

사회연결망상의 우위와 감성 표현과의 관계 분석: 알츠하이머 웹포럼의 적용

이민정*, 우지영**

The Analysis on Users' Centrality in the Social Network and their Sentiment : Applying to Medical Web Forum on Alzheimer's Disease

Min-Jung Lee *, Ji-Young Woo**

요 약

본 연구는 알츠하이머에 대한 의료 웹포럼 콘텐츠의 감성 분석과 사회연결망 분석을 연계하여 감성을 표현하는 정도와 사회연결망상의 영향력 사이의 관계를 살펴보았다. 최근 사회가 고령화가 됨에 따라 치매환자는 증가하고 있는데 이에 환자뿐만 아니라 환자 보호자들은 질병에 관한 정보뿐만 아니라 위로를 얻기 위하여 의료 웹포럼을 이용하고 있다. 웹포럼상에서의 정보적/감성적 지지의 형태로 나타나는 사회연결망을 분석한 결과, 네트워크 중앙에 위치한 사용자의 감성표출이 다른 사용자의 감성 표출보다 높게 나타나는 것을 관찰하였다. 또한 긍정적 단어를 많이 사용하는 그룹이 부정적 단어를 많이 사용하는 그룹에 비해 연결중심성 및 보나시치중심성 값이 높게 나타난 것을 확인할 수 있었다. 본 연구는 웹포럼 상의 감성전달에 있어서 일반 사용자에게 큰 영향력을 줄 수 있는 오피니언 리더를 도출하고, 알츠하이머 환자의 우울증을 완화할 수 있는 정책의 일환으로 활용할 수 있을 것이다.

▶ Keywords : 감정 분석, 사회연결망분석, 알츠하이머, 의료 웹포럼

Abstract

In this study, we aim to analyze the relationship between the centrality in the social network and the sentiment of medial web forum users. In recent, many people use online resources to obtain health and wellness information especially social media resources. In the medial web forum, people give and receive informational supports and emotional supports and this interaction forms the social network. We analyze

• 제1저자 : 이민정 • 교신저자 : 우지영

• 투고일 : 2015. 4. 28, 심사일 : 2015. 5. 18, 게재확정일 : 2015. 6. 5.

* 세종사이버대학교 경영학과(Dept. of Business Administration, Sejong Cyber University)

** 고려대학교 정보보호대학원(Graduate school of information security, Korea University)

the social network, derive node characteristics in terms of centrality and compare the centrality index and the sentiment score derived from users' messages. We found that as more people express their emotion, they possess higher central position in the network. Further, people who express positive emotion in their messages have higher central position in the network than people who have negative emotion. This study will help to identify influentials of emotional supports to others and finally to control the depression of Alzheimer's disease patients and their related ones.

▶ Keywords : Sentiment analysis, Social network analysis, Alzheimer's disease, Medical web forum

I. 서 론

2015년 인터넷 서비스를 이용하는 사용자가 31억명 이상으로 세계인구의 40% 이상에 달한다[1]. PewResearch Center에는 2014년 1월 미국의 경우, 성인 87% 이상이 인터넷을 이용하는 것으로 보고하고 있다[2]. 이와 같이 인터넷 사용자가 급증함에 따라 웹포럼, 블로그, 위키, 트위터 등의 다양한 소셜네트워크 미디어 매체들은 서로간의 정보와 감성 공유를 목적으로 활발하게 사용되고 있다[3, 4]. 최근 고령화 사회로 진입하고, 개인의 삶에 여유가 생기면서 급속도로 성장하고 있는 산업이 헬스케어산업이다. 이와 같이 인터넷의 발전으로 인한 웹 2.0 기술과 의료산업에 대한 관심의 증대는 의료분야에서의 소셜미디어 활용을 촉진하고 있다.

Nolte 외[6]는 의료산업을 발전시키기 위해서 ICT(Information and Communications Technology)가 필요하고, ICT는 혁신적인 의료산업모델을 만드는데 도움을 줄 것이라고 보고했다. Elkin[7]은 성인 미국인들의 약 60%가 건강과 복지 정보를 얻기 위해서 온라인 정보를 사용한다고 했으며, 이때 웹 2.0 기반의 소셜미디어를 이용하는 사용자도 3분의 1 이상이라고 보고했다. 이태희 외[8]는 의료정보학의 대표적인 저널인 의료정보학저널(Journal of Medical Internet Research)을 이용하여 2011년부터 2013년까지 최근 연구동향을 파악했다. 의료 2.0(Medicine 2.0)라는 개념하에 의료산업과 관련한 소셜미디어에 대한 연구가 활발하게 진행되고 있었다. 의료 2.0은 소셜미디어를 이용하여 네트워킹, 참여, 협업, 공유 등을 통해 정보를 주고받는 웹기반 서비스이다. 의료서비스 제공자, 의료서비스 소비

자, 의료 연구자들, 의료산업 투자자들은 소셜미디어 매체인 페이스북, 트위터, 웹포럼, 위키 등을 이용하여 소통과 협업을 도모하고 있다. 이와 같은 의료분야의 소셜미디어를 이용한 의료정보 교환의 장점은 다음과 같다. (1) 의료 정보를 쉽게 획득하고 빠르게 확산시킬 수 있으며, (2) 소셜네트워크를 이용한 감정적 지지를 통해 환자나 환자보호자가 온라인 그룹으로부터 위로를 받을 수 있다[9].

소셜미디어는 유형별로 특성이 다른데[10], 블로그와 페이스북은 글 작성자 지향적인 소셜미디어로 작성자 위주로 글이 작성되기 때문에 콘텐츠가 개인적으로 흘러가기가 쉽다. 위키는 콘텐츠나 문서위주로 작성되어 중립적이고 객관적인 형태로 작성되며 따라서 전문적인 정보위주로 작성되기 때문에 질병에 있어서 감정적인 위로가 필요한 사용자가 사용하기에는 적합하지 않은 소셜미디어이다[5, 10]. 웹포럼의 경우는 특정 토픽에 대해서 의견을 공유하는 자리로 객관적인 지식의 공유뿐만 아니라 감정적인 위로나 공감이 이루어질 수 있다[11]. 특히 의료 웹포럼의 경우는 건강 및 생명이 직접적으로 연결되기 때문에 사용자는 질병관련 정보를 얻기 위해 매우 적극적으로 웹포럼을 이용하며, 이는 웹포럼 사용자간에 있어서 강한 상호작용으로 나타날 수 있다[12]. 건강과 질병을 다루는 의료 웹포럼에서 나타나는 상호관계는 감정적 지지와 정보적 지지가 대표적이다. 건강함을 유지하기 위해서 관련 의료정보를 찾는 행위는 대표적인 정보적 지지 형태라고 볼 수 있다. 이와 같은 정보적 지지 기능은 질병에 관련된 의료정보를 상호 전달함으로써 미지정보를 줄이게 되고, 이는 질병의 공포를 감소시킬 수 있다. 뿐만 아니라 의료 웹포럼은 감정적인 위로가 가능한 소셜미디어 형태의 웹기반 서비스이다. 이와 같은 의료 웹포럼은 의료서비스 제공자, 의료서비스 소비자, 의료연구자들, 의료산업투자자 등 의료산업 관련한

모든 사람들이 전문적인 주제에 대해서 논의하고, 의료산업의 수요자와 공급자의 니즈를 파악할 수 있는 장이라고 볼 수 있을 것이다.

국내 경우는 의료산업의 웹포럼에 대한 연구는 시작단계에 있다(11, 12). 본 연구에서는 의료 웹포럼들 중에서 치매(알츠하이머)에 관한 게시판을 대상으로 분석하고자 했다. 알츠하이머는 퇴행성 뇌질환으로 인지 기능이 저하되다가 성격변화, 초조행동, 우울증 등의 정신행동 증상이 동반되며, 말기에는 보행이상, 경직 등의 신체적인 합병증이 나타나게 되는 질병이다. 호발 연령은 65세 이후로 노인 인구에서 발생 확률이 높으며, 불치질환으로 분류되어 있다. 특히 최근에는 보통 60살 전후의 '젊은 치매'를 의미하는 초로기 치매 환자도 늘고 있다. 알츠하이머 등의 불치병이 발생 시 환자 본인을 포함한 보호자들이 정신적으로 고통을 받는 부분은 알츠하이머에 동반되는 정신행동증상으로 인해서이며, 치매환자를 포함한 보호자들의 자살 및 살해 등의 사건도 증가하고 있다. 치매는 사회가 책임져야 하는 질병이지만 우리 사회에서는 여전히 가족의 몫이 80% 이상이다(13). 따라서 치매에 걸린 환자와 환자보호자들은 웹포럼 게시판을 통해서 치매에 대한 경험적 지식공유 뿐만 아니라 감성적인 공감 및 위로를 통해 어려움을 극복하고 있다(11). 알츠하이머 환자 및 보호자의 어려움에 대해서 파악하고 이들의 감성상태를 관리하기 위해서는 의료 웹포럼 분석은 필요하다.

