

Word2vec와 Label Propagation을 이용한 감성사전 구축 방법

A Novel Method for Constructing Sentiment Dictionaries using Word2vec
and Label Propagation

허 찬, 온승엽¹⁾

Chan Heo, Syng-Yup Ohn

(10540) 경기도 고양시 덕양구 항공대학교로 76 한국 항공대학교 소프트웨어학과
tapril88@kau.ac.kr, syohn@kau.ac.kr

요 약

본 논문에서는 Word2vec을 이용한 새로운 그래프 기반 감성사전 구축 방법을 제안한다. 기존의 연구에서 단어 그래프에서의 변(edge)의 가중치를 계산하기 위하여 널리 사용해 오던 PMI(Point-wise Mutual Information)는 문장의 맥락이나 문맥 상 의미를 통한 단어 간의 관계를 전혀 고려하지 않았다. 본 논문에서는 단어의 의미와 문장에서의 맥락을 고려하여 단어를 벡터로 표현하는 Word2vec 기법에 기반한 변의 가중치 계산 방법을 제안한다. 또한, 새로운 가중치 계산 방법이 감성 사전의 정확도를 개선하는 것을 실증하기 위하여 10,000건의 실제 영화평을 포함하는 데이터를 기반으로 두 가지 LP(Label Propagation) 알고리즘에 새로운 Word2vec 가중치를 적용하여 감성사전을 구축하는 실험을 실행하였다. 이 실험 결과 Word2vec 가중치 계산식을 적용하여 구축한 감성사전을 이용하여 새로운 문장에 대한 감성분석을 하였을 경우 기존의 방법보다 평균적으로 1.5%의 정확도가 향상됨을 알 수 있었다.

Abstract

In this paper we propose a new method for constructing a graph based sentiment dictionary using Word2vec and label propagation. The similarity calculation method used in the previous research to calculate the weight of the edges in a word graph, PMI(Point-wise Mutual Information), does not consider the relationships among the words by the context and the contextual meaning of a sentence for calculating the weights of the edges in a word graph. To improve the accuracy of the weights, we propose a new method using Word2vec, that vectorizes words by considering the meaning of the words and the context of the sentence containing the words. We confirmed the improvement of the performance of our proposed method through the experiments with actual clinical data. In the experiment of constructing sentiment dictionaries based on 10,000 movie

1) 교신저자

reviews, our methods combining Word2vec weight calculation method and label propagation algorithms produced the sentiment dictionaries, which predict new text data with 1.5% more accuracy than the conventional method using PMI.

키워드: 감성분석, 감성사전, 최적화, Word2vec, Label Propagation

Keyword: Sentiment Analysis, Sentiment Dictionary, Optimization, Word2vec, Label Propagation

1. 서론

감성분석은 텍스트에 표현되는 사람들의 감정, 태도, 성향 등과 같은 주관적 데이터를 분석하는 방법으로, 소비자들이 느끼는 것들을 직접 파악 가능하다는 점에서 기업의 마케팅 지표로도 활용되고 있다[1]. 최근의 많은 연구들은 감성분석을 활용할 때, 어휘별 극성을 미리 파악하여 데이터베이스화한 감성사전을 이용하며, 특정 문서에 출현하는 어휘들의 감성 빈도수 등의 기준에 따라 특정 문서의 감성을 양성과 음성의 두 종류로 예측하는 방법을 사용한다[2, 3, 4]. 감성분석과 같은 대부분의 오픈 니언 마이닝분석은 주로 기계학습에 의존하므로 학습을 위한 데이터가 필요하다[5]. 그러나 감성사전이 미리 준비되어 있는 경우, 학습 데이터가 충분하지 않은 상황에서도 충분히 텍스트의 감성을 분석할 수 있다. 감성사전은 일반적인 사용자가 감성을 표현하기 위해 사용하는 객관적인 어휘들을 제공하여 일관된 분석 기준을 규정해 준다[6].

