

# Personalized Movie Recommendation System Using Context-Aware Collaborative Filtering Technique

Min Jeong Kim<sup>†</sup> · Doo-Soon Park<sup>††</sup> · Min Hong<sup>††</sup> · HwaMin Lee<sup>†††</sup>

## ABSTRACT

The explosive growth of information has been difficult for users to get an appropriate information in time. The various ways of new services to solve problems has been provided. As customized service is being magnified, the personalized recommendation system has been important issue. Collaborative filtering system in the recommendation system is widely used, and it is the most successful process in the recommendation system. As the recommendation is based on customers' profile, there can be sparsity and cold-start problems. In this paper, we propose personalized movie recommendation system using collaborative filtering techniques and context-based techniques. The context-based technique is the recommendation method that considers user's environment in term of time, emotion and location, and it can reflect user's preferences depending on the various environments. In order to utilize the context-based technique, this paper uses the human emotion, and uses movie reviews which are effective way to identify subjective individual information. In this paper, this proposed method shows outperforming existing collaborative filtering methods.

**Keywords :** Context-based Technique, Collaborative Filtering, Movie Recommendation, Movie Review

## 상황기반과 협업 필터링 기법을 이용한 개인화 영화 추천 시스템

김민정<sup>†</sup> · 박두순<sup>††</sup> · 홍민<sup>††</sup> · 이화민<sup>†††</sup>

## 요 약

정보의 폭발적인 증가로 사용자들은 원하는 정보를 빠른 시간에 얻는 것이 힘들어졌다. 따라서 이 문제를 해결하기 위한 다양한 방식의 새로운 서비스들이 제공되고 있다. 개인에게 맞는 맞춤 서비스를 제공하는 것이 중요하게 부각되면서 개인화 추천 시스템이 매우 중요하게 되었다. 추천 시스템 중 협업 필터링은 추천 시스템에서 널리 사용되고 있고 개인화 추천 시스템 중에서 가장 성공적인 방법이다. 협업 필터링 방법은 고객들의 프로파일 정보를 기반으로 추천을 하므로 희박성 문제와 cold-start 문제가 있다. 본 논문에서는 개인에게 더 정확하게 추천하기 위해 협업 필터링 기법과 상황기반 기법을 함께 이용하는 방법을 제안한다. 상황기반 기법은 사용자를 둘러싼 시간, 감정, 장소 등과 같은 환경을 고려하여 사용자에게 맞는 아이템을 추천하는 방법으로 상황에 따라 달라지는 사용자의 선호도를 반영할 수 있다. 본 논문에서는 상황기반 기법을 활용하기 위해 상황정보로 감정을 이용하며 이를 위해 개인의 주관적인 정보를 파악하는 데 효과적인 영화 리뷰를 이용한다. 본 논문에서 제안한 방법은 기존의 협업 필터링 방법보다 성능평가 결과, 향상된 성능을 보였다.

**키워드 :** 상황기반 기법, 협업 필터링, 영화 추천 시스템, 영화 리뷰

### 1. 서 론

스마트폰의 보급과 콘텐츠를 시청하고 검색할 수 있는 웹 서비스의 등장으로 콘텐츠에의 접근이 용이해져 시간과 장소에 구애받지 않고 다양한 콘텐츠를 사용자에게 제공하는 것이 가능해졌다. 이에 따라 원하는 콘텐츠를 찾고자 하는

요구가 증가되었다. 하지만 서비스되는 콘텐츠의 종류가 다양해지고, 그 수가 방대하여 이러한 정보의 홍수 속에서 사용자들은 정보가 너무 많기에 자신이 원하는 정보를 찾거나 상품을 선택하여 구매하는 데에 어려움을 겪고 있다. 서비스 제공자는 이러한 문제점을 해결하고자 노력해왔다. 대표적인 예로 사람들이 많이 검색한 내용을 순서대로 보여주는 실시간 검색 제공 서비스를 비롯하여, 쇼핑몰에서는 가장 많이 팔리고 있는 제품 혹은 최신 제품을 가장 눈에 띄는 부분에 배치하고 있다. 이러한 서비스들은 불특정 다수의 선택으로 이는 개개인의 특성을 전혀 반영하고 있지 못하기에 사용자의 실질적인 의사결정에 큰 영향을 끼치지 못하고 있다. 서비스 제공자는 이러한 문제점을 극복하기 위해 선

※ 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2014r1a1a4a01007190).

† 정 회 원 : 순천향대학교 전산학과 박사수료

†† 종신회원 : 순천향대학교 컴퓨터소프트웨어공학과 교수

††† 종신회원 : 순천향대학교 컴퓨터소프트웨어공학과 부교수

Manuscript Received : July 17, 2015

First Revision : August 28, 2015

Accepted : August 31, 2015

\* Corresponding Author : Park Doo-Soon(parkds@sch.ac.kr)

호도에 따른 콘텐츠의 검색과 추천을 사용자들에게 쉽게 제공할 수 있는 방법에 대한 연구를 진행해왔다[1-3].

이를 통해 고객의 개인적인 성향 및 구매 패턴 등을 파악하고 이를 바탕으로 서비스를 제공해주는 이러한 흐름은 개인화 서비스라는 이름으로 각광받고 있다. 개인화 서비스는 개개인에 대한 정보를 바탕으로 개인에 가장 알맞은 서비스를 소비자가 필요로 하는 제품이나 서비스를 명시적으로 묻지 않고 제공하는 것을 뜻한다.

이러한 서비스로는 협업 필터링 방법이 있는데, 고객이 좋아할만한 서비스나 항목을 추천해주는 Amazon이나 CD Now 등 인터넷 쇼핑몰에서 많이 사용되고 있다[2]. 그러나 이러한 방식은 고객들의 프로파일 정보를 기반으로 추천하므로 희박성 문제와 cold-start 문제가 있어 추천의 정확도가 떨어질 수 있으므로 개인에게 더 정확하게 추천하기 위해 협업 필터링 방법과 상황기반 기법을 함께 활용한다.

