

An Analysis of IT Proposal Evaluation Results using Big Data-based Opinion Mining

Hong Sam Kim · Chong Su Kim[†]

Department of Industrial & Management Engineering, Hannam University

빅데이터 분석 기반의 오피니언 마이닝을 이용한 정보화 사업 평가 분석

김홍삼 · 김종수[†]

한남대학교 산업경영공학과

Current evaluation practices for IT projects suffer from several problems, which include the difficulty of self-explanation for the evaluation results and the improperly scaled scoring system. This study aims to develop a methodology of opinion mining to extract key factors for the causal relationship analysis and to assess the feasibility of quantifying evaluation scores from text comments using opinion mining based on big data analysis. The research has been performed on the domain of publicly procured IT proposal evaluations, which are managed by the National Procurement Service. Around 10,000 sets of comments and evaluation scores have been gathered, most of which are in the form of digital data but some in paper documents. Thus, more refined form of text has been prepared using various tools. From them, keywords for factors and polarity indicators have been extracted, and experts on this domain have selected some of them as the key factors and indicators. Also, those keywords have been grouped into into dimensions. Causal relationship between keyword or dimension factors and evaluation scores were analyzed based on the two research models-a keyword-based model and a dimension-based model, using the correlation analysis and the regression analysis. The results show that keyword factors such as planning, strategy, technology and PM mostly affects the evaluation result and that the keywords are more appropriate forms of factors for causal relationship analysis than the dimensions. Also, it can be asserted from the analysis that evaluation scores can be composed or calculated from the unstructured text comments using opinion mining, when a comprehensive dictionary of polarity for Korean language can be provided. This study may contribute to the area of big data-based evaluation methodology and opinion mining for IT proposal evaluation, leading to a more reliable and effective IT proposal evaluation method.

Keywords : Opinion Mining, Big Data, Sentiment Analysis, Evaluation

1. 서 론

정부가 발주하는 정보화 사업의 시장 규모는 연간 3~4조

원 정도이며, 주로 정보시스템 구축 및 운영과 유지보수, BPR/ISP, 서버 및 스토리지 도입 등의 사업이 주류를 이루고 있다. 이들 정보화 사업 프로젝트에서의 사업자 선정 방식은 사업수행 기업이 제출한 제안서를 분야별 전문가들이 평가함으로써 이루어진다. 즉, 제안 참가 기업들이 제출한 제안서를 평가자들이 서면 검토 및 평가하고,

Received 28 November 2017; Finally Revised 8 January 2018;
Accepted 9 January 2018

[†] Corresponding Author : ckim@hnu.kr

제안 업체의 발표 및 질의응답 과정을 거친 후, 제안 요청서에 제시된 평가 기준에 따라 평가 점수를 부여한다. 그리고 평가의 근거 등을 자유 서술 방식으로 기술한 검토 의견을 덧붙인다. 정부 구매 기관은 이와 같이 부여된 점수를 바탕으로 정보화 사업을 수행할 사업자를 선정하게 된다. 즉, 평가 결과를 평가자 개인이 임의로 정한 스케일에 따라 단일의 최종 평가 점수로 표현되며, 이 점수만이 최종 선정에 활용된다. 코멘트 형식의 문장으로 남기는 평가 의견 혹은 검토의견은 주로 점수 부여의 정당성을 뒷받침하는 목적으로 작성되며 사업자 선정에는 직접적으로 영향을 미치지 못한다. 따라서 평가 결과만으로는 평가자의 판단 기준이 무엇인지, 어떠한 기준으로 점수를 부여 하였으며 어떤 요인이 평가 결과에 실질적 영향을 주었는지를 알기 어렵다. 이로 인하여 평가 주관기관이 참가기업의 이의제기에 대하여 납득할 만한 사유를 설명할 수 없는 경우가 많고, 이는 선정의 공정성에 대한 불신으로 이어질 수 있다. 또한, 속성별, 차원별로 점수가 부여되지 않음으로 인하여 제안 참여자 및 정부의 입장에서는 평가 결과에 영향을 미치는 요인을 파악하기 어렵다. 그리고 평가 점수의 스케일이 평가자들마다 다르다는 문제점이 존재한다.

본 연구에서는 이러한 문제점들을 해결하기 위하여 빅데이터 분석 기법의 하나인 오피니언 마이닝을 이용한다. 정성적으로 서술된 평가 의견에서 평가 결과에 영향을 미치는지 요인들을 파악한다. 그리고 비정형 텍스트 형태인 검토 의견을 토대로 정량적 평가 점수를 도출할 수 있는 가능성을 알아본다. 이를 통하여 평가방법의 개선 및 정보화 사업 제안의 질적 향상에 대한 방법을 제시하고자 한다. 통상적으로 인과관계 분석을 수행하기 위해서는 인터뷰나 설문 조사와 같은 방식으로 평가자의 의견을 조사한다[15]. 그러나 공공 정보화 사업 평가에서는 보안상의 이유로 평가자의 명단이 공개되지 않아 평가자들에 대한 접촉 자체가 쉽지 않으며, 평가의 성격상 평가자들의 객관적이고 솔직한 답변을 얻기 어렵다. 따라서 본 연구에서는 비정형적 텍스트로 표현된 검토의견으로부터 오피니언 마이닝을 이용하여 평가자의 의견을 추출한다.

본 연구에서 추구하는 바와 같이 평가자가 자유롭게 기술하는 의견(키보드로 직접 입력한 문장 혹은 토의나 질의응답 과정에서 음성으로 인식한 문장)으로부터 평가자의 정량적 판단을 도출해 낼 수 있다면, 보다 객관적이고 상세한 평가가 가능하다. 우선 얻을 수 있는 정보가 양적으로도 훨씬 더 풍부할 뿐 아니라 항목별, 속성별로 구분된 의견을 얻을 수 있다. 물론 프로젝트 수주와 직접적으로 연결되는 평가의 특성상 오피니언 마이닝으로 얻은 극성 점수만으로는 선정 여부를 결정하기에 불충분한

경우도 많을 것이다. 그렇지만 이러한 경우에도 가중 평 균을 내는 식으로 병행하여 사용하거나, 또는 직접 부여한 점수에 대한 검토 및 검증의 수단으로 사용될 수 있다. 그리고 오피니언 추출의 기술에 대한 신뢰도가 향상되면 자유로운 분위기에서 표현한 의견으로부터 이끌어 낸 결과가 더 나올 수도 있다. 또한, 다양한 요인별, 속성별로 평가의견 추출이 이루어지므로 평가 결과에 대한 설명력을 갖출 수 있고, 평가자 별로 편차가 존재하는 점수 스케일의 문제도 해결할 수 있다.