최근 소셜미디어를 이용한 감성분석이 많이 이루어지고 있는데 이득환(22) 외는 트위터를 감성분석을 했을 때 주성분 분석 결과 크게 긍정성과 부정성으로 감성이 묶일 수 있음을 밝혔다. 정희운 외(14)는 트위터를 이용하여 감성과 사회연결망의 영향력에 대해서 분석하였으나 이들의 연구는 트위터에 국한된 연구였다. 특히 의료분야에 감성분석이 적용된 연구는 국내에는 없는 실정이다.

본 연구에서는 의료 웹포럼을 이용하여 중대질병 환자 및 관계자들의 감성을 측정하고, 웹포럼 메시지의 게시글과 댓글과의 관계를 통한 사회연결망 분석을 실시하고자 한다. 사회연결망 상의 중심성과 감성과의 관계를 파악하여, 중대질병 환자의 감성을 관리할 수 있는 정책의 일환으로, 웹포럼 상의 정보 및 감성 전달에 있어서 일반 사용자에게 큰 영향력을 줄 수 있는 오피니언 리더를 도출하고자 한다. 이들 중 긍정적인 감성 전파 역할이 가능한 웹포럼의 사용자는 향후 해당질환의 좋은 심리상담치료가 후보군이 될 수 있을 것이다.

본 연구에의 2장에서는 의료 웹포럼, 감성분석 및 사회연결망분석 등의 본 연구와 관련된 문헌을 정리하였다. 제 3장에서는 의료 웹포럼의 게시판 글의 수집 방법, 감성분석, 사

회연결망 구성 방법에 대해서 설명하였다. 4장에서는 실제 의료 웹포럼에 본 연구에서 제안한 연구방법론을 적용한 결과를 분석하여 웹포럼에서 상호작용을 많이 하는 사용자와 그들의 감성과의 관계에 대해서 고찰하였다.

II. 관련 연구

2.1 의료 웹포럼과 감성적지지

의료 웹포럼은 의료산업의 의료서비스 제공자(의사, 간호사, 물리치료사 등), 의료서비스 소비자(환자, 환자 보호자 등), 의료기술개발자(제약, 의료기기 등), 의료산업투자자 등의 모든 의료산업 주체들이 만날 수 있는 공간으로 다음과 같은 장점이 있다. (1) 의료 웹포럼은 의료주체에 관해 많이 알고 있거나, 관심이 높은 사용자들이 만나는 공간으로 우수한 품질의 의료정보를 제공하기도 한다. 2) 웹포럼에서는 병을 슬기롭게 극복한 환자나 환자 보호자들은 다른 사용자에게 감성적으로 위로도 할 수 있으며, 그들의 성공적인 투병기를 공유함으로써 감성적 지지를 제공할 수 있다(11).

Cobb(15)과 House 외(16)는 사회적 지지를 의료 정보 및 안내 등의 정보를 제공하는 정보적 지지, 신뢰, 공감, 관심, 사랑, 배려, 친밀감 등을 표현하는 감성적 지지, 금전을 포함한 물질적인 지원을 의미하며 도구적지지, 평가하는 프로세스에 관련된 평가적 지지로 소개하였다. 최근 질병이나 건강에 대한 관심이 커짐에 따라 의료 웹포럼 등의 인터넷 웹보드에 있어서 앞의 사회적 지지와 웹포럼과의 관계를 조사했을 때 정보적 지지와 감성적 지지가 큰 영향력을 미치는 것으로 나타났다(17, 18). 불치병 등의 장기적인 치료가 필요한 질병은 특히 정보적 지지와 감성적 지지가 중요하다(11). 병을 진단하고, 예측하고, 질병을 치료하는 정보를 제공하는 것은 의료 웹포럼의 정보적 지지에 해당한다.

Wenrich(19)외 등은 특히 불치병과 같은 장기치료가 필요한 경우에는 감성적 지지와 물리치료사의 관리 등이 매우 중요하다고 밝혔다. 특히 알츠하이머와 같은 불치병인 경우는 보호자가 집에서 환자를 돌보는 경우가 많아 보호자 역시도 정신적으로나 신체적으로 매우 큰 부담이 있다(20). 따라서 웹포럼의 감성적 지지 역할은 환자와 환자 보호자에게 매우 중요하다. 특히 최근 치매환자 보호자들의 동반 자살 등의 사회적 문제를 해결하기 위해서는 환자뿐만 아니라 보호자의 정신 관리 및 치료는 동반되어야 한다.

본 연구에서는 알츠하이머와 같은 불치병의 환자 및 관계

자들의 감성 관리를 위해 웹포럼을 이용하여 감성적 표현 정도를 분석하고 이에 영향을 미치는 요소를 파악하고자 한다.

2.2 소셜미디어의 감성분석

최근 사람들은 신제품, 영화, 기업, 연예 등에 관련된 정보와 의견을 트위터, 인터넷 뉴스, 웹포럼, 블로그 등에 게시하여 타인과의 의견을 공유하고 있다. 신상품에 대한 리뷰를 감성분석을 통해 제품의 호감도 뿐만 아니라 호감의 요인들을 파악하여 마케팅전략을 수립할 수 있으며, 인터넷 뉴스, 트위터의 내용과 기업성과와의 연관성을 분석한 연구들도 늘고 있다[21-24].

인터넷 사용자들이 소셜미디어의 사용이 증가함에 따라 방대한 데이터들이 생성되었고, 이에 소셜미디어를 전체적으로 평가하는 것은 더 어려워지고 더 중요해졌다. 따라서 최근 소셜미디어에서 생성되는 글들의 주요 토픽, 의도 등을 파악하기 위해서 텍스트 마이닝을 이용한 감성분석과 관련된 연구가 활발히 진행되어 오고 있다. 특히 기업의 성과 중에서 주식 시장 성장 및 투자 가능성에 대한 연구도 최근 늘고 있다. 인터넷 매체의 빅데이터의 감성을 긍정과 부정으로 분류하고, 감성과 주식시장의 성과의 관계를 분석한 연구들이 활발하게 이루어지고 있으며[25, 26, 27], 뉴스 콘텐츠의 감성분석 등을 통해 감성과 투자심리, 주식시장에 영향력과의 관계에 대한 연구들도 있어왔다[22]. 소셜미디어 중 트위터에 대한 관심이 집중되면서, 트위터에서 감성측정에 대한 연구가 이루어졌다. 2011년에 차미영외[28]에서는 트위터 데이터를 분석하여 사람들의 우울감을 측정하였다. 트위터 문서 분석을 통해 음악을 추천하는 시스템을 제안한 연구도 있다[29]. 이득환 외[22]에서는 다음 포털에서 제공받은 트위터를 대상으로 감성분석을 시행하여 추가정보와의 상관여부를 파악하였다. 특히 트위터에 나타나는 감성의 종류인 분노, 미움, 싫음, 두려움, 사랑, 수치심, 슬픔, 바람, 기쁨의 9가지 감성에 대한 주성분 분석을 한 결과 긍정적 감성단어와 부정적 감성단어로 분류됨을 보였다.

텍스트 감성분석 방법은 텍스트 출현 빈도를 측정하여 이미 개발된 감성사전으로 그 횟수를 측정하여 강도를 결정하는 비교사(Unsupervised)방식이 있다[30]. GI(General Inquirer)는 주제별로 부정적 단어와 긍정적 단어 즉 이항(binominal)분류한 가능한 우수한 표준이라고 연구된 바 있다[30-35]. GI는 1966년에 자동 텍스트 분석을 위해서 최초로 개발되어 최근 "Harvard IV-4 TagNeg(H4N)"로 발전했는데, 이 사전은 긍정단어와 부정단어로 전체적으로 2개의 카테고리 분류되며 세부적으로는 긍정, 부정, 강함, 약함, 능동적,

기쁨, 고통 등의 182개의 감성분류를 가지고 있다.

트위터, 인터넷 뉴스 등을 이용한 감성분석 등의 연구는 많이 이루어지고 있지만 의료분야의 웹포럼에 대한 감성분석은 시작 단계이다. 본 연구에서는 주요 토픽별로 작성되는 의료 웹포럼에 H4N을 적용하여 긍정적 단어와 부정적 단어를 분류하여 전체 글수에 대해 긍정단어 사용 정도와 부정 단어 사용 정도를 측정하였다.

