

A Music Recommendation Method Using Emotional States by Contextual Information

Dong-Joo Kim*, Kwon-Mook Lim**

Abstract

User's selection of music is largely influenced by private tastes as well as emotional states, and it is the unconsciousness projection of user's emotion. Therefore, we think user's emotional states to be music itself. In this paper, we try to grasp user's emotional states from music selected by users at a specific context, and we analyze the correlation between its context and user's emotional state. To get emotional states out of music, the proposed method extracts emotional words as the representative of music from lyrics of user-selected music through morphological analysis, and learns weights of linear classifier for each emotional features of extracted words. Regularities learned by classifier are utilized to calculate predictive weights of virtual music using weights of music chosen by other users in context similar to active user's context. Finally, we propose a method to recommend some pieces of music relative to user's contexts and emotional states. Experimental results shows that the proposed method is more accurate than the traditional collaborative filtering method.

▶ Keyword : Music Recommendation, Emotional State, Contextual Information, Hybrid-Filtering, Winnow Algorithm

1. Introduction

장기간의 개인적인 취향이나 관심사에 따라 선택되는 영화나 책, 뉴스 기사 등과 같은 대상과는 달리 음악은 장기간의 개인적 취향뿐만 아니라 단기적 심리 속성이라 할 수 있는 감정 상태에 따라 선택되는 특성을 지닌다. 즉, 즐거울 때 듣고 싶은 음악은 차분할 때 듣고 싶은 음악과 다를 것이고, 화났거나 상기되었을 때 듣고 싶은 음악은 온화한 상태에서 듣고 싶은 음악과 다를 것이다. 그리고 이러한 감정 상태에 따른 음악의 선택은 개인 취향에 따라 장기적으로 쉽게 변하지 않는다. 따라서 이상적인 음악 추천을 위해서 특정 시점에서의 사용자 감정 상태뿐만 아니라 각 감정 상태에 따른 장기적 개인 취향을 알아내는 일은 중요하다.

감정은 자동적 반응 즉, 자신의 의지와는 무관하게 반응하는 일종의 무의식적 행동에 가까우며, 오랜 기간 동안 사회적 학습

과정에서 친화력을 갖는 가족 및 친족, 친구 등 준거집단과 자신이 속한 하위 문화집단들과의 다양한 관계를 통하여 형성된다. 따라서 감정은 개인차 속성에 따라 다르게 발현되는 속성을 지닌다. 인간의 감정 상태는 다양한 원인이 심리적 상태를 자극하여 변경되거나 발현되는데, 인간의 감정 상태를 자극하는 요인은 크게 내적 요인과 외적 요인으로 구분해 볼 수 있고, 외적 요인은 다시 물리적 요인과 상황 요인으로 구분해 볼 수 있다. 내적 요인은 외적 자극에 관계없이 자신의 내적 요인에 의해 지배되는 감정 상태, 혹은 객관적 상황과는 관계없이 한 인간의 내적 상황에 의해 발현되는 정서적 현상이다. 물리적 요인은 시각, 촉각, 청각, 후각, 미각의 오감에 의한 자극을 의미한다. 상황 요인은 개인을 둘러싸고 개인과 이해관계를 가지며 개인의 생활에 직접적으로 의의를 지니는 현실인 상황에서 자극되는 요인을 뜻한다. 이러한 상황은 인간이 처한 물리적 환경 즉, 시간, 위치, 온도, 날씨, 이동속도 등이 될 수 있다.

음악의 선택에는 내적 자극뿐만 아니라 개인적인 취향을 포

• First Author: Dong-Joo Kim, Corresponding Author: Kwon-Mook Lim

*Dong-Joo Kim (djkim@anyang.ac.kr), College of Liberal Arts, Anyang University

**Kwon-Mook Lim (kmlim@anyang.ac.kr), Dept. of Computer Engineering, Anyang University

• Received: 2015. 09. 02, Revised: 2015. 09. 24, Accepted: 2015. 10. 05.

함하여 날씨, 시간, 이동속도에 이르기까지 다양한 사용자의 상황이 영향을 주게 된다. 즉, 화창한 날에 듣고 싶은 음악과 비가 오거나 흐린 날에 듣고 싶은 음악은 확연히 다를 것이다. 또한 조깅할 때와 산책할 때 듣고 싶은 음악 또한 차이가 존재할 것이며, 낮에 듣고 싶은 음악과 밤에 듣고 싶은 음악에도 차이가 존재할 것이다. 물론 특정 상황에서 특정 개인에게 항상 동일한 감정 상태가 유발되는 것은 아니다. 상황 요인들은 내적 요인과 물리적 요인이 결합될 경우 기존의 감정상태가 새롭게 나타나는 감정 상태를 대치할 수도 있다.

본 논문에서는 특정 상황에서 개별 사용자가 선택한 음악으로부터 개별 사용자가 가진 감정을 자극하는 상황 요인과 개인별 감정 상태들 간의 상관관계를 분석하여 사용자의 감정 상태에 적합한 음악을 추천하는 방법을 제안한다. 제안하는 방법에서는 내적요인과 물리적 요인을 배제하고 일반 개인으로부터 현실적으로 획득 가능한 문맥 상황을 측정하고, 해당 상황에서 선택한 음악 유형들로부터 간접적으로 표현한 감정 상태를 분석함으로써 상황에 따른 개인별 음악 선호 성향을 파악하게 된다.