이러한 연구 목적을 달성하기 위하여 우선 기존 연구의 성과와 한계점을 파악하였다. 본 연구의 대상 영역인 정보화 사업 제안 평가는 영화평이나 상품 후기와 같이 주관적 감정이 아닌 공식적인 평가가 강조되는 영역이므로 검토 의견이 보다 절제된 표현이 주를 이룬다. 그러므로 감정 추출 및 극성 판별이라는 감정분석에 있어서 새로운 접근방법이 필요하다. 이러한 특성들을 감안한 연구 모형을 제시하였다. 특정 발주기관의 온라인 평가시스템 및 시기 평가철에 등록되어 있는 검토 의견과 결과 점수를 수집하여, 오피니언 마이닝 기법을 활용하여 비정형 텍스트의 검토의견으로부터 평가에 영향을 미치는 주요 요인 및 감정 키워드를 추출하고 이를 다시 3개의 차원으로 분류하였다. 이를 통해 평가 결과에 미치는 요인의 영향력을 파악하고 검토의견에서 수치화된 평가 점수를 도출할 수 있는 가능성을 검토해보았다. 그리고 이를 바탕으로 정보화 사업 제안의 평가 및 사업 수행 과정을 향상시킬 수 있는 개선책에 대한 시사점을 도출하였다.

2. 기존 연구

2.1 오피니언 마이닝과 극성 판별

인터넷 등에 비정형적 텍스트의 형태로 존재하는 사용자의 의견을 분석, 감정 정보를 추출, 활용하여 사용자들의 인식을 파악하고, 서비스의 만족도를 높이려는 노력이 다양한 분야에서 이루어지고 있다. 이와 같은 연구들은 주로 트위터, 페이스북, 블로그 등 SNS에 대량으로 올려진 의견을 대상으로 하므로 빅데이터 분석의 영역, 구체적으로는 빅데이터 감정분석(Sentiment Analysis)의 영역에 속한다. 감정분석은 텍스트를 통하여 긍정, 부정, 혹은 중립이라는 감정을 정의해 나가는 과정이다[19]. 분석의 핵심은 텍스트에 나타난 감정을 판별하는 방법인데, 수작업적 라벨 부여, 자연언어 처리, 사전 기반 분석 및 기계학습 등이 사용되고 있다. 가장 많이 연구되고 있는 사전 기반 분석은 다시 말뭉치를 이용하는 방법과 사

전을 이용하는 방법으로 구분할 수도 있다[1]. 어느 경우이건 분석을 위해서는 비정형 텍스트를 구성하는 기본 단위인 형태소나 단어, 문장 혹은 문서 전체에 대한 감정 정보가 필요하다[4].

빅데이터 분석과 관련된 이슈는 Cheon and Baek[3]에 잘 정리되어 있다. 빅데이터 분석 기법 중 텍스트 분석 기술을 활용한 분야 중 대표적인 것으로 오피니언 마이닝을 들 수 있다. 오피니언 마이닝은 대상 및 속성에 대한 사람들의 의견, 감정, 평가, 태도, 감정 등을 분석하는 일련의 과정을 의미한다[17]. 즉, 사용자가 다양한 매체를 통해 표출한 의견을 추출하고 분류하여 이를 이해하고 정보자산으로 만들어 가는 과정이다[5]. 의견은 우리의 행위에 영향을 미치는 핵심적 요인이므로[18], 효율적인 의사결정을 위해 타인의 의견을 참고하는 것이 필수적이다. 예를 들어, 기업은 제품 및 서비스 개발의 방향을 결정하기 위해 소비자들의 욕구 및 불만에 대한 의견을 알고 싶어 한다. 이를 위해 과거에는 설문이나 인터뷰 등의 형태로 조사를 수행하였다. 하지만 최근에는 빅데이터 오피니언 마이닝을 활용하여 트위터 혹은 페이스북 등의 소셜 미디어의 데이터를 획득하는 등 소비자들의 의견을 수집하는 방법이 다양화되었다[14]. 가장 기본적인 수준의 오피니언 마이닝은 문서 단위로 이루어지며, 이 차원의 연구는 하나의 문서는 하나의 감성을 표현한다고 간주한다[27]. 보다 세부적 단위의 연구는 문장을 기본 분석 단위로 간주 하며, 주관성 구분과 구와 절 단위의 분석을 주요 이슈로 다루고 있다[10].

오피니언 마이닝을 위해서는 분석의 최소 단위인 어휘의 감정 극성(Sentiment Polarity)이 정해져야 한다. 그러나 어휘의 극성은 사용 문맥과 영역에 의존하며, 의문문에서와 같이 감정 어휘 자체가 중립적인 극성을 나타내는 경우도 있으므로 정하기 쉽지 않다[16]. 영어의 경우에는 텍스트의 극성 분석에 대한 연구가 활발하게 진행되고 있다. Turney and Littman[26]는 긍정, 부정어 각각 7개를 기본 어휘로 삼고 이 어휘들과 함께 사용되는 어휘들의 빈도를 계산하여 확장하는 방식을 제안하였다. 이러한 연구를 바탕으로 극성 어휘 사전인 WordNet을 구축하고, 이를 활용하여 어휘의 극성을 결정하는 방안이 제시되었다[24]. Kamps et al.[13]은 긍정과 부정의 기초 어휘를 설정하고 동의어 관계에서의 최단 길이를 바탕으로 긍정과 부정을 판별하였다.

텍스트에서 감정을 추출해 내기 위해 가장 필요한 것은 감정 사전의 구축이다. 백터공간 모델을 이용하여 문장 내 감정 용어의 빈도를 측정하여 이용하는 연구가 주류를 이루고 있으며, 이외에도 비교급 문장, 조건문, 풍자적 표현, 문장 간 관계에 대한 분석, 용어의 시맨틱을 반영한 분석 등의 접근 방법이 연구되고 있다. 그 중에서

도 가장 대표적인 것은 감정어휘사전인 SentiWordNet의 구축과 관련된 연구이다[26]. 이 방법에서는 14개의 기초적인 긍정, 부정 어휘와 WordNet을 이용하여 감정 속성이 부여된 어휘의 숫자를 늘려간다. 그러나 이것은 영어로 구성되어 있고 영어 단어와 한글 단어의 의미 해석이 문맥에 따라 다르기 때문에 이를 직접 번역하여 한글의 극성에 적용하기에는 한계가 있다[4]. 국내의 오피니언 마이닝 연구는 기존의 연구 성과인 자연어 처리에 대한 이해를 바탕으로 마이닝의 정확도를 향상시키고자 하는 방향으로 활발하게 수행되고 있지만, 극성 분석의 가장 근간이 되는 감정어휘사전의 구축에 대한 연구는 미미한 실정이다[14].

