JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 19, No. 1 : 141~148 Jan. 2015

온라인 마케팅 전략을 위한 SNS와 Web기반 BDAS(Big data Data Analysis Scheme) 설계

정이나[1] · 이병관[1*] · 박석규[2]

An SNS and Web based BDAS design for On-Line Marketing Strategy

Yi-Na Jeong[1] · Byung-Kwan Lee[1*] · Seok-Gyu Park[2]

[1*]Department of Computer Engineering, Catholic Kwandong University, Gangneung 210-701, Korea
[2]Department of Computer & Internet Technique, Gangwon Provincial College, Gangneung 210-804, Korea

요 약

본 논문은 SNS와 Web에서 실시간으로 공유되는 정보를 추출하고, 추출한 데이터를 신속하게 분석하여 고객이 무엇을 원하는 지를 분석해서 온라인 마케팅 전략을 효율적으로 만드는 SNS와 Web기반 BDAS(Big data Data Analysis Scheme)을 제안한다. 제안하는 BDAS는 첫째, SNS와 Web에서 공유되는 데이터를 수집하고, 둘째, 수집된 데이터의 의미를 긍정과 부정으로 분석하여 그 결과를 시각화하여 제공한다. 그 결과, BDAS는 공유되는 SNS와 Web 데이터에 대한 의미를 판단하는데 있어서 평균 90%의 정확성을 보장한다. 따라서 본 논문에서 제안하는 BDAS를 이용하여 소비자의 성향을 정확하게 판단할 수 있으므로 온라인 마케팅에 보다 효율적으로 활용할 수 있을 것이다.

ABSTRACT

This paper proposes the BDAS(Big Data analysis Scheme) design that extracts the real time shared information from SNS and Web, analyzes the extracted data rapidly for customers, and makes an on-line marketing strategy efficiently. First, the BDAS collects the data shared in SNS and Web. Second, it provides the result of visualization by analyzing the semantics of the collected data as positive or negative. Therefore, because the BDAS ensures an average 90% accuracy in judging the semantics about the shared SNA and Web data, it can judge customer's propensity accurately and be used for on-line marketing strategy efficiently.

키워드 : 마케팅, 오피니언 마이닝, 시멘틱 추론, SNS, 스톰

Key word : Marketing, Opinion Mining, semantic analysis, SNS, Storm

접수일자 : 2014. 10. 28 심사완료일자 : 2014. 11. 21 게재확정일자 : 2014. 12. 05

* **Corresponding Author** Byung-Kwan Lee(E-mail:bklee@cku.ac.kr, Tel:+82-01-4441-3373)
Department of Computer Engineering, Catholic Kwandong University, Gangneung 210-701, Korea

Open Access http://dx.doi.org/10.6109/jkiice.2015.19.1.141 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. 서 론

인터넷 및 이동통신 매체들의 발전으로 SNS와 Web Log의 활용이 다양해짐에 따라 사람들은 Face Book, Twitter, Blog, E-News등의 다양한 가상공간에서 실시간으로 자신의 의견 혹은 다양한 정보를 공유하고 있다. 이러한 정보공유의 특징을 이용하여 발전한 마케팅 전략을 온라인 마케팅이라 하는데, 많은 기업에서는 파급력이 높은 이 온라인 마케팅에 많은 노력을 기울이고 있다. 하지만 이러한 파급력은 다른 의미로 언론의 불확실한 추측성 기사 혹은 악성 댓글 등과 같이 기업에 부정적인 이미지를 확산시켜 기업의 이미지에 큰 타격을 입힐 수 있다.

본 논문에서는 SNS 및 Web 마케팅을 보다 효율적으로 활용하고 기업의 이미지에 타격을 입힐 수 있는 데이터를 빠르게 파악하고 대응할 수 있는 “온라인 마케팅 전략을 위한 SNS와 Web기반 BDAS(Big data Data Analysis Scheme)”을 제안한다. 제안하는 BDAS는 SNS와 같이 실시간 공유되는 데이터를 빠르게 분석하여 의미 분석 결과를 시각화시켜 제공한다. 그 결과 부정적인 이미지를 제공하는 데이터에 신속하게 대응하거나 분석결과를 온라인 마케팅 전략에 활용할 수 있도록 정보를 제공하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 관련연구를 설명하고, 3장에서는 본 논문에서 제안하는 온라인 마케팅 전략을 위한 BDAS에 대해 설명한다. 그리고 4장에서 온라인 마케팅 전략을 위한 BDAS의 실험 결과를 설명하고, 5장에서 결론을 맺는다.

