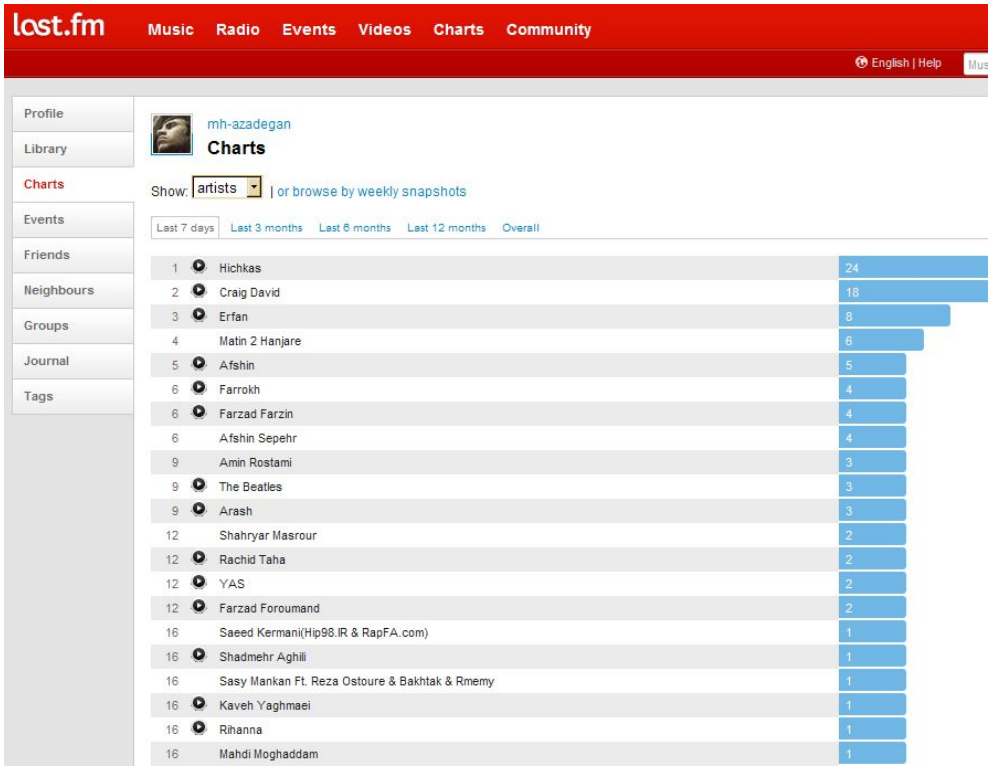
이밖에는 음악의 미디어 콘텐츠의 특징을 이용한 추천 방식을 사용하는 Amplified가 있으며, 애플의 iTunes Genius의 내부는 정확하게 밝혀진 것이 없으나, 애플의 공식 발표와 같이 Pandora와 같은 서비스라는 것과 현재, 인터넷을 통해 게재되어 있는 글을 통해서 장르와 연도에 민

감한 협업 필터링 방식의 추천이라는 정도만 예측이 가능하다. 그러나 Pandora와 마찬가지로 현재 국내 가요를 지원하지 않는다는 점과 잘 알려진 음악에 대한 추천에 강하다는 단점이 있다([그림 1-3]).

국내는 Pandora와 같이 음악 메타 데이터를 구축하여 추천 서비스를 하는 KTH의 Aurora와 같은 서비스가 있다. 이들 서비스는 대체로 내용 기반 음악 검색 시스템과 콘텐츠 필터링을 결합 시킨 형태로 개인의 취향을 반영한 추천 시스템의 의미보다는 개인화된 검색 시스템에 더 가깝다고 할 수 있다.

[그림 1-3] 애플 iTunes Genius

Last.fm은 커뮤니티를 통해 사용자들끼리 음악 정보를 교환하고, 태그를 통해 음악을 손쉽게 찾을 수 있다. 이러한 사용자의 활발한 참여에 의해 음악을 청취한 횟수와 사용자 스스로 추천과 비추천한 음악과 입력한 태그를 바탕으로 실질적인 사용자 기반의 협업 필터링 방식의 음악 추천을 제공하고 있다([그림 1-4]).


last.fm
Music
Radio
Events
Videos
Charts
Community
English | Help
Profile
Library
Charts
Events
Friends
Neighbours
Groups
Journal
Tags
mh-azadegan
Charts
Show: artists | or browse by weekly snapshots
Last 7 days
Last 3 months
Last 6 months
Last 12 months
Overall
1 Hichkas 24
2 Craig David 18
3 Erfan 8
4 Matin 2 Hanjare 6
5 Afshin 5
6 Farrokh 4
6 Farzad Farzin 4
6 Afshin Sepehr 4
9 Amin Rostami 3
9 The Beatles 3
9 Arash 3
12 Shahryar Masrour 2
12 Rachid Taha 2
12 YAS 2
12 Farzad Foroumand 2
16 Saeed Kermani(Hip98.IR & RapFA.com) 1
16 Shadmehr Aghili 1
16 Sasy Mankan Ft. Reza Ostoure & Bakhtak & Rmemy 1
16 Kaveh Yaghmaei 1
16 Rihanna 1
16 Mahdi Moghaddam 1

[그림 1-4] Last.fm

1.2 연구목적

1.2.1 협업 필터링의 문제점

협업 필터링 방식의 음악 추천은 개인의 음악에 대한 선호를 다른 개인의 선호와 비교하여 유사도가 높은 사용자를 찾아 서로 이력이 없는 콘텐츠에 대한 예측을 통해 상호 추천하는 방식이다.

오프라인에서의 추천은 지인을 통해 질문이나 의견을 듣는 형태였다면, 온라인에서의 추천은 시간과 공간의 제약이 없기 때문에 더욱 다양한 형태의 추천이 가능하며, 오프라인에서의 추천의 형태를 더욱 확장시킨 개념이 협업 필터링이다.

현재까지 연구되어온 협업필터링은 온라인 환경에서의 자동화된 개념에 초점이 맞추어져 데이터의 확장성과 희박성을 개선하는 연구(손창환, 2006)와 온라인에서의 속도 개선을 위한 연구 그리고, 추천 정확도 개선을 위한 음악 간의 가중치를 고려한 연구 등이 진행되어 왔다(손창환,2006)(이희춘, 2009)(김동문 & 박교현 & 이동훈 & 이지형, 2007)(이승화, 2008). 그러나 추천의 의미가 온라인에서의 묻고 답하기와 검색에서 더 개인화된 서비스로 진보된 것이라면 다른 관점에서의 협업 필터링의 개선이 필요하다.

실제로 2009년 포털의 지식 검색을 통해서 “비슷한 음악”, “유사한 음악”, “~와 같은 음악”이라는 키워드로 검색한 경우 백만 건 이상의 검색 결과가 나왔다([표 1-2]). 이는 사용자들이 어떤 경로로 알게 된 곡의 유사한 분류에 대한 요구가 높다는 것을 알 수 있다.

[표 1-2] 포털 사이트 검색 결과

사이트 명	키워드	건수
국내포털사이트	비슷한 음악 (곡, 노래)	248,638건
	유사한 음악 (곡, 노래)	27,638건
	같은 음악 (곡, 노래)	1,358,982건
합계		1,635,258건

이와 같이 사용자의 요구사항은 실제로는 잘 모르는 분류의 음악에 대해서 자신이 알고 있는 단편적인 정보를 토대로 접근하려는 것을 알 수 있다. 일반적으로 이렇게 요구되는 분류의 음악은 사용자가 즐겨 듣거나 다량 보유하고 있는 음악이 아닐 확률이 높다. 이러한 경우 기존의 유사한 이력을 포함한 사용자들 간의 협업 필터링 방식은 알맞지 않다.

협업 필터링은 사용자들의 성향에 따라 사용자들을 분류하여 추천을 하는데, 사용자들의 이력이 적거나 보통인 사용자(라이트 사용자)와 사용량이 많거나 전문적 지식을 보유한 사용자(헤비 사용자)로 나눌 수 있다. 만약, 라이트 사용자와 라이트 사용자가 유사도가 높아서 상호 추천을 하게 되는 경우, 서로 보유한 곡의 수가 적어 추천 횟수가 미약하다. 그리고 만약, 헤비 사용자와 헤비 사용자의 유사도가 높아서 상호 추천을 하게 되는 경우는 이미 일정 수준의 전문적 지식을 보유한 분야에 대한 추천이 될 가능성이 높기 때문에 의미가 없을 수 있다. 그러므로 라이트 유저와 헤비 유저의 유사성이 높은 경우에만 효과적인 추천이 가능하다는 단점이 있다. 그리고 실제 서비스에서의 음악 추천의 경우 특수한 상황이나 분위기 그리고 장르별 분류에 의한 추천의 요구가 있을 경우와 새로운 곡에 대한 추천의 요구가 있을 수 있는데, 전자의 경우는 미디어 콘텐츠의 유사 검색이나 이미 기획된 재생 목록을 추천해 주는 콘텐츠 필터링으로 요구를 만족시킬 수 있기 때문에 개인의 취향을 분석하는 협

업 필터링은 후자 즉, 새로운 곡에 대한 요구를 충족시키는 형태로 연구되는 것이 바람직하다.

1.2.2 협업 필터링 개선 방법 제안

앞서 언급한 사용자가 추천 받고자 하는 음악은 해당 분류에 대한 지식이 없을 확률이 높다는 점과, 헤비 사용자로부터 라이트 사용자에 대한 추천이 효과적이라는 점, 그리고 끝으로, 협업 필터링을 통한 추천은 새로운 곡에 대한 반영되는 것이 가장 바람직하다는 세 가지 문제점에 대한 해결을 위해서 다음과 같은 방법을 제안한다.

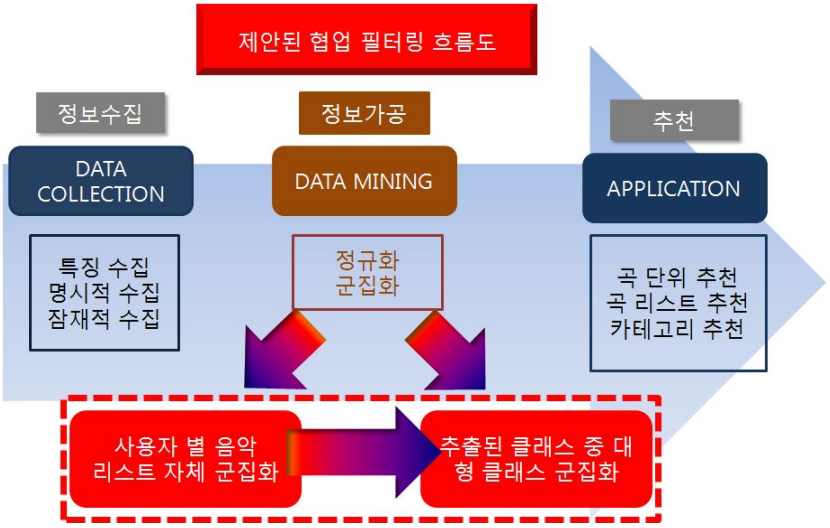
기존의 사용자 기반 협업 필터링에서는 다른 사용자와 유사성을 구하기 위해서 유클리디안 거리 또는 각종 군집화 기법을 이용하여 사용자를 분류하였는데, 여기서는 사용자는 단순한 한가지의 일관된 취향이 아닌 다양하고 복잡한 여러 가지의 취향을 가지고 있다는 점을 전제로 하여, [그림 1-5]의 정보가공 단계와 같이 사용자의 음악 목록을 먼저, 여러 개의 클래스로 나눈다. 나누어진 클래스 중 보유 목록이 많은 클래스는 추천 요구가 낮은 사용자 취향의 주된 부분을 나타내는 즉, 헤비 사용자의 역할을 할 수 있는 클래스이고, 보유 목록이 적은 클래스는 추천 요구가 높은 라이트 사용자의 역할을 하는 클래스라고 가정을 한다. 이때 사용되는 군집 기법은 K-Means 방법을 사용하는데, 그 이유는 실제 서비스에서 개인 사용자의 목록은 전체에 비해 수량이 적고 실시간 변동하기 때문에 가장 속도가 빠르고 간단한 알고리즘을 선택하였다.

