Sense Sentiment Similarity: An Analysis

Mitra Mohtarami[1], Hadi Amiri[1,2], Man Lan[3], Thanh Phu Tran[1], and Chew Lim Tan[1]

[1]Department of Computer Science, National University of Singapore;
[2]NUS Graduate School for Integrative Sciences and Engineering;
[3]Institute for Infocomm Research
{mitra, hadi}@comp.nus.edu.sg; lanman@i2r.a-star.edu.sg; {tancl, thanhphu}@comp.nus.edu.sg

Abstract

This paper describes an emotion-based approach to acquire *sentiment similarity* of word pairs with respect to their senses. Sentiment similarity indicates the similarity between two words from their underlying sentiments. Our approach is built on a model which maps from senses of words to vectors of twelve basic emotions. The emotional vectors are used to measure the sentiment similarity of word pairs. We show the utility of measuring sentiment similarity in two main natural language processing tasks, namely, indirect yes/no question answer pairs (IQAP) Inference and sentiment orientation (SO) prediction. Extensive experiments demonstrate that our approach can effectively capture the sentiment similarity of word pairs and utilize this information to address the above mentioned tasks.

Introduction

This work focuses on the task of measuring sentiment similarity of word pairs. Sentiment similarity reflects the distance between words regarding their underlying sentiments. Many approaches have been proposed to capture the semantic similarity between the words to date; Latent Semantic Analysis (LSA), Point-wise Mutual Information (PMI), and WordNet-based similarity method are some examples of the semantic similarity measures.

These measures are good for relating semantically related words like "*car*" and "*automobile*", but are less effective in relating words with similar sentiment like "*excellent*" and "*superior*". For example, the following relations show the semantic similarity between some sentiment word pairs computed by LSA (Landauer, Foltz, and Laham 1998) and their arranged relations.

$$\begin{aligned} LSA\,(excellent, superior) &= 0.40 \\ &< LSA\,(excellent, good) = 0.46 \\ &< LSA\,(good, bad) = 0.65 \end{aligned}$$

Clearly, the sentiment similarity between these words should be in the reversed order. In fact, although the terms "*excellent*", "*superior*" and "*good*" have the same sentiment orientation (positive), the intensity of sentiment in "*excellent*" is more similar to "*superior*" than "*good*". Thus, ideally, sentiment similarity of "*excellent*" and "*superior*" should be greater than "*excellent*" and "*good*" and as the terms "*good*" and "*bad*" are opposite in sentiment, their sentiment similarity should be zero.

To date, sentiment similarity has not received enough attention. In fact, the majority of previous works employed semantic similarity as a measure to compute sentiment similarity of word pairs (Kim and Hovy 2004; Turney and Littman 2003). In this paper, we propose a principled approach to detect the sentiment similarity of word pairs with respect to their senses and their underlying sentiments. We introduce 12 basic emotions dedicated to sentiment similarity. Our method computes the sentiment similarity of word pairs based on the connection between their lexical semantics and basic emotions. We show that it effectively outperforms the semantic similarity measures that were used to predict sentiment similarity.

Furthermore, we show the utility of sentiment similarity prediction in two NLP tasks, namely, *Indirect yes/no Question Answer Pairs* (IQAPs) *Inference*, *Sentiment Orientation (SO) prediction*. We briefly explain the utility of sentiment similarity for these two tasks:

In IQAPs, the answer of a question-answer pair does not explicitly contain a clear *yes* or *no* word, but rather gives information which can be used to infer such an answer. Therefore, the task is to infer the *yes* or *no* answer for a given question-answer pair, see Table 1 for some examples of IQAPs. In some cases, interpreting the answer is straightforward, e.g. E1, but in many cases the answerer shifts the topic slightly, e.g. E2 and E3. In these cases, the interference task is more difficult.