이를 위해 2장에서 제안하는 방법과 연관된 기반 기술들을 소개한다. 그리고 3장에서는 사용자의 상황 유사도를 계산하기 위한 방법을 제시하고, 사용자가 선택한 음악에 투영된 사용자의 감정 상태를 파악하기 위해 해당 음악의 가사로부터 감정을 대표할 수 있는 감정벡터를 생성하는 방법을 제안한다. 그리고 음악을 추천하기 위해 상황 정보와 감정 상태를 이용한 평가값을 예측하는 방법을 기술한다. 4장에서는 제안하는 방법의 효과를 검증하고, 마지막 5장에서 결론을 맺는다.

II. Related Works

추천시스템이란 잠재적 구매자에게 과거의 구매이력이나 평가이력을 참조하여 관심 있어 할 만한 항목을 추천함으로써 사용자의 만족도를 높이고 기업의 매출을 극대화하기 위한 시스템[1]을 의미하며, 대상 항목에 대해 평가할 것으로 예상되는 값을 예측하는 정보여과시스템(Information filtering system)의 하위 분야라고 할 수 있다[2].

최근 인터넷과 정보기술의 발달과 모바일 기기의 폭발적인 보급으로 생산되는 정보의 양뿐만 아니라 사용자가 접근할 수 있는 정보의 양이 기하급수적으로 증가하였으며, 이로 인하여 사용자는 정보의 선택 문제에 직면하게 되었다. 이러한 정보의 선택 문제는 유사하지만 근본적으로 다른 정보검색 문제와 비교할 수 있는데, 정보검색은 사용자의 정보요구에 부합하는 정보 자원을 제공하는 행위를 의미하며 주로 사건, 인물 등과 같은 지식정보를 요구하는 반면 정보의 선택 문제는 정답을 특정하기 어렵거나 다양한 정답이 존재하는 경우 사용자 개인의 취향에 따라 정보를 선별적으로 제공하는 시스템을 의미한다. 따라서 추천시스템은 개인의 선호도에 따라 그 결과가 분명하게 차이가 나는 음악, 영화와 같은 엔터테인먼트 산업에서 항목의

추천과 선택을 위해 활발히 사용되고 있다.

추천시스템에 관한 연구는 인터넷을 통한 온라인 거래가 활성화되기 시작한 21세기 들어 본격적으로 시작되었으며, 2006년 추천시스템의 연구를 촉발시킨 계기가 된 Netflix Prize[3] 이후 블로그, 도서, 문서, 음악, 영화, 뉴스, 여행 등 다양한 분야[1]에서 폭넓게 연구되고 있다. 추천시스템은 시스템에서 사용되는 핵심 정보원에 따라 크게 협업 필터링(Collaborative filtering)[5, 6], 내용기반 필터링(Content-based filtering)[7-9], 인구통계학적 필터링(Demographic filtering)[10], 문맥기반 필터링(Context-based filtering)[11], 두 가지 이상의 방식을 혼합한 하이브리드(Hybrid) 방식[4]으로 나눌 수 있다. 이들 방법 중 가장 성공적이었고, 현재까지도 다양한 추천 시스템에서 기반 알고리즘으로서 활발히 적용되고 있는 협업필터링 방법은 사용자들의 대상항목에 대한 평가 정보들 간의 상관관계에 기초로 한 방법이다. 즉, 이 방법은 식 (1)과 같이 추천 대상 사용자(Active user)에 대해 이 사용자가 아직 평가하지 않은 대상 항목의 평가값을 예측하기 위해 유사한 평가 성향을 지니는 다른 사용자들의 평가값을 가중치 평균으로 계산한다.

여기서 대상 사용자 a 란 추천 받기를 기대하고 있는 사용자를 의미하며, 대상 사용자 a 가 이전에 평가하지 않았던 항목들 중 다른 사용자 u 가 이전에 평가한 항목 i 을 대상으로 예측 평가값 $\hat{r}_{a,i}$ 을 계산하게 되며, 예측 평가값 $\hat{r}_{a,i}$ 이 높은 순으로 추천이 이루어진다. 식 (1)에서 $W_{a,u}$ 는 대상 사용자 a 와 항목 i 를 평가한 다른 사용자 u 와의 유사도, 혹은 상관계수로 다른 사용자의 평가값을 어느 정도 반영할지를 결정하는 가중치로 사용된다. 또한 다른 사용자에 의해 평가된 항목들을 선택하는데 있어, 가장 유사한 사용자 몇 명만이 평가한 항목들로만 예측 평가값을 계산하기도 한다. 한편으로, 사용되는 유사도 계산 방법에 따라 반대 성향을 지닌 사용자가 평가한 항목의 평가값이 사용하기도 한다. 또한 사용자는 개개인에 성향에 따라 항목에 대한 평가값이 전반적으로 후한 경우도 있고 그 반대의 경우도 있어 식 (1)에서 이를 보정하기 위해 \bar{r}_a 나 \bar{r}_u 의 항을 빼거나 더한다.

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

$$W_{a,u} = \frac{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)^2 (r_{u,i} - \bar{r}_u)^2}} \quad (2)$$

식 (1)에서 다른 사용자의 평가값에 대한 가중치로 사용된 $W_{a,u}$ 는 일반적으로 코사인 유사도(Cosine similarity), 피어슨 상관계수(Pearson's correlation coefficient), 조건부 확률을 이용하여 계산될 수 있다. 본 논문에서 비교 대상으로 삼은 협업필터링 방법에서 사용하였고, 제안하는 방법에서 협업적 특

성을 반영하기 위한 유사도 계산을 위해 사용된 가중치 $W_{a,u}$ 는 식 (2)와 같은 피어슨 상관계수로 계산한다.