Ⅱ. 관련연구

2.1. 오피니언 마이닝

오피니언 마이닝은 사용자가 입력한 문서로부터 문서가 나타내는 의미와 감정을 찾아내기 위해 어휘를 표현할 수 있는 다양한 패턴을 이용해 사용자의 의견이 긍정인지 혹은 부정인지를 찾아내는 마이닝 기술이다[1]. 오피니언 마이닝은 다음과 같은 절차로 이루어진다.

첫째, 특징추출단계: 유용한 정보라고 판단되는 여러 특징과 그 특징의 의미에 대한 어휘정보를 추출한다. 둘째, 판단 및 분석 단계: 추출된 특징과 의견을 나타내는 어휘가 해당 문서에서 어떠한 의미로 사용되었는지를 판단하고 분류한다. 셋째, 요약 및 표현 단계: 의견 정보들을 요약하여 해당 문서의 전체 정보를 효율적으로 사용자에게 전달한다.

오피니언 마이닝은 위와 같은 과정을 거쳐 SNS, Blog, E-mail, News 등 다양한 매체의 정보를 수집해 불특정 다수 사용자들의 의견, 즉 비정형 데이터로부터 보다 가치 있는 정보를 추출한다[2,3]. 특히, SNS는 오피니언 마이닝이 분석하기 가장 적합한 데이터이다. 오피니언 마이닝은 수많은 사람들의 의견을 종합하고 판단하기 위해 의미방향을 분류하는 연구와 언어적 자원을 구축하는 연구로 분류되어 발전되고 있다.

본 논문에서는 이러한 오피니언 마이닝을 이용하여 SNS와 Web상에 기업에 대한 데이터를 정확하게 분석하여 사용자들의 의견을 실시간으로 온라인 마케팅 전략에 반영할 수 있도록 기업에 정보를 제공하고자 한다.

2.2. Storm

Storm은 수집되는 빅 데이터를 실시간으로 분석하는 분산 시스템이다. Hadoop과 동일한 오픈 소스로 제공되고 있지만 Hadoop이 빅데이터 분석에 중점을 두어 배치처리에 초점이 맞춰져 있는 반면, 실시간 데이터 분석 정보를 제공하지 못하는 단점이 있다[4]. 이는 실시간으로 업데이트 되는 Web정보 혹은 SNS 등의 정보들을 활용하는데 있어 큰 문제점을 초래할 수 있다. 예를 들어, 상품을 판매하는 업종의 경우 상품을 비판하는 글이 업로드 되었을 때 실시간 분석이 제공되지 않는다면 이미 여러 사용자들이 비판을 보고 상품에 대한 좋지 않은 평가로 기억할 것이다[5]. 이러한 문제점을 해결하여 빅 데이터를 보다 효율적으로 사용하기 위한 실시간 데이터 정보 처리 기법이 Storm이다[6]. Storm은 토폴로지, 스트림, 스파우트, 볼트로 구성된다. 스트림은 Tuple의 흐름으로 분산 환경에서 신뢰성 있게 다른 스트림으로 전환할 수 있는 기능을 제공하며, 스파우트는 외부로부터 튜플을 읽고 토폴로지로 튜플을 생성해 스트림을 생성하는 과정을 반복하며 볼트는 토폴로지의 모든 처리 작업 즉 필터링, 조인, 데이터베이스 적재 등의 작업을 진행한다. Storm은 다양한 패턴으로 응용이 가능하며 다중 언어를 지원한다는 점에서 매우 활용도가 높다[7-9].

본 논문에서는 Storm을 이용하여 실시간으로 Web혹은 SNS에 업데이트 되는 기업의 관련 데이터를 수집 및

분석하여 마케팅을 목적으로 활용하는 BDAS 설계를 제안한다.