즉 본 연구에서는 선호도를 이용한 협업 필터링 기법과 다양한 개인화 요인 중 사용자의 상황에 따라 달라지는 사용자의 선호도를 반영하기 위해 상황기반 기법의 상황정보이자 개인화 요인으로 개개인이 처한 상황을 반영할 수 있는 영화리뷰의 감정을 이용해 개인에게 더 정확하게 맞는 아이템을 추천한다.

이에 본 연구에서는 사용자에게 더 정확한 맞춤형 추천을 위해 협업 필터링 기법의 선호도 정보뿐 아니라 상황기반 기법의 상황에 따라 달라지는 개인의 상황을 고려할 수 있는 리뷰를 이용해 이 두 개의 기법을 결합한 하이브리드 협업 필터링 기법을 제안한다.

본 논문에서는, 협업 필터링 방법에서는 선호도 분석을 위해 사용자가 남긴 평가치를 사용하였으며 상황 기법에서는 영화 사이트에 남겨진 리뷰의 감정 추론을 위해 영화에 대한 감정 사전인 리뷰 온톨로지를 정의하며 이를 기반으로 영화 리뷰의 감정단어에 대한 감정 정도를 분석하여 협업 필터링 방법과 상황기반 기법을 함께 활용하여 사용자에게 더 정확한 개인화 영화를 추천한다.

## 2. 연구 배경 및 관련 연구

### 2.1 추천 시스템

추천 시스템은 고객이 관심을 가지는 상품에 관한 정보나 인구 통계학적 정보, 과거 구매 행동 분석을 토대로 고객의 요구에 맞는 항목을 추천해주는 시스템이다[4]. Amazon과 Netflix를 시작으로 성과를 거두기 시작하면서 사용자의 취향에 맞는 추천을 위해 기존에 입력했거나 사용했던 기록 및 사용자로부터 정보를 입력받는 등 다양한 방법을 이용한 추천 시스템이 제공되고 있으며, 해외뿐만 아니라 국내에서도 이러한 추천 시스템들이 등장하기 시작하였고 네이버 뮤직 라디오, 왓차 등이 그 대표적인 예라고 할 수 있다.

이렇게 이미 추천 시스템은 우리 주변에서도 쉽게 찾아볼 수 있을 정도로 널리 사용되고 있고, 이와 함께 사용자에게 더 큰 만족도를 주기 위한 연구는 계속되고 있다.

### 2.2 개인화 기법

개인화 기법은 추천 시스템의 한 방법으로 사용자로부터 입력받은 명시적/묵시적 데이터를 바탕으로 개개인의 성향을 반영하여 추천의 질을 높이는 기술을 의미한다[2].

#### 1) 협업 필터링 추천 방식

협업 추천 시스템은 다른 사용자가 이전에 평가점수를 남긴 아이템을 기반으로 추천받는 사용자를 위해 여러 아이템의 가치를 예측한다[2].

협업 필터링 시스템에서 일반적으로 사용되는 k-최근접 이웃 모델은 활성 사용자(active user)가 다른 사용자들과의 거리를 계산하고 거리가 가장 가까운 k명의 사용자들을 이웃으로 선별한다. 사용자들 간의 거리를 구하는 데 피어슨 상관계수를 사용하며 Equation (1)과 같이 얻어진다.

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

$P_{a,u}$ 는 사용자 a와 사용자 u의 유사도를 뜻하고 n은 이웃의 수를 뜻한다. Equation (1)을 통해 예측 점수가 높은 n개의 아이템을 찾는다.

협업 필터링은 여러 분야에서 널리 연구 및 적용되고 있다. Ding[5]은 협업 필터링 방식에 시간에 대한 가중치를 추가하였고, Ungar[6]는 협업 필터링을 위한 클러스터링 방법을 연구했으며, Amazon, CDNow, MovieFinder 등에서 협업 필터링 기법을 사용하여 상품을 추천하고 있다.

협업 필터링은 아이템 정보를 파악할 필요가 없고 실제 사용자가 평가한 점수를 바탕으로 추천해주기 때문에 사용자의 만족도가 높으며, 본 연구에서는 선호도 정보로 평가치를 활용하여 사용자와 사용자 간의 유사점을 분류하는 방식의 방법을 사용했다.

#### 2) 상황기반 추천

상황(Context)이란 한 독립체의 환경을 특정 짓기 위해 사용될 수 있는 어떤 정보를 말한다. 이때 독립체는 사람, 장소, 물리적 또는 컴퓨터 객체가 될 수 있다. 이처럼 상황기반 추천은 사용자를 둘러싼 시간, 감정, 장소 등과 같은 여러 가지 환경을 고려하여 사용자에게 맞는 아이템을 추천하는 방법이다[7].

상황기반 기법의 대표적인 영화 추천 사례는 Ono가 제안한 베이지안 네트워크(Bayesian Network)를 적용한 상황 인식 영화 선호 모델이다. 이 시스템은 사용자의 상황정보(동행 인물, 장소, 감정)를 입력받고 등록되어있는 사용자 정보와 결합, 베이지안 네트워크 추론 엔진을 사용하여 후보 영화들의 평점을 확률적으로 계산하는 방법을 제시하였다[8]. 그러나 사용자의 입력을 요구하며 항상 정보 개념의 변화 없이 같은 개념 수준의 추천 정보만을 제공하는 단점이 있다. 관광 추천 서비스를 목적으로 한 CRUMPET 시스템 역시 사용자의 입력을 요구하며 상황에 따른 사용자의

선호도를 고려하지 못한다는 단점이 있다[9].

상황정보를 고려하는 추천 연구는 시작 단계이고 대부분 시간, 장소 등의 정보를 사용자가 직접 입력해야 하기 때문에 한계가 있다. 따라서 본 연구에서는 사용자의 입력 요구 없이 상황에 따른 사용자의 선호도를 고려하고 상황을 인식할 수 있는 상황정보로 감정을 이용하고 감정을 파악하는데 효과적인 리뷰를 이용해 추천한다.