이와 같이 식 (1), (2)의 계산을 통한 방법은 두 가지 협업 필터링 방법 중 평가 성향이 유사하거나 정반대인 다른 사용자들의 평가값을 이용한 사용자 기반 협업 필터링 방법이다. 그러나 이 방법에서는 사용자의 평가값이 실시간으로 항상 변하기 때문에 평가 예측값을 계산하기 위해 사용자들의 평가값을 찾아 매번 계산해야 하므로 시간이 걸린다는 단점이 있다. 이에 대한 대안으로 사용자 기반 협업 필터링 방법과 같이 사용자들의 평가값을 기준으로 유사한 사용자를 계산하는 것이 아니라, 항목의 평가값을 기준으로 유사한 항목을 찾아 계산하는 항목 기반 협업 필터링 방법[12]이 제시되었다. 이 방법은 전반적으로 자주 변하지 않은 항목의 평가값을 대상으로 하기 때문에 항목들 간의 유사도가 사전에 계산될 수 있어 학습시에는 시간이 오래 걸리지만 예측 평가값을 계산할 때는 매우 빠르게 계산할 수 있다.

내용기반 필터링 방법은 항목을 대표할 수 있는 메타데이터와 사용자 이전 평가 정보를 활용하여 사용자가 좋아할 만한 항목을 추천하는 방법이다. 이때 메타데이터는 항목과 연관된 또 다른 정보로부터 자동으로 추출될 수도 있고, 전문가, 혹은 비전문가에 의해 수동으로 직접 만들어질 수도 있다. 메타데이터의 자동추출을 위한 방법에는 항목에 대한 짧은 설명문으로부터 언어 분석 도구를 사용하여 대표성 있는 단어를 추출하는 방법[10]에서부터 음악, 영화 등과 같은 멀티미디어 항목에 대한 주파수 스펙트럼으로부터 MFCC(Mel-frequency cepstral coefficient)[9]와 같은 저수준의 기술자(Descriptor)를 추출하거나 STFT(Short-Time Fourier Transform)을 통하여 음의 높낮이와 박자를 인식하여 장르를 알아내는 방법[14]도 연구되었다. 전문가 집단에 의해 수동으로 직접 생성하는 경우는 일반적으로 전통적인 분류 방식인 텍소노미(Taxonomy)에 따라 만들어진다. 이 방법은 추천 결과에 대한 품질과 정확도가 매우 높지만 많은 비용과 노력이 든다는 문제점이 존재한다. 반면 비전문가들이 일반 사용자가 자발적으로 협력하여 만들어내는 폭소노미(Folksonomy)의 한 예라고 할 수 있는 소셜 태그(Social tag)[8, 13]는 단순한 개인의 차원이 아니라 사회적, 집단적 차원의 지식을 공유하고자 하는 이상을 실현하는 수단으로서 전문가 집단에 의한 수동 작성 방법의 단점을 보완한 추천 시스템[8]의 구축이 가능하다.

협업 필터링 방법은 아직 평가된 적이 없는 새로 추가된 항목에 대해서는 추천이 불가능한 초기평가자문제(Early-rater problem)와 평가 이력이 없어 성향 분석이 불가능한 사용자에게 대해서는 추천 할 수 없는 초기시작문제(Cold-start problem)로 많은 비판을 받았다. 초기평가자문제는 평가되지 않았더라도 항목을 설명하는 메타데이터를 추출할 수 있다면 내용기반 필터링 방법으로 어느 완화할 수 있다. 초기시작문제를 해결하기 위해 주로 시도하고 있는 대표적인 방법 중 하나가 인구통계학적 필터링 방법이다. 이 방법은 인구통계학적 정보, 즉 성

별, 나이, 사는 지역, 학력, 사회적 지위 등과 같은 정보와 평가 정보를 이용하여 선호하는 항목에 대한 인구통계학적 성향을 파악하여[10], 이를 추천에 활용하는 방법이다. 이 방법은 사용자에게 개인 정보를 요구해야 하는 부담과 더불어 사생활 보호 문제가 존재한다.

오늘날 추천시스템 분야에서의 연구는 각 방법들의 단점을 보완하고 장점을 극대화하기 위해 두 가지 이상의 방법들을 혼합한 하이브리드 방법[4, 15, 16]들이 주류를 이루고 있으며, 최근 사회관계망(SNS)을 활용한 시스템[16, 17]들이 제안되고 있다. 본 논문에서 시도하고 있는 방법 역시 하이브리드 방법으로 분류할 수 있으며, 이 방법은 이동형 기기에서 인식 가능한 상황정보(Contextual information)로부터 사상(Mapping)되는 감정 정보를 추출하고, 상황과 감정 간의 관계를 학습하여 다른 사람들의 상황에 따른 감정 정보로부터 음악을 추천하는 방법이다.

III. The Proposed Method

서론에서도 기술했듯이 청취를 위한 음악의 선택은 영화, 뉴스, 책과 같은 다른 유형의 항목들과는 달리 개인적인 취향뿐만 아니라 청취자, 혹은 사용자의 감정 상태에 큰 영향을 받는다. 물론 사용자가 음악을 선택하는데 있어 음악은 감정 상태를 유지하기 위한 수단이 되기도 하고, 벗어나기 위한 수단이 되기도 하며, 역으로 감정 상태를 유발하기 위한 수단이 되기도 한다. 즉, 감정 상태를 유지하고 그 느낌을 즐기기 위해서는 현재의 감정 상태와 일치하는 감정 정보가 내포된 음악을 선택하겠지만, 그 상태에서 벗어나고자 하는 경우에는 사용자의 감정 상태와는 다른 감정 정보를 내포한 음악을 선택할 수도 있다. 이러한 경향 또한 사용자의 개인적인 성향이라 할 수 있고, 감정에 따른 음악의 선택은 단기 속성이지만 감정에 따른 음악 선택의 전반적인 성향은 개인의 장기적 성향이라 볼 수 있다.

