



Journal of Knowledge Information Technology and Systems

ISSN 1975-7700

<http://www.kkits.or.kr>

A Study of User Text Sentiment Dictionary for Food Recommendation Service on Big Data Environment

Jin-Kwan Cho*

Department of Food Service Management, Kyungdong University

ABSTRACT

In general, it is practice to assess food tastes based on sensory tests, however, this method has a considerable disadvantage in which it is pricey and need to required more time. In addition, important disparity appear in relying on each evaluator. We make a good food tastes up, we are necessary to process as following; Firstly, a pre-taste User text sentiment dictionary is based on establishment a kind of food studies and then gather this information on twitter data of social network service, internet and social media so on. Secondary, many information are based on anticipate in original food tastes data by web-based and mobile-based from their systems. Food name of providing on social network service and internet is different each of names with their food tastes. After all, This data is divided into four different names look over stemming that new savors and foods words are found in order to add keyword to taste sentiment dictionary. Accordingly, frequency measurement of newly formatted taste keyword is important based on sensitivity of filtering data by utilizing a pre-taste word emotional dictionary determined by the weight of the food taste keyword and it also illustrates the taste of foods by taste dictionary.

© 2019 KKITS All rights reserved

KEYWORDS : Food tastes, User text sentiment dictionary, Twitter data of social network of service, Internet and social media, Filtering data

ARTICLE INFO: Received 6 June 2019, Revised 8 July 2019, Accepted 9 August 2019.

*Corresponding author is with the Department of Food Service Management, Kyungdong University, Metropol Campus(Yangju) 27 Kyungdong University-ro, Yangju,

Gyeonggi-do, 11458, KOREA.
E-mail address: jkcho@kduniv.ac.kr

1. 서론

일반적으로 음식의 맛 표현에 관한 연구들은 관능평가에 의해 진행되어 지고 있다 [1-3]. 관능평가란 ‘식품과 물질의 특성이 시각, 후각, 미각, 촉각 및 청각의 5각으로 감지되는 반응을 토대로 측정, 분석, 해석하는 과학의 한 분야’이다 [4-6]. 이에, 본 논문에서는 맛의 키워드를 생성을 위해 키워드를 추출하고자, 소셜 네트워크 서비스에서의 음식의 정보를 수집하고, 수집된 정보 중 데이터 타입을 음식명과 맛의 의미론적 데이터에 따라 새로운 맛의 의미론적 데이터를 추출하고자 한다 [7-9]. 수집된 맛의 의미론적 데이터의 적합성 평정을 위하여 통계적 방법인 상관분석과 다변량 분석을 통해 맛 평정 테이블을 구성하고자 한다. 또한, 맛 표현들을 활용하여 1차 맛 텍스트 감성사전을 구축하고자 한다. 1차 맛 텍스트 감성사전에서의 미각과 질감 및 온도, 후각 등의 맛을 표현하는 데이터와 비교분석함으로써 [10-12], 최종적으로 맛 데이터상의 키워드를 선택한 후 선택된 맛 키워드들을 러프 집합 이론을 토대로 맛 데이터를 측정하고자 한다 [13]. 기존 연구들은 관능평가에 따른 맛 평가를 지역적이고 제한적인 평가로 인해, 지역적인 방법에 대한 문제점을 극복하여 음식(한식) 개발 시 중요한 고려 요인인, 영양에 대한 인지도와 향후 구매 가능성을 조사하여 세계적 음식을 위한 마케팅 전략 수립과 동시에 이에 대한 연구 진행이 반드시 필요하다 [14-16]. 또한, 지역적 한계를 극복하기 위한 관능평가 및 객관적인 평가를 통해 고유의 맛 표현을 측정하는 방법에 대해 살펴보고자 한다. 외국인들에게 인지도가 높은 한국 음식의 대표음식들로는 김치, 음식, 불고기 등이 있으며 특히 음식에서는 밥과 반찬, 양념의 조합으로 한국인의 일상식사에서 쉽게 만들 수 있는 음식으로서의 지역적 차이에 따른 맛의 상태가 존재할 수 있다.