추출된 클래스를 추천에 사용하기 위한 군집화에는 Self-Organizing Map(SOM)을 사용하였는데, SOM은 이미 학습 된 맵을 이용하기 때문에 계산이 많고, 실시간 변동되는 데이터에 대한 시스템의 확장성 문제

를 해결 할 수 있다(Tae Hyup Roha & Kyong Joo Ohb & Ingoo Han, 2003).

학습에는 각 개인마다 음악목록의 수량 차이가 존재하기 때문에 어떤 사용자에게는 라이트 클래스 이지만 다른 사용자의 헤비 클래스 보다 큰 경우도 있을 수 있기 때문에 전체 클래스 중 학습의 의미가 떨어지는 적은 클래스를 제외한 나머지 클래스를 가지고 군집화 하였다.

추천에는 각 개인별로 보유 목록이 적은 추천의 요구가 상대적으로 높다고 판단되는 클래스를 대상으로 추천을 받는다. 기존의 협업 필터링 방식에서 단지 각 개인의 전체 음악목록을 사용하여 사용자 간의 유사도를 측정 하였다면 제안되는 방법은 각 개인의 음악목록에서 사용자의 추천 요구가 큰 클래스를 찾기 위해서 사용자 단위의 유사 분석 전에 각 개인의 음악목록을 분석하는 단계가 추가 되었고, 이로 인해 사용자에 기반을 둔 클래스별 선호도 분석이 되게 되었다.



[그림 1-5] 제안된 협업 필터링 흐름도

1.3 논문의 구성

본 논문의 구성은 총 5장이며 각 장의 주요 내용은 다음과 같다.

2장에서는 본 논문에서 사용자의 음악목록을 분석하기 위해서 사용되는 음악 특징의 선택과 유사 분석을 위한 군집 기법인 K-Means 방법과 SOM을 설명하고 끝으로, 협업 필터링 관련 콘텐츠 기반 필터링(Content-Based Filtering), 아이템 기반 협업 필터링(Item based Collaborative Filtering), 사용자 기반 협업 필터링(User based Collaborative Filtering)을 소개한다. 3장에서는 본 논문에서 제안한 사용자의 재생 목록을 분류하고, 실제 이웃 간의 유사성을 측정하는 방법에 대해서 설명하고 평가 방법에 대해서 기술한다. 4장에서는 실험에서 사용된 데이터와 실험 방법 및 결과에 대해서 기술하고 마지막으로 5장에서는 결론 및 향후 연구 과제에 대해 기술한다.

제 2 장 관련연구

2.1 특징 연구

음악 추천을 위한 특징은 사용자의 음악에 대한 평가 점수를 이용하는 것이 전통적이 협업 필터링 방식에서 사용하는 값이었다. 그러나 마이닝을 통한 선호 점수를 예측하는 방식을 사용하여, 고객의 이력 데이터를 이용하는 방법이 보편화되었다(Nichols, D. M, 1997)(정준 & 김용환 & 이필규, 2000). 또한, 협업 필터링의 단점을 해결하기 위해서 혼합 필터링의 등장으로 사용자 정보, 내용 기반 검색에서 사용하는 디지털 오디오 특징 등의 사용까지도 제안되었다.

2.1.1 사용자 정보

사용자의 나이, 성별, 거주 지역, 학력, 직업 등과 같은 정보와 설문을 통한 취미, 성격 등 다양한 정보를 수집하여 이용한다.

2.1.2 사용자 이력

음악 콘텐츠에 대한 사용자의 선호 점수나 추천 여부 등이 있을 수 있으며, 묵시적인 이력 정보로 대표적인 것은 RFM(Recency, Frequency, Momentary)를 이용하는 방법도 있다. 그 외에 구매여부, 청취 횟수, 댓글 분석, 청취 기간, 청취 시간대 등 다양한 데이터를 마이닝하여 이용한다(강부식, 2002).

2.1.3 음악 콘텐츠 메타 데이터

사용자는 음악을 구입할 때 가수나 앨범 등의 정보에 영향을 받는다(이승화, 2008). 그러므로 메타 데이터에 가중치를 주어 연관성을 분석한다. 해당 메타 데이터로는 가수, 앨범명, 발매일자, 가수 형태, 발매사, 장르, 작곡, 작사, 편곡 등이 사용되고 있다.

2.1.4 내용 기반 특징

일반적으로 장르 검색 또는 음성 인식 등을 위해서 사용되던 특징들로써 Spectral Centroid, Spectral Rolloff, Spectral Flux, ZCR(Zero Crossing Rate), MFCC(Mel-Frequency Cepstrum Coefficent), LPC, PSF(Perceptual Spectral Flux), BPM(Beats Per Minute) 등이 있다. 해당 특징은 내용 기반 검색에 있어서 시간 구간에 민감한 특징이기 때문에 시간이 어긋나는 경우는 전혀 다른 곡으로 오인식 하는 경우가 있다. 그러므로 특정 구간의 통계 값을 특징으로 하는 경우가 많은데, 음악 장르를 구분하기 위한 구간 추출로 MFCC 1-13차 계수를 이용하여 테스트 결과 음악의 10 ~ 60초 사이의 평균과 분산이 가장 좋은 결과 나왔다는 연구결과가 있다(손창환, 2006)(김현주 & 유민준 & 이인권, 2008). Spectral 과 MFCC 등을 이용한 추천 방법에서는 음악의 전후 10%씩을 잘라내고 나머지 구간의 평균과 분산을 음악 추천 특징으로 사용한 경우도 있다(이승화, 2008).

2.1.5 특징 선택

일반적으로 패턴 인식 분야에서는 효율적인 특징을 선택하는 것이 중요한 과정 중 하나이다. 특징의 선택을 위해서는 mRMR(minimum Redundancy Maximum Relevance Feature Selection) 방법과 CMIM(Conditional Mutual Information Measurement), I-Relief, 유전자 알고리즘 등을 사용하는데 본 논문에서는 개인의 음악목록의 각 곡간의 유사성을 분석하기 위해서 시간 구간에 영향을 받지 않는 연속된 값만을 취하여 mRMR 기법을 이용하여 K-Means 알고리즘에서 사용할 가장 최적화된 특징을 선택하였다(Hanchuan Peng & Fuhui Long & Chris Ding, 2005)(Francois Fleuret, 2004)(Yijun Sun & Jian Li, 2006).

mRMR 기법은 특징들 간의 상호정보(Mutual Information)를 기반으로 한 것으로, 클래스와 특징의 연관성(Relevance)을 계산하고 특징과 특징의 잉여성(Redundancy)을 계산하여 두 값의 차이가 클수록 잉여성은 적고 연관성이 커지는 특징을 좋은 특징이라고 보는 방법이다.

식 (1)은 연속된 값에서 잉여성(Minimize Redundancy)을 구하는 식이다.

$$\min W_c, \quad W_c = \frac{1}{|S|^2} \sum_{i,j \in S} |C(i,j)| \quad (1)$$

S는 특징셋

$C(i,j)$는 특징 i와 j의 상호정보

식 (2)는 연속된 값에서 연관성(Maximize Relevance)을 구하는 식이다.

$$\max V_F, \quad V_F = \frac{1}{|S|}\sum_{i \in S} F(i,h) \quad (2)$$

S는 특징셋

$F(i,h)$는 특징i가 클래스h에 대한 $F-statistic$

식 (3)은 두 값의 차이를 계산하여 좋은 특징 값을 계산하는 수식이다.

$$\max(V/W) \quad (3)$$

2.2 군집화 기법

군집화는 동일한 클러스터에 속하는 데이터들 간에 어떤 미리 정의된 어떤 특징들의 유사도 가장 큰 것들 끼리 모아서 여러 개의 집합을 만드는 것을 말한다. 일반적으로 군집화에서 데이터들 간에 유사도 계산 방법에는 거리에 기반을 둔 척도가 주로 사용되고 있다.

2.2.1 유클리디안 거리

유클리디안 거리(Euclidean distance) 측정법은 거리 측정법 중 가장 기본적인 측정법으로 군집간의 거리 또는 데이터 간의 거리를 직선의 길이로 측정한다.

$$d_{ij} = d(X_i, X_j)$$

- $d_{ij} \geq 0$, $d_{ij} = 0$
- $d_{ij} = d_{ji}$
- $d_{ik} + d_{jk} \geq d_{ij}$

을 만족할 때,

$$d_{ij} = \sqrt{(X_i - X_j)'(X_i - X_j)} \quad (4)$$

2.2.2 K-Means 알고리즘

K-Means 알고리즘은 클러스터링을 위해 자주 사용되는 방법으로 구하고자하는 클러스터의 개수를 K라 하면, K개의 임의의 중심 벡터를 선택 한 뒤, 이 벡터들을 기준으로 가장 가까운 거리의 벡터들을 그룹화하는데, 이때 사용되는 거리 측정법은 유클리디안 거리를 사용한다. 모든 과정이 끝난 후에, 다시 각각의 그룹들의 중심 벡터를 계산해서 다시 모든 벡터들을 재계산하여 그룹화하는 과정을 반복하여 중심 벡터의 이동이 없으면 진행을 멈추는 알고리즘이다.

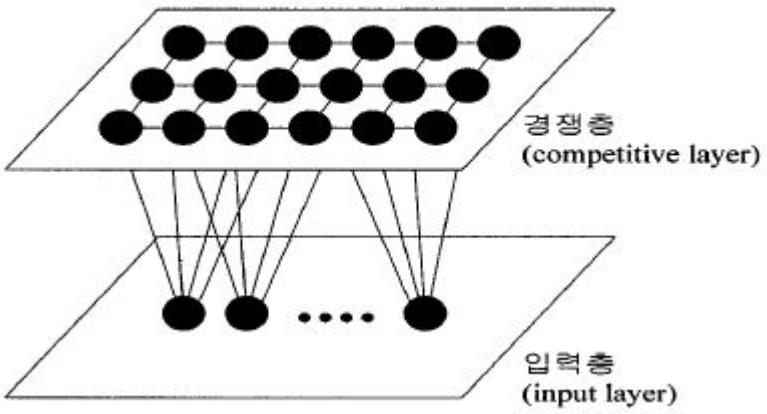
본 논문에서는 사용자 개인의 음악 목록을 분석하는데 협업 필터링에서 자주 사용되는 K-Means 알고리즘을 선택하였다.

