

빅데이터 기반 맛 형용사 감성사전을 통한 음식 추천 시스템

Food Recommendation System Using BigData Based on Scoring Taste Adjectives Sentiment Dictionary

김석진¹, 김용성^{2*}

SukJin Kim¹, Yong-Sung Kim^{2*}

요약

일반적으로 음식 맛의 평가는 관능평가로 이 평가방법은 비용과 시간이 많이 소요되고, 맛의 편차가 크게 나타나는 단점이 있다. 본 연구에서는 이러한 관능평가의 단점을 보완하기 위해서, SNS를 기반으로 빅데이터를 맛 형용사 감성사전을 활용하여 음식 맛을 평가하는데 목적이 있다. 먼저 선행 연구를 토대로 기초 맛 형용사 사전을 구축한 후 맛, 맛집, 음식, 음식점을 키워드로 SNS의 트위터 데이터를 수집한 후 각 트윗을 음식명과 맛 형용사 포함여부에 따라 4가지로 구분한다. 4가지로 구분된 트윗 데이터를 형태소 분석을 통해 새로운 맛 형용사와 음식명을 발굴하여 맛 형용사 감성사전에 추가한다. 새롭게 구성된 맛 형용사 감성사전을 활용하여 필터링한 유의미한 트윗 데이터를 맛 형용사의 빈도수를 측정하고 음식별 맛 형용사의 가중치를 결정하고 음식별 맛 형용사의 맛 지각도를 도식화 시킨다.

핵심어 : 관능평가, 맛 형용사, 추천시스템, 빅데이터, 감성사전, 트위터, k 군집알고리즘

Abstract

It is a general practice to evaluate food tastes based on sensory tests, however, this test method has a disadvantage wherein it is costly and more time is required and significant deviation takes place depending on each evaluator as well. First, a pre-taste Adjectives Sentiment Dictionary is built based on the previous studies and then a collection of Twitter data of SNS is performed by keywords - a food, a restaurant -. Tweet distinguish each of the four different names on food or not to include food taste adjectives. The data that is divided into four tweets go through stemming wherein new flavors and foods adjectives words are discovered in order to add adjectives to taste sentiment dictionary. Thereby measuring the frequency of a newly formed taste adjective is significant based on sensitivity of tweet filtering data by utilizing a pre-taste adjective sentiment dictionary which is determined by the weight of the food taste adjectives and it also illustrates the taste of foods by taste adjectives.

Keyword : Sensory evaluation, Taste Adjective, Recommendation System, BigData, Sentiment Dictionary, Twitter, k-means Algorithm

¹ Department of Computer Engineering, Chonbuk National University 664-14 DuckJinDong, DuckJin-Gu, Jeonju, Korea
e-mail : hanuri00@jbnu.ac.kr

² Department of Computer Engineering, Chonbuk National University 664-14 DuckJinDong, DuckJin-Gu, Jeonju, Korea
e-mail : yskim@jbnu.ac.kr (Corresponding Author)

Received(March 06, 2016), Review(March 19, 2016), Accepted(June 03, 2016), Published(June 30, 2016)

1. 서론

한국 음식의 맛 표현에 관한 연구들은 대부분 관능평가에 의하여 진행되어 지고 있다. 관능평가란 ‘식품과 물질의 특성이 5각(시각, 후각, 미각, 촉각 및 청각)으로 감지되는 반응을 측정, 분석, 해석하는 과학의 한 분야’라고 정의하고 있다[1]. 본 연구에서는 맛 형용사를 추출하기 위해서 SNS의 트위터에서 음식이 들어간 트윗을 수집하여 수집된 트윗 중 쌍(음식명, 맛 형용사)에 따라 새로운 맛 형용사들을 추출하였고, 맛 관련 외식경영잡지 2014년 1월부터 10월호까지 맛 형용사의 표본을 2만 여개 추출하였다. 수집된 맛 표현 형용사들의 적합성 평정을 위해 통계적 방법인 상관분석과 다변량 분석을 통한 맛 평정 테이블을 구성하여 108개의 맛표현 형용사들을 활용하여 1차 맛 형용사 감성사전을 구축하였다.

