

## 오피니언마이닝을 이용한 사용자 맞춤 장소 추천 시스템

최은정<sup>1</sup> · 김동근<sup>2\*</sup>

### Location Recommendation Customize System Using Opinion Mining

Eun-jeong Choi<sup>1</sup> · Dong-keun kim<sup>2\*</sup>

<sup>1</sup>Department of Computer Science, Sangmyung University, Seoul 03016, Korea

<sup>2</sup>Department of Intelligent Engineering Informations for Human, Sangmyung University, Seoul 03016, Korea

#### 요 약

최근 빅데이터 분야의 높아진 관심과 더불어 빅데이터의 처리를 통한 응용 분야에 대한 관심도 높아지고 있다. 개인의 감성을 파악할 수 있는 오피니언마이닝은 사용자 개인 맞춤 서비스 제공 분야에서 많이 이용되고 있는 빅데이터 처리 기법이다. 이를 바탕으로 본 논문에서는 사용자들의 장소에 대한 텍스트 형태의 리뷰를 오피니언마이닝 기법으로 처리하고 k-means 클러스터링 작업을 통해 사용자의 감성을 분석하였다. 클러스터링 작업으로 분류된 비슷한 범주의 감성을 가진 사용자들끼리 동일한 수치 값을 부여한다. 부여된 수치 값으로 협업 필터링 추천 시스템을 이용해 선호도를 예측하고 예측 값이 높은 장소 순으로 지도위에 마커와 함께 내용을 표시하여 사용자에게 추천내용을 보여줄 수 있는 방안을 제안하였다.

#### ABSTRACT

Lately, In addition to the increased interest in the big data field, there is also a growing interest in application fields through the processing of big data. Opinion Mining is a big data processing technique that is widely used in providing personalized service to users. Based on this, in this paper, textual review of users' places is processed by Opinion mining technique and the sentiment of users was analyzed through k-means clustering. The same numerical value is given to users who have a similar category of sentiment classified as a clustering operation. We propose a method to show recommendation contents to users by predicting preference using collaborative filtering recommendation system with assigned numerical values and marking contents with markers on the map in order of places with high predicted value.

**키워드** : 오피니언 마이닝, 감성 분석, 협업 필터링, 추천 시스템

**Key word** : Collaborative filtering, Opinion mining, Recommender system, Sentiment analysis

Received 07 June 2017, Revised 22 June 2017, Accepted 12 October 2017

\* Corresponding Author Dong Keun Kim(E-mail:dkim@smu.ac.kr, Tel:+82-2-2287-5431)

Department of Intelligent Engineering Informations for Human, Sangmyung University, Seoul 03061, Korea

Open Access <https://doi.org/10.6109/jkiice.2017.21.11.2043>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

최근 유무선 인터넷 인프라의 보급과 스마트폰의 등장으로 인터넷 사용이 급속히 늘어나면서 많은 사람들이 커뮤니티, 블로그, 트위터, 등 기타 다양한 소셜 미디어를 통하여 자신의 견해나 의견을 표현하고 있다[1].

이를 통한 사용자들 간의 의견 및 정보 교환은 더욱 활발해지고 있고 이런 정보들을 추천 시스템에 이용하면 개개인의 취향과 성향에 맞는 추천을 할 수 있다[1]. 또한 사용자들이 제공하는 정보의 품질 개선이나 신뢰도에도 큰 기여를 할 수 있다[1]. 또한 사용자들의 감성 분석은 추후 사용자들의 행동에 대해서도 예측 가능할 수 있게 해준다[2].

본 논문에서는 R을 이용하여 장소 데이터 중 음식점 리뷰 데이터를 수집하고 수집한 데이터의 오피니언마이닝과 클러스터링 작업을 진행하여 사용자가 원하는 적절한 장소의 추천 시스템에 대하여 설계하고자 한다. 장소 추천 시스템의 정확도와 사용자 만족도 향상을 위해 사용자의 감성 어휘를 총 4가지 영역에서 분류하였다. 이러한 방식으로 분류된 각 범주의 감성언어들을 바탕으로 사용자가 원하는 맞춤 장소 추천 서비스에 이용할 수 있는 방안의 설계를 목적으로 하고 있다. 본 연구에서는 클러스터링 작업의 수행을 통해 각 군집을 묶을 수 있는 기준 값을 이용하여 사용자들 사이의 감성 유사도를 계산하고 사용자가 원하는 알맞은 장소를 추천할 수 있는 서비스를 개발 하고자 한다. 장소 추천 서비스는 분석된 감성 정보를 이용하여 사용자와 비슷한 감성을 가진 다른 사용자의 선호도가 높았던 장소들을 추천할 수 있다. 따라서 기존의 장소 추천 기법보다 더 만족스러운 결과를 얻을 수 있도록 하였다.

