

SW 교육 뉴스데이터의 감성분석

박선주

광주교육대학교 컴퓨터교육과

요 약

스마트폰의 대중화로 SNS를 통해 유통되는 정보의 내용과 감성을 분석하는 연구가 활발하게 진행되고 있다. 이에 본 논문에서는 SW 교육에 관한 온라인 뉴스데이터를 수집하여 형태소 분석후 단어를 추출하고 뉴스데이터의 감성지수를 산출하여 수집된 뉴스 데이터의 감성분석을 실시하였다. 또한, 산출된 감성점수가 어느 정도 정확한지 정확도를 검토하였다. 분석 결과 수집기간동안 SW 교육 관련 뉴스는 월평균 약 189건 발생되었으며, 감성점수 평균은 0.7로 SW 교육 관련 뉴스는 긍정적임을 알 수 있었다. SW 교육의 중요성 및 정책 실행에는 공감하며 긍정적이었으나 구체적인 실행 방법에는 부정적인 시각이 있었다. 즉, SW 교육환경 및 교육방법 부족 문제, SW 개발자 양성 및 처우개선 문제, 코딩 사교육 증가 문제 등이었다.

키워드 : 빅데이터, 텍스트분석, 감성분석, R, SW교육

Sentimental Analysis of SW Education News Data

SunJu Park

Dept. of Computer Science Education, Gwangju National University of Education

ABSTRACT

Recently, a number of researches actively focus on the contents and sensitivity of information distributed through SNS as smartphones and SNS gained its popularity. In this paper, we collected online news data about SW education, extracted words after morphological analysis, and analyzed emotions of collected news data by calculating sentimental score of each news datum. Also, the accuracy of the calculated sentimental score was examined. As a result, the number of news related to 'SW education' in the collection period was about 189 per month, and the average of sentimental score was 0.7, which signifies the news related to 'SW education' was emotionally positive. We were positive about the importance of SW education and the policy implementation, but there were negative views on the specific method for the realization. That is, a lack of SW education environment and its education method, a problem related to improvement of SW developers and improvement of their labor conditions, and increase of private education in coding were the factors for the negative viewers.

Keywords : Big data, Text analysis, Sentimental analysis, R. SW education

본 논문은 2015년도 광주교육대학교 학술연구비 지원에 의한 것임.

논문투고 : 2016-12-00

논문심사 : 2017-00-00

심사완료 : 2017-00-00

1. 서론

스마트폰의 대중화로 SNS를 통해 사람들은 자신의 생각과 의견 등을 적극적으로 올리고 서로 공유하고 있다. 일반적으로 사람들의 의견이나 생각을 조사할 때 설문지, 면접 및 관찰 등의 방법을 사용하지만, SNS를 통해 유통되는 정보의 양이 많아지고 영향력이 커지므로 빅데이터를 통해 사람들의 생각을 분석하고자 하는 연구가 새롭게 진행되고 있다[1,3].

텍스트 형태의 빅데이터에서 단어들을 추출하고 단어들 간의 관계를 통해 빅데이터를 분석하여 유용한 정보를 찾아내는 텍스트 마이닝 방법과 이를 통해 미래의 흐름을 예측하고 추천하는 분석방법이 연구되고 있다. 또한, 사람들의 감성을 분석하기 위한 감성분석(Sentimental Analysis) 방법도 있는데, 감성 분석은 특정 주제에 대한 감정을 점수화하여 감정의 정도를 예측할 수 있으며, 최근 들어 SNS의 급부상과 함께 다양한 분야에 오피니언 마이닝(Opinion Mining)의 개념으로도 적용되고 있다[1,3,13,16]. 특히 기업에서는 소비자들의 생각을 분석하기 위해 방대한 양의 데이터를 자동으로 처리해주는 시스템을 개발 및 도입하고 있다.

감성분석을 사용한 대표적인 사례로는 2012년 미국 대선[2], 주가예측[3,13], 온라인 쇼핑몰의 상품평 분석[6], 영화흥행성 예측[4] 등이 있다. 또한, 뉴스데이터를 분석한 연구로 문광수 등은 웹 뉴스 댓글의 방향성이 일반 댓글의 동조성에 미치는 효과를 연구하였으며[15], 차은정 등은 뉴욕타임즈의 뉴스정보를 감성분석하여 주식예측을 진행하였으며[3], 최석재 등은 온라인 뉴스 제목을 분석하여 특정 장소에서 이루어질 이벤트 성과를 예측하는 연구[5]를 하였다.