Clearly, the sentiment words of the question and answer of an IQAP are the pivots that determine the final answer as *yes* or *no*. We show that the sentiment similarity between the adjectives in the IQAPs can be used to effectively infer the *yes* or *no* answers. For example, in E1, though the adjective "*acceptable*" has weaker sentiment intensity than the adjective "*great*", the sentiment similarity between the two adjectives is sufficiently high to infer a *weak-yes* answer. However, if the answer contains an adjective with higher sentiment similarity with "*great*", e.g. "*excellent*", then the answer would be inferred as *strong-yes*. This is the same for other examples.

Row	IQAP	Answer
E1	**Q**: Do you think that's a ***great*** idea? **A**: I think it's ***acceptable***.	*weak-yes*
E2	**Q**: Was she the ***best*** one on that old show? **A**: She was simply ***funny***.	*strong-yes*
E3	**Q**: He says he opposes amnesty, but …. Is he ***right***? **A**: He is a bit ***funny***.	*weak-no*
E4	**Q**: … Is that ***true***? **A**: This is ***extraordinary*** and preposterous.	*strong-no*

Table 1. Examples of IQAPs

As the second application, we predict the sentiment orientation of words. Previous research utilized (a) word relations obtained from WordNet (Kim and Hovy 2004; Hassan and Radev 2010), (b) external resources like review rating (Marneffe, Manning, and Potts 2010), and (c) semantic similarity measures for this purpose (Turney and Littman 2003; Kanayama and Nasukawa 2006). We show that sentiment similarity is a more appropriate measure to achieve accurate sentiment orientation of words.

The sentiment similarity may also vary with respect to different senses of the words. For example, in E4, if we use the third sense of the adjective "*extraordinary*", i.e. "*unusual*", we can infer the correct answer, *no*. This is because the sentiment similarity between "*unusual*" and "*true*" is low. This is while the first sense (the most common sense) of "*extraordinary*" means "*bonzer*" that has sufficiently strong sentiment similarity with the adjective "*true*". Therefore the answer will be incorrectly interpreted as *yes* in the latter case.

In summary, the contributions of this paper are follows:

- We propose an effective method to predict the sentiment similarity between word pairs at the sense level,
- We show that such sentiment similarity can better reflect the similarity between sentiment words than semantic similarity measures, and
- We show the utility of sentiment similarity in IQAP inference and SO prediction tasks.

The experiments in sentiment prediction show that our sentiment similarity method significantly outperforms two baselines by 6.85% and 18.1% improvements in F1. It also outperforms the best performing baseline for the IQAP task by 17.93% improvements in F1.

Method: Sense Sentiment Similarity

People often show their sentiment with various emotions, such as "*crying*" or "*laughing*". Although the emotions can be categorized into *positive* and *negative* sentiments, human have different feelings with respect to each emotion. For example, "*anger*" and "*fear*" have *negative* sentiments; however they reflect different feelings. Figure 1 illustrates different human reactions with respect to different emotions. The intensity of sentiment in each emotion is different from others.

desire *Joy* *sadness* *anger* *fear*

Figure 1. Examples of affective emotional states

Human behavior as a result of his emotions can be presented via the look on his face (e.g., Figure 1), the sound of his voice, or opinion words expressed in his writing/speaking. Since the opinion words carry a range of human emotions, they can be represented as a vector of emotional intensities. Emotion intensity values describe the intensity degree of emotions that can be varied from "*very weak*" to "*very strong*". For example, Table 2, adapted from Neviarouskaya, Prendinger, and Ishizuka (2009), shows several sample opinion words and their corresponding intensity values with respect to different emotions. For example, the verb "*regret*" has intensity values of 0.2 and 0.1 with respect to the "*guilt*" and "*sadness*" emotions respectively.

Word	POS	Intensity Values
tremendous	Adj.	surprise:1.0; joy:0.5; fear:0.1
success	noun	joy:0.9; interest:0.6; surprise:0.5
regret	verb	guilt:0.2; sadness:0.1

Table 2. Examples of words with intensity emotions.