음식을 세계화한 국가들은 나라별로 음식의 맛을 느끼는 감성에 대해 연구를 진행하고 있다 [17-19]. 서양의 페스트 푸드, 일본의 스시, 중국의 덩섬, 태국의 쌀국수, 유럽의 레스토랑 음식 등이 세계인의 음식으로 자리 잡고 있다 [20]. 또한 페스트 푸드의 대명사 햄버거의 맛은 나라별, 지역별로 특성에 맞게 약간의 차이점을 갖고 있는 것도 중요한 요건 중의 하나이다. 따라서 본 논문은 다음과 같이 구성된다. 1장에서는 연구의 배경에 대해 설명하고, 2장에서는 빅데이터와 텍스트마이닝, 군집분석에 대해 살펴본다. 또한, 3장에서는 맛 감성사전 구축과 이를 활용한 감성분석 설계에 대하여 살펴본다. 마지막으로 4장에서 결론 및 향후 연구방향에 대하여 기술한다.

2. 관련연구

2.1 빅데이터

빅데이터는 Volume, Variety, Velocity로 3V로 정형, 반정형, 비정형 데이터를 효율적으로 수집·저장·관리하는 동시에 정보와 융합된 다양한 속성 정보에 대해 실시간·통합 분석을 수행하며, 의미 있는 정보를 추출함으로써 미래에 대응할 수 있는 기술을 의미한다. 다양한 스마트 기기나 Social 네트워크 기반 서비스 및 센서와 IoT 등의 다양한 데이터들을 활용하여 다양한 분석기술 기법을 적용하는 등 융합적이고 통합적인 방법론을 통해 가치를 창출하고자 한다. 또한, 융합 플랫폼을 활용한 사용자 중심의 가치있는 서비스를 제공하는 것을 말한다 [21].

2.2 텍스트 마이닝

텍스트 마이닝 (Text Mining)은 비정형이나 반정

형 형태의 텍스트 데이터를 자연어 처리 기술을 이용하여 유용한 정보를 추출, 가공하는 기술을 의미한다. 주가의 상승 하락, 일기예보 등에 기반 한 예측, 소셜 데이터 기반의 효율적인 데이터 마이닝을 이용한 범죄 발생 위험 요소에 대한 연구, 신문 기사의 텍스트 마이닝을 활용한 각 기사별 논조 차이점을 분석하는 연구 등이 있다 [22].

2.3 군집분석

비슷한 특성을 가진 개체를 합쳐가면서 최종적인 유사 특성의 군을 발굴하것을 군집 분석이라 말한다 [23]. 예를 들어 온라인상에서 주로 이미지, 사진 등에 대해 이야기하는 사용자 군이 있을 수 있고, 자동차, 쇼핑 상품에 대해 관심 있는 사용자 군이 있을 수 있다 [24]. 이러한 군집 분석을 통해 관심사나 취미에 따른 사용자군을 분류할 수 있는 기법이다.

3 맛 감성사전

본 연구는 소셜 데이터 서비스 자료 중 트위터 환경 기반의 API를 이용하여 데이터를 수집하였다. 데이터의 수집 기간은 트위터를 가장 많이 활용하는 4계절을 중심으로 약 1년 간의 데이터 중 음식의 맛과 전혀 관련이 없는 무의미한 데이터와 맛을 포함하지 않는 데이터를 1차적으로 제거하여 유의미한 데이터 약 45100건을 추출하였다 [25]. 이후로 데이터의 간결 및 통계처리를 위해 동일인이 동일 내용의 데이터가 2회 이상 반복된 것은 제거하였고, 이렇게 중복된 데이터를 제거한 유효 데이터의 수는 약 21000개이다. 다음의 <그림 1>은 감성사전을 위한 아키텍처이다.