Ⅲ. BDAS 설계

본 논문에서는 SNS 및 Web 마케팅을 보다 효율적으로 활용하고 기업의 이미지에 타격을 입힐 수 있는 데이터를 빠르게 파악하고 대응할 수 있는 “SNS와 Web 기반 BDAS(Big data Data Analysis Scheme)”을 제안한다.

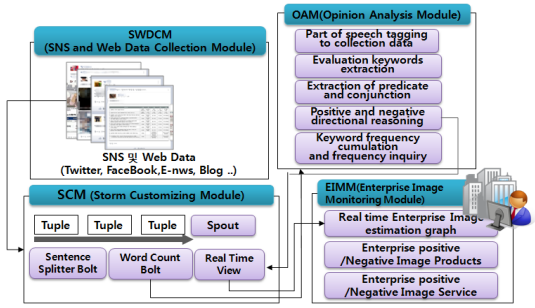

그림 1. BDAS의 전체 구성
Fig. 1 The structure of BDAS

그림 1은 제안하는 BDAS의 구성요소와 전체적인 데이터 흐름을 설명한 것이다. BDAS는 SNS와 Web 데이터를 수집하는 SWDCM(SNS and Web Data Collection Module), 실시간으로 수집된 SNS과 Web 데이터의 단어를 분류하고 카운트 하는 SCM(Storm Customizing Module), 그리고 SCM에서 분류되어 카운트된 단어들을 Opinion Mining 기법을 이용하여 긍정적 혹은 부정적 평가로 분류하여 그 결과를 제공해주는 OAM(Opinion Analysis Module)로 구성된다.

3.1. SWDCM 설계

본 논문에서 제안하는 SWDCM(SNS and Web Data Collection Module)은 특정한 관심이나 활동을 공유하는 사람들 사이의 관계망을 구축해 주는 온라인 서비스인 SNS, E-news, 그리고 블로그와 같은 웹 데이터를 통해 기업에 대한 이미지를 추출할 수 있도록 설계한다.

그림 2와 같이 SWDCM은 FaceBook 혹은 Twitter와 같은 SNS에 실시간으로 업데이트 되는 데이터 가운데 그룹명을 의미하는 Keyword 혹은 그룹의 제품명을 의미하는 Keyword가 포함된 데이터를 추출하여 SCM에 전송한다. 즉, SWDCM이 수집하는 정보는 Keyword, 등록일자 및 시간, 작성자, 내용, 첨부파일로 구성된다.

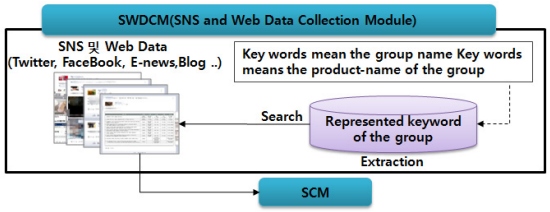

그림 2. SWDCM의 구조
Fig. 2 The component of SWDCM

3.2. SCM 설계

본 논문에서는 초단위로 빠르게 번지는 기업에 대한 악성 평가를 분류하기 위하여 SNS에 기업에 대한 데이터 혹은 기업 제품에 대한 데이터를 수집하여 토폴로지를 통해 분석하는 SCM(Storm Customizing Module)을 제안한다. 제안하는 SCM은 트위터에서 개발한 오픈소스인 Storm에 기반을 두고 있다.

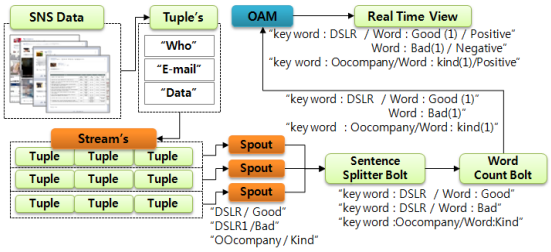

그림 3. SCM의 구조
Fig. 3 The component of SCM

그림 3과 같이 SCM의 토폴로지는 수집데이터를 “작성자”, “E-mail”, “Data” 등 정렬된 목록으로 만들어진 튜플, 이러한 튜플을 연속적으로 수집하는 스트림, 그리고 스트림을 실제 작업을 수행하는 Sentence Splitter 볼트와 Word Count 볼트에게 전달하는 Spouts로 구성된다. 여기서 Sentence Splitter 볼트는 각 키워드 별로 가장 많이 출현한 단어를 카운트하기 위하여 문장에서 단어를 분리하고, Word Count 볼트는 분리된 단어를 카운트하는 역할을 담당한다. 제안하는 SCM의 동작 단계는 다음과 같다.

