Finding Sentiment Dimension in Vector Space of Movie Reviews: An Unsupervised Approach

Youngsam Kim and Hyopil Shin

Seoul National University, Gwanak-gu, Seoul
kys079@snu.ac.kr, hpshin@snu.ac.kr

This study suggests an unsupervised method to find sentiment orienations of the words in Korean movie reviews. The orientations are represented as real values on a sentiment domain, which is derived from high-dimensional vector space for the movie reviews. To search for the dimension, the Point-wise Mutual Information is first used to select a set of words that are close to common modifiers; The phrases comprised of these words often form good/bad associations (e.g., "good acting", "terrible acting"). A neural language model (Word2Vec) is then used to calculate the point-wise similarity distances between the chosen words and, dimensionality reduction algorithms (e.g., PCA, MDS) are employed to find the axis of the sentiment orientations. Finally, the performance of our method is measured by unsupervised classification of the two movie reviews based on the orientation values. According to the results, the best accuracy achieves 66% and 76% for the two datasets.

Key words: *Sentiment Analysis, Vector Semantics, Unsupervised Approach, Word Space Models*

1. Introduction

Previous research of sentiment analysis or opinion mining mostly focus on supervised methods, which require a labelled training data to identify properties of unseen inputs and classify them later. Although such methods have widely been adopted, there are still many situations in which supervised methods are inappropriate for. When a stream of unlabelled

***Journal of Cognitive Science* 18-1: 85-101, 2017**
Date submitted: 02/25/17 Date reviewed: 03/31/17
Date confirmed for publication: 04/01/17

texts on a topic continues to arrive, supervised methods have no way to analyse the unannotated data online. Even if there is a training set ready for the algorithm, since people's word usage is sensitive to the context of a topic, learning from the training data may not be well generalized for unseen texts. For instance, the word 'acting' would not have neutral meaning, but will have negative polarity when viewers of the movie judge the general performance of the actors/actresses as bad. Thus, an approach for unsupervised and topic-oriented opinion mining is worth pursuing.

Our unsupervised algorithm employs VSM (Vector Space Models) as its main component. A multi-dimensional vector space is constructed from relations between words in a corpus. The basic assumption is that a convex region in the multi-dimensional vector space can represent a sentiment dimension of two-sided opinions over the data points. Given the nature of the task of writing movie reviews, participants normally choose a set of words by their chosen stance to the movie. Although the individuals' attitudes towards the movie will vary, it is not strange to assume that the biggest variance over the attitudes will reflect one dimensional positive/negative aspect of them.

The VSM is deeply related to the distributional hypothesis (Turney and Pantel, 2010). The distributional hypothesis states that words in similar contexts tend to have similar meaning (Rubenstein and Goodenough, 1965; Schütze and Pederson, 1995; Deerwester et al., 1990). One well-known problem with this hypothesis is that the meanings of 'similar contexts' and 'similar meaning' are unclear. The problems of determining them (e.g., size of context window, types of semantic similarity) are not easy matters as studied in a series of research (Pádo and Lapata, 2003; Ruge, 1992; Picard, 1999).

Traditionally, the relation of two words in a 'similar context' has been distinguished into two classes: syntagmatic or paradigmatic relations (Murphy, 2003; Sahlgren, 2006). Syntagmatic relations are concerned if two entities are in a co-occurrence relation and paradigmatic relations question if two items can be replaced interchangeably (substitution relation). Many collocation models using N-grams and Point-wise Mutual Information (PMI) analyse the former-type relation of words. The latter type of relations can be modelled using neural language models (e.g., Word2Vec), which implement probability distributions of neighbouring words within the context window. Note that the two types of relations naturally provide two ways of measuring the distance of words, which is compared in our experiment.

Very few studies of sentiment analysis have used dimensionality reduction method to find a sentiment domain in a high-dimensional data. In this study, dimensionality reduction algorithms are implemented to search for the domain, which represents sentiment orientations of selected words that are close enough to most common modifiers throughout a corpus. It is assumed that an axis which has the largest variance is the sentiment dimension in this study. The assumption is tested by observing correlation coefficients between the orientation values on the domain and the signed log-likelihood ratios of the words that are obtained from the Naive Bayesian analysis of the data (Section. 4).

