

## 소셜 데이터의 주된 감성분석에 대한 연구

장 필 식\*

### Study on Principal Sentiment Analysis of Social Data

Phil-Sik Jang\*

#### 요 약

본 논문에서는 대용량의 문서, 인터넷 댓글, 소셜 데이터, 메시지 텍스트 등으로부터 표준, 일상적 언어, 및 은어(隱語), 비속어, 약어, 이모티콘 등을 감성 분석함으로써, 복합적인 감성 중 근간이 되는 주 감성들을 측정하고 평가하는 방법을 제안한다. 제안된 방법론은 IRLBA(Implicitly Restarted Lanczos Bidiagonalization Algorithm)을 활용하여 규모가 큰 희소행렬에 대한 주성분분석을 실시하며, 데이터 취합, 메시지 분석, 감성 평가, 감성 분석 및 통합 그리고 결과물 시각화 모듈로 구성된다. 본 연구를 통해 제안된 방법론은 소셜 데이터의 감성분석의 정확도를 향상시키고 감성분석의 활용범위를 확장시키는데 있어 도움을 줄 수 있을 것으로 기대된다.

▶ Keywords : 소셜 데이터, 빅데이터, 주된 감성분석, SNS, 트위터

#### Abstract

In this paper, we propose a method for identifying hidden principal sentiments among large scale texts from documents, social data, internet and blogs by analyzing standard language, slangs, argots, abbreviations and emoticons in those words. The IRLBA(Implicitly Restarted Lanczos Bidiagonalization Algorithm) is used for principal component analysis with large scale sparse matrix. The proposed system consists of data acquisition, message analysis, sentiment evaluation, sentiment analysis and integration and result visualization modules. The suggested approaches would help to improve the accuracy and expand the application scope of sentiment analysis in social data.

▶ Keywords : Social data, Big data, Principal sentiment analysis, SNS, Twitter

---

•제1저자 : 장필식

•투고일 : 2014. 11. 19, 심사일 : 2014. 11. 23, 게재확정일 : 2014. 12. 19.

\* 세한대학교 정보물류학과(Dept. of IT and Logistics, Sehan University)

※ 이 논문은 2014년도 세한대학교 교내연구비 지원에 의하여 쓰여진 것임

## I. 서론

최근 큰 화두로 떠오르고 있는 '빅데이터'는 기대되는 큰 파급효과로 장밋빛 미래에 대한 예측이 주류를 이루고 있으며 [1-4], 이와 더불어 빅데이터에 대한 글로벌 기업등과 각국의 관심 또한 크게 늘고 있다[5-6]. 지금까지 도입된 빅데이터 기술들이 초점을 맞추고 있는 것은 대부분 검색알고리즘이나 프로세스 기술 및 기반 플랫폼 구축이며, 빅데이터 중 트위터, 블로그 등 소셜 미디어에 대한 분석이 주류를 이루고 있다. 그런데, 최근 들어 이러한 빅데이터 분석에 대한 실제 효용성과 맹신에 대해서는 차츰 의문이 제기 되고 있다. 즉, 현재의 빅데이터와 소셜미디어 데이터에 대한 기계적인 분류 분석은 무가치한 결과 양산할 수 있으며, 중요한 것은 구조화되지 않은 대규모 데이터 속에서 숨겨진 패턴을 찾아내는 것이다[5]. 빅데이터가 생산-소비되는 단계별로 나누어 본다면, 지금까지 대부분 연구들의 초점은 아직까지 빅데이터의 생산(production)과 관리(management)에 맞추어져 있다. 하지만 빅데이터가 실제 환경에서 효용성을 가지려면 소비단계(consumption)인 데이터 마이닝(mining), 데이터 애널리틱스(analytics) 등에서 실제적인 연구결과들이 도출되어야 할 것으로 생각된다.

이들 중 최근 주목받고 있는 새로운 분야가 감성분석(sentiment analysis)인데, 비즈니스 인텔리전스, 고객서비스, 브랜드 평판 등 여러 분야에의 활용이 모색되고 있다. IT 시장조사업체인 가트너(Gartner)의 Bryan Prentice 부사장[7]이 언급한대로, 유행처럼 데이터 수집에만 열을 올리면 결국 쓸모없는 데이터만 손안에 남을 것이며, 감성까지 헤아리는 감성분석을 시도해야 정치인이나 기업 등 빅 데이터를 활용하려는 사람들이 실질적인 도움을 받을 수 있게 된다. 본 연구에서는 기존의 감성분석 방법론에 대해 고찰하고 이들 방법론의 단점을 극복할 수 있는 방안과 구현방법에 대해 살펴보고자 한다. II장에서는 본 연구와 관련된 기존 연구들 즉 감성분석과 감성공학 접근방법 들을 소개하며, III 장에서는 본 연구에서 제안하는 시스템의 구성과 구조에 대해 살펴본다. 그리고 IV 장에서는 제안시스템의 특성과 향후 활용 방안에 대해 논의한다.