이처럼 추천 시스템은 고객이 얼마나 만족할만한 항목을 추천할 수 있는지에 초점을 맞추고 있다. 이 연구에서는 추천 시스템에서 가장 많이 연구되고 있는 협업 필터링 기법과 개개의 상황을 고려한 상황기반 기법을 이용한다.

즉, 본 논문에서는 보다 더 정확한 개인화된 서비스를 위해 협업 필터링의 사용자의 선호도 분석과 상황기반 기법의 상황정보로 사용자의 현재 상황을 파악하는 중요한 변수로 리뷰를 이용해 사용자의 상황정보와 선호정보를 동시에 고려한 협업 필터링 기반 하이브리드 추천 방법을 제안한다.

### 2.3 온톨로지

온톨로지는 단어와 관계들로 구성된 일종의 사전으로 생각할 수 있으며, 그 속에는 특정 도메인에 관련된 단어들이 계층적으로 표현되어있고, 추가적으로 이를 확장할 수 있는 연관관계가 포함되어있어, 웹 기반의 지식처리나 응용 프로그램 사이의 지식 공유 등이 가능하도록 되어있다. 다시 말해 특정 분야에서 사용되는 표준 어휘의 모임이라고도 할 수 있다. 도메인에서 에이전트의 수행을 위한 문장을 만들고, 질의어를 이용한 정보검색을 위해서는 도메인의 개념화를 사용해야 하는데 여기에서 도메인의 개념화를 위해 도메인에 관한 지식을 표현하고 의사소통하기 위한 단어를 제공한다. 음악이나 영화 추천 분야에서도 온톨로지를 활용하여 추천을 제공하는 연구가 있다[10-11].

이처럼 온톨로지 기반 접근 방법을 사용하면 정보의 확장 및 재사용이 용이하다. 본 논문에서는, 획득되는 상황정보인 리뷰는 끊임없이 변화하는 상황을 인식해야 하는 정보로 특정 도메인에 필요한 상황정보를 온톨로지 모델로 표현 및 저장하고 상황정보의 확장 및 재사용이 용이하도록 한다.

즉, 본 논문에서는 상황기반 기법을 이용하기 위해 상황정보의 변수로 리뷰를 사용해 리뷰 온톨로지 모델을 이용한 접근 방법을 제안한다.

### 3. 리뷰를 이용한 상황기반 기법 모델링

상황기반 기법을 이용한 추천 서비스를 위해서는 상황에 대한 모델링이 필수적이다[9]. 본 논문에서는 상황을 표현하고 추천할 수 있도록 온톨로지를 기반으로 모델링 한다. 사용자에게 영화의 연관성을 도출하기 위해 상황정보인 리뷰의 감정단어들을 기반으로 한 감정 사전 온톨로지를 정의해 영화목록을 검색하고 추천할 수 있도록 한다. Fig. 1은 리뷰 온톨로지 모델링 된 정보로부터 상황을 인식하고 상황정

보로부터 상황 필터링을 이용하여 영화를 추천하는 리뷰 온톨로지 모델이다.

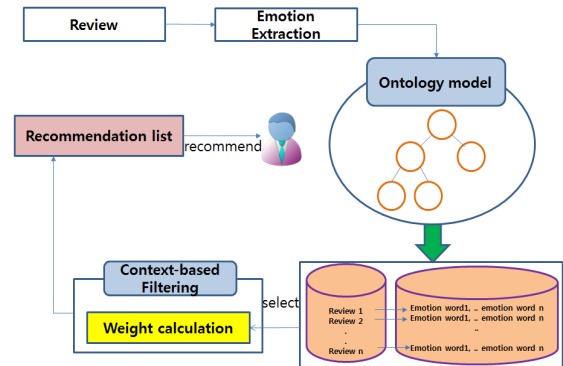


Fig. 1. Reviews Ontology Model Using User Context

Fig. 1의 리뷰 온톨로지 모델은 영화의 연관성을 도출하기 위해 리뷰의 감정과 영화 메타데이터인 키워드 간의 연관성을 표현하기 위한 모델로 구성된다. 온톨로지에 정의된 단어를 이용하여 리뷰에 감정단어가 있는지 분석하고 영화 메타데이터 중 감정과 관련된 키워드가 있는지 분석하여 연결되어있는 영화 목록을 검색한 후 리뷰에서 다시 감정과 관련된 키워드를 제외한 영화 키워드가 있는지 분석한다.

즉, 리뷰 온톨로지를 통해 분류되는 영화의 예를 들면 “아버지 세대 이야기인데 슬프고 감동적이네요. 험난한 역경을 딛고 살아오신 부모님에게 경의를 표합니다.”에서 리뷰 온톨로지에 정의되어있는 기본 감정단어 및 유사 단어, 용언이 있는지 구분하여 분류한다. 위 리뷰의 단어 중 ‘슬프고’는 온톨로지에 정의되어있는 ‘슬픔’의 용언이므로 Ekman의 기본 감정 여섯 가지 중 ‘슬픔’으로 분류된 후 ‘슬픈’이란 키워드를 가진 영화 목록들이 검색되고, 검색된 영화 중에서 리뷰에서 감정 관련 단어를 제외한 영화 키워드인 ‘아버지’가 있으므로 ‘슬픔(슬픈)’과 ‘아버지’를 연결하여 영화를 검색하고 이는 ‘국제 시장’ 등이 해당된다. 리뷰 데이터는 검색 엔진을 통한 데이터가 아닌 세계 최대의 인터넷 영화 데이터베이스(IMDB)로부터 추출한다. 상황정보의 감정을 이용하기 위해 리뷰의 감정을 분석 및 분류했으며 단어 분석 시 중요한 것은 어떤 단어들이 사용자들의 감정을 표현했는지를 찾아내는 것이다. 감정단어에 대해 Ekman[12]이 정의한 여섯 가지 기본 감정(분노, 역겨움, 놀라움, 슬픔, 공포, 행복)을 이용하는데 이 여섯 가지 기본 감정은 국가, 종교, 문화 등에 관계없이 사람의 표정에서 느낄 수 있는 감정을 정의한 것으로, 일반적으로 누구나 같은 감정을 느낄 수 있는 기본 감정이다. 하지만 기본 감정에 대해 기본 감정으로 느낄 수 있는 다른 단어에 대해 범주별로 분류가 되어있지 않아 범주별로 분류한다. 예를 들면, ‘행복’이란 기본 감정을 분류할 때, ‘행복’의 유의어인 ‘만족’ 또한 ‘행복’의 감정이기 때문에 각 기본 감정에 대해 같은 정서 상태를 표현하는 유사 단어로 정의했다. 본 논문에서 유사 단어 정의