2. Related Work

Although the majority of previous studies on sentiment analysis have preferred to use supervised methods, some researchers have tried to develop unsupervised or semi-unsupervised approaches. For instance, Turney (2002) suggests the PMI-IR algorithm to estimate the semantic orientation of a phrase for unsupervised classification of various reviews. He uses two pre-chosen words ('poor' and 'excellent') to calculate the semantic orientation of the target phrases, which is defined as the relative PMI difference of the phrase from the two seed words. His work depends on the theory of the semantic orientation of adjectives by Hatzivassiloglou and McKeown (1997). In this study, the authors find the existence of linguistic constraints on the semantic orientations of words in conjunctions. Turney (2002) also measures the semantic orientation of phrases by the PMI differences toward the reference terms (two seed words). Also, Zagibalov and Carroll (2008) attempt to develop an automatic selection process of the seed words in Chinese texts for the unsupervised classifications.

A major issue for our tasks with Turney (2002) is the availability of corpus to calculate the relevant PMI. When the equipped corpus is not big enough for a PMI analysis, the data sparseness problem will arise and the values of the PMI will be hard to be trusted. Note that Turney (2002) used a search engine (AltaVista) for his experiments, which contained 350 million web pages at the time.

Another unavoidable issue with his method is its insensitivity to the context of an individual corpus. The sole expressions like ‘very evil’ or ‘completely mad’ do not make a bad movie themselves (Turney, 2002: 422). They depend on how the viewers of the movie generally respond to the scenes with the lexical items. A fixed dataset for calculation PMI does not likely represent such context-dependent aspects.

However, the PMI is very useful to estimate the co-occurrence relations of words and this model naturally represents the syntagmatic relation between lexical items. Since semantic orientations of a phrase are usually determined by associations with adjectives (e.g., ‘very good’, ‘very bad’), PMI provides appropriate clues on which words should be considered as parts of associations, given a set of common modifiers. Thus, PMI is adopted in our study to select the set of words in the phrases.

In contrast to collocation models, some researchers attempt to apply neural probabilistic language models to measure the semantic similarities of words based on context-window methods (Mikolov et al., 2013; Collobert and Weston, 2008). Neural language models have advantages of being well generalized to unseen data (Bengio et al., 2003), and the word embedding methods (e.g., Word2Vec) have been found more effective for various tasks in NLP than other traditional techniques (Baroni et al., 2014). Continuous Bag-of-Words (CBOW) models and Skip-gram models used in Mikolov et al., (2013) both highly weigh the similarity of words if they share similar neighbouring entities. We note that the implication of the architecture is very similar to the concept of paradigmatic relation of words. Because the distance between any two words in paradigmatic relation is minimized when they share most similar neighbours, for example, in a minuscule corpus which has only two sentences (“This movie is very good” and “This movie is very bad”), the two words (‘good’ and ‘bad’) will likely have a high cosine similarity in the Word2Vec model.

This aspect can cause unexpected results when such model is employed for clustering a set of items that share similar emotions, because two words in paradigmatic relations often instantiate a contrastive relation (e.g., antonym). However, as noted in Mikolov et al. (2013), nouns seem to have multiple syntactic/semantic relations to each other and the Word2Vec model helps to observe the multiple degrees of similarity for words. In this perspective, we could find a specific relation of them in a subspace of the original vector space if we perform the right implementation of vector calculations.

Conceptual Space theory of Gärdenfors (2000) provides a theoretical framework for understanding the phenomenon. He recommends using geometrical structure (so called 'conceptual space') to represent concepts for exploring of quality dimensions. According to his definitions, the important role of the quality dimensions is to depict various "qualities" of objects, i.e., the dimensions show the different ways the entities are judged to be similar or different (Gärdenfors, 2000: 6). We notice that what the quality dimensions indicate is very similar to the concept of 'multiple degrees of similarity' in Mikolov et al. (2013). Gärdenfors (2000) also suggests applying Multidimensional Scaling (MDS) to the similarity-based vector space to generate an ordering relation for data points on an interested domain. For our study, the 'interested domain' will be the sentiment domain and the 'ordering relation' will refer to an assignment function, mapping each word object to a real valued point indicating the level of its 'positive' or 'negative' significance.