We propose to predict the sentiment similarity between the senses of the words using the words' emotional vectors constructed from their intensities. We follow three steps to achieve this aim:

Designing Basic Emotional Categories

Previous researches showed that there exists a small set of basic (or fundamental) emotions which are central to other emotions (Ortony and Turner 1990; Izard 1971). Though there is little agreement about the number and types of basic emotions, some sets of basic emotions are central and generally accepted (Ortony and Turner 1990).

We employ the basic emotional set studied in (Izard 1971; Neviarouskaya, Prendinger, and Ishizuka 2009) as its basic emotions appear in more number of emotional sets and have higher coverage than others. The basic emotions are: *anger*, *disgust*, *fear*, *guilt*, *sadness*, *shame*, *interest*, *joy*, *surprise*. We considered the first six emotions as negative emotions and the other three as positive. To have a balance number of positive and negative emotions, we also employ three other positive basic emotions adapted from Ortony and Turner (1990): *desire*, *love*, *courage*.

We extend each basic emotion to an emotional category. For this purpose, we use the hierarchical synonyms of the basic emotions; we refer to these words as seeds. For each basic emotion, we pick its synonyms with the following constraints:

1. Relevant Seeds: Having the highest semantic similarity scores (computed by LSA) with the basic emotion, and
2. Balanced Matrix: The total occurrences of all the selected seeds for each category in our corpus remains balanced over the emotional categories

As an example, Table 3 shows some selected seeds and their semantic similarity values with their corresponding basic emotions[1].

Desire	**Joy**	**sadness**
cherished, 0.54	delight, 0.63	depressive,0.55
enthusiasm, 0.47	excitement, 0.6	sad, 0.54
ambition, 0.46	happy, 0.59	weepy, 0.54
honest, 0.46	glorious, 0.58	grief, 0.53
intimate, 0.45	pleasure, 0.57	loneliness, 0.51

Table 3. Examples of seed words in emotional categories

Constructing Emotional Vectors

In this step, we construct an emotional vector like $(I_1, I_2, \dots, I_{12})$ for each word w where each I_k represents the intensity of k^{th} emotion in w. For instance, I_1 represents the intensity of "*anger*" and I_{12} indicates the intensity of "*courage*" emotion in the word w.

We employ the hypothesis that a word can be characterized by its neighbors (Turney and Littman 2003). That is, the emotional vector of a word tends to correspond to the emotional vectors of its neighbors. Therefore, we use the sum of the co-occurrences of w with each seed in an emotional category to estimate the intensity value of w with the corresponding emotion as shown in Equation 1.

$$I_k = Intensity(w, cat_k) = \sum_{seed_j \in cat_k} co\ occur(w, seed_j) \tag{1}$$

[1] Hierarchical synonyms can be obtained from thesaurus.com, and semantic similarity computed by LSA.

where, I_k is the overall intensity value of w with the k^{th} emotional category, cat_k, and $seed_j$ a seed word in cat_k.

Note that employing co-occurrence is critical for words whose emotional meanings are part of common sense knowledge and not explicit (e.g., the terms "*mum*", "*ghost*", and "*war*"). The emotional intensity of such words can be detected based on their co-occurrence patterns with words with explicit emotional meanings, e.g. seeds.

In addition, a problem with the corpus-based co-occurrence of w and cat_k is that w may never (or rarely) co-occur with the seeds of an emotion. This results in a very weak intensity value of w in cat_k. We utilize synsets to tackle this issue. As the synset of a word has the same or nearly the same meaning as the original word, the word can be replaced by any of its synset with no major changes in its emotion. Therefore, we expect the synset to improve the predicted value for intensity of w in cat_k and hence better estimate sentiment similarity between words.

