

## Levenshtein 거리를 이용한 영화평 감성 분류

안광모\*, 김윤석\*\*, 김영훈\*\*\*, 서영훈\*\*\*\*

### 요약

본 논문에서는 레빈슈타인 거리(Levenshtein distance)를 이용한 감성 분류 방법을 제안한다. 감성 자질에 레빈슈타인 거리를 적용하여 BOW(Back-Of-Word)를 생성하고 이를 학습 자질로 사용한다. 학습 모델은 지지벡터기계(support vector machines, SVMs)와 나이브 베이즈(Naive Bayes)를 이용하였다. 실험 데이터로는 다음 영화 사이트로부터 영화평을 수집하였으며, 수집한 영화평은 총 2,385건이다. 수집된 영화평으로부터 감성 어휘를 수작업을 통해 수집하였으며 총 778개 어휘가 선별되었다. 실험에서는 감성 어휘에 레빈슈타인 거리를 적용한 BOW를 이용하여 기계학습을 수행하였으며, 10-fold-cross validation 방식으로 분류기의 성능을 평가하였다. 평가 결과는 레빈슈타인 거리가 3일 때 다항 나이브 베이즈(Multinomial Naive Bayes) 분류기에서 85.46%의 가장 높은 정확도를 보였다. 실험을 통하여 본 논문에서 제안하는 방법이 문서 내의 철자 오류에 대해서도 분류 성능에 영향을 적게 받음을 알 수 있었다.

키워드 : 감성 분류, 오피니언 마이닝, 레빈슈타인 거리, 지지벡터기계, 나이브 베이즈

## Sentiment Classification of Movie Reviews using Levenshtein Distance

Kwang-Mo Ahn\*, Yun-Suk Kim\*\*, Young-Hoon Kim\*\*\*, Young-Hoon Seo\*\*\*\*

### Abstract

In this paper, we propose a method of sentiment classification which uses Levenshtein distance. We generate BOW(Bag-Of-Word) applying Levenshtein distance in sentiment features and used it as the training set. Then the machine learning algorithms we used were SVMs(Support Vector Machines) and NB(Naive Bayes). As the data set, we gather 2,385 reviews of movies from an online movie community (Daum movie service). From the collected reviews, we pick sentiment words up manually and sorted 778 words. In the experiment, we perform the machine learning using previously generated BOW which was applied Levenshtein distance in sentiment words and then we evaluate the performance of classifier by a method, 10-fold-cross validation. As the result of evaluation, we got 85.46% using Multinomial Naive Bayes as the accuracy when the Levenshtein distance was 3. According to the result of the experiment, we proved that it is less affected to performance of the classification in spelling errors in documents.

Keywords : Sentiment Classification, Opinion Mining, Levenshtein Distance, SVMs, Naive Bayes

### 1. 서론

※ 교신저자(Corresponding Author): Young-Hoon Seo

접수일:2013년 12월 04일, 수정일:2013년 12월 25일

완료일:2013년 12월 27일

\* 충북대학교 컴퓨터공학과

Tel: +82-43-261-2319

email: [ahnmo777@gmail.com](mailto:ahnmo777@gmail.com)

\*\* 충북대학교 컴퓨터공학과

\*\*\* 청강문화산업대학교 모바일스쿨

\*\*\*\* 충북대학교 컴퓨터공학과

스마트 기기의 사용이 보편화 되면서 자신의 의견을 표현하는 매우 많은 글들이 시간적, 공간적 제약을 벗어나 웹상에서 손쉽게 작성되고 있다. 특히 댓글이나, SNS 등을 통해 자유롭게

■ 이 논문은 2011년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었음

자신의 의견과 견해를 적극적으로 작성하는 사용자들이 많으며, 이런 환경을 바탕으로 대중의 의견이나 감정 등을 파악하는 오피니언 마이닝(opinion mining)에 대한 연구들이 최근 들어 활발히 이루어지고 있다. 오피니언 마이닝은 기업이나 정치 등의 분야에서 대중의 의견을 파악하여 분석하는 여러 응용 프로그램에서 사용할 수 있는 기술이기 때문에 그 활용에 있어 많은 가능성을 가지고 있다.