한편 2015년 ‘소프트웨어 교육 운영지침’에 의해 중·고등학교는 2018년부터, 초등학교는 2019년부터 SW 교육이 이루어지므로 SW 교육에 관한 사람들의 의견은 어떤지 살펴볼 필요가 있다. 한국인의 약 77%가 인쇄신문이 아닌 온라인을 통해 뉴스를 접하고 있으며 SNS를 통해 다양한 정보를 공유하고 있다고 한다[9]. 그러므로 SW 교육에 관한 온라인 뉴스데이터를 수집하여 SW 교육에 관한 뉴스가 긍정적인지 부정적인지를 살펴보고자 한다.

그러므로 본 논문에서는 SW 교육에 관한 온라인 뉴

스데이터를 수집하여 형태소 분석후 단어를 추출하고 뉴스데이터의 감성지수를 산출하여 수집된 뉴스 데이터의 감성분석을 실시하고자 한다. 또한, 산출된 감성점수가 어느 정도 정확한지 정확도를 검토하고자 한다.

2. 이론적 배경

2.1 감성분석

감성분석은 오피니언마이닝의 한 부분으로 글의 주제와 내용보다는 글쓴이가 글의 주제에 대한 감성을 찾아내는 방법으로 긍정적인 의미인지 부정적인 의미인지를 알아내는 방법이다. 즉, 텍스트에서 감성단어를 추출하여 점수화하는 방법으로 코퍼스(Corpus)를 이용한 기계학습 기반 분석과 사전기반 분석이 있다.

기계학습기반 분석은 인간의 학습절차 모델을 기반으로 컴퓨터가 해석할 수 있도록 훈련시켜 긍정 혹은 부정으로 분류하게 된다. 이 방법은 정확도가 높은 반면 코퍼스 구축이 어렵고 과적합(overfitting) 문제가 발생하고 데이터가 많이 필요하게 되는 단점이 있다. 사전기반 분석은 사용하기 간편하지만 주제에 따라 사전이 달라지는 단점이 있다. 왜냐하면 대상의 속성에 따라 긍정 혹은 부정의 분류가 달라질 수 있기 때문이다.

텍스트마이닝, 감성분석을 통해 더 나아가 선형회귀분석, SVM(Support Vector Machine), RandomForest, Deep Learning 모형을 활용한 예측분석도 실시할 수 있다.

감성분석을 사용한 사례로는 2012년 미국 대선[2], 주가예측[3,13], 온라인 쇼핑몰의 상품평 분석[6], 영화흥행성 예측[4] 등이 있다.

특히 기업의 경우 규모가 커질수록 소비자의 의견을 일일이 분석하는 것은 매우 어려우므로 소셜 데이터를 자동으로 처리하는 감성 분석 솔루션을 제시하고 있다. 감성분석 솔루션으로는 어도비 소셜(Adobe Social)의 예측 퍼블리싱(Predictive Publishing), IBM의 감성 분석 서비스인 IBM 소셜 애널리틱스(Social Analytics), SK플래닛의 OMS 2.0 시스템, SK텔레콤의 Smart Insight, 구글 애널리틱스 프리미엄(Google Analytics Premium), AT 인터넷 애널리라이저 III(Analyzer III), SAS 어드밴스드 커스터머 익스피리언스(Advanced

Customer Experience) 등이 있다[19].

2.2 감성사전

감성분석 과정에서 감성사전이 필요한데 한국어 감성사전은 영어와는 달리 어미와 조사가 발달한 교착어이므로 자연어 처리가 복잡해 한글 감성분석이 어려운 부분이 있다[1].