2.2 적용 분야

의견이나 만족의 정도를 나타내는 감정뿐만 아니라 대상이나 속성에 대한 평가(evaluation)도 감정분석 및 오피니언 마이닝의 대상이 된다. 본 연구의 대상 영역인 정보화 사업 제안 평가도 감정이 아닌 평가를 대상으로 한다. 개인적이고 감성적인 의견이 주를 이루는 통상적인 감정분석의 경우와는 달리, 평가는 대부분의 경우 보다 공식적이고 절제된 표현으로 진행된다. 이는 감정 추출 및 극성 판별을 수행하는 오피니언 마이닝에 새로운 접근방법을 요구한다. 또한, 공정성 시비 등에 휘말릴 수 있고 기업의 영업 기밀과 관련이 있다는 점 때문에 평가 결과의 외부 공개가 쉽지 않아 자료 수집이 어렵다. 그 결과 평가에 대한 연구 사례가 많지 않다. 이 분야의 연구로는 병원에서 환자의 상태를 평가하고 기록한 임상소견에서 환자의 상태에 대한 평가 결과를 이끌어내는 연구가 있다[8]. 수업 및 교사에 대한 학생의 평가를 감정분석 기법으로 분석한 연구에서는 자유롭게 평가 의견을 기술하는 비정형 텍스트 분석이 리커트 척도의 설문조사에 비해 훨씬 다채롭고 유용한 결론을 도출할 수 있다는 결과를 얻었다[23].

평가보다는 주관적인 감정이나 의견, 느낌이 중시되는 분야에서는 오피니언 마이닝을 응용한 다양한 연구와 적용 사례들이 발표되었다. 제품과 서비스에 대한 사용자 리뷰는 구매결정에 강력한 영향을 미친다. 따라서 기업은 이를 분석하여 사용자의 만족도를 측정하기 위한 연구를 진행하고 있다 오피니언 마이닝을 이용하여 호텔 서비스의 품질을 측정한 연구가 대표적인 예이다[6]. 또한 단문 영화평 분석을 통해 영화 평점을 예측하는 연구가 진행되었다[29]. 특히, 상품평에 대한 구조를 분류, 정리하는 연구도 있는데, 한국어 상품평을 분석할 때 의견 단어가 포함되어 있을 확률이 높은 문장 구조를 발견하여 정리하였다[28]. 자료의 수집 용이성도 연구의 발전

방향에 큰 영향을 미쳤다. 지금까지는 트위터와 같이 자료 접근이 용이한 SNS가 연구 및 상업적 목적의 감정분석에 많이 이용되었는데, NFL 배팅과 트위터에서의 팬들의 의견 사이의 관계를 분석하거나[9], 트위터에 나타난 감정이 특정 정책에 대한 여론 탐색에 이용되거나[22], 선거 관련 여론조사의 보조 도구로 활용하는 가능성을 연구한 것이 그 예이다[25]. 트위터에 나타난 영화 평 분석을 통해 매출액을 추정하는 연구는 매우 정형적인 SNS 이용 분야라고 할 수 있다[2]. 다른 형태의 온라인 미디어도 데이터 수집처로 활용되고 있는데, 전자메일을 이용하여 특정 감정 행위에 대한 남녀 간의 차이를 분석한 연구도 있으며[20], 소설과 동화에 나타난 감정을 추적하여 분석하는 연구도 수행되었다[21]. 그리고 마이크로 블로그 텍스트에서 전문적인 투자자들의 의견을 수집하여 주식 투자 성향을 예측하기도 하였다[7]. 또한, 빅데이터로부터 군수품의 품질 정보를 추출하여 제품 개선에 활용하거나[12], 자동차 브랜드에 대한 고객의 인식을 파악하기 위해 유사한 브랜드간의 군집을 도출하고 각 군집을 특징을 기술하는 브랜드 클러스터링에 오피니언 마이닝을 활용하였다[11].

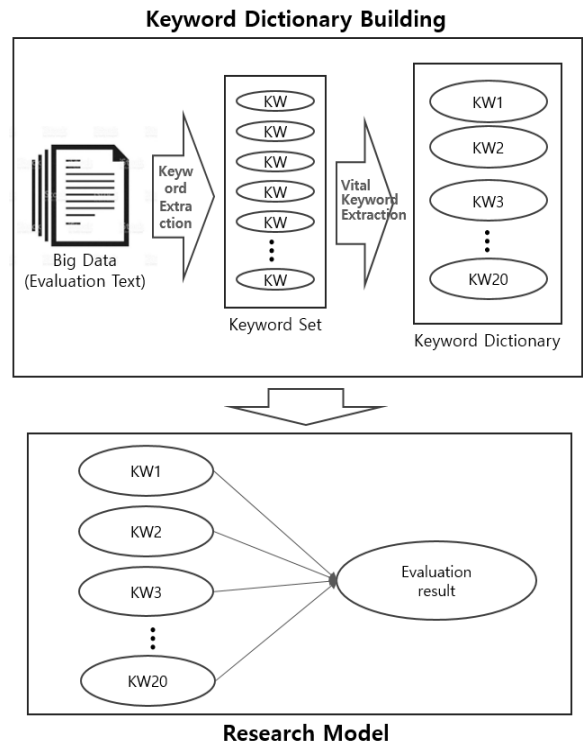
3. 연구 모형 수립 및 데이터 수집

3.1 연구 모형의 제안

앞에서 살펴 본 바와 같이 오피니언 마이닝과 같은 빅데이터 분석 기술을 이용하여 비정형적 텍스트에서 감정을 추출, 평가를 비롯한 다양한 분야에 활용하려고 하는 연구가 활발하게 수행되고 있다. 그러나 기존의 연구에는 아래와 같은 여러 한계점이 있다. 우선, 한글 기반 감정사전 구축에 대한 연구가 미흡하다. 감정 어휘 사전은 텍스트에서 감정을 추출해 내기 위한 가장 핵심적인 요소이다. 그러나 영어와 같은 서구 언어의 경우와는 달리 한국어의 경우에는 사전 구축을 위한 관련 연구가 미흡한 실정이다. 둘째, 지금까지의 연구는 데이터 수집의 편의상 주로 호텔 숙박이나 영화 평가와 같은 분야로 한정되었다. 그러나 자유롭게 작성하는 코멘트나 리뷰와는 달리, 평가 의견은 평가의 특성상 감정 표현이 간접적이고 우회적으로 이루어진다. 이러한 평가 의견을 이용한 오피니언 마이닝 연구는 아직 본격적으로 수행되고 있지 않다.