## II. 관련 연구

### 1. 감성분석(sentiment analysis)

감성분석이란 어떤 사안이나 인물, 이슈, 이벤트에 대한

사람들의 의견이나 평가, 태도, 감정 등을 분석하는 것을 말하며, 오피니언 마이닝(opinion mining)이나 평판 분석이라고도 불린다. 감성분석을 이용하면 소셜 미디어와 같은 온라인상의 여론을 신속하게 파악할 수 있게 되는데, 기존의 오픈라인 여론조사에 비해서 시간과 비용을 줄이고 사람들의 의견을 쉽게 파악하고 예측할 수 있어 그 활용도가 매우 높은 분야로 평가받고 있다[8].

이러한 감성분석은 현재 다양한 분야에 활용이 모색되고 있다. 2011년 11월 미국 슈퍼볼 경기 중 기아자동차의 광고 효과를 분석하는데 이용되었으며, 기존 고객 감성분석 기법보다 한 단계 진화하였다고 평가[9] 받는 미국 WiseWindow사의 MOBI(그림 1 오른쪽)는 다양한 소셜미디어를 분석 대상으로 하며, 긍정과 부정의 범주로 감성을 분류한다. 포레스터 리서치가 발표한 최근 보고서에서 '빅데이터 예측 분석 솔루션' 부문 1위 업체에 오른[10] SAS의 Sentiment Analysis(그림 1 왼쪽)는 감성을 긍정/부정/중립으로 분류한다[11].

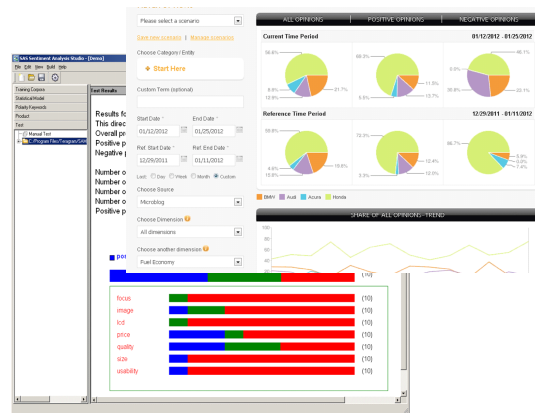


그림 1. Sentiment Analysis(SAS)와 MOBI(WiseWindow)  
Fig. 1. Sentiment Analysis(SAS) and MOBI(WiseWindow)

국내에서는 주로 트위터, 페이스북 등 SNS를 대상으로 감성분석을 실시하고 있으며, 블로그, 웹 카페 댓글 등을 대상으로 하는 경우도 있는데, 모두 긍정/부정/중립 또는 긍정/부정 범주로 분류된 감성을 최종 결과로 제시하고 있다(그림 2).

이러한 국내의 감성기술들은 다음 두 가지 측면에서 문제점을 가지고 있다.

첫째, 단순히 긍정, 부정 두 개 범주 또는 긍정, 부정, 중립의 세 개 범주로 나누는 것은 활용에 있어 큰 제약으로 작용하며, 분석의 정확성을 떨어뜨리게 된다. 예를 들어, '복

고풍의 ‘신세대 감각의’, ‘여성스러운’-‘남성스러운’ 등의 감성은 특정 상품이나 인물에 대한 이미지 파악에 있어 중요한 감성이지만, 단순히 ‘중립’으로 분류되며, 세부 감성은 파악이 불가능하다.