감성사전 구축의 대표적인 예는 PMI(Point-wise Mutual Information)를 활용해 단어의 긍정, 부정, 중립을 평가한 SentiWordNet 연구가 있다. 그러나 범용 감성사전이 주제에 따른 문맥적 의미 변이, 동음이의어 사용 등의 이유로 감성분석의 정확성이 낮아지는 문제가 발생한다[14]. 또한 서울대학교에서 개발한 영어의 MPQA(Multi-perspective Question Answering) 감성사전과 같은 한국어 감성 코퍼스 KOSAC(Korean Sentiment Analysis Corpus), 연세대학교에서 개발한 오픈한글 등의 연구가 있다.

감성사전은 감성분석과정에서 중요한 부분이지만 한글 감성사전에 대한 연구가 많지 않고 계속 진행되고 있으므로 별도로 연구자들이 주제에 맞게 사전을 구축하여 연구하고 있다. 따라서 주제에 적절한 맞춤형 감성사전을 구축하여 감성분석의 정확성을 높이려는 연구들이 진행되고 있다. 이에 이상훈 등은 영화 데이터 특성에 맞는 SO-PMI(Semantic Orientation from Point-wise Mutual Information)를 활용해 장르별 맞춤형 감성사전 구축을 통한 영화리뷰 감성분석을 실시하였다[14].

본 논문에서도 주제에 맞는 감성사전을 구축하기 위해 회귀분석을 통한 긍정어와 부정어를 추출하여 이를 수정 보완하여 활용하였다.

2.3 관련 연구

뉴스데이터의 감성분석에 관한 연구를 살펴보면, 차은정 등은 뉴욕타임즈의 뉴스정보를 감성분석하여 주식예측을 진행하였으며[3], 최석재 등은 온라인 뉴스 제목을 분석하여 특정 장소에서 이루어질 이벤트 성과를 예측하는 연구[5]를 하였다.

이 외에 문화산업에서 아이돌 스타에 대한 감성 분

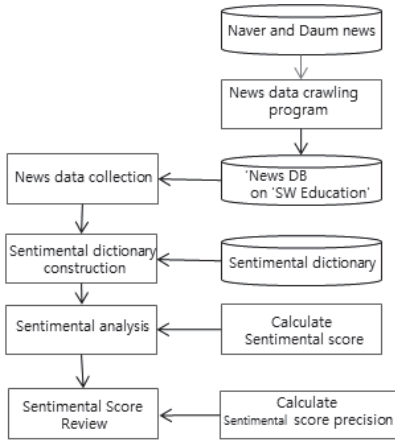
석, 온라인상에 작성된 의료기관의 고객 여론을 분석한 연구, 베스트 댓글의 방향성이 일반댓글의 동조효과에 미치는 영향 연구, 신문기사로부터 추출한 최근동향에 대한 트위터 감성분석 연구 등 다양한 분야로 확대되어 연구되고 있다[15,17].

감성사전 구축에 관한 연구로는 단어별 극성판별에 집단지성을 활용한 연구[1], 단어의 출현 빈도수를 활용해 주제별 감성사전을 구축한 연구[21], 주식관련 콘텐츠에 OAR(Opinion Antonym Rule) 알고리즘을 사용하여 감성사전을 구축한 연구[8], LP(Label Propagation) 방법을 이용해 단어 간의 인접도를 통해 감성사전을 구축한 연구[11] 등이 있으며, 이밖에 TF-IDF (Term Frequency-Inverse Document Frequency) 등의 수치에 기반하거나 PMI 등의 통계적 수치를 활용하여 단어 간의 극성을 정의하는 방법도 있다[7,20]. 안정국 등은 집단지성을 이용한 한글 감성어 사전을 구축하기 위해 오픈한글 사이트를 운영하고 있어[1] 한글 감성분석에 새로운 방향이 되리라 생각한다. 특히 송상일 등은 범용 감성사전을 이용하는 방식보다는 주제에 맞는 감성사전을 활용할 때 감성분석의 정확성이 향상됨을 확인하였다[18].

