

정서 단어 부정어가 정서가의 극성 전환 및 약화에 미치는 영향*

이 신 영 함 준 석 김 미 선 방 그 린 고 일 주[†]
숭실대학교 미디어학과

오피니언 마이닝과 정서 분석에 대한 기존 연구에서는 정서 단어에 부정어를 붙일 경우 긍정, 부정의 극성과 값이 뒤바뀐다고 가정하고 부정어를 처리하였다. 그러나 지금까지 정서 단어에 부정어가 발생했을 때 극성이 어느 정도 바뀌는지에 대한 정량적 연구는 없었다. 따라서 본 논문에서는 한국어 정서 단어와 그 부정형에 대해서 정서가와 각성 차원을 측정하였다. 결과, 정서 단어에 부정형이 올 경우 정서가와 각성 차원의 중간 수준을 기준으로 극성을 뒤바꾸고 값은 약 30~50% 약화되었다. 이 결과를 오피니언 마이닝과 정서 분석 연구에서 부정어를 처리하는 기준으로 제시하였다.

주제어 : 정서 단어, 부정어, 오피니언 마이닝, 정서 분석, 정서가, 각성

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0016540).

[†] 교신저자: 고일주, 숭실대학교 글로벌미디어학부, 연구 세부분야: 전산학
E-mail: andy@ssu.ac.kr

서 론

최근 소셜 네트워크 서비스(SNS)의 급격한 발달과 함께 SNS, 블로그, 댓글 등의 텍스트 데이터로부터 사용자의 의견을 수집하는 오피니언 마이닝이 중요한 이슈가 되고 있다. 오피니언 마이닝은 텍스트로 이루어진 사용자 정보를 수집하여 제품, 영화, 기업 이미지, 정치인에 대한 의견을 긍정, 부정으로 분류한 후 제품 판매량을 늘리거나, 이미지를 개선시키는 등 위험 관리를 하는 것을 주요 목표로 한다. 한편 정서 분석(sentiment analysis)은 오피니언 마이닝의 세부 분야 중 하나다. 오피니언 마이닝이 주로 긍정, 부정의 극성(polarity)을 분류한다면 정서 분석은 텍스트에서 긍정, 부정을 포함해 기쁨, 행복, 슬픔, 혐오 등과 같은 세분화된 정서의 흐름을 분석한다.

단어의 긍정, 부정을 추출할 때에 not, never, 안, 않다 등의 부정어가 쓰이게 되면 단어의 긍정, 부정 정도가 크게 달라진다. 따라서 단어에 부정어가 쓰인 것을 적절히 처리하는 것은 오피니언 마이닝에 있어서 중요하다. 또한 정서 분석에서도 부정어에 따라 분류되는 정서에 영향을 미치기 때문에 부정어에 대한 처리가 필요하다. 정서 분석에서 많이 쓰이는 모형으로 정서의 차원 이론이 있다. 정서의 차원 이론은 정서 단어를 정서가(valence)와 각성(arousal)차원으로 분리하여 측정하고 이것을 정서 분석에 사용한다. 정서가 차원은 긍정, 부정 등 극성을 의미하며 쾌(pleasant)-불쾌(unpleasant) 차원과 동일한 의미로 사용된다. 각성 차원은 각성(arousal)-수면(sleep), 활성화(activation)-비활성(deactivation)과 같은 의미로 사용된다[1-7].

부정어의 영향에 대한 기존의 연구들은 단어에 부정어가 발생했을 때 단어의 극성을 반전시킬 것이라는 가정을 하였다[8-11]. 그러나 부정어가 과연 극성을 반전시키는지, 또 그 값은 어느 정도인지에 대한 정량적인 평가가 없었다. 따라서 본 논문에서는 한국어 정서 단어와 부정형에 대해 정서가를 측정하였다. 또한 부정어가 각성 차원에 미치는 영향도 측정하였다. 측정 결과를 토대로 정서 분석에서 부정어의 처리를 보다 정확하게 하기 위한 기준을 제시할 것이다.

본 논문의 구성은 다음과 같다. 먼저 오피니언 마이닝과 정서 분석에서 부정어에 대한 처리를 한 기존의 관련 연구들을 살펴본다. 다음으로 한국어 정서 단어의 부정형에 대한 측정 방법과 결과를 기술한다. 마지막으로 결론을 기술하고 부정어

가 한국어 정서 단어에 미치는 영향에 관한 논의를 한다.