K-Means 알고리즘은 계산 속도가 빠르고, 단계가 복잡하지 않은 장점이 있기 때문에 각 사용자별 음악 목록을 분석할 때 용이하다(김동문 & 박교현 & 이동훈 & 이지형, 2007).

2.2.3 SOM

SOM 신경망은 1984년 핀란드 헬싱키 공과대학의 Kohonen이 제안한

무감독 학습(Unsupervised Machine Learning) 기법으로, 인간의 대뇌피질은 각 영역별로 서로 다른 기능을 하도록 조직화 되어 있는데, 이러한 영역들을 피질영역이라고 하고, 감각영역, 운동영역, 연합영역 등으로 나뉘어져 있다(Kohonen, T, 1988). 또한, 대뇌피질의 각 영역 내부에서도 세부적인 기능별로 영역이 나뉘어 있는데, 대뇌 시각피질의 경우, 시각자극의 방향에 따라 서로 다른 영역이 반응하도록 되어 있다. 이러한 형태의 분할은 대뇌의 신경 세포들이 학습을 통해 스스로 조직화 되는 과정을 통해 이루어지게 되는 형태를 모방하여 입력 정보들을 서로 유사한 것을 조직화하는 신경망이다. SOM은 입력층과 경쟁층(출력층)으로 구성되어 있고, 경쟁층의 모든 벡터들은 서로 완전 연결되어 있고, 2차원의 격자 형태로 배열된 구조를 가진다.

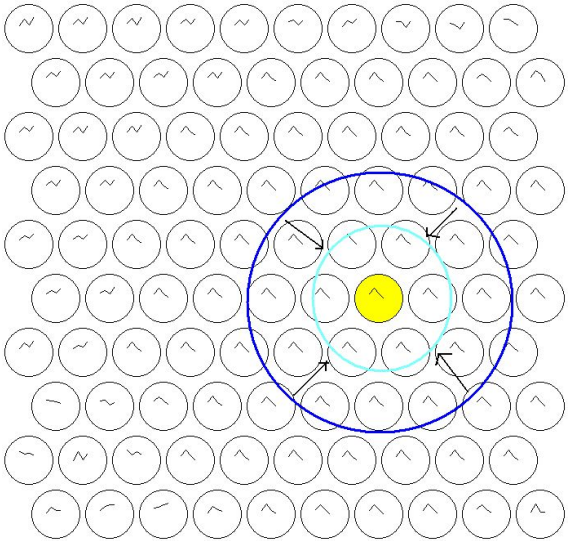


[그림 2-1] SOM의 구성 형태

입력층으로부터 하나의 입력이 주어지면, 경쟁층의 각 노드들은 자신이 가지고 있는 가중치 벡터와 입력이 얼마나 유사한지 계산한다. 이때 사용되는 계산 방법은 유클리디안 거리를 사용하며, 모든 출력 층 노드

들에 대해 입력벡터와의 거리가 계산되면, 그 중 가장 거리가 가까운 즉, 가장 유사한 가중치 벡터를 가진 경쟁노드가 승자로 결정된다.

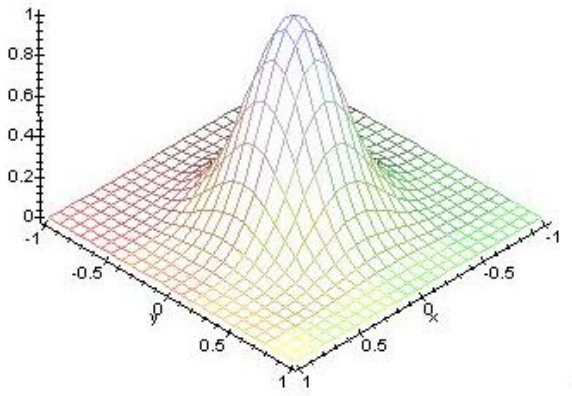
경쟁에 의해 승자가 결정 되면, 승자노드를 중심으로 학습이 일어난다. 학습은 승자로 선택된 경쟁 노드의 가중치가 현재 주어진 입력에 더 가까워지도록 가중치를 수정하는 방식으로 진행된다. 이때 가중치가 크게 변하는 것을 방지하기 위해서 적절한 상수 값을 할당하는데, 이를 학습률이라 부르며, 학습이 진행 될수록 해당 상수 값을 감소시켜 학습 횟수가 증가 할수록 점점 학습률이 감소하게 된다. 이때 승자로 선택된 노드의 이웃들도 함께 가중치를 조정하게 되는데, [그림 2-2]와 같이 이웃반경도 또한 학습이 진행 될수록 감소하여야 한다.



[그림 2-2] SOM 학습 시 이웃 반경의 감소

또한, 이웃반경의 경우 가까운 이웃일수록 더욱 영향을 크게 받아야

하고 멀어질수록 영향을 적게 받아야 하는데 이때 사용하는 이웃 커널 함수로 [그림 2-3]의 형태인 가우시안 함수가 사용된다.



[그림 2-3] 가우시안 함수

SOM의 학습을 다시 정리하여 보면 다음과 같다.

1단계, 초기화

학습을 수행하기 전에 각 경쟁노드의 가중치를 0 ~ 1 사이의 작은 값으로 초기화 한다. 또한, 학습률과 이웃반경 가중치도 초기화하여 초기에 승자노드는 영향을 많이 받고 이웃반경도 크게 잡아 많은 경쟁노드가 학습에 참여할 수 있도록 한다.

$$Vc_{mn} = Random(0,1) \quad (5)$$

$$\eta(t_0) = 0.9$$

$$\sigma(t_0) = 0.9$$

Vc: 경쟁노드벡터, $m \times n$노드수, η: 학습률, σ: 이웃반경

2단계, 승자 노드 결정

입력층에 제시된 노드를 각 경쟁층의 전 노드와 비교하여 가장 유사한 승자 노드를 결정한다. 이때는 앞서 언급하였듯이 가장 가까운 거리의 노드가 승자 노드가 된다.

$$Win(X_i) = \arg min|X_i - Vc_j| \quad (6)$$

X : 입력노드

3단계, 학습

경쟁층의 승자노드와 그 이웃노드의 가중치를 수정한다.

$$Vc_{\neq w}(t+1) = Vc_{old}(t) + \eta(t) \times \lambda_{i.old}(X_i - Vc_{old}(t)) \quad (7)$$

t : 학습횟수, η: 학습률, λ : 이웃반경

이때, 학습률은 t가 증가함에 따라 식 (8)과 같은 형태로 계산된다.

$$\eta(t) = \eta_0 \times \exp(-\frac{t}{Const_\eta}) \quad (8)$$

$Const_\eta$: 상수값

이웃 반경 또한, t가 증가함에 따라 이웃반경이 줄어들면서 영향이 줄어들어야 하기 때문에 입력노드 벡터와 경쟁노드 벡터의 거리를 구하고 이웃반경을 조절하는 값으로 나누어 주면, 식 (9)과 같은 형태로 계산된다.

$$\lambda(t) = \exp(-\frac{dist(X_i - Vc_k)}{2\sigma^2(t)}) \quad (9)$$

σ: 이웃반경조절

이때, 이웃반경을 조절하는 값은 식 (10)과 같은 형태로 구해진다.

$$\sigma(t) = \sigma_0 \times \exp(-\frac{t}{Const_\sigma}) \quad (10)$$

$Const_\sigma : \sigma$에 대한상수값

이러한 학습률과 이웃반경에 대한 조정은 모든 경쟁층의 노드가 가중치 수정이 한 번씩 수행 된 후에 업데이트 되는 것이 바람직하다.

이렇게 해서 3단계를 마치면 2단계로 돌아가서 다시 학습하게 되며, 가중치의 수정이 충분히 작아지면 학습을 멈추게 된다.

본 논문에서는 각 사용자별로 K-Means를 통해서 기 분류된 음악 목록을 추천 대상으로 설정하고 유사 목록을 구분하기 위해서 SOM 알고리즘을 사용하였다. 해당 알고리즘은 학습을 통해 협업필터링의 단점인 확장성 문제를 해결하기 위해 이미 제안된 바 있다(Tae Hyup Roha & Kyong Joo Ohb & Ingoo Han, 2003).

2.3 협업 필터링

2.3.1 내용 기반 필터링

내용 기반 필터링 방법(Content-based Filtering)은 콘텐츠 속성을 임의의 값이나 문자로 정규화 하여 고객의 정보와 유사도가 가장 높은 콘텐츠를 찾는 방식으로 추천이 간편하고, 속도가 빠르나 콘텐츠와 개인의 정보를 사용하여 추천을 하는 단순한 형태이기 때문에 개인의 취향을 고려하지 않은 일관적인 추천만을 할 수 밖에 없다는 문제점이 있다(Claypool M & A Gokhale & T. Miranda & P Murnikov & D Netes & M. Sartin, 1999).

2.3.2 협업 필터링

일반적으로 오프라인 상에서 추천을 받고자 할 때는 주위의 지인들에게 질문을 하거나 추천을 부탁한다. 이러한 개념이 온라인상에서는 시공간의 제약이 없이 다양한 형태로 나타날 수 있는데 지인 뿐 아니라 자신과 유사한 취향을 가진 어떤 사용자들의 집합을 발견하여 정보를 교환할 수 있다는 개념이 협업 필터링의 시작이다.

협업 필터링은 크게 3단계로 나뉘는데, 정보를 수집하는 단계와 정보를 해석하여 가공하는 단계, 그리고 추천 하는 단계로 나눌 수 있다.

정보를 수집하는 단계에서 초기에는 전문가 집단이나 아이템에 대한 점수 또는 사용자들의 평가 정보를 수집 하였는데, 사용자들이 좋아하는 정도에 따라서 5단계에서 7단계의 점수를 통해 평가하도록 하였다(Hill W & L, S & M, Rosenstein & G Furnas, 1995)(Paul, R & Neohphtos, I & Mitesh, S & Peter, B & John R, 1994). 이러한 사용자의 평가 부여 방법은 사용자들의 참여율 문제, 참여 보상에 대한 비용 문제, 그리고 성실한 답변의 진위 여부 등의 문제로 인해 사용자들의 이력 즉, 묵시적

인 정보 수집이 제안 되었다(Nichols, D. M, 1997). 이러한 묵시적 정보로 사용자 행위에 대한 이력을 나타내는 RFM(Recency, Frequency, Monetary)모델을 사용하여 음악 추천을 제공하는 방법이 제안되기도 하였다(Bigus JP & Bigus J, 1998). 그 밖에 음악 선택의 선호 관련 요소인 오디오 특징인 Signal Magnitude, Spectral Centroid, Spectral Roll-off point, Spectral Flux 및 MFCC의 6개 계층의 각각의 평균과 표준편차와 콘텐츠 구매자들의 특징인 성별과 나이의 평균과 표준편차, 그리고 음악 콘텐츠 메타정보인 앨범명, 가수, 발매일, 장르, 앨범 점수(사용자 평가), 앨범 추천인 수, 곡 점수, 곡 추천인 수 및 주 단위 인기곡 선정 여부 등을 특징으로 하여 연관도를 계산하여 사용한 방법이 제안되기도 하였으며, 온톨로지에서의 태그를 이용한 가중치 분석 기법도 있었다(이승화, 2008)(신동민 & 이재원 & 이경종 & 이상구, 2008).