그림 1. 감성사전기반 아키텍처
Figure 1. Sentiment Dictionary-based Architecture

3.1 형태소 분석

R 패키지 기반의 원시 데이터를 활용하여 형태소를 분석하였다. R 패키지의 네이버 사전에는 고유명사인 평정된 맛 테이블을 생성하였다. 맛 낱말의 어근만을 추출함에 있어, 아래 <그림 2>처럼 맛 분석된 형태소의 어근만 추출 처리 한 후 데이터를 랜덤으로 생성된 단어를 추출하였다.

문맥유형	문종명	제목	문맥유형	문종명	문종명	유사	사용여부
33373	텍스트	알뜰방자죽 먹기 + 좋은 삼선부치기 + 겨울매장	1287	25	218.9074165711	Y	
33373	텍스트	알뜰방자죽 먹기 + 좋은 삼선부치기 + 겨울매장	1287	25	218.9074165711	Y	
33373	텍스트	2015년 1월 알뜰 방자죽 먹기 + 겨울매장	1287	25	12.94781392704	Y	
33373	텍스트	알뜰방자죽 먹기 + 좋은 삼선부치기 + 겨울매장	1287	25	24.9555552617	Y	
33373	텍스트	http://m.blog.naver.com/entry.naver?categoryId=44	1287	25	12.94781392704	Y	
33373	텍스트	알뜰방자죽 먹기 + 좋은 삼선부치기 + 겨울매장	1287	25	37.81038262490	Y	
33373	텍스트	1월매장 + 알뜰방자죽 먹기 + 겨울매장	1287	25	12.94781392704	Y	
33373	텍스트	삼선부치기 + 알뜰방자죽 먹기 + 겨울매장	1287	25	12.94781392704	Y	
33373	텍스트	1월매장 + 알뜰방자죽 먹기 + 겨울매장	1287	25	56.40148199826	Y	
33373	텍스트	삼선부치기 + 알뜰방자죽 먹기 + 겨울매장	1287	25	12.94781392704	Y	
33373	텍스트	1월매장 + 알뜰방자죽 먹기 + 겨울매장	1287	25	12.94781392704	Y	
33373	텍스트	삼선부치기 + 알뜰방자죽 먹기 + 겨울매장	1287	25	22.04488667009	Y	
33373	텍스트	알뜰방자죽 먹기 + 좋은 삼선부치기 + 겨울매장	1287	25	42.09519757623	Y	
33373	텍스트	삼선부치기 + 알뜰방자죽 먹기 + 겨울매장	1287	25	42.09519757623	Y	
33373	텍스트	1월매장 + 알뜰방자죽 먹기 + 겨울매장	1287	25	19.01071745911	Y	
33373	텍스트	추석맞이 + 알뜰방자죽 먹기 + 겨울매장	1287	25	87.362914651822	Y	
33373	텍스트	삼선부치기 + 알뜰방자죽 먹기 + 겨울매장	1287	25	74.81651670231	Y	
33373	텍스트	알뜰방자죽 먹기 + 좋은 삼선부치기 + 겨울매장	1287	25	82.88051461187	Y	
33373	텍스트	알뜰방자죽 먹기 + 좋은 삼선부치기 + 겨울매장	1287	25	25.823248492177	Y	
33373	텍스트	삼선부치기 + 알뜰방자죽 먹기 + 겨울매장	1287	25	104.8717915018	Y	
33373	텍스트	삼선부치기 + 알뜰방자죽 먹기 + 겨울매장	1287	25	18.744811951785	Y	

그림 2. 맛 데이터 테이블 샘플
Figure 2. Taste Data Table Sample

추출된 맛 어근을 토대로 맛 데이터 평정 척도 테이블을 생성하여 R 패키지 기반 형태소 분석에 따르는 맛 데이터를 필터링 하였다. 총 100,000개의 데이터 중 2회 이상 중복 및 무의미한 불용어 키워드들을 제거한 후, 유효 데이터 10000여개의

데이터를 불용어 처리 및 형태소를 분석 및 필터링하였다. 그중 빈도수가 가장 많이 나온 단어 상위 1000개의 단어들을 제공한다. 10,000개의 데이터에서 맛과 연관된 단어들 중에서 언급된 빈도수가 두 단어 이상을 이용하여 워드클라우드를 생성하였다.

