

오피니언 마이닝을 이용한 지능형 VOC 분석시스템

김유신

국민대학교 비즈니스IT전문대학원
trust@kookmin.ac.kr

정승렬

국민대학교 비즈니스IT전문대학원 교수
srjeong@kookmin.ac.kr

기업 경영에 있어서 고객의 소리(VOC)는 고객 만족도 향상 및 기업의사결정에 매우 중요한 정보이다. 이는 비단 기업뿐만 아니라 대고객, 대민원 업무를 처리하는 모든 조직에 있어서도 동일하다. 때문에 최근에는 기업뿐만 아니라 공공, 의료, 금융, 교육기관 등 거의 모든 조직이 VOC를 수집하여 활용하고 있다. 이러한 VOC는 방문, 전화, 우편, 인터넷게시판, SNS 등 다양한 채널을 통해 전달되지만, 막상 이를 제대로 활용하기는 쉽지 않다. 왜냐하면, 고객이 매우 감정적인 상태에서 고객의 주관적 의사를 음성 또는 문자로 표출하기 때문에 그 형식이나 내용이 정형화되어 있지 않고 저장하기도 어려우며 또한 저장하더라도 매우 방대한 분량의 비정형 데이터로 남기 때문이다. 본 연구는 이러한 비정형 VOC 데이터를 자동으로 분류하고 VOC의 유형과 극성을 판별할 수 있는 오피니언 마이닝 기반의 지능형 VOC 분석 시스템을 제안하였다. 또한 VOC 오피니언 분석의 기준이 되는 주제지향 감성사전 개발 프로세스와 각 단계를 구체적으로 제시하였다. 그리고 본 연구에서 제시한 시스템의 효용성을 검증하기 위하여 의료기관 홈페이지에서 수집한 4,300여건의 VOC 데이터를 이용하여 병원에 특화된 감성어휘와 감성극성값을 도출하여 감성사전을 구축하고 이를 통해 구현된 VOC분류 모형의 정확도를 비교하는 실험을 수행하였다. 그 결과 “칭찬, 친절함, 감사, 무사히, 잘해, 감동, 미소” 등의 어휘는 매우 높은 긍정 오피니언 값을 가지며, “통명, 뽕니까, 말하더군요, 무시하는” 등의 어휘들은 강한 부정의 극성값을 가지고 있음을 확인하였다. 또한 VOC의 오피니언 분류 임계값이 -0.50일 때 가장 높은 분류 예측정확도 77.8%를 검증함으로써 오피니언 마이닝 기반의 지능형 VOC 분석시스템의 유효성을 확인하였다. 그러므로 지능형 VOC 분석시스템을 통해 VOC의 실시간 자동 분류 및 대응 우선순위를 도출하여 고객 민원에 대해 신속히 대응한다면, VOC 전담 인력을 효율적으로 운용하면서도 고객 불만을 초기에 해소할 수 있는 긍정적 효과를 기대해 볼 수 있을 것이다. 또한 VOC 텍스트를 분석하고 활용할 수 있는 오피니언 마이닝 모형이라는 새로운 시도를 통해 향후 다양한 분석과 실용 프레임워크의 기틀을 제공할 수 있을 것으로 기대된다.

논문접수일 : 2013년 06월 28일 게재확정일 : 2013년 07월 09일

투고유형 : 학술대회우수논문 교신저자 : 정승렬

1. 서론

현대 기업 경영에 있어서 고객과 기업 사이에 펼쳐지는 상호 커뮤니케이션은 기업경쟁력의 근간으로 자리 잡으며, 그 중요성이 증가되고 있다. 고객의 니즈를 파악하고 이를 충족시키기 위해, 고객의 다

양성과 개별적인 특성을 감안하여 고객이 원하는 바를 제공함으로써, 고객의 가치를 높이는 것은 모든 기업이 추구하는 바이다. 특히 고객의 요구를 받아들일 수 있는 여러 전략들 중에서도 VOC는 전화 조사, 고객 패널, 포커스 그룹 인터뷰, 우편조사, 의견조사 엽서 등 다양한 경로를 통해 고객의 의견을

들을 수 있기 때문에 다른 커뮤니케이션 도구보다도 더 큰 의미를 지니고 있다고 볼 수 있다(Choi and Choi, 2011).

이는 비단 기업뿐만 아니라 대고객, 대민원 업무를 처리하는 모든 조직에 있어서도 동일하다. 때문에 최근에는 일반 기업뿐만 아니라 공공, 의료, 금융, 교육기관 등 거의 모든 조직이 VOC를 수집하고 이를 분석하여 상품이나 서비스의 개발에 필요한 정보로 활용하고 있다.

이에 더해 인터넷과 정보기술의 발전은 이제까지 한정되어 있던 VOC수집을 위한 기업과 고객의 접점을 다양한 형태로 확장시켰다(Ju and Hwang, 2004). 인터넷과 모바일, 스마트 서비스의 발달과 함께 VOC 접수창구는 전화, 우편, 방문 등의 기존 오프라인 방식을 넘어 홈페이지, 커뮤니티, 게시판, SNS 등 다양한 온라인 채널로 확대되고 있다. 그러나 채널이 확장된 만큼 더 많은 VOC가 축적되고 있지만, 이를 제대로 활용하기는 쉽지 않다. 왜냐하면, 고객이 매우 감정적인 상태에서 고객의 주관적 의사를 음성 또는 문자로 표출하기 때문에 그 형식이나 내용이 정형화되어 있지 않고 저장하기도 어려우며 또한 저장하더라도 매우 방대한 분량의 비정형 데이터로 남기 때문이다.