정보를 해석하여 가공하는 단계에서는 사용자들 간의 선호 정보를 비교하여 유사 사용자를 찾는다. 아주 일반적인 방식으로 유클리디안 거리(Euclidean distance)가 주로 이용되고 있지만, 피어슨 상관계수를 이용하는 방법도 있다(J.A. Konstan & B.N. Miller & D. Maltz & J.L. Herlocker & L.R. Gordon & J. Riedl, 1997). 모델 기반 방법으로 군집화 모델인 K-Means 방식을 사용한 방법이 제안되기도 하였는데 속도가 빠르기 때문에 작은 군집간의 계산을 위해서 자주 제안되기도 하였으며 본 논문에서 사용자 재생 목록을 분석하기에 적합한 알고리즘으로 채택하였다(김동문 & 박교현 & 이동훈 & 이지형, 2007). 그 밖에, 유사도 가중치를 계산한 후 선호도 예측을 위해서 Naive Bayesian 모델이 제안되기도 하였다(김종훈 & 김용집 & 정경용 & 임기욱 & 이정현, 2007). 그러나 실시간으로 데이터가 변동되는 온라인상에서의 거리를 이용한 방법이나 상관관계를 이용한 계산 방법은 사용자의 증가와 데이터의 변경 시 추천

시스템의 변경에 대한 복잡도 증가로 인해 이러한 문제점을 해결하고자 자기조직화 기법(Self-Organizing Map)을 이용한 추천 시스템이 제안되기도 하였는데 데이터가 충분하다면 학습을 통해서 시스템의 확장성 문제를 해결 할 수 있다(Tae Hyup Roha & Kyong Joo Ohb & Ingoo Han, 2003)(강부식, 2002).

협업 필터링은 사용자 기반 협업 필터링(User-based Collaborative Filtering)과 아이템 기반 협업 필터링(Item-based Collaborative Filtering)으로 나눌 수 있다. 아이템 기반 협업 필터링은 공통으로 평가한 아이템의 평가 정보를 활용하여 유사 선호를 가진 사용자를 찾아서 추천해 주는 방식으로 사용자 기반 협업 필터링의 단점인 데이터 희박성 문제와 초기 추천 문제는 없지만, 사용자 간의 개인 취향을 반영 하지는 못한다(M. Deshpande & G. Karypis, 2004). 사용자 기반 협업 필터링은 개인의 취향을 잘 반영할 수 있다는 장점을 가지고 있지만, 일치하는 아이템이 없으면 추천을 할 수 없다는 단점과 초기 진입한 사용자에 대한 추천 문제와 더 나아가 데이터가 희박한 사용자에 대한 추천 문제가 있다. 이를 해결하기 위해서 두 가지 이상의 필터링 방식을 혼합하여 콘텐츠 기반 필터링과 아이템 기반 협업 필터링을 결합한 하이브리드 협업 필터링 또는 콘텐츠 기반 필터링과 사용자 기반 협업 필터링을 결합한 하이브리드 협업 필터링이 제안되기도 하였고(강의영, 2007)(김정훈, 2006), 데이터 변동에 따른 재계산 비용이 많이 들어가는 확장성 문제 때문에 SOM을 이용한 방법이 제안되기도 하였다(Tae Hyup Roha & Kyong Joo Ohb & Ingoo Han, 2003). 본 논문에서는 사용자 음악 목록을 세분화하는 과정을 통해서 모든 사용자의 음악 목록이 일정 수준의 크기로 나누어지므로, 각 사용자간 데이터의 크기 차이 때문에 빚어지는 희박성 문제를 간접적으로 해결 할 수 있으며, 확장성 문제를 해결하기

위해서는 SOM을 이용한 방법을 채택하여 미리 분류된 사용자 음악 목록을 학습하여 추천 목록을 생성하도록 하였다.

제 3 장 음악 목록의 유사성 분석을 통한 협업 필터링

3.1 특징 선택

본 논문의 실험을 위해서 음악 웹 하드인 솜씨의 각 사용자별 보유 곡을 사용 하였다. 웹 하드의 경우 이미 사용자가 보유하여 웹 하드로 업로드 한 곡은 구매 했다는 전제가 있으므로, 이미 해당 사용자의 취향인 것으로 판단한다. 이에 사용자와 사용자가 보유한 음악 목록에서 특징을 추출하기 위해서는 값이 누락되어 있지 않고, 값이 연속적이며 음악의 유사성과 관계가 모호한 개인 사용자 정보와 시간 구간에 영향을 받는 특징의 사용은 제외 하였으며 mRMR 기법을 이용하여 연관성이 높고 잉영성이 낮은 순서대로 총 4개의 특징을 선별하였다.

다음 [표 3-1]은 선택된 특징들의 mRMR 값을 보여주고 있다.

[표 3-1] mRMR 결과

순위	특징 이름	max(V/W)
1	곡 공통 보유 개수	0.493
2	BPM	0.390
3	가수 형태 분류	0.361
4	앨범 발매일	0.360

다른 논문에서는 각 곡의 인기곡 선정 여부를 특징으로 사용하였다. 이와 비슷한 특징으로 실험 대상이 되는 음악 목록에서 각 사용자별로 같은 곡을 보유하고 있는 경우 해당 곡의 인기도를 반영 한다고 보고 각

곡의 사용자 별 보유 개수를 특징으로 사용하였으며, 음악에서의 특징은 템포(tempo), 음색(timbre), 피치(pitch), 강도(loudness) 등의 특징을 사용할 수 있는데 각 특징 값 중 템포에 해당하는 BPM(Beat Per Minute)이 선택되었다. 메타 데이터 특징 중 가수 형태의 경우 여자 솔로와 남자 솔로로 구간을 나누고 그 가운데 그룹을 배정하여 구간 척도로 변경하여 사용 하였다. 이 밖에 메타 데이터 특징 중 앨범 및 곡의 발매 시기가 연관성이 높고 잉여성이 낮은 특징으로 선택 되었다.

3.2 특징의 정규화

실제 추출된 특징들은 유사 선호를 구하기 위한 학습 데이터로 활용하기 위해 군집화 이전 단계에서 0 ~ 1 사이의 값으로 정규화 하였는데, 정규화 수식은 일반적인 형식의 정규화 과정을 따랐다.

다음 식 (11)은 BPM의 정규화를 계산하는 수식이다.

$$BPM_{normal} = \frac{BPM_i - BPM_{\min}}{BPM_{\max} - BPM_{\min}} \quad (11)$$

i : i번째 곡, BPM_i : i번째 곡의 BPM

$BPM_{\max}$: BPM최대값, $BPM_{\min}$: BPM최소값

다음 식(12)는 곡의 발매년도의 정규화를 위한 수식이다.

$$SD_{normal} = \frac{SD_{year_i} - SD_{year_{\min}}}{SD_{year_{\max}} - SD_{year_{\min}}} + (0.01 \times \frac{SD_{month_i} - SD_{month_{\min}}}{SD_{month_{\max}} - SD_{month_{\min}}}) \quad (12)$$

SD: $SaleDate$의 약자, i : 특정 입력 일자

발매일자의 경우 일자까지는 사용하지 않았으며, 월의 경우 년 단위에 영향을 최소로 주기 위해서 0.01을 곱해 주었다.

다음 식 (13)은 곡의 발매년도의 정규화를 위한 수식이다.

$$Dup_{normal} = \frac{Dup_i - Dup_{\min}}{Dup_{\max} - Dup_{\min}} \quad (13)$$

가수의 형태의 경우 [표 3-2]와 같이 구간을 나눈 후 정규화 된 값을 부여 하였다.

[표 3-2] 가수 형태 정규화

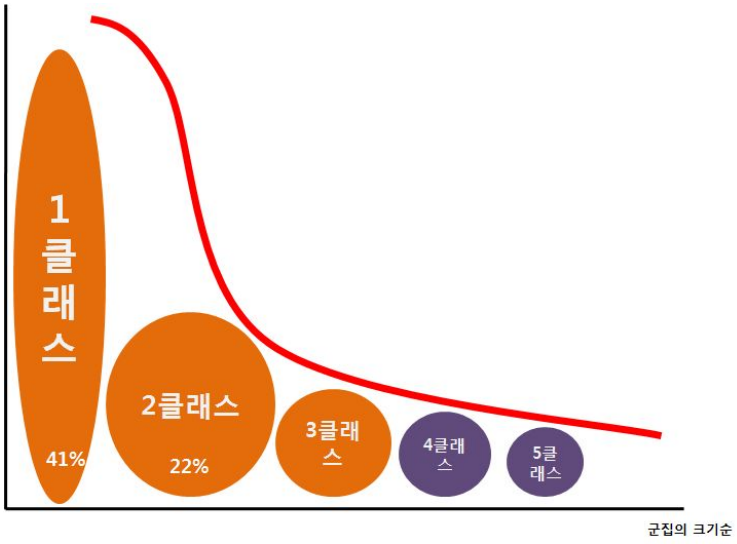
여자솔로	여자그룹	혼성그룹	남자그룹	남자솔로
0	0.25	0.5	0.75	1

3.3 음악 목록 유사성 분석

각 개인의 음악 취향은 하나의 일관적 취향으로 이루어져 있지 않다는 것을 전제로 하여 여러 개의 클래스로 사용자의 음악 목록을 나눈다. 이렇게 나누어진 클래스는 같은 크기가 아니며 상대적으로 작은 클래스는 큰 클래스 보다 더 적은 수량을 보유하고 있기 때문에 추천의 요구가 더

높다. 일반적으로 알고 있는 곡이 아닌 모르는 곡에 대해서 추천의 요구가 높으며, 모르는 곡이라 할지라도 사용자가 원하는 곡을 알기 위해서는 사용자의 이력이나 근거를 통해서 요구를 찾아야만 한다. 가장 근접한 방법은 사용자의 이력을 통해서 가장 적게 보유하고 있는 특정 음악 부류와 유사한 곡을 추천을 하는 것이다. 또한, 추천이 가장 효율적인 형태는 전문가에서 비전문가로 즉, 헤비 유저에서 라이트 유저로의 추천 형태가 가장 바람직하다. 사용자의 클래스가 전문가 클래스와 비전문가 클래스로 나누어지는 경우 추천 요구에 대한 성능이 좋아질 수 있다.