를 위해 사용한 WordNet 3.0은 명사, 동사, 형용사, 부사의 네 개 품사에 속하는 총 147,278개 영어 단어들을 117,659개 동의어 집합(synonym set)들로 수작업 분류해놓은 어휘 사전이다[13]. 또한 실제 정서 자극으로서 특정 단어의 경우, 명사 이외의 형용사와 동사와 같은 다른 품사가 정서를 더 잘 유발시키는 경우도 있어 리뷰로부터 감정단어와 유사 단어에서 명사, 형용사, 동사를 추출한다. 여기서 유사 단어의 예를 들면, Ekman의 기본 감정 ‘슬픔’이 포함된 유사 단어는 비애, 설움, 애통, 비탄, 비극, 애수, 비감 등으로 정의된다. 또한 ‘슬픔’의 용언으로는 슬픈, 슬플, 슬프고, 슬프니 등으로 정의한다. 감정단어들의 가중치 값은 단어의 여러 의미들이 갖는 감정 정도를 파악하기 위해 감성 분석기의 범용성, 도메인 확장성, 의미 수준 고급 처리 등의 측면에서 장점이 많은 SentiWordNet[13]을 사용한다. 본 논문에서 사용한 감정단어에 대한 온톨로지는 Fig. 2와 같다.

리뷰 온톨로지에 정의되어있는 각 감정단어와 영화 간의 연관성을 도출하기 위한 식은 Equation (2)와 같다.

$$RO(w) = hits(NUR \cap \sum_{i=1}^n RO(w_i)) + hits(NUR \cap \sum_{j=1}^n RO(w_j)) \quad (2)$$

$$+ hits(NUR \cap \sum_{k=1}^n RO(w_k))$$

여기서 RO(w)는 리뷰 온톨로지에 정의된 기본 감정단어와 유의어, 용언을 나타내며, hits는 각 단어가 포함된 개수이며, NUR은 nearest\_user\_reviewdataset를 나타낸다.

#### 4. 협업 필터링과 상황기반 기법을 이용한 영화 추천 시스템

본 절에서는 기존의 협업 필터링 방법보다 더 정확한 개인화 추천을 위해 협업 필터링 방법과 상황기반 기법을 같이 이용하는 하이브리드 방법을 제안한다. Fig. 3은 하이브리드 방법으로 기존의 협업 필터링 방식에 사용자의 상황정보로 리뷰를 이용해 사용자와 상황정보 간의 선호관계를 표현하는 상황기반 기법을 포함시킨 하이브리드 추천 시스템이다.

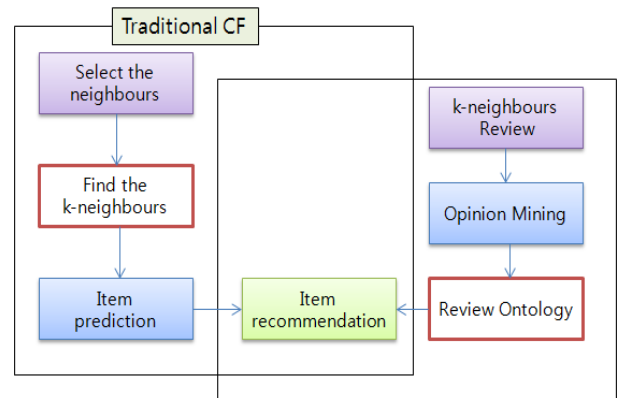


Fig. 3. Hybrid Method that Combines Techniques Collaborative Filtering Methods and Context Based Methods