그렇기 때문에 지금과 같이 인터넷이 발달하고 모바일 기반의 스마트 서비스가 일반화된 상태에서 콜센터, 홈페이지, SNS 등의 다양한 채널로 접수되는 대량의 VOC를 분석할 수 있는 자동화 기법과 시스템에 대한 요구는 더욱 높아질 수 밖에 없다. 또한 담당자가 이런 대량의 VOC 데이터를 수동으로 처리한다는 것 역시 매우 비효율적이고 임의적인 대응방식이라 하겠다. 결국 VOC 데이터를 대량의 데이터, 즉 빅데이터 관점에서 바라보고 이를 효과적으로 분석하고 활용할 수 있는 자동화된 분석 시스템이 요구된다고 하겠다.

본 연구는 이러한 비정형 VOC 데이터를 자동으로 분류하고 VOC의 유형과 극성을 판별할 수 있는 오피니언 마이닝 기반의 지능형 VOC 분석시스템을 제시하였다. 그리고 이러한 시스템의 가능성을 확인하기 위하여 A병원의 홈페이지를 통해 접수된 4,300여건의 VOC 데이터를 오피니언 마이닝으로 분석하여 VOC 유형별 감성어휘 사전과 극성값을 도출하였다.

본 연구의 지능형 VOC 분석시스템은 실시간으로 접수되는 VOC의 유형을 자동 분류함과 동시에 극성을 부여함으로써 VOC 대응의 우선순위를 제시하여 고객 민원에 대한 신속한 대응을 가능하게 하며, 결과적으로 고객 만족도를 향상시켜 조직 운영에 긍정적 효과를 발휘할 수 있을 것으로 기대된다. 또한 VOC 텍스트를 분석하고 활용할 수 있는 오피니언 마이닝 모형이라는 새로운 시도를 통해 향후 다양한 분석과 실용 프레임워크의 기틀을 제공할 수 있을 것으로 기대된다.

본 연구는 다음과 같이 구성되어 있다. 제 2장에서는 VOC와 오피니언 마이닝에 대해서 소개하고, 제 3장에서는 지능형 VOC 분석시스템의 구조 및 VOC에 특화된 감성 사건의 구축 방안에 대해 간략하게 설명한다. 이에 대한 실험 데이터와 결과는 제 4장에서 제시하며 마지막 제 5장에서 결론 및 향후 연구에 대하여 논의한다.

2. 관련 연구

2.1 고객의 소리(VOC)

모든 기업의 서비스 품질향상과 고객만족은 VOC (Voice of Customer)에서부터 시작한다. VOC란 고객의 소리에 귀를 기울여 그들의 요구를 파악하고 이를 수용하여 경영활동을 함으로써 궁극적으로는

고객 만족을 추구하고자 하는 제도이다. 각 기업의 VOC 운영 목적은 기업이 제공하는 제품이나 서비스에 대한 만족 여부를 파악함으로써 구체적으로 고객을 이해하기 위함이며, 변화하는 고객의 니즈와 기대를 파악하여 시장에 대한 끊임없는 주시와 고객 관점에서 나오는 새로운 아이디어의 발견 및 장기적인 차원에서 기업과 고객 간의 유대강화로 고객을 기업에 밀착시키기 위함이다(Choi and Choi, 2011).

그러므로 VOC 분석은 VOC 데이터를 통하여 수시로 변화하는 고객의 요구를 지속적으로 반영하여 상품이나 서비스를 개발할 수 있도록 해주는 것을 목적으로 한다. 즉 다양한 경로를 통하여 수집된 VOC 데이터를 통합하여 고객이나 지역에 따른 고객의 요구변화를 수집하고 이것을 분석하여 상품이나 서비스의 개발에 필요한 정보로 변환하여 제공하는 것이 VOC 분석의 목적이다(Ju and Hwang, 2004).

VOC와 관련된 선행연구들은, VOC 분석시스템을 사후관리 형태가 아닌 고객의 기호나 취향, 불만의 변화를 사전에 감지할 수 있는 적극적인 고객관계관리 기반의 VOC 분석시스템을 제시하거나(Ju and Hwang, 2004), VOC 시스템 운영과 그 결과를 CRM의 하위 지표로서 BSC의 고객관점에 연계를 시도한 연구(Kim, 2012) 등이 있는데, 대체로 고객관점을 분석하고 시스템화하는 연구라고 하겠다. 다른 한편으로 VOC 시스템에 대한 개념 인식, 만족도, 영향력에 대한 인식 등을 이용하여 고객과 직원 간의 인식차이를 분석함으로써 VOC 시스템의 효율적 활용을 유도하고자 한 연구도 있다(Choi and Choi, 2011). Takeuchi et al.(2009)은 단순한 영업 전략을 넘어서 기업의 제품 및 서비스 품질, 운영 효율성 등을 향상시키기 위해서도 고객과의 업무 지향적 대화가 필요함을 강조하면서, 텍스트 마이닝 시스템을 이용해 자동차 렌탈 헬프데스크의 문

의전화 914통을 분석하여 VOC의 중요성을 언급하기도 하였다.

