

의료 웹포럼에서의 텍스트 분석을 통한 정보적 지지 및 감성적 지지 유형의 글 분류 모델

우지영* · 이민정** · Yungchang Ku***

The Informative Support and Emotional Support Classification Model for Medical Web Forums using Text Analysis

Jiyoung Woo* · Min-Jung Lee** · Yungchang Ku***

■ Abstract ■

In the medical web forum, people share medical experience and information as patients and patents' families. Some people search medical information written in non-expert language and some people offer words of comfort to who are suffering from diseases. Medical web forums play a role of the informative support and the emotional support. We propose the automatic classification model of articles in the medical web forum into the information support and emotional support. We extract text features of articles in web forum using text mining techniques from the perspective of linguistics and then perform supervised learning to classify texts into the information support and the emotional support types. We adopt the Support Vector Machine (SVM), Naive-Bayesian, decision tree for automatic classification. We apply the proposed model to the HealthBoards forum, which is also one of the largest and most dynamic medical web forum.

Keyword : Informative support, Emotional support, Medical Web Forum, Classification, Text Analysis

논문투고일 : 2012년 09월 14일 논문수정완료일 : 2012년 10월 25일 논문게재확정일 : 2012년 11월 23일

* 고려대학교 정보보호대학원 연구교수

** 세종사이버대학교 경영학과 교수, 교신저자

*** Yuan Ze University Information Management 박사과정

1. 서 론

최근 인간의 기대수명이 연장됨에 따라 헬스케어 산업은 주목을 받고 있다. 노령화에 따른 의료 산업수요가 증가하고 있으며 기대 수명의 증가로 인해 젊은 층 역시 의료정보에 대해 많은 관심을 가지게 되었다. 웹 2.0시대에 사람들은 건강에 관련된 유용한 정보를 얻기 위해, 블로그, 위키, 웹포럼 등의 웹 기반 소셜미디어를 이용하고 있다. 컴퓨터 기술과 인터넷의 발전은 이러한 현상을 가속화시키고 있다. 최근 IT 기술이 의료 분야에서 중요한 역할을 하게 되면서, 두 산업이 융합되고 있다. 대규모의 의료정보를 효율적으로 활용하고자 하는 Health 2.0은 혁신적인 의료 IT 융합 기술 중의 하나이다[2]. 특히 웹 2.0기술을 이용한 웹포럼¹⁾에서는 특정 주제에 대한 풍부한 정보와 이에 대한 집중적인 논의를 다루고 있기 때문에, 웹포럼은 전문 지식을 구하고 전문가의 견해를 들을 수 있는 중요한 채널 중에 하나이다[3].

Elkin[15]의 조사에 따르면 미국인들의 절반 이상이 의료정보를 얻기 위해 인터넷을 이용하며, 의료정보를 조회하는 사람들의 3분의 1은 위키피디아와 웹포럼 등을 이용하여 질병의 정보, 증상 및 처방에 대해서 알아보는 것으로 나타났다. 특히 웹포럼이나 의료정보사이트의 메시지보드는 전문가의 의료정보를 제공하는 역할에서 더 나아가 사용자들 간의 정보를 공유하는 등의 의사소통의 장으로 활용되고 있다. 일반인은 전문적인 의료정보를 이해하기 어려운 경우가 많아 쉬운 언어로 설명된 전문적인 의료정보뿐만 아니라, 경험적 지식을 얻기 위해서 웹포럼을 이용한다. 또한 같은 질병을 앓는 환자의 글을 통해 해당 질병에 대한 경험적 지식을 얻을 수도 있고 추가적인 질문을 통해 다른 도움을 받을 수도 있다. 또한 에이즈나 치매 등의 불치병 혹은 중대질병을 앓고 있는 경우에는 자신

과 유사한 질병을 가진 사람들로부터 위로 받고 싶어 한다. 이들은 의료 웹포럼을 통해 자신이 처한 상황을 공유하고 다른 이들로부터 위로를 받기도 한다.

이러한 이유들로 인해 많은 의료정보 사이트는 웹포럼, 디스커션 포럼, 헬프라인 등의 상호지원적인 의사소통이 가능한 공간을 제공하고 있다[38]. 상호지원적인 커뮤니케이션은 사용자간의 유대감을 형성하는 기능을 할 뿐만 아니라, 사람들이 그들의 의료상황을 처리하는데 더 좋은 결과를 내도록 하는 역할을 하는 것으로 알려져 있다[29].

의료정보는 사람의 생명과 직접적으로 연결되어 있는 경우가 많아서 의료 웹포럼은 다른 웹포럼에 비해 사용자간의 상호작용이 강하게 일어난다. 이러한 상호작용은 정보적 지지 형태이거나 감성적 지지 형태를 가진다. 웹 2.0 기술 기반의 의료 웹포럼은 환자, 의료진과 의료정보에 관심이 많은 일반인으로 구성된 온라인 지원그룹[28]이 함께 지식을 만들어 나가고, 서로 간의 상호작용을 통해 콘텐츠를 생성하고, 오류를 수정하며, 지식을 확대하여 저장해 놓을 수 있는 하나의 창구가 되었다. 이 의료 온라인 지원그룹은 웹포럼에 지식을 쌓는 주체로서 도움을 얻기 힘든 사람들에게 필요한 정보를 제공하는 역할을 하며[37], 지역적, 시간적 제약 없이 도움을 줄 수 있다[14]. 또한 감추고 싶은 질병을 가지고 있는 환자들은 익명성이 보장됨에 따라 민감한 이슈에 문의할 수 있고, 온라인 지원 그룹이 제공하는 정보를 보다 안전한 환경에서 얻을 수 있다[8]. 이런 익명성은 온라인 지원그룹이 활동하는 웹포럼에서는 환자들은 의료정보뿐만 아니라 그들의 감정에 대해서 자유롭게 표현할 수 있는 환경을 만들어 주는데 중요한 역할을 한다. 웹포럼 상의 사용자간의 이러한 상호작용은 사회적 지지로 설명될 수 있는데, 이러한 사회적 지지[10]는 의료 온라인 지원그룹이 자신은 보살핌과 사랑을 주고받는 조직망 속의 일원이라는 소속감을 준다. 사회적 지지의 주요 4가지 요소로는 정보, 감성, 도구, 평가가 있는데 환자들에 있어서 절실하게 필요한 요소는 다음과 같은 두 가지이다. 첫 번

1) 웹포럼은 토론을 열어서 직접 제작물을 올리는 웹 응용 프로그램이다.