각 사용자별 유사성에 따라 분류된 클래스가 각각 차이를 보이며 [그림 3-1]과 같은 형태로 나누어 질 경우, 긴 꼬리 부분에 해당하는 클래스에 해당하는 음악 부류를 추천 하는 경우 롱테일의 법칙(Long Tail theory)과 연관된 비즈니스 효과를 얻을 수 있다.



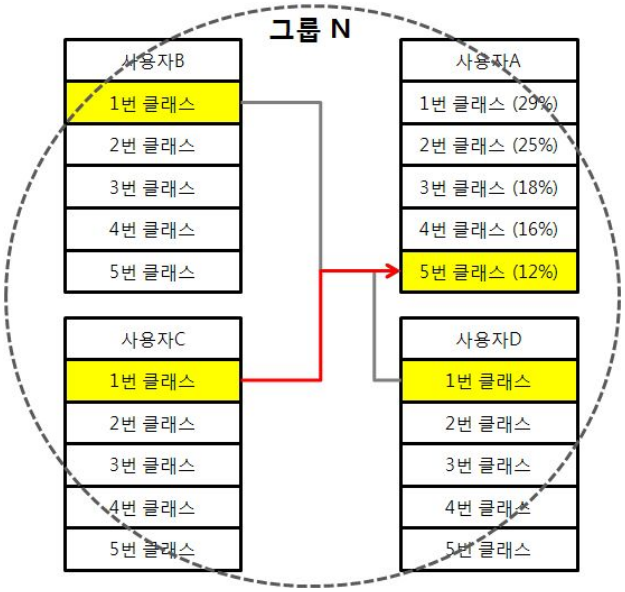
[그림 3-1] 사용자 음악 목록 유사 분석 시 클래스 예시

사용자 클래스의 유사 분석을 위해서 선택된 특징을 이용하여 군집화를 실시한다. 이때 사용하는 각각의 사용자 음악 목록을 분석하기 위한 군집화 알고리즘으로는 K-Means 알고리즘 방법을 사용하는데, 일반적으로 각 사용자에 대해서는 음악 목록의 개수가 적기 때문에 계산량이 적으나, 전체 사용자를 일일이 계산할 경우 계산 속도가 빠르고 고속화 기법으로 주로 사용되는 K-Means를 선택하게 되었다. 적은 사용자 음악 목록을 추천이 가능한 클래스는 10곡미만이 되면 해당 클래스가 가지고 있는 음악의 실질적인 유사성에 대한 특징 값의 분석이 의미가 없기 때문에 실험 결과 상위 사용자와 하위 사용자와의 음악 목록 개수의 차이로 인하여 사용자별 클래스의 개수가 5인 경우 가장 알맞은 크기의 클래스가 나왔다.

나누어진 클래스를 대상으로 SOM 학습을 실행하여 맵을 구성하는데, 학습을 위한 특징으로는 클래스의 분류에서 이미 사용하였던 각각의 특징의 평균값을 이용하여 분류를 하였는데, 각 사용자별 특징 값으로 기분류된 그룹은 이미 해당 특징을 기준으로 분류되었으므로 포함된 음악의 특징 값의 평균이 해당 그룹을 대표한다고 볼 수 있다.

3.4 음악 추천

음악 추천을 위한 맵이 완성이 되면 각 사용자의 클래스 중 상대적으로 긴 꼬리에 해당 하는 클래스를 선택하여 추천을 실행한다. 즉, 사용자가 가장 추천을 받고 싶어 하는 음악은 가장 적게 보유하고 있을 것이라 예측되는 클래스이며 그러므로 제일 작은 클래스를 추천을 받는 클래스로 선택을 하는데 [그림 3-2]와 같은 형태로 설명 할 수 있다.



[그림 3-2] 추천 클래스 선택

협업 필터링에서 상호 추천을 하는 사용자를 찾는 방법은 유클리디안 거리나 피어슨 상관계수를 이용한다. 선택된 사용자는 기준 사용자와 같은 곡을 가장 많이 보유하고 있다. 추천을 할 곡에 대해 각 사용자가 보유한 곡에 대한 점수의 평균과 서로 보유하지 않은 곡에 대한 점수의 차를 이용하여 예측 점수를 구한다. 이때 임계값에 따라 추천을 행하는 방식을 사용한다. 하지만, 본 논문에서는 이미 포함된 그룹에서 헤비 클래스를 선택하여 추천을 하는 방법을 제안 하였고 이미 사용자가 보유한 곡은 사용자의 취향을 반영하는 것으로 전제하여 유사 분석을 하였기 때문에 곡에 대한 추천 클래스는 식 (14)와 같은 형태로 선택한다.

$$Class(i,x) = \max\sum\left(\frac{p_{i,x}}{p_i} \times \frac{p_{i,x}}{p_{all}} + \frac{p_m}{p_{j,y}}\right) \quad (14)$$

i: 추천을 제공하는 사용자, j: 추천을 받을 사용자
x: 추천을 제공하는 클래스, y: 추천을 받을 클래스
m: 사용자 i의 x클래스와 사용자 j의 y클래스가 공통으로 포함한 곡
all: 전체 곡
p: 곡 집합(개수)

위와 같이 추천을 해주는 클래스는 각 사용자별 헤비 클래스에 속하는 보유량이 많은 클래스 중에서 추천을 받을 사용자의 클래스와 공통으로 보유하고 있는 곡이 가장 많은 클래스를 선택한다.

클래스가 선택이 되면 각 클래스에 대한 곡 추천이 필요한데 곡에 대한 추천은 추천을 받는 클래스의 곡을 기준으로 추천을 제공하는 클래스에 포함된 곡의 특징을 식 (15)와 같이 유클리디안 거리 측정법을 사용하여 추천 곡을 선택한다.

$$\min\sum_{i=1}^{n}|f_x(i) - f_y(j)| \quad (15)$$

n: 특징의 개수
x: 추천을 제공하는 클래스, y: 추천을 받을 대상 클래스
$f(i)$: i번째 특징

다음 [그림 3-3]은 사용자 A의 가장 작은 다섯 번째 클래스의 곡을 기준으로 인접한 사용자 C의 가장 큰 첫 번째 클래스의 곡과 대조하여 추천 목록을 만드는 예를 보여 주고 있다.

사용자 A의 5번 클래스

곡 제목	BPM	가수형태	공통 보유 개수	앨범발매일
Heartbeat – 2PM	0.37619	0	0.0294118	0.649885
꿈을 꾸다 – 김태우	0.461905	0.75	0.0147059	0.849452
웨딩드레스 – 태양	0.204762	0.5	0.0147059	0.833579

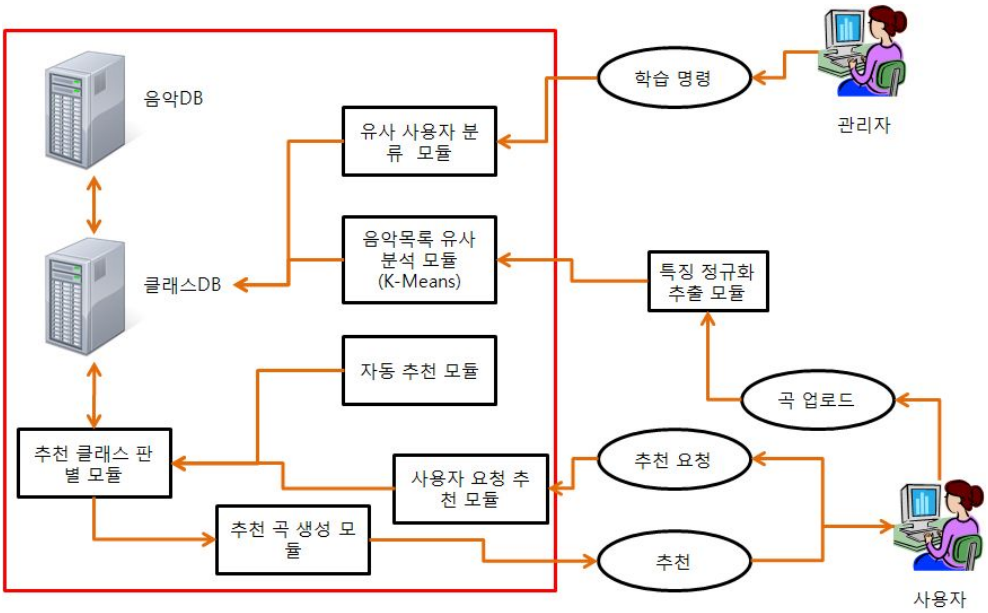
사용자 C의 1번 클래스

곡 제목	BPM	가수형태	공통 보유 개수	앨범발매일
부른다 - 서인국	0.37619	0	0.0294118	0.649885
실화 – 린	0.461905	0.75	0.0147059	0.849452
.....	0.204762	0.5	0.0147059	0.833579

[그림 3-3] 곡 추천

3.5 시스템 구조

본 논문에서 제안하는 사용자 음악 목록 유사성 분석 추천 시스템의 구조도는 [그림 3-4]와 같다.



[그림 3-4] 시스템 구조도

시스템의 구성은 사용자가 곡을 업로드 하게 되면 특징 추출 모듈에서는 특징을 추출하여 정규화한 값을 음악 목록 유사 분석 모듈로 전달한다. 이때 K-Means 알고리즘을 이용하여 실시간으로 사용자의 음악 목록은 클래스 데이터베이스로 분류되어 저장이 된다. 이때, 이미 학습된 맵에 따라 군집이 분류되는데, 유사 사용자 분류 모듈은 학습 모듈로써 관리자가 학습 명령을 실행했을 때에만 동작하도록 되어 있다. 사용자는 추천 요청을 할 수 도 있는데 이때 사용자 요청 추천 모듈을 통해 이미

학습된 클래스를 기반으로 추천을 실행하는데 자동 추천 모듈 또한 주기적으로 클래스 데이터베이스를 점검하여 추천을 행해야 하는 정해진 값에 따라 추천을 제공할 클래스를 선택하는 추천 클래스 판별 모듈을 호출하여 클래스를 선택하게 한다. 추천 클래스 판별 모듈은 클래스가 선택되면 추천 곡 생성 모듈을 호출하여 곡 목록을 작성하고 사용자에게 실제 추천을 실행한다.

제 4 장 실험 및 결과

4.1 실험 데이터

본 논문의 실험을 위해서 음악 웹 하드인 솜씨(www.somsee.com)의 음악 보유 곡을 사용하였다. 솜씨는 자신이 보유한 음악 파일만을 업로드 할 수 있는 웹 하드이며 업로드 되는 디지털 오디오 파일을 분석하여 음악 데이터베이스에 있는 메타 데이터와 매칭을 하기 때문에 업로드 된 곡이 어떤 곡인지 알 수 있다.

실험을 위해서 2008년 11월까지의 데이터 중에서 미디어 파일의 오류로 BPM 추출이 불가한 곡과 메타데이터가 누락된 곡을 제외하고 100곡 이상을 보유한 상위 111명의 데이터를 사용하였다.

음악 웹 하드는 이미 사용자가 보유하고 있어서 업로드 한 곡 이므로, 구매 한 곡으로 전제하고 음악목록 자체가 취향을 반영하므로 해당 목록을 분석하는 것으로 한다.