그러나 VOC가 고객 만족이나 기업경영에 미치는 영향이 매우 큰 것에 비해 대다수 연구가 VOC의 중요성이나 VOC 통합 시스템 구축을 다루고 있을 뿐 VOC의 분석 및 활용과 관련된 선행연구는 많지 않은 실정이다. 더구나 VOC 데이터의 컨텐츠에 대한 연구는 더욱 희박한데 이는 실제 VOC 데이터에는 고객의 불만과 항의가 주를 이루고 있어 쉽게 드러내기 힘든 속성을 가졌기 때문으로 생각된다. 이에 더해 인터넷과 스마트 서비스의 발달로 인해 다양해진 VOC 접수창구는 더 많은 VOC를 축적하고 있지만, 대부분이 비정형 텍스트 데이터로 구성되어 있어 이를 제대로 분석하고 활용하기는 쉽지 않다.

2.2 오피니언 마이닝

이런 관점에서 텍스트 마이닝의 한 분야인 오피니언 마이닝이 대량의 VOC 데이터를 분석하는 기법으로서 유용한 대안이 될 수 있다. 오피니언 마이닝은 온라인 뉴스와 소셜미디어의 코멘트 등 사용자가 다양한 컨텐츠를 통해 표출한 의견을 추출, 분류, 이해, 자산화는 과정을 의미하며 감성분석의 다양한 기술을 활용하여 수행된다(Chen and Zimbra, 2010; Liu, 2010).

오피니언 마이닝을 이용하여, 간단하게는 온라인 쇼핑몰에서 잠재 구매자의 상품평 검색 효율을 높이기 위해 상품평 데이터에 순위를 결정하는데 응용될 수도 있으며(Yang et al., 2009; Yune et al., 2010), 영화관람의 후기를 요약하고 긍정/부정을 평할 수도 있고(Zhuang, 2006), 법률 분야의 블로그를 대상으로 오피니언 마이닝을 이용해 고객의 반응이나 법률적 이슈에 대한 모니터링을 할 수도 있다

(Conrad et al., 2007). 더 나아가서는 소셜 미디어에서 나타나는 오피니언들을 조기 감지하여 기업의 위기 상황을 인지하고 위기에 대응할 수 있는 위기 관리 모델의 핵심정보가 될 수도 있고(Cha et al., 2012), 오피니언을 경제적 관점에서 정량화하여 금액으로 환산할 수도 있으며(Ghose et al., 2007), 뉴스 콘텐츠를 오피니언 마이닝으로 분석하여 주가 예측의 중요 정보로 활용하는 지능형 투자의사결정 모형이 될 수도 있다(Kim et al., 2012).

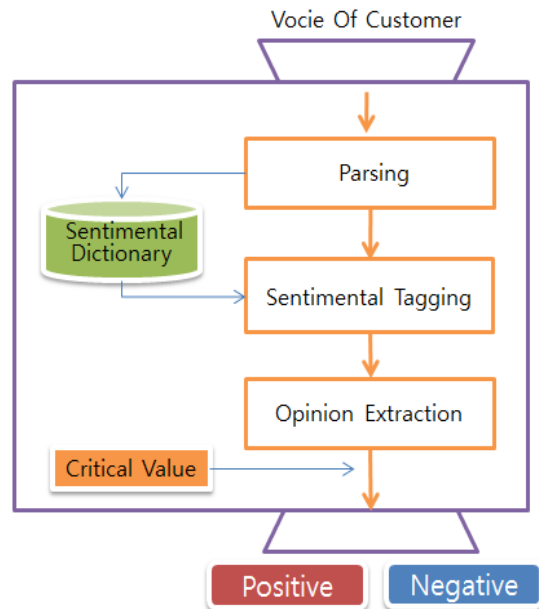
이러한 오피니언 마이닝은 문서에 나타난 의견정보를 추출하기 위해 문장 구조와 문장 사이의 관계, 문장성분의 패턴 정보 등의 언어규칙을 WordNet과 같은 언어사전을 활용하여 감정 분류를 하거나 자연어 처리기법, 또는 기계학습법을 이용하여 감정 극성을 분류하고 감정의 폭을 정량화한다. 특히 언어적 자원을 구축하는 연구에서는 개별 언어의 내용을 대상으로 어휘 또는 어의 수준에서 중립/긍정/부정 평가를 태깅 해놓은 감성사전을 사용하며, WordNet의 각 어휘에 중립/긍정/부정 값을 태깅한 SentiWordNet 기반 연구가 대표적이다.

텍스트에서 정확한 오피니언을 도출하기 위해서는 수집된 언어자원과 그의 감성분류, 즉 감성사전 또는 감성말뭉치라고 하는 어휘 집합이 매우 중요한 요소 중 하나이다. 감성 어휘의 감성이 명확히 정의되어 있다면 텍스트의 감성판별 정확도는 자연스럽게 높아질 것이다. Kim et al.(2012)과 Yu et al.(2013)의 연구에서는 이러한 감성사전을 경제 뉴스로부터 추출하고 감성값을 태깅하여 주가지수의 상승/하락을 예측하는 주식 도메인에 특화된 주제지향 감성사전을 구축하고 이를 활용하는 방안을 제시하였다. 본 연구에서도 VOC 데이터에서 고객의 감성을 추출하는 오피니언 마이닝을 통해 VOC 자동 분류 및 극성 부여를 지원하는 지능형 VOC 분석 시스템을 제시하고자 한다.