패션 산업의 경우에 빈티지 패션(vintage fashion)의 일부로 ‘촌티 패션’이 존재하며, 2000년 전후로 연예인을 필두로 복고풍의 촌티패션이 유행하였다. 따라서 이러한 관련 업체, 디자이너에게는 ‘촌스럽다’는 감성이 ‘부정적’으로 해석되지 않으며, 긍정, 부정 비율이 중요한 것이 아니라 ‘촌스럽다’ 또는 ‘복고풍의’ ‘에스러운’ 등과 같은 비슷한 범주의 감성이 어느 정도 나타나고 있는지가 중요하게 된다. ‘친숙하다’라는 감성도 ‘긍정적’ 범주로 분류되었지만 그 대상이 여자 아이돌 가수의 ‘몸매’라면 이야기가 달라진다. (몸매가 친숙하다는 표현은 일반인의 몸매와 다름이 없으며, 전혀 연예인 같지 않다는 의미가 된다) 물론 자연언어처리의 구분분석단계에서 대상, 또는 주어에 따라 긍정/부정을 다시 평가하도록 할 수도 있겠지만, 연예기획사 입장에서는 긍정/부정 보다 ‘친숙한’, ‘푸근한’ 등의 감성과 ‘섹시한’ 등의 감성이 어느 정도 비율인지, 어떤 감성이 주를 이루는지가 중요하게 된다.

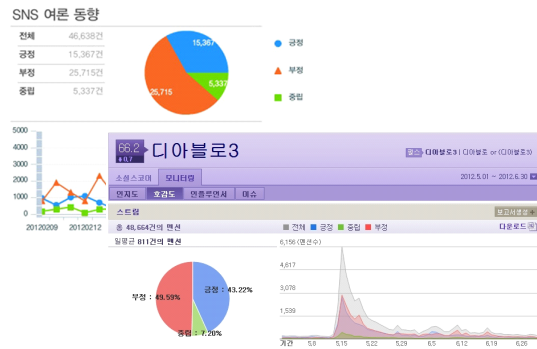


그림 2. 다음소프트와 펄스K의 감성분석  
Fig. 2. Sentiment Analysis of Daumsoft, PulseK

이렇게, 기존 연구, 기술들에서 기본적으로 극성의 판단이 불가능하거나 애매한 감성들과, 사람에 따라 긍정/부정이 다르게 해석될 수 있는 감성들을 두 개(긍정/부정) 또는 세 개(긍정/부정/중립)의 범주에 무리하게 끼워 넣음으로써, 감성 분석의 정확성은 당연히 떨어질 수밖에 없다. 실제로 다양한 자연언어처리 기술을 적용함에도 불구하고, 국내 감성분석의 정확도는 70%~80% 정도인 것으로 평가되고 있으며[12], 국외 감성분석의 정확도 또한 그리 높지 않은 것으로 알려져 있다. 즉, 최종 사용자 입장에서 중요한 것은, 파악하고자 하

는 대상에 대한 감성이 긍정적인가 부정적인가 보다는, 주된 감성이 어떤 것들인지가 중요하므로, 이 구간이 되는, 주된 세부 감성을 추출하는 것이 감성분석의 목적이 되어야 한다.

둘째, 기존 감성분석 기술들에서는 비속어, 은어(隱語), 약어, 이모티콘 등은 필터링하여 제외하고, 표준 어휘만 분석하고 있다. 하지만, SNS 나 홈페이지 댓글, 블로그 등에 사용자들이 올리는 많은 글(텍스트) 들은 구어 및 은어 비속어, 이모티콘을 포함하고 있으며, 이들 텍스트들이 필터링 됨으로써, 다양한 감성을 포함하는 데이터 들이 감성분석과정에서 제외되고 있다.

따라서 감성분석의 정확도를 높이기 위해서는 취합된 문서, 인터넷 댓글, SNS 메시지 텍스트 등으로부터, 비속어, 은어, 약어 등 표준어휘의 범주에 벗어나는 데이터를 포함한, 사람 및 모든 사물에 대한 주된 세부적 이미지와 세부 감성을 자동으로 정량 측정, 평가, 분석할 수 있는 감성분석 방법이 필요하다고 판단된다.

## 2. 감성공학(sensibility engineering)

위에서 언급한 ‘감성분석’과는 별개로, 산업공학 및 인간공학(ergonomics)에서 파생되어 다양한 상품이나 디자인을 대상으로 감성을 직접 정량적으로 측정, 활용하고자하는 ‘감성공학’이라는 분야도 현재 활성화 되어 있다. 감성공학이란 ‘인간-기계 체계 인터페이스 설계에 감성적 차원의 조화성을 도입하는 공학’이라고 정의되며, 일본에서 시작된 감성공학 기법은 인간의 심상(image)을 구체적인 물리적인 설계요소로 번역하여 이를 실현하는 기술을 말한다[13]. 감성공학은 ‘물리적 편리성’을 추구하는 기존 기술 체계와는 달리 ‘정서적 충족’을 추구한다는 점에서 근본적인 차이점을 가지며, 지금까지 자동차의 핸들(steering wheel)과 계기판 등 자동차 인테리어[14], 컬러복사[15], 주택설계[13], 3차원 의자설계[16] 등의 디자인에 이용되었다. 국내에서도 자동차 내장설계 요소에 대한 연구[17], 휴대전화 고급감에 대한 연구[18] 등을 비롯하여 많은 연구가 진행되어 왔으며, 전기자동차 역스테리어[19], 레이저보트 디자인[20], 차량용 HUD 평가 [21] 등에도 활발하게 활용되고 있다.

