

자동 생성된 사용자 프로파일을 이용한 하이브리드 음식 추천 시스템

Hybrid Food Recommendation System Using Auto-generated User Profiles

정주석* · 강신재**†

Ju-Seok Jeong*, Sin-Jae Kang**†

* 대구대학교 대학원 컴퓨터정보공학과

** 대구대학교 정보통신대학 컴퓨터·IT공학부

요 약

본 논문에서는 트위터로부터 자동 생성한 사용자 프로파일을 이용하는 개인 맞춤형 음식 추천 시스템을 제안한다. 사용자 개인의 트위터 문장에서 명사를 추출하고, 감정단어와의 공기 여부에 따라 감정점수를 계산한 후, 사용자 프로파일을 생성한다. 각 음식에 관한 정보는 웹에서 검색한 웹 페이지를 분석, 가공하여 음식별 대표 명사 목록을 만들어 사용한다. 이렇게 구축된 사용자 프로파일과 각 음식별 특징 정보 간 유사도를 계산하여 사용자에게 맞는 음식을 추천 해준다. 제안하는 방법은 추천 시스템을 처음 사용하는 사용자라 할지라도 사용자 프로파일이 자동으로 구축되어 추천에 사용되기 때문에 항상 추천 결과를 얻을 수 있는 장점이 있다.

키워드 : 추천시스템, 사용자 프로파일, 하이브리드 기법, 트위터

Abstract

This paper proposes a personalized food recommendation system using user profiles auto-generated from Twitter. The user profiles are generated by extracting nouns from Twitter, and calculating emotional scores according to whether each noun is collocated with emotion words. Representative noun information for each food is constructed by analyzing web pages relevant to foods. Appropriate foods for users can be recommended by calculating similarities among the extracted resources. The proposed system has an advantage in that it can always recommend foods even if a user is a newcomer.

Key Words : Recommendation system, User profile, Hybrid technique, Twitter

1. 서 론

사람들은 무언가에 대한 선택을 할 때 주변에서 얻을 수 있는 조언이나 추천을 매우 중요하게 생각한다. 일상적인 예를 들어, 사람들은 매끼마다 무엇을 먹을까 고민을 하게 되는데, 이를 해결하는 한 방법으로, 개인의 취향과 개인이 처한 현재의 상황을 자동으로 파악하여, 개인이 좋아할 만한 가장 적당한 음식을 자동으로 추천해주는 시스템이 있다면, 사람들의 고민을 덜 수 있을뿐더러, 만족도도 높일 수 있게 된다.

추천 문제는 기본적으로 평가 구조에 기반을 두고 있으며, 사용자가 접하지 못한 새로운 아이템에 대한 평가의 추정 문제로 볼 수 있다. C 를 모든 사용자의 집합으로, S 를 추천될 수 있는 모든 아이템의 집합(예: 책, 영화, 식당, 음식 등)으로, u 를 사용자 c 에 대한 아이템 s 의 유용성을 평가하는 함수라 가정할 때, 모든 사용자 c 에 대해 각 사용자의 만족도를 최대화할 수 있는 아이템 s' 를 찾는 다음과 같은 식으로 추천 문제를 형식화할 수 있다.

$$\forall c \in C, s'_c = \arg \max_{s \in S} u(c, s) \quad (\text{식 1})$$

접수일자 : 2011년 6월 10일

완료일자 : 2011년 9월 28일

†Corresponding Author : sjkang@daegu.ac.kr

이 논문은 2010학년도 대구대학교 학술연구비 지원에 의하여 연구되었음

이와 같이 추천된 음식이 사용자의 만족도를 높이기 위해서는 사용자의 최근 관심분야, 취향, 성격, 건강상태, 계절, 날씨, 성별, 현재 위치 등 여러 가지 정보를 종합적으로 활용해야 하지만, 이러한 정보들을 수집하는 것은 쉬운 일이 아니다. 수작업으로 구축하여 유지/관리하기에는 비용과 시간의 문제

가 있을뿐더러 불편함 때문에 직접 입력을 꺼리는 사용자도 다수 존재한다. 따라서 이러한 정보를 자동으로 수집, 구축하여 활용하는 방법이 필요하다.

최근 개인의 생각을 단문 형태로 자유롭게 표현할 수 있는 트위터(Twitter)¹⁾와 같은 웹 서비스를 이용하는 사용자가 늘어나고 있다. 이러한 소셜 네트워크 서비스(SNS)에 등록된 사용자의 글에는 사용자의 평상시 취향이나 관심 사항을 나타내는 단어들이 많이 포함되어 있기 때문에, 사용자 프로파일을 자동으로 구축하기 위한 아주 유용한 정보가 될 수 있다.

본 논문에서는 이러한 정보를 수집, 분석하여 사용자 프로파일을 자동으로 구축하고, 이를 활용하기 위한 응용으로 음식 추천을 선택하였다. 기존의 추천 시스템들은 실행 초기에 사용자에게 관한 데이터(사용자가 기 선택한 음식 정보, 사용자가 기 검색한 음식 정보 등) 부족 문제로 인하여 정확한 추천을 할 수 없는 문제가 있다.

본 논문에서 제안하는 추천 시스템은 처음 사용하는 사용

록 관련 데이터가 누적되어 활용되기 때문에 보다 정확한 추천이 가능한 방법을 제안하고자 한다.

최근 아이폰(iPhone), 안드로이드(Android)²⁾ 폰과 같은 스마트폰이 기본적인 휴대전화로서의 기능뿐만 아니라 다양한 소프트웨어 애플리케이션을 설치/운영할 수 있는 컴퓨터의 기능과, 언제 어디서든 사용할 수 있는 이동성을 제공하고 있기 때문에[1], 본 추천시스템의 클라이언트를 안드로이드 플랫폼으로 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서 관련연구에 대해 살펴보고, 3장에서는 본 논문에서 제안하는 추천 방법에 대해 설명한다. 4장에서는 단계별 구현 결과와 실행 결과 화면을 제시하고, 마지막 5장에서 결론 및 향후 계획에 대하여 기술한다.