1차 맛 형용사 감성사전과 미각과 질감 및 온도, 후각 등의 맛 형용사 데이터 87개와 비교분석하여 최종적으로 71개의 형용사를 선택한 후 선택된 맛 형용사들을 러프 집합 이론을 통해 맛 형용사 평정하였다.

기존 연구들은 관능평가에 따른 맛 평가를 지역적이고 제한적인 평가를 진행하였기 때문에 지역적인 방법을 극복하여 한식 개발 시 중요한 고려 요인, 영양에 대한 인지도, 향후 구매 가능성을 조사하여 한식의 세계화를 위한 마케팅 전략 수립에 활용하는 연구도 진행할 필요성이 있다. 또한 관능평가의 지역적 한계를 극복하고 객관적인 평가를 진행하고자 전자혀를 통해 한식의 맛 표현을 측정하는 방안을 수립해 본다.

외국인들에게 인지도가 높은 한국 음식의 대표음식들로는 김치, 음식[2][3], 불고기 등이 있으며 특히 음식[4]은 밥과 반찬, 양념의 조합으로 한국인의 일상식사에서 쉽게 만들 수 있는 음식으로 전국 어디서나 먹을 수 있는 종류[5][6]도 음식이지만 지역적 차이가 존재한다.

일찍이, 음식을 세계화한 국가들은 나라별로 음식의 맛을 느끼는 감성을 연구해 왔다. 서양의 페스트 푸드, 일본의 스시, 중국의 딤섬, 태국의 쌀국수, 유럽의 레스토랑 음식 등이 세계인의 음식으로 자리 잡고 있다. 또한 페스트 푸드의 대명사 햄버거의 맛은 나라별, 지역별로 특성에 맞게 약간의 차이점을 갖고 있는 것도 세계화하는데 하나의 중요한 요건이 되고 있다.

논문의 1장에서는 연구의 배경에 대해 논하였으며, 2장에서는 빅데이터와 텍스트마이닝, 군집분석에 관련된 연구들에 대해서 조사하였고, 3장에서는 맛 형용사 감성사전 구축과 이를 활용한 감성분석 설계에 대하여 설명하였다. 마지막으로 4장에서 결론 및 향후 연구방향에 대하여 기술하였

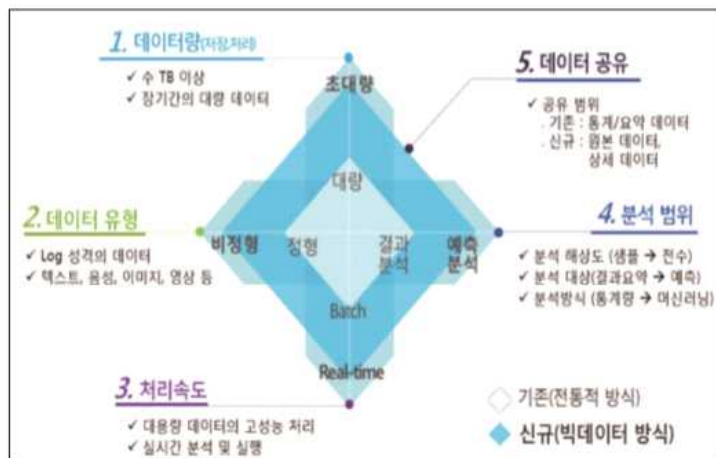
다.

2. 관련연구

2.1 빅데이터

빅데이터의 기본 특징은 3V로 Volume, Variety, Velocity을 말한다. 빅데이터는 기존 데이터에 매우 방대한 규모를 가지며, 정형, 반정형, 비정형 등의 형태를 가지며, 스마트폰이나 SNS, 센서, IoT 등의 다양한 디바이스에 의해 빠른 속도로 생성되는 데이터로서 기존의 일반적인 데이터베이스, 소프트웨어로는 관리하기 어려운 데이터를 의미하며, 나아가 이러한 데이터들을 다양한 기법등을 통해 분석하는 방법 및 플랫폼등을 포함하는 넓은 범위의 용어를 말한다.