본 연구에서는 텍스트로 표현된 평가자의 검토의견으로부터 빅데이터 분석을 통하여 평가 영향을 미치는 요인과 그 요인에 대한 의견(극성)을 추출하고 수치화하기 위하여 오피니언 마이닝 기법을 적용하는 연구를 진행하였다. 전술한 바와 같이 평가 의견의 특징인 간접적이고

우회적인 감정 표현의 자제는 기존의 연구와는 다른 접근 방법을 요구한다. 따라서 이러한 특성을 반영하여 평가 결과에 영향을 미치는 요인을 파악하고, 문장의 극성에 대한 키워드를 추출, 키워드 사전을 구축하여 평가 의견의 수치화 가능성을 모색하는 두 가지 연구 모형을 제시하였다. 키워드 기반 연구 모형(Keyword-based Research Model)은 어휘 혹은 개념어 수준에서 평가자들이 가장 중요하게 판단하는 요인들이 무엇인지를 파악하기 위한 것이고, 차원 기반 연구 모형(Dimension-based Research Model)은 키워드들을 평가 목적에 맞게 결정요인(차원)으로 분류하여 인과관계를 분석하는 모형이다.

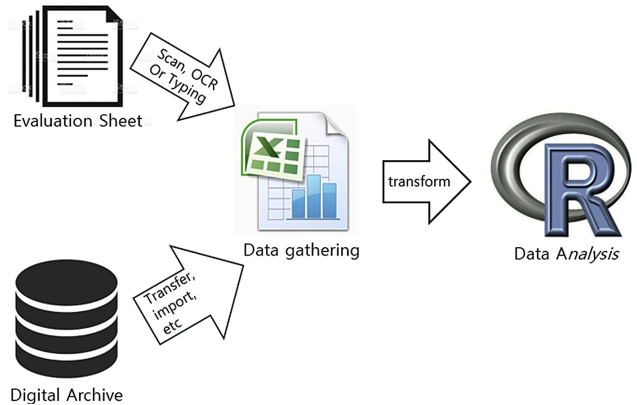


<Figure 1> Keyword-based Research Model

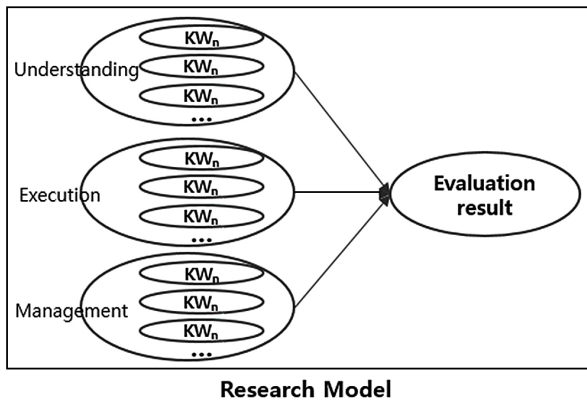
키워드 기반 연구 모형에서의 데이터 분석 과정은 <Figure 1>과 같다. 우선 검토의견 텍스트에서 형태소 추출을 통해 키워드 집합을 만든다. 여기서 빈도수, 명사형 여부 분석을 통해 1차 키워드를 추출하고, 평가 경험이 풍부한 전문가 집단의 검토를 거쳐 평가 결과에 가장 영향을 미치는 20개의 키워드를 키워드 사전으로 선정한다. 인과관계 분석 및 계량화는 20개의 각각의 키워드에 대해 이루어진다.

차원 기반 연구 모형은 키워드 모형을 확장한 것으로, 키워드 기반 요인들을 평가 목적에 맞게 전문가 판단을 거쳐 이해력, 수행력 및 관리력의 3개 차원으로 분류하고 이들과 평가 결과의 인과관계를 분석한다(<Figure 2> 참조).

주요 요인 키워드에 대한 차원 분류를 위하여 조달청의 전자조달시스템인 나라장터 공고정보 검색을 통해 약 100여건의 정보화 사업 제안요청서를 수집하였다. 그 속에 포함된 평가표의 항목들을 수집한 후, 전문가 검토를 통해 이해력, 수행력, 관리력의 3가지 큰 범주로 나누어 본 연구 모형의 차원으로 사용하였다. 이해력은 사업을 수행하는 기업이 해당 정보화 사업의 수행환경에 대해 얼마나 이해하고 있는지를 뜻한다. 수행력은 사업을 수행함에 있어 기술과 인력을 투입하며 조직을 이끌고 정보화 사업을 진행해가는 수행능력 전반을 판단하는 기준이 되는 차원이다. 마지막으로 관리력에서는 각종 위험에 대처하고 전반적인 품질을 유지, 관리해 나가는 능력을 판단한다.



<Figure 3> Data Gathering Procedure



<Figure 2> Dimension-based Research Model

<Table 1> Analysis Data(Example)

No.	Comments	Eval.
1	The RFP has been well incorporated in the proposal, and they implement a comprehensive maintenance plan.)	5
2	Maybe get the functional requirements and provide proper interfacing scheme, but the tech. level of the PM gives some worrying points.)	4

3.2 데이터 수집

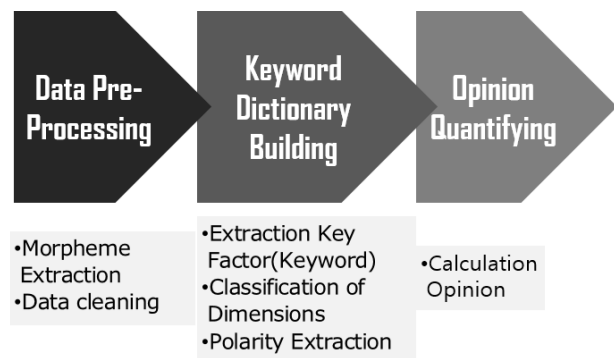
분석을 위한 데이터는 정부 정보화 사업과 진행된 200여 개의 제안 심사를 대상으로 수집하였다(<Figure 3> 참조). 사업 별로 다양한 유형의 평가 절차가 시행되는데, 통상적으로 평가자는 평가 시스템에 접속하여 평가 유의사항, 제안요청서, 제안서 등 평가와 관련된 자료를 검토한다. 이후 제안에 대한 발표가 진행되고 이러한 과정들이 모두 종료되면 온라인화면 혹은 종이로 된 평가표에 검토의견을 등록하고 점수를 부여한다. 점수는 평가 항목별로 1~5 점 사이의 범위 내에서 50~400자 내외의 검토의견을 입력하게 되어 있다. 온라인 방식을 이용하지 않고 수기로 문서를 작성하는 경우도 있다.