검증은 2009년 6월까지 111명의 사용자들이 추가로 업로드 한곡을 비교하기로 한다.

4.2 평가 방법

추천 성능 평가는 Coverage Ratio와 MAE를 사용하였다.

Coverage Ratio는 구매자 i가 구매한 곡 집합 qi 들 중 추천한 곡 집합 pi가 얼마나 있는가를 n명의 구매자에 대하여 평균한 수치이며, 식 (15) 와 같다.

$$CoverageRatio = \frac{1}{n}\sum_{i=1}^{n}\frac{|p_i \cap q_i|}{|q_i|} \quad (15)$$

i: i번째 구매자, n: 전체구매자
p: 추천한 곡 집합, q: 구매한 곡 집합

MAE(Mean Absolute Error)는 N은 총 예측 회수이며, εi는 예측 값과 실제 값의 오차를 나타내고, i는 각 예측 단계를 나타내면 식 (16)과 같다.

$$MAE = \frac{\sum_{i=1}^{N}|\epsilon_i|}{N} \quad (16)$$

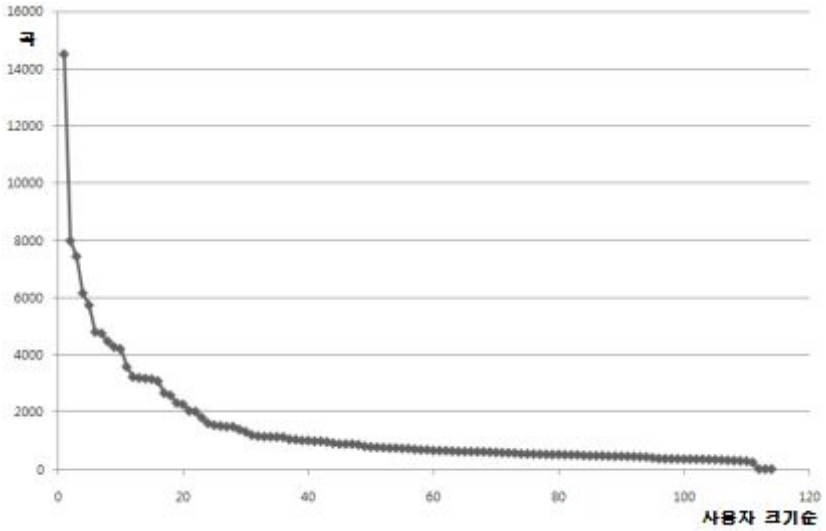
N: 총 예측 회수, ϵ: 예측값

Coverage Ratio는 자동 추천과 검증 데이터의 비교를 통해서 검증을 하고, MAE는 무작위 선출 된 10명의 사용자에게 단계별로 곡 추천을 하여 실제 선호 예측 곡과 평가 값을 토대로 오차율을 계산 한다.

4.3 실험결과

4.3.1 사용자 음악 목록 유사성 분석

초기 실험 데이터의 사용자 보유 목록은 [그림 4-1]과 같다.



[그림 4-1] 사용자별 보유 곡 개수

[표 4-1]은 [그림 4-1]의 데이터를 통계표로 나타낸 것이다.

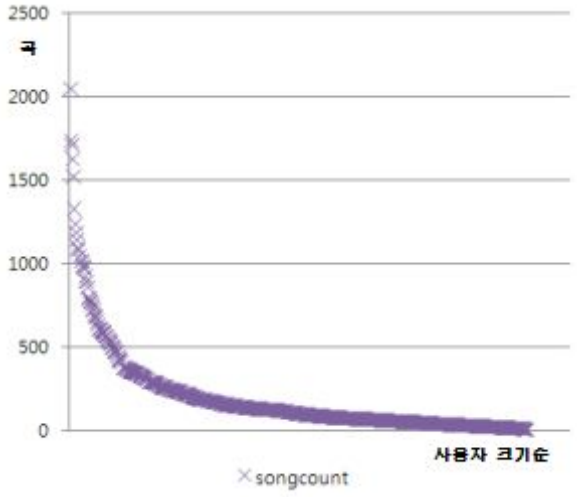
[표 4-1] 사용자별 보유 곡 개수

구분	사용자 수
3000곡 이상	5명
1000곡 이상	30명
500곡 이상	46명
100곡 이상	30명

3,000곡 이상의 사용자가 5명이상이었고, 그 중 10,000곡 이상을 보유한 사용자가 1명, 5,000곡 이상을 보유한 사용자도 4명이나 되었고 500곡 미만의 사용자가 30명이었다. 이러한 경우 협업 필터링을 그대로 적용한다면 보유 곡이 많은 사용자가 선택될 확률이 높아지므로, 성능이 안 좋게 나타날 수 있다. 이로써 실제 서비스에서의 음악 추천은 사용자의 재생 이력이나 보유 목록을 그대로 사용하는 것이 아닌 목록의 분석을 통한 클래스화가 필요하다는 것을 간접적으로 보여준다.

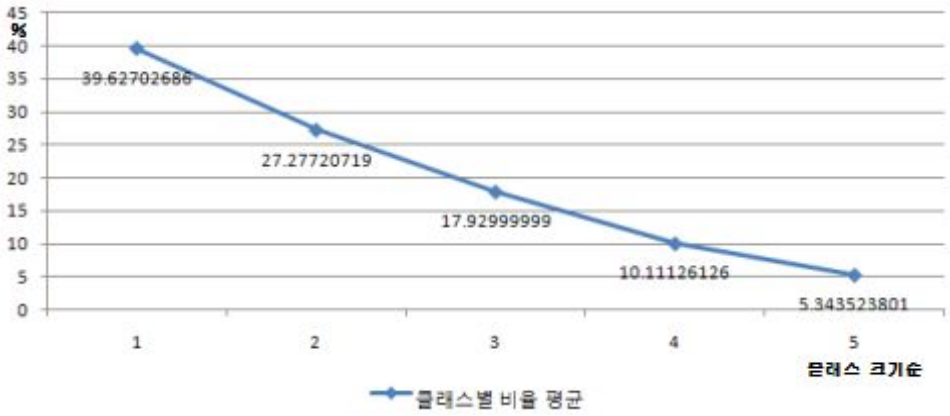
각각의 사용자 별로 K-Means 알고리즘을 이용하여 클래스를 나누었을 때, 상위 20%의 사용자의 보유량과 하위 10%의 사용자의 보유량의 차이로 인해 K의 값을 5로 했을 때 결과가 가장 좋았다.

각 사용자 별로 추출 되어진 클래스의 결과는 [그림 4-2]와 같이 전체 클래스가 2,500곡미만을 보유하게 되어 실제 추천에 적합한 형태로 분류되었음을 알 수 있다.



[그림 4-2] 클래스별 보유 곡 개수

[그림 4-3]에서 클래스별 사용자 기준 비율을 보면 이미 앞서 언급 했듯이 실제 비율이 가장 큰 클래스(1번 클래스)의 평균이 39% 그리고 가장 작은 클래스(5번 클래스)의 사용자 비율이 5%로 나와 사용자가 보유한 곡의 유사 분석에 따른 분류가 각 사용자별로 목록의 크기가 심하게 차이가 나지 않기 때문에 대체로 양호하게 되었음을 알 수 있으며 롱테일의 법칙과 같이 보유량이 가장 적은 클래스를 분석하여 유사한 곡을 추천하기 때문에 음악 추천 시 긴 꼬리 부분에 대한 추천이 가능해 졌음을 알 수 있다.

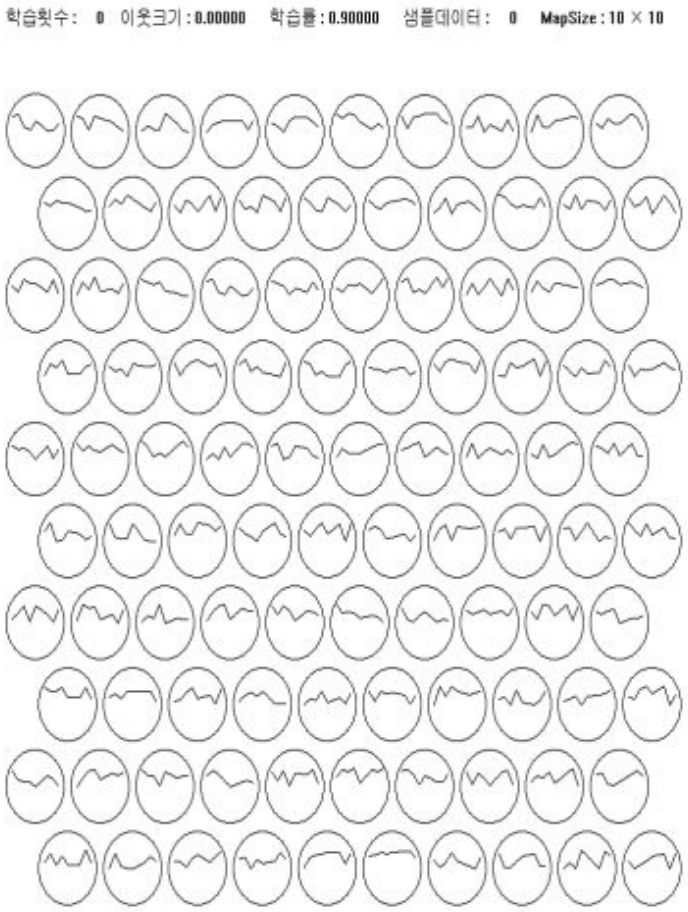


[그림 4-3] 사용자별 클래스 분포 평균

4.3.2 클래스를 이용한 SOM 학습

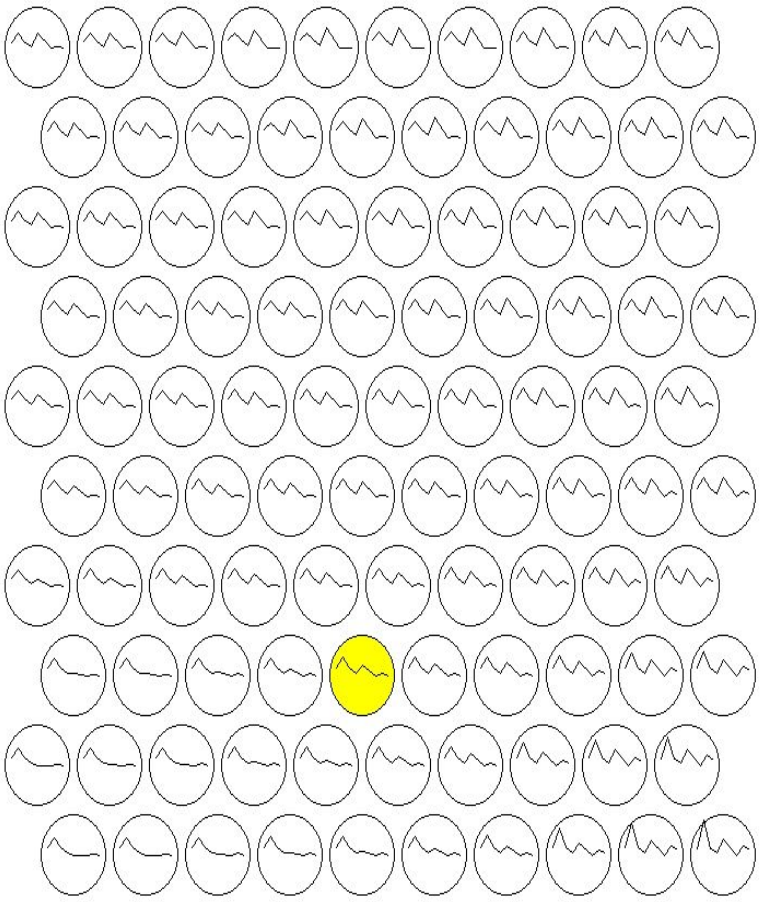
각 사용자의 클래스들 중 학습 결과에 모호한 영향을 줄 수 있는 100곡미만의 클래스를 제외하고 나머지 클래스를 이용해서 학습을 실시하였다.