한편 인간의 감정은 시간, 위치, 온도, 날씨, 이동 속도와 같은 외적 상황 요인으로 자극될 수 있다. 물론 내적 요인이나 외적 요인 중 물리적 자극에 의한 요인이 외적 상황 요인보다 더 즉각적인 변화를 유발하는 강한 자극이 있겠지만, 강한 자극으로부터 유발되는 감정 상태에서 음악을 듣는 일이 드물다는 것을 미뤄볼 때, 실제 음악을 선택하는데 있어 주도적인 역할을 수행하는 것은 외적 상황 요인이라 볼 수 있다. 즉, 사용자가 처한 특정 상황은 인간의 감정을 자극하게 되고 자극된 감정은 음악을 선택하는데 영향을 주게 된다.

따라서 상황에 따른 음악의 장기적인 개인 선택 성향을 파악하여 패턴화한다면 사용자의 감정 상태에 따른 음악을 추천이 가능할 것이다. 이를 위해 본 논문에서는 사용자의 과거 청취 이력으로부터 특정 상황과 그 상황에서 선택한 음악의 감정 정보로부터 추출한 규칙성을 바탕으로 이전에 청취한 이력이 없는 음악을 추천하는 방법, 즉 상황에 따른 감정 상태를 이용한 추천 방법을 제안한다. 제안하는 방법은 사용자가 처한 상황에

서 사용자가 좋아할 만한, 즉 선택 가능성이 높은 음악을 추천하기 위해, 사용자들의 과거 상황에 따른 청취이력으로부터 상황에 따른 감정 상태를 구분한다. 이때 감정 상태는 과거에 특정 상황에서 사용자가 청취를 위해 선택했던 음악의 가사로부터 추출된다. 이러한 과거 청취 이력으로부터 대상 사용자가 선택했던 적이 없는 음악을 대상 사용자에게 추천하기 위해 먼저 대상 사용자의 상황과 유사한 다른 사용자의 과거 상황을 찾는다. 유사한 상황의 다른 사용자가 처했던 상황에서 선택했던 음악에 대한 감정 상태(감정 벡터)에서 단어별 중요도에 대해 가중치 평균을 계산한다. 이때 가중치는 상황 유사도이다.

이어지는 절에서는 상황 유사도를 계산하는 방법과 사용자들이 청취를 위해 과거에 선택한 음악으로부터 감정 상태, 즉 감정 벡터를 추출하는 방법을 제시한다. 그런 다음 상황 정보와 음악으로부터 추출되는 감정 상태를 이용한 추천 방법을 제시한다.

1. Contextual Similarity

상황(Situation), 혹은 문맥(Context)이란 특정 사용자와 또 다른 사용자, 사용자와 시스템간의 상호작용에 영향을 미치는 장소, 시간, 날씨, 온도 등을 의미한다. 이들 상황을 이용한 다양한 분야에서의 연구는 사용자의 상황이 빠르게 변하는 이동형 컴퓨팅 환경에서 상호작용을 연구하는데 있어 무엇보다 중요하다[18]. 음악 추천 분야에서도 이러한 상황은 앞서 설명한 대로 사용자의 감정 상태 변화를 유발하며 음악의 선택에 있어 중요한 역할을 수행하게 된다.

본 논문에서 사용자들의 상황 간의 유사성 정도를 계산하기 위해 이용한 상황 정보는 온도(H), 날씨(W), 시간(T), 이동 속도(S)이다. 대상 사용자 a 의 이동형 기기로부터 인식된 현재 상황과 과거의 다른 사용자 u 의 상황 사이의 유사도는 다음 식 (3)의 $D_{a,u}$ 와 같이 계산된다.

$$D_{a,u} = \alpha H_{a,u} + \beta W_{a,u} + \gamma T_{a,u} + \delta S_{a,u} \quad (3)$$

식 (3)에서 α , β , γ , δ 은 네 가지의 상황 정보 각각에 대한 중요도로 합은 1이며, 이 값들은 실험을 통해 경험적으로 결정된다. 또한 $H_{a,u}$, $W_{a,u}$, $T_{a,u}$ 는 상황 각각에 대한 유사도로 각각 (4), (5), (6)과 같이 평가되며, H_a , W_a , T_a 는 대상 사용자의 상황에 따른 실제 측정값이며, H_u , W_u , T_u 는 비교 대상인 다른 사용자의 측정값이다. 이동 속도에 따른 유사도를 평가하기 위한 식 (7)에서 S_a 와 S_u 는 실제 측정값이 아니라 이동 속도의 유형이다. 이동 속도는 크게 네 가지의 유형으로 구분했으며, km/h 단위로 측정된 속도를 x 라고 했을 때, 네 가지 유형은 멈춰 있는 상태($x=0$), 걷는 상태($0 < x \leq 15$), 뛰는 상태($15 < x \leq 30$), 차량으로 이동 중인 상태($30 < x$)이다.

식 (4)에서와 같이 대상 사용자 온도와 과거 다른 사용자와의 차이가 7도를 넘어가면 유사하지 않은 것으로 간주하였으며, 식 (6)에서와 같이 대상 사용자 시간과 과거 다른 사용자와의 차이가 180분을 넘으면 서로 다른 시간으로 간주하였다. 식

(5)와 (7)에서는 동일 값이거나 동일 유형이면 유사한 것으로 계산한다.

$$H_{a,u} = \begin{cases} \frac{7 - |H_a - H_u|}{7} & \text{if } |H_a - H_u| \leq 7 \\ 0 & \text{if } |H_a - H_u| > 7 \end{cases} \quad (4)$$

$$W_{a,u} = \begin{cases} 1 & \text{if } W_a = W_u \\ 0 & \text{if } W_a \neq W_u \end{cases} \quad (5)$$

$$T_{a,u} = \begin{cases} \frac{180 - |T_a - T_u|}{180} & \text{if } |T_a - T_u| \leq 180 \\ 0 & \text{if } |T_a - T_u| > 180 \end{cases} \quad (6)$$

$$S_{a,u} = \begin{cases} 1 & \text{if } S_a = S_u \\ 0 & \text{if } S_a \neq S_u \end{cases} \quad (7)$$

이렇게 계산된 대상 사용자와 다른 사용자의 상황들 간의 상황 유사도는 다른 사용자가 각 상황에서 선택한 음악으로부터 추출된 감정단어의 중요도에 대한 상황 유사도에 따른 가중치 평균을 계산하는데 있어 가중치로 사용된다.