오피니언 마이닝은 텍스트 마이닝(text mining)의 한 분야로서 의견 추출(opinion extracting), 감정 분석(sentiment analysis), 감정 분류(sentiment classification)로도 불리며, 문서가 가진 감성을 분류한다. 감정 분류에 있어서 기쁨, 분노, 슬픔 등과 같이 여러 감정 도메인으로 분류하는 연구[2]도 있으나 대부분의 연구에서는 감성을 긍정 또는 부정의 극성(polarity)으로 분류하고 있다.

감성을 분류하는 방법에 있어서는 감정 사전을 이용한 의미 분석 기반의 방법과 기계학습 기법을 이용하는 방법으로 크게 나눌 수 있다. 감정 사전을 이용할 경우 보통 형태소 분석을 이용하여 형태소를 분리한 후 감정 사전을 탐색하여 감정 형태소의 빈도수를 구하고 PMI(Pointwise Mutual Information) 방법 등을 이용하여 극성을 판단한다. 하지만 웹상에는 비문법적인 문서가 많아 형태소 분석 오류가 높고 사전에 등재되지 않은 단어들이 문서에 출현할 경우 극성을 분석하기 힘든 단점을 가지고 있다. 따라서 최근에는 기계학습 기법을 이용한 방법들이 많이 연구되고 있다.

본 논문에서는 레빈슈타인 거리(Levenshtein distance, 이하 LD)를 적용하여 비문법적인 문서를 처리하는데 있어 성능을 향상시키는 방법을 제안한다. 감정 사전에 등재된 감정 어휘의 음소를 분리한 후 LD를 적용하여 BOW(Back-Of-Word)를 생성한다. 이렇게 생성된 BOW는 철자 오류 및 띄어쓰기 오류에 대한 영향을 감소시킨다. 그리고 LD를 적용하여 생성된 BOW를 지지벡터기계(SVMs, Support Vector Machines)와 나이브 베이즈(Naive Bayes)에 적용하여 학습을 하였다. 실험 결과 다항 나이브 베이즈(Multi-nominal Naive Bayes)가 가장 좋은 성능을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 오피니언 마이닝에 대한 국내외 연구들에 대해서 살펴본다. 3장에서는 LD를 적용하여 BOW를 생성하고 기계학습에 적용하는 방법에 대하여 기술한다. 4장에서는 3장에서 제안한 방법을 이용하여 실험한 내용을 기술한다. 그리고 5장에서 결론 및 향후 연구에 대하여 논의하고 논문을 마친다.

## 2. 관련 연구

오피니언 마이닝은 문서를 작성한 사용자의 의견이나 감정을 분석하는 기술이다. 그리고 일반적으로 대부분의 연구들에서는 문서의 의견이 부정적인지 긍정적인지를 분류하고 있다.

의견 분류를 수행하는 많은 방법 중 [3]의 연구에서는 PMI 값을 이용하여 의견을 분류한다. PIM란 두 단어 사이의 관계를 수치로 계산하는 방법이며 PIM의 값이 0이면 관계가 없는 단어이며 양수이면 같은 성향의 단어, 음수이면 반대 성향의 단어로 판단한다.

의견을 분류하기 위해 자연언어처리 기술을 활용하거나 통계적인 기법을 활용하기도 한다 [4-7]. 자연언어처리기술 기반의 방법은 철자나 띄어쓰기 오류 등이 많은 웹 문서에 적용에 많은 어려움이 있어 통계적인 방법들을 사용하게 되었다. 그리고 자연언어처리기술과 통계적 방법의 장점을 서로 접목시킨 연구들도 있다[8, 9].