관련 연구

Polanyi와 Zaenen[8]은 문맥 정서가 이동기(contextual valence shifters)를 제안하여 부정어를 처리하였다. 부정어가 발생했을 경우에 단순히 극성을 반전시켰는데, 예를 들면 clever(+2) → not clever(-2)와 같다. 또한 감쇠기(diminisher)를 사용하여 극성을 바꾸지 않고 점수만 줄이기도 하였는데, 예를 들면 efficient(+2) → rather efficient(+1)과 같다.

Hu와 Liu[9]는 고객 평가글에 대한 텍스트 마이닝과 문서 요약을 하였다. 문장의 의견이 긍정인지 부정인지 결정하기 위해서 단순한 규칙을 사용하였는데, 의견이 있는 단어 근처에 부정어가 나오면 단순히 긍정, 부정의 극성을 바꾸었다.

Jia, Yu와 Meng[10]은 문장에서 부정어가 발생했을 때 의견의 극성을 결정하기 위해서 부정어 처리를 하였다. 또한 세 가지 부정어 범위(scope of negation)의 예외를 설정하여 부정어가 발생했지만 부정어 처리를 하지 않는 경우를 설정했다. 첫째, ‘not only’, ‘not just’, ‘not to mention’과 같이 부정어가 발생했지만 부정의 의미가 아닌 경우, 둘째, 부정적 부가의문문일 경우, 마지막으로 조동사 뒤에 부정어가 따르는 경우는 부정의 의미가 아니라고 보고 부정어 범위 예외로 처리하였다. 이 연구 또한 부정어가 발생한 단어는 단순히 극성을 바꾸어서 처리하였다.

정유철, 최윤정, 맹성현[11]은 블로그 문서를 정서 기반으로 분류하면서 부정어 처리를 하였다. 정서 단어로서 ANEW(Bradley & Lang, 1999) 목록을 사용하였다. ANEW 목록에는 1000여개의 정서 단어에 대해 P(pleasure), A(arousal), D(dominance)의 차원값이 설정되어 있다. 여기서 P와 A 차원은 각각 정서가와 각성 차원과 같은 의미이다. PAD값을 [-1, +1]로 정규화한 후, 정서 단어 앞에 no, never, not과 같은 부정어가 포함되었을 경우 P와 A의 극성을 단순 반전시켰다. 자체적인 실험 결과 부정어에 대해 D값을 반전시키면 분류율을 떨어뜨리기 때문에 D의 반전은 제외하였다.

이처럼 정서 분류에서 부정어 처리를 위한 대부분의 연구들은 부정어가 발생했

을 경우 정서 단어의 극성의 값을 단순히 반전시켰다. 그러나 단순 반전만으로는 한계가 있다. 예를 들어 ‘행복하다’는 강한 긍정의 의미이지만 부정형인 ‘행복하지 않다’가 ‘행복하다’와 같은 값을 가지면서 강한 부정의 의미는 아니다. 또한 ‘행복하지 않다’가 ‘행복하다’의 반의어인 ‘불행하다’의 의미도 아니다. 특히 ANEW 목록과 같이 정서 단어를 정서가, 각성의 차원으로 분류한 데이터에서 부정어가 발생했을 경우 단순히 정서가, 각성 차원값을 반전시키는 것이 적절하지 않다.

한국어 정서 단어 부정형의 정서가, 각성 차원 측정

그림 1은 박인조와 민경환[3]의 한국어 정서 단어 목록 중 일부를 정서가와 각성 차원에 표시한 그림이다. 기존의 연구들은 부정어가 발생하였을 때 정서가, 각