3. 지능형 VOC 분석시스템

3.1 지능형 VOC 분석시스템 구조

본 연구에서 제시하는 지능형 VOC 분석시스템은 VOC에 접수된 고객 의견에 나타난 오피니언을 자동으로 분류하고 극성을 부여하는 지능형 오피니언 마이닝 시스템이다. 시스템의 구조는 <Figure 1>과 같이 요약될 수 있다.



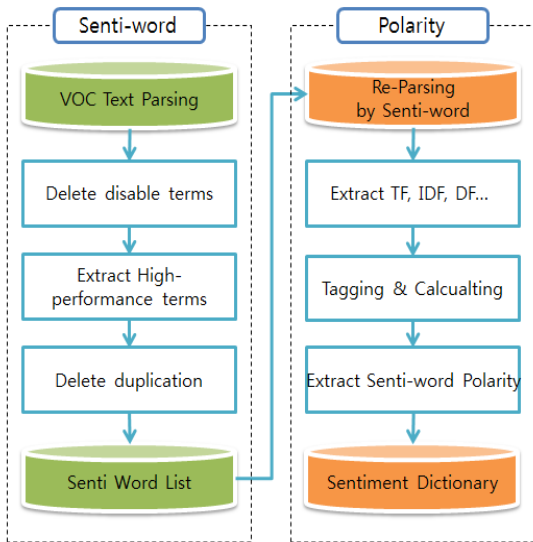
<Figure 1> Intelligent VOC Analyzing System Model

지능형 VOC 분석시스템은 시스템의 입력값으로 들어온 VOC 데이터를 자연어처리(NLP) 프로세스에 의해 파싱하는 것이다. 파싱된 어휘는 감성사전을 참조하여 어휘 극성값을 태깅하고 이를 이용하여 VOC 극성값을 추출한다. 추출된 VOC 극성값은 임계값을 참조하여 긍정/부정의 VOC 오피니언으로 최종 판별된다. 이때 VOC 극성값은 VOC별 우선순위를 비교할 수 있는 기준값으로 활용된다.

3.2 VOC 감성사전 구축

본 연구에서 제시하는 지능형 VOC 분석시스템은 VOC의 콘텐츠를 오피니언 마이닝기법을 분류하여 유형을 분류하고 감성극성을 도출하는 것이다. VOC 오피니언을 도출하기 위해서는 감성극성 값을 가진 감성어휘사전이 필요하며 선행연구에서도 언급된 바와 같이 어휘 자체의 고유 특성으로 부여되기 보다는, 사용되는 도메인과의 관계에서 유연하게 부여되는 것이 바람직하다고 하였다(Yu et al., 2013).

이러한 감성어휘사전을 만들기 위해서는 크게 두 가지 작업이 요구된다. 먼저 감성사전에 쓰일 감성어휘들을 추출하는 작업과 추출된 감성어휘의 감성극성값을 도출하는 것이다. 아래 <Figure 2>는 이러한 감성사전 생성 프로세스를 도식화한 것이다.



<Figure 2> Developing Process of Sentimental Dictionary

감성어휘를 추출하기 위해서는 3단계의 처리 절차를 거친다. 먼저 영어, 숫자, 한 글자 어휘, 특수문

자 등의 불용어 제거작업이 선행된다. 다음으로는 남은 어휘에 대해 TF, TF-IDF 값을 기준으로 일정 기준 이하의 값을 가진 단어들을 제거한다. 마지막으로 일부 중복되는 어휘들을 선별하여 통일성을 부여하는 작업을 통해 감성사전에 사용될 어휘를 최종 확정한다.

이렇게 추출된 어휘에 감성극성을 부여하기 위해서는 감성어휘를 기준으로 2차 파싱을 실시하여 TF, IDF 등을 도출하고 이를 이용하여 극성값을 계산한다. 감성극성값은 감성어휘가 출현한 전체 VOC의 긍정/부정 횟수의 비율로 정의되며, 하나의 VOC의 극성값은 해당 VOC에 출현한 어휘들의 극성값을 평균한 값으로 산출된다. 구체적인 산출 식은 다음 식과 같다.

$$Term(i, j) = \begin{cases} 1 & \left(\begin{array}{l} Doc(j)가 Term(i)를 포함하고, \\ Doc(j)가 긍정인 경우 \end{array} \right) \\ -1 & \left(\begin{array}{l} Doc(j)가 Term(i)를 포함하고, \\ Doc(j)가 부정인 경우 \end{array} \right) \end{cases}$$

$$Term(i).Nm Docs = Term(i)포함하고 있는 VOC의 개수$$

$$Term(i).Opinion = \frac{\sum_{i=1}^n Term(i, j)}{Term(i).Nm Docs}$$

각 VOC의 긍정/부정을 예측하기 위해서는 감성극성값의 구분을 위한 임계값의 설정이 필요하다. 본 연구에서는 감성극성값에 의해 VOC의 긍정/부정 분류의 정확도와 극성값에 의한 우선순위를 도출하는 것을 목적으로 함으로, 임계값의 변화 추이에 따라 예측정확도가 얼마나 달라지며 가장 높은 예측정확도는 어느 정도인지 살펴보도록 한다.