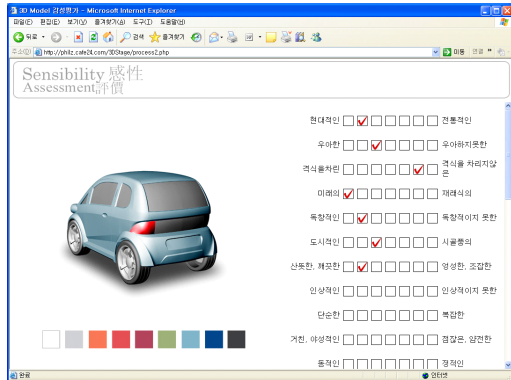


그림 3. 자동차 외장 감성평가(19)  
Fig. 3. Sensibility Evaluation of Car Exterior (19)

하지만 빅데이터 분석에 이러한 감성공학적 기술을 적용하는 데는 다음과 같은 문제점을 가지고 있다.

첫째, 온라인상에서 평가하던, 오프라인에서 평가하던 간에, 감성평가 측정 시마다 평가자에 의한 직접 평가가 이루어져야 하며, 이에 따라 많은 시간과 인적 비용이 발생할 뿐만 아니라 실시간 분석 및 시간추이 분석이 불가능하다.

둘째, 미리 정해진 감성어휘 쌍(SD법)에 의한 평가로써, 정형화된 평가만 가능하며, 문서, 인터넷 댓글, 트위터, 페이스북, 미투데이 등 SNS 메시지 등을 이용한 감성 평가는 불가능하다.

셋째, 감성측정 대상(제품, 브랜드, 사물, 인물) 한 개 당 수십 개의 감성어휘 쌍을 평가해야하는 기법상의 문제로, 평가 가능한 대상의 수와 평가자 수가 한정적이며, 다수 평가자에 의한 다수 대상(제품, 인물 등) 평가가 쉽지 않다.

넷째, 지정된 소수에 의해 비자발적인 평가가 이루어짐으로써, 감성평가 결과가 실제 관심층, 실구매층 집단의 감성을 대변한다고 보기가 어렵다.

즉, 기존 감성공학적 기법은 정형 감성 데이터의 세부 분석 측면에 다양한 장점이 있지만 위에서 언급한 문제점 때문에 빅데이터, 소셜 데이터와 같은 비정형 데이터를 분석하는 데는 한계가 있다. 따라서 비정형 데이터를 분석할 수 있는 감성공학적 기법에 대한 연구와 기존 감성분석 기법과의 융합 연구가 필요하며, 이를 통해 기존의 감성분석의 수준을 한 단계 높일 수 있으리라 생각된다.

### III. 제안 시스템

본 연구의 목표는, 인터넷 댓글, SNS(Social Network

Service) 메시지 텍스트 등으로부터 일상적 언어, 및 은어(隱語), 비속어, 약어, 이모티콘 등을 감성 분석함으로써, 복합적인 감성 중 근간이 되는 주 감성들을 측정하고 평가하는 방법을 개발하는 것이며, 본 연구에서 제안하는 시스템의 구조는 (그림 4)와 같다.