선행연구를 통해 범용 감성사전을 이용하는 방식보다는 주제에 맞는 감성사전을 구축하여 감성분석에 활용하는 방안이 더 효과적임을 알 수 있으며[14], 또한 SW 교육 관련 뉴스 데이터의 감성분석에 관한 연구는 거의 없음을 알 수 있었다,

3. 연구방법

본 연구에서는 SW 교육과 관련된 온라인 뉴스 데이터의 감성분석을 실시하기 위해 수집된 데이터를 형태소 분석하여 단어를 추출한 후 감성사전과의 비교를 통해 텍스트의 감성지수를 산출하여 뉴스 데이터의 감성분석을 실시하였다. 또한, 산출된 감성점수가 어느 정도 정확한지 정확도를 검토하였다. 연구절차는 (Fig. 1)과 같다.



(Fig. 1) The research procedures

3.1 데이터 수집

본 연구에서는 네이버와 다음 검색사이트에서 제공하는 온라인 뉴스를 뉴스 헤드라인과 요약정보를 중심으로 수집하였다. 뉴스 기사 전문은 핵심이 아닌 이해를 돕기 위한 주변의 내용이 들어가는 경우가 있어 극성 판단에 혼란을 줄 수도 있으므로 뉴스 헤드라인과 요약 정보를 수집하였다. 뉴스 헤드라인은 기사의 가장 핵심적인 내용을 담고 있어 본문보다 사실과 의견이 더 정확하게 발현될 뿐 아니라 데이터의 불용어가 작고 간결한 특징을 갖고 있다. 그러나 함축적인 뉴스 제목에 감성을 판단하기가 어려울 수도 있으므로 요약정보를 함께 수집하였다.

이클립스를 통해 2016년 1월 1일부터 2016년 8월 22일까지 데이터를 수집하였으며, 키워드는 'SW 교육'과 '소프트웨어 교육' 으로 검색하여 수집하였다. 수집된 뉴스 데이터(예)는 (Fig. 2)와 같으며, R에서 데이터를 불러올 수 있는 .CSV 형식으로 저장하였다.

A	B	C	D	E
news_date	title	text		
2016.05.26.	[소프트웨어 놀이데이] SW교육콘텐츠 세계가 주목하고 있는 소프트웨어			
2016.07.04.	2016년 미래부 SW창의캠프 개최 이번 기초캠프 참가자들은 SW교육			
2016.07.05	영재 중심 CT-R&E SW 교육...내년 소프트웨어(SW)역량이 미래 사회			

(Fig. 2) Collected news data (example)

3.2 감성분석

해당 문서의 주제별 맞춤형 감성사전을 구축하여 감성분석에 활용하는 방안이 더 효과적이라는 선행연구 [14] 결과를 기반으로 수집된 뉴스데이터의 극성 탐지를 위해 먼저 감성사전을 구축하였다. 먼저 수집된 뉴스데이터를 형태소 분석하여 단어를 추출한 후 회귀분석을 통해 회귀계수를 산출하여 회귀계수가 0보다 크면 긍정 단어, 회귀계수가 0보다 작으면 부정단어로 분류하여 감성사전을 만들었다. 여기에 본 연구의 뉴스 데이터에 적절한 긍정어와 부정어를 추출하여 긍정어 사전과 부정어 사전을 수정 보완하였다.

구축된 감성사전을 기반으로 뉴스데이터의 감성점수를 계산하였다. 감성점수(SC)는 긍정 단어수(PW)의 합에서 부정 단어수(NW)의 합을 뺀 값으로 나타냈다. 즉, 감성점수 SC는 다음과 같다.

$$SC = PW_i - NW_j \quad i = 1 \dots n, j = 1 \dots m$$

이와 같이 뉴스데이터의 감성점수를 계산하여 뉴스데이터에 저장하였다. 감성점수값으로 이 뉴스가 긍정적인지 부정적인지 또는 중립적인지를 구분할 수 있다. 즉, <Table 1>과 같이 감성점수가 1 이상이면 긍정, 감성점수가 0 이면 중립, 감성점수가 0 미만이면 부정으로 판단한다.

<Table 1> Sentimental Condition of Text

Text type	Condition
Positive Text	SC > 0
Neutral Text	SC = 0
Negative Text	SC < 0

3.3 감성점수 정확도 검토

감성사전 기반 방식으로 부여된 감성점수의 정확도를 측정하기 위해 3명이 뉴스데이터를 보고 긍정, 부정, 중립을 판단하였으며 3명의 판단이 차이가 나는 경우는 서로 의견을 나눠 결정했다.