Furthermore, the major advantage of using synset is that we can obtain different emotional vectors for each sense of a word and predict the sentiment similarity at the sense level. Note that, various senses of a word can have diverse meanings and emotions, and consequently different emotional vectors. Using synsets, the intensity value of w in an emotional category is computed by the sum of the intensity value of each word in the synset of w with the emotional category as presented in Equation 2.

$$I_k = \sum_{syn_i \in synset(w, sense(w))} Intensity(syn_i, cat_k) \tag{2}$$

where, $synset(w, sense(w))$ is the synset of a particular sense of *w*, and $Intensity(syn_i, cat_k)$ is computed using Equation 1.

Word Pair Sentiment Similarity

To compute the sentiment similarity between two words with respect to their senses, we use the correlation coefficient between their emotional vectors. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be the emotional vectors of two words. Equation 3 computes their correlation:

$$corr(X, Y) = \frac{\sum_{i=1}^{n} (X_i \quad X)(Y_i \quad Y)}{(n \quad 1) S_X S_Y} \tag{3}$$

where, $n = 12$ is number of emotional categories, X, Y and S_X, S_Y are the mean and standard deviation values of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively.

The above Equation measures the strength of a linear relationship between two vectors. This value varies between -1 (strong negative) and +1 (strong positive). The strong negative value between two vectors means they are completely dissimilar, and the strong positive value means the vectors have perfect similarity.

Having the correlation value between two words, the problem is that how large the correlation value should be such that we can consider the two words as similar in sentiment. We address this issue by utilizing the antonyms of the words. For this purpose, we take an approach similar to the work in (Mohtarami et al. 2011). Since the antonym of a word belongs to the same scalar group (e.g., *hot* and *cold*) and has different sentiment orientation with the word, we consider two words, w_i and w_j as similar in sentiment iff they satisfy both of the following conditions:

1. $corr(w_i, w_j) > corr(w_i, {\sim}w_j), and$
2. $corr(w_i, w_j) > corr({\sim}w_i, w_j)$

where, ${\sim}w_i$ and ${\sim}w_j$ are antonyms of w_i and w_j respectively, and $corr(w_i, w_j)$ is the correlation between the emotional vectors of w_i and w_j obtained from Equation (3). Finally, we compute the sentiment similarity (*SS*) between two words as follows:

$$SS(w_i, w_j) = corr(w_i, w_j) \quad Max\{corr(w_i, {\sim}w_j), corr({\sim}w_i, w_j)\} \quad (4)$$

A positive value of $SS(.,.)$ indicates that the words are sentimentally similar. The large value indicates strong similarity and small value shows weak similarity. Likewise, a negative value of $SS(.,.)$ shows the amount of dissimilarity between the words.

Applications

In this section we explain how sentiment similarity can be used to perform IQAP inference and predict the sentiment orientation of words respectively.

IQAP Inference

In IQAPs, the adjectives in the question and its corresponding answer are the main factors to infer *yes* or *no* answers. We employ the association between the adjectives in questions and their answers to interpret the indirect answers. Figure 2 shows the algorithm we used for this purpose. Note that *SS*(.,.) indicates sentiment similarity computed by our method (see Equation 4). As we discussed before, the positive *SS* between words means they are sentimentally similar which can vary from *weak* to *strong*, this leads to infer *weak-yes* or *strong-yes* response that conveys *yes*. However, negative *SS* indicates that the words are not sentimentally similar and results in *weak/strong-no* which leads to the *no* response.

Input:

SAQ: The adjective in the question of the given IQAP.

SAA: The adjective in the answer of the given IQAP.

Output:

answer $\in \{yes, no, uncertain\}$

Algorithm:

1. if SAQ or SAA are missing from our corpus then
2. answer=*Uncertain*;
3. else if $SS(SAQ, SAA) < 0$ then
4. answer=*No*;
5. else if $SS(SAQ, SAA) > 0$ then
6. answer=*yes*;

Figure 2. Decision procedure of employing sentiment similarity to IQAP inference.

Sentiment Orientation Prediction

We aim to compute more accurate sentiment orientation (SO) using our sentiment similarity method than any other semantic similarity measures.