3. Methods and Data

The purpose of our experiment is to find sentiment orientations of words in a corpus and evaluate the effectiveness of the information by conducting an unsupervised classification on our movie review datasets. However, before the unsupervised evaluation is undertaken, one additional method is invented and applied for observing the significance of the obtained sentiment orientations. The idea is to define our gold-standard data as a set of tuples of a word and its log-likelihood ratio value, which is provided from Naive Bayesian learning system on reviews per movie. And each log-likelihood ratio is signed as negative if the ratio is given with the label of 'negative opinion' and positive if the label is 'positive opinion'. The method provides the set of words that can be sorted on one-dimensional plane towards negative/positive ends. We believe that this data can be considered as evidential representation of the sentiment dimension if the accuracy of the Naive Bayesian analysis is proved to be high.

3.1 Experiment Methods

First of all, our experiment constructs a vector representing the most central modifiers in a corpus and extracts a number of words closest to the vector. In order to select the modifiers, the sentences of the whole dataset are POS-tagged and a Word-by-Document frequency matrix is formed. The matrix is built to calculate the Betweenness centralities of words and the Normalized PMI between words (two words are linked if they co-occurred in a same review). Betweenness centrality is a measure of the number of shortest paths that go through each node in a graph (Freeman, 1978) and its value of a node in the network is given using Equation (1) where means the total number of shortest paths from node i to node j via node k. The measure is chosen since it gives the information on how the word is used widely with other words throughout the whole review posts.[1] By using the metric, we can avoid choosing popular terms in a particular subset of the reviews and prefer to select globally used modifiers.

$$C_k^{Bet} = \sum_i \sum_j \frac{G_{ikj}}{G_{ij}} \tag{1}$$

Since all the words are POS-tagged, it is easy to select top K modifiers (adjectives or adverbs) from the results of the Betweenness centralities. Then the vectors of the K modifiers in the Word-by-Document matrix are summed to produce one representative vector. The calculations of the cosine distance between the reference vector and the vectors of the other words provide the sorted list of words in an ascending order, in that the top M number of the words are determined for the analysis of sentiment dimension. It can be described as a procedure of finding the best point

[1] However, if the network is strongly centralized, simple frequency would be enough to select the modifiers since in such a network, Betweenness centrality becomes equivalent to Degree centrality, which can be translated as occurrences of a node.

candidates for the exploration of the domain. For our main experiment, we choose the two parameters (K=6 and M=50) after conducting a series of tests, which will be reported in Section 4.

After the candidates are obtained, Normalized PMI or Word2Vec model is implemented to build a pointwise distance matrix for the selected words, resulting in a 50-by-50 matrix when the parameter M is set to 50. Then, three algorithms (PCA, Nonmetric-MDS, t-SNE) of dimensionality reduction are used and the results are compared. Principal Component Analysis (PCA) and NMDS are mathematically close models if the input distance data is Euclidean (Bishop, 2006), but t-distributed stochastic neighbor embedding (t-SNE) is a different technique from the other two algorithms due to its local structure-oriented approach against data (van der Maaten and Hinton, 2008).

All three dimensionality reduction algorithms require the parameter D (number of dimensions to project) to be filled before the computations begin (If D is larger than one, we choose the dimension which has a higher variance). We only use 1 or 2 for the parameter with the following reasons: First, relations between words are usually very complex so the meaning with resultant axes using the parameter (D 3) would be difficult to interpret. Second, we assume that dimensions of the relations would represent one of the two types of the relations of words (syntagmatic/paradigmatic) when the parameter is less than or equal to the largest value ().

The underlying logic for the second assumption is as follows: If similarity distances of words are solely based on their co-occurrences (e.g., PMI), the dimension of the greatest variance would represent syntagmatic relations. Even if semantic distances contain the information on syntagmatic and paradigmatic relation at the same time, the situation does not change. Because we already narrowed down the search space to the subspace for the entities that are found in the phrases of modifiers, the selected M words all share the common adjectives/adverbs near them. Thus, the semantic similarities of the words would not be random in the perspective of paradigmatic relation because they all have at least one common neighbor (the vector of K modifiers). On the other hand, the similarities between the words would have a higher variance in the view of syntagmatic relation (i.e., co-occurrence relation) since some of them could never co-occur in the context window. For instance, think of the small corpus of two sentences in Section 2. The two phrases, ‘very good’ and ‘very bad’ both have the adverb (‘very’) as their common part, but ‘good’ and ‘bad’ have no co-occurrence relation. This condition allows us to find the dimension representing the syntagmatic relation between the words when a vector rotation of the greatest variance in the space is known. Considering that lexical items that express a similar feeling in movie reviews would co-occur more often than words that have an opposite sentiment, the domain found in the analysis would automatically capture the two-sided emotional aspects of the words.