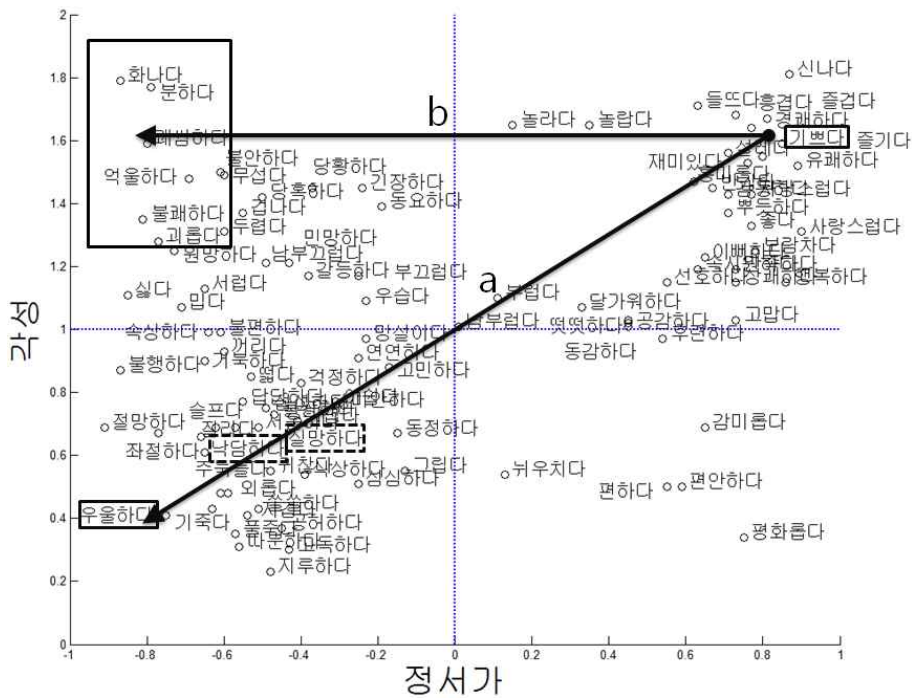


그림 1. 한국어 정서 단어의 정서가, 각성 차원 분포

성 차원의 값을 모두 반전시키거나(그림 1의 a) 정서가 차원의 극성만 반전시켰다(그림 1의 b). 그림 1의 우측 상단의 ‘기쁘다’를 이러한 부정어 처리 가정에 적용하여 정서가와 각성 차원 모두 변하는 a와 같은 변환을 하면 ‘기쁘다’의 부정형은 ‘우울하다’이어야 한다. 또한 정서가 차원의 극성만 반전시키는 b와 같은 변환을 하면 ‘기쁘다’의 부정형은 ‘화나다’, ‘분하다’, ‘괴씸하다’, ‘억울하다’, ‘불쾌하다’, ‘괴롭다’ 중 하나여야 한다. 그러나 이러한 단어들은 ‘기쁘다’의 부정형인 ‘안 기쁘다’, ‘기쁘지 않다’와는 의미가 많이 다르다. 오히려 그림 1의 좌측 중하단의 점선으로 표시한 ‘실망하다’, ‘낙담하다’와 유사하다. 이러한 관찰을 토대로 정서 단어 부정형의 정서가, 각성 차원의 값은 반전되면서 값도 같이 줄어든다는 것을 예상할 수 있다. 따라서 본 논문에서는 이러한 가설을 검증하기 위해 한국어 정서 단어와 그 단어의 부정형에 대해서 각각 정서가, 각성 차원의 값을 측정하였다.

참가자

수도권 소재 대학교의 대학생 50명(남성 20명, 여성 30명)으로 구성하였다. 평균 나이는 25.1세(SD=2.73)였다. 설문에 소요된 시간은 약 20분이었으며 참가자는 자율적인 신청을 받아 설문에 참여하였다. 설문이 완료된 후 문화상품권으로 보상이었다.

자료

자료는 박인조와 민경환[31]의 한국어 정서 단어 목록의 일부를 사용하였다. 이 목록은 연세대학교 언어정보 개발연구원의 ‘현대 한국어 어휘 빈도 자료집’을 기초로 하여 최근 10년 동안에 출판된 신문, 잡지, 소설, 수필, 취미, 교양, 수기, 전기, 국어 교과서, 희곡, 시나리오 등을 대상으로 434개의 정서 단어 목록을 작성한 것이다. 이 목록에는 ‘궁지’, ‘의혹’과 같은 부정어를 붙이기 힘든 명사형 정서 단어와, ‘기겁다’, ‘셀쪽하다’와 같이 잘 쓰이지 않는 단어들이 포함되어 있기 때문에 본 논문에서는 434개 중 부정어를 붙였을 때 친숙한 101개의 단어를 다시 선정하였다. 단어 선정을 위해서 숭실대학교 인공 정서 연구실에서 정서 연구를 하는 박

사과정 연구원 4명이 434개 정서 단어의 부정형에 대해 친숙 정도에 따라 0~3점의 점수를 부여하였고 평균 2점 이상의 점수를 받은 정서 단어 101개를 최종 선정하였다.