## II. 관련 연구

### 2.1. 감성 분석 및 감성 사전 구축

감성분석(sentiment analysis)은 텍스트마이닝 분석의 한 분야로 특정 문서의 긍정, 부정, 중립 등에 대한 감정을 추측하고 분류한다[3]. 각 문서의 최소단위인 단어의 감성극성(sentimentpolarity)에 기반으로 감성분석이 이루어진다[3].

텍스트에서 정확한 오피니언을 도출하기 위해서는

수집된 언어 자원과 그에 해당하는 감성분류, 즉 감성 사전이라고 하는 어휘 집합이 중요한 요소로 적용된다 [4]. 감성 어휘의 감성이 정의되어 있다면 텍스트의 감성판별 정확도는 높아지지만 언어의 특성에서 발생하는 다양한 의미 변이와 동음이의어 등 텍스트의 감성을 판별하기 위해서는 여러가지 고려사항이 필요하다 [4]. 이와 같은 문제를 해결하기 위해 리뷰에 사용된 언어의 형태소를 분석하여 긍정, 부정 및 중립적 의미를 얼마나 가지고 있는지를 분류하는 극성 분류 작업을 수행하고, 분류된 형태소들을 기반으로 감성사전을 구축한다[5].

극성을 분류하는 방식은 PMI 방식과 이를 개선한 SO-PMI 방법이 있는데 PMI방식은 기준 단어를 선택하여 이를 다른 단어와 특정 문서 내에서 출현할 확률을 구하여 극성을 분류하는 방법이다[5]. PMI 방식에서 비슷한 의미의 극성을 가지면 확률 값은 양수를 나타내고, 다른 의미의 극성을 가지면 확률 값은 음수를 나타낸다 [4]. SO-PMI 방법은 미리 긍정 어휘, 부정 어휘들 및 중립 어휘를 정의하고 어휘의 극성을 분류하는 방법이다. SO-PMI 값이 음수가 나오면 부정으로, 양수가 나오면 긍정으로 분류한다[5].

본 연구에서는 SO-PMI 방식을 통해 리뷰의 긍정 단어, 부정 단어를 파악하고 각 사용자별 리뷰의 긍정적 의미, 부정적 의미의 단어 사용 개수를 파악하여 리뷰의 긍정도, 부정도, 중립도를 계산하였다. 이러한 계산을 바탕으로 사용자의 리뷰를 추천 시스템에 적용할 수 있는 새로운 방안을 연구하였다.

### 2.2. R을 이용한 오피니언 마이닝

오피니언 마이닝은 감성 분석을 위한 빅데이터 기법으로 감성 사전을 통한 단어들 간의 극성을 수치화하여 극성도 즉, 부정도와 긍정도를 측정하는 연구 방법이며 많은 연구자들이 이를 활용하고 있다. 리뷰어를 잘 정리하면 좋은 극성 사전이 만들어지고, 잘 만들어진 극성사전은 오피니언 마이닝에서 중심적 역할을 한다[6].

R은 통계, 해석 기법을 위해 이용하는 프로그래밍 언어 중 하나로 R을 통해 텍스트 분석을 진행 하며 오피니언마이닝의 감성사전의 설계를 비롯한 앞에서 언급한 일련의 과정들을 가능하게 해준다.

오피니언의 정확도는 감성사전의 구축의 방식을 통하여 차이를 보일 수 있고 해외의 다양한 감성분석과

데이터마이닝은 영어 문법 기반의 텍스트마이닝을 위주로 분석을 수행하고 있다[7]. 이에 따라 한국어 문법에서는 영어문법과 비교하여 다양한 접속어 어휘들이 등장하고, 이로 인하여 문장 전체가 긍정적인 구조일지라도 부정 어휘의 쓰임에 의해서 전혀 상반된 결과를 나타내기도 한다[7].

본 연구에서는 오피니언 마이닝을 통한 감성분석에서 더 나아가 감성을 클러스터링을 통해 분류 하고 군집화 하여 사용자들의 감성 군집을 만들었다. 또한 감성분석을 통해 만들어진 군집들을 통해 추천시스템에 적용할 수 있는 새로운 방안을 고려하였다.

### 2.3. 협업 필터링 추천 시스템

인터넷과 컴퓨터의 발달로 온라인을 통하여 활용할 수 있는 정보들이 다양해짐에 따라, 사용자에게 불필요한 내용을 제거하고 필요한 정보만을 제공해주는 정보 필터링 기법이 제안되며 추천 시스템(recommendation system)과 같은 응용 프로그램에 대한 관심이 집중되고 있다[8].