최근의 국내의 연구들을 간단히 살펴보면 [2]에서는 한글 문서를 기반으로 기계학습모델을 적용하여 7개의 감정으로 분류하는 연구를 진행하였다. 여러 학습 모델을 사용하여 다항 나이브 베이즈에서 가장 높은 성능을 보였고, 영화 감정 비율 및 영화 장르별 감정 특성을 분석하는데 응용하였다. [10]의 String kernel을 한국어에 맞게 변형하여 감정 분석을 수행한 연구들도 있는데, [11]의 연구에서는 음절 커널을 사용하였고, [12]에서는 음절 커널을 음운 커널로 변형하여 감정 분류에 사용하였다. [13]의 연구에서는 어절을 이용한 빈도수 기반의 점수가 부여된 극성 패턴을 추출하고, 극성 패턴과 그것의 점수를 규칙에 의해 조정하는 방법으로 긍/부정 극성을 분류하였다.

### 3. LD를 이용한 자질 추출과 기계학습

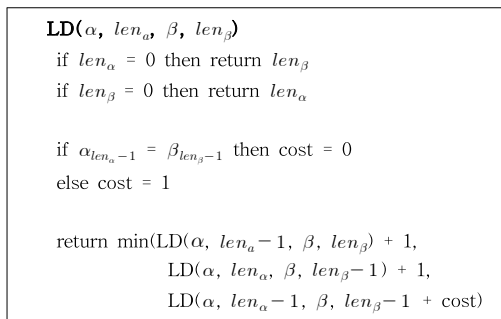
이 장에서는 LD에 대하여 간단히 설명하고, 본 논문에서 제안하는 BOW에 LD를 적용하여 생성하는 방법에 대하여 서술한다. 제안하는 방법에 의해 생성된 BOW는 기계학습의 자질로서 사용된다.

#### 3.1 LD(Levenshtein Distance)

LD는 두 문자열 사이의 차이를 계산하기 위한 알고리즘이다. 하나의 문자열에 수정(edit) 연산을 하여 다른 문자열과 같게 할 때 가장 적은 수정 연산 횟수를 구하는데, 이 횟수가 LD이다. 수정 연산은 삽입(insertion), 삭제(deletion), 치환(substitution)이 있으며, 문자의 수정 횟수를 이용하여 두 문자열 사이의 거리를 구하기 때문에 edit distance라는 용어를 사용하기도 한다.

두 문자열  $\alpha$ 와  $\beta$ 가 주어졌을 때 두 문자열의 LD를 구하는 알고리즘은 (그림 1)과 같다.

(그림 1) LD 알고리즘



(Figure 1) Levenshtein distance algorithm

위 알고리즘에서  $len_w$ 는 문자열  $w$ 의 길이를 의미하며  $w_i$ 의 형태는 문자열  $w$ 의  $i$ 번째 문자를 의미한다. 위의 알고리즘은 문자열  $\alpha$ 를 문자열  $\beta$ 와 같도록 하는 거리를 계산하는 알고리즘을 의미한다.

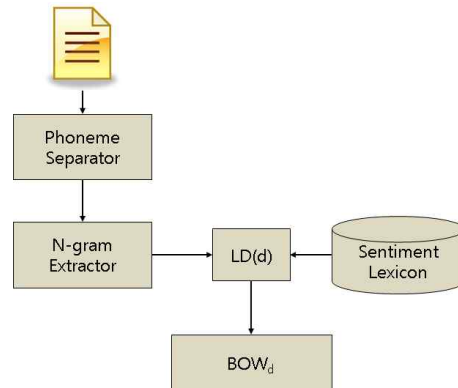
본 논문에서는 두 문자열 사이의 LD가 작을수록 유사한 문자열일 가능성이 많다고 가정하고 이를 이용하여 BOW를 생성하고 이를 기계학습에 적용한다.

#### 3.2 LD를 이용한 자질 추출 및 기계학습

본 논문에서 사용하는 자질은 포털 사이트 다음의 영화 서비스에서 수집된 영화평을 바탕으로 추출한 감성 어휘들이다. 감성 어휘는 명확하게 긍정과 부정을 표현하는 단어들로 선별하였으며 각 감성 어휘는 어근만을 자질로 선별하였다. 예를 들어 ‘좋아하다’와 같은 감성 어휘는 ‘좋아’만을 자질로 사용하였다. 선별된 감성 어휘는 음소가 분리되어 감성 사전에 저장된다.