절차

먼저 101개 정서 단어에 대해서 정서가와 각성 수준을 측정하였다. 정서가와 각성의 개념이 일반인에게는 어려우므로 설문 참가자에게 정서가와 각성의 개념을 설명하였다. 제시되는 정서 단어에 대한 느낌이 “긍정적인, 행복한, 기쁜, 만족한, 희망적인”인지 아니면 “부정적인, 불행한, 짜증나는, 불만족한, 우울한, 절망적인”지 점수를 부여하게 하였다. 그리고 각성 수준을 측정하기 위해서 “흥분한, 자극적인, 열광적인, 각성된, 긴장한”인지 아니면 “침착한, 이완되는, 고요한, 지루한, 졸린”인지 점수를 부여하게 하였다. 정서가는 긍정, 부정의 의미이고 이것은 양수, 음수의 개념과 유사하므로 +3에서 -3점까지 7점 척도로 선택하게 하였다. 또한 각성은 높고 낮은 상태만 있으므로 1점에서 7점까지 7점 척도로 선택하게 하였다. 이어서 ‘기쁘지 않다’, ‘슬프지 않다’와 같이 앞서 측정한 정서 단어의 부정형에 대해서 정서가와 각성 수준을 같은 방식으로 측정하였다. 측정된 정서가와 각성 값은 다시 각각 [-1, +1]와 [0, +2]로 정규화하였다.

결과

그림 2는 정서가가 양수인 긍정 정서 단어와 부정형 값의 변화를 나타낸다. ‘o’ 기호는 정서 단어의 정서가, 각성 차원의 값을 나타내며 ‘x’ 기호는 부정형일 때 변화한 값을 나타낸다. ‘◆’ 기호는 각각의 군집의 중앙을 나타낸다. 긍정 정서 단어의 정서가 평균은 0.66(SD=0.17)이고 각성 평균은 1.24(SD=0.38)이었다. 긍정 정서 단어 부정형의 정서가 평균은 -0.48(SD=0.12)이고 각성 평균은 0.73(SD=0.2)이었다. 긍정 정서 단어의 부정형은 정서가의 극성이 바뀌어 정서가가 음수인 영역으로 측정되었다. 그러나 정서가 값은 완전히 반대가 아니고 각 군집의 중앙을 기준으로 볼 때 약 30%가 감소하였다. 또한 각성 차원을 기준으로 볼 때 각성 값이