4. 실험

4.1 실험 데이터 소개

본 절에서는 제 3장에서 제안한 지능형 VOC 분석시스템의 분류기준이 되는 VOC 데이터의 속성과 감성사전 구축 및 모형의 예측력을 평가하기 위해, A종합병원의 홈페이지를 통해 접수된 4,300여건의 VOC 데이터를 대상으로 실험을 수행하고자 한다. VOC의 분류 및 극성값에 대한 예측 대상이 되는 목표 변수는 VOC 등록 시 고객이 선택한 VOC 유형으로 “친절 및 감동사례”, “불친절사례”, “불편사항”, “제안 및 건의사항”으로 구성되어있다. 이러한 VOC 유형은 긍정 또는 부정으로 변환하여 감성사전을 추출하고 VOC의 오피니언 극성값을 도출하는 기준값으로 활용된다.

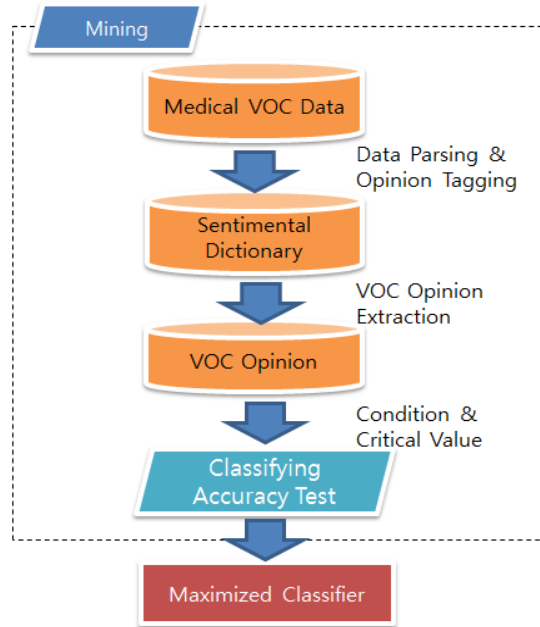
수집한 VOC 데이터는 총 4,305건으로 각 유형별 개수는 아래 <Table 1>과 같다.

<Table 1> VOC Data

VOC 유형	건수	비율
친절 및 감동사례	675	15.7%
불친절사례	887	20.4%
불편사항	1,129	26.2%
제안 및 건의사항	1,624	37.7%
계	4,305	100%

4.2 실험 모형

본 연구에서는 오피니언 마이닝을 이용한 지능형 VOC 분석시스템의 유효성을 검증하기 위하여 실제 VOC 데이터를 오피니언 마이닝으로 분류한 결과의 예측 정확도를 측정하는 실험을 수행한다. 실험의 전체 과정은 <Figure 3>과 같이 요약될 수 있다.



<Figure 3> Process of Performance Evaluation

본 실험의 대상이 되는 의료기관 홈페이지에 접수된 VOC 데이터를 이용해 병원에 특화된 감성어휘와 감성 극성값을 도출하여 감성사전을 구축하고 이를 통해 구현된 VOC 분류 모형의 정확도를 비교하는 실험이다. 이때 VOC의 오피니언을 분류하는 임계값이 중요하므로 임계값의 변화에 따라 VOC의 긍정/부정 오피니언의 예측 정확도가 변하는 양상을 살펴보고 최적의 임계값을 도출하였다.

4.3 결과 분석

VOC 오피니언마이닝을 위한 감성사전을 만들기 위해 수집된 VOC 데이터 4,305건의 내용 텍스트를 SAS 텍스트 마이너를 이용하여 파싱하였다. 파싱을 통해 도출된 어휘 리스트는 총 60,800여 개로 동사, 명사, 형용사, 영어, 숫자, 특수문자 등이 모두 포함되어 있으며, 각 어휘의 Frequency(TF), Weight

(TF-IDF), Numdocs(DF) 값을 가지고 있다. 아래 <Table 2>는 감성 어휘들 중 유형별 고빈도 어휘를 추출한 예시이다. VOC 유형별 고빈도 어휘를 살펴 보면, “입원, 부탁”과 같은 특정 단어는 거의 모든 유형에서 나타나는 어휘이지만, “감사, 친절, 덕분에, 감동, 미소” 등은 친절 및 감동사례에서만 나타나고 있으며, “불친절, 불편, 고통, 취소” 등은 불친절 사례나 불편사항에서 두드러지게 보여진다. 또한 “인턴”이나 “신종플루”, “연말정산”, “확인서” 등은 특정 대상에 대한 오피니언을 강하게 반영하고 있는 것으로 보여진다.

<Table 2> Example of High Frequency Terms

No.	친절	불친절	불편	제안
1	감사	불친절	부탁	부탁
2	부탁	입원	입원	입원
3	입원	부탁	불편	관련
4	친절	필요	고통	감사
5	덕분	불편	필요	불편
6	감동	그러면서	관련	필요
7	주셨습~	인턴	취소	가능
8	칭찬	가라고	신종플루	연말정산
9	드리며	말하는	감사	취소
10	미소	물었더니	요구	확인서

감성사전구축 단계에서 설명한 감성어휘 추출 및 감성극성값 도출 과정을 거쳐 최종 확정된 어휘는 1,243개로 TF \geq 30, TF-IDF \geq 0.5의 기준을 적용하였다. 이렇게 추출된 감성어휘를 SAS 텍스트 마이너의 Start List로 입력하여 2차 파싱을 실시한다. 2차 파싱은 감성어휘의 극성값을 도출하기 위한 것으로 각 어휘의 긍정/부정 극성이 태깅된다. 본 연구에서는 VOC의 4가지 유형을 긍정/부정의 오피니언으로 도출하기 “친절 및 감동사례”와 제안 및 건의 사항”을 긍정으로, “불친절 사례”와 “불편 사항”을

부정으로 간주하여 각 유형에서의 어휘 출현 비율을 확률계산하였다.