째로는 질병의 정보적 지지를 통해 의료정보를 알아내는 것이며 두 번째로는 감성적 지지를 통해서 환자가 질병으로부터 고통 받고 있는 상황을 다른 온라인 지원그룹으로부터 위로 받고, 질병이 치료 될 것이라는 희망을 갖게 되는 것이다[29]. 이에 따라 웹포럼 사용자의 글은 주로 정보적 지지의 성격인 글과 감성적 지지의 성격인 글로 나눌 수 있다.

본 연구에서는 텍스트 특성을 이용하여 대량의 의료 웹포럼의 게시글을 정보적 지지, 감성적 지지 유형으로 자동적으로 탐지하여 분류하는 방법론을 제시하고자 한다. 제 2장에서는 의료정보의 정보적 지지와 감성적 지지에 대한 연구들과 본 연구와 관련된 텍스트 마이닝 방법론을 정리했다. 제 3장에서는 의료 웹포럼의 게시판의 글들을 자동으로 정보적 지지 유형과 감성적 지지 유형으로 탐지하기 위해서 텍스트 특성을 추출하여 의료정보를 정보적 지지 및 감성적 지지 형태로 분류할 수 있는 모델을 제시하였다. 제 4장에서는 제시하는 모델을 실제 의료 웹포럼에 적용한 결과를 비교분석하여 가장 나은 모델을 제시하였다.

2. 문헌 연구

2.1 의료정보의 정보적 지지와 감성적 지지

Weiss[36]와 Cobb[10]가 사회적 지지의 주요 요소를 정보적 지지(Information support), 감성적 지지(Emotional support), 도구적 지지(Instrument support)와 평가적 지지(Appraisal support)로 소개하였다[12]. 정보적 지지는 충고나 안내 등의 필요한 정보를 제공하며[1, 12, 13], 감성적 지지는 신뢰하면서 공감하는 등의 감성적인 표현을 의미하며, 개인들이 관심과 사랑의 감정을 느낄 수 있는 친밀감, 밀착, 동정, 관심, 배려 등을 포함한다. 감성적 지지는 위기 상황에 있는 개인에게 다른 사람이 도와 줄 것이라는 확신을 주며, 두려움이나 불안감을 감소시켜주는 역할을 한다. 도구적 지지는

금전을 포함한 물질적인 지원을 의미하며, 마지막으로 평가적 지지는 긍정적 칭찬 등을 제공함으로써 피드백을 하는 평가 프로세스를 나타낸다.

사회적 지지가 개인의 심리에 미치는 영향에 관한 많은 연구들에서 개인이 사회적 지지로부터 받는 정신건강, 스트레스 대처, 그리고 인생의 여러 가지 변화에 대한 적응에 긍정적인 영향을 미친다고 밝혀왔으며, 또한 사회 구성원들과의 사회적 관계에 의해 제공되는 많은 예방적인 기능을 가진다고 보고하였다[5, 12, 13]. Gottlieb et al.[17]는 사회적 작용은 상호적 표현과 상호작용을 하는 개인들이 영향을 통해 지원 행위로부터 나타나는 관계임을 밝혔다.

이러한 사회적 지지에 대해서 국내에서도 많은 연구가 진행되었는데 임화영 외[4]는 근로자를 대상으로 사회 심리적 스트레스 관련 요인과 사회적 지지 간의 관련성을 살펴보고, 근로자의 스트레스를 줄이기 위해서는 사업장 내에서 동료나 상사의 지지 및 지역사회에서 형성되는 사회적 지지가 증대되어야 한다고 제시하였다.

권예지 외[1]는 건강 관련 정보습득 과정에서의 정보원의 공신력과 이성/감성적 메시지가 이용자의 사회적 지지(감성적 지지와 정보적 지지), 태도, 건강행동[22]에 미치는 영향을 분석하여 온라인에서의 건강정보에 대한 효율적인 지식답변 운영 및 온라인에서 요구되는 의사의 역할, 사회적 지지와 건강행동의도 형성 경로 등에 대해 논의하였다. 특히 건강과 관련된 메시지는 정보위주의 이성적 지지와 감성적 지지의 부분이 핵심적이라는 점을 고려하여 연구했다.

건강함을 유지하기 위해 혹은 건강해지기 위해서 의료정보를 찾는 행위는 건강행동 중에 하나이며[1, 23] 이전의 연구들은 건강행동과 건강행동의도에 영향을 주는 대표적인 요인으로 사회적 지지를 뽑을 수 있으며 사회적 지지와 건강행동 간의 상관관계가 존재함을 밝히고 있다[19, 20]. 특히 의료정보를 제공받음으로 인해서 질병으로 인한 두려움을 줄여서 건강해질 수 있으며, 질병에 대한

두려움과 슬픔을 함께 나누며 이를 극복할 수 있다는 것을 많은 연구에서 보여주고 있다[9, 34]. 즉 많은 연구에서 건강하기 위해서 의료정보들을 찾아보는 행위는 건강행동의도로 볼 수 있으며 사회적 지지 중에서 감성적 지지와 정보적 지지가 밀접하게 관련되어 있다는 것을 보여주고 있다.

1990년대의 사회적 지지는 전화상담 서비스나 전문가들이 제공하는 고객맞춤형 서비스의 형태였으나, 2000년대에 접어들면서 많은 웹사이트에서 질병에 대해 지원하는 커뮤니티들²⁾이 생겨 오프라인에서 행해지던 사회적 지지가 온라인에서도 가능하게 되었다. 환자와 그의 가족들은 질병과 치료에 대한 지식과 경험을 웹에서 공유하면서 서로에게 도움이 되고 있다. 이러한 웹사이트를 통한 메시지 교환은 다양한 메시지 유형으로 정보적 및 감성적 지지를 얻는 과정에 영향을 준다[30]. 2005년 이후의 사회적 지지 커뮤니티들은 웹 2.0기술을 기반으로 동적 콘텐츠, 협업 지식[6]을 구축하는 체계를 구축하였고, 개인들이 정보, 경험, 의견들을 공유하고 이들 간의 상호 작용들을 통해 풍부한 지식들을 보유할 수 있게 되었다. 의료 산업 분야 역시 웹을 통한 정보교환이 더욱더 활성화 되고 있다.

생명을 위협할 수 있는 질병들은 심하게 감성적으로 고통을 환자들에게 가하기하며, 치료과정에서 따른 우울증 등으로 나타나기도 한다[7]. 따라서 의료정보 웹사이트에 접속하는 고객들은 물리적으로 질병을 치료하고자 하는 욕망을 가질 뿐만 아니라 유사한 질병을 극복한 환자들의 포스팅한 글이나 자신이 게시한 글에 응원을 보내는 다른 이들의 글들을 보면서 질병에 대한 강한 회복의 의지를 가질 수도 있다.