실제 문서의 긍정/부정 결과와 감성분석에 의해 예측된 결과의 관계를 <Table 2>와 같이 나타낼 때 정밀도는

예측한 분류의 결과가 얼마나 정확한가를 나타낸다[3].

<Table 2> Relationship between real document and predicted results

		prediction	
		Positive	Negative
real	True	<i>TruePositive</i>	<i>TrueNegative</i>
	False	<i>FalsePositive</i>	<i>FalseNegative</i>

정밀도(Precision)는 긍정이라고 예측(Positive prediction)한 사례 중 실제긍정(real True)인 비율을 의미하며 구하는 공식은 다음과 같다.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

그리고 실제로는 긍정인데 부정으로 예측한 경우(TrueNegative), 실제 부정인데 긍정으로 예측한 경우(FalsePositive)는 오류라고 할 수 있다.

4. 연구결과

이클립스를 통해 2016년 1월 1일부터 2016년 8월 22일까지 네이버와 다음에서 제공되는 ‘SW 교육’ 관련 온라인 뉴스데이터를 수집하여 중복 데이터는 제거한 데이터 총수는 1,507개 이다.

데이터 전처리를 위해 R의 KoNLP 패키지를 사용하여 형태소를 분석하였다. 감성사전 구축을 위해 회귀계수를 통한 긍정어와 부정어를 추출한 후 연구 데이터에 맞는 긍정어와 부정어를 수정 보완하였다. 구축한 감성사전을 기반으로 긍정 단어수의 합에서 부정 단어수의 합을 빼서 각 데이터의 감성점수 값으로 부여하였다. 감성점수를 계산하는 R 코드는 <Table 3>과 같다.

<Table 3> R code to calculate sentimental score

```
score.sentiment =
function(sentences,pos.words,neg.words)
{
scores = lapply(sentences ,
function(sentence,pos.words, neg.words)
{
word.list=str_split(sentence,"\\s+")
words = unlist(word.list)
pos.matches = match(words,pos.words)
neg.matches = match(words,neg.words)
pos.matches = lis.na(pos.matches)
neg.matches = lis.na(neg.matches)
score =
sum(pos.matches) - sum(neg.matches)
return(score)
}, pos.words , neg.words )
scores.df =
data.frame(text=sentences , score=scores)
return(scores.df)
}
```

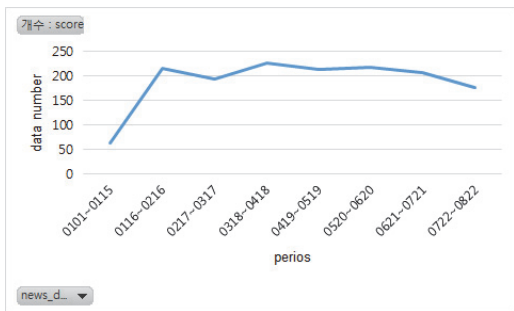
부여된 감성점수의 정확도를 점검하기 위해 사람이 부여한 감성점수를 기반으로 감성사전을 수정하여 다시 실행한 결과, 실제 긍정 데이터가 220개인데 239개를 긍정 데이터로 예측하여 92퍼센트 정확도를 나타냈다.

사람이 부여한 감성점수와 감성사전에 의하여 부여된 감성점수의 부호가 차이가 있다는 것은 긍정을 부정으로, 부정을 긍정으로 예측했다는 의미로 감성사전 기반 감성점수의 오류라고 할 수 있다. 즉, 감성점수의 오류율은 두 개의 감성점수의 부호를 비교하면 되므로 감성점수의 부호가 달라진 경우는 1,507개 중 60개가 나왔다. 예를 들어, ‘않은’, ‘않다’, ‘없이’ 등은 부정어이지만 ‘차질없이’와 같이 긍정으로 쓰여지는 경우가 있어 긍정인데 부정으로 판단하는 경우가 있었다. 즉 실제로는 긍정인데 부정으로 예측했고, 실제로는 부정인데 긍정으로 예측한 건수가 60개로 오류율은 3.98퍼센트라고 할 수 있다. 정밀도와 오류율을 볼 때 감성사전을 사용한 본 연구의 감성분석 결과는 신뢰 할 수 있다고 판단된다.