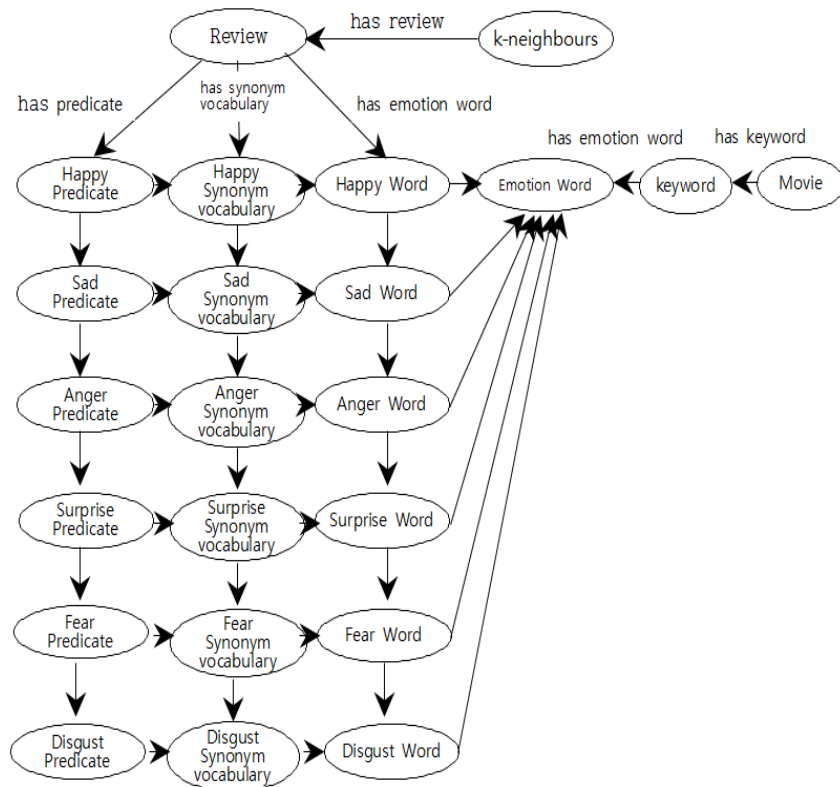


Fig. 2. Ontology Model for Emotion Words

제안된 방법은 협업 필터링 방법의 데이터 준비, 사용자 평가 분석, 이웃 분석, 아이템 분석을 포함하며 상황기반 기법은 데이터 준비와 리뷰 감정 분석, 감정 분석을 통한 아이템 분석을 포함한다. 협업 필터링 방법에서는 사용자의 평가치를 활용해 선호도 분석을 하였으며 상황기반 기법에서는 3절에서 제안된 리뷰 온톨로지에 정의된 각 단어에 대한 가중치를 부여하는 방법을 사용하였다. 제안하는 방법에서 사용하는 알고리즘은 협업 필터링 방법을 이용해 선정된 선호도가 비슷한 사용자들의 리뷰 데이터를 이용해 리뷰에서 감정 분석을 하며 추출된 감정단어를 분류해 Equation (4)의 가중치를 적용 후 Equation (3)을 이용해 가중치를 계산한다. 각 감정단어에 대한 중요도 차이를 두기 위해 검색된 단어에 따라 가중치를 다르게 가지며 가중치는 Equation (3)과 같다.

$$W(M_w) = \sum_{i=1}^n W(word_i) + \sum_{j=1}^n W(word_j) + \sum_{k=1}^n W(word_k) \quad (3)$$

$M_w$ 는 영화 리뷰에 출현하는 특정 단어의 가중치 합이고, 각 word의 가중치는 Equation (4)와 같다.

$$W(word) \begin{cases} 0.28 & word_i = Emotion \ word \\ 0.23 & word_j = Synonym \ word \\ 0.27 & word_k = Predicate \ word \end{cases} \quad (4)$$

단어 가중치의 값은 각 감정단어에 대해 온톨로지에 정의된 감정단어들을 SentiWordNet에서 검색 후, 긍정과 부정 감정 강도 차이들의 평균값과 긍정과 부정 및 중립값의 각각 감정에 대한 평균값을 구한 뒤 MAE값이 가장 작게 산출된 가중치를 적용하며 가중치 산출 알고리즘은 Algorithm 1과 같다.

Algorithm 1. Emotions Word Weight Calculation Algorithm

- 1 : in the nearest\_user\_review find the emotion words.
- 2 : search SentiWordNet
- 3 : get Pos value, values of Neg and Neu
- 4 : calculate the difference between the values of the Pos and Neg
- 5 : calculate the average of the above formula
- 6 : calculate the average values of the Pos
- 7 : calculate the average values of the Neg
- 8 : calculate the average values of the Neu
- 9 : calculate the MAE values of the above 5, 6, 7, 8
- 10 : application to the smallest MAE value of the above 6

상황기반 기법을 사용하기 위한 알고리즘은 Algorithm 2

Algorithm 2. Context Based Algorithm Techniques

- 1 : get nearest\_user\_review data(U)
- 2 : nearest\_user\_reviewdataset  $\neq$   $\emptyset$
- 3 : extracts an emotion words that are defined in an ontology in the reviews
- 4 : distinguish words
- 5 : select  $W(Word)$  using equation(4)
- 6 : calculate  $W(M_w)$  using equation(3)

와 같다.

검색된 영화 목록은 사용자에게 추천되기 위한 후보 영화가 된다. 상황기반 기법을 통해 후보 영화가 정해지면, 본문에서 제안한, 협업 필터링 방식에 상황기반 기법을 결합한 하이브리드 추천 시스템에서 사용자 평가를 기반으로 한 협업 필터링 방법과 사용자의 상황정보인 리뷰를 기반으로 한 상황기반 기법으로 검색된 아이템 예측점수  $P(CR)$ 는 Equation (5)와 같다.

$$P(CR) = \sum_{x=1}^i W(M_x) + P \quad (5)$$

$P$ 는 Equation (1)의 예측치를 나타내며, Equation (5)의 값은 최종적으로 사용자에게 추천될 영화의 순위를 가지게 되고 사용자는 순위가 정해진 영화에 따라 영화를 추천받는다. 결과 값이 큰 순서부터 Top N 리스트를 생성하게 되고 최종 리스트의 결과 값에 따라 병합한 뒤 정렬하여 추천 결과를 보여지게 된다. 이는 Algorithm 3과 같다.

Algorithm 3. Movie Recommendation System that Combines the Collaborative Filtering Techniques and Context-based Techniques

- 1 : INPUT :  $U_i$ :user, $R$ :rating,  
RU: recommended receive user
- 2 : count rating (RU, R)
- 3 : user clustering( $RU, U_i$ )
- 4 : find nearest RU
- 5 : the first selected candidate movies
- 6 : get  $U_i$ \_review of nearest RU
- 7 : in the  $U_i$ \_review find the emotion words.
- 8 : calculate the weight of each word
- 9 : applying review ontologies
- 10 : the second selected candidate movies
- 11 : combination Collaborative Filtering and Contextual-based techniques
- 12 : recommended movie list with 20 movie.

## 5. 성능 평가

본 절에서는 리뷰 온톨로지에 정의되어있는 각 단어들의 가중치를 계산하기 위한 성능평가를 먼저 하고, 다음으로 제안하는 방법과 협업 필터링 방법에 대해서 성능평가한다. 감성 강도 중에 가장 영향이 큰 주요 감성들을 분석하기 위해 4개의 감성관계를 이용하여 분석하고, 각 방법들에 대한 성능평가는 MAE를 통해 성능평가를 시행한다. 또한 이 연구에서 제안하는 추천 시스템을 평가하기 위해 협업 필터링 알고리즘과 제안하는 알고리즘에 대해 정확도(Precision)와 재현율(Recall)을 기준으로 평가를 수행하였다.