추천 시스템은 사용자의 선호와 기호에 알맞게 필터링하여 선호도가 높을 거 같은 콘텐츠를 제안하는 시스템이다. 추천 시스템은 크게 협업 필터링 추천방법과 내용 기반 추천 방법으로 나뉜다[9]. 협업 필터링은 특정 아이템에 대해 유사한 성향을 보였던 사용자들은 다른 아이템에 대해서도 유사한 성향을 보일 것이라는 가정을 바탕으로 추천을 수행한다[10]. 사용자 기반의 협업적 필터링은 최초의 자동화된 협업적 필터링 방법이며, 사용자-사용자간 협업적 필터링 또는 k-NN 협업적 필터링으로도 알려져 있다[11]. 사용자 기반의 협업적 필터링은 특정 사용자와 유사한 평가를 한 사용자를 찾고, 사용자들의 평점을 현재 사용자가 만족할 만한 아이템을 예측한다[11]. 사용자 A가 평점을 매기지 않은 아이템에 대한 선호도를 예측할 때, 사용자 기반의 협업적 필터링에선 사용자 A와 높은 경향성을 가진 사용자들에 주목한다[11].

본 연구에서는 감성분석을 통한 감성 군집의 결과를 이용하여 협업 필터링 추천 시스템에 적용할 수 있는 방안을 연구하였다. 사용자의 감성을 분석하여 기존의 협업필터링 추천 시스템에 적용하였다는 측면에서 사용자에 대한 만족도를 더욱 높일 수 있도록 하였다.

## III. 사용자 맞춤 장소 추천 시스템 설계

### 3.1. 웹기반 오피니언마이닝

본 장에서는 웹 기반 오피니언마이닝 시스템의 설계에 대하여 설명한다. 웹으로부터 오피니언마이닝에 필요한 데이터들을 수집하고 관리하는 ‘리뷰 수집 모듈’, 수집된 데이터에서 형태소 분석을 진행하고 감성 어휘를 추출하는 ‘데이터 분석 모듈’, 분석된 데이터로 사용자의 감성과 감성사전 어휘 사이의 일치도를 높일 수 있도록 클러스터링 작업을 진행하는 ‘클러스터링 모듈’ 총 3단계로 구성된다.

그림 1은 앞에서 설명한 웹 기반 오피니언마이닝 시스템의 구성도이다. 웹 서버를 이용해 리뷰 수집 모듈에서 크롤링을 진행한다. 크롤링을 통해 받은 리뷰데이터는 형태소 분석과 감성 사전을 통한 감성 분석을 진행한다. 감성 분석을 거친 데이터는 클러스터링 작업을 진행하여 사용자의 감성을 군집화 하였다.

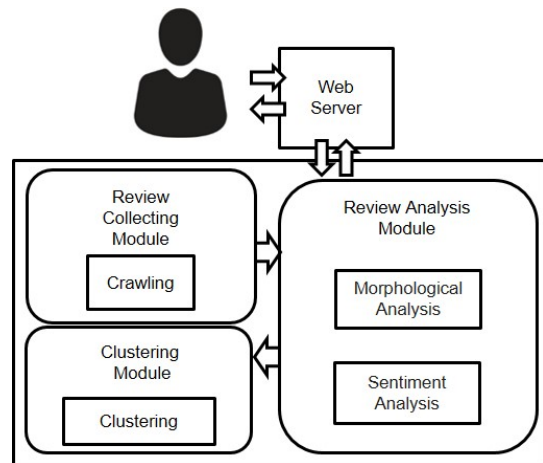


Fig. 1 Composition of Web-based Opinion Mining System

#### 3.1.1. 데이터 수집 및 전처리 시스템

데이터 수집 단계에서는 오피니언마이닝을 위한 사용자의 의견을 텍스트의 형태로 수집한다. 파이썬을 이용한 크롤링 시스템을 구축하여 웹을 통한 실시간 데이터의 수집이 가능하도록 한다. 수집하고자 하는 데이터가 있는 URL을 입력하면 해당 웹사이트의 크롤링이 가능하다. 사용자들의 의견 파악의 정확도를 높이기 위해 사용자들이 입력한 평점 지수와 리뷰 문장이 동시에

수집 가능하도록 한다. 웹을 통해 수집한 텍스트 형태의 데이터를 R의 Konlp 라이브러리 패키지를 통해 R-Studio에서 형태소 분석을 진행하였다[12-14].