본 논문에서는 새로운 그래프 기반 감성사전 구축 방법을 제안한다. 기존의 그래프 기반 감성사전 구축 방법에서 그래프 변(edge)의 가중치 계산에 사용되어 오던 유사도 계산 방법인 PMI(Point-wise Mutual Information)는 문장에서의 맥락이나 문맥상의 의미를 통한 단어들 간의 관계를 전혀 고려하지 않는다. 본 논문에서는 이러한 한계를 가진 가중치 계산방법을 개선하기 위해 단어의 의미와 문장에서의 맥락을 고려하여 단어를 벡터로 표현하는 Word2vec[7,8]을 이용한 변의 가중치 계산식을 제안한다. 또한, 새로운 가중치 계산 방법이 감성사전의 정확도를 개선하는 것을 실증하기 위

하여 10,000 건의 실제 영화평 데이터를 기반으로 두 가지 LP(Label Propagation) 알고리즘에 새로 제안하는 Word2vec 가중치 계산식을 적용하여 감성사전을 생성하는 실험을 실행하였고 이 감성사전들을 기반으로 한 새로운 텍스트에 대한 감성분석 결과에서 감성 예측의 정확도가 기존의 방법보다 1.5% 향상됨을 확인하였다.

본 논문은 다음과 같이 구성된다. 제 2장에서는 감성사전 구축과 Word2vec과 관련된 기존 연구들을 소개한다. 제 3장에서는 개선된 그래프의 가중치 계산식을 제안하고, LP 알고리즘을 이용한 감성사전 구축 방법에 대해 소개한다. 제 4장에서는 제안하는 방법의 성능 향상을 검증하기 위해 기존의 PMI 가중치 계산식[6]과 새로 제안하는 Word2vec 기반의 가중치 계산식을 각각 두 가지 LP 알고리즘에 적용하여 감성사전을 구축하여 감성사전의 정확도를 계산하고, 구축된 감성사전을 기반으로 새로운 문장의 감성분류 실험을 실행하며 실험 결과를 비교 해석한다. 끝으로, 제 5장에서는 결론으로 마무리한다.

2. 관련 연구

텍스트는 자연어 처리의 대상으로서 분석의 목적에 따라 계층, 벡터, 매트릭스 등의 여러 가지 형태로 표현되며[9], 텍스트 기반 감성 분석인 오픈 니언 마이닝은 문장 또는 문서 단위로 이루어진다.

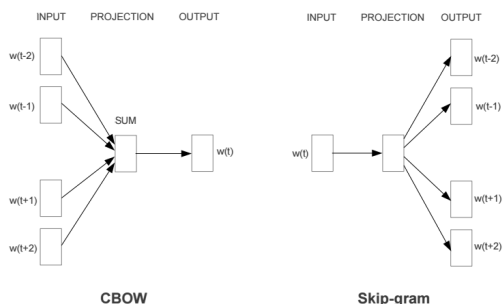
감성사전 구축 방법들은 크게 사전기반 방법과 말뭉치기반 방법으로 분류된다. 사전기반 방법은 WordNet 등의 일반적인 사전으로부터 생성한 시드 어휘와 다른 어휘들 간의 유사도 및 유클리디언

거리(Euclidian Distance) 등을 이용해 감성사전을 구축하는 방식이다. 말뭉치기반 방법은 목적하는 사전의 해당 영역으로부터 수집한 문장들로부터 추출된 말뭉치들을 구문 분석하여 감성사전을 구축하는 방식이다[6].

Word2vec은 말뭉치를 입력으로 받아서 말뭉치의 단어를 벡터화(Vectorization)하는 방법으로서 Word2vec의 학습에 사용되는 신경망 알고리즘은 Harris의 분포가설[10]에서 시작된 자연어 처리 모델이다. Word2vec는 벡터가 단순한 수치적 개념이 아닌, 의미적 개념의 집합으로 표현되기 때문에 하나의 의미를 내포하는 기능을 가진다[11].