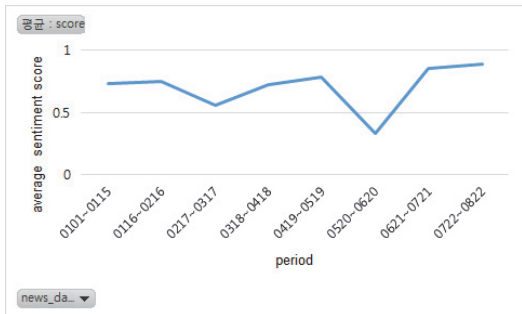
수집된 뉴스 데이터의 월별 갯수 및 감성점수 월별 평균은 <Table 4>, (Fig. 3), (Fig. 4)와 같다. 데이터 갯수는 중복 뉴스데이터를 제거한 건수이며, 감성점수는 1부터 -1 사이의 값으로 변환하여 1값에 가까울수록 긍정적이며 -1에 가까울수록 부정적이며 0에 가까울수록 중립적

<Table 4> The number and average sentimental score of monthly news data

period	data number	average sentimental score
2016.1.1~1.15	63	0.73
2016.1.16~2.16	215	0.75
2016.2.17~3.17	193	0.57
2016.3.18~4.18	226	0.73
2016.4.19~5.19	213	0.78
2016.5.20~6.20	216	0.33
2016.6.21~7.21	205	0.85
2016.7.22~8.22	176	0.89
average	189	0.70
total	1,507	



(Fig. 3) Number of monthly news data



(Fig. 4) Average sentimental score of monthly news data

이라고 할 수 있다. 수집기간동안 SW 교육 관련 뉴스는 월평균 약 189건 발생되었으며, 감성점수 평균은 0.7로 SW 교육 관련 뉴스는 긍정적이었음을 알 수 있었다.

(Fig. 4)를 살펴보면 분석기간 동안 감성점수가 모두 0 이하로 내려가지 않았으며 2016년 2월 17일 ~ 3월 17일 기간과 2016년 5월 20일 ~ 6월 20일 기간을 제외하고는 거의 평균 점수를 유지하고 있었다. 이를 통해 SW 교육에 관하여 부정적인 시각보다는 긍정적인 시각

이 더 많음을 유추할 수 있다.

특히, (Fig. 4)에서 평균 감성점수 0.7보다 낮아진 기간인 2월 17일~3월 17일 기간의 부정적인 의견을 정리하면, SW 교육이 지속될 수 있는 여건 조성이 부족하며, HW 기반보다는 SW 인재 양성에 더욱 노력해야 하고, SW 전문교사가 부족하다. 또한, SW 교육정책을 교육현장에 실행하는데는 차이가 있으며, 인공지능 발달로 인간 존엄성을 중시하는 새로운 윤리가 필요하다는 의견이 있었다. 또한, 5월 20일~6월 20일 기간의 부정적인 의견을 정리하면, SW 교육의 세부방침이 미흡하여 고가의 코딩 학원이 양산되며, SW 개발자 양성문제가 시급하고, 수도권과 지방의 SW 교육 격차가 존재하며, SW 교육이 SW 취업으로 연계될지 여부 등에 관한 의견이 있었다.

분석결과 긍정적인 의견을 종합 정리하면, SW 교육이 강조되면서 SW 선도학교가 확대되고 다양한 창의캠프와 SW 동아리 지원 등으로 재미있는 코딩, 로봇, 드론 등 체험형 SW 교육을 받을 수 있어 좋다. 성공한 창업가들 덕분에 코딩교육이 인기이며, 놀이하듯 자연스럽게 SW 원리를 익혀 문제해결력 향상 등 사고력 신장에 도움이 된다. SW 교육 강화가 미래를 준비하기 위한 포석이고 미래산업과 일자리 창출 전략에 밑거름이 된다 등의 의견이 있었다.