이렇게 입력된 평가 결과는 적절한 형태로 변환되어야 한다. 우선, 수기로 작성되어 보관 중인 문서는 문자 인식 등 다양한 방법으로 디지털 형태로 변환하였다. 시스템에 등록된 Archive 형태의 데이터는 SQL 등의 도구를 이용하여 스프레드시트 형태로 변환시켜 수집하였다. 이러한 과정을 거쳐 총 9,264개의 평가 결과(검토 의견, 평가 점수)를 수집하였다(<Table 1> 참조).

4. 분석

4.1 데이터 분석 방법 및 절차

수집된 데이터를 분석에 이용하기 위해서는 몇 단계의 데이터 정제 및 가공이 필요한데, <Figure 4>와 같이 단계를 거치게 된다. 데이터 사전처리 단계에서는 수집된 검토의견에서 형태소(키워드)를 추출하고, 추출된 키워드를 정제한다. 초기에 추출된 집합에는 의존명사, 숫자 등 분석에 불필요한 형태소들이 다수 포함되어 있다. 이들은 분석의 방해 요인이 되므로 사전에 제거해 주어야



<Figure 4> Data Analysis Procedure

한다. 두 번째 단계에서는 키워드 사전 구축을 위해 앞 단계에서 추출된 형태소를 대상으로 극성 키워드를 추출한다. 차원 모형에 사용할 주요 키워드에 대한 차원 분류도 이 단계에서 이루어진다.

본 연구에서는 데이터 정제, 가공 및 분석을 위해서 다양한 도구들을 사용하였다. 우선 통계 및 빅데이터 분석에는 R을 주로 사용하였다. R의 빅데이터 분석용 패키지 중에서 텍스트 분석을 위한 TM, 한글 텍스트 마이닝 도구인 KONLP, 그리고 각종 통계 분석용 패키지를 플러그인하여 사용하였다. 텍스트에서 주요 요인을 추출하기 위해서는 문장 형태소 추출이 필요하다. 본 연구에서는 형태소 사전으로 국립국어원의 우리말샘 사전을 바탕으로 제작되고 신조어와 법률, 의료 등 6대 분야별 전문 단어를 추가하여 빈도수 및 연관 분석이 가능한 NiaDic을 KONLP의 기초 사전으로 이용하였다. 이외에 텍스트 정제 및 가공을 위하여 Visual Basic for Applications와 Python, 주요 요인 키워드 및 차원에 대한 인과관계 분석을 위하여 통계 패키지인 SPSS를 부가적으로 사용하였다.

4.1.1 데이터 사전 처리

데이터 사전 처리는 분석의 가장 기초가 되는 단계로 데이터를 분석 가능한 상태로 만든다. 우선 비정형 텍스트 형태인 검토의견을 R에 등록하고 한글 형태소 분석 도구인 KONLP 패키지에 한글 사전 NIADic을 플러그인하여 사전에 등록된 단어와 매핑, 이를 통해 명사형 키워드를 추출하여 형태소 세트를 만든다. 이렇게 만들어진 형태소 세트에서 분석에 불필요한 숫자, 구두점, 공백 등 여러 가지 불필요한 단어들을 정제하기 위하여 TM 패키지의 여러 함수들을 이용한다. 또한 추출된 명사형 키워드에 대해 빈도수 측정을 위하여 정제된 데이터를 데이터 프레임 형태로 변환, 출현 빈도수에 따라 도수 분포표로 변환하는 과정을 거치게 된다.

일련의 데이터 정제 과정을 거쳐 전체 검토 의견에서 총 4,674개의 초기 키워드를 추출했다. 데이터 정제과정을 거쳐서 불필요한 단어를 제거 했지만, 아직 완전한 정제가 이루어진 상태는 아니다. 정밀한 분석을 위해 직접 추출된 단어를 수작업으로 정제를 해야 한다. 우선 자주 사용하지 않는 키워드(빈도수 5회 이하)를 제거 하고, ‘함’, ‘적’과 같은 의존명사, 숫자 등을 제거하여 1차적으로 1,024개의 분석 대상 키워드를 선별하였다.

4.1.2 키워드 사전 구축 및 차원 분류

데이터 사전 처리를 통해서 추출된 1,024개의 키워드에서 평가에 영향을 미치는 주요 키워드를 제안 평가 경험이 있는 전문가들이 개인별 선별 작업과 토론을 통하여 최종적으로 20개를 선정했다. 또한 차원 기반 연구

모형 적용을 위해서 20개의 주요 키워드를 이해력, 수행력, 및 관리력의 3개 차원으로 분류하였다.

- 이해력 : 이해(Comprehension), 방안(Method), 전략(Strategy), 기획(Planning), 경험(Experience), 실적(References)
- 수행력 : 제안(Proposal), 계획(Plan), 인력(Manpower), 수행(Execution), 기술(Technology), 일정(Schedule), 투입(Input), 방법(How-to), 개발(Development), 능력(Capability), 조직(Organization)
- 관리력 : PM(Manager), 관리(Management), 품질(Quality)

오피니언 마이닝을 위해서는 주요 요인 키워드 뿐 아니라 감정의 방향, 즉 극성에 대한 키워드 사전도 구축할 필요가 있다. 앞서 추출된 1,024개의 키워드에서 전문가 검토 및 토의를 거쳐 21개가 추출되었다. 또한, 추출된 극성 키워드에는 감정의 정도에 해당하는 긍정, 중립, 부정을 나타내는 키워드가 섞여 있는데, 이 키워드를 다시 전문가 토의를 통해 긍정, 중립, 부정을 의미하는 키워드로 분류하였다.

- 긍정 : 탁월, 훌륭, 우수, 참신, 적절, 충실, 충분, 충족, 풍부, 양호, 적정
- 중립 : 평범, 평이, 무난, 부합
- 부정 : 아쉬움, 필요, 부족, 미흡, 적음, 우려, 무능

4.1.3 검토 의견의 계량화

극성 키워드들이 같은 그룹으로 분류된다고 하더라도 동일한 정도의 극성을 나타낸다고 할 수 없다. 예를 들어, ‘탁월’과 ‘양호’와 같이 같은 긍정의 극성을 지닌 키워드라 할지라도 실제 평가자가 느끼는 긍정의 정도는 ‘탁월’이 더 크다. 이에 본 연구에서는 이를 반영하여 개별 극성 키워드에 극성 점수를 부여했다. 전문가 집단의 검토를 통해 동일 그룹의 극성 키워드에 대해 순위를 정하고, 가장 긍정적인 극성을 1점, 가장 부정적인 극성을 0.5점, 그리고 극성 키워드의 순위에 따라서 0.01~0.02점 간격으로 점수를 부여 했다. 예를 들어, 긍정 키워드 중 1순위인 ‘탁월’은 1.0, 2순위인 ‘훌륭’은 0.98로 점수가 부여되었다.