의료 웹포럼의 경우에 있어서는 의료 서비스 관련 전문가, 환자와 그 가족들이 주로 이용함에 따라 질병과 치료에 대한 정보를 얻는 것은 정보적 지지 관계로 볼 수 있으며, 질병을 극복한 사례

라든지, 다른 사람들이 질병을 극복할 수 있도록 응원의 메시지들을 공유하는 것은 감성적 지지 관계로 볼 수 있을 것이다. Mo and Coulson[28]는 온라인 HIV/AIDS 그룹에 대해서 사회적 지지의 유형별로 콘텐츠 분석을 하여 정보적 지지와 감성적 지지를 제공하는 메시지들은 존중감 지지와 네트워크 지지를 따른다는 것을 밝혔다. 즉 의료 웹포럼의 주된 이용 목적은 도구적 지지나 평가적 지지 보다는 질병을 치료하기 위한 방법을 공유하는 정보적인 지지와 질병을 극복하기 위한 감성적인 지지를 받는 것이 주된 목적이라고 했다. 즉 건강행동을 할 때는 사회적 지지가 영향을 주며, 사람들이 질병을 가졌을 때, 다른 사람의 위로를 받거나, 치료나 회복을 위해서 실질적인 정보를 얻고 싶어 하는 것을 알 수 있다.

이와 같이 의료 관련 웹, 포럼, 게시판 등의 다양한 온라인 채널 등의 정보들과 사회적 지지와의 관계들을 살펴본 연구들은 수행되어 왔지만, 온라인 상의 정보들을 감성적 지지와 정보적 지지로 분류하는 방법론에 대한 연구는 거의 없었다.

2.2 웹 기반의 텍스트 분석

텍스트 마이닝은 비정형적인 글들을 자연어 처리 기술과 문서 처리 기술을 이용하여 유용한 정보를 추출하고 찾아내는 기술이다. 최근에는 많은 사람들이 신제품, 영화, 기업, 드라마, 연예 등에 관련한 개인의 생각 및 관련 정보들을 웹포럼이나 블로그등에 게시하고 있고, 이들 글에 대한 의도를 분석하기 위해서 텍스트 정보 분석에 대한 연구가 활발히 진행되고 있다. 웹블로그나 웹포럼으로부터 상품평에 대한 정보를 파악하고[16], 블로그에서 해당상품에 대해 언급한 수 및 블로그간의 링크구조와 판매와의 관련성[18] 등에 대해 파악하여 텍스트 정보와 기업의 성과와의 연관성을 찾아내는 연구들이 활발히 진행되고 있으며 웹포럼의 글을 텍스트 특성을 분석하여 스팸글을 자동탐지하는 모델도 제시되었다[3]. Koppel et al.[24]은 ex-

2) American Social Health Association, American Cancer Society, American Diabetes Association.

ponential gradient algorithm을 이용하여 허구와 실화의 문서에서 성별을 분류하는 모델을 만들었다. 해당 모델은 기능어와 POS(Part of Speech) 태그를 조합한 특성 조합을 통해 허구 문서에 대해서는 79.5%의 정확도를 나타냈으며, 실화 문서에서 82.6%의 정확도를 나타냈다. 저자는 특성 선택을 통해서 98%까지 정확도를 개선시켰다. Zhang[40] 외는 웹포럼 글을 텍스트 특성 분석을 통해서 글을 게재한 사람의 성별을 최고 86%의 정확도로 분류할 수 있는 모델을 제시하였다.

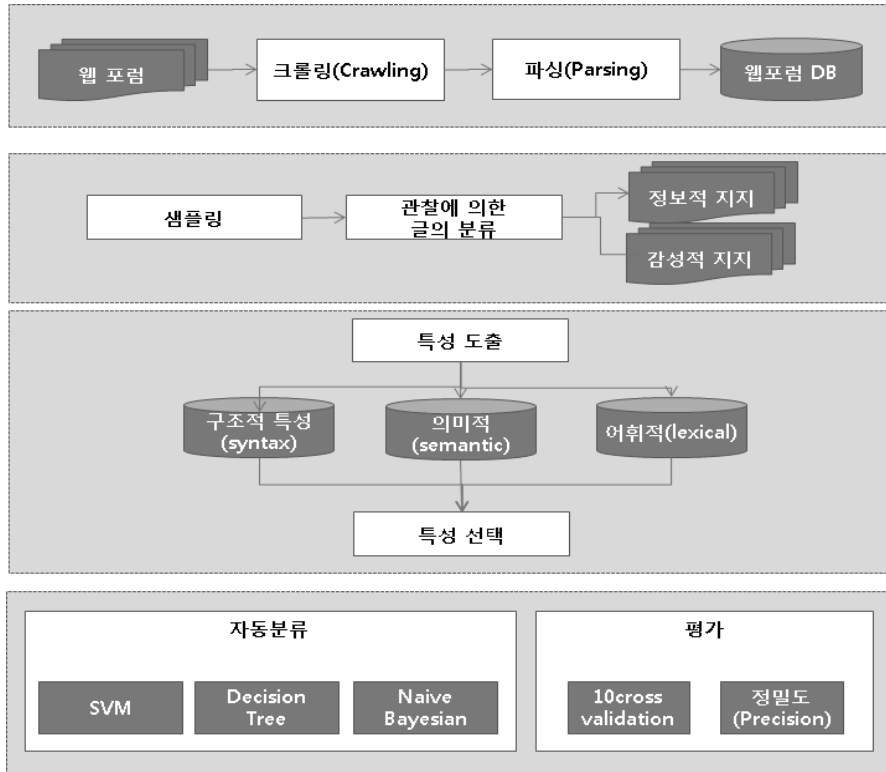
최근에는 의료정보에 대한 관심이 높아짐에 따라 의료 웹포럼상의 메시지들을 분석하는 연구들이 진행되어 오고 있다. Mo and Coulson[28]은 질병에 대한 블로그를 자동적으로 텍스트 분석을 통해 분류함으로써 사람들이 암에 관한 요구하는 정보는 무엇이고, 서로 간에 어떻게 관련 정보를 공

유하며 도움을 받는지에 대해 살펴보았다. Hwang [21]은 비만과 관련한 온라인 포럼, 설문 등의 분석을 통해 사회적 지지에 대해서 분석을 수행하였다. 이와 같이 최근 웹 기반의 미디어의 텍스트 분석은 의료 분야에서도 활발하게 진행되고 있으나 텍스트 분석의 정확성에 관해 언급한 연구는 찾기 어려운 실정이다.

본 연구에서는 의료 웹포럼의 글들을 사회적 지지 중에서 정보적 지지와 감성적 지지를 자동으로 분류하는 텍스트 분석 모델을 개발하였다.

3. 연구방법

본 연구에서는 의료 웹포럼에서 글의 특성을 자동 구분하는 모델을 제안한다. 제안하는 모델의 전체적인 프로세스는 [그림 1]과 같다.



