

온습도에 따른 대중의 감성(감정+감각) 활동 변화

양중기*, 김근영**, 이영호**, 강운구**
가천대학교 일반대학원 IT융합공학과*, 가천대학교 컴퓨터공학과**

A change of the public's emotion depending on Temperature & Humidity index

Junggi Yang*, Geunyoung Kim**, Youngho Lee**, Un-Gu Kang**

Dept. of IT Convergence Engineering, Gachon University*

Dept. of Computer Science, Gachon University**

요약 소셜 미디어 데이터를 통해 파급되는 형태를 분석하여 국내 외 정치, 경제, 보건, 사회 문화현상을 대응하고자 하는 연구가 활발히 진행 중이다. 본 연구는 한국인이 가장 많이 사용하는 검색 서비스인 검색 정보를 알 수 있는 네이버 트렌드와 소셜 데이터인 네이버 블로그, 네이버 카페와 Open Data(API)를 사용하고 기상청의 온도, 습도 데이터를 사용하였다. 사람의 감성을 나타내는 감정 어휘와 감각을 표현하는 감각어휘 중 미각 어휘를 분석하여 대중의 감성 활동 변화를 연구하였다. 적합도 검증과 계층적 군집분석으로 군집의 개수를 정하여 비 계층적 군집분석으로 군집화 하였다. 군집분석 결과 8개의 군집으로 군집화되어 감성어휘를 알 수 있었다. 판별분석에 의하면, 군집분석에서 결정된 8개의 그룹은 98.9% 정확성을 갖는 것으로 나타났다. 본 연구에서 연구한 감성 활동 변화는 온도와 습도에 의해 감성 활동을 예측 할 수 있어 감성을 공유하고 대중의 기분을 파악하여 서로 공감대를 형성 할 수 있다.

주제어 : 소셜 미디어, 소셜 네트워킹 서비스, 소셜 데이터, 오픈 API, 오픈 데이터

Abstract Many researches about the effect on politics, economics and Sociocultural phenomenon using the social media are in progress. Authors utilized NAVER Trend most famous web browsing service in korea, NAVER Blog social media, NAVER Cafe service and Open Data(API) and also used temperature, humidity index data of Korea Meteorological Administration. This study analyzed a change of the public's emotion in korea using Cluster analysis of vocabulary of taste among its of feelings and senses. K-means clustering was followed by decision of the number of groups which was used Chi-square goodness of fit test and ward analysis. Eight groups was made and it represented sensitive vocabulary. By Discriminant analysis, eight groups decided by Cluster analysis has 98.9% accuracy. The change of the public's emotion has capability to predict people's activity so they can share sensibility and a bond of sympathy developed between them.

Key Words : Social Media, Social Networking Service, Social Data, Open API, Open Data

* 본 논문은 2014년 미래창조과학부 사회문제해결형 기술개발사업의 일환으로 수행하였음

Received 29 June 2014, Revised 27 July 2014

Accepted 20 October 2014

Corresponding Author: Un-Gu Kang(Gachon University)

Email: ugkang@gachon.ac.kr

© The Society of Digital Policy & Management. All rights reserved. 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.

1. 서론

문화와 경제는 고유가, 고물가, 이상기후 등 사회적인 현상들과 맞물려 새로운 경험들을 생산해 내고 파편화된다. 기업이 마케팅 활동을 통하여 고객을 끌어 모으는 구조가 아닌 소비자의 입맛에 맞는 맞춤형 시장이 목소리를 높여 가고 있는 것은 바로 소셜 미디어의 혁신 때문이다. 소셜 미디어란 스마트 IT 혁명에 맞는 스마트폰, 태블릿 PC, 클라우드 네트워킹 등 SNS(Social Network Service)로 대변되는 시간과 공간의 제약 없이 실시간으로 정보를 생산해내고 다양한 경로로 공유하는 미디어를 말한다. 사회적인 사건·사고의 양과 질에 따라 생성되는 많은 이슈들은 소셜 공간을 떠돌아다닌다. 특히 그 중에서도 사람들의 먹거리에 대한 이슈는 우리의 건강에 직결되므로 매우 조심스럽게 처리되어야 함에도 불구하고 그 이슈의 전달은 이슈의 본질의 여부와 관계없이 기하급수적으로 확대 재생산된다. 최근 홈페이지 게시판에는 뉴스의 정보를 빠르게 전달할 수 있는 SNS 버튼을 마련하여 현재 이슈가 될 수 있는 글을 버튼 클릭으로 쉽게 전달할 수 있기에 그 파급효과는 매우 크다[1].