부정적인 의견을 종합 정리하면, SW 교육 의무화에 따른 준비가 아직 미흡하며, SW 교육을 하고 싶어도 SW 교육 수업자료를 구하기가 힘들다. SW 교육 강화뿐만 아니라 SW 개발자들의 처우 개선도 시급하며, SW 교육을 필수화해도 당장 효과를 보기는 시간이 소요되므로 SW 개발자 양성부분도 시급히 이루어져야 한다. SW 교육 의무화에 따른 학부모의 불안감 해소 방안이 필요하며, 프로그래밍 교육보다는 미래 꿈을 생각할 수 있는 희망을 줄 수 있는 SW 교육이 되어야 미래가 밝다고 했다. 그리고 필요할 때마다 주먹구구식으로 과목과 시간을 늘리기보다는 필요할 때 과감한 투자로 속도감 있게 차별화 전략이 필요하다는 등의 의견이 있었다.

5. 결론

인터넷과 스마트폰의 발달로 SNS에 정보들이 증가

하면서 SNS를 활용한 빅데이터 분석 연구가 활발히 진행되고 있다. 특히 사람들의 감성을 분석하기 위한 감성 분석은 글의 주제와 내용을 넘어 감성을 찾아내는 방법으로 기업에서는 소비자들의 생각을 분석하여 제품 개발 및 홍보에 적극적으로 반영하고 있다.

본 논문에서는 2015 개정 교육과정에 의해 2018년부터 초·중·고등학교에 SW 교육이 적용되므로 SW 교육 관련 온라인 뉴스데이터를 수집하여 감성분석을 실시하였다. 그 결과 수집기간동안 SW 교육 관련 뉴스는 월 평균 약 189건 발생되었으며, 감성점수 평균은 0.7로 SW 교육 관련 뉴스는 긍정적이었음을 알 수 있었다. 분석결과 의견을 정리하면, SW 교육의 중요성 및 정책 실행에는 공감하며 긍정적이었으나 구체적인 실행 방법에는 부정적인 시각이 있었다. 즉, SW 교육 환경 및 교육방법 부족 문제, SW 개발자 양성 및 처우개선 문제, 코딩 사교육 증가 문제 등이었다.

본 연구의 한계점은 첫째, 실제로 감성을 긍·부정으로만 나타낼 수는 없다. 감성분석은 문서에 긍정어와 부정어 포함여부를 점검하여 종합적으로 긍정적인지 부정적인지 또는 중립인지를 구분한다. 이때 감성사전의 긍정어와 부정어의 출현빈도로 감성을 측정하는 방법을 일반적으로 사용하고 있으나 감성을 긍정, 부정으로만 나타낼 수 없기 때문에 제한점이 존재한다고 할 수 있다. 둘째, 감성사전을 일반화하기 어렵다. 현재 한글 감성사전에 대한 연구가 부족하여 범용 한글 감성사전을 적용하기가 어려운 실정이다. 본 논문에서도 수집데이터에 맞는 감성사전을 구축하여 적용하였으므로 일반화하기 어려우며, 이러한 한계점을 해소하기 위해 한글 범용 감성사전 구축이 시급하다고 하겠다. 셋째, 데이터의 부족이다. 본 논문에서는 약 8개월 기간의 온라인 뉴스데이터의 헤드라인과 요약정보를 사용하여 분석하였다. 뉴스 헤드라인은 본문보다 사실과 의견이 더 정확하게 발현될 수 있지만 뉴스데이터의 전문에 다른 감성정보가 전달될 수도 있다는 한계점을 가지고 있다.

추후 뉴스데이터 뿐만 아니라 다양한 SNS 정보들을 수집하여 SW 교육에 대한 사람들의 의견을 감성분석하여 SW 교육 정책수립 및 실행에 피드백으로 활용되기를 기대한다.