Word2vec의 학습 방법에는 CBOW(Continuous bag-of-words)와 Skip-gram 모델이 있다. CBOW 방식은 특정 단어의 주변에 위치한 단어들이 구성하는 맥락을 이용하여 특정 단어를 예측하는 것이고 Skip-gram은 특정 단어를 기준으로 주변에 위치할 수 있는 단어를 예측하는 것이다. 아래의 (그림 1)은 Word2vec의 설계 모델을 도식화한 것이다[12].

Saggion[13]은 오피니언 마이닝 기법이 감성분석에서 언어의 문법적 부분만을 활용하기 때문에 어용론적·의미론적 특징을 고려하지 못한다는 한계를 극복하기 위해서 폭 넓은 의미 자질을 고려하는 Word2vec 방법을 이용하였다. 또한 형용사를 감정을 표출하는 어휘로 분류하고, Word2vec 방법을 이용해 추출된 어휘의 관련 어휘 중 명사를 해당 감정을 일으키는 요소라고 정의했다.



(그림 1) Word2vec의 계산 구조[12]

3. Word2vec을 이용한 감성사전 구축 방법

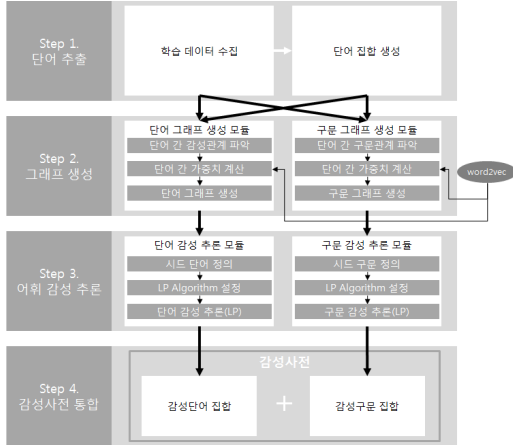
본 장에서는 Word2vec 기법에 기반 한 새로운 가중치 계산방법을 제안하고, 감성사전을 구축하기 위하여 이 기법을 LP (Label Propagation) 알고리즘에 적용하기 위한 방법에 대하여 설명한다.

아래의 (그림 2)는 감성사전을 구축하기 위한 단계들을 나타낸다. Step 1에서는 텍스트 자료로부터 단어와 구문을 추출하게 된다. 이 단계에는 자연어 추출 기법이 적용된다. Step 2에서는 추출한 단어와 구문으로 단어 그래프와 구문 그래프를 구성한다. 단어 그래프에서의 정점은 하나의 단어를 포함하며 변은 단어 간의 감성 관계를 가중치로 가지게 된다. 두 단어가 동일한 감성을 가지는 동감관계일수록 가중치는 높아지게 되고 반대되는 감성을 가지는 반감관계 일수록 낮아지게 된다. 구문 그래프에서는 정점은 하나의 단어를 포함하며 두 정점 사이의 변의 가중치는 두 정점에 포함된 단어들로 구성되는 구문의 감성을 나타낸다. Step 3에서는 Step 2에서 정의된 그래프에 LP 알고리즘을 적용하여 시드 단어 및 구문의 감성을 기반으로 단어와 구문의 감성을 추론하게 된다. Step 4에서는 Step 3에서 추론된 단어와 구문의 감성 집합을 통합하여 감성사전을 생성한다.

3.1 Word2vec을 이용한 가중치 계산법

그래프 기반 텍스트 마이닝에서 변의 가중치를 계산하는 방법들은 대표적으로 PageRank, TextRank, HITS(Hyperlink-Induced Topic Search), PMI, TF-IDF 등이 있다. 이들 가운데 TF-IDF와 PMI는 두 개체간의 연관 정도를 단순히 단어가 문장에서의 빈출 횟수를 계산하여 단어들의 관계를 파악한다[14,15]. 이러한 계산 방법들은 단어들이 문장에서의 맥락이나 문맥 상 의미를 통한 단어들 간의 관계를 전혀 고려하지 않은 유사도 계산 방법이기 때문에 성능의 한계를 가진다[16]. 본 논문에서는 이 같은 한계를 가진 유사도 계산 방법들을 개선하기 위하여 단어의 의미와 문장에서의 맥락을 고려

하여 단어를 벡터로 표현하는 Word2vec을 이용하여 변의 가중치를 계산하는 방법을 제안한다.