장소 리뷰의 경우 웹 사이트 “망고플레이트”에서 제공하는 각 지역별 음식점 리뷰를 이용하였다[14]. 여러 지역의 장소리뷰 중 서촌과 북촌 지역의 음식점들의 리뷰를 데이터 자료로 이용하였다. 각 음식점 리뷰는 2016년 10월부터 2017년 5월 현재까지의 리뷰들 중 각 음식점 별 25개 내외의 리뷰를 데이터로 이용하였다.

```

>
>
>
> library(KonLP)
> library(wordcloud)
> useSejongDic()
Backup was just finished!
87007 words were added to dic_user.txt.
> library(RColorBrewer)
> txt = readLines("mango.txt")
> txt<-sapply(review,extractNoun,USE.NAMES = F)
> txt2<- unlist(txt)
> txt2<-gsub("\\d+", "",txt2)
> txt2<-gsub("\\|", "",txt2)
> txt2<-gsub("\\|", "",txt2)
> txt2<-gsub("\\|", "",txt2)
> txt2<-gsub("\\|", "",txt2)
> txt2<-gsub("\\|", "",txt2)
> txt2<-gsub("\\|", "",txt2)
> txt2<-gsub("[A-Za-z]", "",txt2)
> txt2<-gsub("\\|", "",txt2)
> write(txt2,"review2.txt")
> review2=read.table("review2.txt")
warning message:
In scan(file, what, nmax, sep, dec, quote, skip, nlines, na
.strings, :
  EOF within quoted string
> nrow(review2)
[1] 1546
> wordcount = table(review2)
> wc<-head(sort(wordcount,decreasing = T),100)
> palette = brewer.pal(9,"set1")
> wordcloud(names(wordcount),freq=wc,scale=c(5,1),rot.per=0
.25,min.freq=2,random.order=F,random.color=T,colors=palette)
There were 50 or more warnings (use warnings() to see the f
irst 50)
>
> |
    
```

Fig. 2 R-Studio console

그림 2는 서촌의 음식점 중 특정 음식점 하나의 리뷰를 크롤링하여 텍스트 파일로 저장한 자료를 R프로그램을 이용하여 오피니언 마이닝을 진행하였다. 형태소 분석을 통한 명사 추출과 추출된 명사의 빈도수 순으로 100개의 단어를 워드클라우드의 형태로 나타내기 위한 과정을 보여주는 R의 콘솔화면이다.

### 3.1.2. R을 이용한 데이터 분석 시스템

수집되어 형태소 분석이 된 텍스트 데이터의 감성 어휘 추출을 위해 데이터 분석 과정을 진행한다. 한글 데이터의 분석을 용이하게 할 수 있도록 텍스트를 숫자로 표현하는 TermDocumentMatrix 의 형태로 지정하여 형태소 분석이 진행된 어휘들을 긍정과 부정, 중립의 세 가지 범주로 분류 한다[15]. 긍정의 정도와 부정의 정도, 중립의 정도를 통하여 한글의 특성상 동일한 어휘의 쓰임에도 전체 문장은 완전히 다른 의미를 보일 수 있기 때문에 형태소 분석이 실시된 각 단어들 사이의 연관도 검사를 수행 한다. 이는 미리 구축해 놓은 감성사전을 이용한 실제 장소 리뷰 데이터의 긍정과 부정정도의 수치계산과 제시된 주제어의 연관도 수행의 결과가 어느 정도의 관련을 보이는지 계산하여 긍정적 의미의 단어가 부정적 의미로 쓰였는지, 반대로 부정적 의미의 단어가 긍정적 감정을 나타내기 위해 쓰였는지의 확인과 더불어 미리 구축된 감성 사전의 수정 필요 여부를 알아보기 위하여 진행하였다. 이와 같은 작업을 통해 리뷰 데이터 분석에 대한 정확성을 더 높였다.

그림 3은 연관도 분석을 통한 리뷰 데이터의 분석 과정을 보여주는 순서도이다.