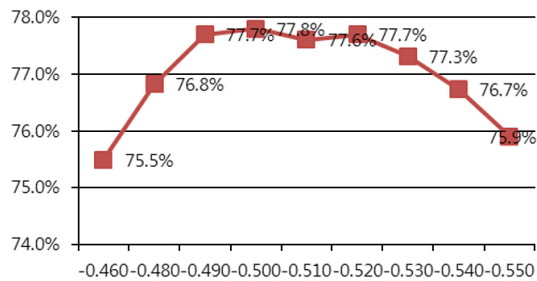
다음의 <Table 3>은 긍정/부정 양 극단에 나타난 어휘와 극성값 예시를 보여주는 것이다. <Table 3>에서 보여지는 바와 같이 “칭찬, 친절함, 감사, 무사히, 잘해, 감동, 미소” 등의 어휘는 매우 높은 긍정 오피니언 값을 가지고 있으며, 이와 반대로 “통명, 뭉니까, 말하더군요, 무시하는” 등의 어휘들은 우리가 일상생활에서도 매우 부정적으로 사용하는 부정 오피니언 어휘들로 강한 부정의 극성값을 나타내고 있다.

<Table 3> Example of Domain-Specific Sentiment Dictionary

감성어휘	극성값(+)	감성어휘	극성값(+)
칭찬	0.6286	통명	-1.0000
친절함	0.5278	이니까	-0.9643
감사	0.5151	뭉니까	-0.9444
무사히	0.5000	대우	-0.9348
기원	0.4375	말하더군요	-0.9259
잘해	0.4063	소비자	-0.9200
친절하시고	0.4063	기다리게	-0.9184
감동	0.3793	무시하는	-0.9118
드리며	0.3750	화중	-0.9091
미소	0.3721	하는거	-0.9063
편안하게	0.3214	기다린	-0.9032
깊은	0.3125	마디	-0.9024
덕분	0.2879	더구나	-0.9000
해주셔서	0.2821	왔다고	-0.9000
밝은	0.2692	하나고	-0.9000
드려요	0.2542	물었습니다	-0.8974

다음은, 이렇게 구축된 VOC 감성사전을 이용해 VOC 분류 예측 정확도 실험을 실시하였다. 앞서 추출된 감성어휘의 값들을 이용하여 VOC 각각의 감성극성을 계산하였다. VOC의 감성극성은 감성

어휘의 극성값 평균으로 계산되며, 이렇게 도출된 VOC의 극성값이 정해진 임계값보다 큰 경우 긍정 VOC로 예측하고 그렇지 않은 경우 부정 VOC로 예측하였다. 그리고 전체 예측 건수 중 긍정/부정을 정확하게 예측한 비율을 예측 정확도로 정의하였다. 이렇게 수행된 실험의 결과가 <Figure 4>에 나타나있다.



<Figure 4> Prediction Accuracy

<Figure 4>에서 나타나듯 VOC 극성값이 -0.50 일 때 가장 높은 분류 예측 정확도 77.8%를 나타내고 있다. 이는 VOC 극성값 -0.5에서 VOC의 오피니언이 긍정적인지 부정적인지 가장 잘 분류한다는 것을 의미한다. 다시 말하면, VOC 감성사전을 이용하여 새롭게 접수된 VOC를 자동 분류한다면 긍정 또는 부정 오피니언을 정확히 판별할 확률이 약 78%에 이른다는 것이다. 여기에 더해 VOC의 오피니언 극성 값까지를 고려한다면 극성값이 높은 VOC를 자동으로 분류 선별해 줌으로써 보다 신속한 고객 대응을 지원할 수 있을 것으로 판단된다.

5. 결론

본 연구에서는 텍스트 마이닝의 하나인 오피니언 마이닝 기법을 이용하여 고객의 다양한 의견을 담고 있는 VOC 데이터를 자동으로 분류하고 우선순

위의 제공하는 오피니언 극성값을 추출하는 지능형 VOC 분석시스템을 제안하였다. 또한 VOC 오피니언 분석시스템의 기준이 되는 주제지향 감성사전 개발 프로세스와 각 단계를 구체적으로 제시하였다. 그리고 본 연구에서 제시한 시스템의 효용성을 검증하기 위하여 의료기관 홈페이지에서 수집한 4,300여건의 VOC 데이터를 이용하여 병원에 특화된 감성어휘와 감성극성값을 도출하여 감성사전을 구축하고 이를 통해 구현된 VOC 분류 모형의 정확도를 비교하는 실험을 수행하였다. 이때 VOC의 오피니언을 분류하는 임계값이 중요함으로 임계값의 변화에 따라 VOC의 긍정/부정 오피니언의 예측 정확도가 변하는 양상을 살펴보고 최적의 임계값을 도출하였다.