각 감성 어휘들은 BOW의 자질로 사용이 되는데 변환되는 과정에서 LD가 적용된다. LD의 값을  $d$ 로 적용한 BOW를  $BOW_d$ 로 본 논문에서는 표현하는데 여기서  $d$ 는 LD 값을 의미한다. 특정 문서가 들어올 경우 문서를  $BOW_d$ 를 추출하는 과정은 (그림 2)와 같다.

(그림 2) BOW<sub>d</sub>의 생성 과정



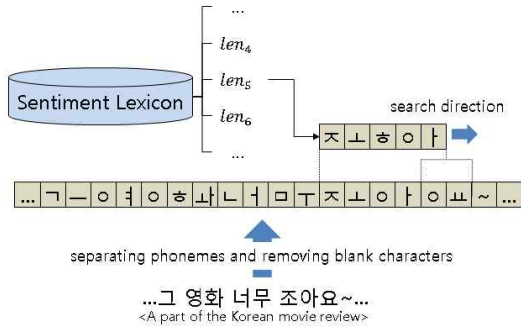
(Figure 2) The process generating BOW<sub>d</sub>

입력 문서는 우선 음소분리기를 거쳐 음소가 분리가 되며 이 과정에서 공백 문자는 제거가 된다. 공백 문자를 제거하는 이유는 웹상의 문서에는 띄어쓰기 오류가 많이 존재하기 때문에 이로 인한 오류를 완화하기 위함이다.

그 다음 단계로 n-gram 추출기를 이용하여 n-gram의 문자열을 추출하는데 이것은 감성 사전의 감성 어휘에 존재하는 길이에 따른 n-gram을 추출한다. 예를 들어 감성 어휘의 길이가 3~13이라면 n-gram 추출기는  $n = 3 \sim 13 \pm d$ 의 n-gram의 문자열을 추출한다. 여기서  $d$ 의 범위를 주는 이유는 LD의 값에 따라 n-gram 문자열이 해당 감성 어휘를 판단하는 유사성 정도가 되는데  $d$ 의 범위만큼 철자오류가 발생할 것

이라고 보기 때문이다. 예를 들어 4-gram 문자열 ‘조아(조아)’와 감성 어휘 ‘좋아(조아호아)’의 LD는 1이 되는데, d를 1로 설정하게 되면 길이가 4인 ‘조아’라는 어휘도 길이가 5인 감성 어휘 ‘좋아’에 해당하는 자질 값이 될 수 있게 된다. 다음 (그림 3)은 d의 값이 1일 때, 감성 어휘 ‘좋아’의 n-gram 탐색 범위이다.

(그림 3) 감성 어휘 ‘좋아’의 탐색 과정 (LD가 1인 경우)



(Figure 3) The search process of a Korean sentiment lexicon when LD is 1

d 값을 설정하는데 있어 한 가지 제약이 있는데 문자열의 길이가 짧으면 철자오류가 발생의 정도가 심하지 않을 것이라 가정하여 길이가 L인 감성 어휘에 적용되는 d 값은 최대  $\lfloor L/3 \rfloor$  까지만 적용이 된다. 즉, 길이가 5인 ‘좋아’에 해당하는 n-gram 문자열을 문서에서 탐색할 경우 적용되는 d는 최대 1까지만 적용이 된다.

이렇게 생성된  $BOW_d$ 는 이용하여 학습 데이터 및 테스트 데이터로 사용이 되는데 본 논문에서 사용한 기계학습 모델은 SMV, 다항 나이브 베이즈(Multi-nomial Naive Bayes), 베르누이 나이브 베이즈(Bernoulli Naive Bayes)이며 SVM의 커널로는 선형 커널(linear kernel), 다항 커널(polynomial kernel), RBF 커널(Radial Basis Function kernel)을 사용하였다.

#### 4. 실험 및 결과 분석

본 장에서는 논문에서 제안한 LD를 적용한 BOW를 이용하여 분류기를 학습하고, 학습된 분류기의 성능을 평가한다. 그리고 실험 결과를 통

해 제안한 방법이 비문법적인 문서에 대하여 감성 분류 성능을 높임을 보인다.