소셜 미디어에서 얻어지는 데이터는 크기, 속도, 형태를 분석할 때 그 의미가 있는 빅 데이터다. 빅 데이터에 대한 관심이 최근에 급격히 증가하면서 이에 대한 검색, 저장, 분석을 하고자 하는 시도가 늘고 있다[2]. 빅 데이터의 정의를 보면 기존 방식으로 저장, 관리, 분석하기 어려운 큰 규모의 데이터를 총칭하는 말이다. 다른 한편으로 빅 데이터는 데이터의 크기, 데이터가 유입되는 속도, 유입되는 데이터의 형태로 정의되어지기도 한다. 즉, 정보의 양이 얼마나 큰지, 실시간 스트림을 통한 데이터인지, 정형화된 데이터인지 비정형화된 데이터인지 또는 이 모든 게 합쳐진 데이터형태 등을 통해 빅 데이터를 정의하기도 한다[3]. 이러한 형태와 크기가 다양한 빅데이터의 활용은 새로운 가치 창출의 기회가 되고 있다[4]. 최근 데이터를 통해 파급되는 형태를 분석하여 국내 외 정치, 경제, 보건, 사회 문화현상을 대응하고자 하는 연구가 활발히 진행 중이다[5]. 수천억건의 국민건강보험공단 각종정보들에 대한 빅데이터 활용 연구가 활발하게 진행 중이다. 미국에서는 보험상품의 보조금 수령인과 서비스 공급자를 파악하는 빅데이터 분석을 통해 보험사기 범죄를 예방하는 시도가 있었다[6]. 정책당국은 빅데이터 분

석으로 교통체증이 심한 지역의 교통량을 예측하기도 한다.

기상청에서는 기상 자료 데이터를 바탕으로 생활에 직접 활용 가능한 생활기상지수를 개발하여 제공하고 있다. 이는 우리나라 전역에 분포하고 있는 관측지점에 대하여 실시간 관측 데이터와 수치예보 자료를 이용하여 사람의 불쾌함과 관련된 감성 상태를 예측하거나, 식중독 발생 가능성, 식품의 부패 정도 등과 같은 식이 상태를 예측하여 제공해주는 지수이다[7].

본 연구에서는 2013년 1월부터 12월까지 기상청 Open Data를 활용하여 획득한 기상 자료 데이터와 한국인이 가장 많이 사용하는 검색 서비스인 검색 정보를 알 수 있는 네이버 트렌드와 소셜데이터인 네이버 블로그, 네이버 카페, 공공데이터인 Open API, Open Data에 사람의 감성을 나타내는 감성 어휘와 미각을 표현하는 미각 어휘를 활용하여 관련성을 파악하였다. 소셜 미디어를 분석한 소셜 분석을 활용해 얻을 수 있는 버즈 모니터링(Buzz Monitoring), 감성 분석(Sentiment Analysis)하여 대중 분석 및 메시지를 분석하여 새로운 의미를 찾는 소셜 분석을 통해 사회적 관심이 나타난 시기, 증폭된 시기, 사회적 관심도의 변화가 일어나는 시기의 위험요인들의 의미를 분석하여 얻을 수 있는 상관관계를 나타내어 한국인의 감성 활동 변화를 연구하였다[8].

본 연구에서 개발된 감성 활동 변화는 평균기온, 상대 습도에 따라 감성을 예측 할 수 있어 선호하는 감성을 공유하여 대중의 기분을 파악 할 수 있고 감성을 공유하고 대중의 기분을 파악하여 서로 공감대를 형성 할 수 있을 것이다.

2. 선행연구

2.1 소셜 분석 관련 연구

빅 데이터 시대를 맞이하여 데이터 이슈 탐지 및 예측 분석 기술이 큰 주목을 받고 있다. 이와 관련된 연구가 국내·외로 활발하게 진행 중이다. 이러한 소셜 미디어 분석을 통한 효과가 입증되면서 다각적으로 소셜 미디어 분석을 통한 이슈 탐지의 중요성이 더욱 커지고 있다.

2012년에는 숭실대학교 대학원에서 기업이 마케팅하는데 있어서 소셜 미디어 분석을 활용할 수 있는 구체적

인 방법으로 SNS를 마케팅에 이용하는 기업들을 대상으로 SNS 고객 분석 모형과 캠페인 대상이 되는 고객을 최적으로 선정하는 모델을 제시했다[2].

한국보건교육·건강증진 학회에서 소셜 데이터인 2011년 1월 1일부터 2011년 12월 31일까지 언급된 자살 관련 토픽(스트레스, 음주, 운동, 자살)을 추출하여 오픈라인 통계데이터의 연결을 통하여 우리나라 자살감색의 예측요인을 다변량 분석을 통해 연관성을 파악, 자살 예측 시스템을 개발하였다[9].

2013년에 동의대학교 대학원에서 2011-2012 국내 남자 프로배구 경기의 문자중계데이터와 소셜 네트워크 분석과 텍스트 마이닝을 활용하여 국내 남자프로배구 구단의 공격, 패스 패턴을 찾아내고, 배구 경기력과 관련된 핵심 키워드를 추출하여 경기력을 평가하고 구단의 경기진력을 상승시키기 위한 연구를 진행하였다[10].

최근에는 국내 개봉한 영화들 중에서 공식 SNS사이트를 보유한 영화와 그렇지 않은 영화의 관객추세 분석을 토대로 소셜미디어가 잠재 고객에게 감성적 도구로 활용되었는지 연구하였다[11].

