

종속성 네트워크 기반의 온라인 고객리뷰 분석시스템 설계 및 구현

Designn and Implementation Online Customer Reviews Analysis System based on Dependency Network Model

김근형
제주대학교 경영정보학과

Keun-Hyung Kim(khkim@jejunu.ac.kr)

요약

특정 제품이나 서비스에 대한 네티즌의 의견들은 고객뿐만 아니라 기업 입장에서 마케팅이나 경영전략을 수립하기 위한 중요한 자료가 될 수 있기 때문에 온라인 고객리뷰를 분석하는 것은 매우 중요하다. 본 논문에서는 온라인 고객리뷰를 분석하기 위한 도구인 종속성 네트워크 모델을 제안하였고 종속성네트워크 모델을 기반으로 한 분석시스템을 설계하고 개발하였다. 종속성 네트워크모델은 고객리뷰 내의 주관적 문장과 객관적 문장을 분석대상으로 하며, 명사들 사이의 상대적 중요성과 연관성을 나타낼 수 있다. 시스템 구현 결과, 종속성네트워크 모델은 오피니언마이닝 기술에 의하여 도출할 수 없는 새로운 정보 즉, 추출된 특징들 사이의 상대적 중요성 및 연관관계 등을 추출할 수 있음을 알 수 있었다.

■ 중심어 : | 온라인고객리뷰 | 종속성네트워크 | 빈발도 | 관련도 | 오피니언마이닝 |

Abstract

It is very important to analyze online customer reviews, which are small documents of writing opinions or experiences about products or services, for both customers and companies because the customers can get good informations and the companies can establish good marketing strategies. In this paper, we did not propose only dependency network model which is tool for analyzing online customer reviews, but also designed and implemented the system based on the dependency network model. The dependency network model analyzes both subjective and objective sentences, so that it can represent relative importance and relationship between the nouns in the sentences. In the result of implementing, we recognized that relative importance and relationship between the features of products or services, which can not be mined by opinion mining, can be represented by the dependency network model.

■ keyword : | Online Customer Review | Dependency Network | Frequency | Relationship | Opinion Mining |

I. 서론

인터넷 상에서의 참여와 공유, 확대 재생산을 표방하

는 웹 2.0시대를 맞이하여 인터넷 사용자들이 생성하는 온라인 고객리뷰들(online customer reviews)인 비정형 데이터(unstructured data)는 매우 많고 또한 계속 증가

* 이 논문은 2009년도 제주대학교 학술연구지원사업에 의하여 연구되었음

(This work was supported by the research grant of the Jeju National University in 2009)

접수번호 : #100813-001

심사완료일 : 2010년 10월 26일

접수일자 : 2010년 08월 13일

교신저자 : 김근형, e-mail : khkim@jejunu.ac.kr

하고 있다. 웹 2.0의 국내 주요 사례로 싸이월드, 네이버 지식인, 다음카페 등이 있지만 IT 선진국인 미국에서는 더욱 활성화 되어 있다. 미국의 최대 온라인 관광 사이트인 TripAdvisor(www.tripadvisor.com)에는 관광객들이 직접 기록해 놓은 수십만 건의 관광정보와 경험 지식들이 비정형 데이터 형태로 게시되어 있고 많은 관광객들이 유용한 정보로써 활용하며 확대 재생산하고 있다. 국내에서도 웹 2.0 마인드를 표방하는 전문 사이트들이 개설되어 운영되고 있으며 네티즌들의 참여가 점차 확대되고 있는 추세이기 때문에 네티즌들의 의견, 경험, 지식 등을 포함하는 비정형 데이터는 더욱 증가 될 것이다.

웹2.0 기술과 관련하여 Ajax(Asynchronous Javascript and XML)나 매시업(Mashup) 등 온라인 콘텐츠의 작성과 공유기법 등에는 많은 관심을 기울이고 있지만 온라인 고객리뷰(customer reviews) 등과 같은 비정형 데이터를 분석하여 좋은 정보를 생산하기 위한 기법에 대한 연구는 그에 미치지 못하는 것으로 판단된다. 데이터를 생성하고 공유하고자 하는 욕구의 다음 단계는 이 데이터를 분석·요약하여 효율적으로 활용하려는 욕구가 생기게 마련이다. 웹 사이트에 게시되어 있는 온라인 고객리뷰 등 대량의 비정형 데이터들을 분석함으로써 유용한 정보나 지식을 도출할 수 있으며 이러한 정보나 지식은 네티즌과 일반 사용자뿐만 아니라 특히 기업 등에서 매우 유용하게 활용할 수 있다.