(그림 2) 감성사전의 구축 단계

3.2 Word2vec을 이용한 가중치 계산 방법

Word2vec은 단어의 의미를 벡터형태로 표현하는 계량기법으로 의미적으로 유사한 단어들은 거리상 근접한 곳에 위치하게 된다. 이러한 특성을 이용하여 단어 Word2vec에 기반하여 변의 가중치를 계산한다. 각 단어에 대한 Word2vec의 벡터화 결과로 나타내어진 단어벡터로부터 코사인 유사도 계산식 (1)을 이용하여 단어들 간의 유사도를 계산한다.

$$Word2vec(u, v) = \frac{\sum_{i=1}^n (u_i v_i)}{\sqrt{\sum_{i=1}^n (u_i)^2} \sqrt{\sum_{i=1}^n (v_i)^2}} \quad (1)$$

- u_i : 단어벡터 u 의 i 번째 요소

- v_j : 단어벡터 v 의 j 번째 요소

(코사인 유사도를 편의상 Word2vec으로 표기함.)

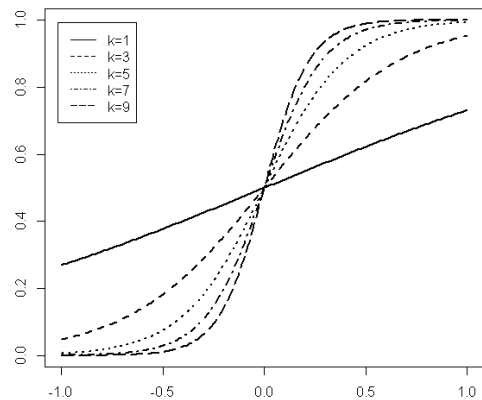
<표 1>은 본 논문의 실험에서 실제 사용된 단어들 중 대표적인 사례들의 Word2vec 코사인 유사도를 나타낸 표이다. 여기서 각 단어 벡터는 300차원의 공간에서 표현 되었으며 각 요소는 $[-1, 1]$ 의

범위로 정규화되었다.

LP 알고리즘에서 변의 가중치에 대한 허용 범위는 $[0, 1]$ 이며 Word2vec 유사도를 허용범위내로 변환하기 위하여 식 (2)의 로지스틱 함수를 사용한다. 아래의 (그림 3)은 k 값에 따른 로지스틱 함수의 그래프 형태를 나타내며, 여기서 $x_0 = 0$, $L = 1$ 로 정하였으며 일반적인 시그모이드 함수의 형태가 된다. 본 논문에서는 감성 사전을 k 값을 기반으로 최적화하기 위하여 다양한 k 값을 가지는 로지스틱 함수를 이용하여 가중치를 계산하였다.

<표 1> Word2vec 단어 벡터의 유사도계산 사례

감성관계	단어 1	단어 2	유사도
강한동감	재미있/VA	좋/VA	0.91
	관심/VA	좋아하/VV	0.86
	재미있/VA	어설프/VA	0.88
강한반감	좋아하/VV	쪽팔리/VV	-0.82
	이쁘/VA	욕먹/VV	-0.72
	재미있/VA	재미없/VA	-0.87



(그림 3) 로지스틱 시그모이드함수

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \quad (2)$$

x_0 : x 값의 중점

L : 곡선의 최대값

k : 곡선의 첨도

Word2vec의 코사인 유사도 계산식 (1)과 로지스틱 함수의 식 (2)를 이용하여 (3)의 식을 얻을 수 있으며, 이를 이용하여 단어 u 와 v 사이의 변의 가중치 w 를 계산한다.

$$w = \frac{1}{1 + e^{-\text{Word2vec}(u, v)}} \quad (3)$$

가중치가 1에 가까울수록 두 단어의 관계가 강한 동감관계라고 하며, -1에 가까울수록 강한 반감 관계이다. <표 2>은 로지스틱 함수의 k 가 3일 경우, 실험에 사용된 강한 동감 관계와 강한 반감 관계를 가지는 대표적인 단어들 간의 가중치 값을 계산하여 나타낸 표이다.