텍스트 마이닝(Text Mining) 비정형이나 반정형형태의 텍스트데이터에서 자연어처리(Natural Language Processing) 기술에 기반하여 유용한 정보를 추출, 가공하는 것을 목적으로 하는 기술이다. 뉴스 텍스트 마이닝을 통해 주가의 상승 하락을 예측[7], SNS 데이터의 효율적으로 데이터 마이닝을 통한 범죄 발생 위험 요소에 대한 연구[8], 신문기사의 텍스트 마이닝을 통한 신문사별 논조 차이점을 분석한 연구 등이 있다.



[그림 1] 빅데이터 특징 및 기술사례

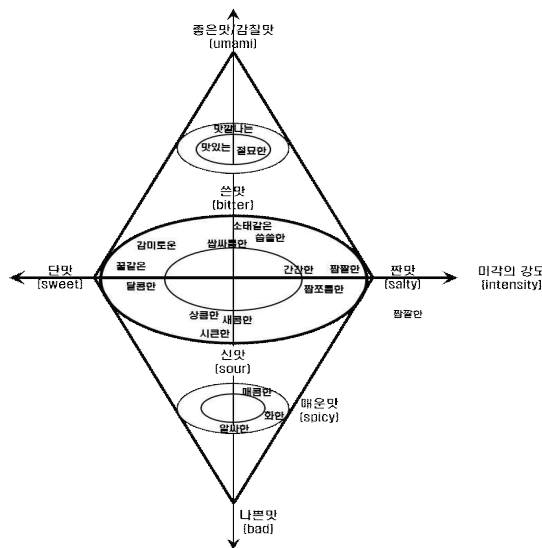
[Fig 1] Big Data features and technical case

군집 분석(Cluster Analysis)은 비슷한 특성을 가진 개체를 합쳐가면서 최종적으로 유사 특성의

군(Group)을 발굴하는데 사용된다. 예를 들어 트위터 상에서 주로 사진, 카메라에 대해 이야기하는 사용자군이 있을 수 있고, 자동차 에 대해 관심 있는 사용자군이 있을 수 있다. 이러한 관심사나 취미에 따른 사용자군을 군집 분석을 통해 분류할 수 있는 기법이다.

2.2 맛 형용사

일반적으로 미각의 질은 5가지의 단맛(sweet), 신맛(sour), 짠맛(salty), 쓴맛(bitter), 감칠맛(umami)으로 정의하지만, 한국인의 식탁에서 빠질 수 없는 매운맛(spicy)을 한식의 맛에 부여하여 6가지로 구분하기도 한다. 맛 표현 형용사들을 쌍원추 모델 이용하여 날말밭으로 구조화하였는데 쌍원추 모델의 수평축은 맛의 강도를 나타내는데, 중심에서 멀어질수록 맛의 강도가 증가함을 보여준다. [그림 2]에서보면 맛의 3차원 날말밭과 같이 맛 형용사들을 분류하는데 3개의 매개변수인 미각의 질(taste quality), 미각의 강도(taste intensity), 쾌감이 결정적인 역할을 수행한다고 표현하였다.



[그림 2] 맛의 3차원의 날말밭

[Fig 2] 3D fields of taste

3. 맛 형용사 감성사전 기반 추천시스템

본 연구에서는 트위터 자료를 twitter API를 이용하여 2015년 1월 1일부터 6월 31일까지 데이터

를 수집하였다. 본 30여만건의 트윗중에서 맛 형용사와 전혀 관련이 없는 무의미한 트윗과 맛 형용사를 포함하지 않는 트윗을 1차적으로 제거하여 유의미한 트윗 45076건을 추출하였다. 이후로 데이터의 간결 및 통계처리를 위해 같은 내용의 tweet이 2회 이상 반복된 트윗과 동일 인물이 2회 이상 올린 tweets를 제거하였고 이렇게 중복된 tweets를 제거한 유효 tweet의 수는 21038개이다.