구분	참고문헌	추천대상	특징
내용 기반	김성기, 김영학 [4]	학습방법	<ul style="list-style-type: none"> 항목과 항목간의 관련성을 기반으로 한 항목기반 추천 코사인 유사도 계산 기법 이용 학습 요소에 따른 학습 방법에 대한 데이터 이용 학습 요소에 따른 방법의 가중치 설정(설문지)
	한현구, 서의현 [5]	음식	<ul style="list-style-type: none"> 학습기반 추천 사용자의 프로파일과 음식 주문 내용을 분석 음식 주문 내용을 기반으로 A priori 알고리즘 이용한 연관 규칙 사용 지지도와 선호도를 사용하여 추천
	S. K. Ray, S. Singh [6]	블로그	<ul style="list-style-type: none"> 블로그 내용을 기반으로 한 추천 온톨로지와 워드넷을 사용하여 블로그의 주제를 확장 블로그 콘텐츠 추출 및 랭킹 후, 사용자 취향에 맞는 블로그 선택
협력 기반	이제식, 이진천 [7]	음악	<ul style="list-style-type: none"> 사례기반 추론기법(Case-Based Reasoning) 사용 상황 인식을 기반으로 한 CBR모델로 음악 추천 시스템 개발 현재 상황 정보 구축(계절, 날씨, 요일, 기온 등) CBR모델에서 사용자간 유사도 계산
	오유수, 최아영, 우운택 [8]	음식	<ul style="list-style-type: none"> 생체 신호, 주변 환경 센싱 정보를 통합하는 맥락 인식 기법 사용 휴대 기기를 이용한 사용자 생체 신호 센서를 활용, 사용자의 건강 모니터링 사용자 병력, 센싱 정보를 토대로 프로파일 생성 센싱 데이터 변화와 음식 연관관계를 한의학적 해석법에 근거하여 추천
	이제식, 박석두 [9]	영화	<ul style="list-style-type: none"> 협업 필터링 피어슨 상관계수와 k-NN 유사도 계산 기법 사용 평가치들을 가중 평균하여 상위 N개의 장르 선정
	Y. Ying [10]	범용	<ul style="list-style-type: none"> k-NN 알고리즘을 사용하여 사용자 클러스터링 비어있는 등급 정보를 스무딩(smoothing) 스무딩을 반영한 사용자 클러스터링을 통하여 시스템 확장성 확보
하이브리드	김기수 [11]	인터넷 쇼핑물	<ul style="list-style-type: none"> 사용자간 유사도와 상품간의 유사도를 고려하는 하이브리드 방법 수정된 피어슨 상관계수를 사용하여 사용자간 유사도 계산 수정된 코사인 유사도 계산 기법을 사용하여 상품간 유사도 계산 두 유사도 값의 평균을 최종 선호도 값으로 사용
	C. Liu, C. Sun, and J. Yu [12]	모바일 상거래	<ul style="list-style-type: none"> 일반 사용자 프로파일과 모바일 디바이스 프로파일을 사용 여러 추천 알고리즘의 결과에 가중치를 부여하여 결합하는 하이브리드 추천 알고리즘 사용 다중 에이전트 구조로 추천시스템 설계

표 1. 추천기법에 따른 분류

자라 할지라도 사용자 프로파일이 자동으로 구축되어 추천에 사용되기 때문에 항상 추천 결과를 얻을 수 있고, 또한 하이브리드 추천 기법을 사용하여 추천 시스템을 사용하면 할수

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

2) <http://www.android.com>

2. 관련 연구

추천시스템은 추천이 어떻게 만들어지느냐에 따라 크게 내용 기반(content-based)³⁾, 협력 기반(collaborative), 그리고 이 둘을 결합한 하이브리드(hybrid) 형태의 세 가지 접근법으로 분류할 수 있다[2-3].

내용 기반의 추천 방법은 대부분의 사람들이 과거에 자신이 좋아했던 상품과 비슷한 상품이면 좋아하는 경향이 있고 반대로 싫어했던 상품과 비슷한 상품이면 싫어하는 경향이 있다는 점을 이용하여 구현된다. 즉, 한 사용자가 특정 아이템을 선호한다면, 이 아이템과 관련성이 높은 아이템을 추천하는 방식이다. 추천의 대상이 되는 아이템이 자주 바뀌지 않을 때 적합한 방법이다.

협력 기반 방법은 사용자들의 과거 경향이 미래에서도 그대로 유지될 것이라는 가정 하에, 비슷한 취향을 가진 고객들에게 서로 아직 구매하지 않은 상품들은 교차 추천하는 방법을 말한다. 추천의 대상이 되는 아이템이 자주 바뀔 때 유용한 방법이다.

하이브리드 방법은 내용 기반과 협력 기반 방법을 따로 구현하여 운영하고, 각각의 결과를 적절하게 조합해서 추천하는 방식이다. 추천에 필요한 데이터의 유무에 따라 융통성 있게 처리할 수 있는 장점이 있어 최근 연구에서 많이 채택하고 있는 방법이다.

관련 연구들을 추천기법에 따라 요약하여 정리한 내용을 <표 1>에 제시하였다. 제시된 추천 방법들은 기본적으로 많은 사용자들로부터 수집한 정보를 사용한다는 것이 특징인데, 만약 어느 사용자에게도 평가받지 못한 새로운 아이템이 있거나, 지금까지 한 번도 평가를 하지 않은 사용자가 있는 경우에는 제대로 추천을 하지 못한다는 문제가 있다.

따라서 데이터 부족 문제를 해결하여 추천시스템의 실행 초기 단계에서도 추천이 가능한 방법이 필요하다.