[그림 4-4]는 그래프 맵을 이용하여 학습을 처음 시작 하는 단계의 초기화된 맵의 모습이다.



[그림 4-4] SOM 초기화

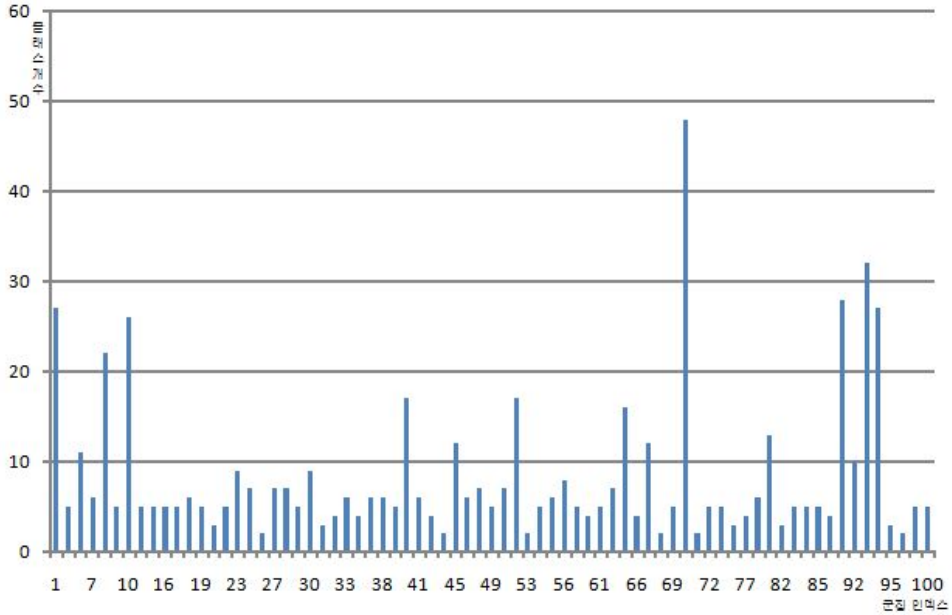
학습은 총 10,000번을 진행하였으며 [그림 4-5]는 학습이 끝난 상태이다.



[그림 4-5] SOM 학습 종료

학습은 추출된 각 클래스의 사용되어진 특징의 평균을 이용하였다.

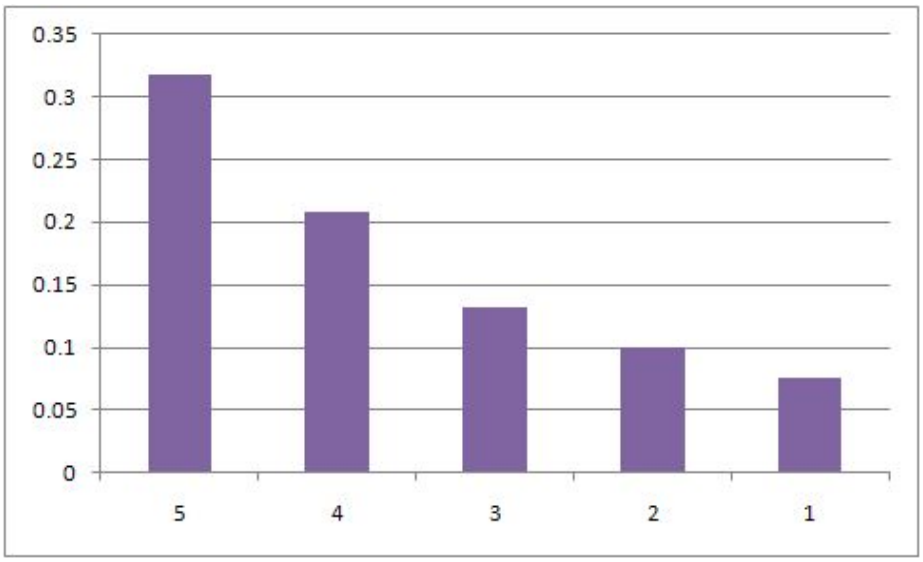
100X100 그래프 맵에서 평균을 이용한 학습결과 [그림 4-6]과 같이 총 22개의 군집에 각 클래스가 분포되었다.



[그림 4-6] SOM 학습 결과에 따른 클래스 분포

4.3.3 추천 결과

[그림 4-7]은 2008년 11월까지의 데이터를 기준으로 실제 추천 요구가 많을 것으로 예상되는 가장 작은 클래스를 대상으로 추천을 한 결과를 2009년 6월을 기준으로 그 기간 안에 실제로 사용자가 업로드 한 곡에 대한 Coverage Ratio의 값이다. 가장 큰 클래스를 임의로 1번 클래스 그리고 순서대로 가장 작은 클래스를 5번 클래스로 정하고 각각의 클래스에 대하여 추천 결과가 다음과 같이 나왔다.



[그림 4-7] 클래스별 추천 결과

실제로 사용자의 클래스 중 가장 작은 클래스인 5번 클래스에 곡을 추천 한 결과 사용자들은 6개월 뒤에 추천된 곡을 가장 많이 업로드 한 것을 볼 수 있다.

[표 4-2] 클래스별 Coverage Ratio 결과

클래스 번호	Coverage Ratio
5번 클래스	30.177
4번 클래스	20.773
3번 클래스	13.195
2번 클래스	10.027
1번 클래스	7.517

[표 4-2]와 같이 실제 곡 추천 시 가장 작은 5번 클래스부터 30.177, 4번 클래스 20.773, 3번 클래스 13.195, 2번 클래스 10.027, 가장 큰 1번 클래스 7.7517 이 나왔다. 클래스의 크기가 커질수록 추천 결과가 낮아짐을 알 수 있다. 여기서 유추해 볼 수 있는 것은 실제로 사용자들은 자신이 잘 알고 있는 곡의 유형 즉, 부류에 대해서는 더 이상 곡을 구매하거나 수집하지 않았으며, 그러므로 실제 자신이 잘 모르는 음악에 대한 추천 요구가 더욱 높다는 가정이 성공적이었음을 알 수 있다. 그리고 실제 추천된 곡의 30%를 사용자들이 이후 구매 했다는 것은 어쩌면 추천 성능이 높지 않다고 볼 수 있으나, 2008년 11월에서 2009년 6월까지 업로드된 곡을 분석한 결과 약 40% 정도가 2008년 11월 이후에 발매된 신곡이었다. 그러므로 실제 추천 결과가 높게 나오지 않을 수밖에 없다고 볼 수 있다.

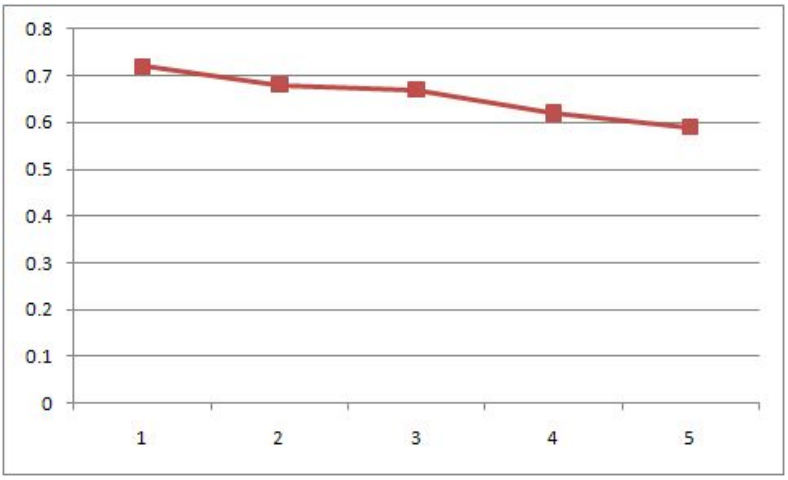
[그림 4-8]은 실제 사용자들의 추천 평가를 측정하기 위해서 자신의 음악 목록을 클릭하면 클래스를 선택할 수 있고, 자신이 가진 5개의 클래스 중에서 하나를 선택하면 각 곡에 가장 유사한 곡을 하나씩 대입하여 보여주고, 사용자에게 추천 곡에 대해서 만족 하는지 불만족 하는지를 집계하기 위한 테스트 페이지이다.

사용자: 486 구룹 5

원곡> 김동률 - 아이처럼	추천곡> Acid House Kings-This Heart Is A Stone	만족 불만족 이미알고있음
원곡> 넬 (Nell) - 기억을 걷는 시간	추천곡> Acid House Kings-This Heart Is A Stone	만족 불만족 이미알고있음
원곡> 마이티 마우스 (Mighty Mouth) - 사랑해 (Feat. 윤은혜)	추천곡> V.A. CF Drama Movie-Brown Eye Girls (영화 '적과의 동침')	만족 불만족 이미알고있음
원곡> Epik High - Breakdown	추천곡> V.A. CF Drama Movie-Brown Eye Girls (영화 '적과의 동침')	만족 불만족 이미알고있음
원곡> Epik High - One (Feat. 지선)	추천곡> Aqua (아쿠아)-Cartoon Heroes	만족 불만족 이미알고있음
원곡> Epik High - 연필깎이 (Feat. Kebee)	추천곡> Aqua (아쿠아)-Cartoon Heroes	만족 불만족 이미알고있음
원곡> Epik High - Girl (Feat. 진보)	추천곡> V.A. CF Drama Movie-Brown Eye Girls (영화 '적과의 동침')	만족 불만족 이미알고있음
원곡> Epik High - Eight By Eight (Feat. Dynamic Duo, Dok2, Double K, TBNY)	추천곡> V.A. CF Drama Movie-Brown Eye Girls (영화 '적과의 동침')	만족 불만족 이미알고있음
원곡> Epik High - Decalcomanie	추천곡> V.A. CF Drama Movie-Brown Eye Girls (영화 '적과의 동침')	만족 불만족 이미알고있음
원곡> Epik High - 낙화 (落花)	추천곡> Acid House Kings-This Heart Is A Stone	만족 불만족 이미알고있음
원곡> Epik High - 우산 (Feat. 윤하)	추천곡> V.A. CF Drama Movie-Brown Eye Girls (영화 '적과의 동침')	만족 불만족 이미알고있음
원곡> Epik High - B-Side 01 : Breakdown (Supreme Mix)	추천곡> V.A. CF Drama Movie-Brown Eye Girls (영화 '적과의 동침')	만족 불만족 이미알고있음
원곡> MC 몽 - 서커스 (Feat. 임유경 - 달래 음악단 , $howgun)	추천곡> Aqua (아쿠아)-Cartoon Heroes	만족 불만족 이미알고있음
원곡> MC 몽 - 아홉번째 구름 (Feat. 양파)	추천곡> Acid House Kings-This Heart Is A Stone	만족 불만족 이미알고있음
원곡> MC 몽 - 죽도록 사랑해 (Feat. 박정현)	추천곡> Acid House Kings-This Heart Is A Stone	만족 불만족 이미알고있음
원곡> MC 몽 - 미치겠어 (Feat. M.A.C)	추천곡> Acid House Kings-This Heart Is A Stone	만족 불만족 이미알고있음

[그림 4-8] 추천 테스트 페이지

해당 페이지를 통해 추천에 대한 만족도를 설문한 결과 다음 [그림 4-9]과 [표 4-3] 과 같이 나왔다.