3.1 형태소 분석

원시데이터를 R의 KoNLP 패키지를 활용하여 형태소 분석을 처리했다. KoNLP 에 포함된 세종 사전에는 고유명사를 비롯한 맛 관련 형용사는 포함되어 있지 않아 평정된 맛 형용사 테이블을 따로 추가한 후 맛 형용사와 낱말의 어근만을 추출하였다. 아래 [그림 4]은 맛 형용사 및 단어어근만 추출처리 후 데이터를 랜덤으로 100개의 단어를 보여준 것이다.

번호	형태소	어근	어근	어근
110	맛있는거	맛	맛	맛
111	맛있음	맛	맛	맛
112	맛있어	맛	맛	맛
113	맛있	맛	맛	맛
114	맛있	맛	맛	맛
115	맛있	맛	맛	맛
116	맛있	맛	맛	맛
117	맛있	맛	맛	맛
118	맛있	맛	맛	맛
119	맛있	맛	맛	맛
120	맛있	맛	맛	맛
121	맛있	맛	맛	맛
122	맛있	맛	맛	맛
123	맛있	맛	맛	맛
124	맛있	맛	맛	맛
125	맛있	맛	맛	맛
126	맛있	맛	맛	맛
127	맛있	맛	맛	맛
128	맛있	맛	맛	맛
129	맛있	맛	맛	맛
130	맛있	맛	맛	맛
131	맛있	맛	맛	맛
132	맛있	맛	맛	맛
133	맛있	맛	맛	맛
134	맛있	맛	맛	맛
135	맛있	맛	맛	맛
136	맛있	맛	맛	맛
137	맛있	맛	맛	맛
138	맛있	맛	맛	맛
139	맛있	맛	맛	맛
140	맛있	맛	맛	맛
141	맛있	맛	맛	맛
142	맛있	맛	맛	맛
143	맛있	맛	맛	맛
144	맛있	맛	맛	맛
145	맛있	맛	맛	맛
146	맛있	맛	맛	맛
147	맛있	맛	맛	맛
148	맛있	맛	맛	맛
149	맛있	맛	맛	맛
150	맛있	맛	맛	맛

[그림 3] 단어어근 추출
[Fig 3] Word of root extract

빈도수	빈도수	빈도수	빈도수	빈도수	빈도수
2371	맛	맛	맛	맛	맛
2041	맛	맛	맛	맛	맛
1889	맛	맛	맛	맛	맛
2011	맛	맛	맛	맛	맛
937	맛	맛	맛	맛	맛
395	맛	맛	맛	맛	맛
491	맛	맛	맛	맛	맛
302	맛	맛	맛	맛	맛
292	맛	맛	맛	맛	맛
261	맛	맛	맛	맛	맛
241	맛	맛	맛	맛	맛
236	맛	맛	맛	맛	맛
235	맛	맛	맛	맛	맛
230	맛	맛	맛	맛	맛
212	맛	맛	맛	맛	맛
195	맛	맛	맛	맛	맛
183	맛	맛	맛	맛	맛
182	맛	맛	맛	맛	맛
179	맛	맛	맛	맛	맛
171	맛	맛	맛	맛	맛
163	맛	맛	맛	맛	맛
161	맛	맛	맛	맛	맛
159	맛	맛	맛	맛	맛
158	맛	맛	맛	맛	맛
155	맛	맛	맛	맛	맛
152	맛	맛	맛	맛	맛
149	맛	맛	맛	맛	맛
141	맛	맛	맛	맛	맛

[그림 4] 빈도수 상위 100개 단어
[Fig 4] Words of 100 high-frequency

추출된 맛 형용사 및 어근에 맛 형용사 평정 척도 table를 이용하여 맛 형용사가 있는 트위터 데이터를 필터링 한다. 총 45076개의 데이터에서 2회 이상 중복되거나 무의미한 불용어만으로 된 tweets들을 제거한 유효 데이터 21038여개의 tweets를 불용어 처리 및 형태소 분석 후 필터링하였다. 그중 빈도수가 가장 많이 나온 단어 상위 100개의 단어들은 [그림4]이다. 21038개의 트윗에서 맛 형용사 와 단어들 중에서 언급된 빈도수가 2인 단어들을 이용하여 wordcloud를 구성하였다. [그림 5]은 R를 통해 wordcloud를 구성한 것이다. 비빔밥에 관한 언급이 가장 많았고 전주에 대한 언급과 음식, 칼국수, 찌개 등에 대한 언급도 많이 보인다.