3. 음식 추천 시스템

본 논문에서는 사용자 프로파일을 자동으로 구축하고, 이를 이용하여 초기 단계부터도 추천을 할 수 있고, 시스템을 사용할수록 추천의 성능을 높일 수 있는 하이브리드 기반의 추천시스템을 제안한다.

3.1 시스템 구성

음식 추천 시스템은 크게 웹 서버, 모바일 클라이언트(안드로이드폰)의 두 가지 구성요소로 이루어진다. <그림 1>에 전체 시스템의 구성도를 제시하였다. 웹 서버는 트위터 서버에 주기적으로 접속하여 사용자가 등록한 글로부터 프로파일 정보를 자동으로 구축하게 되며, 또한 추천할 음식에 관한 정보를 관련 웹 사이트 검색을 통하여 미리 수집해 DB로 저장해 두게 된다. 모바일 클라이언트는 단순 트위터 작업을 할 때에는 트위터 서버에 접속하여 통신을 하게 되며, 음식 추천 기능을 이용할 때에는 웹 서버에 접속하게 되고, 웹 서버는 사

용자에게 적합한 추천 메뉴를 실시간으로 선정하여 클라이언트로 전송한다.

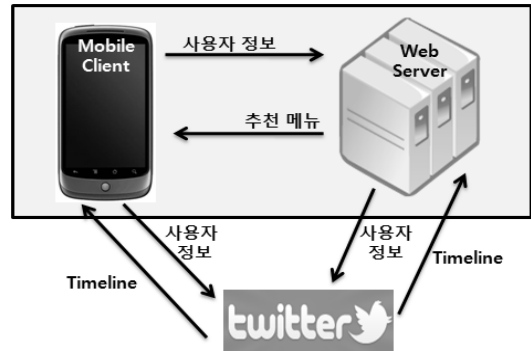


그림 1. 전체 시스템 구성도

3.2 음식 추천 시스템 구현 단계

사용자 프로파일을 자동으로 생성하고, 이를 기반으로 음식을 추천하는 과정은 다음과 같은 3가지 단계로 이루어진다.

• 단계 1: 추천 대상(음식)의 관련정보 수집

본 논문에서는 음식을 추천 대상으로 선정하였기 때문에 이 단계에서 음식에 관한 기본 정보와 음식별 대표명사 목록을 구축하게 된다. 대표명사 목록을 구축하는 과정은 <그림 2>와 같다. 먼저, 위키피디아와 네이버 백과사전과 같은 구조적 정보(테이블 등)에서 “음식 관련 명사 목록”을 추출하고, 여기에 [13]에서 제시한 음식 관련 명사 목록을 수작업으로 추가하여 약 2,400여개의 목록을 수집하였다. 이 정보는 통제 어휘(controlled vocabulary)의 역할을 하여, 음식과 관련된 명사들만 음식별 대표명사로 선택될 수 있도록 한다.

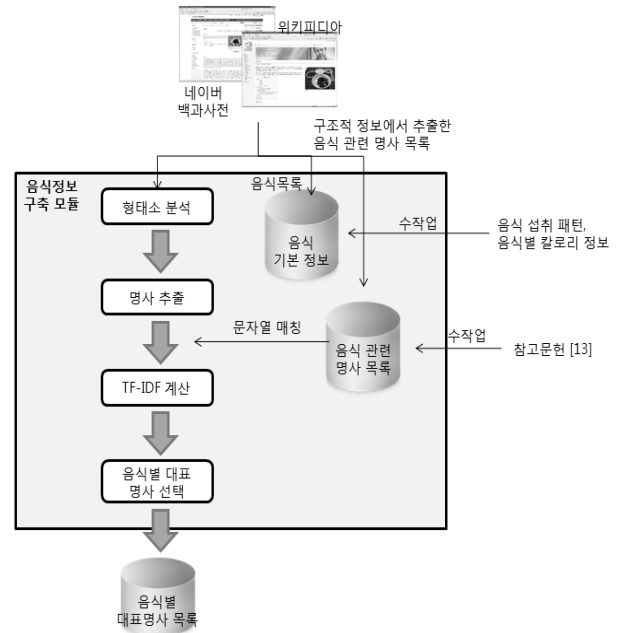


그림 2. 음식정보 구축 과정

“음식별 대표명사 목록”을 구축하기 위해서는 위키피디아와 네이버 백과사전에 존재하는 음식에 관한 웹 페이지를 수

3) 콘텐츠 기반 또는 항목 기반이라고도 한다.

4) 사용자 기반이라고도 한다.

집하고, 형태소 분석을 하여 후보 명사들을 추출한다. 여기에는 음식 관련 명사뿐만 아니라 일반 명사들도 포함되어 있기 때문에, 기 구축한 “음식 관련 명사 목록”에 등록된 명사만을 선택한다. 이 가운데 각 음식을 대표하는 명사, 다시 말해 음식 간의 구분을 용이하게 할 수 있는 변별력 있는 명사를 선택하기 위해 명사별 가중치를 계산할 목적으로 각 명사의 출현 빈도수를 계산한다.

명사의 문서 내에서의 가중치는 한 문서에 출현한 명사의 빈도수를 문서의 전체 길이로 정규화한 tf (term frequency)를 이용한다. 문서에서 많이 나타나는 명사가 그 문서를 대표하는 중요한 명사이므로 가중치를 높게 주겠다는 의미이다.

$$tf = f_{ik} / |D_i| \quad (\text{식 2})$$

f_{ik} : 문서 i 에서 단어 k 의 빈도
 $|D_i|$: 문서 D_i 의 길이

명사의 문서 간 가중치는 전체 문서집합 가운데 해당 명사가 나타난 문서가 어느 정도인지를 수치화한 idf (inverse document frequency)를 이용한다. 이는 특정문서에만 나타나는 명사는 가중치를 높이고, 대부분의 문서에 나타나는 범용 명사는 가중치를 낮추는 역할을 한다.

$$idf = \log_2 \left(\frac{N}{d_k} \right) \quad (\text{식 3})$$

N : 전체집합에 포함된 문서의 수
 d_k : 단어 k 를 포함하는 문서의 수

위의 2가지 가중치를 함께 고려하기 위해 각 문서에 속한 명사의 최종 가중치는 $tf \times idf$ 로 결정되며, 이 값이 높은 명사들을 각 음식별 대표 명사목록으로 선택하게 된다.