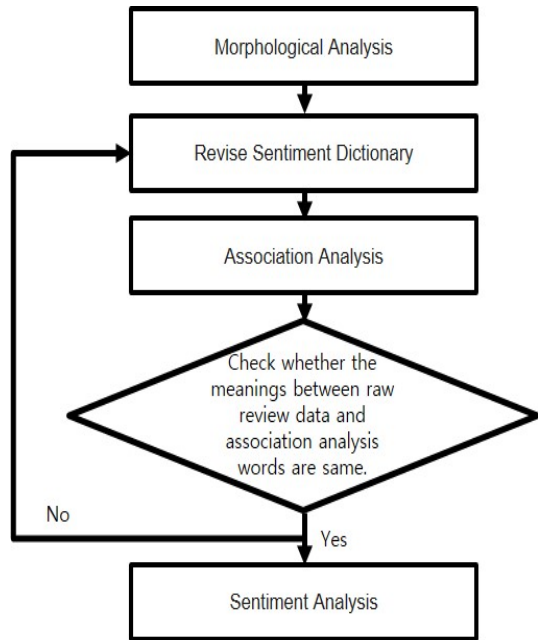


Fig. 3 Data sentiment analysis flowchart

### 3.2. 장소 추천 시스템

사용자 오피니언마이닝을 이용한 협업 필터링 기반 장소 추천 시스템은 분석된 사용자 감성정보를 클러스터링으로 군집화 시킨 후 각 군집들의 수치 값을 추천 모듈을 통해 유사도를 계산하도록 적용하였다. 사용자의 리뷰를 통해 리뷰에 쓰인 각 군집 간 단어들의 각각의 수치 값의 평균치를 사용자에게 부여하고 사용자에게 부여된 수치 값은 다른 사용자들 간의 감성 유사도를 비교하는 계산을 위해 사용된다.

**Table. 1** Example of restaurants list

No.	Name	Location	Category
1.	Se*****	**, Jahamun-ro 16-gil, Jongno-gu, Seoul, Republic of Korea	Italian Food
2.	Go***	**, Jahamun-ro 12-gil, Jongno-gu, Seoul, Republic of Korea	Burger
3.	To*****	**, Jahamun-ro, Jongno-gu, Seoul, Republic of Korea	Desert Cafe
4.	Gal*****	**, Jahamun-ro 13-gil, Jongno-gu, Seoul, Republic of Korea	Italian Food
5.	Tap*****	**, Jahamun-ro 9-gil, Jongno-gu, Seoul, Republic of Korea	Spanish Food
6.	Fo***	**, Jahamun-ro 9-gil, Jongno-gu, Seoul, Republic of Korea	Chinese Food
7.	Go*****	**, Jahamun-ro, Jongno-gu, Seoul, Republic of Korea	Noodle
8.	Og*****	**, Ogin-gil, Jongno-gu, Seoul, Republic of Korea	Pizza
9.	O***	**, Ogin-gil, Jongno-gu, Seoul, Republic of Korea	Korean Food
10.	for*****	**, Ogin 3-gil, Jongno-gu, Seoul, Republic of Korea	Japanese Food

장소를 지역별 음식점으로 한정하여 보여 주기 때문에 본 연구에서는 웹 사이트 “망고 플레이트”에서 제공하는 서촌과 북촌의 음식점 리스트를 이용하였다[16]. 음식점에 대한 리스트를 만들어서 각 장소에 대한 사용

자들의 선호도를 예측한다.

표 1 은 본 연구를 위해 사용한 서촌 지역의 음식점들 리스트 중 일부이다. 상호명의 직접적 언급을 피하기 위해 음식점의 정보 중 일부는 공개하지 않았다. 사용자들은 다음의 음식점 리스트들 내의 음식점 중에서 일부의 음식점에 대한 추천을 받고 추천 받은 음식점에 대한 위치나 전화번호 등과 같은 간단한 정보를 지도 위에서 받을 수 있다.

#### 3.2.1. 감성 유사도 계산

사용자와 다른 사용자 사이의 감성 유사도를 계산할 때는 Pearson-Correlation을 이용한다. 아래 식은 사용자 간의 유사도를 계산하기 위한 Pearson-Correlation 식이다.

$$R_{xy} = \frac{\sum_{i=1}^m (R_{xi} - \bar{R}_x) * (R_{yi} - \bar{R}_y)}{\sigma_x * \sigma_y} \quad (1)$$

$R_{xy}$ 는 사용자 x와 사용자 y가 공통 장소에 대하여 느낀 감성 정보 값을 바탕으로 계산한 유사도를 나타낸다 [5]. i는 m개의 장소 중 하나의 장소이고,  $R_{xi}$ 은 특정 사용자 x가 장소 i에 대하여 느낀 감성 값이고,  $R_{yi}$ 은 특정 사용자 y가 장소 i에 대하여 느낀 감성 값이다[9].  $\bar{R}_x$ 와  $\bar{R}_y$ 는 특정 사용자 x와 사용자 y가 모든 장소에 대해 느낀 감성 값의 평균이다[9]. 분모는 사용자 x와 사용자 y가 모든 장소에 대해 느낀 감성 값의 평균의 표준 편차이다[6]. 계산에 의해서 나온 값은 -1에서 1까지의 값을 가지는데 1에 가까울수록 높은 유사도를 보이고 -1에 가까울수록 낮은 유사도를 보인다[17].

#### 3.2.2. 사용자 감성 장소 선호도

예측 사용자에게 추천되어질 장소에 대하여 사용자의 선호도를 예측하는 식은 (2)와 같다.