When the dimensionality reduction phase is completed, it is possible to observe a correlation coefficient between values on a detected domain and the gold-standard dataset (the signed log-likelihood ratios of the words). Pearson-r and Spearman-r are both used to measure the correlations in our study. Since signs of the coefficients are irrelevant to our purpose, its absolute forms are only considered.

With the procedure explained above, we can determine two sets of words divided from the origin (0.0) on the domain, which is given by the dimensionality reduction. And, by summing the vectors of the words in the original vector space per each set, a left-ward and right-ward vector is constructed, which is used to calculate the sentiment orientations of the whole words in the corpus with Equation (2).

$$W_k^{s\_ort} = CosDist(LeftVec, W_k) - CosDist(RightVec, W_k) \tag{2}$$

While observing correlation coefficients between the orientations and the signed log-likelihood ratios again, an unsupervised classification is performed in a manner similar to Turney (2002). If a mean of the sentiment orientations in a review post is less than the average of all orientations in the dataset, it is labelled as a ‘left-ward post’ or a ‘right-ward post’ if it is greater than the average. To know the meaning of the label, we used the reference word, ‘최고/NNG’ (meaning ‘the best’ in English). Therefore, if a left-ward or right-ward vector is closer to the reference word than to the other vector, the sentiment direction of the referent vector is interpreted as ‘Positive’. The simplified procedure of our method is presented in Procedure 1.

3.2 Data

Our data consists of two sets of movie reviews: Tidal Wave (2009) directed by Je-Kyoon Yoon and Thirst (2008) by Chanwook Park. They

Procedure 1: Dimensionality reduction based unsupervised sentiment classification

```
Data: Input set I = {$p_1$, $p_2$, $p_i$}, where $p_i = (w_1, w_2, w_j)$
        A seed word believed to have positive polarity (e.g., 'excellent')
Result: Labelled set $L$ = {($p_{1,}$, $t$), ($p_{2,}$, $t$), ($p_{i,}$, $t$) where $t \in$ {'positive', 'negative'}
Begin
    Build word-by-post matrix from dataset I
    Calculate Betweenness centralities of words (using Equation 1)
    Choose top K words with a modifier POS-tag according to the centralities
    Construct one vector by summing the vectors of the K modifiers in the word-by-post matrix
    Compute top M words in the order of being close to the modifier vector using cosine distance
    Build M × M pairwise distance matrix using N-PMI or Word2Vec (distance = 1– similarity)
    Perform a dimensionality reduction algorithm (e.g., PCA or NMDS) on the M × M distance matrix
    Take a dimension which has the largest variance over the entities (if D > 1)
    Group the entities into two sets from origin (0.0) on the plane
    Construct LeftVector and RightVector by summing the vectors of the words per set
    Find which vector has positive direction by calculating a similarity to the seed word
    For $j$ = 1 to J do
        Calculate $w_j^{s\_ort}$ (using Equation 2)
    Let the average of $\sum_j w_j^{s\_ort}$ be θ
    For $i$ = 1 to I do
        Let $S$ = 0
        For $w_j^{s\_ort}$ in $p_i$
            $S \leftarrow S + w_j^{s\_ort}$
        If Mean(S) θ and LeftVector's polarity is not positive Then
            Label $p_i$ as 'negative'
        If Mean(S) θ and LeftVector's polarity is positive Then
            Label $p_i$ as 'positive'
        If Mean(S) θ and RightVector's polarity is not positive Then
            Label $p_i$ as 'negative'
        If Mean(S) θ and RightVector's polarity is positive Then
            Label $p_i$ as 'positive'
End
```

are provided for research purposes from the movie review site of Naver Corporation.[2] Each dataset contains 10,000 review posts and every review is annotated with an integer point ranging from 1 to 10. Because we want the classification task to be binary (negative or positive), reviews with five points are excluded from the datasets. Also, all texts of the reviews are POS-tagged and only nouns/modifiers whose frequency is greater than two are extracted per review. The diminished number of the posts for Tidal Wave and Thirst is 9031 and 9155 respectively. A 10-fold cross-validation was run on the datasets with Naive Bayesian algorithm using NLTK library.[3] The accuracy for Tidal Wave was 83% (RMSE: 0.014) and the result for Thirst was 75% (RMSE: 0.022). Since the accuracies are high, the obtained log-likelihood ratios of the words are believed to be reliable and they are regarded as the gold-standard data in our study.[4]