#### 4.1 실험 데이터

실험 데이터는 다음 영화 서비스에서 관객수가 천 만 명 이상인 영화로부터 영화평을 수집하였다. 수집된 8,182개의 영화평 중 긍정과 부정 의견이 확실한 데이터 2,385개를 선별하여 긍정/부정 태깅 작업을 수행하였다. 긍정/부정 태깅을 수행하는 과정에서 긍정과 부정을 표현하는 감성 어휘를 수집하여 감성 사전을 구축하였으며 수집된 감성 어휘는 총 778개이다. 감성 어휘는 어근을 추출하고 음소를 분리하여 사전에 저장하고 ‘X하다’, ‘X되다’, ‘X시키다’와 같은 형태는 X만을 감성 어휘로 추가하였다. 다음 <표 1>은 분리된 음소 각각을 길이 1로 보았을 때 길이별 빈도수이다.

<표 1> 감성 어휘의 길이별 빈도수

Length of Sentiment Lexicon	Frequency
3	13
4	61
5	206
6	214
7	64
8	78
9	55
10	40
11	33
12	13
13	1
<b>Total</b>	<b>778</b>

<Table 1> The frequency by length of sentiment lexicon

길이가 5~6인 감성 어휘가 가장 높은 빈도를 차지하였으며, 이것은 감성 어휘의 음절수가 약 2~3인 경우가 가장 많은 것과 같다. 수집된 감성 어휘를 이용하여 영화평 데이터를  $BOW_d(0 \leq d \leq 4)$ 의 형태로 변환하였다. d를 4까지 한 이유는 수집된 감성 어휘의 최대 길이가 13이기 때문에 제안된 방법에 의해서 적용되는 최대 LD가 4이기 때문이다.

#### 4.2 실험 결과

실험의 평가는 정확도(accuracy)로 측정하였

으며 평가식은 다음과 같다.

$$Accuracy = \frac{\sum_{i,j \in \{n,p\}} N_{i=j}}{N}$$

$N$ 은 테스트 문서의 수이고  $i$ 와  $j$ 는 정답이  $i$ 인 문서를 분류기가  $j$ 로 예측한 결과이다. 위 식을 이용하여 총 2,385개의 영화평 데이터를 10-folds-cross validation을 수행하여 평가하였으며 평가 결과는 다음 <표 2>와 같다.

<표 2> 분류기별 LD 값에 따른 정확도

Classifier	LD	Accuracy(%)
SVM(linear)	0	82.353
	1	83.109
	<b>2</b>	<b>83.571</b>
	3	83.403
	4	83.067
SVM(poly)	0	77.521
	1	77.521
	2	77.605
	3	77.521
	4	77.479
SVM(rbf)	0	77.521
	1	77.563
	2	77.647
	3	77.563
	4	77.479
Multinomial NB	0	83.487
	1	85.252
	2	85.210
	<b>3</b>	<b>85.462</b>
	4	84.790
Bernoulli NB	0	82.101
	1	79.622
	2	79.706
	3	79.118
	4	79.622

<Table 2> The accuracy referring to the value of LD dependent on the classifiers

실험 결과를 보면 다항 나이브 베이즈가 가장 높은 성능을 보이고 있으며 다음으로 선형 커널을 이용한 SVM이 높은 성능을 보이고 있다. 다

항 나이브 베이즈 분류기의 경우 LD가 3일 때, 85.462%로 가장 높은 성능을 보이고 있으며, 선형 커널 SVM이 LD가 2일 때 83.571% 가장 높은 성능을 보이고 있다. 다른 분류기의 경우는 LD의 값에 따라 큰 분류 성능 차이를 보이지 않았으며 베르누이 나이브 베이즈의 경우는 오히려 LD가 0일 때가 가장 높은 성능을 보였다. 따라서 본 논문의 방법을 적용하기 적합한 분류기는 다항 나이브 베이즈라 볼 수 있다.

### 4.3 결과 분석

실험 결과를 분석한 결과 본 논문에서 제안하는 방법을 적용할 경우 철자 오류 및 띄어쓰기 오류에서 가장 높은 성능 향상이 있었다. 최근 들어 스마트 폰을 통한 입력이 많아 이러한 오류는 그 빈도가 더 많아졌다. 크게 분류를 하면 다음과 같은 경우가 있다.