본 연구에서 사용한 데이터는 MovieLens 1M dataset로 사용자의 수가 6040명, 항목, 즉 영화의 수가 3883편, 선호도 평가 데이터의 수가 약 1,000,000개이다. 또한 영화 리뷰는 IMDB로부터 수집해 사용한다.

5.1 가중치 분석을 위한 성능평가

예측의 정확도는 일반적으로 MAE(Mean Absolute Error)와 RMSE(Root Mean Squared Error) 같은 지표를 통해 측정된다. 본 논문에서는 각 감성단어의 가중치 분석을 위해 성능평가는 MAE를 통해 값을 구했다. 본 연구에서 사용한 성능 측정 지표 MAE는 Equation (6)과 같다.

$$MAE = \frac{\sum_{k=1}^n |P_{a,i} - P(CR)|}{N} \tag{6}$$

즉, 주어진 선호도 평가 데이터가 들어있는  $P_{a,i}$ 와 Equation (5)에 의해 구해진 선호도 P(CR)의 차이를 누적하고 데이터의 수로 나누어 평균을 구한다. 본 논문에서는 가중치를 선정하기 위해 4가지의 방법에 대해 MAE값을 구했으며 첫 번째 방법은 Equation (7)이다.

$$W(word) = \frac{\sum_{k=1}^n |Pos|}{N} \tag{7}$$

Equation (7)은 긍정 강도 값들의 평균이고 Pos는 긍정 강도 값이며 N은 용어 샘플의 총수를 나타낸다.

두 번째 방법인 Equation (8)은 긍정과 부정 감성 강도의 차이들의 평균을 구하였으며, 여기서 Pos는 긍정의 강도 값을, Neg는 부정의 강도 값을 나타내고, N은 용어 샘플의 총수를 나타낸다. 세 번째 방법 Equation (9)는 부정 값들의 평균을 구하였으며, 네 번째 방법인 Equation (10)은 중립 값들의 평균을 구한다.

$$W(word) = \frac{\sum_{k=1}^n |Pos - Neg|}{N} \tag{8}$$

$$W(word) = \frac{\sum_{k=1}^n |Neg|}{N} \tag{9}$$

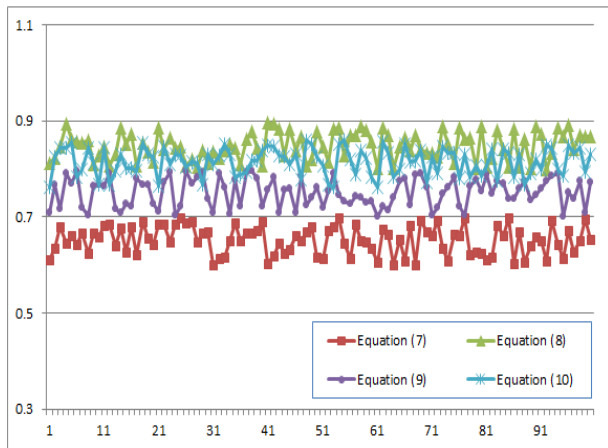


Fig. 4. Performance Evaluation for Weight(part)

$$W(word) = \frac{\sum_{k=1}^n |Neu|}{N} \tag{10}$$

Fig 4는 방정식 4개(Equation (7), Equation (8), Equation (9), Equation (10))를 적용한 Equation (5)에 대한 성능을 MAE 지표로 나타낸 것이다.

이 그래프는 약 3900개의 데이터에 대해 MAE를 구한 그래프의 일부이며, 그래프를 보면 Equation (8)의 가중치를 적용했을 경우 MAE값이 전체적으로 높아 성능이 낮다는 것을 확인할 수 있으며, Equation (7)의 가중치를 적용했을 경우가 전체적으로 MAE값이 낮아 가중치 식 4개 중 성능이 가장 좋은 것을 확인할 수 있다.

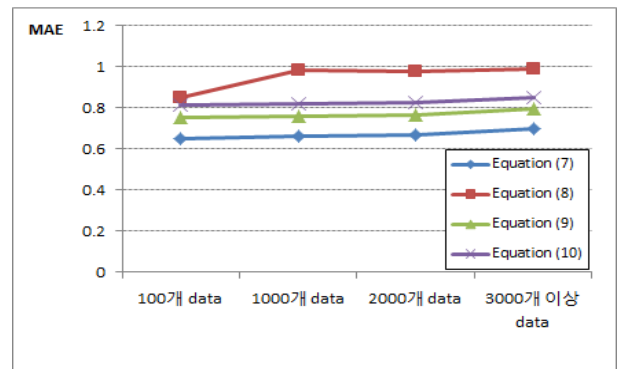


Fig. 5. MAE Graph Comparing the Average of the Weight Change

Fig. 5는 가중치 변화에 따른 MAE 평균값에 대한 성능 비교 그래프이며, 그래프를 보면 데이터 양에 따라 비교한 결과 데이터가 적거나 많아도 가중치 Equation (7)의 성능이 좋은 것을 볼 수 있다. Table 1은 가중치와 데이터 개수에 따른 MAE 성능 비교 결과표이다.

Table 1. Experimental Results

	100 data	1000 data	2000 data	3000 data 이상
Equation (7)	0.652	0.663	0.671	0.697
Equation (8)	0.848	0.980	0.975	0.989
Equation (9)	0.751	0.760	0.768	0.794
Equation (10)	0.815	0.821	0.828	0.848

다양한 가중치값을 적용하여 MAE 방식으로 실험한 결과 긍정의 강도값을 용어 샘플의 수로 나눈 방법인 Equation (7)로 가중치를 주었을 경우에 가장 높은 성능을 보여 최종 가중치로 선정하였다.