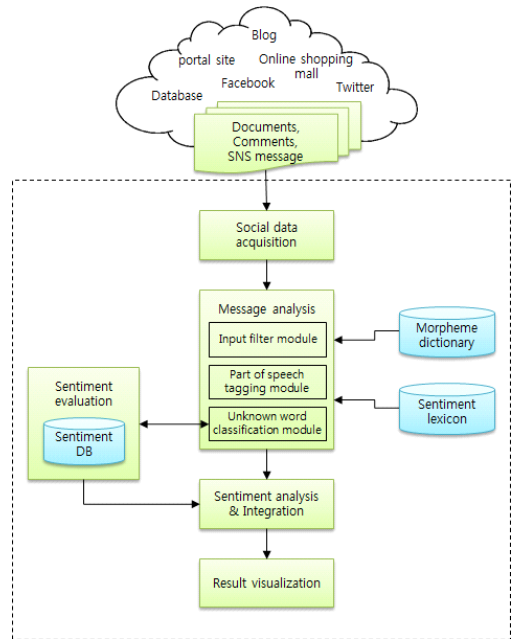


그림 4. 시스템 구조  
Fig. 4. System Structure

#### 1. 데이터 취합(data acquisition)

데이터를 직접 입력받거나 블로그(Blog), 카페(Cafe), 포털(Portal), 쇼핑몰, 트위터 등 인터넷 망을 통해 콘텐츠의 데이터를 URL Request 방식, 검색엔진을 이용하는 방식, 브라우저를 직접 실행하는 방식 등을 이용하여 수집하고, 저장한다.

#### 2. 메시지 분석(message analysis)

수집, 저장된 텍스트와 메시지들을 형태소 단위로 분석하여 품사를 구분하고, 각각의 형태소에 해당되는 품사를 태깅(tagging)하며, 개체명을 인식하고, 패러프레이징(paraphrasing) 및 구문을 분석한다. 또한 의미 없는 특수 기호의 반복 사용이나 띄어쓰기 없는 짧은 패턴의 반복 등을 필터링하여 제외하거나, 축약하여 분석하며, 형태소 DB 및 상기 감성어휘 DB에 저장되어 있지 않아서 품사 태깅이 되지

않은 비속어, 은어, 약어, 이모티콘 어휘는 감성 평가부로 전달한다.

### 3. 감성 평가(sentiment evaluation)

메시지 분석단계의 미등록어 처리 모듈에 의하여 품사태깅이 되지 않은 비속어, 은어, 약어, 이모티콘 등의 미등록 어휘에 대하여, 인간 평가자에 의한 감성평가를 통해 대체할 수 있는 표준어 조합을 선정한다. 감성평가는 SD법(Semantic Differential method)을 사용하여 평가하는데, 예를 들어, '뽀대 있는'이라는 비속어를 표준어로 매핑(또는 번역)하기 위해서 다음 (그림 5)와 같이 감성을 나타내는 형용사들을 피실험자들에게 쌍(서로 반대 의미를 가지는)으로 제시하고(보통 20~40개 쌍) 평가 하도록 한다.

뽀대 있는				
편안해 보이는	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	불편해 보이는
단순한	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	복잡한
안정된	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	불안정한
격식을 차린	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	격식을 차리지 않은
나이 들어 보이는	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	젊은 취향의
부드러운	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	각진, 딱딱한
균형 잡힌	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	불균형의
남성적인	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	여성적인
품위 있는, 우아한	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	품위 없는, 우아하지 못한
.....				

그림 5. 비속어 감성평가  
 Fig. 5. Sentiment evaluation of slang

은어, 약어, 비속어에 대한 감성평가 후, 분석은 주성분 분석(Principal Component Analysis)을 사용하며, 주성분을 2~3 개 추출하고 그 비율과 감성 표준어 조합을 감성DB에 저장한다. 예를 들어 '뽀대 있는'이라는 비속어는 0.4\*(멋진) + 0.3\*(권위 있는) + 0.3\*(폼 나는) 와 같이 표준어를 가중치 조합하여 나타낼 수 있다.

### 4. 감성 분석 및 통합 (sentiment analysis and integration)

감성 측정 대상에 대한 감성 어휘들과 감성 어휘들의 빈도 및 감정적 강도 등을 조합하고 이를 입력자료로 하여 주성분 분석을 실시함으로써, 주된 감성 성분을 추출한다. 그런데, 인터넷이나 SNS에서 수집된 데이터 들은 충분한 감성어휘 쌍에 대한 평가자료를 얻을 수 없으며, 취합, 분류된 데이터는 희소 행렬(sparse matrix) 형태이다. 따라서 규모가 큰

희소행렬에 대한 주성분분석을 위해, IRLBA(Implicitly Restarted Lanczos Bidiagonalization Algorithm)[22]를 활용한다. 주성분 분석에는 일반적으로 SVD(singular value decomposition)를 사용하나, IRLBA는 크고 희소한 행렬의 partial singular value decomposition(SVD)을 빠르고 효율적으로 수행할 수 있다.