Turney and Littman (2003) proposed a method in which the sentiment orientation of a given word is calculated from its contextual/semantic similarity with seven positive words like "*excellent*", minus its similarity with seven negative words like "*poor*" as shown in Figure 3.

Input:

$Pwords$: seven words with positive sentiment orientation

$Nwords$: seven words with negative sentiment orientation

$A(.,.)$: similarity function that measures the similarity between its arguments

w: a given word with unknown sentiment orientation

Output:

P: sentiment orientation of *w*

Algorithm:

1. $P = SO\_A(w) =$

$$\sum_{pword \in Pwords} A(w, pword) - \sum_{nword \in Nwords} A(w, nword)$$

Figure 3. Procedure to predict sentiment orientation (SO) of a word based on the similarity function *A*(.,.)

As the similarity function, A(.,.), they employed point-wise mutual information (PMI) and LSA to compute the similarity between the words. We utilize the same approach, but instead PMI or LSA we use our *SS*(.,.) measure as the similarity function.

PMI between two words measures the mutual dependence of them and is defined as follows (Turney and Littman 2003):

$$PMI(w_1, w_2) = \log_2\left(\frac{P(w_1, w_2)}{P(w_1)\,P(w_2)}\right) \quad (5)$$

where $P(w_1, w_2)$ is the probability that w_1 and w_2 co-occur, and $P(w_1)$ and $P(w_2)$ are the probability of w_1 and w_2.

LSA performs several steps to compute the semantic similarity between two words. First, it forms a matrix with documents as rows and words as columns. Cells contain the number of times that a given word is used in a given document. Second, it attempts to reduce the high dimensional semantic space and compute the similarity of two words by the cosine between their corresponding vectors in the semantic space. LSA and to some extent PMI only utilize the semantic space and ignore the emotional space, whereas our SS measure effectively utilizes the emotional space.

Evaluation and Results

In this section we first explain the datasets used, and then report the experiments conducted to evaluate our approach.

Data and Settings

We used the review dataset developed by Maas et al. (2011) as the development dataset to compute the co-occurrences of word pairs. This dataset contains 50k movie reviews and 90k vocabulary. We consider a window of 10 words to compute co-occurrences.

We also employed the standard TASA corpus to compute the semantic similarity of word pairs for LSA. This corpus contains around 61K documents and 155K vocabulary. We believe that LSA with TASA produces better performance than our development dataset. This is because our corpus is smaller than TASA and it contains user generated text which is known to be grammatically week with many spelling errors and slangs. However, TASA is adapted from 6,333 textbooks and does not have the above issues.

For the evaluation purpose, we used two datasets: the MPQA (Wilson, Wiebe, and Hoffmann 2005) and IQAPs (Marneffe, Manning, and Potts 2010) datasets. The MPQA dataset is used for SO prediction experiments, while the IQAP dataset is used for the IQAP experiments. For MPQA dataset, we ignore the neutral words and use the remaining 4000 opinion words with their sentiment orientations. The IQAPs dataset contains a 125 IQAPs and their corresponding *yes* or *no* labels as the ground truth as described in (Marneffe, Manning, and Potts 2010).

Experimental Results

IQAP Inference Evaluation

Table 4 shows the evaluation results for the task IQAPs. The first row presents the result obtained by the approach proposed by Marneffe, Manning, and Potts (2010). This is our baseline and obtained an accuracy of 60% on the IQAP dataset. As explained in Section "*Related Works*", their decision procedure is based on the individual sentiment orientation of the adjectives in question and its corresponding answer and does not consider the correlation between the two adjectives. However, our approach is able to directly infer *yes* or *no* responses using sentiment similarity between the adjectives and does not require computing sentiment orientation.