3.2 소셜 데이터 기반 트윗 수집

본 연구에서는 소셜 데이터 서비스 기반으로 트윗 환경에서의 트윗 수집은 사용자가 요구하는 검색 키워드와 사용자의 경험적 데이터를 활용한다. 대부분 유무선 환경기반 기기들에서 활용되는 장비에 사용자에게 요구되는 정보로는 오직 키워드이다. 이러한 시스템은 사용자에게 검색 키워드와 사용자의 피드백 데이터, 사용자의 경험적 데이터를 통해 트위터 API에 활용할 질의를 구성한다. 트윗에 포함된 각각의 정보들은 사용자가 원하는 음식의 맛에 대한 결과 정보를 보여줄 때 활용되며, 이를 통해 사용자는 자신이 원하는 음식을 토대로 이와 유사하거나 새로운 음식에 대해 관심을 갖도록 한다. 입력받은 데이터 중 사용자, 키워드, 음식, 맛 등 빅데이터 환경 기반 데이터베이스에 저장하여 동적으로 시스템을 사용할 수 있도록 한다. 시스템은 트윗 수집을 위해 사용자의 검색어와 이전 데이터 정보를 조합하여 질의를 구성한다.

3.3 감성 분석

본 연구에서 감성 분석을 위해 위에서 언급한 감성사전을 구축하였다.

표 1. 감성사전의 감성태그
Table 1. Sentiment tag of Sentiment Dictionary

Word	Tag	positive/negative
맵다	ncps	N
짜다	ncpss	N
달다	ncpsw	N
쓰다	ncpb	N
맛있다	ncpns	N
...
고소하다	ncpsp	P

감성사전은 기존의 다양한 연구들에서 작성된 맛에 대한 한국어 감성 단어 400개를 바탕으로 소셜 데이터 기반 트윗에서 추출한 감성단어 100여개를 추가하여 구축하였다. 사전은 형태소 분석된 형태에 태그를 추가하여 긍정을 나타내면 ‘Positive’, 부정을 나타내면 ‘Negative’로 태그하였다. 위의 <표 1>과 같이 음식의 맛에 대한 감성 단어에 대한 샘플 예를 보여준다. 트윗에서 추출한 감성단어들 중에는 의미론적 데이터를 표현 할 수 없는 형태, 예를 들면 ‘므헹헹’, ‘뽕미’, ‘아쓰짬맛’ 등을 제외한 의미론적 데이터의 형태의 감성 단어에 준한 단어들을 생성하였다. 또한, 트위터 상의 단어들 중 반정형화된 맛에 대한 의미론적 데이터를 감성사전에 또한, <표 2>에 추가하였다.

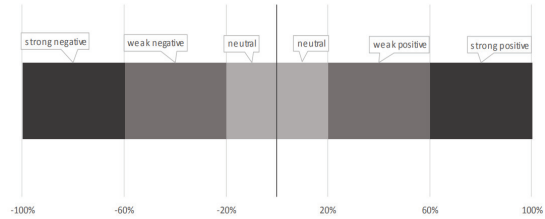
표 2. 감성사전의 감성 태그에 추가된 트위터 단어
Table 2. Tweet word added to the Sentiment tag of Sentiment Dictionary

Word	Tag	positive/negative
꿀맛	ncps	P
대박	ncps	P
맛나	ncpsw	P
척오	ncpb	P
개꿀맛	ncpns	P
...
망	ncpsp	N

위의 <테이블 1과 2>에서 본 것과 같이 수집된 트윗 집합 $T=\{t_1, t_2, \dots, t_n\}$ 이고, 임의의 트윗 t_i 에 존재하는 감성단어의 집합 $W_i = \{w_1, w_2, \dots, w_n\}$ 이라 할 때, 긍정과 부정의 함으로 이루어진 임의의 트윗의 추천 점수를 반영할 수 있다. 이에 다음 <그림 4>와 <그림 5>와 같이 나타난다.