최근에 부상하고 있는 오피니언마이닝(opinion mining)은 웹사이트에 게시되어 있는 온라인 고객리뷰들을 분석 대상으로 하는 텍스트마이닝(Text Mining)의 한 분야로써 온라인 고객리뷰들에 대한 긍정(positive)과 부정(negative)의 분포 등을 분석할 수 있지만[1-3] 그 분석능력이 단순하기 때문에 그 유용성에 한계가 있다. 즉, 오피니언마이닝 기술에 의하여 특정제품이나 서비스의 각 특징들에 대한 고객의견 분포를 전반적으로 파악할 수 있다할 지라도, 고객들이 보다 중요하게 생각하는 특징들은 무엇이고 그 특징들 사이의 상대적 중요성은 어떤지 등에 대한 정보는 보여줄 수 없다.

본 연구에서는 온라인고객리뷰들 중에서 고객들이 중요하게 생각하는 특징들과 각 특징 사이의 연관관계

등을 표현할 수 있는 종속성네트워크 모델을 제안한다. 또한, 이러한 종속성네트워크 모델을 기반으로 한 온라인 고객리뷰 분석시스템을 설계하고 구현하여 그 성능을 평가한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 관련연구를 고찰하고 제 3장에서는 종속성 네트워크 모델을 제안하고 4장에서 종속성 네트워크를 기반으로 한 분석시스템을 설계한다. 제 5장에서는 구현시스템의 성능을 평가하고 제 6장에서 결론을 맺는다.

II. 관련연구

1. 문서요약

기존의 문서요약 기술은 2가지 유형으로 나누어진다. 하나는 원형틀(template, 원형관)을 채워 넣은 방식이고, 다른 하나는 핵심문장을 추출하는 방식이다[4][5]. 원형틀을 채워 넣는 방식은 문서안의 핵심 개체(entity)나 사실(fact)을 식별·추출하여 원형틀의 각슬롯(slot)에 할당한다. 이러한 방법은 원형틀이 먼저 만들어져야 하기 때문에 해당 도메인(domain)에 대한 사전지식이 필요하며 따라서, 도메인 의존적 기법이라는 한계가 있다. 핵심문장 추출방식은 문서내용 중에서 가장 대표적인 문장이나 단락을 추출함으로써 문서내용을 간략화한다.

2. 오피니언마이닝

오피니언마이닝은 상품 평이나 고객리뷰를 요약한다는 측면에서 기존의 문서요약과 유사한 점이 있지만, 온라인 고객리뷰가 대량의 다중문서로 구성되고 마이닝 대상이 상품특성과 의견이라는 측면에서 기존의 문서요약기법과 차이가 있다. [6]에서는 기계학습 및 자연어처리기술을 활용하여, 온라인고객리뷰 데이터에 대한 감성분석과 분석결과 요약기법을 제시하고 있으며, Opinion Observer라는 시스템을 개발하였다. 미국 카네기멜론 대학교에서는 Redopal 시스템을 개발한 사례가 있으며[7], 이는 고객리뷰 데이터와 사용자 평가점수를 활용하여 요약보고서를 생성하는 기법을 제안하였

다. [8]에서는 문장구조와 문장 사이의 관계, 문장성분의 패턴정보 등의 언어규칙을 이용한 통계학적 방법으로 오피니언마이닝에 접근하고 있다. [9][10]에서는 워드넷을 활용하여 어휘의 긍정이나 부정적 의미를 판단하고 이를 센티워드넷(SentiwordNet)으로 응용하여 감정의 폭을 정량화하는 방법을 제시하고 있다.