$P_{xj}$ 는 감성 유사도를 토대로 새롭게 정리된 장소 리스트 k개의 장소 중에서 j번째 장소에 대하여 사용자 x의 선호도 예측 값을 나타낸다[9].  $\bar{R}_x$ 와  $\bar{R}_y$ 은 사용자 x와 y가 모든 장소에 대해 느낀 감성 값의 평균이고,  $R_{yj}$ 은 특정 사용자 y가 장소 j에 대하여 느낀 감성 값이다[6].  $\sum_{i=1}^k R_{xy}$ 은 모든 사용자들의 감성을 두 명씩 비

교한 감성유사도 값의 합계를 의미한다.

$$P_{x,j} = \overline{R_x} + \frac{\sum_{j=1}^k (R_{yj} - \overline{R_y}) * R_{xy}}{\sum_{j=1}^k R_{xy}} \quad (2)$$

$R_{xy}$ 는 공통 장소에 대하여 사용자 x와 사용자 y가 느낀 감성 값이다[9]. 사용자 본인의 감성 값은 항상 1.0을 기본으로 하고 있으며 감성 유사도 값이 높은 다른 사람들의 예측된 장소 결과 값을 통해 예측도가 높은 순서대로 장소를 사용자에게 추천한다. 추천 장소는 사용자 인터페이스를 통해 나타낼 수 있는 방안을 고려하였다.

#### IV. 결 과

##### 4.1. 클러스터링 시스템

연관도 검사의 결과를 바탕으로 사용자의 감성을 총 3번의 K-means 클러스터링으로 분석하였다. 하나의 리뷰에 대한 총 3번의 클러스터링 작업으로 각각 3가지, 4가지, 5가지의 범주로 분류하여 결과들을 서로 비교, 분석하여 클러스터링을 통한 분류가 가장 잘 된 범주의 개수를 파악하고자 하였다.

표 2,3,4는 앞서 제시하였던 장소 리뷰의 데이터를 이용한 3가지, 4가지, 5가지의 범주의 클러스터링 결과이다.

Table. 2 Clustering Words (K=3)

Sections	Sentiment Words
clustering 1	taste, many, a lot
clustering 2	kind, very, owner
clustering 3	amount, proper, mood

표 2의 내용을 보면 clustering 1, clustering 2는 각각 맛과 서비스에 대한 긍정적 의미를, clustering 3은 중립적 의미를 나타내고 있다. 이와 같은 분류로 볼 때 대부분의 리뷰데이터가 매우 긍정적인 의미를 나타내고 있음을 알 수 있고 3개의 범주로 진행된 클러스터링은 긍정적 의미, 부정적 의미, 중립적 의미의 측면에서 볼

때 만족스럽지 못한 결과를 보임을 알 수 있다.

4개의 범주로 클러스터링을 진행한 표 3의 내용을 보면 clustering 1은 맛에 대한 긍정적 인식을 나타내고, clustering 2는 음식점 서비스에 대한 중립적 의미의 단어들을 포함하고 있다. clustering 3은 음식점의 직원들 및 사장에 대한 긍정적 평가를 보여준다. 앞서 설명한 3개의 군집과는 다르게 clustering 4는 음식점에 대한 아쉬운 점과 부정적 단어들을 중심으로 이루어져 있음을 알 수 있다.

Table. 3 Clustering Words (K=4)

Sections	Sentiment Words
clustering 1	taste, so, perfect
clustering 2	basic, food, amount
clustering 3	kind, very, owner
clustering 4	nothing special, inconvenience, parking

다음으로 5개의 범주로 클러스터링을 진행한 표 4의 clustering 1과 clustering 3, clustering 4는 각각 맛, 서비스, 음식에 대한 긍정적 의미를 나타내고, clustering 2는 서비스와 메뉴에 대한 중립적 의미를 나타내고 있다. clustering 5는 표 3과 같은 부정적 의미의 결과를 나타낸다.

Table. 4 Clustering Words (K=5)

Sections	Sentiment Words
clustering 1	taste, so, perfect
clustering 2	service, basic, menu
clustering 3	many, people, kind
clustering 4	temperature, amount, good
clustering 5	not good, parking, inconvenience

실제 소비자들이 남기는 리뷰의 특성상 긍정적 의미가 다수를 차지하므로 클러스터링의 범주의 개수가 증가할수록 다수를 차지하는 긍정적 의미의 범주가 계속 증가하는 경향을 보였다.