2. Learning Classifier

특정 상황에서 무의식적으로 음악에 투영된 사용자의 감정 상태는 해당 상황에서 선택된 음악에 내포되어 있는 감정 정보로 대신할 수 있다. 이 절에서는 과거 사용자가 특정 상황에서 선택한 음악의 가사로부터 감정을 대표할 수 있는 감정벡터를 생성하는 방법을 제시한다.

음성, 표정, 행동, 글 등으로 표현된 감정을 인식하는 연구는 1990년대 중반 인간 컴퓨터 상호작용(Human computer interaction) 분야에서 감정의 중요성을 강조한 감성 컴퓨팅(Affective computing)[19, 20]이 소개되면서부터 시작되었다. 이들 연구 중 자연언어 텍스트에 담긴 감정을 인식하는 기술은 TTS (Text-to-speech) 시스템에서 설득력 있는 연설이나 자연스럽게 친밀한 대화를 위해 요구되는 기술이며, 퍼지의 미분류(Fuzzy semantic typing)에 기초한 연구[21]를 시작으로 2000년대 들어 본격적으로 관심을 받기 시작하였다. 텍스트의 감정 인식에 관한 연구는 키워드에 기반한 연구를 포함하여 규칙기반, 분류 및 학습 기반 연구 등의 다양한 연구가 이루어져 왔으며, 많은 연구의 공통점은 기본적으로 감정 사전(Affect lexicon)[21]의 구축을 필수로 하고 있다. 즉, 감정 단어와 이 단어에 대한 품사, 의미 정보와 같은 일반적인 사전 정보 외에 감정의 범주(Category), 감정의 강도(Intensity) 정보를 가지며, 부가적으로 긍정, 부정, 혹은 중립과 같은 극성(Polarity) 정보를 갖기도 한다. 그러나 이러한 감정 사전의 구축에는 많은 시간과 비용이 들 뿐만 아니라, 감정 사전의 구축과 확장에 관한 문제[22]는 독립적인 연구 주제로 한 영역을 차지하고 있을 정도로 아직 활발히 연구되고 있는 주제이다. 본 논문에서는 가사로부터 음악에 담겨 있는 감정 정보를 추출하기 위해 감정 사전의 어휘 정보를 사용하지 않고 일반적인 사전에서 얻을 수 있는 품사정보만을 사용한다.

감정 단어는 객관적인 사실을 표현할 때 사용하기보다 주관적인 견해를 표현할 때 주로 사용된다. 주관적인 견해를 표현할

때 사용하는 단어의 분포는 품사에 따라 현저히 차이가 난다. 즉, 단어의 품사에 따른 주관성 정도를 조사한 결과에 따르면 영어의 경우 부사, 형용사, 동사 순으로 주관성 정도가 높다 [24]. 가요를 대상으로 하고 있는 본 논문에서 한국어의 경우 주관적 표현에 사용되는 술어의 상당수가 동사가 아닌 형용사나 명사에 접미사나 서술격조사가 붙어 형성됨을 상기하였을 때, 본 논문에서는 단어의 품사가 형용사, 부사, 명사인 단어만을 감정 단어로 간주하였다. 구체적으로 품사가 상태성 명사, 성상 형용사, 성상 관형사, 일반 부사인 형태소만을 감정 단어로 간주하였다. 감탄사 중 일부는 적극적인 감정 표현에 사용되기도 하지만 그 수가 매우 제한적일 뿐만 아니라, 대부분의 감정 표현을 위한 감탄사는 긍정 표현이나 부정 표현 양쪽 모두에서 두루 사용되어 분별력을 떨어뜨리는 작용을 한다. 따라서 감정을 표현하는 단어일지라도 감탄사를 제외하였다.

가사로부터 이들 형태소를 추출하기 위해 먼저 기호, 의성/의태어를 제거하는 전처리 과정을 거쳤으며, 그런 다음 형태소 분석을 수행한다. 형태소 분석은 한자성 복합명사를 제외하고 모든 어절을 완전 분석, 즉 한자성 복합명사를 제외하고 복합어를 허용하지 않는 분해할 수 있는 최소 단위인 형태소로 분해하였다. 이로부터 어근에 해당하는 형태소 중 앞서 기술 한 네 가지 형태소만 감성 단어로 추출하였다.

그런 다음 추출된 감성 단어들 중 각 음악을 대표할 만한 단어만을 선별하는 처리를 수행하게 된다. 이 과정에서 이전 단계에서 추출된 감성 단어를 대상으로 TF/IDF 가중치를 계산하여 높은 값을 갖는 단어만을 선택한다. TF/IDF는 정보검색분야에서 문서를 대표할 수 있는 단어의 가중치를 계산하는 수단으로 사용되는 단어빈도(Term frequency)와 역문헌빈도(Inverse document frequency)의 비율이다. 감정 단어들 중 TF/IDF가 높은 순으로 상위 일정 수의 단어들을 각 음악을 대표할 수 있는 단어로 선정하게 된다.