[1단계] SWDCM 모듈로부터 실시간으로 수집되는 SNS의 데이터 가운데 기업의 제품 혹은 기업을 대표하는 단어를 키워드로 하여 관련된 데이터를 추출하여 수집한다.
[2단계] 수집된 데이터를 튜플의 형태, 즉 정렬된 목록으로 재구성한다. SCM은 작성자, SNS의 ID에 해당하는 E-mail, 그리고 SNS에 작성된 Data로 구분하여 {Who,E-mail,Data}형태의 튜플을 생성한다.
[3단계] SCM은 실시간으로 생성되는 튜플 들로 구성된 Stream을 Spout을 통해 Sentence Splitter Bolt에게 전달한다.
[4단계] Sentence Splitter Bolt는 전송받은 튜플의 Data를 형태소 분석기를 통하여 단어들만을 추출하고 기업에 관련된 Keyword를 기준으로 어떤 단어가 존재하는지를 판단한다. 그리고 그 결과는 Word Count Bolt에게 전달한다.
[5단계] Word Count Bolt는 Sentence Splitter Bolt로부터 전달받은 각 Keyword별 Word 데이터를 취합하고, 각 KeyWord별로 동일한 단어를 카운트한 후, 이 결과를 긍정적인 데이터 혹은 부정적인 데이터로 분류 및 평가하고 그 결과를 기업에 제공하기 위해 OAM으로 전송한다.

3.3. OAM 설계

본 논문에서는 SCM으로부터 전달받은 SNS와 Web 데이터에 대한 문장의 형태소를 분석하여 사용자의 평가의견을 분석하고, Keyword별 카운트 결과를 이용하여 긍정 혹은 부정에 대한 평가의견의 출현 빈도를 파악하여 그 결과를 제공하는 OAM(Opinion Analysis Module)을 제안한다.

3.3.1. 평가 주제어 추출 및 형태소 분석기를 통한 품사 Tagging 절차

그림 4는 OAM이 형태소 분석기를 통해 주제어를 추출하고 태깅을 수행하는 흐름을 설명한 것이며, 그 단계별 절차는 다음과 같다.
[1단계] OAM은 SCM으로부터 SNS 또는 Web 데이터인 “DSLR-01은 정말 깨끗하게 찍힌다.”를 Value 변수에 저장한다.
[2단계] Value 변수에 저장된 문장을 오픈 소스인 형태소분석기를 통해 품사를 Tagging 한다.
[3단계] 분석된 형태소에 주격조사 “은”, “는”, “이”, “가”를 탐지하여 주격조사 앞에 명사를 임시 주제어로 선정한다. 단, 여기서 임시 주제어가 될 수 있는 조건은 SCM에서 Keyword로 출현한 단어여야 한다. 이는 SCM의 Keyword는 수집한 데이터 가운데 기업의 제품 혹은 기업을 대표하는 단어로 선정되기 때문이다.

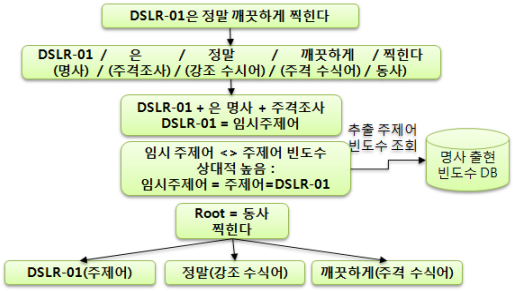

그림 4. 주제어 탐색 및 태깅 절차
Fig. 4 The keyword search and tagging process

[4단계] 임시 주제어의 명사 출현 빈도수가 상대적으로 높은 빈도수를 갖는 임시 주제어를 주제어로 선정한다.
[5단계] Tagging된 형태소 중 동사를 기준으로 트리를 구성 한다.

OAM은 위의 단계를 거쳐 트리화된 데이터를 이용하여 긍정 및 부정에 대한 의미방향 추출과정을 수행한다.