<표 2> Word2vec 기반의 가중치 계산 사례

감성관계	단어 1	단어 2	가중치
강한동감	재미있/VA	좋/VA	0.94
	관찮/VA	좋아하/VV	0.93
	재미없/VA	어설프/VA	0.93
강한반감	좋아하/VV	쪽팔리/VV	0.09
	이쁘/VA	욕먹/VV	0.10
	재미있/VA	재미없/VA	0.07

3.3 LP 알고리즘을 이용한 감성사전 구축

단어 그래프에서는 하나의 단어가 정점(vertex)을 차지하게 되고, 두 정점 사이의 변은 각 정점을 차지하는 두 단어 간의 감성관계를 나타내는 가중치를 가진다. 단어 감성사전 구축을 위해서는 단어 그래프에서 단어들의 감성을 구하기 위하여 준지도적 학습 방법인 LP 알고리즘을 이용한다. 최초의 그래프에는 감성이 이미 알려진 시드 단어들의 정점에만 감성이 정하여져 있고 각 변에는 변의 양 끝에 해당되는 단어들 간의 감성관계로서 식 (3)으로부터 계산된 가중치가 포함되어 있다. LP 알고리즘은 반복적으로 그래프의 정점을 차지하는 각 단어들의 감성을 인접한 다른 단어들의 감성으로부터

유추해 낸다.

구문 그래프에서는 단어 그래프와 비슷하게 단어를 정점으로 하지만, 두 단어가 함께 사용되어 구문을 이루는 구문관계(phrase relation)일 때 단어가 차지하는 두 정점이 서로 변으로 연결된다. 구문 그래프의 경우도 단어그래프와 마찬가지로 LP 알고리즘을 적용하여 구문의 감성을 추론한다.

3.4 Label Propagation

LP를 이용한 단어와 구문 감성 추론은 각각 알고리즘 1 과 2 와 같은 절차대로 이루어지며 자세한 과정과 수렴(Convergence)에 대한 증명은 [17, 18]에 나타나 있다. 두 LP 알고리즘에서 단어 그래프와 구문 그래프 생성 방법은 동일하며 반복 과정에서 가중치 갱신(update) 방법에 차이가 있다.

알고리즘 1. 단어의 감성 추론

- (1) 단어의 감성 매트릭스 정의 및 초기화
- (2) 변의 가중치 매트릭스를 정의
- (3) 감성 매트릭스를 갱신하여 각 단어의 감성을 추론
- (4) (3)의 작업을 그래프가 수렴할 때까지 반복

알고리즘 2. 구문의 감성 추론

- (1) 구문의 감성인 매트릭스 정의 및 초기화
- (2) 단어 그래프의 가중치 매트릭스를 통한 각 단어의 감성을 추론. (수직 방향 LP)
- (3) (2)의 단계를 일정한 값을 가지는 매트릭스에 대해 수렴할 때까지 반복.
- (4) 수렴된 구문의 감성 매트릭스의 각 행 내에서 단어 그래프의 가중치 매트릭스를 통해 각 단어의 감성을 추론. (수평 방향 LP)
- (5) (4)의 단계를 일정한 값을 가지는 매트릭스에 수렴할 때까지 반복

4. 감성사전 구축 실험

본 논문에서는 그래프 기반 감성사전 구축 실험을 실행하기 위해 10,000건의 영화평을 포함하는

데이터를 사용했다. 이 실험에서는 PMI 가중치 식과 Word2vec 가중치 식을 두 가지 LP 알고리즘에 적용하여 감성사전(총 4개의 조합)을 구축하고, 감성사전의 성능을 최적화하기 위하여 로지스틱 함수의 k 값의 변화에 따라 정확도를 계산하였다. 각 LP 알고리즘은 그래프가 수렴할 때까지 반복 수행되었다.