2.2 감성 분석 관련 연구

SNS보급에 따른 인간의 감성을 이용한 감성마케팅 사례가 늘어가고 있다. 감성(感性)이란 자극이나 자극의 변화를 느끼는 성질이나 이성(理性)에 대응되는 개념으로, 외계의 대상을 오관(五官)으로 감각하고 지각하여 표상을 형성하는 인간의 인식 능력으로 정의한다. 2007년 홍익대학교 국제디자인전문대학원에서 무분별하게 사용되고 있는 감성과 감성관련 용어의 문제점을 지적하고 그 개념을 재 정의하는 것을 목적으로, 감성 연구에서 주로 사용하고 있는 감성 어휘를 영어와 국문학 관점에서 제시된 형용사 분류 체계와의 관계 속에서 감성 어휘 분류 시스템을 구축 하였다[12].

상해교통대학 공업디자인학과에서 디자인의 일부분으로써 한글을 사용할 때 다양한 한글글꼴에 대한 감성과 선호도에 대한 객관적인 정보를 제공하여 한글글꼴을 효과적으로 사용하기 위해 방향을 제시하였다[13].

서울대학교 언론정보 연구소에서는 한국인이 자신의 정서 표현에 자주 사용하는 단어를 기본 목록으로 삼아 개인마다 정서의 기본 범위를 확인하고 정서의 기본 구성 차원을 밝히는 연구를 진행하였다[14].

소리 정보중에 특정 주파수대역의 볼륨 값을 활용하여 콘텐츠 안에 감성지수를 추출하는 방법에 대한 연구를 진행하였다[15]. 또한 컨설팅을 받은 경험이 있는 기업의 관리자급들의 수백건의 사례를 분석하여 컨설턴트의 감성지능이 컨설팅서비스 품질에 어떠한 영향을 주는 지 분석하는 연구를 진행하였다[16].

2.3 감성과 감각(미각) 관련성 관련 연구

음식에 의해서 감정이 결정되고, 감정이 먹을 것을 결정한다는 관점이 이슈 되기 시작하면서 관련된 연구가 진행되고 있다. 사람이 음식을 선호하고 레스토랑을 선택하는 각각의 속성, 그리고 식사량 변화에 미치는 영향을 조사하고 이러한 음식 선호와 레스토랑 선택에 대한 감정의 영향이 차이가 있는지를 파악하였다[17].

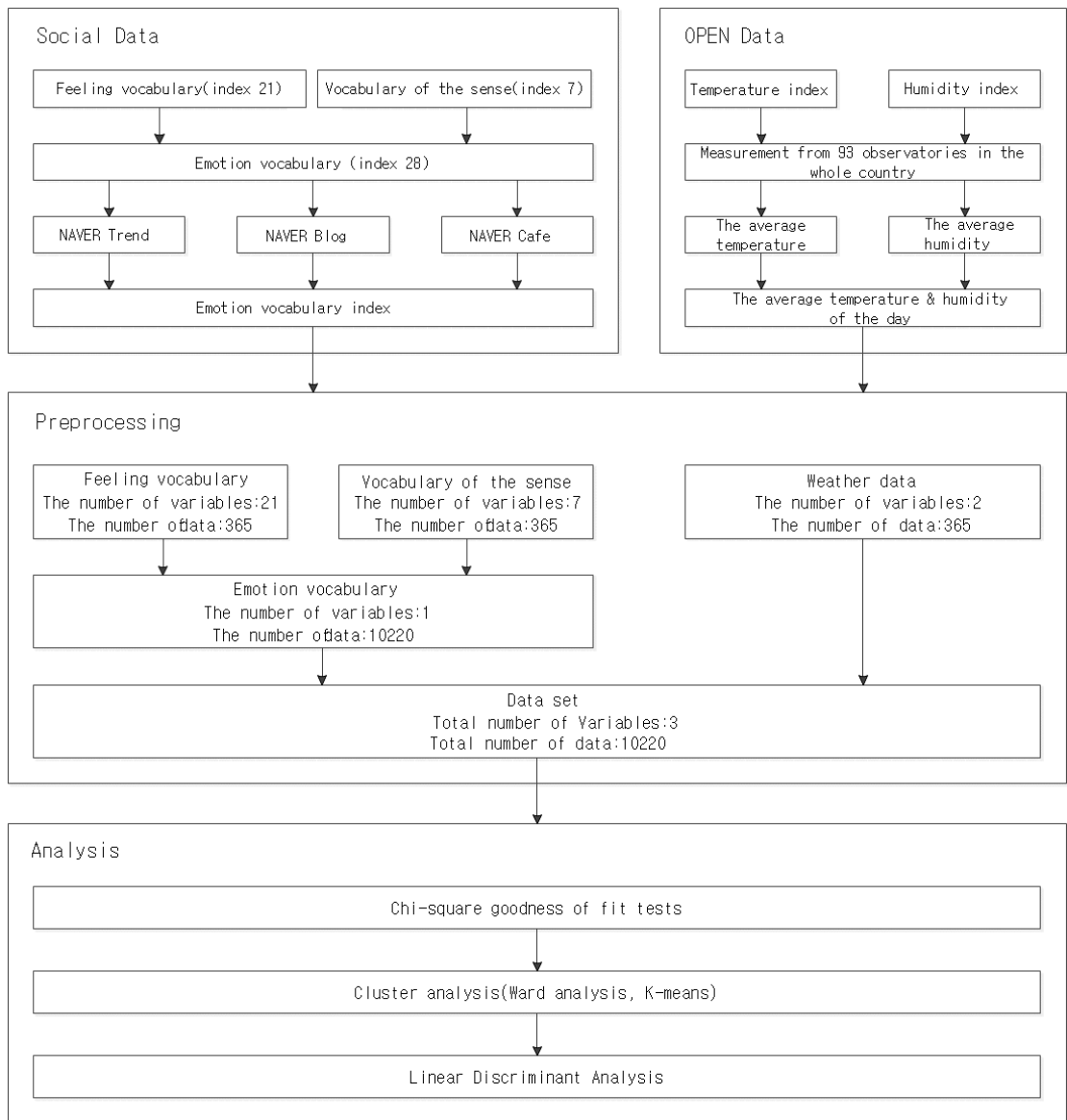
2012년에 한국 정보 과학회에서 감성분석 기반 음식 추천 어플리케이션 설계를 위한 연구를 진행하였다. 기존처럼 현재위치로부터 가까운 거리로 추천하거나 인터넷 블로그 등에 소개되었던 맛집과 같은 정보를 활용하여 음식을 추천하는 방식이 아니라 개인의 기분, 체중, 현재의 날씨와 위치 같은 개인적인 정보와 함께 음식에 관한 선호도 조사와 요인분석을 통해 만들어진 감성지도를 기반으로 여러 상황에 맞는 음식을 추천한다[18].