3.3.2 긍정 부정 판단을 위한 의미 방향 추출 설계

OAM은 긍정, 부정 판단을 위해 3.3.1절에서 생성한 형태소 트리를 기준으로 하위 계층의 서술어의 극성값을 이용하여 의미방향을 추론한다. 그림 4에 형성된 트리를 기준으로 OAM이 의미방향을 추론하는 과정은 다음과 같다.
[1단계] OAM은 트리의 하위 계층 가운데 주격 서술어인 “깨끗하다”를 선택하고, 데이터베이스를 검색하여 “깨끗하다”라는 표현이 긍정적으로 출현한 횟수와 부정적으로 출현한 횟수와 긍정, 부정 서술어의 총 출현 빈도수를 추출한다. 그리고 OAM은 식(1)을 이용하여 극성값을 산출한다.

$$\text{극성값} = \frac{\text{긍정표현빈도} - \text{부정표현빈도}}{\text{총긍정표현빈도수} + \text{총부정표현빈도수}} \quad (1)$$

"깨끗하다"를 기준으로 긍정 표현 빈도수가 400회이고 부정표현 빈도수가 60회 그리고 총 긍정표현 부정표현의 빈도수를 합산한 결과 38000이라는 값이 추출되었을 경우 "깨끗하다"의 극성값은 약 0.00895이 된다.

[2단계] OAM은 1단계에서 산출한 극성값을 양수와 음수로 구분하여 양수인 경우 극성값을 +1로 음수인 경우 -1로 대체한다.

[3단계] OAM은 트리의 하위 계층 가운데 주격 서술어 앞에 강조 서술어가 있는 경우 서술어의 극성값을 2배 증가시킨다. 그림 5의 경우, 트리의 하위 계층 주격 서술어 "깨끗하다" 앞에 강조 서술어 "정말" 이라는 단어가 있기 때문에 "깨끗하다"의 극성값은 {+1×2}로 총 +2의 극성값을 갖는다.

[4단계] 그림 4의 문장과 다르게 만약 역정 접속사인 "하지만", "그러나", "그렇지만"과 같은 부정적인 접속사가 붙은 경우에는 기업에 대한 긍정적 평가와 부정적 평가가 동시에 존재한다. 이때, OAM은 그림 5와 같이 접속사를 판단하여 각각의 극성값을 산출한다. 따라서 OAM은 기업 평가 문장에 존재하는 긍정적인 평가와 부정적인 평가를 효율적으로 판단할 수 있다.

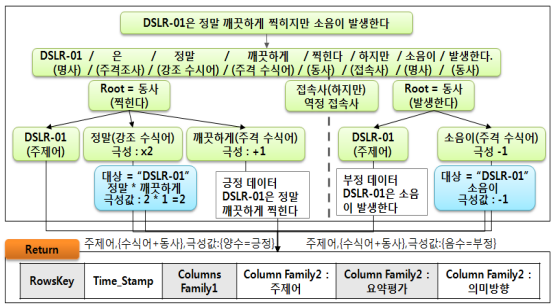

그림 5. 접속사가 존재하는 경우의 의미방향 추출
Fig. 5 The extraction example of the semantic direction where a conjunction exists

3.3.3. EIMM(Enterprise Image Monitoring Module)

본 논문의 기업 이미지 모니터링 모듈은 OAM으로부터 긍정 혹은 부정적 평가로 분류된 데이터와 SCM에서 파악한 긍정 혹은 부정적 평가의 출현 빈도수를 전달받아 이를 시각화 하는 역할을 수행한다.

기업 이미지 모니터링 모듈은 "DSLR"에 대해 검색하면 해당 Keyword가 포함된 모든 데이터의 긍정 혹은 부정 데이터 출현 횟수를 양성 혹은 음성으로 그래프화 하여 제공하며, 혹은 모든 상품 혹은 기업의 Keyword의 긍정 혹은 부정의 평균값으로 그래프화 하여 시각화를 제공한다. 또한 실시간 분석을 통해 긍정 혹은 부정적인 데이터가 지속하여 전달되는 제품 혹은 기업 서비스에 대한 평가를 파악할 수 있기 때문에 기업 이미지 마케팅에 효율적으로 활용할 수 있다.