rank	recommend score	distance	latitude, longitude coordinate	tweet contents
1	2	7.0610	37.55624574 126.92703	lunch time 다들 맛있는 점심 드셨죠? 이제부터 홈입해보겠습니다. 하하 #점심 #먹방 #lunch #eat #foodpic #pic #dailycook #cook...
2	1	2.3709	37.56826 126.97783	#피자왕볼로 #100% #오늘은 좋은 일이 가득할것입니다 #점심 #피자 #맛난피자 #먹고보자 #대한항공 #이리저리 #9000miles #pizza #행복하세요 @ 아이카삼
3	1	3.7265	37.53863101 127.0010210	맛있는 점심~ @ 비제나
4	1	7.3882	37.55615285 126.923225	이 밥이 얼마나 맛있는지 알아요? #점심 #규동 @ 흥대어디가
5	1	7.6223	37.50383 127.0049737	Minions macaron! 하하 너무 귀엽네 미니언즈 마카롱! 율회사 스타일과 행복한 점심시간 #일상 #점심 #마카롱 #미니언즈 #맛스타그램 #맛스타그램 #디저트 #데일리...
6	0	1.5240	37.57754976 126.9886474	분어 는 못먹어도 분어 #카르파치오 는 먹지요 #점심 #종로 #종로맛집 #맛스타그램 #먹스타그램 #foodstagram #firstfood @ Dining in SPACE
7	0	1.8066	37.56969652 126.984036	눈여겨봤던 #랜스데이크 집에 나왔음! 사촌동생 청첩 #점심 " " #점심부터 지금저글 꼬기 남남 랜스데이크집 한 세군데를 가봤던것 같은데 여기가 쎄 맛나네 스테이크 리어스는 좋...
8	0	2.1332	37.5636473 126.9825027	점심 #seoulsearching @ Todai
9	0	2.3709	37.56826 126.97783	표정어. #먼치킨 #munchkin #catstagram #cat #고양이 #아이메고르다 #점심 @ 아이카삼
10	0	2.4147	37.5604964 126.9810446	점심2차 □ @ 신세계 본점 Shinsegae

그림 4. 감성사전 기반 추천 결과
Figure 4. The result of Recommendation Sentiment Dictionary



Type of dictionary	MAE	Paired sample T-test (p-value)	Improvement
Original dictionary	2.315	0.000***	2.76%
Modified dictionary	2.251		
Original dictionary	2.144	0.165	-1.17%
Modified dictionary	2.169		
Original dictionary	2.440	0.000***	2.21%
Modified dictionary	2.386		
Original dictionary	2.393	0.000***	1.80%
Modified dictionary	2.350		
Original dictionary	2.516	0.030*	2.03%
Modified dictionary	2.465		
Original dictionary	2.572	0.000***	4.16%
Modified dictionary	2.465		
Original dictionary	2.411	0.000***	2.82%
Modified dictionary	2.343		

그림 5. 감성사전 기반 추천 실험결과
Figure 5. The Evaluated result of Recommendation Sentiment Dictionary