5.2 추천 시스템의 성능평가

협업 필터링을 변형한 많은 알고리즘은 MAE값이 대체로 0.7을 기준으로 크게 향상되는 것은 아니다.

협업 필터링 방법과 제안하는 방법의 MAE 평균값에 대한 성능평가는 Fig. 6과 같다. 데이터가 1000개 정도까지는



기존의 협업 필터링 방법의 MAE값이 낮게 나타나지만, 데이터가 증가할수록 제안하는 방법의 MAE값이 낮게 나타난다. 상황기반 기법을 이용한 상황 보인 리뷰의 특성상 리뷰 데이터의 양은 계속 증가할 수밖에 없으므로 제안하는 방법으로 성능을 개선할 수 있다. 결과적으로 약 3900개 데이터에 대해 MAE 성능평가한 결과, 제안하는 방법은 기존 협업 필터링 방법에 비해 MAE가 0.028 낮게 나타난다.

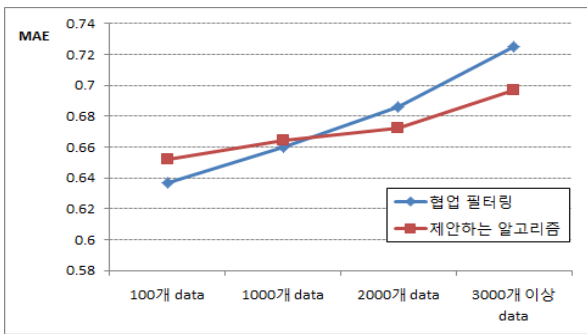


Fig. 6. The Proposed Method and the Collaborative Filtering Method MAE Compared to the Average Performance Graph

제안하는 방법은 기존의 협업 필터링 방법에 비해 성능을 약 3.85% 개선한다. 3.85%의 성능은 결코 적은 수가 아니다. Fig. 7은 정확도와 재현율을 기존의 협업 필터링 방법과 제안하는 하이브리드 추천 시스템과 비교한 그래프이며, Fig. 8은 추천 목록의 개수에 따른 제안하는 하이브리드 추천 시스템과 기존의 협업 필터링 방법의 정확도와 재현율의 성능 비교 그래프이다.

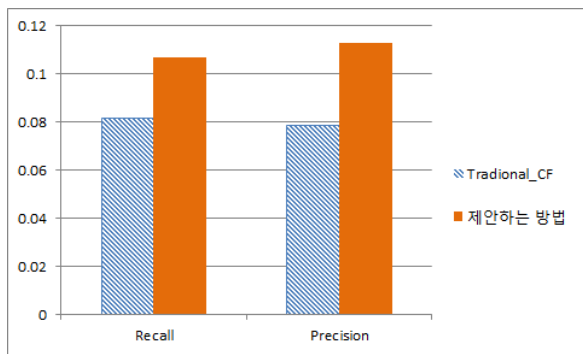


Fig. 7. The Precision and Recall of Collaborative Filtering and Proposed Method

Fig. 8은 제안하는 추천 시스템의 성능이 기존의 협업 필터링 추천 시스템보다 정확도는 약 0.034, 재현율은 약 0.026 더 나은 결과를 보이고 있어 성능이 향상된 결과를 확인할 수 있다.

본 논문에서 가중치 분석을 위한 MAE와 추천 시스템의 성능을 평가하기 위해 정확도, 재현율 방식으로 성능평가를 수행한 결과, 모든 평가에서 기존 협업 필터링 추천 시스템보다 협업 필터링과 상황기반 기법을 이용한 하이브리드 추천 방식의 성능이 향상된 결과를 보였다.

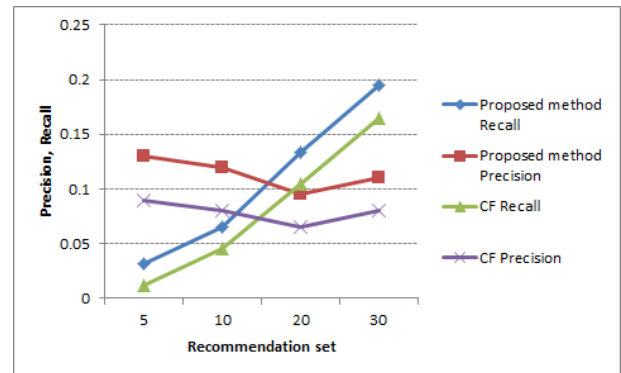


Fig. 8. The Precision and Recall Compared to the Performance of the Recommended List

## 6. 결 론

정보의 폭발적인 증가로 사용자들은 원하는 정보를 빠른 시간에 얻는 것이 힘들어져 이를 해결하는 다양한 방법이 제공되고 있지만, 직접적인 의사결정에 큰 영향을 주지 못하고 있어 의사결정에 도움을 주는 추천 시스템이 주목을 받으면서 그 중요성 또한 부각되고 있다.

이러한 추천 시스템 중 영화 추천에서 주로 사용하는 기법은 데이터마이닝 기법 중 가장 성공한 알고리즘 기법인 협업 필터링 알고리즘이 많이 사용되고 있다. 본 논문에서는 추천 시스템에서 가장 성공적인 방법인 협업 필터링 방법을 이용한다. 협업 필터링 방법은 고객들의 프로파일 정보를 기반으로 추천하는 방법으로, 개인에게 더 정확하게 추천하기 위해 협업 필터링의 선호도 정보와 사용자의 상황 정보를 이용한 상황기반 기법을 함께 이용한다. 본 논문의 시스템은 온톨로지를 기반으로 상황을 모델링 하였으며, 현재 감정이 슬픈지, 행복한지에 따라 아이템을 다르게 선택하는 경우도 있기 때문에 상황정보를 고려하기 위해 사용자의 상황정보로 다양한 요소 중에서 수많은 사용자들이 작성한 인터넷상의 의미 있는 텍스트 정보이고 감정이 풍부하게 드러낸 영화 리뷰를 이용한다. 본 논문에서 제안하는 기법을 평가하기 위해 상황기반 기법의 상황정보인 리뷰를 바탕으로 영화에 대한 감정 사진인 리뷰 온톨로지를 구축해 협업 필터링 방법과 결합하여 하이브리드 방법으로 사용자에게 더 정확하게 영화를 추천해주는 시스템을 구축하고 성능 평가하였다.