3. 기존 연구들의 한계

문서를 요약하는 방법 중의 하나인 핵심문장 추출 방식은 길이가 긴 단일문서의 간략화를 목적으로 하기 때문에 길이가 짧은 대량의 다중문서로 구성된 온라인 고객리뷰를 분석하기 위한 방법으로는 적합하지 않다.

오피니언마이닝 연구의 핵심은 고객리뷰내의 주관적 문장들에 대해서 긍정 혹은 부정의 비율을 자동으로 분석하는 것이다. 그러나, 고객의 관심은 주관적 의견을 표현하는 문장뿐만 아니라 객관적 사실을 나타내는 문장을 통해서도 표현되므로 온라인 고객리뷰의 보다 정확한 분석은 주관적 관점과 객관적 관점 모두를 분석대상으로 설정할 필요가 있다.

III. 종속성 네트워크 모델

이번 장에서는 온라인 고객리뷰를 보다 정확하고 심층적으로 분석하기 위한 종속성네트워크모델을 새롭게 제안한다.

종속성 네트워크는 온라인 고객리뷰와 같은 대량의 다중문서에 대하여 고객들이 관심을 갖는 상품특성들과 이들 사이의 관계성을 표현하는 형태로 다중문서를 간략화 시킨다. 온라인 고객리뷰에서 언급되는 상품특성들은 주로 명사들로 표현된다. 오피니언마이닝은 주관적 의견이 반영된 명사들만을 다루는데 반하여 종속성 네트워크는 주관적 의견이 반영되지 않은 명사들까지 고려하기 때문에 정보표현력이 더 크다.

온라인 고객리뷰에서 자주 언급되는 명사들은 고객들의 주요 관심사항임을 나타내는 것이므로 종속성네트워크에서는 자주 출현하는 명사들을 우선적인 분석 대상으로 설정한다.

<정의 1> 빈발도(degree of frequency)

온라인 고객리뷰에서 명사 A가 얼마나 자주 출현하는지를 나타내는 척도

$$f(A) = \frac{n(A)}{N}$$

(N: 전체 리뷰 개수, n(A): 명사 A의 출현 빈도) □

온라인 고객리뷰에서 고객들이 관심을 갖는 상품특성들 사이의 관계성이나 연계성은 의미있는 정보가 될 수 있다. 예를 들어, 디지털카메라와 관련된 온라인 고객리뷰에서 '크기'라는 단어와 '가격'이라는 단어가 동시에 언급되는 빈도가 많은 경우 카메라의 크기에 관심이 있는 고객들은 가격에도 관심이 있음을 의미한다.

<정의 2> 관련도(degree of relationship)

온라인 고객리뷰에서 명사 A와 명사 B가 서로 얼마나 관련되어 있는지 나타내는 척도

$$r(A, B) = \frac{n(A \cap B)}{n(A)} \quad (\text{단, } n(A) \geq n(B)) \quad \square$$

관련도가 있는 명사들 사이에서는 서로 종속관계가 존재할 수 있다. 예를 들어, 디지털카메라의 특성 중에서 '크기'와 '가격'이 서로 관련도가 있는 것으로 분석되었을 때, '가격'의 빈발도가 '크기'의 빈발도보다 더 크다면 '가격' 특성이 '크기' 특성보다 더 독립적이라 할 수 있다. 즉, '가격'은 개체 개념이 되고 '크기'는 가격 개체에 종속되면서 영향을 미치는 속성 개념이 되며, '크기'는 '가격'에 종속된다. '크기'가 '가격'에 종속될 경우, '크기'는 '가격'에 영향을 미치는 여러 요소 중에서 하나가 된다는 의미이다.

<정의 3> 종속성(dependency)

$B \rightarrow A$: 명사 A와 명사 B에 대하여 $f(A) \geq t_1$, $f(B) \geq t_1$, $f(A) \geq f(B)$, $r(A, B) \geq t_2$ 일 때, B는 A에 종속된다고 한다(단, t_1 은 최소빈발도, t_2 는 최소관련도) □

<정의 4> 종속성 네트워크

N에 속하는 모든 명사들 간의 종속성 관계를 나타낸

다이어그램

{ $B \rightarrow A \mid A \in N, B \in N, N$ 는 리뷰 안에 있는 명사들의 집합 } □

[그림 1]은 종속성네트워크를 나타내고 있다. N_i 는 명사들이고 f_i 는 해당 명사의 빈발도, r_i 는 관련도를 의미한다.