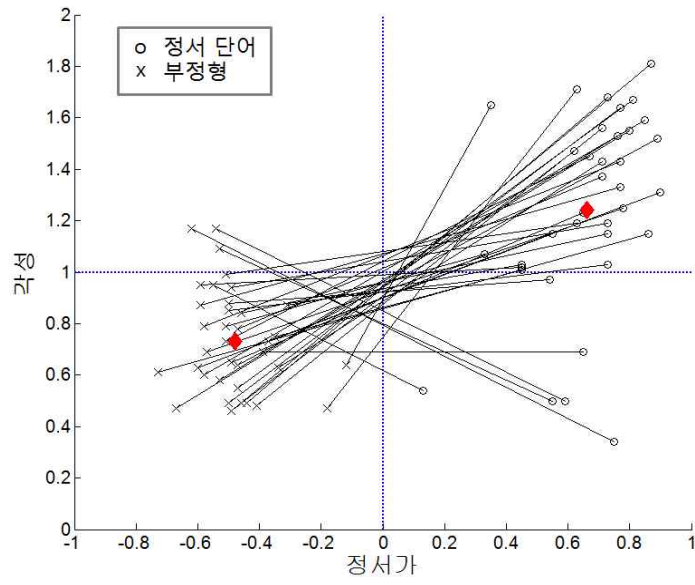


그림 2. 긍정 정서 단어와 부정형의 정서가, 각성 차원값

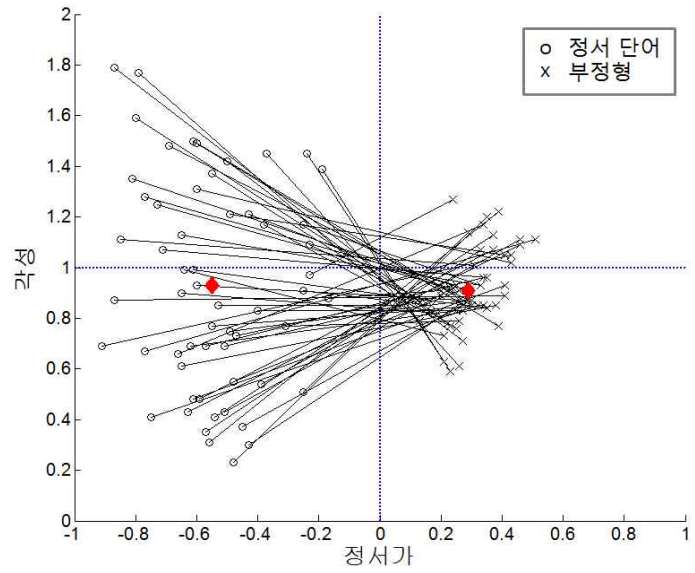


그림 3. 부정 정서 단어와 부정형의 정서가, 각성 차원값

높은 단어는 각성 값이 낮아졌고, 각성 값이 낮은 단어는 오히려 각성 값이 높아졌음을 알 수 있다. 즉, 각성 차원의 중간 수준(약 0.9)을 기준으로 바뀌지만 값도 역시 약화되었다. 그림 3은 정서가가 음수인 부정 정서 단어와 부정형의 변화를 나타낸다. 부정 정서 단어의 정서가 평균은 -0.55(SD=0.19)이고 각성 평균은 0.93(SD=0.41)이었다. 부정 정서 단어 부정형의 정서가 평균은 0.29(SD=0.1)이고 각성 평균은 0.91(SD= 0.15)이었다. 마찬가지로 부정 정서 단어의 부정형도 정서가의 극성이 바뀌어서 양수인 영역으로 측정되었지만 각 군집의 중앙을 기준으로 볼 때 값은 약 50%가 감소하였다. 또한 각성 차원도 마찬가지로 각성의 중간 수준(약 0.9)을 기준으로 값이 바뀌면서 감소하였다. 이 결과를 바탕으로 긍정 정서 단어, 부정 정서 단어, 그리고 각각의 부정형의 군집의 중앙값을 중심으로 수식을 구성하면 식1과 같다.

$$\begin{aligned}
 valence_{Negation} &= \begin{cases} -0.48 - 0.12(2valence - 1) & , valence \geq 0 \\ 0.29 - 0.1(2valence - 1) & , valence < 0 \end{cases} \\
 arousal_{Negation} &= \begin{cases} 0.73 - 0.2(arousal - 1) & , valence \geq 0 \\ 0.91 - 0.15(arousal - 1) & , valence < 0 \end{cases}
 \end{aligned}
 \tag{식 1}$$

여기서 *valence*와 *arousal*은 각각 정서 단어의 정서가와 각성 값이고, $valence_{Negation}$ 과 $arousal_{Negation}$ 은 각각 정서 단어 부정형의 정서가와 각성 값이다. 정서 단어 부정형의 정서가 값($valence_{Negation}$)은 긍정 정서 단어($valence \geq 0$)일 때 -0.48 ± 0.12 의 범위로 이동하며, 부정 정서 단어($valence < 0$)일 때 0.29 ± 0.1 의 범위로 이동하는 것을 의미한다. 즉, 긍정 정서 단어를 부정하면 정서가 수준은 약 -0.48의 부정 의미로 값은 약화되며 극성이 바뀔을 의미하며, 부정 정서 단어를 부정하면 정서가 수준은 약 0.29의 긍정 의미로 값은 약화되며 극성이 바뀔을 의미한다.

또한 정서 단어 부정형의 각성 값($arousal_{Negation}$)은 긍정 정서 단어($valence \geq 0$)일 때 0.73 ± 0.2 의 범위로 이동하며, 부정 정서 단어($valence < 0$)일 때 0.91 ± 0.15 의 범위로 이동하는 것을 의미한다. 즉, 긍정 정서 단어를 부정하면 각성 수준은 0.73 근방으로 수렴하고, 부정 정서 단어를 부정하면 각성 수준은 0.91 근방으로 수렴함을 의미한다. 이 수식의 의미는 결국 정서 단어의 부정형은 정서가, 각성 수준의

중간 수준을 기준으로 극성이 바뀌면서 값이 약화됨을 의미한다.