[그림 1] 연구모형

3.1 데이터 수집

대부분의 웹포럼은 API(Application Program Interface)를 통해 데이터플랫폼을 외부에 공개하기 때문에 이를 이용하여 웹포럼으로부터 자동화된 데이터 수집이 가능하다. 웹포럼상의 글을 자동으로 수집하기 위해서 웹 크롤러(Crawler)를 이용한다. 웹 페이지는 각각의 페이지가 하이퍼링크로 연결되어 있기 때문에 웹 크롤러는 웹의 그래프 구조를 이용하여 데이터를 수집하도록 구현된다. 웹 크롤러를 이용해 수집된 데이터는 html 형태인데, 이 html 형태의 데이터로부터 필요한 필드를 추출하여 데이터베이스화하는 과정을 거친다. 이 과정을 파싱(parsing)이라고 한다. 본 연구에서는 글의 텍스트적 특성만을 이용하기 때문에 글의 번호 및 글 내용의 두 개의 필드를 파싱 과정을 거쳐 도출하고, 이를 데이터베이스화 한다.

3.2 변수 추출

본 연구는 글의 텍스트 특성을 이용해 게시글을 정보적 지지와 감성적 지지 유형으로 구분하고자 한다. 이를 위해 언어학에서 사용되는 글 분류 특성을 이용하여 텍스트 변수를 도출한다. 본 연구에서 사용하는 글의 텍스트 변수는 구조적(syntactic), 의미적(semantic), 어휘적(lexical) 특성이다. 기존의 텍스트 분석에 관련한 연구에서는 의미적 특성 즉, 단어나 단어의 조합을 주로 이용하였다. 본 연구에서는 이외에도 문장의 구조적 특성 및 어휘적 특성을 추가하여 다양한 관점에서의 텍스트 특성을 나타내는 변수를 구축하도록 한다.

구조적 특성이란 단어가 결합하여 형성되는 구나 절, 문장의 구조와 기능을 측정하는 변수이다. 본 연구에서는 아래와 같은 특성을 구조적 특성으로 사용한다. 문장 구두점(punctuation)은 문장의 끝에 붙는 부호로 마침표, 쉼표, 의문표, 따옴표 종류에 따라 글을 분류하는 변수이다. 제외어를 포함하고 있는지를 나타내는 변수도 사용한다. 제외어란 전치사, 관계대명사 등 의미가 확연히 구분

되지 않아 텍스트 분석에서 제외하는 단어를 말한다. 해당 변수는 글이 얼마나 많은 정보를 포함하고 있는지를 나타낸다. 어근(stemming)은 형태는 다르지만 같은 어근을 갖는 단어가 문장 내 나타나는 빈도를 측정하는 변수이다. 글의 길이(line length) 또한 글의 깊이와 상세함을 나타내는 평가 기준이다. 의미적 특성이란 뜻이 서로 다른 단어를 나타낸다. 의미적 특성은 보통 몇 개 길이의 단어 조합이 사용되는데, 본 연구에서는 unigram과 bigram을 사용한다. Unigram은 하나의 단어를, bigram은 두 개의 연속된 단어의 조합을 의미한다. 예를 들어, 감성은 unigram, 감성적 지지는 bigram이다. 어휘적 특성이란 문법적으로 의미를 가지는 최소 형태 즉, 형태소의 결합 형태를 의미한다. 예를 들어 “지원한다”는 “지원”과 “한다”의 형태소의 결합이다. 즉, 문법적으로 의미를 가지는 두 개의 단어, “지원”이라는 명사와 “한다”라는 동사의 결합으로 이루어진 것이다. 어휘적 특성은 형태소의 결합, 즉 Part of speech(POS)로 측정한다. POS는 bigram과 유사하나 문법적으로 구분되는 단어의 조합이라는 점에서 다르다. 이는 글의 구문론적 특성을 나타낸다. 구문론적 구분은 보통 동사, 명사, 형용사, 부사 등이다. 따라서 POS는 동사+부사, 형용사+명사, 명사+명사 등의 조합인지를 나타내는 특성이다.

본 연구에서는 포럼 글의 텍스트를 분석하여 글별로 구조적, 의미적, 어휘적 특성을 변수로 추출한다. 이를 위해 본 연구에서는 Java로 텍스트 특성 추출 도구를 구현하였다. 글의 텍스트적 특성을 도출하면 많은 변수가 생성되는데, 글별로 구축된 변수 set은 특성별로 해당 글에서 발생한 빈도로 변수를 형성한다. 즉 예를 들어 구조적 변수인 “~”, “?”가 글 1번에서 몇 번 발생했는지, 의미적 변수인 “웹사이트”, “병원”등이 해당 글에서 몇 번 발생했는지, 어휘적 특성인 “명사+명사”, “동사+부사” 등의 특성이 몇 번 발생했는지를 변수로 한다.

텍스트 특성 추출 단계를 거치고 나면 특성 선택(feature selection) 과정을 거친다. 글의 텍스트

적 특성을 도출하면 많은 변수가 생성되는데, 이중 일부는 노이즈이거나 다른 변수와 중복되거나 또는 필요 없는 변수이다. 또한 이렇게 많은 변수를 이용하여 분류기 알고리즘을 학습시키면 분류 모델이 과잉 학습되어 예측력을 저하될 수 있다 [27]. 이와 같은 이유로 데이터의 차원을 줄이고 필요한 변수만을 추출하고자 특성 선택 과정을 거치는 것이 필요하다.

3.3 학습 모델 구축

글을 자동으로 분류하는 모델을 학습시키기 위해서는 글의 클래스(class)를 사전에 분류한 학습 데이터가 필요하다. 학습 데이터는 전체 데이터 중 일부를 샘플링하여 학습 데이터로 활용한다. 학습 데이터는 글의 클래스가 사전에 분류되어 있어야 하는데, 이를 위해서 평가자가 글의 특성을 평가하도록 한다. 평가자의 주관에 의해 발생하는 편차를 최소화하고, 평가자에 의한 평가의 객관성을 부여하고자 2명 이상의 평가자에 의해 평가가 이루어지게 한다. 여러 평가자의 결과가 유사하게 나타난다면, 평가자에 의해 이루어진 평가가 객관적으로 신뢰할 만하다고 할 수 있다. 평가자에 의한 분류 결과의 유사성을 측정하기 위해 일치도를 측정한다[11].

학습 데이터가 구축되고 나면, 글의 특성을 분류하는 자동화된 모델을 구축하기 위하여, 분류기(classifier)를 학습시킨다. 글의 클래스를 분류하는 모델은 이진 분류 학습으로 구현된다. 글의 클래스를 정보적 지지와 감성적 지지로 분류하고, 이 클래스를 구분하는 규칙을 글의 특성을 통해 도출하는 것이다.