다음 단계는 텍스트로 구성되어 있는 검토 의견에 대한 계량화를 진행하는 단계이다. 우선 검토 의견에 주요 요인 키워드와 극성 키워드가 포함되어 있는지의 여부를 검색하였다. 둘 다 포함되어 있는 검토 의견에 대해서는 극성 키워드의 점수를 부여했다. 만약 하나의 검토 의견에 복수 개의 극성 키워드가 포함되어 있을 경우에는 검토 의견에 포함된 각각의 극성 키워드의 평균을 구하여 점수를 산출했다. 차원 기반 모형 분석을 위한 차원별 점수도 유사한 방식으로 산정하였다.

4.2 분석 결과

앞에서 분석 가능한 상태로 가공한 데이터에 대하여 엑셀과 VBA를 이용하여 키워드별, 차원별 출현 빈도 및 비율과 같은 기본적 통계 분석을 진행하였다. 또한 주요 키워드와 차원의 인과 관계를 통계 분석 도구인 SPSS를 이용하여 분석하였다.

4.2.1 키워드 기반 모형

검토의견 내의 주요 키워드 및 차원별 출현 빈도를 분석한 결과 <Table 2>와 같은 결과가 도출 되었다. 전체 키워드에 대해 비교적 고른 분포를 하고 있지만 ‘이해’, ‘방안’, ‘제안’, ‘관리’, ‘계획’과 같은 키워드의 빈도가 가장 높았다. 이는 이들 키워드가 평가를 진행함에 있어 평가자들이 가장 관심을 가지는 항목임을 의미한다.

<Table 2> Frequency Analysis Results

Dim.	Keyword	Freq.	Freq.	Ratio
Understanding	Comprehension	573	2,228	7.46%
	Method	581		7.57%
	Strategy	383		4.99%
	Planning	158		2.06%
	Experience	399		5.20%
	References	134		1.75%
Execution	Proposal	713	4,309	9.29%
	Plan	562		7.32%
	Manpower	563		7.33%
	Execution	510		6.64%
	Technology	424		5.52%
	Schedule	305		3.97%
	Input	258		3.36%
	How-to	513		6.68%
	Development	183		2.38%
	Capability	156		2.03%
Management	Organization	122	1,139	1.59%
	Manager	50		0.65%
	Management	698		9.09%
	Quality	391		5.09%

또한 <Table 3>과 같이 검토의견 내의 극성 키워드 출현 빈도를 분석한 결과 긍정이 부정에 비하여 4.7배 높은 결과가 나왔다. 평가 결과 낙찰자 선정시 대부분 1~2% 정도의 미세한 차이로 결정된다는 점과 평가자가 평가 점수 입력 및 의견을 등록할 때 감정의 표현을 상대적으로 자제한다는 특성 등을 고려하면, 이러한 분석 결과는 긍정 키워드 그룹에서 긍정의 정도를 매우 정밀하게 분류할 필요가 있음을 의미한다. 극성이 없거나 중립적 극성을 표하는 의견은 전체의 40% 정도이다. 이 부분

도 평가자가 검토의견을 작성할 때 낮은 점수를 부여하지 않고 감정적인 표현을 자제하는 경향을 데이터로 확인한 결과라 할 수 있다.

<Table 3> Frequency for Polarity Keywords

Polarity		Freq.	Ratio
Polar.	Positive	4,529	47.57%
	Neutral	279	2.93%
	Negative	1,011	10.62%
Not Polarized		3,701	38.88%

키워드별 산출 점수와 실제로 부여한 평가 점수 간의 인과관계를 상관계수로 분석한 결과는 <Table 4>와 같다. 키워드별로 0.291~0.558사이의 상관관계 값을 보이며, 특히 방안, 전략, 기획, 실적, 계획, 기술, PM과 같은 키워드의 상관관계 값이 높다. 또한, 출현 빈도수가 높은 4개 키워드와 전문가들이 가장 중요한 키워드로 꼽은 ‘PM’에 대하여 회귀분석을 실시하였다(<Table 5> 참조). 회귀식의 설명력을 의미하는 결정계수 의 값은 0.135~0.311의 값인데, 단일 값으로 보면 비교적 낮은 수치이지만 20개 키워드 중 하나만으로 구성된 회귀모형의 설명력이라는 점을 감안한다면 본 연구의 목적 중의 하나인 검토의견에서 평가 점수를 도출할 수 있는 가능성을 검증할 수 있는 근거가 된다.

<Table 4> Correlation Analysis for Key Factors

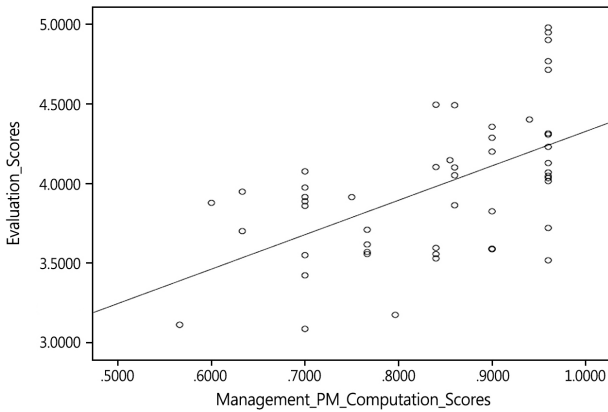
Keyword	Correl. Coeff.	Keyword	Correl. Coeff.
Comprehension	0.376	Technology	0.434
Method	0.413	Schedule	0.282
Strategy	0.455	Input	0.371
Planning	0.501	How-to	0.291
Experience	0.312	Development	0.373
References	0.475	Capability	0.36
Proposal	0.37	Organization	0.357
Plan	0.435	Manager	0.558
Manpower	0.377	Management	0.368
Execution	0.293	Quality	0.348

주요 키워드 선정 당시 ‘PM’은 전문가들이 가장 중요하게 생각하는 요인 중 하나였다. 그러나 실제 분석 결과 빈도수가 50으로 타 키워드에 비해 현저히 낮았다. 그런데 회귀분석 결과에서는 다른 키워드에 비해 훨씬 높은 회귀계수와 설명력을 보인다. 이는 PM이라는 키워드가 타 요인들에 비해 자주 언급되지는 않지만, 일단 언급되면 평가 점수에 큰 영향을 미치는 중요 키워드로서, 검토 의견의 수치화 가능성 파악이라는 본 논문의 연구 목적에 비추어 볼 때 가장 관심을 가져야 하는 키워드라 할 수 있다. PM 키워드와 평가 점수에 대한 산점도는 <Figure 5>와 같다.