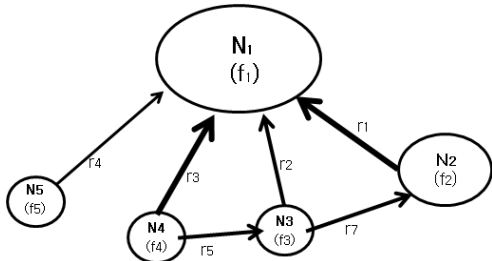


그림 1. 종속성 네트워크의 예

IV. 시스템설계

이번 장에서는 종속성 네트워크 모델을 기반으로 하는 온라인 고객리뷰 분석시스템을 설계한다.

1. 종속성 네트워크 생성과정

[그림 2]에서는 종속성 네트워크의 생성과정을 나타내고 있다. 웹사이트 상에 게시된 온라인 고객리뷰들은 텍스트문서 형태로 고객리뷰 데이터베이스에 저장된다. 텍스트문서는 비정형 데이터이므로 전처리과정(preprocessing)을 통하여 정형 데이터인 관계형 파일로 변환된다. 전처리과정 중에서 텍스트문서내의 각 한글문장들은 한국어 구문분석기에 의하여 각 단어들에 품사가 부여된 형태의 구문구조트리로 변환된다. 구문구조트리 파일로부터 명사에 해당하는 단어들 추출되어 테이블 형태의 관계형 파일에 저장된다.

종속성 네트워크 생성과정 동안에는 관계형 파일안의 각 명사들에 대하여 빈발도가 계산되고 또한 명사들 사이의 관련도가 계산되면서 종속성 네트워크가 만들어진다.

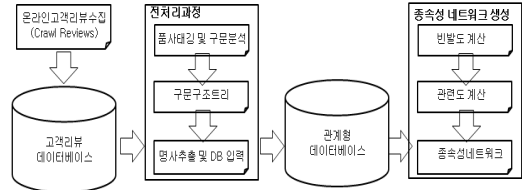


그림 2. 종속성 네트워크 생성과정

2. 시스템 구조도 및 알고리즘

[그림 3]은 온라인 고객리뷰 분석시스템의 구조도를 나타내고 있다.

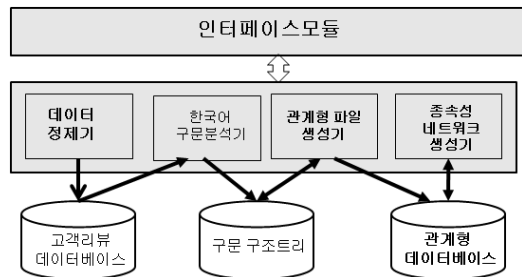


그림 3. 시스템 구조도

데이터정제기에 의하여 고객리뷰 문서내의 특수문자들이 삭제된다. 정제된 고객리뷰 문서들 내의 각 한글문장들은 한국어구문분석기에 의하여 구문구조트리 파일로 변환된다. 구문구조트리는 고객리뷰의 각 단어에 품사가 할당된 형태의 구문분석된 고객리뷰들의 집합이다. 관계형파일 생성기는 구문구조트리를 입력파일로 하여 명사들을 추출하고 각 고객리뷰에 속하는 명사들을 관계형 파일의 각 행에 대응시키면서 관계형 파일을 생성한다. 종속성 네트워크 생성기는 고객리뷰내의 각 명사들이 저장된 관계형 파일을 입력으로 하여 각 명사들의 빈발도와 다른 명사와의 관련도를 계산하면서 종속성 네트워크를 생성한다.

[그림 4]는 관계형파일생성기의 개략적인 알고리즘을 나타내고 있다. 구문분석된 고객리뷰들의 집합 C의 각 고객리뷰 c_i 로부터 명사들 n_{ij} 를 추출하고 관계형파일 N의 레코드 i에 삽입하고 있다.