2.3 사회연결망분석

사회연결망분석은 정보를 전달하는 도구로 전염 효과의 크기나 향후 분포를 예측하는데 이용될 수 있다[36, 37, 38]. 최근 소셜미디어의 확산으로 이와 같은 사회연결망분석에 대한 연구가 가속화되고 있다. 최근에는 웹상의 소셜네트워크 분석이 이슈가 되면서, 네트워크 효과를 이용한 바이럴 마케팅에 대한 연구가 활발하게 진행되고 있다[39, 40]. 소셜네트워크 상에서의 영향력이 높은 유력자를 파악하는 연구 및 네트워크상에서 정보나 영향력이 확산되는 현상을 관찰하는 연구가 새로운 연구 주제로 떠오르고 있다. 일반적으로 네트워크상에서 사람들의 다른 사람에 대한 영향력을 측정하기 위해서 중심성 지수들을 이용한다. 중심성 지표는 권력과 영향력과 연관되어 자주 언급이 되는데 경험적 분석에서 중심성이 높은 노드는 특별한 경제적 지위를 가지거나, 중심성이 높을수록 생존률이 높거나 기업성과 관련 지수가 높은 것으로 알려져 있다[41]. 중심성 관련 지표는 연결중심성(degree centrality), 인접중심성(Closeness), 매개변수중심성(betweenness centrality), 보나시치중심성(Bonacich power centrality) 등이 있다. 연결중심성은 연결된 노드수를 의미하는 것으로 지역적인 중심성을 측정하는 지표이며, 같은 연결망 그룹사이내에서 연결망들의 밀도를 비교할 때 적절하다. 인접중심성은 다른 점들과의 최단 거리, 경로거리를 의미하며, 매개변수중심성은 한 노드가 연결망 내의 다른 노드 사이에 위치하는 정도로 최단거리를 연결하는 선 위에 위치할수록 이 값이 높아지며 그룹간의 매개역할을 하는 것으로 알려져 있다. 보나시치중심성(위세중심성)은 연결된 노드의 중요성에 가중치를 두어 측정된 지수로 보나시치권력지수라고도 불린다[42]. 해당 노드의 연결정도 중심성으로부터 발생하는 영향력과 본인에 연결된 타 노드의 영향력을 합하여 권력 중심성을 결정하는 값이다. 정보가 확산되는 영향력을 파악하기 위해서는 위치적 접근보다는 관계적 접근이 보다 타당한데 최단거리에 집착하는 경우는 위치적 접근이라고 볼 수 있을 것이다. 따라서 본 연구에서는 연결중심성과 위세중심성을 이용하여 네트워크의 영향력을 측정하였다.

본 연구에서는 알츠하이머 웹포럼 게시판을 대상으로 감성 분석을 통해 긍정적인 단어와 부정적인 단어를 사용 정도와 네트워크 중심성과의 관계를 파악하여 향후 의료 웹포럼의 활용성에 대해서 고찰하였다.

III. 연구방법

3.1 가설 수립

소셜미디어를 통해 형성되는 네트워크의 연결적, 결속적 속성이 사회 자본에 긍정적인 영향이 있다는 것은 기존의 연구에서 보여졌다. 구체적으로 소셜미디어는 기존의 오프라인에서의 인간관계를 강화시키기도 하고, 새로운 인간관계를 형성하여 약한 연계를 만들기도 하면서 사회적 자본 증진에 기여한다[43, 44]. 기존의 연구에서 사회적 자본은 사람의 감성적 건강에 긍정적 영향을 미친다는 것을 보였다[45, 46]. 또한 앞서 언급한 대로 온라인커뮤니티에서 사람들은 감성, 적 또는 정보적 형태로 사회적지지를 주고받게 되는데[17, 18], 이러한 사회적지지는 사람의 정신건강에 좋은 영향을 준다는 것이 기존의 연구에서 밝혀졌다. 최근의 연구를 살펴보면 페이스북에서 우울한 사용자는 다른 사용자와의 상호작용이 적은 것으로 나타났다[47].

위의 이론적 토대를 배경으로 의료 웹포럼에서도 사람들간의 상호관계의 결과로 나타나는 네트워크상에서의 우위가 사용자의 감성과 관련을 가지는지를 분석하기 위해 다음의 두 가지 가설에 대한 검정을 수행한다.

가설 1. 감성표현을 많이 하는 사람이 온라인 의료 커뮤니티에서 사람들과의 관계네트워크에서 중심에 존재한다.

가설 2. 긍정적인 감성을 표현하는 사람이 부정적인 감성을 표현하는 사람보다 사람들과의 관계 네트워크에서 중심에 존재한다.

3.2 가설 검증 모델

본 연구에서는 의료 웹포럼에서 긍정단어의 사용 정도에 따른 네트워크상의 영향력을 검증하기 위한 연구절차는 그림 1과 같다.

첫 번째로는 데이터 수집단계로서 웹 포럼에서 작성된 글들은 일반적으로 게시글(post)과 게시글에 답글을 다는 댓글

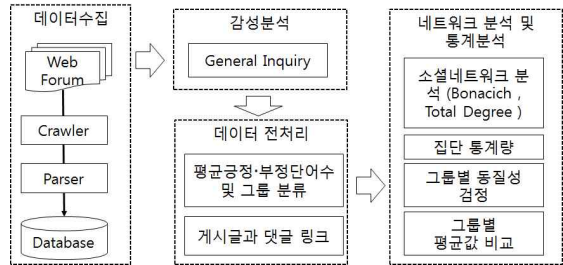


그림 1. 연구 절차
Fig. 1. The Diagram of Research Process

(reply)형태로 작성되어 있으며 이와 같은 웹 페이지는 HTML로 작성되어 있으며, HTML 소스를 외부에 공개한다. 이를 이용하여 웹 크롤러(crawler)를 개발하고, 웹 포럼상의 글을 자동으로 수집한다. 이를 통해 수집된 HTML 형태 데이터는 작성자, 콘텐츠 작성 시간, 제목, 내용 등이 태그(tag)화 되어 있다. 이 HTML을 파싱(parsing)과정을 통해 필요한 필드를 추출하고, 이를 표준화하여 저장하는 데이터베이스화 과정을 거친다.

두 번째는 감성분석단계로서 웹포럼의 메시지의 긍정적 단어 수를 측정하기 위해서 General Inquirer[30-34]를 이용하여 개별 메시지의 단어수를 측정하고 긍정단어와 부정단어 수를 Harvard IV-4 TagNeg(H4N) 사전 기반으로 도출한다. 이 H4N는 메세지의 긍정·부정, 강한·약한 어조, 즐거움, 활동적, 고통 등과 같은 감성을 포함한 182개의 태그 카테고리 사전을 이용하여 콘텐츠 분석을 수행한다. 본 연구에서는 H4N 사전의 긍정 부정 단어 카테고리를 이용하여 의료 웹포럼의 개별 메시지 별로 긍정단어와 부정단어 전체수를 도출하였다.

세 번째로는 데이터전처리 단계로서 사회연결망분석 및 통계분석을 위해서 데이터 통합, 변환, 범주화 등의 과정을 수행한다. H4N사전을 통해 도출된 긍정으로 분류된 단어수 (GI_p)의 총합을 메세지수로 나눈 값을 평균긍정단어수로 식(1)과 같이 정의하고, 도출된 부정으로 분류된 단어수 (GI_n)의 총합을 메세지수로 나눈 값을 평균부정단어수로 식(2)과 같이 정의하였다.

$$Avg_{\text{긍정}}(v) = \frac{\sum_{i=1}^{N_v} GI_{\text{긍정}}(v, i)}{N_v} \dots\dots\dots\text{식(1)}$$

$$Avg_{부정}(v) = \frac{\sum_{i=1}^{N_v} GI_{부정}(v, i)}{N_v} \dots\dots\dots\text{식(2)}$$

v : 노드(사용자)

N_v : v 노드의 메시지수

$Avg_{긍정}(v)$: 노드의 메시지에 대한 평균긍정단어수

$Avg_{부정}(v)$: 노드의 메시지에 대한 평균부정단어수

사용자가 메시지 평균긍정단어수를 $Avg_{긍정}(v)$, 메시지 평균부정단어수를 $Avg_{부정}(v)$ 로 표기한다. 이때 $Avg_{긍정}(v)$ 값을 내림차순으로 정렬하여 20% 단위로 Pos1, Pos2, Pos3, Pos4, Pos5로 분류하고 긍정단어그룹이라고 하였다. $Avg_{부정}(v)$ 값을 내림차순으로 정렬하여 20% 단위로 Neg1, Neg2, Neg3, Neg4, Neg5로 분류하고 부정단어그룹이라고 하였다. 마지막으로 웹포럼의 메시지별로 게시글(post)와 댓글(reply)관계를 데이터베이스화하였다.