참고문헌

- [1] An, Jungkook, Kim, HeeWoong(2015), Building a Korean Sentiment Lexicon Using Collective Intelligence, *Journal of Intelligence and Information Systems*, 21(2), 49~67.
- [2] Bollen, J., A. Pepe, and H. Mao(2009), "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," arXiv preprint arXiv:0911.1583.
- [3] Cha, EunJeong, Hong, TaeHo(2016), Stock Index Prediction Using SVM and News Sentimental Analysis, *Proceedings of the Korean Society of Management Information Systems Conference*, 2016(6).
- [4] Cho, S. Y., Kim, H. K., Kim, B. and Kim, H. W.(2014), "Predicting Movie Revenue by Online Review Mining: Using the Opening Week Online Review," *Information Systems Review*, 16(3), 111~132.
- [5] Choi Sukjae, Lee Jaewoong, Kwon Ohbyung(2015), A Morphological Analysis Method of Predicting Place-Event Performance by Online News Titles, *The Journal of Society for e-Business Studies*, 21(1), 15~32.
- [6] Jang, J.-Y.(2009), "A Sentiment Analysis Algorithm for Automatic Product Reviews Classification in On-Line Shopping Mall," *The Journal of Society for e-Business Studies*, 14(4), 19~33.
- [7] Jin W., H. H. Ho and R. K. Srihari(2009), OpinionMiner: A Novel Machine Learning System for Web Opinion Mining and Extraction, *KDD Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* 1195~1204.
- [8] Jo H. J., Seo, J. H., and Choi, J. T.(2015), OAR Algorithm Technology Based on Opinion Mining Utilizing Stock News Contents, *Journal of Korean Institute of Information Technology*, 13(2), 111~119.

- [9] Jwa, BoKyung, Paek, HyeJin, Seol, Pil Kyo(2014), A Content Analysis of Online News and Comments about Anti-smoking Policy, *Journal of Public Relations*, 18(3).
- [10] Khan, F. H., S. Bashir, and U. Qamar(2014), "TOM:Twitter opinion mining framework using hybrid classification scheme," *Decision Support Systems*, 57, 245~257.
- [11] Kim J. H., Oh, Y. J. and Chae, S. H.(2015), The Construction of a Domain-Specific Sentiment Dictionary Using Graph-based Semisupervised Learning Method, *Korean Journal of the Science of Emotion and Sensibility*, 18(4), 97~104.
- [12] Kim, Jungho, Chae, Sohoan(2014), Automatic Construction of Korean Polarity Dictionary using Graph-based Semi-supervised Learning, *Proceedings of the Korean Society for Internet Information Conference*, 2014(5).
- [13] Kim, Yoosin, Kim, Namgyu, Jeong, SeongRyoul(2012), Stock-Index Invest Model Using News Big Data Opinion Mining, *Journal of Intelligence and Information Systems*, 18(2).
- [14] Lee, SangHoon, Choi, Jung, Kim, JongWoo(2016), Sentiment analysis on movie review through building modified sentiment dictionary, *Journal of Intelligence and Information Systems* 22(2), 97~113.
- [15] Moon, Kwangsu, Kim, Seul, Oah, Shezeen(2013). An Effect of the Valence of Best Reply on the Conformity of General Reply, *Journal of the Korea Contents Association*, 13(12), 201~211.
- [16] Pang, B., and L. Lee(2008), Opinion mining and sentiment analysis, *Foundations and trends in information retrieval*, 2(1-2), 1~135.
- [17] Park, SungGeon, Won, GyuSik, Lee, SooWon(2015), Web News Comment-based Sentiment Analysis of the South Korean National Team Members in the 2014 Brazil World Cup, *Korean Journal of Sport Management*, 20(2).
- [18] Song S. I., Lee, D. J. and Lee, S. G.(2010), *Identifying Sentiment Polarity of Korean Vocabulary Using PMI*, *Proceedings of the Korean Information Science Society Conference*, 37(1), 260~265.
- [19] Sung, JunMo(2015), A study on the convergence of SNS and storytelling emotional marketing, Master's dissertation, Graduate school in Hanyang University.
- [20] Turney P. D. and M.L. Littman(2002), Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus, National Research Council, Institute for Information Technology, Technical Report, ERB-1094.
- [21] Yu E. J., Kim, Y. S., Kim, N. Y. and Jeong, S. R.(2013), Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary, *Journal of Intelligent Information Systems*, 19(1), 95~10.

저자소개



박 선 주

1995 전남대학교 전산통계학과 (이학박사)

2003 George Mason University 객원교수

1996~현재 광주교육대학교 컴퓨터교육과 교수

관심분야 : 컴퓨터교육, SW교육, 앱개발, 빅데이터

E-mail : sjpark@gnue.ac.kr