Method	Precision	Recall	F1
Marneffe et al. (2010)	60.00	60.00	60.00
PMI	60.61	58.70	59.64
LSA	66.70	54.95	60.26
SS (w/o WSD)	75.03	77.85	76.41
SS (with WSD)	**76.69**	**79.75**	**78.19**

Table 4. Performance on IQAPs

The second and third rows of Table 4 show the results of using PMI and LSA as the sentiment similarity (SS) measures in the algorithm explained in Figure 2. The last rows, *SS (with WSD)* and *SS (w/o WSD)* indicates the results when we use our sentiment similarity measures with and without WSD respectively. *SS (w/o WSD)* is based on the first sense (most common sense) of the words, whereas *SS (with WSD)* utilizes the real sense of the words. We manually annotate the sense of the adjectives to investigate the importance of WSD in a perfect setting. The results show that they significantly improve the performance of the best performing baseline (LSA) by 16.15% and 17.93% F1 improvements. Furthermore, as it is clear in Table 4, using correct sense of the adjectives increase the performance from 76.41% to 78.19%. However, this difference is not significant because only 14% of the adjectives are assigned senses different from their first senses. The efficiency of the WSD could have been more highlighted, if more IQAPs contain adjectives with senses different from their first senses.

Evaluation of Sentiment Orientation Prediction

Table 5 shows the results of word sentiment prediction. The results in the table are based on the algorithm in Figure 3 where PMI, LSA and SS (our method) are used for calculating the similarity between two words respectively. As it is shown, LSA significantly outperforms PMI. It was expected since PMI is known as a contextual similarity measure which is based on co-occurrence of word pairs. Furthermore, our development dataset is relatively small and this leads to poor co-occurrence information. The SS method utilizes the fist sense of the words here and significantly outperforms the two baselines. It outperforms PMI and LSA by 18.1% and 6.85% respectively. The SS method, in contrast to PMI, does not require big development dataset to perform well.

Method	Precision	Recall	F1
SO-PMI	56.20	56.36	55.01
SO-LSA	66.31	66.89	66.26
SO-SS	**73.07**	**73.89**	**73.11**

Table 5. Performance on SO prediction

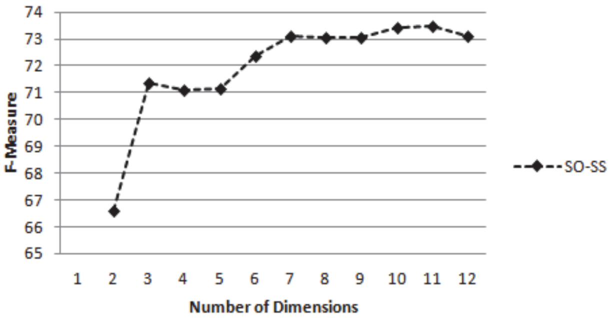

Figure 4. Dimensions reduction

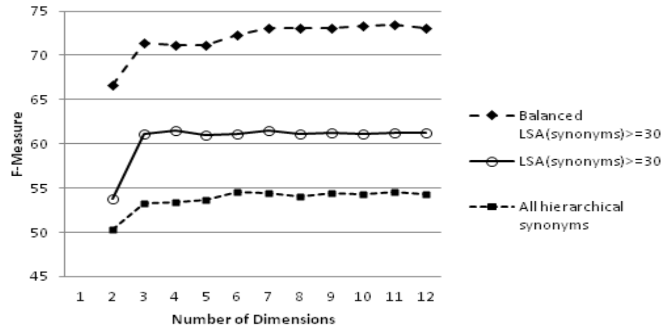

Figure 5. Selection of emotional categories

Analysis and Discussion

In this section, we explore the role of using singular value decomposition (SVD) and different emotional categories. In addition, we study the effect of synsets and antonyms of words for predicting their sentiment similarity. We investigate these factors on the sentiment prediction task.