최적의 학습 모델을 찾기 위해서, 다양한 분류기를 테스트하는 것이 필요하다. 데이터마이닝 및 비정형화된 데이터를 다루는 텍스트 마이닝에서 그 성능이 입증된, SVM(Support Vector Machine)[35] 중에 한 종류인 SMO(Sequential Minimal Optimization)를 기본 모델로 하고, 동작 방식이 서로 다른 여러 개의 분류기를 테스트한다. 일반적으로

SVM은 우수한 성능을 보이지만 계산량이 많은 단점을 지니며 이를 개선하기 위한 방법으로 SMO 방식이 우수한 성능을 나타낸 것으로 알려져 있다 [31]. 본 연구에서는 테스트 분류기로 나이브-베이시안(Naïve Bayesian)[26], 의사결정나무(Decision Tree)[32]를 이용하였다. 나이브-베이시안은 변수 간 독립성을 가정한 확률 모형인데, 많은 연구에서 복잡한 시스템에서의 우월한 성능이 입증되었다. 해당 방법의 장점은 적은 수의 데이터로도 학습이 가능하다는 점이다. 의사결정나무는 엔트로피를 감소시키는 방향으로 변수의 나무구조(tree)를 구성하는 알고리즘이다. 위의 두 모델과는 달리 모든 변수를 사용하는 것이 아니라 엔트로피 감소량을 기준으로 트리를 구성하는 변수가 선택되는 특성을 가진다.

많은 변수를 사용할 경우 분류 모델은 학습 데이터에만 최적화되어 학습 데이터에서는 좋은 성능을 보이지만, 테스트 데이터에서의 성능은 저조하게 나타나는 현상이 발생한다. 이를 방지하기 위해 10-fold cross-validation 방식을 도입한다[33]. 10-fold cross-validation에서는 학습 데이터를 10개의 그룹으로 분류하고, 하나의 그룹을 테스트 데이터로 나머지는 학습 데이터로 활용된다. 10개의 그룹을 분류할 때 데이터를 무작위로 그룹에 할당하고, 그룹의 크기는 서로 균등할 수 있도록 한다. 테스트 데이터 셋을 변경해가면서, 이러한 학습 과정을 10번 반복하여, 분류기가 특정 데이터에만 적합하게 학습되는 현상을 방지한다.

학습에 의해 구축한 분류모델의 성능을 평가하기 위해 분류 모델에서 사용되는 평가 척도로는, 정확도를 사용한다. 각 클래스의 수가 비슷한 경우에는 정확도 값으로 학습모델을 평가하는 것이 타당하다. 정확도는 평가자에 의해 분류된 글을 학습 모델로 일치되게 분류하는 정도를 나타낸다. 식은 아래와 같다.

$$\text{정확도(Accuracy)} = \frac{\text{제대로 분류된 글의 수}}{\text{전체 글의 수}} \quad (1)$$

4. 의료 웹포럼에의 적용

본 연구에서 제시하는 모델을 의료 웹포럼 중의 하나인 HealthBoards³⁾내 소개시판인 Alzheimer's Disease and Dementia 게시판의 데이터에 적용한다. 해당 게시판에서 2000년도 12월 29일부터 2007년도 5월 6일까지 작성된 150개의 원글과 답글 638개를 포함하여 총 799개의 글을 추출하였다. 치매는 뇌기능의 손상으로 인해 인지 기능이 저하되는 질병이다. 이 질병은 노인 인구에서 발생 확률이 높으며, 현재까지는 불치질병으로 분류되어 있다. 질병 발생 시 환자 본인뿐만 아니라 주변 가족에게 정신적 물질적으로 큰 과장을 주기 때문에 환자 가족들의 정신적인 충격도 심각하다. 이러한 이유로 환자 및 환자의 가족들은 치매에 대한 경험적인 지식을 얻고, 또한 감성적인 공감 및 위로를 얻기 위해 해당 포럼을 이용한다.

웹포럼 글의 사회적 지지의 자동화된 분류 모델 구축을 위해서 두 명의 평가자를 통해 글을 코딩하게 하였다. 본 연구는 텍스트를 분류하는 연구[22, 39]에서 사용된 코딩 과정을 따랐다. 코딩은 자료 처리를 자동화하기 위해 일정한 규칙에 따라 품목별로 대상번호 또는 문자를 부여하는 것을 말하며, 주로 질적 데이터(qualitative data)에 label을 부여하기 위해 사용된다. 기존의 연구에서 사용된 일반적인 코딩 기법은 사전에 분류에 대한 프로토타입을 구성하고, 두 명 이상의 평가자를 교육한 후 코딩이 이루어지게 하며, 결과의 신뢰도를 확인하기 위해 평가자간의 일치도를 측정한다.

본 연구에서는 포럼 글을 추출하여 감성적 지지와 정보적 지지로 분류한 프로토타입을 만들었다. 코딩 과정은 평가자가 해당 웹포럼에 대한 배경지식을 습득한 후, 감성적 지지 및 정보적 지지의 의미를 파악하게 한 후 실시하였다. 사전 테스트를 거쳐 평가자의 글 분류의 성능을 확신한 후에 평가를 진행하였다. 평가의 객관성을 평가하고자 두

평가자의 평가 결과의 일치도를 Cohen Kappa[11] 지수로 측정하였다.

〈표 1〉 평가자 일치도

	데이터 수	확률	
일치도 (agreement)	550	68.84%	p(agreement)
오류(error)	249	31.16%	p(error)

이때 p(agreement)는 평가자의 일치도를 의미하고 p(error)는 우연적으로 평가가 일치할 확률을 의미한다. Cohen Kappa 지수는 다음의 식 (2)에 의해 계산된다.

$$k = \frac{p(\text{agreement}) - p(\text{error})}{1 - p(\text{error})} \quad (2)$$

두 평가자에 의한 결과는 68.84%로 일치하는 것으로 나타났다. 이를 이용해 Cohen Kappa 값을 구하였고, 그 값은 54.7%로 나타났다. 이 값의 수준은 중간정도의 일치도를 보이는 수준이다.

Landis and Koch[25]의 연구에 따르면 Kappa 값이 0이면 평가자간의 일치 수준이 0이며, 0.20 이하일 경우에는 일치 수준이 조금 있는 수준이고, 0.40 이하인 경우에는 fair 수준이고 0.41~0.60은 적절한 수준, 0.61~0.80은 상당한 수준이고, 0.81~1은 거의 완벽한 일치한다고 할 수 있다.