4. Results

To decide the parameters of the K modifiers and the M closest words, we observed if the correlation coefficients of our results severely changed in different settings of K (3~10) and M (30~60). In the results, the coefficients increased when K 4 and M 40, but slightly decreased as the parameters approach its limit. Based on the observation, we set the parameters (K=6 and M=50) and experimented the effects of the two factors, distance measure (1Normalized PMI or Cosine distance of Word2Vec) with the dimensionality reduction method (PCA, NMDS or t-SNE).[5] Thus, there are 6 experimental conditions (23) in total. Tables (1~4) describe the correlation coefficients between the orientation values of M words and the gold-standard ratios, depending on 6 combinations per movie (note that signs of

[2] http://movie.naver.com/movie/point/af/list.nhn

[3] http://www.nltk.org/_modules/nltk/classify/naivebayes.html

[4] Note that when two ratios exist for an identical word a bigger absolute value is used.

[5] We used Scikit-learn toolkit for implementing of the three algorithms (PCA, MDS, and t-SNE) and Gensim for Word2Vec. See details for the tools at http://scikit-learn.org/stable/ and https://radimrehurek.com/gensim/models/word2vec.html

Table 1. Correlation coefficients of combined models for Tidal Wave (D=1, K=6, M=50)

Model	Pearson	Spearman
N-PMI, PCA	0.395	0.678
N-PMI, NMDS	0.402	0.435
N-PMI, t-SNE	0.104	0.158
W2V, PCA	**0.754**	**0.853**
W2V, NMDS	0.633	0.659
W2V, t-SNE	0.162	0.137

Table 2. Correlation coefficients of combined models for Tidal Wave (D=2, K=6, M=50)

Model	Pearson	Spearman
N-PMI, PCA	0.395	0.678
N-PMI, NMDS	0.635	0.765
N-PMI, t-SNE	0.230	0.355
W2V, PCA	**0.754**	**0.853**
W2V, NMDS	0.672	0.727
W2V, t-SNE	0.006	0.157

Table 3. Correlation coefficients of combined models for Thirst (D=1, K=6, M=50)

Model	Pearson	Spearman
N-PMI, PCA	0.071	0.025
N-PMI, NMDS	0.049	0.071
N-PMI, t-SNE	0.082	0.036
W2V, PCA	**0.487**	**0.588**
W2V, NMDS	0.281	0.265
W2V, t-SNE	0.125	0.001

the coefficients are ignored).

As it can be seen in the tables, the combination of Word2Vec and PCA outperforms the other models in all cases. Since PCA uses an orthogonal transformation to find linearly uncorrelated variables (a principal component) and the first component always contains the largest variance of the data, increasing D does not affect the results. Using the N-PMI distance

Table 4. Correlation coefficients of combined models for Thirst (D=2, K=6, M=50)

Model	Pearson	Spearman
N-PMI, PCA	0.071	0.025
N-PMI, NMDS	0.409	0.554
N-PMI, t-SNE	0.019	0.022
W2V, PCA	**0.487**	**0.588**
W2V, NMDS	0.472	0.576
W2V, t-SNE	0.064	0.262

Table 5. Correlation coefficients of all words and the gold-standard data (D=2, K=6, M=50)

Movie	Model	Pearson	Spearman
Tidal Wave	W2V-PCA	0.287	0.333
	W2V-NMDS	**0.293**	**0.340**
Thirst	W2V-PCA	0.201	0.213
	W2V-NMDS	**0.207**	**0.224**

Table 6. Accuracy of unsupervised classification results (D=2, K=6, M=50)

Movie	Model	Accuracy
Tidal Wave	W2V-PCA	**0.76**
	W2V-NMDS	**0.76**
Thirst	W2V-PCA	0.63
	W2V-NMDS	**0.66**

measure did not improve coefficients except for using NMDS, but the improvements appear only when the parameter D is set as 2. Interestingly, NMDS seems not show its strength when D is equal to 1. We notice that t-SNE always records the poor results.