마지막으로 사회연결망분석 및 통계분석을 통해 감성분석과 사회연결망분석과의 관계에 대해서 분석하였다. 각 사용자를 메시지별로 게시글과 댓글과의 관계를 네트워크 구조로 도출하고, 이로부터 도출되는 영향력 지수를 긍정단어사용정도 와 연관지어 살펴보았다. 해당 연구에서는 사용자의 영향력을 측정하기 위해서 정보의 확산 횟수를 이용하였다. 정보의 확산은 게시글/댓글 관계에서 이루어진다는 가정하에 정보확산 네트워크상에서의 연결을 중심으로 한 영향력 지표를 활용하였다. 이를 통해 웹포럼에 있어서 상호관계가 높은 사람일수록 긍정적 감성이 높은지를 파악하였다. 구체적으로는 연결중심성, 보나시치권력지수[37, 38]를 도출하여 노드(사용자) 간의 영향력을 측정하였다. 연결중심성 $C_D(v)$ 은 한 점에 연

결된 다른 점의 수를 고려하여 한 점의 중심성을 측정하는 값으로 이 값이 높은 점은 다른 점들과의 직접 연결이 많으며, 이 값은 여러 점들 중에서 중심적인 역할을 할 수 있는 것을 의미한다. 웹포럼에서의 연결중심성이 높은 노드(사용자)는 메시지를 주고받는 노드(사용자)가 많은 것을 의미한다.

$$C_D(v) = \frac{\text{deg}(v)}{n-1} \dots\dots\dots\text{식(3)}$$

$C_D(v)$: 연결중심성

n : 점(node) 수

$\text{deg}(v)$: 노드(node)와 직접적으로 연결된 노드(node)들의 수

보나시치(Bonacich)권력지수는 자신의 연결정도로부터 발생하는 영향력과 자신과 연결된 타 행위자의 영향력을 합하여 위세중심성(prestige index)[41, 42]을 결정하는데 이 값은 연결된 상대방의 중요성에 가중치를 두는 값이다(식4). 행위자(v)는 위세가 높은 사람들 j 로부터 많은 관계를 수신할수록 위세가 높아지는 것으로 해석할 수 있다. 즉 웹포럼에서의 중요한 노드(사용자)를 많이 두는 노드(사용자)를 의미하는 값이다.

$$C_B(v) = \sum_{j=1}^{N-1} P_j Z_{jv}, 0 \leq p_v \leq 1 \dots\dots\dots\text{식(4)}$$

$C_B(v)$: 보나시치중심성

IV. 의료 웹포럼에의 적용

본 연구에서 3장에서 제시하는 연구방법론을 알츠하이머(Alzheimer)에 대해 의견들을 게재하는 대표적인 의료 웹포

표 1. 웹포럼의 메시지별 데이터 감성분석 결과 매핑(일부)
Table 1. Message Mapping of Sentiment Analysis

code	path	time	questioner	Author	title	msg	positive	negative	total
TID0001	TID0001-TID0001	01-03-2009, 08:34 PM	UPB****	UPB****	my sympathies petal*pusher	I always like to read your posts ****	6	2	52
TID0002	TID0001-TID0002	01-05-2009, 07:12 AM	UPB****	Shi****	Music and Alzheimer's	I am really interested in the benefits *****	13	3	198
TID0003	TID0001-TID0003	01-06-2009, 10:04 PM	UPB****	Mic****	A Quest for Information	Syndrome due to alchcohol abuse ****	17	5	202

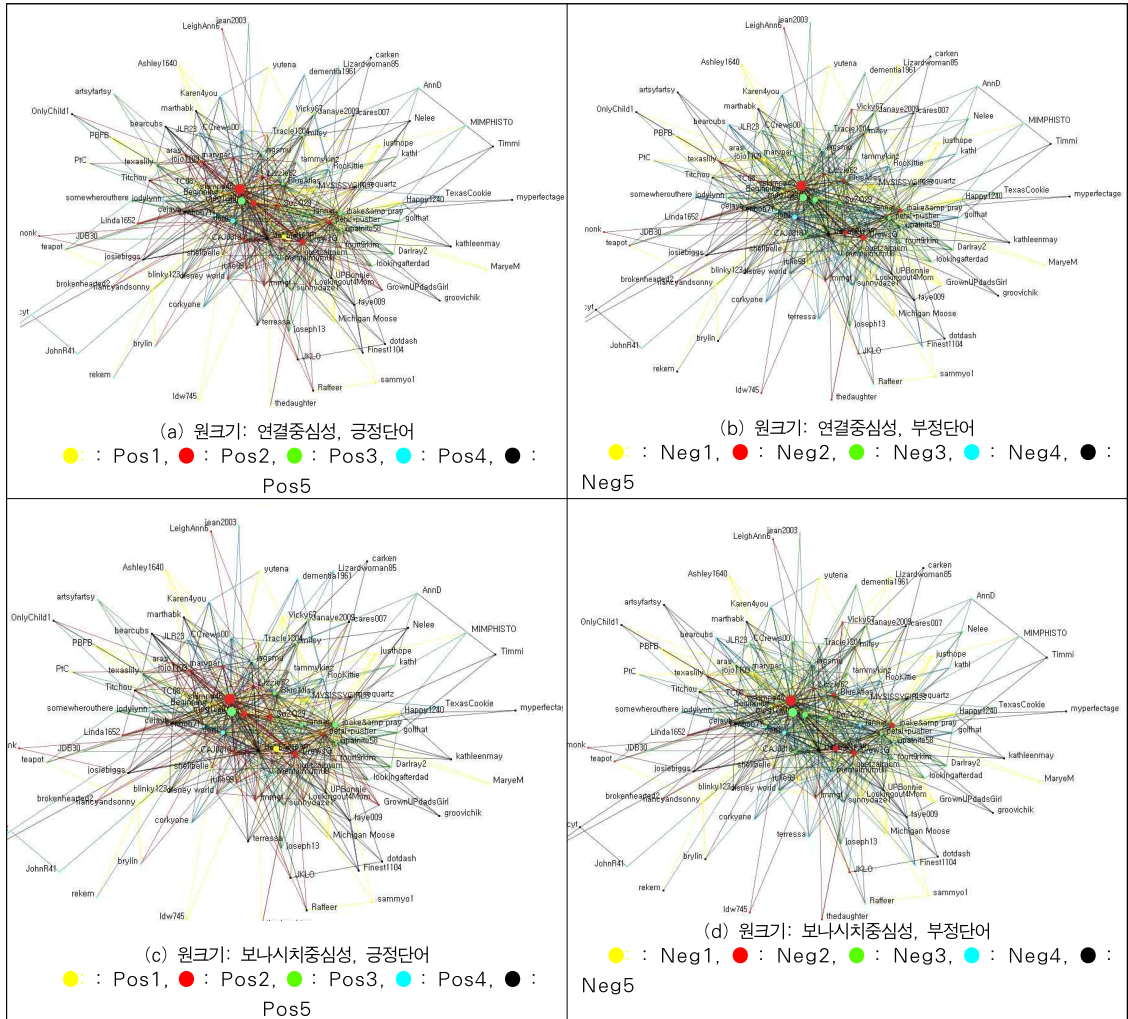


그림3. 연결중심성, 보나시치중심성 지수에 따른 긍정단어 및 부정단어 사용그룹의 사회연결망
 Fig3. Network Graphs by Sentiment Groups based on Degree Centrality and Bonacich power centrality

럼인 HealthBoards²⁾의 Alzheimer's Disease & Dementia 소게시판의 데이터에 적용한다. 해당 게시판에서 1년 동안의 데이터 글을 원글(158개)과 답글(7,867개)을 포함하여 총 8,025개의 글을 추출하였다. 원글과 댓글을 작성한 사용자는 261명이었으며 웹פור럼에서 메시지를 2회 이하 혹은 비정상적인 대화 메시지를 작성한 사용자를 제외한 123명 데이터를 이용하여 분석하였다.