학습 데이터 생성은 두 평가자가 감성적 지지 글로 일치하게 평가한 글만을 감성적 지지 글로 분류하였고, 기타 글은 정보적 지지 글로 분류하였다. 여러 실험 조합에 대해 테스트 한 결과 위와 같은 조합에 대해 학습 정확도가 가장 높게 나타났다. 결과적으로 학습 데이터는 799개의 글로 구성되었으며, 이중 282개의 감성적 지지 글로 517는 정보적 지지 글로 나타났다.

글의 텍스트 변수를 추출한 결과 총 2,235개의 변수가 추출되었다. 본 연구에서 사용된 변수는 다음과 같은 형태를 지닌다.

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

<표 2> 텍스트 특성으로 구성된 변수

글 번호	구조적 변수				의미적 변수				어휘적 변수		
	?	!	글의 길이	...	web site	new_drug	sad	...	명사 + 명사	동사 + 부사	“ ”
#1	4	5	2		1	2	0		5	2	
#2	1	0	3		0	0	3		1	3	
...											

특성 선택 후에는 변수가 440개로 줄어들었다. 본 연구에서는 계산의 편의성을 위해 가장 간단한 특성 선택 방법을 선택하였다. 발생 빈도가 낮은 특성을 학습에서 제외하는 특성 선택 방법을 이용해 학습에 사용된 변수 수를 감소시켰다. 학습 모델은 제 3장에서 언급한 SVM, 의사결정나무, 나이브-베이지언을 선택하였다. 본 연구에서 사용한 알고리즘은 WEKA⁴⁾ 소프트웨어에서 구현된 모듈을 사용하였다.

<표 3> 학습 모델의 정확도 비교

Feature Set	SVM	나이브-베이지언	의사결정 나무
모든 특성(의미적 +구조적+어휘적)	학습안됨	70.09%	65.46%
FS+모든 특성(의미적 +구조적+어휘적)	76.22%	70.59%	66.58%
FS+의미적	74.84%	73.46%	64.70%
FS+의미적+구조적	76.22%	73.22%	63.70%
FS+의미적+어휘적	74.47%	70.21%	66.08%
FS+구조적+어휘적	67.83%	68.59%	66.02%

주) FS : feature selection(특성 선택).

<표 3>는 학습 모델의 정확도를 비교한 결과이다. 세 개의 학습 모델 중에 SVM의 성능이 76.22%로 가장 높게 나타났다. 다른 텍스트 마이닝 연구에서 성능이 입증된 SVM이 본 연구에서도 가장 좋은 성능을 나타냈다. 확률 기반의 모델인 베이지언 모델은 SVM 모델보다 약간 성능이 떨어지

는 것으로 나타났으며, 의사결정나무의 정확도 값은 최대 66.58%로, 앞의 두 모델에 비해 성능이 떨어지는 것으로 나타났다. 이는 의사결정나무 모델은 선택된 변수 중 일부만을 이용해 분류 모델을 생성하기 때문에 모든 변수를 이용하는 앞선 두 모델에 비해 성능이 떨어지는 것으로 보인다.

각 그룹별 특성의 중요도를 파악하기 위해 가장 기본이 되는 변수 집합인 의미적 특성을 기초로 하고, 다른 특성을 더했을 때 정확도가 얼마나 상승하는지를 살펴보았다. 글의 감성적 지지 성격과 정보적 지지 성격을 대부분 글에 사용된 단어 즉, n-gram에 의해 분류가 될 것이기 때문에 기본 변수 집합으로 의미적 특성을 사용하였다. 가장 좋은 성능을 보인 SVM을 이용한 학습에서 의미적 특성만으로는 74.84%의 정확도가 나왔고, 여기에 구조적 특성을 더한 경우 정확도가 76.22%로 상승, 어휘적 특성을 추가한 경우 정확도가 74.47%로 오히려 감소하였다. SVM과 의사결정나무에서는 의미적, 구조적, 어휘적 특성을 모두 사용하여 특성 추출 후 학습시킬 경우 가장 좋은 성능을 보였다. 하지만 나이브-베이지언의 경우에는 모든 특성을 사용하는 것보다는 의미적 특성만을 사용한 경우에 더 좋은 성능을 나타내는 것으로 나타났다. 이의 결과에서 의료 웹포럼 상의 글의 특성을 분류할 때는 의미적 특성과 구조적 특성만으로도 충분하다는 것을 알 수 있다.

특성추출의 효과를 비교하기 위해 모든 특성을 사용한 경우를 기준으로 특성 선택을 한 경우와 하지 않은 경우를 비교해 보면, 모든 학습 모델에서 특성 추출 과정을 거친 경우에 학습 정확도가 상승한 것을 볼 수 있다. 그리고 실험 결과에서 보듯이 글의 구조적 특성과 어휘적 특성으로 학습한 결과는 의미적 특성만으로 학습한 결과보다 모든 분류기에서 그 성능이 떨어지는 것을 볼 수 있다. 이는 글의 특성 분류에서 가장 중요한 변수는 의미적 특성임을 의미한다.

가장 좋은 성능을 보였던 SVM의 결과에서 글의 특성을 구분하는데 중요하게 사용되는 특성을

4) WEKA : <http://www.cs.waikato.ac.nz/~ml/weka>.

살펴보고자 글의 특성별로 상위 20개의 특성을 나열하였다. <표 4>은 감성적 지지 글을 분류하는데 중요하게 사용되는 상위 단어를 나타내는데 의문부호(Question_mark)를 제외하고는 모두 의미적 특성이 중요하게 나타났다.

<표 4> 정보적 지지 글을 분류하는 중요 변수

변수	중요도	특성 분류
website	0.6894	의미적, unigram
statin(약 이름)	0.4333	의미적, unigram
special	0.4548	의미적, unigram
QUESTION_MARK(?)	0.4723	구조적, 문장부호
progress	0.4361	의미적, unigram
physic	0.4694	의미적, unigram
patient	0.6246	의미적, unigram
new_drug	0.4618	의미적, bigram
memory	0.4482	의미적, unigram
link	0.6578	의미적, unigram
knew	0.5963	의미적, unigram
input	0.658	의미적, unigram
info	0.4537	의미적, unigram
haldol(약 이름)	0.4661	의미적, unigram
Grandpa	0.4432	의미적, unigram
drug	0.4227	의미적, unigram
diagnosis	0.4217	의미적, unigram
check	0.4073	의미적, unigram
Hello	0.4485	의미적, unigram
40	0.4575	구조적, 단어 수

<표 5>는 정보적 지지 글을 분류하는데 중요하게 사용되는 상위 단어를 나타내며 대부분이 의미적 특성이 중요한 특성으로 나타났다.