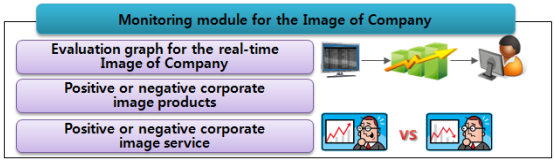

그림 6. EIMM의 전체 구조
Fig. 6 The component of EIMM

Ⅳ. 성능분석

본 논문에서 제안한 온라인 마케팅 전략을 위한 SNS와 Web기반 BDAS(Big data Data Analysis Scheme)에 대한 실험데이터는 Web 사이트의 검색을 통해 블로그나 카페 혹은 개인 웹페이지에 업로드 되어있는 S전자에 대한 댓글 혹은 평가 데이터 10건을 이용하였다.

BDAS는 Visual Basic 6.0으로 구현하였으며, KAIST의 한나눔과 KOMORAN과 같은 어휘 분석기를 이용하여 데이터의 텍스트를 품사 태깅 과정을 수행하였고, 이때 한음절 어휘는 제거하였다.

표 1. 실험에 사용한 입력 데이터
Table. 1 Input data used for the experiment

Identifier	Data
lov**	안녕하세요 너무 화가나서 글을 작성합니다 동생이,중학생인데 제가 핸드폰**을 해준지 한달도 안되어 폰안 액정이 파손이 되었습니다 ..생략... 파손 액정 달라고 했더니 못준다고 할인해서 액정은 못준다고 하더라구요 정말 말도 너무 기본나쁘게 얘기하고 진짜, **전자 실망스러웠습니다, 다음부터 **전자 안쓸려고 맘을 먹었습니다.
pos***	Re : 삼성전자서비스 불만 : **전자서비스...고객에게 신뢰를 잃어가네요...
hea****	삼성전자 불만접수 어디서 해야 하나요? 저는 3주전 인터넷으로 **전자 23인치 모니터와 타사 본체를 함께 구입한 고객입니다.처음 제품 사용후 모니터에서 물결이 치는 불량으로 인해 타사본체(**)의 문제로 여겨타사에 AS를 요청하여 몇번 방문하여 본체 메인보드 및 그래픽카드등을 바꿔주셨는데 그래도 해결이 안되어 이틀후 제품교환을 해주셨습니다. 제품교환후 확인결과 모니터의 불량으로 확인되어 **전자에 AS신청하여 당기사님이 확인결과 모니터 불량으로 확인되었습니다. 저는 개인적으로 직장생활 퇴근시간이 일정치 않아 그간 AS방문 시간등을 맞추기위해 스트레스를 많이 받았으며 좋아하는 컴퓨터 사용도 3주간 못했습니다. ..생략.. 그간 불량으로 고생하고 이
***광	해외(***) **지역 **전자 서비스 센타 불만 있어요 제가 *** 핸드폰2 **지역에서 샀는데 액정값이 한국의 3배를 달라구 하네요 도무지 이해할 수 없어요 직원들이 사기 치는건지 진짜 3배 차이 나는지 센타내에서 한국인도 없고
Jang*****	** 망해야 정신 차려
최**	**전자 보안AP제품 심플하니 이쁘다!
pl**	**전자 모바일 기프트 형사 영화 해택 짱이네요!
news	진정 LCD 대비 LED 화질을 차이나게 느낄만한 경우가 얼마나 있겠어? 그 몇번 안되는 경우때문에 백만원 가까운 돈을 더 내라구?
news	그돈이면 최신 스마트폰 한두대는 덤으로 사겠다 라는 생각으로 아예 고려를 안했었는데요, 최근 LED TV 를 체험하면서 또 정보를 탐색하다보니 조금씩 생각이 바뀌었습니다.
news	물론 단순히 LED TV 의 화질이 더 좋아서는 절대 아니구요, 좋다는건 알겠는데 지금은 너무 비싸다는 생각에 변화가 생겼는데요 그 이유는 결코 LED TV 가 비싼게 아닐수 있겠다 라는 생각이 들어서였습니다. 바로 LED TV 가 LCD

실험에서는 표 1의 데이터와 표 2의 사전을 이용하였고, 실제 데이터에 대한 극성값은 입력된 데이터가 긍정적인 데이터인 경우에는 1.0이고, 부정적인 데이터인 경우에는 -1.0로 설정하였다. 그리고 OAM이 산출한 극성값이 부등호와 실제 데이터의 극성 값인 부등호를 비교하여 OAM의 정확도를 평가 하였다.