• 단계 2: 사용자 프로파일 자동 생성

사용자 프로파일을 자동으로 구축하는 과정은 <그림 3>과 같다.

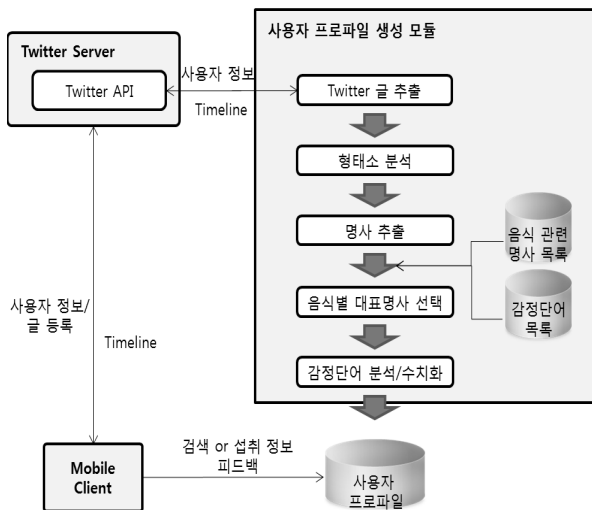


그림 3. 사용자 프로파일 자동구축 과정

먼저, 사용자 프로파일을 구축할 사용자의 정보(Twitter ID, access token, access secret)를 이용하여 트위터 서버에 접

속하고, 사용자가 올린 정보를 주기적으로 가져오기 위해 TImeline API를 호출한다[14]. 이러한 과정을 거쳐 사용자가 트위터에 올린 글을 추출하여 형태소 분석을 한 후, 후보 명사 목록을 만든다. 이 가운데 기 구축한 “음식 관련 명사 목록”에 등록된 명사만을 대상으로 고빈도 명사 순으로 선택한다. 선택된 명사 목록과 최대빈도 명사의 빈도수로 나누어 정규화한 각 명사의 빈도정보⁵⁾를 사용자 프로파일에 함께 저장하게 된다. 이 정도의 정보만으로도 사용자 프로파일의 역할을 할 수는 있겠지만, 사용자의 감정에 따른 보다 정교한 처리를 더하기 위해서 [15]에서 제시한 긍정적 또는 부정적 감정단어들을 “감정단어 목록”으로 구축하여 이용한다. 이 목록에는 감정 단어별로 점수가 부여되어 있는데 4점 이상인 단어는 긍정단어로, 4점 미만인 단어는 부정단어로 분류하고 있다. <표 2>에 감정단어들을 예시하였다.

긍정단어

단어	점수
행복하다	6.16
상쾌하다	5.99
기쁘다	5.94
흡족하다	5.77
황홀하다	5.72

부정단어

단어	점수
역겹다	1.67
혐오하다	1.80
분노하다	1.84
울화	1.87
괴롭다	1.97

표 2. 긍정 및 부정 감정단어 예시

위 목록에 있는 모든 감정단어가 음식 관련 명사들과 함께 사용되지는 않지만, 일부 감정단어들이 음식 관련 명사와 함께 나타나는 경우가 있기 때문에, 이를 다음과 같이 분석하고 수치화하여 사용자 프로파일에 추가 정보로 저장한다. 음식 명사들의 감정점수 초기값을 0으로 설정한 다음, 특정 음식 명사와 함께 사용된 감정단어가 긍정적 분류일 경우에는 1점을, 부정적 분류일 경우에는 -1점을 감정점수에 더한다. 이러한 과정을 반복하면서 각 음식 명사별 감정점수를 누적 합산한다. 만약 음식 명사와 함께 사용된 감정 단어가 없었던 초기값 0을 그대로 갖게 된다. 이렇게 구한 단어별 감정점수는 절대값이 가장 큰 감정점수의 값으로 각각 나누어 정규화⁶⁾한 후, 프로파일에 추가 저장된다.

• 단계 3: 하이브리드 기반의 음식 추천

1, 2 단계에서 구축된 정보를 이용하여 음식을 추천하는 과정은 <그림 4>와 같이 4가지 과정으로 나뉜다. 본 논문에서 제안하는 추천 방법은 하이브리드 기반의 방법으로, ②, ④과정에서 내용 기반의 추천 기법을 사용하고, ③과정에서 협력 기반의 추천 기법을 이용하게 되며, 각 과정의 결과들을 종합하여 최종 추천 결과를 생성하게 된다.

5) 정규화한 빈도정보(freq)는 $0 < freq \leq 1$ 의 값을 가진다.

6) 정규화한 감정점수(ES)는 $-1 \leq ES \leq 1$ 의 값을 가진다.

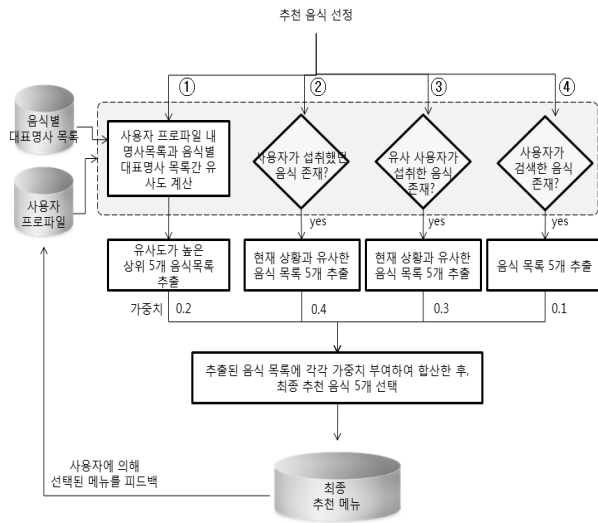


그림 4. 음식 추천 과정

추천과정에서 ①은 항상 실행하게 되고, ②, ③, ④과정은 해당 정보가 있는 경우에만 실행된다. 즉, 추천 시스템을 처음 사용하는 사용자인 경우에는 ①과정만을 통해 추천 메뉴를 결정하게 되고, 추천 시스템을 일정 기간 사용한 사용자인 경우에는 ①, ②, ③, ④ 전 과정을 통하여 사용자에게 보다 적합한 추천 메뉴를 결정하게 된다.