트위터로부터 사용자 프로파일을 이용하여 개인 맞춤형 음식 추천 시스템을 연구했다. 사용자가 사용한 개인의 트위터 문장에서 명사를 추출하고, 감정단어와 공기여부에 따라 감정점수를 계산하고 프로파일을 생성한다. 음식에 대한 정보는 웹 검색을 통해서 검색한 페이지를 분석하고 가공하여 음식마다 대표 명사 목록을 만들어 추천한다[19].

3. 연구방법

3.1 데이터 수집 및 전처리

본 연구는 2013년 1월 1일부터 2013년 12월 31일까지 1년 동안의 감성, 감각(미각) 한글 Keyword 데이터를 수집하였다. 감성을 표현할 수 있는 감정 어휘 21개와 감각을 나타내는 미각 어휘 7개를 사용하여 총 28개의 감성키워드를 네이버 트렌드(PC, 모바일)의 검색 버즈량, 네이버 블로그, 네이버 카페에 게시된 글의 한글 Keyword를



[Fig. 1] Structure map of a change of the public's emotion

Count하였다.

감성 어휘 Count값을 동일한 가중치를 주기 위해 1~10의 값으로 Level화 하였다. 이를 통해 네이버 트렌드, 네이버 블로그, 네이버 카페 변수는 각각 1:1:1비율로 동일한 가중치 값을 갖는다. Keyword를 Count 하여 수집한 시기 이후 일부 사용자가 글을 삭제하거나 수정할 수 있으므로 미세한 데이터의 변동이 있을 수 있다.

기상청 기상 자료는 기상청에서 제공하는 OPEN DATA를 이용하여 기온과 상대습도 데이터를 수집하였다. 이를 전국 93개 관측 값을 기준으로 평균화 하여 일별 평균 기온 및 습도 데이터로 변환하였다.

[Fig. 1]는 감성 활동 변화를 위한 데이터 습득, 전처리 과정과 분석 진행의 전체 구조도이다.

<Table 1> Searching feeling vocabularies keyword

Group	Feeling vocabulary	Blog Keyword Count	Cafe Keyword Count	NAVER Trend Count (PC+Mobile)
Happiness	Enjoyable	199,544	52,334	52,522
	Happy	109,889	20,745	54,432
	Wonderful	97,899	19,203	54,631
Anger	Furious	13,902	3,326	700
	Distressed	16,350	3,427	49,396
	Blame	177,045	68,187	48,988
Sadness	Unfair	200,496	104,663	55,134
	Disappointed	600,476	220,252	56,901
	Frustrated	195,182	82,612	46,536
Fear	Scary	352,183	124,598	1,946
	Nervous	578,340	578,340	51,926
	Embarrassed	628,634	304,998	54,705
Favorance	Like	1,986,707	393,945	55,457
	Impressed	10,484,778	349,539	45,256
	Boast	1,382,546	630,181	56,357
Aversiveness	Hate	224,938	97,413	12,437
	Inconvenient	1,389,814	981,560	55,347
	Doubt	624,069	310,209	53,201
Hope	Wish	259,845	105,965	10,737
	Complain	639,306	273,272	48,617
	Conflict	300,333	97,041	39,472

<Table 2> Searching vocabularies of the sense keyword

Group	Vocabulary of the sense	Blog Keyword Count	Cafe Keyword Count	NAVER Trend Count (PC+Mobile)
A salty taste	Salty	27,506	6,041	50,409
A bitter taste	Bitter	176,116	122,748	52,587
A vinegary taste	Vinegary	92,945	61,184	47,121
Sweet taste	Sweet	47,822	10,412	57,169
Astringent taste	Astringe	682	125	46,656
Hot taste	Spicy	26,904	4,376	50,627
Bland taste	Bland	6,640	797	52,544

취합된 데이터로 적합도 검증(Chi-square goodness of fit test)과 계층적 군집분석(Ward analysis)으로 군집의 개수를 정하여 비 계층적 군집분석(K-means)으로 군집화 하였다. 이를 판별분석(Linear Discriminant Analysis)으로 검증하였다.