Based on the results above, we adopted the two models (W2V-PCA and W2V-NMDS) for our unsupervised classification task following the steps of Procedure 1. During the process, the correlation coefficients between the orientations of the whole words and the signed log-likelihood ratios are recorded. As a result, severe decreases of coefficients emerged due to the cost of the extrapolative process, that is, expanding the scope of the sentiment orientations to all items of the corpus (See Table 5).

One way to see whether the sentiment orientations in Table 5 are useful is to perform unsupervised sentiment classifications for the movie reviews using the values. Following Procedure 1, the accuracies calculated in our experiment are presented in Table 6. It clearly shows that the orientations are helpful in predicting the polarity of a review post.

5. Discussion and Conclusion

Given the results of Table 1~4, one of the questions naturally arising is why PMI-based distance records a poorer performance compared to Word2Vec-based distance records. One plausible explanation is that the size of the dataset was relatively small to apply PMI models. As noted in Section 3.2, the dataset per movie is around 9,000 reviews and the length of a review is usually short (one or two sentences). Co-occurrence based methods often encounter data sparseness problems, and in such a case, point-wise distances between words would not be differential. A neural language model, Word2Vec seems to have more strength on the subject, showing stronger results at most times.

The superiority between PCA and NMDS is hard to determine with the results we obtained. Table 5 and Table 6 show that W2V-NMDS slightly outperforms the W2V-PCA, but the difference is marginal. Particularly, NMDS needs random initialization to start its computation, thus the result varies from time to time and a large-scale experiment is required to answer the question.

NMDS also attempts to preserve point-wise distances between entities as well as possible in lower space; On the other hand, PCA finds the first component that accounts for the largest variance of data. This might explain why the outcomes of NMDS improved dramatically when the number of dimensions to project was more than one. Since PCA does not have such a constraint, giving one for the parameter of the PCA would suffice to the task. However, it is worth to note that when D was set as two, NMDS found a domain which shows the strongest correlation with the gold-standard set in all PMI-distance conditions. It might indicate that NMDS is more powerful than PCA to capture the degrees of mutual-relations of entities

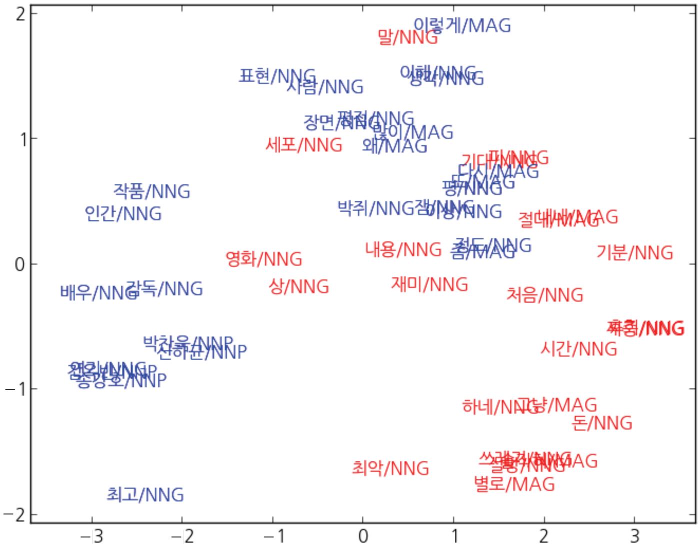

Figure 1. Scatter plot of the W2V-PCA results on the 50 candidates in Thirst (D=2): Different colors represent the signs of the log-likelihood ratios indicating its polarity

with an impoverished similarity data.

The superior performances of PCA and NMDS to that of t-SNE might be explained by the difference in architecture. Classical scaling methods such as PCA and MDS try to preserve the distances between widely separated data points rather than preserving the distances between nearby points (van der Maaten and Hinton, 2008). In contrast to the classical methods, t-SNE aims to describe the local structure of the data. Therefore, for data that has a high intrinsic dimensionality, which is the case for our data, the local linearity assumption t-SNE makes would be violated (van der Maaten and Hinton, 2008: 2599). Based on the theoretical facts and the experimental results, we conclude that dimensionality reduction algorithms which look for the global structure of the data would be preferable for our method.