①과정은 기본적인 사용자 정보 즉, 트위터의 글에서 추출한 사용자 취향 정보와 음식 정보의 유사도를 계산하여 추천 리스트를 만들게 된다. 사용자의 취향 정보와 음식별 대표 명사 목록 정보는 모두 명사로 표현되어 있기 때문에, 두 목록간 중복된 명사(N)를 대상으로 추천점수를 계산한다. (식 4)는 사용자 프로필 u와 음식 f의 대표명사 목록을 가지고 추천점수(recommendation score, RS)를 계산하는 수식이다.

$$RS(u, f) = \sum_{\forall n \in N} (tfidf_{f,n} \times freq_{u,n} + ES_{u,n}) \quad (식 4)$$

$tfidf_{f,n}$: 음식 f의 대표명사목록에 저장된 명사 n의 $tf * idf$ 값
 $freq_{u,n}$: 사용자프로파일 u에 저장된 명사 n의 정규화된 빈도수
 $ES_{u,n}$: 사용자프로파일 u에 저장된 명사 n의 정규화된 감정점수

이 수식은 음식을 구분하는데 도움이 되는, 변별력이 높은 명사일수록(tfidf), 사용자가 트위터에서 자주 사용한 명사일수록(freq), 사용자가 긍정적인 감정을 갖는 명사일수록(ES) 높은 값을 갖게 되며, 또한 u와 f에 중복된 명사(N)가 많을수록 높은 추천점수를 갖게 된다. 이러한 방식으로 특정 사용자의 프로필과 모든 음식의 대표명사 목록간 추천점수를 각각 계산한 후, 상위 5개의 음식을 ①과정에 의한 추천 후보로 잠정 선택한다.

②, ③, ④과정은 사용자가 기 섭취한 음식 정보가 있거나, 사용자와 유사한 사용자가 있거나, 사용자가 기 검색했던 음식 정보가 있는 경우에 한해 진행된다. 각 과정에서는 최대 5개의 음식을 후보로 선정한다. 상황 정보는 사용자의 위치 정보 등을 추출하여 사용한다. ②과정에서는 사용자가 현재와

같은 상황에서 기 섭취한 음식이 있는지를 사용자 프로필에서 검색하여 많이 섭취했던 순으로 최대 5개까지 후보 음식 목록을 선정한다. <표 3>은 ②과정을 위한 알고리즘이다.

②과정 알고리즘

1. 사용자의 현재와 같은 상황(시간, 날씨 등)에서 기 섭취했던 음식의 정보(음식명, 섭취횟수 정보)를 검색
2. 1의 검색 결과가 존재하면, 섭취횟수를 기준으로 내림차순 정렬
3. 상위 음식 5개 선택
4. 선정된 음식에 가중치 0.4 부여

표 3 사용자가 기 섭취한 음식 정보를 이용한 추천 과정

③과정은 협력 기반의 추천 기법을 적용하는 단계로, 사용자와 유사한 음식을 섭취했던 사용자 그룹을 찾고, 그 사용자들이 선택했던 음식 가운데 상위 5개의 후보 음식을 선정한다. 유사 사용자 정보를 이용한 추천 과정은 <표 4>와 같다.

③과정 알고리즘

```
// 1. 사용자 정보 로드
while (시스템에 등록된 사용자 수만큼 반복) {
    각 사용자가 섭취한 음식명, 섭취횟수를 사용자 프로파일에서 검색하여 사용자배열에 저장;
}

// 2. 음식 섭취 횟수의 최대값 검색
while (시스템에 등록된 음식 수만큼 반복) {
    각 음식 섭취횟수의 최대값을 검색하여 저장;
}

// 3. 음식 섭취 횟수 정규화
while (시스템에 등록된 사용자 수만큼 반복) {
    사용자배열에 저장된 섭취횟수를 각 음식 섭취횟수의 최대값으로 나누어 정규화;
}

/* 4. 추천 대상 사용자 i와 나머지 사용자간 유사도를 계산하여 유사 사용자 그룹 생성 */
while (전체사용자 수 - 1) {
    if (추천 대상 사용자 i와 비교 사용자 j가 공통으로 섭취한 음식이 있을 때) {
        (식 5)를 이용하여 피어슨 상관관계수 r 계산;
        계산된 결과를 결과배열에 저장;
    }
    else {
        결과배열에 0 저장;
    }
}

5. 계산된 결과배열을 내림차순으로 정렬;
6. 상위 5명을 선택하여 사용자 i의 유사 사용자 그룹으로 설정;
7. 유사 사용자 그룹이 섭취한 음식을 대상으로 ②과
```

7) 추천 시스템의 모바일 클라이언트를 안드로이드 플랫폼으로 구현하였기 때문에 위치 정보의 인식이 가능하다.

정 알고리즘(1~3)을 적용하여 상위 음식 5개 선택;
8. 선정된 음식에 가중치 0.3 부여;

표 4 유사 사용자 정보를 이용한 추천 과정

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \quad (\text{식 5})$$

X: 추천 대상 사용자 i가 섭취한 음식 정보
Y: 비교 대상자 j가 섭취한 음식 정보
N: 두 사용자가 공통으로 섭취한 음식의 수

(식 5)에서 제시한 피어슨 상관계수 r은 -1 ~ 1의 값을 가지는데 계산 결과가 1에 가까울수록 X와 Y가 유사함을 의미하고, -1에 가까울수록 반대로 연관됨을, 0일 경우는 서로 연관이 없음을 의미한다.