본 연구에서 사용한 감정을 정의하는 어휘 21가지는 형용사 분류 체계에 따라 나뉜 감각 어휘, 감정 어휘, 평가 어휘, 존재 어휘 중 감정 어휘 110개를 토대로 정의하였다[12]. 110개의 감정 어휘는 기쁨, 노여움, 슬픔, 두려움, 좋아함, 싫어함, 바쁨 총 7가지로 분류 되며, 각 분류 별로 소셜 미디어(네이버 트렌드, 네이버 블로그, 네이버 카페)에서 가장 많이 사용하는 상위 3가지의 Keyword로 정의하였다. 감정 어휘 검색 Keyword 중 2013년 1월 1일

부터 12월 31일까지 네이버에서 가장 많이 사용하는 각 감정을 대표하는 감정 어휘 상위 3가지는 <Table 1>과 같다.

감각 어휘는 미각 평가 어휘 중 기본 미각어를 사용하여 사람이 느끼는 짠맛(짜다), 쓴맛(쓰다), 신맛(시다), 단맛(달다), 뚝은맛(뚝다), 매운맛(맵다), 싱거운맛(싱겁다) 7개를 토대로 정의하였다. 정의된 기본 미각어의 수집된 데이터 수는 <Table 2>와 같다.

네이버트렌드, 네이버 블로그, 네이버 카페를 이용해 수집한 총 28개의 어휘와 기상청의 일별 평균 온, 습도 데이터를 바탕으로 데이터에 대한 통계적 검증을 위해 적합도 검증을 진행하였다. 적합도 검증은 어떤 조건에서 기대되는 빈도에 관측빈도가 얼마나 적합한지에 대한

검증 방법이다. 28개 변수에 대한 적합도 검증 결과는 <Table 3>와 같다.

검증 통계량의 카이제곱값은 197.233~730.066의 값을 가지며, P-value=.000이므로 수집된 데이터의 각 변수 별 버즈량의 차이가 있는 것으로 나타났다.

검증된 데이터를 바탕으로 데이터 분석을 위한 데이터 전처리 과정을 거쳤다. 감정어휘와 감각 어휘에 공통 특성에 대한 케이스 설명 기준 변수를 만들기 위해 2가지 어휘 변수를 하나로 통합하였다. 통합된 감성 활동 어휘는 총 10,220건으로 28가지 어휘 각각 365개의 1년의 데이터가 존재하게 된다. 통합된 1년의 감성 활동 어휘와 기상 데이터인 온, 습도 데이터를 합쳐 최종 데이터 셋을 구성하였다. 총 10,220건으로 3가지의 변수인 감성 활동 어휘 Count값과 온도와 습도로 구성된다.

<Table 3> Goodness of fit for 35 variables

Variables	Chi-square test	Degree of freedom	Approximation of the significant probability
Enjoyable	330.392 ^a	5	.000
Happy	219.625 ^b	3	.000
Wonderful	482.052 ^a	5	.000
Furious	709.690 ^a	5	.000
Distressed	380.658 ^c	4	.000
Blame	290.822 ^c	4	.000
Unfair	243.822 ^b	3	.000
Disappointed	299.164 ^b	3	.000
Frustrated	472.258 ^d	6	.000
Scary	419.589 ^f	4	.000
Nervous	357.679 ^a	5	.000
Embarrassed	197.233 ^c	4	.000
Like	362.315 ^a	5	.000
Impressed	459.564 ^a	5	.000
Boast	276.384 ^c	4	.000
Hate	599.907 ^d	6	.000
Inconvenient	231.836 ^c	4	.000
Doubt	207.068 ^c	4	.000
Wish	730.066 ^c	8	.000
Complain	347.781 ^c	4	.000
Conflict	232.984 ^f	7	.000
Salty	516.866 ^d	6	.000
Bitter	310.301 ^c	4	.000
Vinegary	396.274 ^c	4	.000
Sweet	519.630 ^a	5	.000
Astringe	301.526 ^a	5	.000
Spicy	262.384 ^c	4	.000
Bland	373.953 ^a	5	.000

3.2 군집 분석(Cluster Analysis)

본 연구에서는 적절한 군집의 수를 도출하기 위해 사전에 계층적 군집분석(Ward analysis)을 실시하였다. 계층적 군집 분석은 군집내의 관측치들은 가급적 유사하고, 상이한 대상들을 그룹핑하는 방법으로 본 연구에서는 군집분석 중에서 가장 널리 사용하고 있고 유사성 거리측정 방법에 따라 다양한 기법들을 비교 분석하기 위해 군집들간의 거리(Euclidean Distance)를 이용하여[20] 가장 가까운 군집거리 단계적으로 결합하여 최종적으로 모든 관측치가 한 군집으로 형성하게 하는 병합적 계층 군집분석(Agglomerative hierarchical clustering analysis)으로 군집을 형성하였다.