관계형파일 생성기 알고리즘
입력: 구문분석된 고객리뷰들의 집합 C 출력: 관계형파일 N
<pre> 구문분석된 고객리뷰들로부터 명사들을 추출하여 관계형파일 N 생성 BEGIN while each c_j in C { /* c_j는 C안에 있는 각 고객리뷰들 */ find $n_{j1}, n_{j2}, \dots, n_{jm}$ in c_j; /* n_{jk}는 c_j에 있는 명사 */ /* m은 N의 속성 수 */ insert into N values($n_{j1}, n_{j2}, \dots, n_{jm}$); } END </pre>

그림 4. 관계형 파일 생성기

[그림 5]는 종속성네트워크생성기의 빈발도 계산모듈에 대한 개략적인 알고리즘을 나타내고 있다. 관계형 파일 N의 각 레코드 r_i 는 하나의 고객리뷰와 대응되고 n_{ij} 는 I번째 고객리뷰 내의 j번째 명사에 대응된다. 관계형파일 N의 각 레코드 r_i 의 각 명사들 n_{ij} 의 빈발도를 계산하기 위해서 N의 각 속성 a_k 에 대하여 SQL의 select 명령과 update명령, like연산자를 이용하고 있다(7행에서 9행). 10행에서 f는 명사 n_{ij} 의 출현빈도수를 의미하고 total은 고객리뷰 개수를 나타내므로 dof(degree of frequency)는 n_{ij} 의 빈발도가 된다. 11행에서 각 명사의 빈발도는 2개의 속성 noun과 freq로 구성된 새로운 관계형 파일 NC에 삽입된다.

빈발도 계산 알고리즘
입력: 관계형파일 N 출력: 관계형파일 NC
<pre> 관계형파일 N에 있는 각 명사들의 출현빈도수를 계산하여 관계형파일 NC에 입력 BEGIN 1 total = rec# of N 2 while each r_i in N /* r_i는 N의 각 레코드 */ 3 while each n_{ij} in r_i /* n_{ij}는 r_i의 각 속성 값 */ 4 if ($n_{ij} \neq null$){ 5 f = 0; 6 for (k = 1; k ≤ m; k++){ 7 /* m은 레코드의 속성 수 */ 7 select count(*) as f from N where a_k like n_{ij} 8 /* a_k는 레코드의 각 속성 */ 8 update N set $a_k = null$ where a_k like n_{ij} 9 f = f + t; /* f는 명사 n_{ij}의 출현 빈도수 */ 9 } /* end of if */ 10 dof = f/total; // dof : degree of frequency 11 insert into NC($noun, freq$) values(n_{ij}, dof); 12 } /* end of while */ 13 } /* end of while */ 14 } /* end of while */ END </pre>

그림 5. 빈발도 계산

[그림 5]는 종속성네트워크생성기의 관련도 계산모듈에 대한 개략적인 알고리즘을 나타내고 있다. 관계형 파일 NC의 각 레코드 r_i 는 고객리뷰 상의 각 명사 $r_i.noun$ 에 대한 빈발도 $r_i.freq$ 를 포함한다. $r_i.noun$ 과 관련된 명사들을 선택하고 그 관련도를 계산하기 위하여 먼저 관계형파일 N으로부터 $r_i.noun$ 에 대한 뷰 NV를 생성한다(4행). NV에서 $r_i.noun$ 이외의 명사들 $r_j.noun$ 의 출현빈도 f를 계산한 다음(6행에서 11행), NV의 전체레코드 수인 total과 함께 $r_i.noun$ 와 $r_j.noun$ 사이의 관련도 dor(degree of relationship)을 계산할 수 있다(13행). dor이 최소관련도 t_2 보다 크면 $r_i.noun$ 과 $r_j.noun$ 는 dor과 함께 종속성 네트워크에 포함된다.