크롤링과 파싱을 통한 데이터 수집 후 Harvard IV-4 TagNeg(H4N) 사전을 이용한 General Inquirer를 이용한 감성분석 결과는 <표 1>과 같다. 웹פור럼 글은 원글과 댓글과

의 형태로 되어 있으며 크롤링과 파싱작업을 통해 코드(code), 경로(path), 글입력시간(time), 원글저자 ID(questioner), 답글저자ID(author), 글제목(title), 글메세지수(msg), 전체단어수(total) 형태로 저장되어 있다. 이때 경로(path) 컬럼은 '원글코드-댓글코드'와 같은 형태로 저장되도록 하였다. 이 데이터를 기반으로 하여 General Inquirer를 이용한 감성분석을 통해 긍정단어수(positive)와 부정단어수(negative)를 추출하였다.

<표 1> 데이터를 기반으로 사용자(node)별로 사용한 긍정단어수, 부정단어수, 전체단어수, 전체메세지수를 합하고, 메시지별 평균긍정단어수($Av_{g\text{ 긍정}}(v)$)와 메시지별 평균부정단어

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

어수($Avg_{부정}(v)$)도 도출하였다. 노드별 평균긍정단어수와 평균부정단어수를 기준으로 <표2>와 같이 그룹화하였다.

본 연구에서는 감성분석과 네트워크 분석의 비교대상을 절대값으로 비교하지 않고, 그룹화하여 상대비교를 실시하였다. 이러한 상대비교 방식은 감성지수와 네트워크의 중요도 값을 절대 비교 시에 둘 간의 선형 관계를 파악하기 어려울 수 있는 경우, 두 변수간의 관계를 파악할 수 있다는 장점을 가진다.

<표 1>의 'path' 컬럼을 이용하여 원글과 댓글과의 관계를 매트릭스로 <그림 2>와 같이 도출하였다.

	m	sunnydaze1	Happy1240	Beginning	JLR23	frustratedb.	upatnile58	jannar	Lizzie62
teasaily		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
disney wo...		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
four9rkim		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sunnydaze1		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Happy1240		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Beginning		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
JLR23		0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
frustratedb.		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
upatnile58		3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
jannar		1.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0
Lizzie62		0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
manyar		0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0
mentaimu...		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TC08		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CAJ0818		1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
cejayb		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jagsmu		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
kenbob71		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
inio1109		0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

그림 2. 원글과 댓글의 매트릭스 결과
Fig. 2. Matrix Result between posts and replies

먼저 웹포럼 메시지에서 사용자가 작성한 평균긍정단어수와 평균부정단어수의 차이를 비교하기 위해서 t 검정을 한 결과는 <표3>와 같다. $Avg_{긍정}(v)$ 와 $Avg_{부정}(v)$ 의 상관계수는 0.859로 두 변수의 상관관계는 매우 높고, 두 값의 평균치에 대한 차이는 양측검정에서 '유의확률($sig.$)=4.9E-30'($\alpha = 0.05$)으로 귀무가설을 기각하여 메시지별 평균부정단어수와 평균긍정단어수는 차이가 있는 것으로 나타났다. 알츠하이머 병에 대해 평균긍정단어수가 평균부정단어수보다 약 3.93개 높게 나타났으며 웹포럼에 긍정적인 글을 다소 더 작성하는 것으로 나타났다. <그림 3>은 <그림 2>의 매트릭스 결과물을 이용하여 사회연결망 분석을 통해 도식화한 결과이다. <그림 3-(a)>에서는 원의 크기가 연결중심성 값을 나타내며 ●은 메시지별로 평균긍정단어수 최상위 그룹(Pos1)을 나타내며, ●는 상위그룹(Pos2), ●는 중위그룹(Pos3), ●는 하위그룹(Pos4), ●는 최하그룹(Pos5)을 나타낸다. 연결중심성이 높은 노드는 최상위그룹(Pos1)보다는 상위그룹(Pos2)에 분포하는 것을 확인할 수 있었다.

표 2. 사용자별 전처리 데이터 결과 (일부)
Table 2. Result of Data Pretreatment (Sample)

node	메세지 합계	긍정 단어 합계	부정 단어 합계	평균 긍정 단어수	긍정 그룹	평균 부정 단어수	부정 그룹
Ran***	3	70	41	13.7	Pos1	23.3	Neg1
jus***	9	169	102	11.3	Pos1	18.8	Neg1
Ash***	8	185	166	20.8	Pos1	23.1	Neg1
Mar***	5	118	96	19.2	Pos1	23.6	Neg1
nan***	7	162	145	20.7	Pos1	23.1	Neg1
phy***	3	59	48	16	Pos1	19.7	Neg1

표 3. 긍정단어그룹과 부정단어그룹의 t 검정 및 상관관계 분석
Table 3. T test between positive and negative word groups

	평균 부정단어수	평균 긍정단어수
평균	8.360858	12.28603
분산	18.99742	30.53906
관측수	123	123
피어슨 상관 계수	0.859093	
가설 평균차	0	
자유도	122	
t 통계량	-15.2475	
P(T<=t) 양측 검정	4.9E-30*	
t 기각치 양측 검정	1.9796	

<그림 3-(b)>에서는 마찬가지로 원의 크기가 연결중심성 값을 나타내며 ●은 메시지별로 부정단어 사용률이 최상위 그룹(Neg1)을 나타내며, ●는 상위그룹(Neg2), ●는 중위그룹(Neg3), ●는 하위그룹(Neg4), ●는 최하그룹(Neg5)을 나타낸다. 연결중심성이 높은 노드는 최상위그룹(Neg1)보다는 상위그룹(Neg2)에 분포하는 것을 확인할 수 있었다. 긍정단어 사용그룹 분포나 부정단어 사용그룹 분포 모두 최상위그룹 보다는 상위그룹의 연결중심성이 높게 나타나 보였다. 연결중심성은 연결된 결점의 수를 의미하는 연결정도를 의미하며 국지적인 지역중심성을 측정하는 좋은 지표로 알려져 있다. 연결중심성이 큰 노드는 본인이 웹포럼내의 다른 사용자와 대화가 많은 사용자라고 볼 수 있을 것이다. 긍정·부정 단어를 매우 많이 사용하는 최상위그룹(Pos1)을 제외한 경우에는 긍정감성과 부정감성의 지지도가 클수록 인터렉션에 더 영향을 미치는 것으로 보여진다. <그림 3-(b)>은 <그림 3-(a)>에 비해 중·하위위그룹의 비중이 높고, 상위그룹비중은 낮은 것을 확인할 수 있다. 이는 긍정단어를 사용하는 그룹이 네트워크의 중심성이 보다 높은 것으로 판단할 수 있다.

<그림 3-(c)>에서는 원의 크기가 보나시치중심성을 나타내며 긍정그룹별 분포를 보여준다. 긍정단어 저사용자는 분포하지 않았으며, 긍정단어 고사용자는 연결중심성 값이 높은

상위에 분포하는 것을 확인할 수 있었다. 연결중심성 노드 분포와 유사하게 보나시치중심성이 높은 노드도 최상위그룹(Pos1)보다는 상위그룹(Pos2)에 분포하는 것을 확인할 수 있었다. 보나시치중심값 역시 긍정단어 사용그룹 분포나 부정단어 사용그룹 분포 모두 최상위그룹 보다는 상위그룹의 연결중심성이 높게 나타났다. 상위그룹(Pos1)에서는 급격하게 감소하는 것을 확인할 수 있다. 보나시치중심성은 위세중심성으로도 표기되며 영향력 있는 노드(사용자)와 연결되어 있는 노드가 높은 값을 가지게 된다. 웹포럼에서 보나시치중심성 값이 높은 사용자의 경우는 웹포럼에서 중요한 사람들과 네트워크활동을 활발히 하는 것으로 볼 수 있다. 위의 연결중심성 결과와 유사하게 긍정·부정 단어를 매우 많이 사용하는 그룹(Pos1, Neg1)을 제외한 경우에는 긍정감성과 부정감성의 지지도가 클수록 인터랙션에 더 영향을 미치는 것으로 보여진다. 연결중심성과 마찬가지로 보나시치중심성 값도 <그림 3-(d)>은 <3-(e)>에 비해 중/하위 그룹의 비중이 높고, 상위그룹 비중은 낮은 것을 확인할 수 있다. 이는 긍정단어를 사용하는 그룹이 네트워크의 위세중심성이 보다 높은 것으로 판단할 수 있다.