실험의 결과로서 “칭찬, 친절함, 감사, 무사히, 잘해, 감동, 미소” 등의 어휘는 매우 높은 긍정 오피니언 값을 가지며, “통명, 뿔니까, 말하더군요, 무시하는” 등의 어휘들은 강한 부정의 극성값을 나타내고 있음을 확인하였다. 이는 일상생활에서도 매우 빈번히 사용되는 단어이면서 그 극성이 비교적 극명히 인지되는 어휘들로 도출된 어휘 극성값이 매우 개연성을 가지고 있음을 확인할 수 있었다. 또한 VOC의 긍정/부정 오피니언 판별의 예측 정확도를 검증한 결과, VOC 극성값이 0.50일 때 가장 높은 분류 예측 정확도 77.8%를 나타내고 있음을 확인하였고, 이는 VOC 감성사전을 이용하여 새롭게 접수된 VOC를 자동 분류한다면 긍정 또는 부정 오피니언을 정확히 판별할 확률이 약 78%에 이른다고 할 수 있다. 결국 본 지능형 VOC 분석시스템을 통해 VOC의 실시간 자동 분류 및 대응 우선순위를 도출하여 고객 민원에 대한 신속한 대응한다면, VOC 운용부서에서는 VOC 대응 인력을 보다 효율적으로 배치하면서도 고객민원의 강도에 따라 유연하면서도 신속한 대응이 가능할 것으로 기대된다. 또한

VOC 텍스트를 분석하고 활용할 수 있는 오피니언 마이닝 모형이라는 새로운 시도를 통해 향후 다양한 분석과 실용 프레임워크의 기틀을 제공할 수 있을 것으로 기대된다.

이러한 연구결과와 기대효과에도 불구하고 본 연구의 한계와 후속 연구의 필요성은 첫째, 본 연구를 위해 수집한 VOC 4,300여건은 의료기관 홈페이지를 통해서만 수집된 것으로 병원에 편중된 오피니언만을 다루었다는 한계와 둘째, 방문접수, 콜센터, SNS 등 VOC 접수 채널의 다양성을 확보하지 못한 점, 셋째, 보통의 VOC 데이터 건에 비해 4,300건은 그다지 많은 양이라 보기 어렵다는 점이다. 그러나 VOC를 대용량의 데이터, 즉 빅데이터 관점에서 오피니언 마이닝을 이용해 자동으로 유형을 분류하고 우선순위를 도출하는 지능형 VOC 분석시스템을 제시함으로써 보다 신속한 VOC 대응을 통한 고객 만족도 향상과 운용인력의 효율적 안배를 지원하는 것은 의미가 있다고 하겠다. 그럼으로 향후 연구에서는 다양한 채널에서의 VOC 데이터를 통해 분석의 품질을 높이고, 공공기관, 금융기관 등 다양한 분야에서의 비교연구를 통해 활용성을 높이는 후속 연구가 수행되어야 한다.

참고문헌

- Ghose, A., P. G. Ipeirotis, and A. Sundararajan, "Opinion Mining Using Econometrics : A Case Study on Reputation Systems," *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, (2007), 416~423.
- Cha, S., J. Kang, and J. Choi, "A Study on Social media Opinion Mining based Enterprise Crisis Management," *Proceedings of KIISE Conference*, Vol.39, No.1(2012), 142~144.
- Chen, H. and D. Zimbra, "AI and Opinion Mining," *IEEE Intelligent Systems*, Vol.25, Issue.3 (2010), 74~80.
- Choi, Y.-J. and H. Choi, "A Study on the Customer Satisfaction Strategies of the Online Company Using VOC," *Journal of Korean Industrial Economics and Business*, Vol.3, No.1(2011), 73~93.
- Hu, M. and B. Liu, "Mining and summarizing customer review," *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, 168~177.
- Conrad, J. G. and F. Schilder, "Opinion Mining in Legal Blogs," *Proceedings of the 11th ACM International Conference on Artificial Intelligence and Law*, 2007, 231~236.
- Ju, J.-M. and S.-G. Hwang, "Establishment of VOC analysis system for efficient CRM," *Journal of the Korean Society for Quality Management*, Vol.32, No.1(2004), 75~91.
- Kim, Y., "Case Analysis of Specific Unification CRM and BSC," *Journal of the Korea Service Management Society*, Vol.8, No.3(2007), 277~292.
- Kim, Y., *News Big Data Opinion Mining Model for Predicting KOSPI Movement*, Kookmin University Graduate School of Business IT, 2012.
- Kim, Y., N. Kim, and S. R. Jeong, "Stock-Index Invest Model Using News Big Data Opinion Mining," *Journal of Intelligence and Information Systems*, Vol.18, No.2(2012), 143~156.
- Takeuchi, H., L. V. Subramaniam., T. Nasukawa, and S. Roy, "Getting insights from the voices of customers : Conversation mining at a contact center," *Information Science*, Vol.179 Issue.11 (2009), 1584~1591.