- 단어를 변형하여 쓰는 경우: 착오(최고), 조아(좋아), 짱(짜증) 등
- 입력 오류: ㅈ | 루, 장안(장난) 등
- 맞춤법을 제대로 알지 못하고 쓰는 경우: 별루(별로), 어의없(어이없) 등

위의 경우는 본 논문에서 제안하는 방법으로 쉽게 해결이 되는 문제들이다. 예를 들어 ‘착오’의 경우는 ‘최고’라는 단어를 변형하여 사용한 경우로 LD의 값은 1이 된다. BOW에서 ‘착오’는 ‘최고’의 자질 값으로 사용될 수 있게 된다. ‘ㅈ | 루’ 경우는 단어를 입력하는 과정에서 일어난 오타로 모바일 환경의 키패드로 입력하는 경우 자주 발생하는 형태이다. ‘ㅈ | 루’의 경우는 제안하는 분류기에서 음소를 분리하고 공백 문자를 제거하여 처리하기 때문에 크게 문제가 될 것이 없는 형태이다. 나머지의 경우는 앞서 예들 든 ‘착오’의 경우와 같이 LD를 이용하여 쉽게 해결되는 형태이다.

다음으로 본 논문의 방법으로 발생하는 오류에 대하여 살펴보면 우선 비슷한 형태의 감성 어휘에 의한 오류이다. 예를 들어 감성 어휘 ‘상쾌’, ‘불쾌’는 두 단어 사이의 LD 값이 1이다. 만약, ‘상쾌’와 같은 어휘가 입력되더라도 LD를 1로 설정할 경우 두 감성 어휘와 LD 값이 같기 때문에 같은 감성 어휘로 판별이 되게 된다. 그

리고 두 단어는 서로 반대의 극성을 가진 어휘이기 때문에 분류의 성능을 떨어뜨리는 원인이 된다. 이러한 문제를 해결하기 위해 입력 단어를 LD 값이 가장 작은 감성 어휘로 결정할 수 있도록 하는 방법으로 해결이 가능하다. 다른 예로 ‘예쁘’와 ‘기쁘’의 경우 ‘이쁘’와 같은 단어가 입력되었을 때, 같은 감성 어휘로 판별하게 된다. 이러한 경우는 초성에 가중치를 주는 방법으로 해결이 가능하다. 이유는 단어를 변형하거나 맞춤법이 틀리게 쓰더라도 초성이 달라지거나 틀리는 경우는 적기 때문이다.

마지막으로 본 논문의 방법으로 해결이 힘든 문제는 구문적인 구조에 의한 문제이다. 예를 들어 ‘그 영화는 하나도 재미있지 않다.’와 같은 문장에서 단순히 감성 어휘를 자질로 사용할 경우 ‘재미’라는 감성 어휘 때문에 긍정으로 판단하게 된다. 이러한 예를 정확히 분석하기 위해서는 ‘하나도’와 ‘~지 않다’와 같은 구문적인 요소를 자질로 사용하여야 한다. 이러한 문제를 해결하기 위해 구문적인 자질을 적용하는 것은 향후 연구로 남겨둔다.

## 5. 결론 및 향후 연구

본 논문에서는 LD를 이용하여 감성 분류 방법을 제안하였다. LD를 이용하여 BOW를 생성하고 이를 여러 기계학습 방법에 적용한 결과에서는 LD가 3일 때 다항 나이브 베이즈 분류기에서 가장 높은 정확도를 보일 수 있었다. 그리고 실험을 통해 본 논문에서 제안하는 방법은 철자 오류 및 띄어쓰기 오류의 영향을 감소시키고 감성 분류의 성능을 높일 수 있음을 보였다. 따라서 본 논문에서 제안한 방법을 응용하면 비교적 철자 오류나 띄어쓰기 오류가 많은 웹문서의 분석에도 활용할 수 있을 것이다.

향후 연구로는 감성 분류에 구문적인 구조를 이용하는 것이다. 본 논문에서는 구문적인 구조에 의한 문제는 해결할 수 없었으며, 이를 위해 구문적인 자질을 적용할 수 있는 연구가 더 필요하다.