평균긍정단어수 그룹에 따른 연결중심성과 보나시치중심성과의 동질성 및 기초 통계량은 <표 4>와 같다. 평균긍정단어사용그룹별 연결중심성의 평균값은 Pos2인 경우가 0.0034로 가장 큰 값으로 나타났다. 최하위그룹(Pos5)부터 상위그룹(Pos2)으로 메시지당 긍정단어 사용수가 증가함에 따라 연결중심성이 커지고, 이는 그룹별 평균값이 커지는 것을 확인할 수 있었다. 이는 <그림 3-(a)>, <그림 3-(c)>에서 Pos2 그룹의 연결중심성 및 보나시치중심성 값이 큰 노드가 두드러지게 나타나는 현상과 일치한다고도 볼 수 있다.

표 4. 긍정단어그룹별 연결중심성과 보나시치중심성 집단 통계량
Table 4. The Summary of Degree Centrality and Bonacich Power Centrality by Positive Word Group

평균긍정단어사용그룹	연결중심성		보나시치중심성	
	평균	표준편차	평균	표준편차
Pos1(25)	0.0010	0.0020	0.0343	0.0729
Pos2(24)	0.0034	0.0060	0.1094	0.1788
Pos3(25)	0.0020	0.0045	0.0658	0.1500
Pos4(24)	0.0011	0.0024	0.0336	0.0606
Pos5(25)	0.0008	0.0017	0.0211	0.0454
전체(123)	0.0016	0.0038	0.0525	0.1167

평균부정단어사용그룹별로 연결중심성과 보나시치중심성의 기초통계량이 <표 5>에서 제시되어 있다. 평균긍정단어사용그룹과 유사한 형태로 연결중심성, 보나시치중심성의 평균

이 Neg2에서 최고값을 가지는 것을 보여주며, 이는 <그림 3-(b)>, <3-(d)>에서 Neg2 그룹의 연결중심성 값이 큰 노드가 확연하게 두드러지는 것과도 일치하는 결과이다.

표 5. 긍정단어그룹별 연결중심성과 보나시치중심성 집단 통계량
Table 5. The Summary of Degree Centrality and Bonacich Power Centrality by Negative Word Group

부정긍정단어사용그룹	연결중심성		보나시치중심성	
	평균	표준편차	평균	표준편차
Neg1(25)	0.0007	0.0010	0.0250	0.0333
Neg2(24)	0.0030	0.0061	0.0970	0.1832
Neg3(25)	0.0024	0.0048	0.0825	0.1570
Neg4(24)	0.0014	0.0024	0.0411	0.0656
Neg5(25)	0.0007	0.0018	0.0184	0.0457
전체(123)	0.0016	0.0038	0.0525	0.1167

<표 4>와<표5>의 집단 간의 평균차이가 통계적으로 유의한지를 확인하기 위해서 집단 평균의 동질성에 대한 검증을 시행하였다. Wilks'람다(=집단내 총분산/총분산) 값을 이용하여 집단 동질성에 대한 유의확률을 구한 결과는 <표 6>,<표 7>과 같다. Wilks'람다 값이 작으면 F통계량이 값이 커지며, F 통계량값이 클수록 전체분산비율이 높으므로 판별력이 높아지게 된다. 연결중심성과 보나시치중심성은 F 통계량 유의확률 <0.1 이므로 평균단어사용그룹에 따라 평균차이는 유의하다고 볼 수 있다. 평균부정단어사용 그룹에 따른 연결중심성의 동질성에서는 유의확률이 0.106으로 평균차이가 거의 유의하다고 볼 수 있다. 즉 극단적인 웹포럼 감성적 단어 사용그룹을 제외하였을 때 감성적인 단어를 많이 사용할수록 상호작용을 많이 하여 네트워크의 중심역할을 한다고 볼 수 있다.

표 6. 긍정단어그룹에 대한 연결중심성과 보나시치 중심성의 동질성에 대한 검정

Table 6. The Summary of Degree Centrality and Bonacich Power Centrality by Positive Word Group

	Wilks' 람다	F	df1	df2	Sig.
연결중심성	0.932	2.147	4	118	0.079
보나시치중심성	0.926	2.374	4	118	0.056

표 7. 부정단어그룹 에 대한 연결중심성과 보나시치중심성의 동질성에 대한 검정

Table 7. The Summary of Degree Centrality and Bonacich Power Centrality by Negative Word Group

	Wilks' Lambda	F	df1	df2	Sig.
연결중심성	0.938	1.953	4	118	0.106
보나시치중심성	0.927	2.323	4	118	0.061

지금까지 결과를 요약하면, 평균긍정단어수와 평균부정단어수는 높은 상관성을 보이며, 연결중심성과 보나시치중심성은 유사하게 Pos2, Neg2 그룹의 연결중심성과 보나시치중심성값이 최대치를 가지면서 줄어드는 형태를 보였다. 긍정단어 사용도와 부정단어사용정도가 높아짐에 따라 연결중심성과 보나시치중심성이 높아지는 현상은 인터랙션이 많을수록 감성표현을 더 많이 한다고 보여진다. 하지만 긍정적이거나 부정적인 감성표현이 극단적으로 많은 그룹에서는 오히려 웹포럼에서 메시지 작성 횟수가 적어 네트워크상에서 영향력이 적었다. 네트워크를 원클릭/댓글의 관계로 구성하였기 때문에 메시지 수와 네트워크 영향력이 완전히 비례하지는 않지만, 영향력 지수값이 높으려면 어느 정도 이상의 메시지 수를 갖고 있어야하는데, 메시지 수가 극히 적은 경우에는 해당 네트워크에서 영향력 지표가 높게 나타날 수 없다.

메시지별로 긍정/부정 감성을 가장 많이 표현하는 Pos1, Neg1 그룹에서는 연결중심성과 보나시치중심성이 극감하는 이유를 파악하기 위해서 평균긍정단어의 Pos1그룹(=메시지별로 긍정단어 17개 이상 작성자)인 평균긍정단어수의 메시지 분포를 보았을 때 이 그룹 사람들은 상대적으로 낮은 메시지수를 작성하는 것을 확인할 수 있었다(그림 3). 메시지수가 낮은 그룹은 네트워크상의 중심지수를 나타내는 높은 연결중심성, 보나시치중심성을 갖기 어렵다. 마찬가지로 평균부정단어수와 메시지수도 유사한 분포를 보였다.

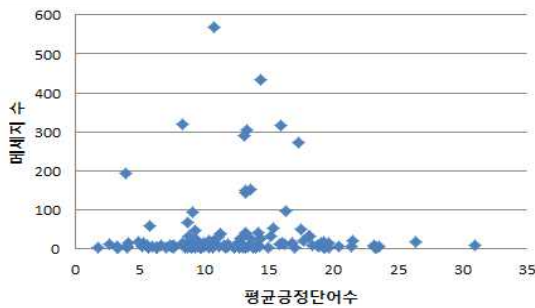


그림 3. 평균긍정단어수 사용자와 메시지수와의 분포
Fig. 3. The Distribution between the Number of Messages and The Average number of Positive Words

이때 긍정단어 사용그룹과 부정단어 사용그룹과 연결중심성 평균값을 살펴보면 긍정단어를 많이 사용한 최상위(Pos1), 상위(Pos2)그룹에서는 긍정단어사용그룹의 연결중심성 평균값은 부정단어를 많이 사용하는 Neg1, Neg2 그룹보다 높았고, 반대로, 중위이하의 그룹에서는 부정단어사용그룹의 연결중심성 평균값이 더 높게 나타났다(그림 4(a)). 즉, 웹포럼에

서 긍정적 감성단어를 더 많이 사용하는 사람들이 부정적 감성 단어를 더 많이 사용하는 사람들보다 연결중심성 값이 높게 나타났다. 이는 긍정단어를 많이 사용하여 긍정적 감성을 표현하는 사람들이 부정적 감성을 표현하는 사람들에 비해 웹포럼에서 직접적인 인터랙션이 더 높다는 것을 나타낸다. 마찬가지로, 웹포럼의 메시지에 긍정단어를 많이 작성하는 사용자가 부정단어를 많이 작성하는 사용자에 비해 높은 보나시치중심성 값을 나타낸다(그림 4(b)). 이도 역시 긍정적 단어를 많이 사용하여 긍정적 감성을 표현하는 사람들이 부정적 단어를 사용하는 사람들에 비해 주변에 중요한 노드(사용자)를 두고 있다는 것을 나타내며, 이는 영향력 높은 노드를 주변에 많이 두고 있다는 것으로 해당노드를 통해 정보/감성의 전달이 더 효과적으로 진행될 수 있을 것으로도 판단할 수 있다.