관련도 계산 알고리즘
입력: 관계형파일 NC 출력: 종속성 네트워크
<pre> NC 상에 있는 각 명사들 사이의 관련도를 계산하여 종속성 네트워크를 생성한다. BEGIN 1 while each r_i in NC 2 /* r_i는 NC의 각 레코드, 1 ≤ i ≤ rec# of NC */ 2 if ($r_i.freq \geq t_1$) { /* t_1은 최소빈발도 */ 3 j = i + 1; 4 create 뷰 NV(select * from N 5 where a_k like $r_i.noun$ from N; 6 /* a_k는 N의 레코드의 각 속성 */ 7 total = rec# of NV; 8 while each r_j in NC { 9 /* i+1 ≤ j ≤ rec# of NC */ 10 f = 0; 11 for (k = 1; k ≤ m; k++){ 12 /* m은 레코드의 속성 수 */ 12 select count(*) as f from NV 13 where a_k like $r_j.noun$; 14 f = f + t; // f는 명사 $r_j.noun$의 출현 빈도수 15 } /* end of for */ 16 dor = f/total; /* dor : degree of relationship */ 17 if (dor ≥ t_2) /* t_2는 최소관련도 */ 18 add ($r_i.noun, r_j.noun$) with ($r_i.freq, dor$) 19 to 종속성네트워크; 20 } /* end of while */ 21 } /* end of if */ 22 } /* end of while */ END </pre>

그림 6. 관련도 계산

V. 시스템 구현 및 성능평가

본 논문에서는 4장에서 제시한 시스템구조도와 알고리즘을 바탕으로 온라인 고객리뷰 분석시스템을 개발

하였다. 프로그래머언어로는 비주얼베이직을 사용하였고 DBMS는 MS SQL server 2008을 사용하였다. 한국어 구문분석을 위하여 국내의 대표적인 형태소분석기인 KLT[12]을 사용하였다. 실험용 데이터는 네이버랩(lab.naver.com)에서 제공하는 영화 “해운대” 40자평 데이터셋을 사용하였다. 데이터셋에는 약 1만개의 고객리뷰가 포함된다. 개발된 시스템은 CPU 1.73GHz, 메모리 1GB, 윈도우XP가 탑재된 환경에서 수행되었다.

[그림 7]은 한국어구문분석을 위한 형태소분석 프로그램(parse.exe)을 실행하는 화면을 나타내고 있다.

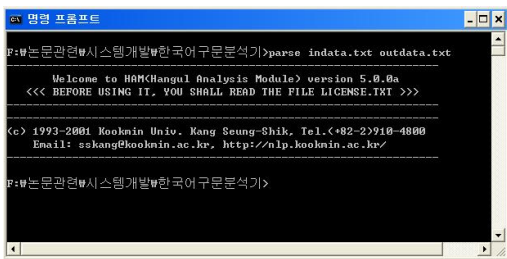


그림 7. 한국어 구문분석 프로그램 실행화면

[그림 8]은 비주얼베이직과 MS SQL server 2008을 사용하여 개발한 시스템 구현 화면을 나타내고 있다.

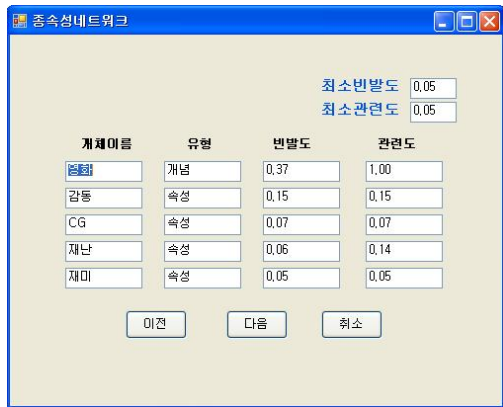


그림 8. 시스템구현 화면

[그림 9]는 최소빈발도와 최소관련도를 0.05로 설정하였을 때 구현한 시스템을 통하여 도출될 수 있는 종속성네트워크를 나타내고 있다. ‘영화’라는 단어가 자주

출현하고 있으며 “영화”와 관련하여 고객들은 ‘CG (Computer Graphic)’, ‘재미’ 등과 같은 속성에 관심을 나타내고 있다.