4. 결 론

본 연구는 소셜 네트워크의 데이터에 대한 맛 데이터의 판별을 통해 감성사전 구축에 대한 연구를 수행하였다. 연구결과 감성사전 사용 시, 기존의 맛에 대한 예측 정확도가 높아질 수 있음을 볼 수 있었다. 또한, 본 논문에서는 맛 단어의 극성을 새로 정의할 수 있었으며, 실질적 성능 향상을 확인한 것이 본 연구의 가장 큰 결과라고 할 수 있다. 하지만, 맛에 대한 음식별 감성분석에 필요한 맞춤형 감성 사전 구축을 위해 본 연구는 제한적 한계가 있다. 첫째, 맛 데이터에 대한 맞춤형 감성사전 구축을 위해서는 모든 음식 관련 기업에서 사용하는 음식 재료 및 음식에 대한 빅데이터 정보가 요구된다. 또한, 맛 데이터에서의 동사, 명사, 형용사 등 다른 여러 품사의 단어에도 감성분석이 활용될 수 있기 때문에 이에 대한 알고리즘 개발이 요구된다. 둘째로, 개인 맞춤형 맛 데이터를 제공하는데 있어서는 정확도가 개인별 환경 및 바이오 리듬에 따라 달라질 수 있기 때문에 정확도가 개선되지 않은 이유를 분석하여 감성 사전 구축 방안을 개선할 필요가 있다. 셋째로, 맛 데이터에 대한 비정형 데이터들 긍정과 부정이 아닌 다른 맛의 평가 형태에 대한 비율을 고려하지 않았기에, 부정어 처리를 진행하지 않은 문제점에 대해 감성 분석을 통한 맛 예측 정확도 방법을 제시해야만 한다. 마지막으로, 데이터의 특성이 명확하게 구분되는 맛 데이터를 활용해야만, 감성사전에 대한 확정성이 요구될 수 있다. 이를 위해 보다 정량적이고 정성적인 다양한 데이터 확보와 감성사전에 적합한 모델링을 통해 추가적인 연구가 진행되어야 한다.

References

- [1] K. Dave, S. Lawrence, and D. M. Pennock, 2003. *Mining the peanut gallery: opinion extraction and semantic classification of product reviews*. Vol. 2, No.3, pp. 519-528, 2003.
- [2] C. Ding T. Li, W. Peng, and H. Park. *Orthogonal nonnegative matrix t-factorizations for clustering*. In KDD '06, pp. 126-135, 2006.
- [3] C. Ding, X., He, and H. D. Simon. *On the equivalence of nonnegative matrix factorization and spectral clustering*. In Proceedings of the SIAM Data Mining Conference, pp. 606-610, 2005.
- [4] C. Ding, T. Li, and W. Peng. *On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing*. *Computation Statistics and Data Analysis* Vol. 52, No. 8, pp. 3913-3927, 2008.
- [5] A. Esuli, and F. Sebastiani. *Sentiwordnet: A publicly available lexical resource for opinion mining*. In LREC06, Vol. 3, No. 2, pp. 417-422, 2006.
- [6] Hatzivassiloglou. and McKeown. *Predicting the semantic orientation of adjectives*. In EACL '97, pp. 174-181, 1997.
- [7] Hofmann. *Unsupervised learning by probabilistic latent semantic analysis*. *Machine Learning*, Vol. 42 No. 1, pp. 177-196, 2001.
- [8] M. Hu, and B. Liu. *Mining and summarizing customer reviews*. In KDD '04, pp. 168-177, 2004.
- [9] K. Marx, R. Mokken, and M. Rijke. *Using wordnet to measure semantic orientation of adjectives*. In National Institute. pp. 1115-1118, 2004.