행렬  $A \in \mathbf{R}^{\ell \times n}$ ,  $\ell \geq n$  의 SVD가 다음과 같다면,

$$A = \sum_{j=1}^n \sigma_j u_j v_j^T, \quad v_j^T v_k = u_j^T u_k = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{o.w.,} \end{cases}$$

where

$A$  : large sparse matrix

$\mathbf{R}^{\ell \times n}$  : set of  $\ell \times n$  real matrix

$u_j$  : left singular vectors

$v_j$  : right singular vectors

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$  : singular value of  $A$

$A$  의 partial SVD는  $1 \leq k < n$  일 때,

$$A_k := \sum_{j=1}^k \sigma_j u_j v_j^T$$

로 정의된다.

희소행렬 주성분 분석에 이용되는 IRLBA 기법을 단계별로 요약하면 다음과 같다.

#### ① Partial Lanczos Bidiagonalization

벡터  $P_l$ 부터 시작하여 Lanczos의  $m$  단계를 다음과 같이 계산

$$\begin{aligned} AP_m &= Q_m B_m \\ A^T Q_m &= P_m B_m^T + r_m e_m^T, \end{aligned}$$

where

$P_l$  : initial vector

$Q_m$  : output matrix with orthogonal columns

$B_m$  : upper bidiagonal matrix

$r_m$  : residual vector

$$P_m^T P_m = Q_m^T Q_m = I_m,$$

$$r_m \in \mathbf{R}^n, P_m^T r_m = 0,$$

$$P_m = [p_1, p_2, \dots, p_m].$$

where

$I_m$  : identity matrix of order  $m$

② Partial SVD 근사계산

$$\begin{aligned}
 A^T A P_m &= A^T Q_m B_m \\
 &= P_m B_m^T B_m + r_m e_m^T B_m, \\
 A A^T Q_m &= A P_m B_m^T + A r_m e_m^T, \\
 &= Q_m B_m B_m^T + A r_m e_m^T.
 \end{aligned}$$

$B_m$ 의 SVD를 계산 :

$$B_m = \sum_{j=1}^m \sigma_j^B u_j^B (v_j^B)^T.$$

(i.e.,  $B_m v_j^B = \sigma_j^B u_j^B$ , and  $B_m^T u_j^B = \sigma_j^B v_j^B$ .)

정의 :  $\tilde{\sigma}_j := \sigma_j^B$ ,  $\tilde{u}_j := Q_m u_j^B$ ,  $\tilde{v}_j := P_m v_j^B$ .

then :

$$\begin{aligned}
 A \tilde{v}_j &= A P_m v_j^B \\
 &= Q_m B_m v_j^B \\
 &= \sigma_j^B Q_m u_j^B \\
 &= \tilde{\sigma}_j \tilde{u}_j,
 \end{aligned}$$

그리고

$$\begin{aligned}
 A^T \tilde{u}_j &= A^T Q_m u_j^B \\
 &= P_m B_m^T u_j^B + r_m e_m^T u_j^B \\
 &= \sigma_j^B P_m v_j^B + r_m e_m^T u_j^B \\
 &= \tilde{\sigma}_j \tilde{v}_j + r_m e_m^T u_j^B.
 \end{aligned}$$

여기에서  $r_m e_m^T u_j^B$ 은 정확한 SVD 값과 현 단계 계산 값과의 차이(error)를 의미하며, 반복계산을 통해 이 error term의 norm을 최소화한다.

5. 결과물 시각화(result visualization)

감성 분석단계에서 추출되는 수개의 주된 감성 성분으로부터 2차원 내지 3차원의 요인축을 구성하고, 감성측정 대상의 특정 시점에 대한 이미지 공간상의 위치를 실시간으로 표시하거나, 특정 기간 동안의 시간의 추이에 따른 분석결과를 애니메이션 형식으로 제시한다. 시간에 따른 추이는 감성의 시간적 변화를 보여주는데, 예를 들어 자동차 양산, 출시 전 컨셉카(concept car) 사진들에 대한 감성과 출시 후 실제 모델들을 보거나 타보고 난 후의 감성 변화를 보여 줄 수 있다. 정치인이나 연예인들의 경우도 여러 가지 사건이나 루머, 정치적 이슈에 따라 감성평가 결과가 시간에 따라 달라진다.