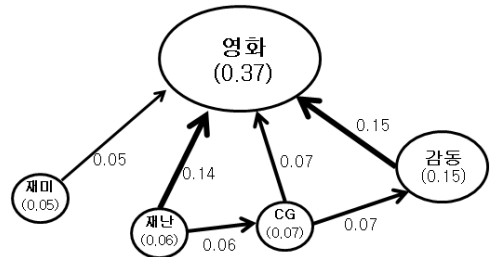


그림 9. 종속성네트워크 예(최소 빈발도 : 0.05, 최소관련도 : 0.05)

[그림 10]은 최소빈발도를 0.03, 최소관련도를 0.01로 설정하였을 때 생성된 종속성네트워크를 나타내고 있다. 고객들은 ‘국가대표’라는 또 다른 영화와 비교하면서 관심을 나타내고 있음을 알 수 있다.

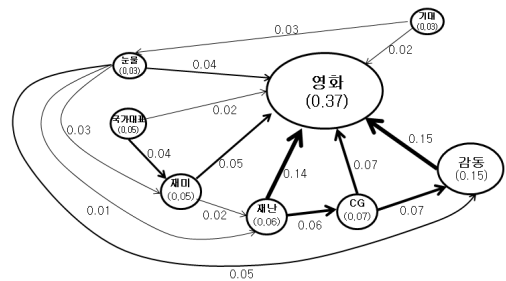


그림 10. 종속성네트워크 예(최소 빈발도 : 0.03, 최소관련도 : 0.01)

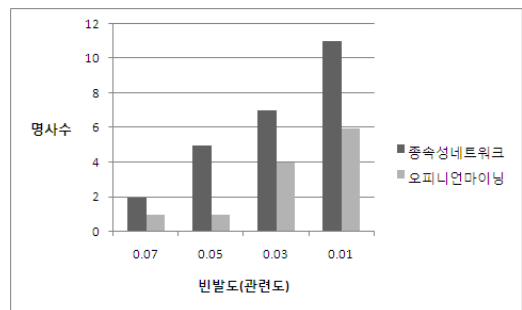


그림 11. 명사수의 비교

[그림 11]에서는 종속성네트워크와 기존의 오피니언 마이닝 기법을 비교하고 있다. 최소빈발도와 최소관련도의 변화에 따라 요약정보에 포함되는 명사들(개념 또는 속성)의 수가 어떻게 달라지는지 보여주는 그래프이다. 종속성네트워크에 포함된 명사들의 수가 기존 오피니언마이닝 기법에 의하여 생성되는 명사 수보다 더 많음을 알 수 있다. 명사들의 수가 많을수록 보다 풍부한 정보를 표현할 수 있으므로 종속성네트워크의 정보 표현력이 더 크다는 것을 알 수 있다.

[그림 12]는 최소빈발도와 최소관련도의 변화에 따라 종속성 네트워크에 포함된 명사들의 수가 어떻게 달라지는지 보여주는 그래프이다. 최소빈발도와 최소관련도를 낮게 설정할 수록 종속성 네트워크에 포함되는 명사들의 수는 증가하고 있음을 알 수 있다.

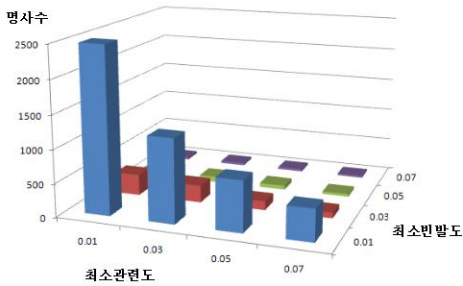


그림 12. 종속성네트워크의 명사수

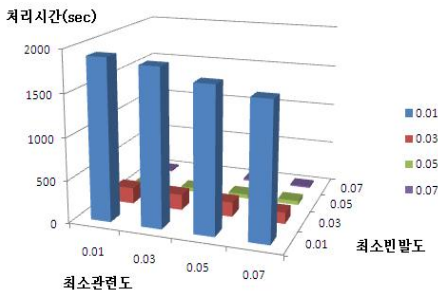


그림 13. 종속성네트워크의 생성시간

[그림 13]은 최소빈발도와 최소관련도의 변화에 따라 종속성 네트워크를 생성하는 시간이 어떻게 달라지는지 보여주는 그래프이다. 보다 다양한 정보를 도출하기 위해서는 최소빈발도와 최소관련도를 낮게 설정할 필

요가 있지만, 유의미하지 않은 단어들이 종속성 네트워크에 포함될 소지가 있으며 종속성 네트워크의 생성시간이 더 길어질 수 있다.