[그림 4-9] 클래스별 추천 MAE 결과

[표 4-3] 클래스별 MAE 결과

클래스 번호	MAE
5번 클래스	0.59
4번 클래스	0.62
3번 클래스	0.67
2번 클래스	0.68
1번 클래스	0.72

실험 시 만족과 불만족 표시 할 때 이미 아는 곡의 경우는 추천의 효과가 없다고 판단하여 따로 표기 하도록 유도하여 해당 결과를 불만족 결과로 처리하였고, 실제 사용자들이 대체적으로 점수의 차이를 크게 두지 않고 가장 크기가 큰 1번 클래스의 경우와 3번 클래스까지 만족도의 큰 차이가 없음을 미루어 보아 긍정적인 평가를 한 것으로 예측되므로 전체적인 MAE의 값은 큰 차이가 없었다. 그러나 가장 작은 클래스인 5번과 그 다음 작은 4번 클래스가 다른 클래스와 좀 더 큰 차이로 사용자들이 추천에 대한 만족을 나타낸 것을 보더라도 실제로 사용자들이 원하는 추천은 자신이 보유한 음악들 중 더 작은 부류 즉, 잘 모르기 때문에 수집량이 적은 클래스에 대해 더욱 큰 만족이 있음을 알 수 있다.

제 5 장 결론 및 향후 연구

음악 추천은 현재도 꾸준히 연구되고 있으며 실제 서비스에서 사용하려는 요구도 더욱 늘어나고 있다.

기존 서비스에서 협업 필터링은 상호 공통된 곡의 수가 많을 경우 추천 조건이 성립하기 때문에 실질적으로 개인에게 추천을 했을 경우 단순하게 추천보다는 같은 경향의 사용자 간의 음악 교환 형식이 되어 잘 알지 못하는 새로운 음악을 접할 수 있는 기회가 떨어진다.

본 논문에서는 사용자 음악 목록의 유사성을 분석하는 단계를 추가하여 실제 추천의 요구가 클 것으로 예측되는 음악 분류의 곡을 추천 해줌으로써 사용자의 요구를 충족시켜 주려는 시도를 하였다.

실제로 사용자의 음악 목록을 분석 결과 가장 큰 클래스의 음악 목록에서 차지하는 비율이 평균 40% 정도 이었으며 가장 낮은 클래스의 비율은 평균 5%로 사용자의 추천 요구가 높은 클래스와 이미 잘 알고 있는 클래스를 실험 전 예측 했던 비율에 근접했다.

실제 음악 추천을 통해서 가장 낮은 클래스에 대해서 추천을 한 음악의 경우 추후 실제 구입하여 업로드한 곡 Coverage Ratio의 값이 30% 이상을 차지하였으며 가장 큰 클래스는 잘 알고 있는 음악이기 때문에 접근이 용이하고 구입도 편함에도 불구하고 추천 결과 7%라는 더 낮은 결과를 보여줬다. 또한, 클래스의 크기에 따라서 점점 더 낮은 비율의 결과를 보여줌을 알 수 있었다.

실제 사용자 실험을 통해서도 마찬가지로 사용자 음악 목록의 클래스 중 가장 작은 클래스 즉, 추천 요구가 클 것으로 예상되었던 클래스에서 높은 만족도를 보여줬다. 즉, 사용자 요구의 더 개인화된 추천 서비스가

가능하다고 볼 수 있다.

개인의 이력이나 재생 리스트를 토대로 음악 추천을 하고자 하는 경우 실제 서비스에서는 한 사용자가 재생하는 방대한 이력 때문에 추천을 하기 위한 과정에서 시간과 비용이 계속 증가한다.

그러나 본 논문에서 제시한 음악 목록의 유사성을 이용하여 클래스를 나누었을 경우 사용자간 음악 데이터량의 차이 때문에 실제 추천 시 기준이 모호해지는 현상과 늘어나는 데이터에 대한 확장성 문제 또한 간접적으로 해결 할 수 있다는 장점이 있다.

향후 연구과제로는 본 논문에서는 사용한 데이터는 111명의 사용자에 대한 약 6만 여곡의 데이터를 사용하여 실험 하였으나, 더욱 많은 사용자와 다양한 음악을 이용하여 실험하는 경우 좀 더 명확한 결과가 나올 수 있을 것으로 기대된다. 또한, 본 논문에서 사용자가 보유한 음악 목록을 분석하기 위해서 사용자의 취향 파악에 용이한 특징의 발굴 작업이 필요하다. 사용자가 가지고 있는 음악 목록의 취향 구분을 본 논문에서는 임의의 개수로 정하였다. 하지만 실제로는 각 사용자 마다 좋아하는 음악의 취향이 다양하기 때문에 유동적인 클래스의 수가 나올 수 있도록 사용자별 음악 목록 구분 시 개선된 형태의 클러스터링 알고리즘이 필요할 것으로 예상된다.

앞으로 사용자의 음악 목록을 구분하고 그에 따라 추천하는 다양한 방법이 제안 될 것을 기대해 본다.

참고문헌

한국 소프트웨어 진흥원(2008), 디지털 콘텐츠 산업백서.

Nicolas Scaringella & Giorgio Zoia & and Daniel Mlynek(2006), Automatic Genre Classification of Music Content, IEEE, IEEE Signal Processing Magazine, Special Issue on Semantic Retrieval of Multimedia.

손창환(2006), Web상에서 개인화된 상품추천을 위한 Hybrid 추천시스템에 관한 연구, 영남대학교 대학원, 박사학위논문.

이희춘(2009), 전자상거래 추천시스템에서 협력적 필터링의 성능 향상, 강원대학교 대학원, 박사학위논문.

김동문 & 박교현 & 이동훈 & 이지형(2007), 공간 분할 기법을 사용한 고속화된 사용자 적응형 음악 추천 시스템, 한국지능시스템학회, 한국지능시스템학회 학술발표 논문집, 한국퍼지 및 지능시스템학회 2007년도 춘계학술대회 학술발표논문집 제17권 제1호, pp. 109-112.

이승화(2008), 음악 추천을 위한 연관성 분석 방법, 숭실대학교 대학원, 석사학위논문.

Tae Hyup Roha & Kyong Joo Ohb & Ingoo Han(2003), The collaborative filtering recommendation based on SOM cluster-indexing CBR, Expert Systems with Applications 25, pp. 413 - 423.

Nichols, D. M(1997), Implicit Rating and Filtering, the 5's DELOS Workshop on Filtering and Collaborative Filtering, pp. 10-12.

정준 & 김용환 & 이필규(2000), 사용자의 묵시적인 정보를 이용한 추천 시스템, 한국정보과학회, 한국정보과학회 학술발표논문집

한국정보과학회 2000년도 봄 학술발표논문집 제27권 제1호(B), pp. 289-291.

강부식(2002), 세계 초일류를 향한 전자정부 및 e-biz 전략 - 자기조직화신경망을 이용한 협력적 필터링 방식의 웹 개인화 방안, 한국경영정보학회 추계학술대회, pp. 409-415.

김현주 & 유민준 & 이인권(2008), 음악 장르 구분을 위한 대표 샘플 추출 방법, 한국정보과학회, 한국정보과학회 학술발표논문집 한국정보과학회 2008 가을 학술발표논문집 제35권 제2호(B), pp. 165-168.

Hanchuan Peng & Fuhui Long & Chris Ding(2005), Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8.

Francois Fleuret(2004), Fast Binary Feature Selection with Conditional Mutual Information, Journal of Machine Learning Research 5 1531-1555.

Yijun Sun and Jian Li(2006), Iterative RELIEF for Feature Weighting, the 23rd International Conference on Machine Learning.

Kohonen, T(1988), An introduction to neural computing, Neural Networks 1, pp. 3-16.

Claypool M & A Gokhale & T. Miranda & P Murnikov & D Netes & M. Sartin(1999), Combining content-based and Collarborative Filters in an online newspaper, ACM SIGIR Workshop on Recommender System, Algorithms and Evaluation. University of California, Berkeley.

Hill, W,L & S,M, Rosenstein & G Furnas(1995), Recommending and Evaluating Choices in A Virtual Community of use, In Proceeding of SIGCHI Conference on Human factors in computing systems, pp. 194-201.

Paul, R & Neohphtos, I & Mitesh, S & Peter, B & John R(1994), Group Lens : an Open architecture for collaborative filtering of netnews, In Proceedings of ACM CSCW '94 Conference on Computer Supported Cooperative Work, pp. 175-186.

Bigus JP & Bigus J(1998), Constructing intelligent agents with Java, New York, Wiley.

신동민 & 이재원 & 이경종 & 이상구(2008), 확장된 태그 기반 협력적 필터링, 한국정보과학회, 한국정보과학회 학술발표논문집 한국정보과학회 2008 종합학술대회 논문집 제35권 제1호(C), pp. 91-94.

J.A. Konstan & B.N. Miller & D. Maltz & J.L. Herlocker & L.R. Gordon & and J. Riedl(1997), GroupLens: Applying Collaborative Filtering to Usenet News, Comm. ACM, vol. 40, no. 3, pp. 77-87.

김종훈 & 김용집 & 정경용 & 임기욱 & 이정현(2007), "분류 속성과 Naive Bayesian을 이용한 사용자와 아이템 기반의 협력적 필터링", 한국콘텐츠학회, 한국콘텐츠학회논문지 제7권 제11호, pp. 23-33.

M. Deshpande & G. Karypis(2004), Item-Based Top-N Recommendation Algorithms, ACM Trans. Information Systems, vol. 22, no. 1, pp. 143-177.

강의영(2007), 하이브리드 필터링 기반 개인화된 추천 프레임워크의

개발 및 적용, 제주대학교 대학원, 석사학위논문.
김정훈(2006), 혼합필터링 기법을 이용한 건강기능식품
상품추천시스템의 개발, 경상대학교 대학원, 석사학위논문.