## References

- [1] V. I. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals," *Soviet Physics Doklady*, Vol.10, pp.707-710, 1965
- [2] C. Lee, D. Choi, S. Kim, J. Kang, "Classification and Analysis of Emotion in Korean Microblog Texts," *KIISE : Databases*, Vol.40, No.3, 2013 (in Korean)
- [3] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," In *Proceeding of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, pp.417-424, 2002
- [4] T. Nasukawa, J. Yi, "Sentiment Analysis: Capturing Favorability using Natural Language Processing," In *Proceedings of the K-CAP-03, 2nd International Conference on Knowledge Capture*, pp.70-77, 2003
- [5] J. Yi, W. Niblack, "Sentiment Mining in Web-Fountain," *International Conference on Data Engineering (ICDE'05)*, pp.1073-1083, 2005
- [6] N. Godbole, M. Srinivasaiah, S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs," *Intel AA AI Conference on Weblogs and Social Media (ICWSM 2007)*, 2007
- [7] E. Boiy, P. Hens, K. Deschacht, M. Moens, "Automatic Sentiment Analysis in On-line Text," *ELPUB2007 Conference on Electronic Publishing*, 2007
- [8] M. Gamon, A. Aue, S. Corston-Oliver, E. Ringger, "Pulse: Mining Customer Opinions from Free Text," In *Lecture Notes in Computer Science*, Vol.3646, Springer Verlag (IDA 2005), 2005
- [9] X. Ding, B. Liu, "The Utility of Linguistic Rules in Opinion Mining," pp.811-812, *SIGIR2007*, 2007
- [10] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, C. Watkins, "Text Classification using String Kernels," *Journal of Machine Learning Research*, Vol.2, pp.419-444, 2002

[11] S. Kim, S. Park, S. Park, S. Lee, K. Kim, "A Syllable Kernel based Sentiment Classification for Movie Reviews," Journal of Korean Institute of Intelligent Systems, Vol.20, No.2, pp.202-207, 2010 (in Korean)

[12] H. Kim, S. Lee, "The Phoneme Kernel Technique based on Support Vector Machine for Emotion Classification of Mobile Texts," Journal of KIISE : Software and Application, Vol.40, No.6, pp.350-355, 2013 (in Korean)

[13] J. Kim, S. Lee, H. Yong, "Automatic Classification Scheme of Opinions Written in Korean," Journal of KIISE : Databases, Vol.38, No.6, pp.423-428, 2011



**김 영 훈**

1994년 : 충북대학교 (학사-컴퓨터공학)

1997년 : 충북대학교 대학원 (공학석사-컴퓨터공학)

2012년 : 충북대학교 대학원 (공학박사-컴퓨터공학)

1997년~2005년: LG전자 선임연구원

2007년~2011년: LG U+ 부장

2005년~2013년: 동양미래대학교 겸임교수

2013년~현 재: 청강문화산업대학교 초빙교수

관심분야 : 자연언어처리, 임베디드, 모바일 서비스 등



**안 광 모**

2007년 : 충북대학교 (학사-컴퓨터공학)

2009년 : 충북대학교 대학원 (공학석사-컴퓨터공학)

2009년~현 재: 충북대학교 컴퓨터공학과 박사과정  
관심분야 : 자연언어처리, 정보검색, 기계학습 등



**서 영 훈**

1983년 : 서울대학교 (학사-컴퓨터공학)

1985년 : 서울대학교 대학원 (공학석사-컴퓨터공학)

1991년 : 서울대학교 대학원 (공학박사-컴퓨터공학)

1994년~1995년: 미국 Camegie Mellon 대학 기계번역센터 객원교수

1988년~현 재: 충북대학교 컴퓨터공학과(교수)

관심분야 : 자연언어처리, 한영기계번역, 정보검색, 질의응답시스템 등



**김 윤 석**

2012년 : 충북대학교 (학사-컴퓨터공학)

2012년~현 재: 충북대학교 컴퓨터공학과 석사과정  
관심분야 : 자연언어처리, 기계학습 등