<Table 5> Regression Analysis for Key Factors

Keyword	Regr. Coeff.	R ²
Comprehension	0.376**	0.141
Method	0.413**	0.171
Proposal	0.370**	0.137
P.Manager	0.558	0.311
Management	0.368	0.135

** p < 0.05.



<Figure 5> Scatter Plot for PM

4.2.2 차원 기반 모형

<Table 2>의 차원별 빈도 분석 결과, 빈도수가 수행력, 이해력 그리고 관리력 순으로 높게 나타났다. 이는 해당 차원의 중요도 보다는 포함된 키워드 수에 따른 빈도 차이를 반영한 것으로 보이며 각 차원별 선호도의 차이는 크지 않은 것으로 생각된다. 차원별 산출 점수와 실제 부여 점수 간의 인과관계 분석을 실시하였는데, 우선 차원별 상관계수 값은 이해력, 수행력, 관리력에 대해 각각 0.402, 0.377, 0.378로 나타났다. 이는 키워드 모형에서의 상관계수 값과 유사한 수준이다. 차원별 회귀분석 결과는 <Table 6>과 같다. 회귀계수는 0.3~0.4 정도의 수준이며 차원별 설명력을 의미하는 결정계수 값은 14~16% 정도이다. 또한 차원별 산점도 및 회귀선을 분석해본 결과 실제 점수의 분포가 3~5점 사이에 많은 데이터가 분포하고 있는 것을 확인할 수 있는데, 이는 일반적으로 평가자가 점수를 부여할 때 1~2점의 극단적으로 낮은 하위 점수보다는 3~5점 사이의 중 상위 점수를 중점적으로 많이 부여하고 평가 기업별로 점수를 부여할 때 작은 차이로 점수를 부여하는 경향이 있기 때문이다.

<Table 6> Regression Analysis for Dimensions

Dimension	Regr. Coeff.	R ²
Understanding	0.402**	0.162
Execution	0.377**	0.142
Management	0.378**	0.143

** p < 0.05.

5. 논의 및 시사점

우선 본 연구에서는 빅데이터 기반의 오피니언 마이닝을 적용하기 위한 일련의 데이터 준비 과정(데이터 수집, 디지털화, 데이터 정제 등)이 연구 영역인 정보화 사업 제안 평가에서 원활하게 수행될 수 있음을 확인할 수 있다. 이는 연구의 대상이 되는 비정형 텍스트 형태의 검토 의견이 트위터와 같은 SNS 텍스트에 비해 보다 공식적이고 감정이 절제된 형태로 작성되기 때문으로 생각된다.

평가에 영향을 미치는 요인을 파악하는 작업에서, 데이터 분석을 통해 평가 결과에 영향을 미치는 요인들에 대한 분석을 진행해본 결과 20개의 주요 요인 키워드 중 특히 이해, 방안, 제안, 관리, 계획, 인력과 같은 키워드가 검토의견 내에서 가장 빈도가 높았다. 차원별 분석 결과에서는 수행력, 이해력, 그리고 관리력의 순으로 빈도수가 나타났다. 또한 극성 키워드에 대한 분석 결과에서 전체 검토의견 중 긍정 의견의 빈도가 부정 의견에 비하여 약 5배 정도 높았다. 이로부터 평가자들이 가장 관심을 기울이는 요인을 파악할 수 있었고 검토의견에서 부정적 의견은 우회적이고 간접적으로 피력하는 경향이 있음을 알 수 있다.

요인과 평가 점수 사이의 인과관계 분석 결과를 살펴보면, 우선 차원 기반 모형에서의 분석값(상관계수 및 회귀계수, 설명력 등)이 개별 키워드 기반 모형과 큰 차이가 없다. 따라서 굳이 차원별 요인으로 그룹화하여 인과관계를 파악할 필요는 없다. 그리고 각각의 키워드 인과관계 모형이 모두 어느 정도의 설명력을 보여주고 있다. 구체적으로는 방안, 전략, 기획, 실적, 계획, 기술, 그리고 PM 키워드가 여타 요인(키워드)에 비해 다소 높은 상관관계와 설명력을 보여줄 수 있다. 또한 극성에 대한 분석 결과가 보여 주듯이 여러 요인 중에서도 평가자가 관심을 기울이고 있는 특정 키워드에 대한 언급 여부가 평가 점수에 큰 영향을 미침을 알 수 있다. 이는 한글의 극성 판별 방법론이 확립되면, 소수의 키워드 집합을 가지고도 검토의견으로부터 신뢰할만한 평가 점수를 산출할 수 있음을 의미한다.

인과관계 분석과 더불어 본 연구의 또 다른 목적은 검토 의견의 수치화 가능성에 대한 평가이다. 단어와 평점의 상관관계 분석을 통하여 영화평 평점을 예측하고자 한 연구와 비교할 때[29], 본 연구의 결과가 보다 양호한 결과를 보여준다고는 할 수 없다. 이는 빅데이터 분석을 하기에는 검토 의견이 질과 양적인 측면에서 충분하지 않았고, 신뢰할만한 극성 어휘 사전의 부재로 인하여 단순 극성 키워드의 예측 값으로 점수를 산출하였으며, 주요 요인 키워드에 대한 중요도 또한 연구 목적상 차별화시키지 않았기 때문으로 생각된다. 그러나 본 연구가 빅

데이터 기반 제안 평가의 가능성 여부에 대한 판단을 목적으로 하고 있으므로 향후 연구에서 보다 많은 데이터를 수집하고, 제안 평가 영역에 특화된 정밀한 키워드 사전을 구축하여 정확도를 높일 수 있다면, 본 연구 결과도 제시된 접근 방법이 충분히 유효한 결과를 도출할 수 있을 것으로 판단된다.