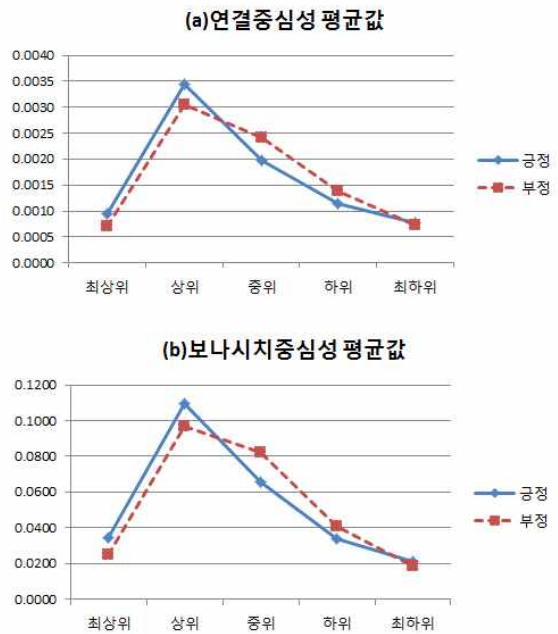


그림 4. 연결중심성과 보나시치중심성 평균값의 그룹별 비교
Fig. 4. The Comparison between positive and negative groups: Degree Centrality(a) and Bonacich power centrality(b)

IV. 결론 및 향후 연구

5.1 결론

본 연구에서는 웹포럼의 게시글과 댓글의 사용자의 상호관

게 분석을 통해 긍정적 단어를 많이 사용하는 웹포럼 사용자와 부정적 단어를 많이 사용하는 웹포럼 사용자들간의 파워중심성과 연결중심성을 비교함으로써 감정과 사용자간의 네트워크의 중심성 및 영향력을 파악하였다. 본 연구를 통해 연결중심성, 보나시치중심성 값을 이용하여 웹포럼 사용자들에 영향을 미칠 수 있는 오피니언 리더 후보자 리스트를 도출할 수 있다. 본 연구는 의의는 다음과 같다. (1)국내에서 아직까지 알츠하이머에 대한 웹포럼을 이용한 감정분석 및 감정과 사회연결망연결 분석에 대한 연구는 거의 없는 실정으로, 알츠하이머 웹포럼의 네트워크 활동에 대한 기초자료를 제공했다. (2)또한 의료 웹포럼에 대한 감정분석과 네트워크 분석을 통해 극단으로 감정표현을 많이 하는 그룹을 제외한 경우, 감정 표현이 많아질수록 네트워크 노드가 영향력이 높고, 인터랙션이 많아지는 현상을 규명했다.

긍정적인 단어나 부정적인 단어를 극단적으로 많이 사용하는 그룹을 제외했을 때는 긍정적인 단어, 부정적인 단어 모두 많이 사용하는 그룹일수록 네트워크 상의 연결밀도를 의미하는 연결중심성 값이 높게 나타났다. 연결중심성이 큰 노드는 본인이 웹포럼 내의 다른 사용자와 인터랙션이 많은 사용자라고 볼 수 있을 것이다. 중요한 그룹을 많이 확보하고 있는 노드를 의미하는 보나시치파워중심성 값도 긍정적인 단어나 부정적인 단어를 극단적으로 많이 사용하는 그룹을 제외했을 때는 긍정적인 단어, 부정적인 단어 모두 많이 사용하는 그룹일수록 높게 나타났다. 보나시치파워중심성 값이 높은 노드는 본인이 웹포럼내의 다른 중요한 사용자와 인터랙션이 많은 사용자라고 볼 수 있을 것이다.

긍정·부정 단어를 매우 많이 사용하는 최상위그룹(Pos1)을 제외한 경우에는 긍정감과 부정감정의 지지도가 클수록 노드간의 인터랙션에 더 영향을 미치는 것으로 보여진다. 보나시치 파워중심성, 연결중심성 모두 긍정단어나 부정단어를 극단적으로 많이 사용하는 최상위 그룹인 경우는 웹포럼에서 네트워크상의 영향력을 미치기에는 매우 적은 수의 메시지를 작성한 것을 확인 할 수 있었다. 이는 감정적인 글을 메시지에 너무 많이 작성하는 경우는 웹포럼에 메시지 작성을 많이 하지 않는다는 것으로도 볼 수 있다.

긍정단어그룹과 부정단어그룹을 연결중심성과 보나시치파워중심성 값에 대해서 비교해 보면 두 중심성 값 모두 유사한 패턴을 보였다. 단어 사용이 많은 최상위, 상위 그룹에서 긍정단어사용자 그룹의 중심성 값이 더 높게 나타났으며 단어 사용이 중위 이하인 그룹에서는 부정단어사용자 그룹의 중심성 값이 더 높게 나타났다. 이는 긍정그룹인 경우가 부정그룹에서 보다 네트워크에서 영향력이 더 높은 지위를 가진다고도

볼 수 있을 것이다.

5.2 연구한계 및 향후 연구과제

본 연구의 다음과 같은 한계점을 가지고 있으며, 이를 극복하기 위한 향후 연구방향은 다음과 같다. 첫째, 알츠하이머의 웹포럼 게시판을 지정하여 연구조사를 진행하였다는 점이다. 추후 장기적인 질병과 단기적인 질병에 대해서 감정분석과 네트워크 분석을 진행하여 비교해 볼 필요가 있다. 둘째, 2009년의 12개월간의 웹포럼 데이터를 이용하여 분석 한 점이다. 이를 보완하기 위해서 시계열 분석을 수행하여 감정과 네트워크 구조간의 관계를 규명할 필요가 있다. 셋째, 의료 웹포럼에 사용된 감정분석과 영향력을 설명할 수 있는 새로운 지수를 개발할 필요가 있다. 현재 사회연결망분석을 통해 도출되는 지표는 감정분석을 충분히 설명할 수 있는 지표가 없다. 따라서 감정분석과 네트워크 영향력을 설명할 수 있는 모델을 연구하는 연구도 추후 수행할 예정이다. 넷째, 본 연구는 연구대상을 웹포럼에 한정하고 있다. 일상생활에서 이용하는 소셜미디어에 나타나는 질병관련 글은 목적과 커뮤니케이션 대상이 명확한 의료 웹포럼에 비해 주제와 표현이 제한될 수 밖에 없다는 한계를 가진다. 또한 다른 소셜미디어에서 특정 질환의 환자에 대한 정보를 대량으로 얻기 어렵기 때문에 웹포럼에 한정하여 데이터 수집을 진행하였다. 하지만 향후 연구에서는 다른 소셜미디어의 데이터를 종합하는 과정을 통해 연구 결과를 좀 더 일반화할 필요가 있다.

마지막으로, 본 연구는 감정분석에 사람의 감성이 글을 통해 완전히 나타나지는 않을 수도 있고, 사람에 따라서는 감성을 완곡하게 표현하는 경우가 있어 감성을 글로만 측정하는 것에는 한계를 가진다. 하지만 많은 사람들의 감성을 쉽게 측정하는 것이 쉽지 않으므로, 본 연구에서 사용한 방법이 대안이 될 수 있다. 또한 모든 사람들의 글에 대해 동일한 기준으로 감성을 측정하였는데, 사람에 따라서 감성의 표현 강도가 다를 수 있다는 점은 본 연구의 한계이다.

Acknowledgement

본 연구는 한국연구재단의 지원(NRF-2013R1A1A3011816)을 받아 수행되었음.

REFERENCES

[1] Global Internet Usage, <http://www>.