선정된 단어들은 특정 사용자가 특정 상황에서 선호하는 음악인지 선호하지 않는 음악인지를 판단하기 위해 분류를 위한 학습 자질(Feature)로 사용한다. 또한 이 자질을 사용한 선형 분류기(Linear classifier)의 가중치를 학습하기 위한 방법으로 Winnow 알고리즘[24]을 사용하였다. Winnow 알고리즘은 많은 수의 자질을 가지면서 각 자질의 통계량이 적은 대상을 빠르게 학습할 수 있는 알고리즘으로, 본 논문에서 대상으로 하는 가사의 특성에 적합하다 할 수 있다. 즉, 이전 단계에서 선정된 각 음악의 가사 X 를 대표하는 n 개의 대표 단어들을 자질 $X = \{x_1, x_2, \dots, x_n\}$ 이라고 하고, 각 단어에 대한 자질의 가중치를 $W = \{w_1, w_2, \dots, w_n\}$ 이라고 했을 때, 학습기는 식 (8)을 만족하도록 배수규칙(Multiplicative rule)을 통해 가중치를 갱신하게 된다.

$$\Omega(X) = \sum_{i=1}^n w_i x_i > \Theta \quad (8)$$

여기서 Θ 는 임계치이고, x_i 는 가사에서 해당 자질들의 존재 유무로, 존재하는 경우 1이고 존재하지 않으면 0의 값을 갖는

블리언 자질이다. 학습 과정에서 먼저 w_i 는 1로 초기화되고, 특정 사용자가 선호하는 음악과 선호하지 않는 음악의 가사에서 대표 단어에 대한 w_i 를 갱신하기 위한 배수규칙은 다음과 같다.

- 가) 가사 X 에 존재하는 단어 자질의 가중치 합 $\Omega(X)$ 를 식 (8)로 계산
- 나) 선호 음악의 가사 X 에 대한 $\Omega(X)$ 가 Θ 이상이면, 가중치 w_i 를 갱신하지 않음
- 다) 선호하지 않는 음악의 가사 X 에 대한 $\Omega(X)$ 가 Θ 미만이면, 가중치 w_i 를 갱신하지 않음
- 라) 선호 음악의 가사 X 에 대한 $\Omega(X)$ 가 Θ 미만이면, 가사에 존재하는 단어에 대한 w_i 에 대해 $w_i = w_i \times 2$ 로 갱신
- 마) 선호하지 않는 음악의 가사 X 에 대한 $\Omega(X)$ 가 Θ 이상이면, 가사에 존재하는 단어에 대한 w_i 에 대해 $w_i = w_i / 2$ 로 갱신

특정 사용자의 과거 청취 이력으로부터 특정 상황에서 모든 선호 및 비선호 음악의 가사 X 에 대해 갱신은 가중치 값의 변화가 없을 때까지, 혹은 정확도에 변화가 없이 10번 이상 반복되기 전까지 배수규칙 가)~마)를 반복한다.

3. Recommendation

과거 청취이력으로부터 각 사용자에게 따라 상황과 선호 여부에 따른 모든 가중치 집합이 계산되면 이를 이용하여 대상 사용자 a 에게 특정 상황에서의 과거에 청취 이력이 없던 새로운 곡을 추천하게 된다. 이를 위해 먼저 식 (3)을 이용하여 대상 사용자의 현재 상황 정보와 과거 청취 이력에서 상황 정보의 유사도를 계산한다. 현재 상황 c 에 따른 대상 사용자 a_c 의 현재 상황 c 와 과거의 다른 모든 상황 k 에 대해 계산된 유사도 $D_{a_c, u}$ 는 식 (8)과 배수규칙으로 학습된 다른 사용자 u 의 과거 특정 상황 k 에서 특정 곡에 대한 W 의 가중치의 평균을 식 (9)와 같이 계산한다. 식 (9)의 가중치는 특정 곡에 대한 자질의 가중치 W 가 아니라 유사도 $D_{a_c, u}$ 를 의미한다.

$$\hat{W}_{a_c, u} = \frac{\sum_{u=1}^n \sum_{k=1}^m W_{u, k, i} \times D_{a_c, u}}{\sum_{u=1}^n \sum_{k=1}^m D_{a_c, u}} \quad (9)$$

식 (9)에서 $\hat{W}_{a_c, u}$ 는 추천하려고 하는 미지의 곡에 대한 추정 가중치 집합으로 현재 사용자의 감정 상태를 다른 사용자들의 과거 청취이력을 바탕으로 추정한 것이다. 다시 말해 이 추정치는 현재의 대상 사용자가 처해 있는 상황에 가장 이상적인 미지의 음악에 대한 가사의 대표 감정 단어의 가중치 집합이라고 할 수 있다. 따라서 실제 추천되는 음악은 현재의 대상 사용자가 과거에 선호 여부를 표시하지 않았던 음악들 중 추정 가중치 집합 $\hat{W}_{a_c, u}$ 와 가장 유사한 가중치 집합을 갖는 음악이 된다.