4.1 감성사전 구축 실험 개요

실험 데이터는 네이버에서 제공하는 40자 영화평 데이터이다. 본 실험에서 사용한 영화평 데이터와 그래프에 사용된 단어 및 시드 의 수는 <표 3>와 같다.

<표 3> 실험에 사용된 영화평 데이터의 구성

데이터 구성(단위: 건)				시드(단위: 건)			
텍스트 데이터		단어 및 구문		단어		구문	
양성	음성	단어	구문	양성	음성	양성	음성
5,000	5,000	200	100	5	5	5	5

4.2 영역별 어휘의 감성 추론

실험에서는 단어의 범위를 감성을 표현할 때 주로 사용되는 명사(NNG), 형용사(VA), 동사(VV)로 제한하였고, 구문은 bi-gram 구문만 고려하였다. <표 4>는 단어와 구문은 감성사전을 구축하기 위해 사전에 정의한 시드 단어와 시드 구문을 나열한 것이다.

<표 4> 시드 단어 와 구문

감성	단어	구문
양성	재미있/VA	(스토리/NNG, 좋/VA)
	좋/VA	(스토리/NNG, 재미있/VA)
	관심/VA	(웃음/NNG, 주/VV)
	좋아하/VV	(많/VA, 재미/NNG)
	이쁘/VA	(눈시울/NNG, 붉어지/VV)
음성	재미없/VA	(연기/NNG, 어설프/VA)
	어설프/VA	(재미/NNG, 없/VA)
	나쁘/VA	(역지/NNG, 많/VA)
	쪽팔리/VV	(역지/NNG, 많/VA)
	욕역/VV	(스토리/NNG, 뻔하/VA)

4.1.1 단어 감성사전 실험

실험에서는 아래와 같은 두 가지 LP 알고리즘과 두 가지 가중치 계산 방법을 조합하여 네 가지의 구축 실험을 실행하였다.

방법 1 - Word2vec 가중치 + D. Zhou의 LP 알고리즘

방법 2 - Word2vec 가중치 + X. Zhu의 LP 알고리즘

방법 1' - PMI 가중치 + D. Zhou의 LP 알고리즘

방법 2' - PMI 가중치 + X. Zhu의 LP 알고리즘

<표 5>는 PMI 가중치 계산식을 이용하여 구축한 단어 감성사전의 정확도를 측정한 실험 결과를 나타내고, <표 6>는 영화평 데이터를 Word2vec 가중치 계산식으로 계산하여 구축한 단어 감성사전의 정확도를 측정한 실험 결과를 나타낸다.

<표 5> PMI를 사용한 경우의 단어 감성사전의 정확도(단위: %)

감성	방법 1'	방법 2'
정확도	80.63	77.65

<표 6> Word2vec를 사용한 경우의 단어 감성사전의 정확도(단위: %)

k	정확도	
	방법 1	방법 2
1	89.23	85.77
3	89.53	84.93
5	89.74	85.99
7	90.02	85.45
9	89.34	85.32

실험결과, 두 가지 LP 알고리즘에 Word2vec 가중치를 적용한 경우가 PMI 가중치를 적용했을 경우보다 감성사전의 정확도가 모든 k 값의 경우에 대하여 평균적으로 8.9%, 7.8% 만큼 향상되는 것으로 나타난다.

4.2.2 구문 감성사전 실험

<표 7>은 PMI 가중치 계산식을 이용하여 구축한 구문 감성사전의 정확도를 측정한 실험 결과를 나타내고, <표 8>은 영화평 데이터를 Word2vec 가중치 계산식으로 계산하여 구축한 구문 감성사전의 정확도를 측정한 실험 결과를 나타낸다.