④과정에서는 사용자가 음식을 검색했던 이력 정보를 이용하여, 최대 5개의 후보 음식을 선정한다.

④과정 알고리즘

1. 사용자가 검색했던 모든 음식명과 검색횟수를 검색
2. 검색횟수를 기준으로 내림차순 정렬
3. 상위 음식 5개 선택
4. 선정된 음식에 가중치 0.1 부여

표 5 사용자가 검색한 음식 정보를 이용한 추천 과정

지금까지 설명한 4가지 추천 과정은 ②, ③, ①, ④의 순서로 추천에 유용하다고 판단하여, 각각 다른 가중치를 부여하고 가중치의 합은 1이 되게 하였다. 따라서 ①, ②, ③, ④ 과정의 결과에 각각 0.2, 0.4, 0.3, 0.1의 가중치를 부여하여 모든 결과를 합산한 후, 최종적으로 상위 5개의 음식을 선택하게 된다.

4. 실험 및 분석

4.1 단계별 구현 결과

본 장에서는 3장에서 설명한 리소스 구축 및 음식 추천의 과정을 실제 사례를 들어 설명한다.

• 단계 1: 추천 대상(음식)의 관련정보 수집

<그림 2>의 과정을 통해 수집된 음식 데이터 가운데, 일상생활에서 가장 많이 먹는 음식 54개를 선택하고 각 음식의 대표명사 목록을 구축하였다. <표 5>에 음식 대표명사 목록의 일부를 제시하였다.

비빔밥 보쌈

대표명사	tfidf	대표명사	tfidf
비빔밥	0.204	보쌈	0.194
밥	0.157	돼지고기	0.161
지단	0.098	무	0.085
다시마	0.082	배추	0.078
고사리	0.082	미나리	0.056

국밥

볶음밥

대표명사	tfidf	대표명사	tfidf
국밥	0.897	밥	0.374
밥	0.286	달걀	0.043
콩나물	0.124	새우	0.036
순대	0.075	완두콩	0.036
갈비탕	0.075	부추	0.036

라면

오므라이스

대표명사	tfidf	대표명사	tfidf
라면	0.303	오믈렛	0.440
수프	0.121	밥	0.255
인스턴트식품	0.061	토마토	0.127
건면	0.061	오므라이스	0.107
인스턴트	0.050	달걀	0.043

표 6 음식별 대표명사 목록 일부

• 단계 2: 사용자 프로필 자동 생성

트위터에 실제 등록된 글 가운데 음식 추천을 위한 사용자 프로필 구축에 유용하게 사용될 수 있는 문장의 일부를 <표 7>에 제시하였다. 이 가운데 음식과 관련된 단어가 사용자의 감정을 나타내는 단어와 함께 나타나는 경우가 가장 유용하며, 그냥 음식명만 나타나는 문장도 음식에 관한 사용자의 관심을 나타내는 것이므로 유용한 정보가 된다.

치킨왕을 오븐에 넣을까 하다가 복숭아 하나 꺾어먹은 나는야 이 밤의 승리자..
깊은 침연에서 우리나라는 듯한 향. 돼지고기도 향기가 있다는걸 처음 느낌.
속이 너무 느끼해서 아이스크림 음료수 바로 대량구매 - 행복하네요 ㄱ
울릉도 오징어는 인터넷 쇼핑에서 시켜먹으셈~ 호박엿도 맛있음 ㄱㄱ
시원한 맥주 좋아하시는 분들 보세요~ 맥스 스페셜 호프가 나왔네요!
오늘 먹은 거 두유 한 모가 전부..
오늘 같이 비가 추적추적 오는 날엔 짬뽕이 땡기는구나.. ^^*
말복을 앞두고 수박을 사주셨어요. 역쉬 여름엔 수박이 최고! 내일도 비 소식이 있어서 복날 기분은 못낼 것 같네요.ㄷㄷ
국민트친님! 밤에도 잠 못드는 열대야가 지속되고 있습니다. 여름에 시원한 모카빙수 어떠세요? 경북구역 앞에 있는 '일층카페' 모카빙수 소개해 드립니다.

표 7 사용자 프로파일 생성에 도움이 되는 트위터 문장 예시

<그림 3>의 과정을 통해 자동 구축된 사용자 프로파일의 일부를 <표 8>에 제시하였다. 사용자 A의 경우, 음식명과 감정단어를 자주 함께 사용하였기 때문에 0이 아닌 감정점수를 가진 명사가 상위에 포함되었음을 알 수 있고, 사용자 B의 경우는 감정단어를 자주 사용하지 않아서 상위 10개 명사 목록의 감정점수가 0이 되었음을 미루어 짐작할 수 있다.

사용자 A (총 133개 대표 명사 중 상위 10개)

대표 명사	freq _A	ES _A
커피	1.0	0
밥	0.9	0
밥	0.875	0
물	0.625	0
김	0.475	0
라면	0.425	0.25
우유	0.35	0
맥주	0.35	0
고추	0.225	-1.0
고사리	0.025	0.25

사용자 B (총 50개 대표 명사 중 상위 10개)

대표 명사	freq _B	ES _B
커피	1.0	0
옥수수	1.0	0
밥	0.929	0
돼지고기	0.714	0
물	0.571	0
김	0.5	0
사과	0.357	0
달걀	0.286	0
우유	0.286	0
사태	0.214	0

표 8 구축된 사용자 프로파일 일부

• 단계 3: 음식 추천

단계 1, 2를 거쳐 구축된 음식별 대표명사 목록과 사용자 프로파일을 이용하여, 사용자 A, B를 위한 추천 음식을 ①과정에 따라 실제로 계산하여 선택된 상위 5개의 결과는 <표 9>와 같다.