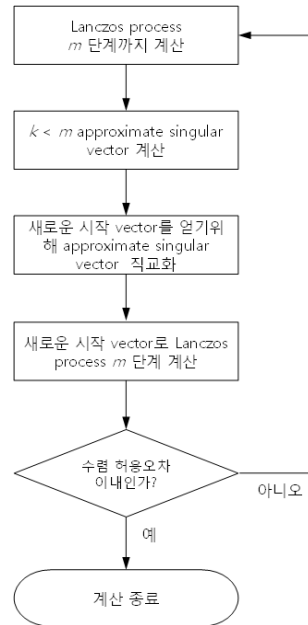


그림 6. IRLBA 프로세스  
Fig. 6. IRLBA process

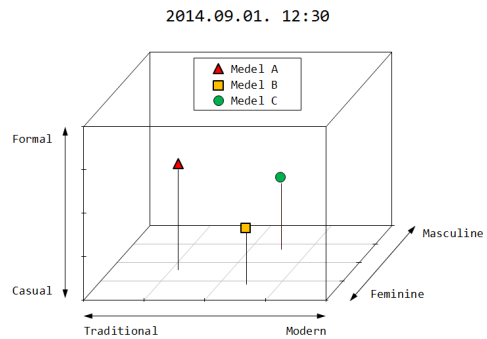


그림 7. 자동차 외장에 대한 주된 감성 시각화  
Fig. 7. Visualization of Principal Sentiment of Car Exteriors

또한 감성측정 대상 여러 개를 주된 감성 축 상에 표현함으로써, 공통의 감정 및 감성 성분들로 구성된 이미지 공간상에 대상들을 동시에 표시 가능하게 된다. 예를 들어, 자동차 3개 모델(A, B, C)을 대상으로 감성분석을 실시하여, 주된 감성성분을 두 개 추출한 경우, 2D 차트에 표시하고, 주된 감성 성분을 세 개 추출한 경우는 (그림 7)과 같이 삼차원 차트에 표현 가능하다.

#### IV. 결 론

본 연구에서는 은어(隱語), 비속어, 약어, 이모티콘 등을 포함하는 소셜데이터를 감성 분석함으로써 복합적인 감성 중간이 되는 주 감성들을 측정하고 평가하는 방법을 개발하고 시스템을 구축, 제안하였다. 제안된 방법론은 긍정/부정 두 개 범주, 또는 긍정/부정/중립의 세 개 범주로만 구분 짓는 기존 감성분석 방법들과는 달리, 근간이 되는, 숨어있는 주된 감성들을 추출할 수 있으며, 표준어를 포함한 은어, 비속어, 약어에 대한 감성분석을 포함함으로써 분석의 정확성을 향상시키고, 분석결과의 효용성을 높일 수 있을 것으로 기대된다.

그리고 기존의 감성공학적 기법의 경우, 평가자들로부터 설문 등을 통해 감성평가를 수행하고 미리 정해진 정형화된 데이터 분석만 가능하나, 제안된 방법론은 자동화된 방식에 의해 소셜데이터로부터 실시간 분석 및 시간차이 분석이 가능하며, 자발적 감성표현을 기초로 분석이 시행되므로 분석의 범위와 신뢰도 측면에서 유리한 것으로 판단된다.

물론 감성분석을 행하고자 하는 대상에 대한 소셜데이터가 적을 경우, 신뢰도와 정확도 측면에서 불리한 점은 있으나, 충분한 데이터가 확보되고, 한국어 구문분석 알고리즘의 정확성이 좀 더 제고된다면 다양한 활용효과를 기대할 수 있을 것이다. 즉, 이러한 감성분석 서비스는 소셜평판 관리, 품평 모니터링 및 입소문 분석, 브랜드 마케팅 등에 활용되어, 소비자 감성적 구조 파악을 통해 기업 평판을 개선하는데 도움이 될 것이며, 잠재 소비자 군 도출하고, 감지하기 어려웠던 소비자 감성을 빠르게 실시간으로 분석하는데 활용 가능할 것으로 기대된다. 추후 연구로는, 본 연구에서 제안한 시스템의 분석결과와 실제 인간 피 실험자가 평가한 감성 분석결과를 비교하여, 어느 정도 인간의 감성평가와 유사한지를 정량적으로 평가하는 단계가 필요할 것이며, 기존 감성분석방법 및 연구 등과의 정확도 비교 평가가 시행될 것이다. 또한 제안된 시간차이 분석결과를 포함한 다양한 분석결과를 직관적, 효과적으로 전달할 수 있는 방안에 대한 연구가 필요할 것으로 판단되며, 여기에는 다양한 시각적 표현이 가능한 인포그래픽스(inforgraphics) 기술의 활용이 바람직할 것으로 사료된다.