Role of using SVD: To study the role of SVD, we construct an *emotional matrix* using the emotional vectors of words and their antonyms with respect to their senses.

$$\begin{array}{r c} & \begin{array}{cccc} anger & disgust & \dots & courage \end{array} \\ \begin{array}{r} w_1, sense(w_1) \\ w_2, sense(w_2) \\ \vdots \\ w_i, sense(w_i) \end{array} & \begin{bmatrix} I_1 & I_2 & \cdots & I_{12} \\ I'_1 & I'_2 & \cdots & I'_{12} \\ \vdots & \vdots & \dots & \vdots \\ I''_1 & I''_2 & \dots & I''_{12} \end{bmatrix} \end{array}$$

Our *SS* measure works based on the co-occurrence between words and emotional categories. Thus, some inappropriate words may add some noise to the vectors and emotional matrix. Running SVD allows us to collapse the matrix into a smaller dimensional space where highly correlated items are captured as a single feature. In other words, it makes the best possible reconstruction of the matrix with the least possible information and can potentially reduce the noise coming from the co-occurrence information. It can also emphasize the strong patterns and trends.

We repeat the experiments on the sentiment prediction task using SVD with different dimensional reductions. Figure 4 shows the results. As it is shows, higher performances can be achieved with greater dimensions. The highest performance occurs in the dimension 11 which is 73.50%. The results also show that the dimensions lower than three results in great reductions in the performance, whereas there are no big performance reductions in the greater dimensions. We believe this is because of the use of synsets that can highly resist against the co-occurrence noise in the data.

Role of emotional categories: As explained in Section "*Designing Basic Emotional Categories*", we construct emotional categories from hierarchical synonyms of the basic emotions (we referred to them as seeds). Here, we repeat the experiments on the sentiment prediction task by three sets of emotional categories to illustrate the importance of the two constraints.

Figure 5 shows that if we use "*all hierarchical synonyms*" as seeds, the performance of sentiment prediction is poor. The reason is that some irrelevant seeds may enter into the emotional categories solely due to their distance in the hierarchical synonyms.

To only utilize the relevant seeds of each emotional category, we considered the first constraint which is the selected seeds should be semantically close to the basic emotions. Thus, we construct the second emotional set employing the hierarchical synonyms which have high semantic similarities (LSA) with the basic emotions, i.e. "*LSA(synonyms)* $\geq$ *30*". Here we set 30 as the threshold. This is because we aim to keep a sufficient number of seeds in each category and at the same time preserve the semantic similarities between seeds and their corresponding emotion categories. As Figure 5 shows, this constraint improves the performance of sentiment prediction over all the dimensions.

The emotional vectors may also being biased toward the category that has the highest occurrences of seeds in the development corpus. Thus, the second constraint requires the categories to be balanced with respect to their seeds occurrences (frequencies) in the development corpus. We balanced the second set in such a way that the sum of the frequencies of all the seeds in each category remains the same among the categories (balance matrix). We also manually removed a few ambiguous or irrelevant seeds from each category. For example, the emotional category "*interest*" has mainly two sets of synonyms related to *interestingness* and *finance*; however we only consider the *interestingness* set as it reflects the target sentiment. As Figure 5 shows, using two constraints results in the best performance in any dimension. This experiment indicates that an accurate result can be obtained, if only relevant seeds that results in a balance matrix are selected.

Role of using synsets and antonyms of words: We show the important role of antonyms and synsets of words which we explained in Section "*Constructing Emotional Vectors*". For this purpose, we repeat the experiment for SO prediction by computing sentiment similarity of word pairs without using the synonyms and antonyms.

Table 6 shows the results. As it is clear, the highest performance can be achieved when antonyms and synonyms are used, while the lowest performance is obtained without using them. Table 6 also shows that using only synsets is more effective than using only antonyms. This could be because of the higher probability of the existence of synonyms than antonyms for a word.