[그림 5] 음식 및 맛 형용사 단어

[Fig 5] Food and taste adjective words

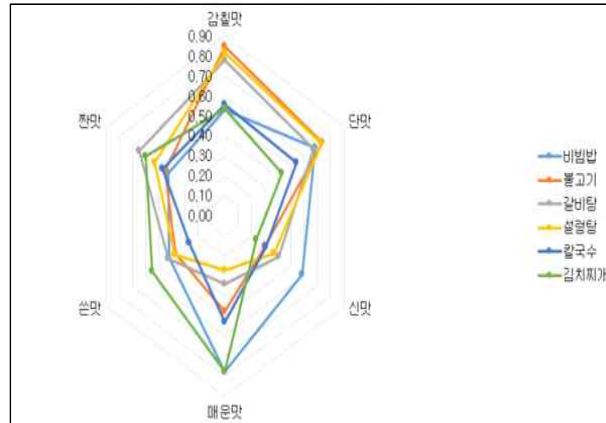
3.2 clustering

필터링한 데이터를 단맛, 짠맛, 신맛, 쓴맛, 감칠맛의 오미와 매운맛으로 처리한 후 clustering 한다. 음식에 대한 언급된 상위 20 단어들의 분포를 아래 그림과 같다. 가장 많은 언급은 ‘비빔밥’으로 11.7%이고, 전주 10.0%, 음식, 7.2%, 칼국수 6.7%, 짜개 6.1% 순으로 음식 관련 단어들로 나타났다. 비빔밥에 대한 언급이 많은 이유로는 비빔밥에 대한 종류가 다양한 것으로 육회비빔밥, 영양돌솥비빔밥, 돌솥비빔밥, 새싹비빔밥, 산채비빔밥 등을 비빔밥으로 통합 분석한 것에 기인한다[8]. [표 1]은 수집 트위터에서 가장 많이 언급된 음식 6가지에 대한 맛 평점표에 기반하여 분석처리한 수치값이다. 비빔밥에 대한 맛의 대표적인 분류는 매운 음식으로 볼 수 있는데 매운 정도가 0.78로 가장 높게 나타났고 단맛 0.68, 신맛 0.59, 감칠맛 0.53, 짠맛 0.43, 쓴맛 0.42 로 나타났으며 불고기는 감칠맛이 0.85로 높게 나타났다. 맛 지각도는 아래 그림 [그림 6,7]과 같다.

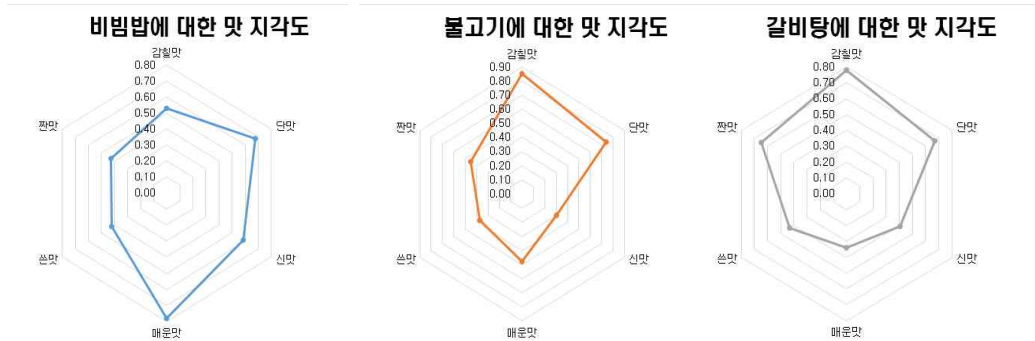
[표 1] 음식(6개)에 대한 맛 지각도

[Table 1] Recognition of food and taste

맛	감칠맛	단맛	신맛	매운맛	쓴맛	짠맛
비빔밥	0.53	0.68	0.59	0.78	0.42	0.43
불고기	0.85	0.74	0.30	0.48	0.37	0.45
갈비탕	0.78	0.67	0.41	0.34	0.43	0.65
설렁탕	0.82	0.73	0.37	0.27	0.38	0.53
칼국수	0.56	0.54	0.31	0.53	0.27	0.47
김치찌개	0.54	0.43	0.24	0.78	0.55	0.60