표 2. 실험에 사용한 긍정 및 부정 사전
Table. 2 Positive and negative dictionary used for the experiment

Positive	할인, 해결, 좋아, 심플, 예쁘다, 짱이다, 혜택, 최신, 변화, 효과, 절약, 좋았다, 행복, 친절 ,튼튼, 웃는, 따뜻
negative	어렵다, 불친절, 불량, 교환, 파손, 불만, 화나, 안 된다, 못준다, 나쁘다, 실망, 안써, 잃다, 사기, 망해, 비싸, 그돈

실제데이터와 비교하여 보았을 때, 첫 번째 데이터는 총 32개의 긍정 및 부정 사전 가운데 긍정 데이터의 출현 빈도에 따른 결과 수치인 1과 부정 데이터의 출현 빈도에 따른 결과 수치인 10을 통해 1-10/32로 최종 극성값이 －0.291인 음수 값이 나온다. 따라서 실제 데이터 또한 음수 값(-1)으로 OAM이 산출한 극성 값의 부등호와 일치한다는 것을 알 수 있다. 또한, 실제로 긍정적인 데이터인 10번째 데이터의 경우 긍정적인 명사 출현과 그에 따른 결과값 6과 부정적인 명사 출현에 따른 결과 값 2를 통해 6-2/32로 최종 극성값 0.0631이 산출되며, 양수 값으로 실제 긍정적인 데이터와 결과가 일치한다.

표 3. OAM과 실제 극성 값의 비교
Table. 3 the comparison of OAM result with actual polarity value

The polarity value for the result of the OAM	Actual polarity value (Extreme positiveness /Extreme negative)	Result
-0.281	-1(negative)	Correct
-0.031	-1(negative)	Match
-0.281	-1(negative)	Match
-0.125	-1(negative)	Match
-0.031	-1(negative)	Match
0.063	1 (positive)	Match
0.063	1 (positive)	Match
0.000	-1(negative)	Incorrect
0.094	1 (positive)	Match
0.125	1 (positive)	Match

8번째의 데이터의 경우 OAM에서 인식하지 못하여 극성값을 0으로 판단하였다. 그 이유는 8번째 입력된 데이터가 “진정 LCD 대비 LED 화질을 차이나게 느낄 만한 경우가 얼마나 있겠어? 그 몇 번 안 되는 경우 때문에 백만원 가까운 돈을 더 내라구?”으로 부정적인 데이터이지만, 실험에 사용한 표 1의 사전에 존재하는 명사가 없기 때문에 해당 문장을 분석하지 못하였다. 하지만, 실제로 의미 분석을 위한 사전을 작성할 때에는 기업 혹은 제품에 대해 수집한 데이터를 어휘 분석기를 통해 형태소를 분석하고, 명사만을 추출하여 긍정 부정 사전을 생성하기 때문에 이와 같은 오류는 발생할 확률이 매우 적다고 할 수 있다.

그림 7의 데이터와 표 1의 사전을 이용한 실험결과는 총 10개의 입력 데이터 가운데 90%의 분석이 실제 의미 분석과 일치하였다. 그 결과를 표 3과 그림 7에서 설명하고 있다.

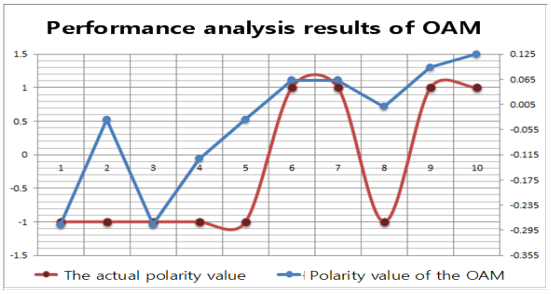

그림 7. OAM의 성능분석 결과
Fig. 7 the performance analysis of the OAM

그림 8은 실험 데이터를 순차적으로 증가하여 실험 결과를 설명한 것이다. 데이터의 수가 증가할 때, 실험에 사용하는 긍정 및 부정 사전의 단어들도 증가하기 때문에 그림 8과 같이 분석의 일치도가 증가하는 것으로 나타났다.