5. 결론 및 향후 연구방안

최근 건강에 대한 관심이 높아짐에 따라 의료정보에 대한 관심 역시 높아졌다. 따라서 웹포럼은 전문가들이 의료정보를 제공하고, 일반인들이 그 정보를 조회하고, 상호간의 질병에 대한 경험 공유하

<표 5> 감성적 지지 글을 분류하는 중요 변수

변수	중요도	특성 분류
my	0.8723	의미적, unigram
great	0.7398	의미적, unigram
deal	0.6787	의미적, unigram
betty	0.5796	의미적, unigram
TILDE(~)	0.5715	구조적, 문장기호
jerk	0.5185	의미적, unigram
awhile	0.5086	의미적, unigram
compute	0.4956	의미적, unigram
pray	0.4827	의미적, unigram
time	0.4727	의미적, unigram
hear	0.4704	의미적, unigram
rene (프랑스어로 reborn)	0.4656	의미적, unigram
cry	0.4514	의미적, unigram
feel	0.4508	의미적, unigram
today	0.4438	의미적, unigram
okay	0.4342	의미적, unigram
test	0.433	의미적, unigram
gizmo_gizmo (장치에 대한 은어)	0.4328	의미적, unigram
least	0.4314	의미적, unigram
respond	0.4303	의미적, unigram

는 역할을 수행함으로써 정보적 측면뿐만 아니라 감성적 측면으로 지원하는 채널 역할을 하고 있다.

본 연구에서는 방대한 양의 의료 웹포럼상의 글을 자동으로 수집하여 텍스트의 구조적, 의미적, 어휘적 특성을 도출하여 글을 정보적 지지 유형과 감성적 지지 유형으로 자동으로 분류하는 모델을 제시하였고, 모델의 정확도를 높이고자 하였다. 의사결정나무, 나이브-베이시안 모델에 비해 SVM과 의미적 특성과 구조적 특성을 이용할 때, 의료 웹포럼의 게시글을 감성적 지지 유형과 정보적 지지 유형으로 분류한 정확도가 76.22%로 가장 높게 나타났다. 본 연구를 통해 의료 웹포럼 글을 정보적 지지 유형과 감성적 지지 유형으로 분류할 수 있는 더 나은 모델을 도출했다는 점에서 의의를 찾을 수 있다.

본 연구의 모델을 이용하면 의료 웹포럼 정보들

을 정보적 지지, 감성적 지지 측면의 정보를 자동으로 분류할 수 있으며, 의료 웹포럼을 이용하는 환자, 환자 가족, 의료진들에게 맞춤형 정보적 지원과 감성적 지원을 제공할 수 있는 기반을 구축할 수 있을 것이다. 그리고 대량의 웹포럼 글을 자동으로 분류함에 따라 효율적인 웹 운영을 가능하게 하는데 기여할 것이다.

하지만 본 연구의 학습모델의 정확도는 76.22%로 더 높은 정확도를 가진 모델로 개선시킬 수 있는 여지가 있다. 새로운 텍스트 특성을 찾아서 학습모델을 만들거나 새로운 알고리즘을 개발한다면 학습모델은 더 개선될 수 있을 것이다. 예를 들어 의료 웹포럼 사용자들의 네트워크 분석을 통해 지수화하여 텍스트 특성에 추가하여 사용할 수도 있을 것이다. 해당 연구를 기초로 웹포럼 사용자들을 정보적 지지를 우선시하는 사람과 감성적 지지를 우선시하는 사용자로 분류하여 감성 커뮤니티와 지식 커뮤니티를 비교 분석할 수도 있다. 그리고 시계열적으로 의료 웹포럼 내의 감성 커뮤니티와 지식 커뮤니티의 분석을 통해 커뮤니티의 최초로 형성되고 성장하고 성숙하는 과정에서 강한 네트워크를 형성하는 원인들을 파악하여 성공적인 의료 웹포럼의 성공 요인들을 도출하는 것도 의미 있는 향후 연구가 될 것이다.

참 고 문 헌

- [1] 권예지, 나은영 “정보원의 공신력과 이성적, 감성적 메시지 유형이 사회적 지지, 태도, 건강행동의도에 미치는 영향”, 『韓國言論學報』, 제55권, 제5호(2011).
- [2] 김승환, “의료 IT 융합기술 동향”, 『전자통신 동향분석』, 제26권, 제6호(2011).
- [3] 우지영, “텍스트 마이닝을 이용한 웹포럼 불량 글 탐지 모델”, 『한국지식정보기술학술회논문지』, 제7권, 제1호(2012), pp.159-166.
- [4] 임화영, 김형수, 최영환, 장성훈, 이건설, 정최경희, 오원기, 최재욱, 정춘화, “지역사회에서 형성된 사회적 지지와 근로자의 사회심리적 스트레스간의 관련성”, 『대한산업의학회지』, 제18권, 제4호(2006).
- [5] 장미경, “청소년의 사회망 지향 : 사회적 지지 과정과 심리적 적응을 중심으로”, 『청소년상담 연구』, 제12권, 제1(2004), pp.3-14.
- [6] 채승병, 예지은, 진현, 박성민, 김진성, 김병완, 『가치창출의 새로운 원천, 집단지성』, 삼성경제연구소, 2010.
- [7] 추연화 “항암화학요법에 관한 개별교육이 암 환자의 삶의 질에 미치는 효과”, 『종양간호학회지』, 제4권, 제2호(2004), pp.103-109.
- [8] Buchanan, H. and N. S. Coulson, “Accessing dental anxiety online support groups : An exploratory qualitative study of motives and experiences”, *Patient Education and Counseling*, Vol.66, No.3(2007), pp.263-269.
- [9] Claire, F. S., “Gendered Cybersupport : A Thematic Analysis of Two Online Cancer”, *Journal of Health Psychology*, Vol.8, No.1 (2003), pp.83-103.
- [10] Cobb, S., “Social Support as Moderator of Life Stress,” *Psychosomatic Medicine*, Vol. 38, No.5(1976), pp.300-314.
- [11] Cohen, J., “Weighted kappa : nominal scale agreement with provision for scaled disagreement or partial credit”, *Psychological Bulletin*, Vol.70(1968), pp.213-220.
- [12] Cohen, S., B. Gottlieb, and L. Underwood, “Social relationships and health”, *American Psychologist*, (2004), pp.676-684.
- [13] Cohen, S., L. Underwood, and B. Gottlieb (Eds.), *Measuring and intervening in social support*, NY : Oxford press, 2000.
- [14] Eastin, M. S. and R. LaRose, “Alt.support : Modeling social support online”, *Computers in Human Behavior*, Vol.21, No.6(2005), pp. 977-992.