도출 되는 군집 결과를 토대로 비 계층적 군집분석(K-means)방법으로 온, 습도에 따른 감점, 감각(미각) 키워드간의 유사한 동류 집단으로 분류를 하기 위해 K-means 방식의 군집 분석을 진행하였다. 이 방식은 관측 자료들을 겹치지 않는 상호 이질적인 몇 개의 군집으로 구분하는 집락분석방식으로, 몇 개의 군집(K)으로 구분할 것인지를 지정해 주고 반복적인 과정을 거쳐 K개의 군집 해를 찾게 된다[21]. 각 자료는 그 평균값이 가장 가까운 군집으로 할당되며, 이렇게 생긴 새로운 군집의 평균은 나머지 자료들이 할당되는 기준으로 사용된다. 그리하여 모든 자료들의 군집이 바뀌지 않게 될 때 K 집단이 정의된다[22].

집단의 수와 구조에 대한 가정이 없으며 오직 대상들 사이의 유사성, 근접성에 근거하여 군집을 형성하고 공통자질을 파악하여 군집들 사이의 관계를 파악하기 케이스 설명 기준 변수는 감성 어휘 Count값으로 하고 반복적으로 탐색적인 분석을 진행하였다. 그 중 8개의 군집 해가 온, 습도의 구분 값이 가장 뚜렷하고, 각 군집 별 감성 변화의 구분이 명확하였다.

3.3 판별 분석(Linear Discriminant Analysis)

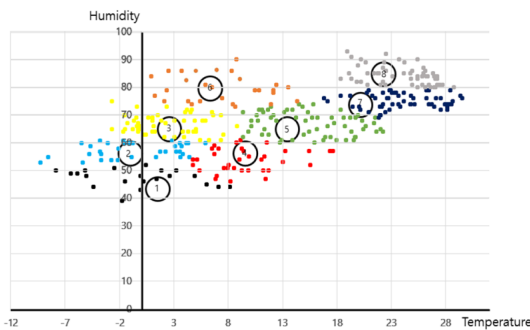
본 논문에서 판별분석은 군집분석의 정확성을 평가하는 것이며, 정성적인 종속변수와 일련의 독립변수가 판별분석을 위하여 필요하다. 해당 모델에 대한 집단의 분류가 제대로 됐는지를 파악하기 위해 판별분석을 수행하여 검증하였다.

Fisher[23]에 의해 처음 도입된 판별분석은 클래스간 분산과 클래스내 분산의 비율을 최대화하여 사회현상의

여러 특성들을 토대로 하여 주어진 상황에서 응답자들이 어떻게 행동할 것인지를 예측하는 하나의 통계기법이다. 판별분석은 하나의 집단변수(Group variable) 개개의 변수값을 판별하는데 중요한 역할을 하는 판별변수(Discriminant variable)의 선형결합(Linear combination)인 판별함수(Discriminant function)를 만든 후, 이 판별함수에 새로운 개체의 특성을 대입하여 어떤 집단에 속할지를 판별하는 것이다. 판별분석은 두 개 이상의 모집단에서 추출된 다변량 관측치의 정보를 이용하여 관측치가 어느 모집단에서 추출된 것인가를 결정해 줄 수 있는 기준을 찾는 분석으로, 이를 통해 군집분석 결과를 검증할 수 있다.

4. 결과

온, 습도에 따른 감정 키워드를 계층적 군집분석(Ward analysis)으로 군집의 개수를 정하여 비 계층적 군집분석(K-means)으로 군집한 결과 8개의 군집으로 분류 되었다. 이와 관련하여 온습도에 따른 감정 키워드 군집된 8개의 분포는 [Fig. 2]와 같다.



[Fig. 2] Keyword of emotion clustering by temperature and humidity

X축과 Y축은 각각 온도와 습도를 나타내며 온, 습도에 따른 군집 별 분포를 나타낸다. 군집 1~8까지로 총 8개의 군집을 표현한다. 8개의 군집은 온, 습도에 따른 감정 키워드간의 유사한 동류 집단으로 분류 되어 있으며, 각 군집 별 온, 습도 범주, 감정 키워드는 다른 특성을 지니고 있다. 각 군집 범주에 따른 온, 습도 범주 및 특성 키워드는 <Table 4>과 같다.

<Table 4> Temperature of 8 kinds of cluster and humidity index & keyword

Cluster(K)	Temperature (°C)		Humidity (%)		Keyword	
	Minimum value	Maximum value	Minimum value	Maximum value	Keyword name	%
1	-7.9	8.1	39	52	Sweet	60
2	-9.3	5.9	53	61	Furious	20
3	-2.7	8.8	61	75	Scary	40
4	4.7	17.6	44	61	Embarrassed	60
5	9.3	22.2	57	74	Hate	40
6	0.9	14.4	73	91	Unfair	80
7	16.8	29.5	69	79	Scary	40
8	15.5	28.7	80	93	Disappointed	80

8개의 군집 별로 서로 다른 감정, 감각 증상을 보인다. 군집 1은 100명중 60명은 “달다”라는 감성을 느끼는 날이고 군집 8은 100명중 80명은 “실망하다”라는 감성을 느끼는 날이다. 군집분석 결과 제대로 군집화 됐는지를 파악하기 위해 판별분석을 수행하여 검증하였다.