결론 및 논의

본 논문에서는 오피니언 마이닝과 정서 분석 연구에서 부정어가 발생했을 경우 긍정, 부정의 극성이 바뀐다는 기존 연구들의 가정을 검증하기 위해 한국어 정서 단어와 정서 단어의 부정형에 대해 정서가와 각성 수준을 측정하였다. 그 결과 정서 단어의 부정형은 정서가와 각성 차원의 중간 수준을 기준으로 극성이 바뀌는 결과를 얻었다. 그러나 두 차원의 값은 모두 정확히 반대의 값은 아니었고 약화되는 결과를 얻었다. 즉, ‘기쁘다’는 긍정 값이 매우 큰 단어인 반면 ‘기쁘지 않다’는 부정 값이 크지 않다. 특히 정서가 차원을 기준으로 봤을 때 긍정 정서 단어는 약 30%, 부정 정서 단어는 약 50%로 값이 감소하였다. 긍정 단어의 부정형보다 부정 단어의 부정형을 상대적으로 더 부정적으로 평가하는 경향이 있었다. 예를 들면, ‘기쁘지 않다’는 꽤 부정적으로 느끼지만 ‘화나지 않다’는 별로 긍정적으로 느끼지 않는다는 의미이다. 결과적으로 오피니언 마이닝과 정서 분석 연구에서 부정어가 발생했을 때 긍정, 부정의 극성을 바꾸어서 처리하되 강도는 긍정 단어의 경우 약 30%, 부정 단어의 경우 약 50% 약화시키는 것이 적절하다. 정서 단어에 대한 정서가, 각성 차원의 측정값은 언어, 문화, 집단마다 조금씩 차이가 있기 때문에 이것을 설명하기 위해서는 다른 문화와 언어 사용자에 대한 후속 연구가 필요하다. 또한 긍정, 부정 정서 단어 모두 부정형의 군집의 크기가 원래 정서 단어들의 군집의 크기보다 작아졌다. 이것은 인간은 넓은 범주의 정서가와 각성 차원의 정서 단어를 표상하고 있지만, 문장 안에서 부정어로 인한 정서 단어의 부정형에 대해서는 좁은 범주로 표상함을 의미한다.

참고문헌

[1] 김영아, 김진관, 박수경, 오경자, 정찬섭 (1998). 정서관련 어휘 분석을 통한 내

- 적 상태의 차원 연구. **감성과학**, 1(1), 145-152.
- [2] 김진관, 문혜신, 오경자 (1998). 감성 개념 이차원 구조의 안정성. **감성과학**, 2(1), 43-52.
- [3] 박인조, 민경환 (2005). 한국어 감성단어의 목록 작성과 차원 탐색. **한국심리학 회지: 사회 및 성격**, 19(1), 109-129.
- [4] 유 상, 조정자, 한광희 (2006). 손동작과 정서 차원 분석. **감성과학**, 9(2), 119-132.
- [5] Bradley, M., Lang, P. (1999). Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- [6] Russell, J. (1978). Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology*, 36, 1152-1168.
- [7] Russell, J., Barrett, L. (1999). Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant. *Journal of Personality and Social Psychology*, 76(5), 805-819.
- [8] Polanyi, L., Zaenen, A. (2004). Context Valence Shifters. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- [9] Hu, M., Liu, B. (2004). Mining and summarizing customer reviews, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 168-177.
- [10] Jia, L., Yu, C., Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness, *Proceedings of the 18th ACM conference on Information and knowledge management*, 1827-1830.
- [11] 정유철, 최윤정, 맹성현 (2008). 감정 기반 블로그 문서 분류를 위한 부정어 처리 및 단어 가중치 적용 기법의 효과에 대한 연구, **인지과학**, 19(4), 477-497.

1 차원고접수 : 2012. 2. 13

2 차원고접수 : 2012. 3. 14

최종게재승인 : 2012. 3. 16

(*Abstract*)

The effect of negated emotional words on polarity reversal and weakening value in valence

Shin-Young Rhee Jun-Seok Ham Mi-Sun Kim Green Bang Il-Ju Ko

Department of Media, Soongsil University

Previous studies on opinion mining and sentiment analysis have supposed that the polarity and value of an emotional word is reversed when a negation word is attached. However, there are no quantitative studies on how much the polarity is changed when a negation word is following. Therefore, we measured the valence and arousal dimensions for Korean emotional words and their negations. Consequently, the polarity of valence and arousal was reversed on their intermediate level. Also, the value was reduced by about 30% to 50%. We propose this result as a guideline for processing negation words for studies on opinion mining and sentiment analysis.

Key words : emotional word, negation word, opinion mining, sentiment analysis, valence, arousal