- [15] Elkin, N., *How America Searches : Health and Wellness*, iCrossing, 2008.
- [16] Glance, N., M. Hurst, K. Nigam, and M. Siegler, *Deriving Marketing Intelligence from Online Discussion*, KDD, 2005.
- [17] Gottlieb, B. H. and A. E. Bergen, "Social support concepts and measures", *Journal of Psychosomatic Research*, Vol.69(2010), pp. 511-520.
- [18] Gruhl, D., R. Guha, R. Kumar, and J. Novak, *The predictive power of online chatter*, KDD, (2005), pp.78-87.
- [19] House, J. S., K. R. Landis, and D. Umberson, "Social Relationships and Health", *Science New Series*, Vol.241, No.4865(1988), pp.540-545.
- [20] House, J. S., D. Umberson, and K. R. Landis, "Structures and Processes of Social Support", *Annual Review of Sociology*, Vol. 14(1998), pp.292-318.
- [21] Hwang, K. O., A. J. Ottenbacher, J. F. Lucke, J. M. Etchegaray, A. L. Graham, E. J. Thomas, and E. V. Bernstam, "Measuring Social Support for Weight Loss in an Internet Weight Loss Community", *Journal of Health Communication*, Vol.16, No.2(2011), pp. 198-211.
- [22] Joshi, K. D., L. Chi, A. Datta, and S. Han, "Changing the competitive landscape : Continuous innovation through IT-enabled knowledge capabilities", *Information Systems Research*, Vol.21(2010), pp.472-495.
- [23] Kasl, S. V. and S. Cobb, "Health Behaviour, Illness Behaviour and Sick Role Behaviour", *Archives of Environmental Health*, Vol.12 (1966), pp.246-266.
- [24] Koppel, M., S. Argamon, and A. Shimoni, "Automatically categorizing written texts by author gender", *Literary Linguistic Compute*, Vol.17, No.4(2002), pp.401-412.
- [25] Landis, J. R. and G. G. Koch, "The measurement of observer agreement for categorical data", *Biometrics*, Vol.33, No.1(1977), pp.159-217.
- [26] Lewis, D., "Naive(Bayes) at forty : The independence assumption in information retrieval", *Machine Learning*, (1998), pp.4-15.
- [27] Meiri, R. and J. Zahavi, "Using simulated annealing to optimize the feature selection problem in marketing applications", *European Journal of Operational Research*, Vol. 171, No.3(2006), pp.842-858.
- [28] Mo, P. K. H. and N. S. Coulson, "Exploring the communication of social support within virtual communities : A content analysis of messages posted to an online HIV/AIDS support group", *CyberPsychology and Behavior*, Vol.11(2008), pp.371-374.
- [29] Mookadam, F. and H. M. Arthur, "Social support and its relationship to morbidity and mortality after acute myocardial infarction : Systematic overview", *Archives of Internal Medicine*, Vol.164, No.14(2004), pp. 1514-1518.
- [30] Peterson, J. L., "You Have to be Positive," *Social Support Processes of an Online Communication Studies*, Vol.60, No.5(2009), pp.526-554.
- [31] Platt, J. C., *Sequential Minimal Optimization : A Fast Algorithm for Training Support Vector Machines*, Microsoft Research Technical Report MSR-TR-98-14, 1998.
- [32] Quinlan, J. R., "Induction of decision trees", *In Machine Learning*, (1986), pp.81-106.
- [33] Ron, K., "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Mo-

- del Selection”, *Appears in the International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.
- [34] Tamar, G., “Online Participation : A Content Analysis of Differences in Utilization”, *Health Communication*, Vol.23(2008), pp.1-12.
- [35] Vapnik, V. N., *The nature of statistical learning theory*, New York : Springer-Verlag, 1999.
- [36] Weiss, R. S., *The provisions of social relationships*, In Z. Rubin (Ed.), *Doing unto others*, NJ : Prentice-Hall, (1974), pp.17-26.
- [37] Wellman, B. and M. Gulia, *Virtual communities as communities : Net surfers don't ride alone*, *Communities in cyberspace*, Berkeley, CA : Routledge, (1999), pp.167-194.
- [38] White, M. H. and S. M. Dorman, “Online support for caregivers : analysis of an Internet Alzheimer mailgroup”, *Computers in Nursing*, Vol.18(2000), pp.168-176.
- [39] Yang, B., J. Lim, W. Oh, A. Animesh, A. Pinonneault, “Using Real Options to Investigate the Market Value of Virtual World Businesses”, *Information Systems Research, Articles in Advance*, 2012.
- [40] Zhang, Y., Y. Dang, and H. Chen, “Gender Classification for Web Forums”, *IEEE Transactions on Systems, Man and Cybernetics-Part A : Systems and Humans*, Vol. 41, No.4(2011).

◆ 저 자 소 개 ◆



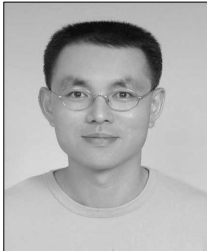
우 지 영 (jywoo@korea.ac.kr)

고려대학교 정보보호 대학원에 재직중이다. 한국과학기술원(KAIST) 산업공학과 학사, 석사, 박사 학위를 취득하였다. 삼성화재 과장으로 재직하였으며, 미국 아리조나 대학 AI 연구실에서 박사후 연구과정을 수행하였다. Expert system with application에 서비스 품질관리 시스템 관련 논문을 다수 게재하였으며, Applied Mathematics and Information Sciences, Computers and Mathematics with Applications 등의 저널에 온라인게임 보안 논문을 게재하였다. 또한, IEEE ISI, ACM Siggraph Asis 등의 국제 학회에 참여하여 정보 확산에 관련된 논문을 발표하였다. 주요 관심분야는 소셜 미디어 분석 및 정보확산 모델링, 게임 보안 등이다.



이민정 (mjleekorea@gmail.com)

KAIST 재료공학과에서 학사, 석사 학위를 받고, 산업공학과에서 박사학위를 받았다. 삼성 SDS, 한국산업기술진흥원, 엔씨소프트에서 실무경력이 있으며, 현재 세종사이버대학교 경영학과, 세종사이버대학원 경영 MBA 조교수로 재직 중이다. Expert system with application, Total Quality Management, Information 등의 저널에 서비스 품질 관련 논문을 다수 게재하였다. 관심분야는 MIS, CRM, 품질, 기술경영 등이다.



Yungchang Ku (ycku1230@gmail.com)

현재 타이완의 Yuan Ze 대학의 정보경영(information management) 학과의 박사과정으로 재학중이며, Central Police 대학에서 기술 전문가로 재직중이다. 타이완 National Central University의 정보경영학과에서 석사학위를 취득하였다. 박사과정중에 미국 아리조나 대학의 AI 연구실에서 연구원으로 근무 하였다. 주요 관심분야는 의료정보 시스템 및 소셜미디어 분석 등이다.