성능평가 결과, 본 논문에서 제안하는 협업 필터링과 상황기반 기법을 이용한 추천 알고리즘은 기존의 협업 필터링 방법을 이용한 추천 방법과 비교하여 평균 오류를 0.028 낮추는 결과를 보였고 성능 정도를 3.85% 개선하였으며, 협업 필터링 추천 시스템과 협업 필터링과 상황기반 기법을 이용한 추천 시스템의 성능평가 결과 정확도와 재현율의 평균이 각각 0.007, 0.027 향상된 결과를 보여 기존 추천 시스템의 정확도와 재현율을 개선해 개인에게 더 정확한 추천을 할 수 있다.

향후 연구로 감정단어의 분류기준을 최적화할 수 있는 연구와 사용자의 리뷰 이력 분석을 통하여 사용자의 특정 아이템에 대한 선호도를 분석하여 최적화된 개인화 추천을 제공받기 위한 기법을 연구하고자 한다.

### References

- [1] Fernando Ortega, Jose-Luis Sanchez, Jesus Bobadilla, and Abraham Gutierrez, "Improving collaborative filtering- based recommender systems results using Pareto dominance," *Information Sciences*, Vol.239, pp.50-61, 2013.
- [2] W. H. Jeong, S. J. Kim, D. S. Park, and J. Kwak, "Performance Improvement of a Movie Recommendation System based on Personal Propensity and Secure Collaborative Filtering," *Journal of Information Processing Systems*, Vol.9, Issue.1, Mar., 2013.
- [3] Subhash K. Shinde and Uday Kulkarni, "Hybrid personalized recommender system using centering-bunching based clustering algorithm," *Contents lists available as SciVerse ScienceDirect*, pp.1381-1387, 2012.
- [4] Zui Zhang, Hua Lin, Kun Liu, Dianshuang Wu, Guangquan Zhang, and Jie Lu, "A hybrid fuzzy-based personalized recommender system for telecom products/services," *Information Sciences* 235, pp.117-129, 2013.
- [5] Y. Ding, "Time weight collaborative filtering," *Proceedings of the 14<sup>th</sup> ACM international conference*, pp.485-492, 2005.
- [6] L. H. Ungar and D. P. Foster, "Clustering Methods for Collaborative Filtering," in *AAAI Workshop on Recommendation System*, pp.114-129, 1998.
- [7] P. Sessing, "Towards a Better Understanding of Context and Context-Awareness," *Computing System*, pp.304-307, 1999.
- [8] C. Ono, M. Kurokawa, and Y. Motomura, "A context-aware movie preference model using a Bayesian network for recommendation and promotion," *User Modeling* 2007, pp.247-257, 2007.
- [9] M. H. Ahn and J. H. Kwon, "Ontology based Context-Aware Recommendation System using Concept Hierarcht," *Korea Network Information Society*, Vol.8, No.5, Oct., 2007.
- [10] Gamon M, Aue A, Corston S, and Ringer E, "Mining customer opinions from free text," *International symposium on intelligent data analysis(IDA)*, pp.121-132, 2005.
- [11] Wong, W., Liu, W., and Bennamoun, M., "Ontology learning from text: A Look Back and into the Future," *ACM Computing Surveys(CSUR)*, Vol.44, Issue.4, 2012.
- [12] P. Ekman, "Expression and the nature of emotion," *Approaches to Emotion*, pp.319-343, 1984.
- [13] InSu Kang, "A Comparative Study on Using SentiWordNet for English Twitter Sentiment Analysis," *Journal of Korean Institute of Intelligent System*, Vol.23, No.4, Aug., pp.317-324, 2013.



### 김민정

e-mail : kmj6906@hanmail.net  
 2001년 순천향대학교 정보처리공학(학사)  
 2004년 순천향대학교 전산학과(석사)  
 2006년 순천향대학교 전산학과 박사수료  
 관심분야: 데이터마이닝, 병렬처리



### 박두순

e-mail : parkds@sch.ac.kr  
 1988년 고려대학교 전산학전공(이학박사)  
 2015년~현 재 한국정보처리학회 회장  
 2014년~현 재 순천향대학교 중앙도서관  
 관장  
 2014년~현 재 순천향대학교 웰니스코칭  
 서비스연구센터 센터장  
 1985년~현 재 순천향대학교 컴퓨터소프트웨어공학과 교수  
 관심분야: 병렬처리, 데이터마이닝, 병렬처리, 웰니스



### 홍민

e-mail : mhong@sch.ac.kr  
 1995년 순천향대학교 전산학과(공학사)  
 2001년 University of Colorado at Boulder  
 (공학석사)  
 2005년 University of Colorado at Denver  
 (이학박사)  
 2006년~현 재 순천향대학교 컴퓨터소프트웨어공학과 교수  
 관심분야: 컴퓨터 그래픽스, 다이나믹 시뮬레이션, 바이오인포매틱스, 연상처리



### 이화민

e-mail : leehm@sch.ac.kr  
 2000년 고려대학교 컴퓨터교육과(학사)  
 2002년 고려대학교 컴퓨터교육학과(석사)  
 2006년 고려대학교 컴퓨터교육학과(박사)  
 2006년~2007년 특허청 전기전자심사본부  
 통신사무관  
 2007년~현 재 순천향대학교 컴퓨터소프트웨어공학과 부교수  
 관심분야: 클라우드컴퓨팅, 모바일컴퓨팅, 웰니스, IoT, 컴퓨터교육