- [10] D. Lee, and S. H. Seung. *Learning the parts of objects by non-negative matrix factorization*. Nature 401(6755), pp. 788-791, 1999.
- [11] T. Li, C. Ding, and M. Jordan, *Solving consensus and semi-supervised clustering problems using nonnegative matrix factorization*. In ICDM '07, pp. 577-582, 2007.
- [12] B. Pang, L. Lee, and Vaithyanathan. *Thumbs up?: sentiment classification using machine learning techniques*. In EMNLP '02, pp. 79-86, 2002.
- [13] P. D. Turney, *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*. In ACL '02, pp. 417-424, 2002.
- [14] T. Wilson, J. Wiebe, and P. Hoffmann. *Recognizing contextual polarity in phrase-level sentiment analysis*. In HLT '05, pp. 347-354, 2005.
- [15] T. Wilson, J. Wiebe, and P. Hoffmann. *Recognizing contextual polarity in phrase-level sentiment analysis*. In HLT '05, pp. 347-354, 2005.
- [16] W. Xu, X. Liu, and Y. Gong, *Document clustering based on non-negative matrix factorization*. In SIGIR '03, pp. 267-273, 2003.
- [17] A. Esuli, and F. Sebastiani. *SentiWordnet: A publicly available lexical resource for opinion mining*. in In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06), pp. 417-422, 2006.
- [18] S. Mohammad, and P. D. Turney. *Crowdsourcing a word-emotion association lexicon*. Computational Intelligence, Vol. 29, No. 3, pp. 436-465, 2013.
- [19] X. Vum and S. Park. *Construction of Vietnamese Sentiwordnet by using Vietnamese dictionary*. The 40th Conference of the Korea Information Processing Society, pp. 745-748, 2014.
- [20] S. Baccianella, A. Esuli, and F. Sebastiani. *Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining*. Proceedings of LREC Vol. 10, pp. 17-23, 2010.
- [21] M. Hu, and B. Liu. *Mining and Summarizing customer reviews*. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD04), pp. 168-177, 2004.
- [22] P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie, 1968. *The general inquirer: A computer approach to content analysis*. American Journal of Sociology, Vol. 73, No. 5, pp. 634-635, 1968.
- [23] B. T. Kieu, and S. B. Pham. *Sentiment analysis for Vietnamese*. 2010 Second International Conference on Knowledge and Systems Engineering, pp. 152-157, 2010.
- [24] T. Vu, D. Q. Nguyen, M. Dras, and M. Johnson. *VnCoreNLP: A Vietnamese Natural language processing toolkit*. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, Demonstrations, pp. 56-60, 2018.
- [25] T. D. Nguyen, X. B. Ngo, and M. P. Tu. 2014. *An empirical study on sentiment analysis for Vietnamese*. 2014 International Conference on Advanced Technologies for Communications (ATC 2014), pp. 309-314, 2014.

빅데이터 환경에서의 음식 추천 서비스를 위한 사용자 텍스트 감성사전에 관한 연구

조진관

경동대학교 외식사업경영학과 부교수

요 약

일반적으로 감각 테스트에 기초한 음식 맛을 평가하는 것은 실용적이지만, 이러한 평가는 비용과 시간이 많이 소요되는 단점이 있다. 또한, 각 평가자에게 의존하는데 있기에 불일치가 나타나기도 한다. 따라서, 우리는 맛있는 음식을 맛보고, 다음과 같이 처리하고자 한다. 첫째로, 사전에 맞볼 수 있는 사용자 텍스트 감성 사전이 일종의 음식 연구에 가장 기여하고 있으며, 소셜 네트워크 서비스, 인터넷 및 소셜 미디어의 트위터 데이터를 수집함으로써 음식 연구에 또한 기여한다. 둘째로, 많은 정보들은 그들의 시스템으로부터 웹기반과 모바일기반에서 본연의 음식 맛 데이터를 기대하는데 기반을 두고 있다. 소셜 네트워크 서비스 및 인터넷에서 제공되는 음식 이름은 각기 다른 이름으로 그 음식맛이 다르다. 결국, 이러한 데이터들은 4개의 다른 이름으로 나뉘어져 새로운 맛과 음식 단어가 맛 감성 사전에 키워드를 추가할 뿐 만 아니라, 이를 발견하기 위한 형태소 분석을 해야한다. 따라서, 새로이 형성된 맛 키워드의 빈도를 측정하는 것은 음식 취향 키워드의 가중치에 의해 결정되고, 사전 취향의 단어 감성 사전을 이용함으로써 필터링된 데이터의 감도에 기초하기에 그 중요성이 있으며, 감성 사전에 의한 음식의 맛을 제공한다.

including Hanyang University, Kyunghee University and Kwangwoon University from 1991 to 1997. He has been a professor in the Department of Food Service Management in Kyungdong University since 1998. His current research interests include English phonetics and phonology, big data, and food service industry. He is a member of the KKITS.

E-mail address: jkcho@kduniv.ac.kr



Jin-Kwan Cho received the bachelor's degree from the department of English Language and Literature in Hanyang University in 1988. He received the M.S. degree and the Ph.D. degree from the Department of English Language and Literature in Hanyang University in 1990 and 2001 respectively. He was an instructor in the various universities