유사한 가중치 집합을 갖는 음악은 식 (10)의 코사인 유사성 척도를 이용하여 계산하고, 가장 큰 값으로 계산되는 음악들을 사용자에게 순서대로 추천한다.

$$SIM(\hat{W}_{a,u}, W_i) = \frac{\sum_{j=1}^n \hat{w}_j \times w_j}{\sqrt{\sum_{j=1}^n (\hat{w}_j)^2} \times \sqrt{\sum_{j=1}^n (w_j)^2}} \quad (10)$$

IV. Experiments and Evaluations

실험을 위해 구축한 데이터는 19~25세 사이의 58명의 대학생들로부터 수집된 청취이력이다. 청취 대상이 되는 음악은 2006년 모 방송국에서 13세 이상 남녀 1,500명을 대상으로 조사한 ‘한국인이 좋아하는 가요 100곡’을 선정하였으며, 인터넷 스트리밍을 통하여 각 곡의 앞 1분만 청취 가능하게 했다. 실험 대상자들에게 모든 곡에 대해 다양한 상황에서 최소 한 번 이상 각 곡의 선호도를 평가점수로 표시하도록 요구했다. 평가점수는 1~5까지 표시하는 것이 일반적이지만 이 실험에서는 선호(1), 비선호(-1)로만 평가하도록 했으며, 각각의 음악에 대해 본인의 평가 여부는 피실험자가 조회할 수 있지만 상황에 따른 평가 여부와 평가값은 조회할 수 없도록 했다.

전체 100곡에 대해 일정기간 동안 다양한 상황에서 선호도 평가를 실시한 결과 피실험자 58명 중 12명이 모든 곡을 평가하였으며, 1인당 평균 43곡을 평가하였다. 피실험자가 평가한 곡들 중 2번 이상의 상황에서 평가된 곡의 비율은 평균 89.7%였으며, 10번 이상의 상황에서 평가된 곡의 비율은 평균 23.5%였고, 전체 평가 횟수는 12,478회로 곡당 평균 124.7회였다. 본 논문에서 청취이력으로 기록되는 상황은 온도, 날씨, 시간, 이동속도의 네 가지에 관한 상황이며, 58명의 전체 학생 모두 시간적으로 연속되지 않는 8가지 이상의 상황에서 청취이력을 남겼다.

정확률의 평가를 위해 피실험자 별로 평가한 이력이 있는 음악 중 임의로 60%를 선택하여 학습 데이터로 사용하였으며, 나머지 40%의 곡은 평가 데이터로 사용하였다. 학습 데이터로 선택된 곡은 사용자의 상황 정보, 상황에 따른 평가정보를 모두 남겨놓은 반면, 평가 데이터로 선택된 곡에 대해서는 상황에 따른 평가정보는 제거하고 상황정보만 남겨놓았다. 학습 데이터는 상황에 따른 음악의 사용자 평가정보에 대해 각 음악의 대표 단어(감정 단어)를 추출하고 이 단어에 대해 가중치를 학습하였다.

정확률의 평가는 임의로 선택된 나머지 40%의 평가 음악에 대한 상황정보로부터 식 (9)를 통하여 추정가중치 집합을 계산하였으며, $D_{a,u}$ 의 계산을 위한 네 가지 상황별 유사도에 대한 중요도는 각각 $\alpha = 0.15$, $\beta = 0.3$, $\gamma = 0.25$, $\delta = 0.3$ 로 설정하였다. 식 (10)을 통하여 추정가중치와 유사한 학습데이터로부터 계산된 음악에 대한 가중치와의 유사도를 계산하였다. 이들 중 추정가중치와 가장 유사도가 높은 상위 3개의 가중치를 갖

는 음악을 선택하여 추천하였다. 추천한 3개의 음악 중 2개 이상의 음악이 구축된 평가 이력에서 해당 사용자가 선호한다고 평가를 했다면 정확하게 판단한 것으로 간주하였다.

Table 1. Experimental Results

	collaborative filtering	proposed method
correctness	61.2%	63.7%

이에 대한 분할 실험을 모든 피실험자 58명에 대해 10번씩 반복하여 총 580회를 실시하였고, 비교 실험을 위해 전통적인 협업필터링방법을 통하여 동일 실험을 반복 실시하였다. Table 1에서 보는 바와 같이 제안하는 방법의 정확률은 63.7%로 측정된 반면, 협업필터링 방법의 정확률 61.2%보다 높은 것으로 측정되었다. 협업필터링 방법으로 측정한 정확률도 제안하는 방법과 불과 2.5% 차이로 비교적 우수한 성능을 보였다. 협업필터링 방법의 특성상 모든 피실험자의 선호도 평가값이 대부분의 항목이 다양한 상황에 대해 충분히 평가하도록 요구했기 때문일 뿐이고, 실험 환경과는 달리 모든 사용자가 다양한 상황에서 많은 음악에 대해 평가하지는 않기 때문에 실제 환경에서 협업필터링 방법은 현저히 낮은 정확률을 보일 것으로 예상된다.

IV. Conclusions

음악의 선택은 사용자의 상황에 따른 감정의 무의식적 투영이라 할 수 있다. 이러한 특징을 반영하기 위해 본 논문에서는 특정 상황에서 사용자가 선택한 음악의 가사를 형태소 분석하여 감정이 담겨 있을 가능성이 높은 대표 감정 단어를 추출하였다. 추출된 단어가 가지는 기여도, 즉 사용자가 처한 감정 상태를 가장 잘 반영하고 있는 정도를 계산하기 위해 선형 분류기의 학습 방법 중의 하나인 Winnow 알고리즘을 사용하였다. 그런 다음 추천 대상 사용자가 처한 상황과 과거 다른 사용자가 처한 상황의 유사도를 계산하여 상황 유사도에 따른 감정 단어별 가중치 합을 계산하였다. 이렇게 계산된 감정 단어별 가중치들은 미지의 음악에 대한 추정 가중치로 볼 수 있다. 추정 가중치의 값은 실제 음악의 가중치와의 유사도 계산을 통하여 추천 음악을 선택할 수 있었다.