사용자 A	
RS(A, 국밥)=(0.075×0.025+0) +(0.897×0.125+0.375) +(0.0488×0.025+0) +(0.286×0.9+0) +(0.124×0.025+0.25)	//갈비탕 //국밥 //김치 //밥 //콩나물 =1.001
RS(A, 비빔밥)=(0.082×0.025+0.25) +(0.157×0.9+0) +(0.204×0.025+0.25) +(0.023×0.075+0) +(0.054×0.025+0) +(0.068×0.025+0.25)	//고사리 //밥 //비빔밥 //최고기 //오이 //콩나물 =0.904
RS(A, 커피)=(0.013×0.05+0) +(0.011×0.05+0) +(0.04×0.35+0) +(0.378×1.0+0)	//과일 //얼음 //우유 //커피 =0.393
RS(A, 라면)=(0.027×0.125+0) +(0.303×0.425+0.25) +(0.033×0.25+0)	//국수 //라면 //면 =0.391
RS(A, 스테이크)=(0.024×0.075+0) +(0.019×0.05+0) +(0.037×0.775+0) +(0.031×0.075+0) +(0.031×0.025+0.25) +(0.031×0.275+0)	//감자 //당근 //대구 //최고기 //시금치 //옥수수 =0.293

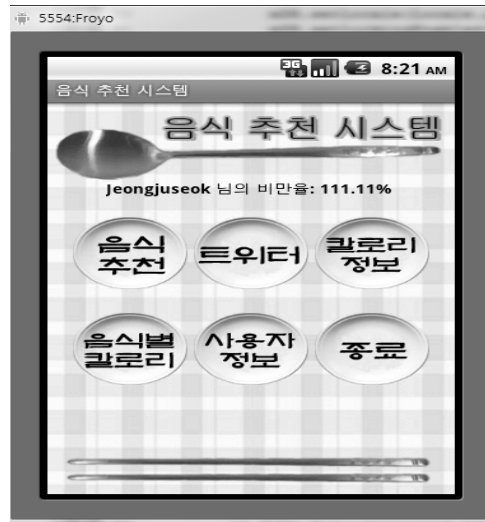
사용자 B	
RS(B, 커피)=(0.013×0.071+0) //과일 +(0.040×0.286+0) //우유 +(0.378×1.0+0) //커피 =0.390	
RS(B, 볶음밥)=(0.043×0.286+0) //달걀 +(0.018×0.071+0) //당근 +(0.374×0.929+0) //밥 +(0.023×0.143+0) //양파 =0.364	
RS(B, 국밥)=(0.286×0.929+0) //밥 =0.266	
RS(B, 오프라이스)=(0.043×0.286+0) //달걀 +(0.255×0.929+0) //밥 +(0.034×0.143+0) //양파 =0.254	
RS(B, 보쌈)=(0.161×0.714+0) //돼지고기 +(0.056×0.071+0) //된장 +(0.085×0.143+0) //무 +(0.012×0.571+0) //물 +(0.078×0.071+0) //배추 +(0.194×0.071+0) //보쌈 +(0.039×0.214+0) //사태 +(0.028×1.0+0) //커피 =0.194	

표 9 사용자별 ①과정 추천 결과

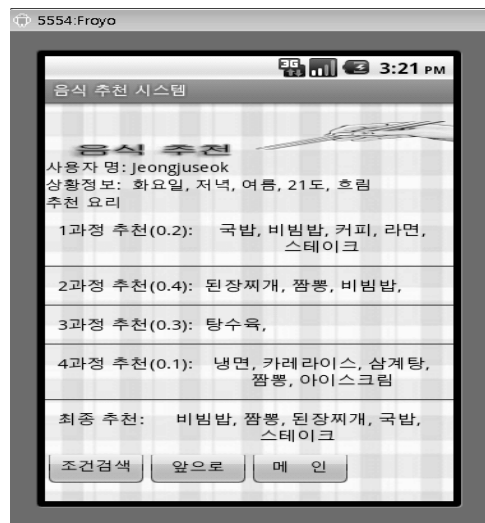
위 결과에 따라 ①과정의 추천 후보 음식이 1차적으로 결정되고, 다음 단계로 ②, ③, ④과정을 거치게 된다. 만약 ②, ③, ④과정에서 필요로 하는 추가 정보가 없는 경우에는 ①과정의 결과가 최종 추천결과가 된다. ②, ③, ④과정의 처리를 위한 추가 정보가 있는 경우에는 <그림 4>와 같이 각 과정의 추천 결과에 가중치를 부여하여 합산한 후, 최종 추천 음식을 결정하게 된다.

4.2 실행 화면

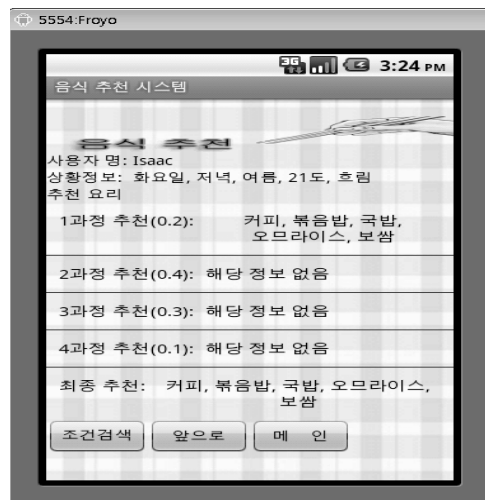
추천 프로그램은 Java와 안드로이드 API로 구현하였고, Tomcat과 Mysql로 웹 서버와 DB 서버를 구축하였다. 추천 시스템의 클라이언트는 상황정보 수집과 이동성, 활용성을 높이기 위해 안드로이드 애플리케이션으로 개발하였다. 추천 기능 외에 트위터 사용 기능과 음식별 칼로리 검색 기능, 사용자별 섭취 칼로리 관리 기능, 상황정보를 직접 입력하여 음식을 검색하는 조건검색 기능도 구현하였다. <그림 5>에 모바일 클라이언트의 초기 실행 화면과 음식 메뉴의 추천 결과 화면을 예시하였는데, 결과 화면 중 각 과정의 추천 결과는 독자의 이해를 돕기 위해 추가로 표시한 정보이며, 실제 사용 시에는 최종 추천 결과만 표시하면 된다. 사용자 A의 경우는 4가지 과정의 모든 결과를 종합하여 최종 추천 결과를 구하였지만, 사용자 B의 경우에는 사용자가 선택하거나 검색했던 이력 정보가 없어서 ①과정의 결과만을 이용하여 추천 결과를 구한 것이다. 이처럼 본 시스템은 어떤 경우에도 추천 결과를 제시할 수 있는 장점이 있다.