표 2, 3, 4의 클러스터링 결과의 각 결과값을 비교해 볼 때 표 3의 4개의 범주를 이용한 클러스터링 결과가 긍정적 의미, 부정적 의미, 중립적 의미의 측면에서 볼 때 가장 알맞은 결과를 나타냄을 알 수 있다. 이를 통해 본 연구에서는 클러스터링을 통해 묶인 군집들에 일정

한 값을 부여하여 오피니언마이닝을 통한 감성분석을 실시하였다. 또한 이를 바탕으로 각 군집에 해당하는 감성을 지닌 사용자들을 다시 분류하여 장소 추천 시스템의 감성유사도 검사에 이용하였다. 각 군집은 4개의 군집으로 분류하였기 때문에 각 군집에 해당하는 적절한 value 값을 부여하여 감성 유사도를 계산 할 수 있도록 하였다. value 값은 본 연구에서 사용된 전체 사용자들의 음식점 리뷰의 감성사전 단어들의 극성 즉, 긍정도, 부정도, 중립도를 각각 백분율로 계산하였다. 그 결과 긍정도는 72%, 중립도는 19%, 부정도는 9%에 해당하는 결과를 나타냈고 이를 척도로 하여 각 군집에 해당하는 value 값을 나타낸 클러스터링 결과는 표 5와 같다.

표 5의 value 값은 사용자의 감성 유사도 계산에 이용되며 이 값을 바탕으로 사용자와 감성이 유사한 다른 사용자들이 선호하는 장소들이 사용자에게도 추천되어 질 수 있도록 하였다.

본 연구에서 이용한 음식점 리뷰 데이터의 각 사용자들의 리뷰에 다시 오피니언 마이닝을 진행하여 표5의 4개의 군집들 중 가장 높은 상관도를 보이는 군집들을 계산하여 사용자들에게 각 군집에 해당하는 value 값을 부여하였다.

**Table. 5** Clustering Words (K=4) with Measurement Values

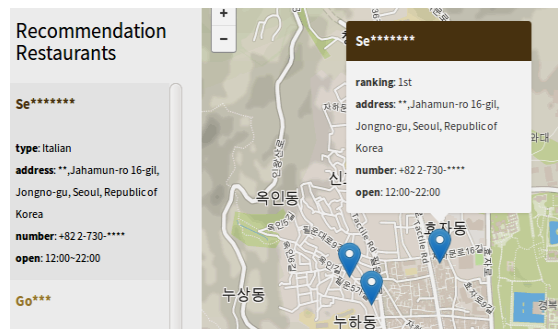
Sections	Sentiment Words
clustering 1	taste, so, perfect
clustering 2	basic, food, amount
clustering 3	kind, very, owner
clustering 4	nothing special , inconvenience, parking

**4.2. 장소 추천 결과**

선호도 예측을 통한 장소추천의 내용을 지도를 통해 표현하였다. 지도의 마커에는 추천 순위와 함께 날짜 및 위치가 같이 표시되도록 하였다. 장소 추천의 내용은 사전에 구축된 서촌과 북촌의 음식점 리스트를 이용하여 각 음식점 리뷰를 오피니언 마이닝을 통한 감성 분류를 실시하고 클러스터링을 통해 군집으로 묶은 뒤 해당 군집과 사용자의 감성 유사도 값이 일치하거나 비슷한 구간을 찾아 리스트내의 장소들 중 선호도 순으로 사용자의 화면에 보여준다. 같은 감성 군

집내의 장소 목록들만을 결과화면에 보여줌으로써 사용자는 좀 더 높은 만족도의 결과를 얻을 수 있다. 본 논문에서는 특정 지역의 음식점 정보를 이용한 장소 추천을 진행을 위해 특정 지역의 음식점 25개를 선정하여 리스트로 정리하고 각 음식점의 리뷰를 20개씩 크롤링 하였다.

그림 4는 표 2 의 서촌 지역 음식점 리스트를 중심으로 감성분석을 통한 장소 추천 시스템의 사용자 결과 화면이다.



**Fig. 4** Example of Location Recommendation results

**V. 결론**

본 논문에서는 기존의 다른 추천 시스템에서 사용자의 만족도나 감성 파악의 측면에서 미흡했던 부분에 대한 해결을 위해 사용자 리뷰를 이용한 감성 어휘 추출과 감성 분석을 통한 시스템을 설계하고 분석된 감성들을 처리할 수 있는 시스템 설계로 보완하였다는 측면에 의의를 둔다.

아울러, 본 논문에서는 기존의 감성사전을 활용한 오피니언마이닝을 진행하였다. 향후 연구에서는 본 논문에서 제시한 오피니언마이닝을 통한 데이터의 분석과 장소 추천 시스템 기술에 이용할 수 있는 독자적 감성사전 연구와 감성어의 이산적의미를 표현하기 위한 이측기반의 표현법을 위한 연구가 행해져야 할 것으로 보인다. 또한 모바일 기기에서도 적용해 볼 수 있는 어플리케이션 설계를 위한 연구도 필요할 것으로 보여진다.