<표 7> PMI를 사용한 경우의 구문감성사전의 정확도(단위: %)

감성	방법 1'	방법 2'
정확도	77.45	70.78

<표 8> Word2vec를 사용한 경우의 구문 감성사전의 정확도(단위: %)

k	정확도	
	방법 1	방법 2
1	78.04	76.28
3	78.78	77.19
5	79.42	77.19
7	79.42	76.98
9	79.09	76.98

단어 사전과 마찬가지로 구문 사전에서도 모든 경우에 Word2vec 가중치를 사용한 감성사전의 정확도가 PMI 가중치를 사용한 감성사전의 정확도보다 높았다. 또한 대부분의 경우에 k값이 3, 5, 7일 때 가장 높은 성능을 보였다.

4.3 감성사전을 이용한 텍스트 감성분석

영화평 데이터로부터 구축된 단어 및 구문 감성사전을 기반으로 새로운 텍스트에 대한 감성분석을 실행하기 위하여 Naive Bayesian[19], 로지스틱 회귀분석[20], 신경망 [21], Random Forest 분류기를 사용하였다. 텍스트 감성분석에서는 5겹 교차 검증(5-fold cross validation) [22]을 사용하였다. 또한, 모든 실험에서 감성 사전을 구축하기 위하여 k값은 5로 설정하였다.

<표 9>는 새로운 텍스트에 대한 감성 분석의 결

과를 정리한 표이다.

<표 9> 새로운 텍스트에 대한 감성 분석 실험 결과(단위: %)

방법 분류기	방법 1'	방법 1	방법 2'	방법 2
Naive Bayesian	71.4	73.9	64.5	66.0
로지스틱 회귀분석	72.9	73.5	74.8	75.2
신경망	76.8	79.4	73.3	75.6
Random Forest	76.9	78.5	74.1	74.9

새로운 텍스트에 대한 감성 추론 실험에서 Word2vec 가중치를 사용하였을 경우 방법 1에서와 같이 D. Zhou의 알고리즘을 적용한 경우에 정확도가 평균적으로 1.8%, 방법 2와 같이 X. Zhu의 알고리즘을 적용한 경우에 1.3% 만큼 향상되었고, 평균적으로는 정확도가 1.5% 개선되었다.

5. 결론

본 논문에서는 새로운 그래프 기반 감성사전 구축 방법을 제안하였다. 이 방법에서는 감성사전의 정확도를 향상시키기 위해 단어의 의미와 문장에서 맥락을 고려하여 단어를 벡터로 표현하는 Word2vec을 단어 그래프에서 변의 가중치 계산식에 적용하였다.

또한, 실제 데이터를 기반으로 제안하는 Word2vec 가중치 계산방법을 감성사전 구축에 적용하여 기존의 방법보다 감성 사전의 성능이 향상됨을 실증하였다. 실제 영화평 데이터를 기반으로 한 실험에서 단어 감성사전과 구문 감성사전의 정확도가 기존의 PMI 방법보다 각각 8.4%, 3.8% 만큼 향상되는 것을 확인하였다. 또한, 구축된 감성사전으로 새로운 텍스트에 대한 감성분류를 실험하였을 때도 Word2vec 가중치를 사용하여 구축한 감성사전을 사용했을 때, 평균적으로 1.5% 향상되었다.