실험을 통해 기존의 전통적인 협업필터링 방법에 비해 우수한 성능을 보였다. 향후 연구과제로 선호도 평가가 없거나 매우 부족한 사용자에 대한, 그리고 선호도 평가가 충분히 이루어지지 않은 항목에 대한 협업필터링 방법의 초기평가자문제와 초기시작문제를 어떻게 극복되었는지에 관한 면밀한 분석이 필요할 것이다. 또한 논문에서 대상으로 삼고 있는 네 가지 상황, 즉 온도, 날씨, 시간, 이동 속도에 대한 기여도 분석을 통하여 상황에 따른 가중치를 달리하거나 불필요한 상황 요소를 배제하는 시도가 필요할 것이다.

REFERENCE

- [1] Li-Hua Li, Rong-Wang Hsu, and Fu-Min Lee, "Review of Recommender Systems and Their Application," *International Journal of Advanced Information Technologies*, Vol. 6, No. 1, pp. 63-87, June 2012.
- [2] K. Satya Reddy, "Improving an aggregate recommendation diversity Using ranking-based tactics," *International Journal of Computer Trends and Technology*, Vol. 4, Issue 9, pp. 3178-3183, Sep. 2013.
- [3] Robert M. Bell, Yehuda Koren, and Chris Volinsky, "All Together Now: A Perspective on the Netflix Prize," *Chance*, Vol. 23, No. 1, pp. 24-29, April 2010.
- [4] Robin Burke, "Hybrid Web Recommender Systems," *The Adaptive Web*, Vol. 4321, pp. 377-408, May 2007.
- [5] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry, "Using collaborative filtering to weave an information tapestry," *Communications of ACM*, Vol. 35, No. 12, pp. 61-70, Dec. 1992.
- [6] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pp. 175-186, Oct. 1994.
- [7] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro, "Content-based Recommender Systems: State of the Art and Trends," *Recommender Systems Handbook*, pp. 73-105, Jan. 2011.
- [8] Ivan Cantador, Alejandro Bellogin, and David Vallet, "Content-based Recommendation in Social Tagging Systems," In *Proceedings of the 4th ACM Conference on Recommender Systems*, pp. 237-240, Sep. 2010.
- [9] Jonathan T. Foote, "Content-Based Retrieval of Music and Audio," In *Proceedings of SPIE Multimedia Storage Archiving Systems II*, vol. 3229, pp. 138-147, Oct. 1997.
- [10] Bruce Krulwich, "Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data," *Artificial Intelligence Magazine*, vol. 18, no. 2, pp. 37-45, July 1997.
- [11] Gediminas Adomavicius and Alexander Tuzhilin, "Context-Aware Recommender Systems," *Recommender Systems Handbook*, pp. 217-253, Oct. 2011.
- [12] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Item-based Collaborative Filtering Recommendation Algorithm," In *Proceedings of the 10th International Conference on World Wide Web*, pp. 285-295, May 2001.
- [13] Paul Lamere, "Social Tagging and Music Information Retrieval," *Journal of New Music Research*, Vol. 37, No. 2, pp. 101-114, June 2008.
- [14] George Tzanetakis and Perry Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 5, pp. 293-302, July 2002.
- [15] Robin Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, Vol. 12, No. 4, pp. 331-370, Nov. 2002.
- [16] Kyoung-Jae Kim and Hyun-chul Ahn, "Hybrid Recommender Systems using Social Network Analysis," In *Proceedings of World Academy of Science, Engineering and Technology*, pp. 879-882, Oct. 2012.
- [17] Faustino Sanchez, Marta Barrilero, Silvia Uribe, Federico Alvarez, Agustin Tena, and Jose Manuel Menendez, "Social and Content Hybrid Image Recommender System for Mobile Social Networks," *Mobile Networks and Applications*, Vol. 17, Issue 6, pp. 782-795, Dec. 2012.
- [18] Anind K. Dey and Gregory D. Abowd, "Towards a Better Understanding of Context and Context-Awareness," In *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, pp. 304-307, June 1999.
- [19] Rosalind W. Picard, "Affective Computing," MIT Media Laboratory Perceptual Computing Section Technical Report, No. 321, Nov. 1995.
- [20] Rosalind W. Picard, *Affective Computing*, The MIT Press, Oct. 1997.
- [21] Pero Subasic and Alison Huettner, "Affect Analysis of Text Using Fuzzy Semantic Typing," *IEEE Transactions on Fuzzy Systems*, Vol. 9, Issue 4, Aug. 2001.
- [22] Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi

Chen, "Building Emotion Lexicon from Weblog Corpora," In Proceedings of the 45th Annual Meeting of the ACL, pp. 133-136, June 2007.

[23] Andrea Esuli and Fabrizio Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," In Proceedings of the 5th Conference on Language Resources and Evaluation, pp. 417-422, May 2006.

[24] Nick Littlestone, "Learning Quickly When Irrelevant Attributes Abound: A New Linear-threshold Algorithm," Machine Learning, Vol. 2, Issue 4, pp. 285-318, April 1988.

Authors



Dong-Joo Kim received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Hanyang University, Korea, in 1996, 1998 and 2007, respectively.

Dr. Kim joined the faculty of the College of Liberal Arts at Anyang University, Korea, in 2008. He is currently a Professor in the College of Liberal Arts, Anyang University. He is interested in natural language processing, opinion mining, critiquing system, machine learning, and information retrieval.



Kwon-Mook Lim received the B.E. degree in Electronic Engineering in 1980 from Kyunghee University, Korea, the M.S. degree in Computer Science in 1987 from Western Illinois University,

USA, and Ph.D. degree in Computer Science in 1996 from Yonsei University, Korea.

Dr. Lim joined the faculty of the Department of Computer Engineering at Anyang University, Korea, in 1991. He is currently a Professor in the Department of Computer Engineering, Anyang University. He is interested in morphological analysis, big data, and machine learning.