<Table 5> Wilks' lambda and significant probability

	Wilks' lambda	F	df1	df2	Sig.
Temperature	.122	10509.370	7	10212	.000
Humidity	.093	14170.653	7	10212	.000

<Table 5>는 집단 간 차이가 있는 변수들이 판별분석이 이용되므로, 이를 위하여 집단평균의 동질성 검정을 실시하였다. 검정 결과 유의확률은 0.000으로 변수들의 평균차이가 유의하게 나타났다. 습도는 온도에 비해 Wilks' lambda 값이 작고 분산분석 F값은 크다. 이는 온도에 비해 습도 요인이 판별력이 높음을 알 수 있다.

<Table 6> Eigen value and canonical correlation

Function	Eigen value	Variance %	Cumulant %	Canonical correlation coefficient
1	16.795 ^a	89.0	89.0	.971
2	2.083 ^a	11.0	100.0	.822

<Table 6>에서 정준상관계수를 제공하면 (.971)²와 (.822)²로 이는 판별점수 분산의 94.28%와 67.56%로 변수들에 의해 설명됨을 의미한다. 정준상관계수는 0.971과

0.822로 1.0에 가까우므로 판별함수의 판별력이 매우 우수한 것으로 나타났다.

<Table 7> Result of the discriminant analysis by emotion keyword depending on temperature and humidity

Cluster	Predicted affiliated group								Total	
	1	2	3	4	5	6	7	8		
	1	896	0	0	28	0	0	0	28	952
	2	0	784	0	0	0	0	0	0	784
	3	0	0	1680	0	0	0	0	0	1680
	4	0	0	0	1512	0	0	0	0	1512
	5	0	0	0	28	1764	0	0	0	1792
	6	0	0	0	0	0	1848	0	0	1848
	7	0	0	0	0	0	0	1036	28	1064
	8	0	0	0	0	0	0	0	588	588
%	1	94.1	.0	.0	2.9	.0	.0	.0	2.9	100.0
	2	.0	100.0	.0	.0	.0	.0	.0	.0	100.0
	3	.0	.0	100.0	.0	.0	.0	.0	.0	100.0
	4	.0	.0	.0	100.0	.0	.0	.0	.0	100.0
	5	.0	.0	.0	1.6	98.4	.0	.0	.0	100.0
	6	.0	.0	.0	.0	.0	100.0	.0	.0	100.0
	7	.0	.0	.0	.0	.0	.0	97.4	2.6	100.0
	8	.0	.0	.0	.0	.0	.0	.0	100.0	100.0

군집의 판별적중률(Hit ratio)은 회귀분석의 적합도를 나타내는 r2의 개념과 유사한 것으로 판별식이 대상을 얼마나 잘 분류하는가를 나타내는 정도라고 할 수 있다.

<Table 7>과 같이 군집1의 판별적중률은 94.1%, 군집 7의 판별적중률은 97.4로 군집1과 군집 7을 제외한 군집의 판별적중률은 100.0%이며, 전체 군집분석 분류 적중률은 98.9%로 올바르게 분류되었다.

5. 결론

본 연구는 기상 데이터(온도, 습도)를 이용해 대중의 감성과 다양한 감성 욕구를 파악하여 감성 활동을 예측할 수 있어 감성을 공유하고 대중의 기분을 파악하여 서로 공감대를 형성할 수 있고 특별한 일이나 기념일 등에 최적화된 일자를 선별하여 이벤트를 할 수 있을 것이다. 또한 기상청 API를 사용하여 온도, 습도에 대한 데이터를 실시간으로 수집이 가능하고 이에 맞춰 실시간으로 업데이트된 감성 활동 변화 및 시각화 된 서비스 모델을 제공할 수 있다.

감성이라는 용어의 의미는 단순히 신체 반응만을 동반하는 감정에 비해 직관적, 무의식적, 복합적인 감정으로 의미가 확대된 개념이라 할 수 있다. 향후 연구에서 더 많은 요인을 추가한다면 다양한 환경에서 감성에 더욱 적합하게 대응할 수 있을 것이다. 본 논문은 1년간의 데이터로 분석을 하였지만 계속 축적되는 Social 데이터로 분석을 하면 대중의 감성을 좀더 명확히 파악할 수 있을 것이다. 또한 자연어 처리로 Keyword의 의미를 정확히 한다면 Social 데이터의 정확한 의미를 파악하여 데이터의 질을 높일 수 있을 것이다

ACKNOWLEDGMENTS

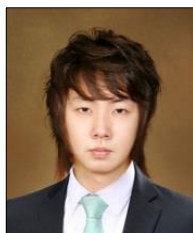
This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP)(NRF-2013M3C8A2A02078403).