참고문헌

- [1] Kim, S. and Hovy, E., "Determining the sentiment of Opinions", Proceedings of International Conference on Computational Linguistics, p1367, 2004.
- [2] 이철성, 최동희, 김성순, 강재우, "한글 마이크로블로그 텍스트의 감정 분류 및 분석", 정보과학회논문지:데이터베이스, 제40권 제3호, pp.159-167, 2013.
- [3] 김명규, 김정호, 차명훈, 채수환, "텍스트 문서 기반의 감성 인식 시스템", 감성과학, 제12호 제4권, : pp.433-442, 2009.
- [4] 김승우, 김남규, "오피니언 분류의 감성사전 활용효과에 대한 연구"지능정보연구, 제20호 제1호 pp.133-148, 2014.
- [5] 홍초희, 김학수, "트윗 감정 분류를 위한 다양한 기계학습 자질에 대한 비교 연구", 한국콘텐츠학회논문지, 제12권 제12호, pp.471-478, 2012.
- [6] Kim, J. H. "The graph-based method for construction of domain-oriented sentiment dictionary", Ph.D. Dissertaion, Korea Aerospace University, Seoul, Korea, 2014.
- [7] Goldberg, Yoav, Omer Levy, "Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method, arXiv preprint arXiv:1402.3722, 2014.
- [8] Xue, B., Fu, C., Shaobin, Z, "A study on sentiment computing and classification of sina weibo with word2vec", In Big Data, BigData Congress, IEEE International Congress on, pp.358-363, IEEE, 2014.
- [9] Stavrianou, Anna, Periklis Andritsos, Nicolas Nicoloyannis, "Overview and semantic issues of text mining" ACM Sigmod Record Vol.36 No. 3 pp.23-34, 2007.
- [10] Harris, Zellig S., "Distributional structure", Word, Vol.10 No. 2-3, pp. 146-162, 1954.
- [11] 안주영, 배정환, 한남기, 송민, "텍스트 마이닝을 이용한 감정 유발 요인 'Emotion Trigger'에 관한 연구", 지능정보연구, 제21권 제2호, pp.69-92, 2015.
- [12] Mikolov, Thomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781, 2013.
- [13] Saggion, Horacio, and A. Funk. "Interpreting SentiWordNet for opinion classification." Proceedings of the seventh conference on international language resources and evaluation LREC10, 2010.
- [14] 장재영, "텍스트 마이닝을 위한 그래프 기반 텍스트 표현 모델의 연구 동향", 한국인터넷방송통신학회 논문지, 제13권 제5호, pp.37-47, 2013.
- [15] 오연주, 채수환. "영화평과 평점을 이용한 감성 문장 구축을 통한 영화 평점 추론", 인터넷정보학회논문지, 제16권 제2호, pp.41-48, 2015.
- [16] Xue, B., Fu, C., and Shaobin, Z. "A study on sentiment computing and classification of sina weibo with word2vec", In Big Data, BigData Congress, IEEE International Congress on, pp.358-363, IEEE, 2014.
- [17] Zhou, Dengyong, et al. "Learning with local and global consistency." Advances in neural information processing systems, Vol. 16, No. 16, pp.321-328, 2003.
- [18] Zhu, Xiaojin, and Zoubin Ghahramani. "Learning from labeled and unlabeled data with label propagation." Technical Report CMU-CALD-02-107, Carnegie Mellon University, 2002.
- [19] 이현규, 최민국, 신병석, 이상철, "명암도 조절을 통한 베이스 분류기 기반 영상분할기법", 한국차세대컴퓨팅학회 논문지, 제8권 제5호, pp.73-86, 2012. 10.
- [20] 홍정훈, 안성만, 임철수, "데이터마이닝에서 연관분석을 이용한 변수선택", 한국차세대컴퓨팅학회 논문지, 제7권 제3호, pp.20-27, 2011. 6.
- [21] 윤경목, "유전알고리즘과 신경망 파라미터 값들의 변

동에 의한 영화 추천시스템 성능향상”, 한국차세대컴퓨팅학회 논문지, 제10권 제5호, pp.56-64, 2014. 10.

[22] 아흐람 말락, 이상웅, “알츠하이머병 진단을 위한 커널 분류기에서의 PCA의 효과”, 한국차세대컴퓨팅학회 논문지, 제12권 제2호, pp.25-32, 2016. 4.

■ 저자소개

◆ 허찬



- 2013년 한국항공대학교 컴퓨터공학 학사
- 2016년 한국항공대학교 컴퓨터공학 학사
- 관심분야: 자연어처리, 데이터마이닝

◆ 온승엽



- 1984년 서울대학교 전기공학 학사
- 1996년 뉴욕대학교 컴퓨터공학 박사
- 2000년~현재 한국 항공대학교 소프트웨어학과 교수
- 관심분야: 데이터사이언스, 컴퓨터비전