더 나아가 현재의 시스템의 경우 특정 지역의 음식점에만 한정하여 정보를 제공하였지만 나아가 전국의 음



식점 뿐 아니라 다른 여행지나 특정 장소에 대한 정보 제공도 이루어 질 수 있도록 응용작업이 필요할 것으로 보인다.

## ACKNOWLEDGMENTS

This study is a study consisting year received support school funding Sangmyung 2017, thanks to the ministries.

## REFERENCES

- [ 1 ] X. Han, "For social information referral techniques Big Data Model," M.S. Theses, Ewha Womans University, Seoul, 2012.
- [ 2 ] Ankit Gurura, "Mining User-Aware Uncommon Consecutive TopicPatterns in Report Streams," *Asia-pacific Journal of Convergent Research Interchange*, vol. 2, no.4, pp. 15-21, December 2016.
- [ 3 ] S.H. Lee, J. Choi, J. W. Kim, "Sentiment analysis on movie review through building modified sentiment dictionary by movie genre," *Journal of Intelligence and Information Systems*, vol. 22, no. 2, pp. 97-113, June 2016.
- [ 4 ] J. H. Seo, H. J. Jo, J. T. Choi, "Design for Opinion Dictionary of Emotion Applying Rules for Antonym of the Korean Grammar," *Journal of Korean Institute of Information Technology*, vol. 13, no. 2, pp. 109-117, Feb. 2015.
- [ 5 ] S. E. Kim, E. K. Kim, Y. G. Kim, "Cosmetic Recommendation System using Fuzzy Inference and Building Sentiment Dictionary," *Journal of Korean Institute of Intelligent Systems*, vol. 27, no. 3, pp. 253-260, June 2017.
- [ 6 ] J. H. Seo, "Design of Opinion Sensitivity Dictionary Model for Big Data Management," Ph. D. dissertation, Incheon University, Incheon, MS, 2014.
- [ 7 ] J. H. Lee, H. S. Lee and H. K. Lee, "A Study on Customer Reviews about Domestic and Imported Clothes Products through Opinion Mining," *Korea Internet Electronic Commerce Association*, vol. 15, no. 3, pp. 223-234, June 2015.
- [ 8 ] S. J. Lee, T. R. Jeon, G. D. Baek, S. S. Kim, "A Movie Rating Prediction System of User Propensity Analysis based on Collaborative Filtering and Fuzzy System," *Journal of the Korea Institute of Intelligent Systems*, vol.19, no.2, April 2009.
- [ 9 ] B. Jeong, D. K. Kim, "Design and Implementation of Location Recommending Services using Personal Emotional Information based on Collaborative Filtering," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 20, no. 8, pp. 1407-1414, Aug. 2016.
- [10] H. C. Shin and S. B. Cho, "A Location-based Collaborative Filtering Recommender using Quadtree," *Journal of the Korea Institute of Information and Communication Engineering : Computing Practices and Letters*, vol. 19, no. 1, pp. 15-22, Jan. 2013.
- [11] N. I. Woo, "Shopping recommendation system using collaborative filtering," M.S. Theses, Inha University, Incheon, 2014.
- [12] The members of the R Development Core Team. The R environment [Internet]. Available: <https://www.r-project.org/>.
- [13] H. W. Jeon, T. K. Kim. Korean NLP Package [Internet]. Available: <https://cran.r-project.org/web/packages/KoNLP/index.html>.
- [14] RStudio, Inc. IDE for R [Internet]. Available: <https://www.rstudio.com/>.
- [15] I. Feinerer, K. Hornik, Artifex Software, Inc. Package 'tm' [Internet]. pp. 40-41. Available: <https://cran.r-project.org/web/packages/tm/tm.pdf>.
- [16] MangoPlate Co., Ltd. [internet]. Available: <https://www.mangoplate.com/>.
- [17] C. H. Lee, S. Y. Lee, T. C. Chung and S. H. Yoon, "Application recommender system based on personalized collaborative-filtering using user's emotion information from smartphone," *Journal of Korea Institute of Information Scientists and Engineers*, vol. 39, no. 1A, pp. 224-226, June 2012.





최은정(Eun Jeong Choi)

2016년 상명대학교 미디어소프트웨어학과 학사  
2016 ~ 현재 상명대학교 컴퓨터과학과 석사과정  
※관심분야 : 빅데이터 분석, 오피니언 마이닝, 감성분석



김동근(Dong Keun Kim)

2003년 : 연세대학교 의료정보 석사  
2008년 : 연세대학교 생체공학 박사  
2009년 ~ 현재 : 상명대학교 미래융합공학대학 휴먼지능정보공학과 부교수  
※관심분야 : 생체정보공학, 헬스케어, 감성공학, HCI