- Yang, J.-Y., J. Myung, and S.-G. Lee, "A Sentiment Classification Method using Context Information in Product Review Summarization," *Journal of KIISE : Databases*, Vol.36, No.4(2009), 254~262.
- Yu, Y., Y. Kim., N. Kim, and S. R. Jeong, "Predicting the Direction of the Stock Index by Using a Domain Specific Sentiment Dictionary," *Journal of Intelligence and Information Systems*, Vol.19, No.1(2013), 92~110.
- Yune, H., H.-J. Kim, and J.-Y. Chang, "An Efficient Search Method of Product Review using Opinion Mining Techniques," *Journal of KIISE : Computing Practices and Letters*, Vol.16, No.2(2010), 222~226.
- Zhuang, L., F. Jing, and X. Y. Zhu, "Movie Review Mining and Summarization," *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, (2006), 43~50.

Abstract

Intelligent VOC Analyzing System Using Opinion Mining

Yoosin Kim^{*} · Seung Ryul Jeong^{**}

Every company wants to know customer's requirement and makes an effort to meet them. Cause that, communication between customer and company became core competition of business and that important is increasing continuously. There are several strategies to find customer's needs, but VOC (Voice of customer) is one of most powerful communication tools and VOC gathering by several channels as telephone, post, e-mail, website and so on is so meaningful. So, almost company is gathering VOC and operating VOC system.

VOC is important not only to business organization but also public organization such as government, education institute, and medical center that should drive up public service quality and customer satisfaction. Accordingly, they make a VOC gathering and analyzing System and then use for making a new product and service, and upgrade. In recent years, innovations in internet and ICT have made diverse channels such as SNS, mobile, website and call-center to collect VOC data.

Although a lot of VOC data is collected through diverse channel, the proper utilization is still difficult. It is because the VOC data is made of very emotional contents by voice or text of informal style and the volume of the VOC data are so big. These unstructured big data make a difficult to store and analyze for use by human. So that, the organization need to automatic collecting, storing, classifying and analyzing system for unstructured big VOC data.

This study propose an intelligent VOC analyzing system based on opinion mining to classify the unstructured VOC data automatically and determine the polarity as well as the type of VOC. And then, the basis of the VOC opinion analyzing system, called domain-oriented sentiment dictionary is created and corresponding stages are presented in detail. The experiment is conducted with 4,300 VOC data collected from a medical website to measure the effectiveness of the proposed system and utilized them to develop the sensitive data dictionary by determining the special sentiment vocabulary and their

* Digital-Data and Analytics, Technology, Accenture Korea

** Corresponding Author: Seung Ryul Jeong

Graduate School of Business IT, Kookmin University

77 Jeongneung-ro, Seongbuk-gu, Seoul 136-702, Korea

Tel: +82-2-910-4018, Fax: +82-2-910-4017, E-mail: srjeong@kookmin.ac.kr

polarity value in a medical domain.

Through the experiment, it comes out that positive terms such as “칭찬, 친절함, 감사, 무사히, 잘해, 감동, 미소” have high positive opinion value, and negative terms such as “통명, 뭉니까, 말하더군요, 무시하는” have strong negative opinion. These terms are in general use and the experiment result seems to be a high probability of opinion polarity. Furthermore, the accuracy of proposed VOC classification model has been compared and the highest classification accuracy of 77.8% is conformed at threshold with -0.50 of opinion classification of VOC.

Through the proposed intelligent VOC analyzing system, the real time opinion classification and response priority of VOC can be predicted. Ultimately the positive effectiveness is expected to catch the customer complains at early stage and deal with it quickly with the lower number of staff to operate the VOC system. It can be made available human resource and time of customer service part.

Above all, this study is new try to automatic analyzing the unstructured VOC data using opinion mining, and shows that the system could be used as variable to classify the positive or negative polarity of VOC opinion. It is expected to suggest practical framework of the VOC analysis to diverse use and the model can be used as real VOC analyzing system if it is implemented as system.

Despite experiment results and expectation, this study has several limits. First of all, the sample data is only collected from a hospital web-site. It means that the sentimental dictionary made by sample data can be lean too much towards on that hospital and web-site. Therefore, next research has to take several channels such as call-center and SNS, and other domain like government, financial company, and education institute.

Key Words : VOC, Voice of Customers, Opinion Mining, Sentimental Analysis, Text Mining, Big Data

저자 소개



김유신

국민대학교 정보관리학과에서 학사, 비즈니스IT전문대학원에서 경영정보학 석사 및 박사를 취득하였고, 현재 액센츄어 코리아에서 데이터 사이언티스트로 재직 중이다. 금융과 의료, e비즈니스 분야에서 다수의 IT 컨설팅 및 정보시스템 구축 프로젝트를 수행하였으며, 현재는 빅데이터 디스커버리, 소셜 미디어 인텔리전스, 텍스트 오피니언 마이닝, 고객 세분화와 추천 알고리즘 개발 등의 Analytics 분야에 주력하고 있다.



정승렬

서강대학교에서 경제학사, 미국 위스컨신 대학에서 경영정보학 석사, 그리고 사우스 캐롤라이나 대학에서 경영정보학 박사를 취득하였다. 현재 국민대학교 경영정보학부 및 비즈니스IT전문대학원 교수로 재직 중이며, Journal of MIS, Communications of the ACM, Information and Management, Information Systems Management, Journal of Systems and Software, Lecture Notes on Computer Science, APJIS, 경영과학, 한국 경영과학회지, ISR, 정보처리학회지 등의 국내외 저널에 프로세스 관리, ERP, 정보자원관리, 시스템 구현, 정보시스템 감리 등의 주제와 관련하여 많은 논문을 발표하였다.