VI. 결론

웹 2.0 시대에 네티즌들의 경험과 의견 등이 표현된 온라인 고객리뷰나 상품리뷰 등은 매우 많고 더욱 증가하고 있다. 특정 제품이나 서비스에 대한 네티즌의 의견들은 고객뿐만 아니라 기업 입장에서 마케팅이나 경영전략을 수립하기 위한 중요한 자료가 될 수 있기 때문에 온라인 고객리뷰를 분석하는 것은 매우 중요하다.

본 논문에서는 온라인고객리뷰들 중에서 고객들이 중요하게 생각하는 특징들과 각 특징 사이의 연관관계 등을 표현할 수 있는 종속성네트워크 모델을 제안하였고, 이러한 종속성네트워크 모델을 기반으로 한 온라인 고객리뷰 분석시스템을 설계하고 구현하였다.

구현한 시스템의 성능을 평가한 결과, 최소빈발도와 최소관련도가 작을수록 종속성 네트워크는 보다 많은 단어들을 포함하고 보다 많은 정보를 표현할 수 있지만 네트워크 생성시간은 길어짐을 알 수 있었다.

종속성 네트워크 모델은 주관적 의견이 반영되지 않은 명사들까지 고려하기 때문에 기존의 오피니언마이닝 기술에 비하여 보다 많은 정보를 표현할 수 있지만 특징들에 대한 긍정이나 부정 정도는 나타낼 수 없다. 이는 본 연구의 한계로써, 향후 오피니언마이닝의 표현력을 내재할 수 있도록 종속성 네트워크 모델을 더 확장시키는 연구가 필요하다.

참고 문헌

- [1] Xiaowen Ding, "A Holistic Lexicon-Based Approach to Opinion Mining," Proc. of the international conference on web search and web mining, pp.231-240, 2008.
- [2] Jerzy Stefanowski and Dawid Weiss, "Comprehensible

and Accurate Cluster Labels in Text Clustering," Conference RIAO 2007, Pittsburgh PA, USA, May, pp.34-45, 2007.

- [3] Zamir and Etzioni, Grouper: a dynamic clustering interface to Web search results. Computer Networks(Amsterdam, Netherlands), 31, pp.11-16, 1999.
- [4] G. Salton, A. Singhal, C. Buckley, and M. Mitra, Automatic Text Decomposition using Text Segments and Text Themes," ACM Conference on Hypertext, pp.56-65, 1996.
- [5] B. Boguraev and C. Kennedy, "Salience-Based Content Characterization of Text Documents," Proc. of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization, pp.76-87, 1997.
- [6] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the Web," Proc. of the 14th international conference on WWW, pp.10-14, 2005.
- [7] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, Chun Jin, "Red Opal: Product-Feature Scoring from Reviews," Proc. of the 8th ACM conference on Electronic commerce, pp.11-15, 2007.
- [8] Xiaowen Ding and Bing Lui, "The Utility of Linguistic Rules in Opinion Mining," SIGR pp.811-812, 2007.
- [9] E. Courses and T. Surveys, "Using SentiWordNet for multilingual sentiment analysis," Data Engineering Workshop ICDEW, pp.102-110, 2008.
- [10] Mingqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews," KDD'04, pp.168-177, 2004.
- [11] Xiaowen Ding, Bing Liu and Philip S. Yu, "A Holistic Lexicon-Based Approach to Opinion Mining," WSDM'08, pp.231-239, 2008.
- [12] <http://nlp.kookmin.ac.kr/HAM/kor/download.html>

저 자 소 개

김 근 형(Keun-Hyung Kim)

정회원



- 1990년 2월 : 서강대학교 컴퓨터학과(공학사)
- 1992년 2월 : 서강대학교 컴퓨터학과(공학석사)
- 2001년 2월 : 서강대학교 컴퓨터학과(공학박사)
- 2001년 9월 ~ 현재 : 제주대학교 경영정보학과 부교수
<관심분야> : 데이터마이닝, 비즈니스인텔리전스