#### 참고문헌

[1] McKinsey, "Big Data: The next frontier for innovation, competition, and productivity,"

May. 2011.  
 [2] Gartner, "Hype Cycle for Emerging Technologies," 2011.  
 [3] H. Y. Kim, D. H. Jeon and S. H. Jee, "Bigdata Analysis Project Development Methodology," Journal of The Korea Society of Computer and Information, Vol. 19, No. 3, pp.73-85, Mar. 2014.  
 [4] S. S. Kim, "A study on development method for practical use of Big Data related to recommendation to financial item," Journal of The Korea Society of Computer and Information, Vol. 19, No. 8, pp.73~81, Aug. 2014.  
 [5] etnews, <http://www.etnews.com/201203020076>  
 [6] H. W. Park, "Data as Social Culture, Networked Innovation and Government 3.0," Asian Hub Conference Triple Helix & Network Sciences (DISC) 2014, Dec. 2014.  
 [7] dongA.com, <http://news.donga.com/3/all/20120202/43753364/1>  
 [8] helloworld, <http://helloworld.naver.com/helloworld/1860>  
 [9] Mary M. Flory, "How are You Feeling? Sentiment Analysis Aims to Find Out," AMA Access: Marketing Researchers, Apr. 2011.  
 [10] etnews, <http://www.etnews.com/201301210369>  
 [11] SAS Institute Inc., [http://www.sas.com/en\\_us/software/analytics/sentiment-analysis.html](http://www.sas.com/en_us/software/analytics/sentiment-analysis.html)  
 [12] J. S. Noh, "Find the Meaning in the Sea of Big Data," Microsoft, pp.130-133, Aug. 2012.  
 [13] M. Nagamachi, "An image technology expert system and its application to design consultation," International Journal of Human-Computer Interaction, Vol. 3, No. 3, pp.267-279, Jul. 1991.  
 [14] T. Jindo and K. Hirasago, "Application studies to car interior of Kansei engineering," International Journal of Industrial Ergonomics, Vol. 19, No. 2, pp.105-114, Feb. 1997.  
 [15] K. Fukushima, H. Kawata, Y. Fujiwara and H. Genno, "Human sensory perception oriented

- image processing in color copy system," International Journal of Industrial Ergonomics, Vol. 15, No. 1, pp.63-74, Jan. 1995.
- [16] T. Jindo, K. Hirasago and M. Nagamachi, "Development of a design support system for office chairs using 3-D graphics," International Journal of Industrial Ergonomics, Vol. 15, No. 1, pp.49-62, Jan. 1995.
- [17] Y. H. Lee, M. J. Jeong, M. W. Yoon and S. H. Kang, "Development of Luxuriousness Models for Automobile Crash Pad based on Subjective and Objective Material Characteristics," Journal of Ergonomics Society of Korea, Vol. 25, No. 2, pp.187-195, May. 2006.
- [18] I. K. Kim, C. Lee, and M. W. Yoon, "A Comparison of Modeling Methods for a Luxuriousness Model of Mobile Phones," Journal of Ergonomics Society of Korea, Vol. 25, No. 2, pp.161-171, May. 2006.
- [19] P. S. Jang, C. H. Choi, "Exterior Design of Electric Vehicle Using Web Based Human Sensibility Assessment," Journal of Ergonomics Society of Korea, Vol. 25, No. 4, pp.63-69, Nov. 2006.
- [20] C. H. Choi, P. S. Jang and S. M. Seo, "Digital Design Process of Marine Leisure Boat Using Human Sensibility Evaluation," Journal of Ergonomics Society of Korea, Vol. 29, No. 4, pp.693-699, Aug. 2010.
- [21] W. J. Choi, W. J. Lee, S. H. Lee and Y. K. Park, "Vehicle HUD's cognitive emotional evaluation - Focused on color visibility of driving information", Korean Journal of The Science of Emotion & Sensibility, Vol. 16, No. 2, pp.195-206, Jun. 2013.
- [22] J. Baglama, L. Reichel, "Restarted Block Lanczos Bidiagonalization Methods," Numerical Algorithms, Vol.43, pp.251-272, Nov. 2006.

## 저 자 소 개



### 장 필 식

1990: 서울대학교  
조선공학과 공학사.  
1992: KAIST  
산업공학과 공학석사.  
1998: KAIST  
산업공학과 공학박사  
현 재: 세한대학교  
정보물류학과 교수  
관심분야: HCI, 감성공학  
Email : phil@sehan.ac.kr