REFERENCES

- [1] Hwang, Yun Chan, KOH, Chan. "Analysis of Opinion Social Data on the SNS (Social Network Service) by Analyzing of Collective Damage Reply." The Society of Digital Policy & Management 11.5 (2013): 41-51.
- [2] Choi, Sungbin, et al. "Semantic concept-enriched dependence model for medical information retrieval." Journal of biomedical informatics 47.(2014): 18-27.
- [3] Han Sowol, Lee Minsu. "A Big Data Model for Social Information Recommendation Techniques." Korea information science society: database 39.6 (2012): 380-386.
- [4] Lee Seonghun and Lee Dongu. "Current Status of Big Data Utilization." The Korea Society of Digital Policy and Management 11.2 (2013): 229-233.
- [5] Jung-Gi Yang, Jae-Kwon Kim, Un-Gu Kang, Young-Ho Lee, "Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis", Pers

- Ubiquit Comput, 17.7 (2013): 1315-1572
- [6] Noh Gyuseong. "A Study on Utilization Strategy of Big Data for Local Administration by Analyzing Cases." The Korea Society of Digital Policy and Management 12.1(2014): 89-97.
- [7] "Development of weather index Korean life." National Institute of Meteorological Research, (2009)
- [8] Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." Journal of Computational Science 2.1 (2011): 1-8.
- [9] Song Taemin, et al. "Multivariate Analysis of Factors for Search on Suicide Using Social Big Data." Korean Society for Health Education 30.3 (2013): 59-73.
- [10] Kang Byeonguk, "Performance analysis of volleyball games using the social network and text mining techniques." Dong-Eui University. (2013)
- [11] Song Myeongbin, Lee Sangho. "A Study on the Policy Implication on the Relation of Social Media & Movie industry : Focusing on Emotional Perception & Audience Trends." The Korea Society of Digital Policy and Management 12.1 (2014): 295-303.
- [12] Na Geon. "A Study on the Meaning of Sensibility and Vocabulary System for Sensibility Evaluation." Journal of Ergonomics Society of Korea 26.3 (2007): 17-25.
- [13] Yoon Hyeonggeon. "Sensitivity Analysis on Korean Fonts Between Korean and Chinese." Korea Society for Emotion and Sensibility 14.4 (2011): 637-644.
- [14] Lee Junung, et al. "Classification of Emotion Terms in Korean." Korea Communication Association 52.1 (2008): 85-116.
- [15] Kwon Yeonghun, Chang Jaegeon. "Emotion Extraction of Multimedia Contents based on Specific Sound Frequency Bands." The Korea Society of Digital Policy and Management 11.11 (2013): 381-387.
- [16] Kim Duyeol, Lee Seongyu, Kang Eungu. "A Study on the Relationships between Emotional Intelligence of Consultant and Consulting Service Quality." The Korea Society of Digital Policy and Management 11.7 (2013): 41-50.
- [17] Lee Eunyeong. "Effect of Emotions on the Food Preference and Restaurant Selection of Female University Students." Ewha Womans University. (2008)
- [18] Lee Ganghyeon, et al. "A Design of Food Recommendation Application Based on Kansei Analysis." Korea Information Science Society (2012): 528-530
- [19] Jung Juseok, Kang Sinjae. "Hybrid Food Recommendation System Using Auto-generated User Profiles." Journal of Korean Institute of Intelligent Systems 21.5 (2011): 609-617.
- [20] Eisen, Michael B., et al. "Cluster analysis and display of genome-wide expression patterns." Proceedings of the National Academy of Sciences 95.25 (1998): 14863-14868.
- [21] Kang Eunjeong. "Clustering of Lifestyle Behaviors of Korean Adults Using Smoking, Drinking, and Physical Activity." KIHASA 27.2 (2007): 44-66.
- [22] Huang, Zhexue. "Extensions to the k-means algorithm for clustering large data sets with categorical values." Data Mining and Knowledge Discovery 2.3 (1998): 283-304.
- [23] Fisher, Ronald A. "The use of multiple measurements in taxonomic problems." Annals of eugenics 7.2 (1936): 179-188.

양 중 기(Yang, Junggi)



- 2011년 2월 : 가천의과학대학교 의료공학(공학사)
- 2013년 2월 ~ 현재 : 가천대학교 IT융합공학과 석사 과정
- 관심분야 : 데이터마ining, 데이터분석, 유헬스케어
- E-Mail : jygang@gachon.ac.kr

김 근 영(Kim, Geunyoung)



- 2008년 2월 ~ 현재 : 가천대학교 정보공학부 재학중
- 관심분야 : 정보 시각화
- E-Mail : presskgy@naver.com

이 영 호(Lee, Youngho)



- 1996년 2월 : 한국외국어 대학교 응용전산학과 (이학석사)
- 2005년 8월 : 아주대학교 의과대학 의료정보학과 (이학 박사)
- 1999년 2월 ~ 2002년 1월 : IBM Korea BI&CRM EM
- 2002년 2월 ~ 현재 : 가천대학교 IT대학 컴퓨터공학과 부교수
- 관심분야 : 데이터 마이닝, 모바일 헬스케어
- E-Mail : lyh@gachon.ac.kr

강 운 구(Kang, Un gu)



- 2001년 2월 : 인하대학교 전자계산공학과 (공학박사)
- 1994년 2월 ~ 현재 : 가천대학교 컴퓨터공학과 교수
- 2012년 3월 ~ 현재 : u-헬스케어 연구 소장
- 관심분야 : u-헬스케어, 메디컬인포매틱스 의료 IT융합
- E-Mail : ugkang@gachon.ac.kr