초기 실행 화면



사용자 A의 추천 결과 화면



사용자 B의 추천 결과 화면

그림 5 모바일 클라이언트 실행 화면

5. 결론 및 향후 연구과제

최근 유행하는 소셜 네트워크 서비스인 트위터를 이용하여 사용자의 성향을 나타내는 프로파일을 자동으로 구축하고, 이를 기반으로 음식을 추천해 주는 하이브리드 시스템을 제안하였다. 기존의 추천 시스템들은 실행 초기에 데이터 부족 문제로 인하여 정확한 추천을 할 수 없는 문제가 있으나, 본 연구에서는 트위터 계정을 가지고 있다면 사용자가 추천 시스템을 처음 사용한다 할지라도 사용자 프로파일이 자동으로 구축되어 추천에 사용되기 때문에 항상 추천 결과를 얻을 수 있었다. 또한 추천 시스템을 사용하면 할수록 관련 데이터가 누적되어 내용 기반 및 협력 기반을 결합한 하이브리드 형태로 활용되기 때문에 보다 정확한 추천이 가능하게 된다.

추천 문제는 사용자 자신도 원하는 답이 무엇인지 모르는 경우가 많기 때문에 정확도와 같은 척도로 시스템의 결과를 평가하기는 쉽지 않다. 다만 사용자의 만족도를 대안으로 사용할 수는 있겠다. 향후에는 트위터 뿐만 아니라 블로그나, 미투데이, 싸이월드 같은 시스템의 Open API를 활용하여 보다 정확한 사용자 프로파일을 구축하는 방법과 이를 개인화 시맨틱 정보검색에 활용하는 방법에 대해 연구하고자 한다.

참 고 문 헌

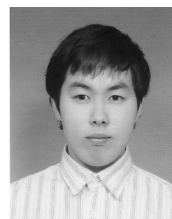
- [1] 김수동, 라현정, “안드로이드 기반 모바일 서비스 애플리케이션의 아키텍처”, *한국정보과학회지*, vol. 28, no. 6, pp.25-34, 2010.
- [2] G. Adomavicius, and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, June 2005.
- [3] 토비 세가란, *집단지성 프로그래밍*, O'Reilly, 2008.
- [4] 한현구, 서의현, “학습기반 개인 맞춤형 음식 추천 에이전트”, *한국정보처리학회 논문지*, vol. 16, no. 2, pp.313-314, 2009.
- [5] 김성기, 김영학, “항목기반 패턴을 사용한 학습 방법 추천 시스템의 설계 및 평가”, *한국콘텐츠학회 논문지*, vol. 9, no. 5, pp.346-354, 2009.
- [6] S. K. Ray and S. Singh, “Blog Content Based Recommendation Framework Using WordNet and Multiple Ontologies”, *Proceedings of the International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, pp.432-437, 2010.
- [7] 이재식, 이진천, “상황 인식을 이용한 사례기반 음악추천시스템”, *한국지능정보시스템학회 논문지*, vol. 12, no. 3, pp.111-123, 2006.
- [8] 오유수, 최아영, 우운택, “차세대 건강관리 밥상 시스템: 맥락 인식 기반 개인화된 음식 추천 시스템”, *한국정보처리학회 논문지*, vol. 15, no. 1, pp.51-61, 2008.
- [9] 이재식, 박석두, “장르별 협업필터링을 이용한 영화 추천시스템의 성능 향상”, *한국지능정보시스템학회 논문지*, vol. 13, no. 4, pp.65-78, 2007.
- [10] Y. Ying, “A Personalized Collaborative Recommendation Algorithm based on User Clustering Smoothing”, *Proceedings of the International Conference on Future*

- BioMedical Information Engineering*, pp.383-385, 2009.
- [11] 김기수, “개인화된 추천을 위한 하이브리드 협업 필터링 알고리즘 및 인터넷 전자상거래에서의 활용”, *인터넷전자상거래연구*, vol.8, no.4, pp.1-20, 2008.
- [12] C. Liu, C. Sun, J. Yu, “The Design of an Open Hybrid Recommendation System for Mobile Commerce”, *Proceedings of the International Conference on Computational Intelligence for Modelling Control & Automation (CIMCA)*, pp.861-866, 2008.
- [13] 이홍식, “음식 명사의 사전 기술에 대하여”, *한국어학회지*, vol. 38, no. 1, pp.307-344, 2008.
- [14] K. Makeie, *Twitter API: Up and Running: Learn How to Build Applications with the Twitter API*, O'Reilly, 2009.
- [15] 박인조, 민경환, “한국어 감성단어의 목록 작성과 차원 탐색”, *한국심리학회지*, vol. 19, no. 1, pp.109-129, 2005

저 자 소 개

강신재 (Sin-Jae Kang)

한국지능시스템학회 논문지, 제 21권 제 2호 참조



정주석 (Ju-Seok Jeong)

2010년 : 대구대학교 컴퓨터IT공학부
공학사

2011년~현재 : 대구대학교 대학원
컴퓨터정보공학과 석사과정

관심분야 : 자연어처리, 정보검색,
시맨틱 웹

Phone : 053-850-4464

E-mail : jsjeong626@gmail.com