6. 결 론

본 연구에서는 정부 발주 정보화 사업 평가에 빅데이터 분석 기법인 오피니언 마이닝을 적용하였다. 텍스트 형태로 표시된 평가자의 검토의견으로부터 평가 결과에 영향을 미치는 요인들을 파악하는 인과관계 분석을 실시하였다. 그리고 정성적인 검토 의견에서 정량적 평가 점수를 산출하여 평가 및 선정에 활용할 수 있는 가능성을 알아보았다. 우선 인과관계 분석은 요인들을 차원으로 그룹화하거나 소수의 키워드 집합을 그대로 요인으로 사용하는 두 가지 방식으로 수행하였는데, 양쪽 접근방식 모두 비슷한 결과를 보여 주었다. 또한 검토 의견으로부터 평가 점수를 산출하여 직접적으로 부여한 점수와 비교한 결과, 평가 요인 별, 항목 별로 세분된, 신뢰할만한 점수를 산출할 수 있는 가능성이 있음을 알 수 있었다. 그러나 실제 평가에 활용되기 위해서는 보다 정교한 분석 방법이 필요하다고 판단된다. 이를 위해서는 여러 분야에서의 폭 넓은 기반 연구, 특히 한글과 관련된 키워드 사전 구축과 극성 키워드 수치화에 대한 연구가 요구된다.

Acknowledgement

This study has been partially supported by a 2017 Research Fund of Hannam University, Daejeon, Korea.

References

- [1] Ahmed, K., El Tazi, N., and Hossny, A., *Sentiment Analysis over Social Networks : An Overview*, IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, 2015.
- [2] Asur, S. and Huberman, B., *Predicting the future with social media.*, Arxiv preprint arXiv:1003.5699, 2010.
- [3] Cheon, M. and Baek, D., An Assessment System for Evaluating Big Data Capability Based on a Reference Model, *Journal of Society of Korea Industrial and Systems Engineering*, 2016, Vol. 39, No. 2, pp. 54-63.
- [4] Choi, S. and Kwon, O., The Study of Developing Korean SentiWordNet for Big Data Analytic : Focusing on Anger Emotion, *The Journal of Society for e-Business Studies*, 2014, Vol. 19, No. 4, pp. 1-19.
- [5] Dave, K., Lawrence, S., and Pennock, D., *Mining the Peanut Gallery : Opinion Extraction and Semantic Classification of Product Reviews*, WWW '03 Proceedings of the 12th international conference on World Wide Web, 2003, pp. 519-528.
- [6] Duan, W., CaQ, Q., Yu, Y., and Levy, S., Mining Online User-Generated Content-Using Sentiment Analysis Technique to Study Hotel Service Quality Hawaii International Conference On System Sciences, 2013, pp. 3119-3128.
- [7] Feldman, R., Ronenfeld, B., Bar-Haim, R., and Fresko, M., The Stock Sonar-Sentiment Analysis of Stocks Based on a Hybrid Approach., *Proceedings of 23rd IAAI Conference on Artificial Intelligence(IAAI-2011)*, 2011.
- [8] Goeriot, L., Kelly, L., Jones, G., Müller, M., and Zobel, J., Report on the SIGIR 2014 Workshop on Medical Information Retrieval (MedIR), *ACM SIGIR Forum*, 2014, Vol. 48, No. 2, pp. 78-82.
- [9] Hong, Y. and Steven, S., The Wisdom of Bookies? Sentiment Analysis vs. the NFL Point Spread., *Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2010)*, 2010.
- [10] Hu, H. and Liu, B., Mining and summarizing customer reviews, *KDD '04 Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 168-177.
- [11] Hwang, H.-S., Clustering Corporate Brands based on Opinion Mining : A Case Study of the Automobile Industry, *Journal of the Korea Academia-Industrial Cooperation Society*, 2016, Vol. 17, No. 11, pp 453-462.
- [12] Jeon, S., Lee, D., and Bae, M., A Study on the Application Method of Munition's Quality Information based on Big Data, *Journal of the Korea Academia-Industrial Cooperation Society*, 2016, Vol. 17, No. 6, pp. 315-325.
- [13] Kamps, J., Marx, M., Mokken, R.J., and Rijke, M.D., Using *WordNet* to Measure Semantic Orientation of Adjectives, In *Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation*, 2004, Vol. 4, pp. 1115-1118.
- [14] Kim, S. and Kim, N., A Study on the Effect of Using Sentiment Lexicon in Opinion Classification, *Journal of Intelligence and Information Systems*, 2014, Vol. 20, No. 1, pp. 133-148.

- [15] Lee, J. and Baek, D., The Effect of Smartphone Purchasing Determinants on Repurchase Intention, *Journal of Society of Korea Industrial and Systems Engineering*, 2017, Vol. 40, No. 2, pp. 1-12.
- [16] Liu, B., Sentiment Analysis and Opinion Mining, *Computational Linguistics*, 2014, Vol. 40, No. 2, pp 511-513.
- [17] Liu, B., Sentiment Analysis and Opinion Mining, Morgan and Claypool Publishers, 2012.
- [18] Liu, B., *Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data*, Springer, 2011.
- [19] Medhat, W., Hassan, A., and Korashy, H., Sentiment analysis algorithms and applications : A survey, *Ain Shams Engineering Journal*, 2014, Vol. 5, No. 4, pp. 1093-1113.
- [20] Mohammad, S. and Yang, T., Tracking Sentiment in Mail : How Genders Differ on Emotional Axes, Proceedings of the ACL Workshop on ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA-2011), 2011.
- [21] Mohammad, S., From Once Upon a Time to Happily Ever After : Tracking Emotions in Novels and Fairy Tales., Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), 2011.
- [22] O'Connor, B., Balasubramanyan, R., Routledge, B., and Smith, N., From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series., Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM 2010), 2010.
- [23] Rajput, Q., Haider, S., and Ghani, S., Lexicon-Based Sentiment Analysis of Teachers' Evaluation, Applied Computational Intelligence and Soft Computing, 2016.
- [24] Roh, J., Kim, H., and Chang, J., Improving Hypertext Classification Systems through WordNet-based Feature Abstraction, *Journal of Society for e-Business Studies*, 2013, Vol. 18, No. 2, pp. 95-110.
- [25] Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I., Predicting elections with twitter : What 140 characters reveal about political sentiment, Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2010), 2010.
- [26] Turney, P. and Littman, M., Measuring Praise and Criticism : Inference of Semantic Orientation from Association, *ACM Transactions on Information Systems*, 2003, Vol. 21, No. 4, pp. 315-346.
- [27] Turney, P., Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Proceedings of Annual Meeting of the Association for computational Linguistics, 2002, pp. 417-424.
- [28] Wang, X. and Park, B., Extraction of Feature and Opinion Words from Product Reviews Using the 12-Structure, *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, 2014, pp. 896-897.
- [29] Yoon, D. and Kim, K., Prediction of Rating Score from Short Comments on Movies using Word-Rating Correlation Analysis, *HCI Korea*, 2011, No. 1, pp. 484-486.

ORCID

Hong Sam Kim | <http://orcid.org/0000-0001-7728-7263>

Chong Su Kim | <http://orcid.org/0000-0002-0172-3679>