Implementing dimensionality reduction algorithms seems to reveal the co-occurrence relation of words even for a data that is generated by Word2Vec which adopts the Skip-gram/CBOW model. Figure 1 graphically presents an example of capturing the aspect of the syntagmatic/paradigmatic relation types. In our dataset of Thirst, the cosine distance between '최고/NNG' (the best) and '최악/NNG' (the worst) is somewhat close (0.49) by Word2Vec. Interestingly, the distance between the two words on the x-axis (the first component) is greater than the distance on the y-axis (the second component) as predicted in our method (see the lower left region of the figure). This property shows why dimensionality reduction is very useful for our task, especially when Word2Vec-based distance measure is used.

We believe that our suggested method could be useful for various opinion mining tasks, which need to extract particular information on contexts of an unannotated corpus. For instance, one might wonder how people judge the performance of actors/actresses in a movie. Using our method, it is possible to observe if a word referring to 'acting' is positive or negative *within* the corpus. In our data of Tidal Wave, the sentiment orientation of '연기/NNG' (acting) is 0.13 in negative direction while its signed log-likelihood ratio records . On the other hand, the orientation of the word in Thirst turns out to be 1.23 in positive direction and the log-likelihood ratio accordingly marks 2.5. In this spirit, our method can be regarded as a within-corpus approach.

To our knowledge, prior studies of sentiment analysis have not focused on using dimensionality reduction methods to find a sentiment domain in a high-dimensional vector space. We think this approach has several advantages compared to the previous methods. First, it needs less data as shown in our experiment. Second, it requires minimal lexical knowledge for the task (a seed word of positive meaning). However, if we want to focus on contrastive usage of phrases, even the minimal information would not be needed. Third, the method is unsupervised and context-independent since the degree of closeness between two words would be determined intrinsically. In addition, we believe our method is applicable to most other languages other than Korean because of the minimal requirement for linguistic components.

Finally, several warnings or limitations of our study should be noted. Our method assumes that there is a dimension that represents sentiment orientations of words in a vector space. The Tables (1~4) obviously indicate the existence of such dimension for the selected candidates, but we observed decreases in the correlation results after getting orientations for words corresponding to all the gold-standard ratios (Table 5). Although the result might mean that extending the scope of the dimension into all words would cost some noises, the decline could partially be due to the poor differentiability of the signed log-likelihood ratios. Many of the ratios have discrete values such as 1 or -1, and they might reduce the correlation coefficients because those cases would make the general pattern of data-points look less linear. If such is the case, why Spearman's coefficients were generally higher than Person's throughout Table 1~5 is explained because Spearman's rho only concerns rank correlations. A technique that can find more generalizable sentiment domain from the original multi-dimensional space is left for future study.

Also, because our method uses the Bag-of-words model and vector representation, the suggested approach is vulnerable to situations where combinatorial complexity of lexical items is very high. One of our movies, Tidal Wave is a disaster film which has a simple story structure and characters, thus the sentiment types of comments on the movie are easily distinguished. This factor appeared indeed in the difference of the accuracy results (76% of Tidal Wave vs. 66% of Thirst). Note that this gap is comparable to the difference of the Naive Bayesian classification results (83% vs. 75%).

Reference

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Paper presented at the ACL* (1).

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, 3(Feb), 1137-1155.

Bishop, C. (2007). *Pattern Recognition and Machine Learning* (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn: Springer, New York.

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. *Paper presented at* *the Proceedings of the 25th international conference on Machine learning*.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6), 391.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*: MIT press.

Hatzivassiloglou, V., & McKeown, K. R. 1997. Predicting the semantic orientation of adjectives. Paper presented at the *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov), 2579-2605.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*: Cambridge University Press.

Padó, S., & Lapata, M. (2003). Constructing semantic space models from parsed corpora. *Paper presented at the Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume* 1.

Picard, J. (1999). Finding content-bearing terms using term similarities. *Paper presented at the Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*.

Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Communications of the ACM*, 8(10), 627-633.

Ruge, G. (1992). Experiments on linguistically-based term associations. *Information Processing & Management*, 28(3), 317-332.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, 20(1), 33-54.

Schütze, H., & Pedersen, J. O. (1995). Information retrieval based on word senses. *Paper presented at the Proceedings of the 4th Annual Symposium on Document Analysis and Information Retrieval*, 161-175.

Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *Paper presented at the Proceedings of the 40th annual meeting on association for computational linguistics.*

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1), 141-188.

Zagibalov, T., & Carroll, J. (2008). Automatic seed word selection for unsupervised sentiment classification of Chinese text. *Paper presented at the Proceedings of the 22nd International Conference on Computational Linguistics-Volume* 1.