- internetlivestats .com /internet-users/
- [2] Internet User Demographics, <http://www.pewinternet.org/data-trend/internet-use/latest-stats/>
- [3] W. Suh, N. Syamsul, J. Hong and J. Park, "An Exploratory Study on Employee Use of Social Media and Policy: Focused on Facebook", *The Journal of Business and Economics*, Vol. 28, No. 2, pp. 131-159, 2012.
- [4] S. Hong, Y. Chung and J. Lee, "Semi-supervised learning for sentiment analysis in mass social media", *Journal of Korean Institute of Intelligent Systems*, Vol. 24, No. 5, pp. 482-488, 2014.
- [5] J. Dang and S. Lee, "Health Care and Social Media :Building relationships via Social Networks", *Korean journal of hospital management*, Vol 17, No. 3, pp. 99-109, 2014.
- [6] E. Nolte, M. McKee (Eds.), "Caring for People with Chronic Conditions - A Health System Perspective", McGraw-Hill International, 2008.
- [7] N. Elkin, *How America searches: health and wellness*, iCrossing 2008.
- [8] T. Lee and Y. Jang, "IT Convergence Research Trends of Health Care, HEALTH-WELFARE POLICYFORUM", *Korea Institute for Health and Social Affairs*, vol. 03 pp. 36-49, 2014.
- [9] T. H. Van De Belt, L. J. Engelen, S. A. Berben, and L. Schoonhoven, "Definition of Health 2.0 and Medicine 2.0: A Systematic Review", *J. of Medical Internet Research*. Vol. 12, No. 2, e18, 2010.
- [10] Differences among Discussion Boards, http://www.adelaide.edu.au/myuni/staff/resources/tutorials/content/Differences_between_Discussion_Boards_Blogs_and_Wikis.html
- [11] J. Woo, M. Lee, Y. Ku, and H. Chen, "Modeling the dynamics of medical information through web forums in medical industry", *Technological Forecasting and Social Change*, In press, 2013.
- [12] J. Woo and M. Lee, "The Informative Support and Emotional Support Classification Model for Medical Web Forums using Text Analysis", *Korea Society of IT Services*. Vol. 11, pp. 139-152. 2012.
- [13] <http://www.hani.co.kr/arti/society/rights/686880.html>
- [14] H. Jeong, S. Ji, H. Yang, K. Kim and K. Kim, "Measuring Influence of Social Networks based on Sentiment Analysis", *Proceedings of Korean Association of Computer Education*, Vol. 19, No. 1, 2015.
- [15] S. Cobb, "Social support as a moderator of life stress, *Psychosom. Med.* Vol. 38 pp. 300 - 314, 1976.
- [16] J.S. House, C. Robbins, H.L. Metzner, "The association of social relationships and activities with mortality: prospective evidence from the Tecumseh Community Health Study", *Am. J. Epidemiol.* Vol. 116, No.1, pp. 123 - 140, 1982.
- [17] N.S. Coulson, H. Buchanan and A. Aubeeluck, "Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group", *Patient Educ. Couns.* Vol. 68, No. 2, pp. 173 - 178, 2007.
- [18] N.S. Coulson, "Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome", *Cyberpsychol. Behav.*, Vol. 8, pp. 580 - 584, 2005.
- [19] M.D. Wenrich, J.R. Curtis, D.A. Ambrozy, J.D. Carline, S.E. Shannon and P.G. Ramsey, "Dying patients' need for emotional support and personalized care from physicians: perspectives of patients with terminal illness, families, and health care providers", *J. Pain Symptom Manag.* Vol. 25 No. 3, pp. 236 - 246, 2003.
- [20] J. Georges, S. Jansen, J. Jackson, and et al. "Alzheimer's disease in real life - the dementia carer's survey", *Int J Geriatr Psychiatry*, Vol. 23, No. 5, pp 546 - 551, 2008.
- [21] H. Cho, Y. Chung, J. Lee and J. Lee, "Sentiment Analysis Using News Comments for

- Public Opinion Mining”, Proceedings of KIIS Spring Conference, Vol. 23, No. 1, 2013.
- [22] D. Lee, H. Kang, S. Kim and C. Lee, “Autocorrelation Analysis of the Sentiment with Stock Information Appearing on Big-Data”, The Korean Journal of Financial Engineering, Vol. 12 No. 2, pp. 79-96, 2013.
- [23] J. Yeon, D. Lee, J. Shim and S. Lee, “Product Review Data and Sentiment Analytical Processing Modeling”, The Journal of Society for e-Business Studies”, Vol. 16 No. 4, 2011.
- [24] S. Kweon and Y. Choi, “Rating Analyses Used Big Data-Empathy Word in Network Analysis Method”, Proceedings of KOREA INFORMATION SCIENCE SOCIETY, 2014.
- [25] J. Bollen, H. Mao, and X.J. Zeng, “Twitter Mood Predicts the Stock Market”, Journal of Computational Science, Vol. 2 No. 1, pp. 1-8, 2010.
- [26] B. Malcolm, and J. Wurgler “Investor Sentiment in the Stock Market”, Journal of Economic Perspectives, Vol 21, pp. 129-151, 2007.
- [27] Y. Kim, S. Jeong, S. Lee, “A Study on the Stock Market Prediction Based on Sentiment Analysis of Social Media”, Entrue Journal of Information Technology, Vol, 13, No, 3, pp. 59-69, 2014.
- [28] M. Park, M. Cha, and C. Cha, “Depressive moods captured in Twitter, an online social network”, In Proc. of the 1st Congress for Qualitative Health Research, 2011.
- [29] H. Choi, E. Hwang. “Emotion-based Music Recommendation System based on Twitter Document Analysis.” Journal of KISS (C): Computing Practices and Letters, Vol. 18, No. 11, pp. 762-767, 2012
- [30] R. Feldman, “Techniques and applications for sentiment analysis”, Communications of the ACM, Vol. 56, no. 4, pp. 82-89, 2013.
- [31] General Inquirer, <http://www.wjh.harvard.edu/inquirer/>
- [32] G. Psathas, “The general inquirer: Useful or not?”, Computers and the Humanities, Vol. 3 No. 3, pp. 163-174, 1969.
- [33] P.J. Stone, D.C. Dunphy, M.S. Smith, and D.M. Ogilvie, “The General Inquirer: A computer approach to content analysis. Cambridge”, MA: MIT Press, 1966.
- [34] T. Loughran, and B. McDonald, “When is a liability not a liability? Textual analysis, dictionaries, and 10 Ks”, The Journal of Finance, Vol. 66 No. 1 pp. 35-65, 2011.
- [35] M. Guerini, L. Gatti and M. Turchi, “Sentiment analysis: How to derive prior polarities from SentiWordNet”, arXiv preprint arXiv:1309.5843, 2013.
- [36] J. H. Fowler and N. A. Christakis, “Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais”, Journal of health economics, Vol. 27, No. 5, pp. 1400-1405, 2008.
- [37] W. Lee, M. Cha, and H. Yang, “Network characteristics of influentials in social media.” Journal of Communication Research, Vol. 48, No. 2, pp. 44-79, 2011.
- [38] Y. Ki, J. Woo, and H.K. Kim, “Identifying spreaders of malicious behaviors in online games”, In Proceedings of the companion publication of the 23rd international conference on World wide web companion pp. 315-316, 2014.
- [39] M.S. Kim, and W.S. Woo, “Analysis of SNS(Social Networking Service) functions applicable to electronic commerce for building regular relationship with customers”, Journal of the Korea society of computer and information , Vol. 20 No. 4, pp.131-138, 2015
- [40] M.J. Lee, J.W. Kim, “Design and Implementation of the Menu Navigation using Social Network Analysis among the Menus of Management Information System”, Journal of the Korea society of computer and information, Vol. 19, No. 9, pp. 151-160, 2014
- [41] Y. Kim, Social Network Anaysis, Pakyoungsa,

- 2011.
- [42] P. Bonacich, "Power and centrality: A family of measures", American journal of sociology, pp. 1170-1182, 1987.
- [43] J. Donath, and D. Boyd, "Public displays of connection", BT Technology Journal, Vol. 22, No. 4, pp. 71-82, 2004.
- [44] B. Wellman, A.Q. Haase, and J. Witte, and K. Hampton, "Does the Internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment," American Behavioral Scientist, Vol. 45, No. 3, pp. 436, 2001.
- [45] R. Rose, "How much does social capital add to individual health?", Social science & medicine, Vol. 51, No. 9, pp. 1421-1435, 2000.
- [46] C. Steinfield, N.B. Ellison, and C. Lampe, "Social capital, self-esteem, and use of online social network sites: A longitudinal analysis", Journal of Applied Developmental Psychology, Vol. 29, No. 6, pp. 434-445, 2008.
- [47] S. Park, S.W. Lee, J. Kwak, and M. Cha, & B. Jeong, B, "Activities on Facebook reveal the depressive state of users", Journal of medical Internet research, Vol. 15, No. 10, 2013.

저 자 소 개



이 민 정

1999: KAIST 재료공학과 공학사.
 2001: KAIST 재료공학과 석사.
 2008: KAIST 산업공학과 박사.
 현 재: 세종사이버대학교
 경영학과 교수
 관심분야: MIS, CRM, 전자상거래,
 품질경영, 기술경영,
 데이터마이닝
 Email : mjleekorea@gmail.com



우 지 영

2000: KAIST 산업공학과 공학사.
 2002: KAIST 산업공학과 석사.
 2006: KAIST 산업공학과 박사.
 현 재: 고려대학교
 정보보호대학원 연구교수
 관심분야: 데이터사이언스,
 온라인게임, 소셜미디어,
 정보보호
 Email : jywoo@korea.ac.kr