Strategies	Precision	Recall	F1
w/o Ants and Syns	67.79	68.47	67.57
with Syns	71.47	72.25	71.43
with Ants	68.34	69.04	68.12
with Ants and Syns	**73.07**	**73.89**	**73.11**

Table 6. Role of synsets and antonyms

Related Works

Sentiment similarity has not received enough attention to date. Most previous works employed semantic similarity of word pairs to estimate their sentiment similarity (Kim and Hovy 2004; Turney and Littman 2003).

Turney and Littman (2003) proposed a method for automatically inferring the sentiment orientation (SO) of a word from its semantic similarity with a set of positive and negative seed words. To calculate the semantic similarity, they used PMI and LSA. Hassan and Radev (2010) presented a graph-based method for predicting SO of words. They constructed a lexical graph where nodes are words and edges connect two words with large semantic similarity based on a Wordnet (Fellbaum 1998) similarity measure. Then, they propagated the SO of a set of seed words through this graph.

In IQAPs, Marneffe, Manning, and Potts (2010) attempted to infer the *yes*/*no* answers using sentiment orientation (SO) of the adjectives. They calculated the probability of rating given the adjective (of question or answer) over the corpus to compute the SO of the adjective. If the computed SOs for the adjectives in an IQAP have different signs, the answer conveys *no*. Otherwise, if the absolute value of SO for the adjective in question is smaller than the absolute value of the adjective in answer, then the answer conveys *yes*, and otherwise *no*.

Mohtarami et al. (2011) employed semantic similarity measures (PMI and LSA) to infer *yes*/*no* from indirect answers in a given IQAP. They showed that measuring the association between the adjectives in question and answer can be a main factor to infer a clear response from an IQAP. Their experiments showed that LSA can achieve a better performance than PMI. It is notable that our experiments showed that we can achieve an accurate SO and better infer *yes*/*no* answer from a given IQAP by employing sentiment similarity measure instead of semantic similarity.

Conclusion

In this paper, we propose an effective method to compute sentiment similarity from a connection between semantic space and emotional space. We show the effectiveness of our method in two NLP tasks namely, indirect question-answer pair inference and sentiment orientation prediction. Our experiments show that sentiment similarity measure is an essential pre-requisite to obtain reasonable performances in the above tasks. We show that sentiment similarity significantly outperforms two popular semantic similarity measures, namely, PMI and LSA.

References

Fellbaum., C. 1998. WordNet: An Electronic Lexical Database. *Cambridge, MA: MIT Press*.

Hassan, A., and Radev, D. 2010. Identifying Text Polarity Using Random Walks. *ACL*, 395–403.

Izard, C. E. 1971. The face of emotion. *New York: Appleton-Century-Crofts*.

Kanayama, H., and Nasukawa, T. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. *ENMLP*, 355–363.

Kim, S. M., and Hovy, E. 2004. Determining the sentiment of opinions. *COLING*. 1367–1373.

Landauer, T. K.; Foltz, P. W.; and Laham, D. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes*. 259-284.

Maas, A. L.; Daly, R. E.; Pham, P. T.; Huang, D.; Ng, A. Y.; and Potts, C. 2011. Learning Word Vectors for Sentiment Analysis. *ACL*. 142-150.

Marneffe, M.; Manning, C.; and Potts, C. 2010. "Was it good? It was provocative." Learning the meaning of scalar adjectives. *ACL*. 167–176.

Mohtarami, M.; Amiri, H.; Lan, M.; Tan, C. L. 2011. Predicting the uncertainty of sentiment adjectives in indirect answers. *CIKM*. 2485-2488.

Neviarouskaya, A.; Prendinger, H.; and Ishizuka, M. 2009. SentiFul: Generating a Reliable Lexicon for Sentiment Analysis. *ACII*. 363-368.

Ortony, A., and Turner, T. J. 1990. What's Basic About Basic Emotions. *American Psychological Association*. 97(3), 315-331.

Turney, P., and Littman, M. 2003. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *ACM Transactions on Information Systems*, 21(4), 315–346.

Wilson, T.; Wiebe, J.; and Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. *HLT-EMNLP*. 347–354.