[그림 6] 음식 맛 지각도
 [Fig 6] Recognition of food and taste



[그림 7] 음식들의 맛 지각도
 [Fig 7] Recognition of food and taste

오미와 매운맛을 포함한 6분류의 맛 형용사를 고려하여 초기 클러스터링의 수를 6으로 정하였다. 제한된 맛 형용사에 의해서 클러스터링 분포가 선형성을 띄게 되었다. 더 많은 맛 형용사와 트윗 데이터를 통해 분석할 필요가 있다.

4. 결론

본 논문은 한식을 세계화하기 위해서 한식 중에서 우리나라 사람들이 가장 선호하는 음식이 무엇이고, 왜 한국인의 기호 식품이 되었고, 왜 아직도 외국인들이 좋아하지 않는 지, 국가마다 연령마다 맛을 느끼는 감성이 다르다는 데 착안하여 연구에 출발점을 두었다. 먼저 맛을 나타내는 맛

형용사들을 선별하기 위해서 현재 활용하고 있는 맛 형용사 단어들을 기존 연구 및 SNS에서 추출하여 기초 맛 형용사 사전을 구축한 후 기초 맛 형용사 사전을 활용하여 음식명과 맛 형용사의 4가지 쌍집합을 통해 새로운 맛 형용사와 음식을 추가하여 1차 맛 형용사 감성사전을 구축하였다. 그리고 맛-형용사 테이블에 속하는 트윗을 형태소 분석과 테이블 매핑을 통해서 선별한 후 이들중에서 빈도수가 가장 높은 한식 중에서 비빔밥, 불고기, 갈비탕을 k-means 알고리즘에 의해 선별하여 추천하였다. 또한 대표 음식들의 맛의 강도를 측정하기 위해서 5가지 맛에 대한 지각도를 나타냄으로써 한국인의 대표음식에 대한 맛의 감성부분을 구할 수 있었다. 본 논문의 공헌도는 명사위주로 되어진 감성사전에 한국의 맛 형용사 감성사전을 활용한 것과 평정된 맛 형용사 감성사전을 이용해서 대표 한식을 추천할 수 있는 알고리즘을 제안하고, 이를 활용할 수 있는 기반을 제공했다는 점이다. 향후연구로는 특정 지역의 트위터 데이터를 수집한 것을 전국적으로 시행하여 각 시, 도의 맛-감성지도를 작성하는 것과 외국인들의 맛지도를 작성하여 한식을 세계화 할 수 있는 맛과 레시피를 제안하는데 있다.

References

- [1] Korean Ministry for Food, Agriculture, Forestry and Fisheries, (2010).
- [2] krdic.naver.com, Retrieved: (2016) March 10.
- [3] Y. Joo, "A Study on Evolution and Discourse of BIBIMBAB", *Society and History*, vol. 87, (2010), pp. 5-38.
- [4] ko.wikipedia.org, Retrieved: (2016) March 10.
- [5] S. Ahn and S. Cho, "Stock Prediction Using News Text Mining and Time Series Analysis", *Korea Computer Congress*, vol. 37, no. 1, (2010), pp. 364-369.
- [6] J. Lee, K. Song, J. Kang and J. Hwang, "A study on the efficient extraction method of SNS data related to crime risk factor", *Journal of The Korea Society of Computer and Information*, vol. 20, no. 1, (2015), pp. 255-263.
- [7] M. Kam and M. Song, "A Study on Differences of Contents and Tones of Arguments among Newspapers Using Text Mining Analysis", *Journal of Intelligence and Information Systems*, vol. 18, no. 3, (2012), pp. 53-77.
- [8] Y. Park and H. Kim, "A Semantic Foundation for the study of Taste-Adjectives", *Korean Association for Semiotic Studies*, vol. 36, (2013), pp. 9-36.