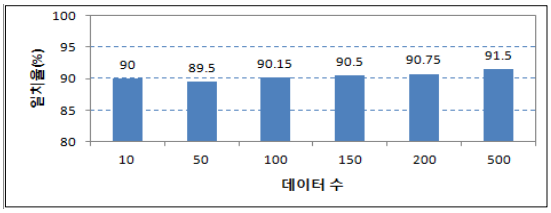

그림 8. 데이터 수에 따른 성능분석 결과
Fig. 8 the performance analysis according to data

Ⅴ. 결 론

본 논문에서는 SNS 및 Web 마케팅을 보다 효율적으로 활용하고 기업의 이미지에 타격을 입힐 수 있는 데이터를 빠르게 파악하고 대응할 수 있는 "온라인 마케팅 전략을 위한 SNS와 Web기반 BDAS(Big data Data Analysis Scheme)"을 제안하였다. 제안하는 BDAS의 효율성은 다음과 같다.

첫째, BDAS는 SNS와 같이 실시간 공유되는 데이터를 빠르게 분석하여 의미 분석 결과를 시각화시켜 제공할 수 있다. 둘째, BDAS는 기업과 기업제품에 대한 부정적인 이미지를 제공하는 데이터에 신속하게 대응할 수 있다. 셋째, BDAS는 소비자의 평가를 분석한 결과를 온라인 마케팅 자료로 활용할 수 있도록 정보를 제공할 수 있다. 그 결과, BDAS는 OAM을 통해 정확하게 판단한 소비자의 긍정적인 평가와 부정적인 평가를 온라인 마케팅 전략에 보다 효율적으로 활용할 수 있을 것이다.

REFERENCES

[1] K.H. Kim, "Design and Implementation of Opinion Mining System based on Association Model," *KIICE,* vol.14 no.1, pp.133-139, Jan. 2011.

[2] Gamgarn Somprasertsri and Pattarachai Lalitrojwong, "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization," *Journal of Universal Computer Science*, vol.16, no.6, pp.938-955, Jun 2010.

[3] J. S. Song and S. W. Lee, "Automatic Construction of Positive/ Negative Feature-Predicate Dictionary for Polarity Classification of Product Reviews," *KIISE: Software and Application*, vol.38, no.3, pp.157-168, Mar. 2011.

[4] Babcock et al. "Models and issues in data stream systems," *in Proceedings of the twenty-first ACM SIGMOD- SIGACT-SIGART symposium on Principles of database systems*, pp.1-16, 2002.

[5] I. S. Hwang, "Data Distribution and Task Scheduling for Improving MapReduce Performance," M.S. dissertation, University of Inha, Feb. 2011.

[6] Data, data everywhere. Available at: https://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf

[7] L. Golab and M. Tamer Őzsu, "Issues in data stream management," *SIGMOD Record*, vol.32, no.2 pp.5-14, June. 2003.

[8] Hellerstein and Stonebraker. Chapter 10. Stream-Based Data Management, Readings in Database Systems, Fourth Edition, MIT Press.

[9] Hyde. "Data in Flight," *Communications of the ACM*, vol.53(1) pp.48-52, Jan. 2010.

정이나(Yi-Na Jeong)

2011년 2월 가톨릭관동대학교 컴퓨터학과 공학사
2011년 3월 ~ 현재 가톨릭관동대학교 컴퓨터공학과 박사과정
※관심분야 : 빅데이터, IoT

이병관(Byung-Kwan Lee)

1979년 2월 부산대학교 기계설계학과 공학학사
1986년 2월 중앙대학교 전자계산공학과 공학석사
1990년 2월 중앙대학교 전자계산공학과 공학박사
1988년 3월 ~ 현재 가톨릭관동대학교 공과대학 컴퓨터공학과 교수
※관심분야 :네트워크 보안, 빅데이터, IoT

박석규(Seok-Gyu Park)

2005년 2월 경상대학교 컴퓨터과학과 공학박사
2001년 3월 ~ 현재 강원도립대학 컴퓨터인터넷과 부교수
※관심분야 : 소프트웨어신뢰성, 시스템분석