# Topic-Based Stance Mining for Social Media Texts

Wei-Fan Chen[1], Yann-Hui Lee[2], Lun-Wei Ku[1]

Institute of Information Science, Academia Sinica

[1]{viericwf, lwku}@iis.sinica.edu.tw; [2]andycyrus@gmail.com

**Abstract.** Recent techniques of opinion mining have succeeded in analyzing sentiment on the social media, but processing the skewed data or data with few labels about political or social issues remains tough. In this paper, we introduce a two-step approach that starts from only five seed words for detecting the stance of Facebook posts toward the anti-reconstruction of the nuclear power plant. First, InterestFinder, which detects interest words, is adopted to filter out irrelevant documents. Second, we employ machine learning methods including SVM and co-training, and also a compositional sentiment scoring tool CopeOpi to determine the stance of each relevant post. Experimental results show that when applying the proposed transition process, CopeOpi outperforms the other machine learning methods. The best precision scores of predicting three stance categories (i.e., *supportive*, *neutral* and *unsupportive*) are 94.62%, 88.86% and 10.47%, respectively, which concludes that the proposed approach can capture the sentiment of documents from lack-of-label, skewed data.

## 1 Introduction

Opinion mining has drawn much attention for both research communities and industries for its high relevance to knowledge mining applications. In the web environment, opinions are largely provided by users, especially in the form of texts. In the past, opinions have been classified into several categories, e.g., positive, neutral and negative. In some researches, scores which show the degree of valence for text segments were also determined [13].

The successful results of applying opinion mining in business encourages the attempt at automatically collecting opinions from the Internet for government or political parties to make decisions. However, it is more challenging either from the perspective of relevance judgment for posts or from telling the standpoint, compared to the review analysis of products, movies or travel experience, which is considered a success in business. As to the relevance, real controversial posts need to be identified from all posts related to the issue. The pure introductions, descriptions, and the facts about the issue should be excluded, which are usually separated from real comments in review forums but mixed up in media like Facebook, blogs and political forums where opinions towards public issues are found. As these issues are usually very specific, performance of finding relevant posts tends to report high recall but low precision. In addition, typical decision making involves telling supportive instances from

adfa, p. 1, 2011.

unsupportive ones, which is considered more easily prone to errors than telling positive instances from negative ones. A typical example could be a post which criticizes (negative content) the behavior of the unsupportive party and makes the standpoint of the whole post supportive.

At least three requirements need to be fulfilled for support determination. First, opinions that government is keen to know are usually towards new issues that trigger a heat debate. As these issues are new, data analytics usually have difficulties to find sufficient labeled data to create their models. Therefore, we need an approach which can easily start from a small set of labeled data. Second, data collected from the Internet are very often highly skewed. However, the minority opinions are of vital importance as they are challenging to be retrieved but valuable for decision making. Therefore, the proposed approach should be able to retrieve the minority. Third, as there will be many supportive and unsupportive documents, precision is very important for the proposed method when reporting evidence to the decision makers.

In this paper, we aim to extract supportive and unsupportive evidence from Facebook data of two characteristics: highly skewed but with little labeled training data. Hence, the public issue "Anti-reconstruction of Lungmen Nuclear Power Plant[1,]", is selected in order to demonstrate the challenge of this research problem. The fate of the power plant is to be decided in a future country-wide referendum, and whether having the referendum or not is a government decision. Successfully analyzing opinions on Internet can definitely help government to make a right decision. However, there is no labeled data for experiments. Moreover, from the testing data we generate, we know the unsupportive evidence is only about 1.25% of the whole, which is quite little. Here "supportive" denotes the "support" of "anti-construction" and having this literally unsupportive topic title implies that it might be unlikely to extract evidence by only determining the polarity of documents.

To meet the requirements of this research problem, we propose models which can start from a very small set of seed words, i.e., no more than 5 words. Working with these few seed words, we illustrate how to find supportive and unsupportive evidence based on an existing sentiment polarity determination tool or the SVM models. Then their performances are compared with each other. Results show that seed words working with the sentiment analysis tool together with a transition process from polarity to standpoint significantly outperforms the commonly adopted SVM models, i.e., pure SVM model or SVM co-training models, when having little and skewed training data.

## 2 Related Work

As we mentioned, researchers have been applying sentiment analysis techniques on political, social or public issues. For example, some researches tried to predict the results of American president election, analyze the aspects of candidates [12], or more specifically, show the influence of the speech of Obama on the election [5]. Wang [16] adopted the Naïve Bayes model to analyze people's attitude towards DOMA

[1] http://en.wikipedia.org/wiki/Lungmen_Nuclear_Power_Plant

(Defense of Marriage Act). However, most of them were working on existing balanced data or a certain amount of balanced data generated for experiments, which is not cost effective when many different issues are to be analyzed.

Both content based and knowledge based approaches have been tested for sentiment analysis, and we propose one approach for each of them. The knowledge based approaches usually have issues acquiring necessary resources or being applied to data of different languages. In this paper, we adopted a Chinese sentiment analysis tool CopeOpi [7], which provides sentiment scores for words, sentences and documents. As to the content based approach, we adopt the commonly used SVM model to solve the proposed research problem. SVM model has been adopted from the very beginning of the history of tackling sentiment analysis problems [10]. It has served as a good baseline. However, requiring labeled data for training makes it difficult to be applied on a large amount of various unlabeled Internet data.

This paper focuses on unsupervised or semi-unsupervised methods as it is usually difficult to find stance labels for web posts. Bollen [2] used Profile of Mood States (POMS), which detected each tweet's emotion on six dimension without machine learning technique. The results on the timeline matched the global social, political and economic events happened in this time period. Hu [4] used post-level and word-level models to detect the polarity of a tweet. This unsupervised method can be considered as an alternative of LSA or word vector. On the other hand, Blum and Mitchell [1] proposed co-train algorithm to utilize labeled and unlabeled data together when training models, and it is also widely used for sentiment analysis [8], [15]. It seems that the co-training model can decrease the pain of having too little labeled training data. Recently, Gao [3] used co-training to build models to construct bilingual sentiment lexicon. Their co-training process is modified and utilized in this paper.

## 3 Materials

We collected a total of 41,902 Facebook documents from related fans groups in one year period of time for experiments. Each document contains the post time, title, message, number of likes, number of shares and number of comments. From them, we randomly selected 4,000 documents and labeled them as supportive, neutral or unsupportive for testing, while the other 37,902 unlabeled documents were left for training. Documents are classified into 5 types as shown in Table 1 when they are collected, where documents of the *status* type are from the status or shared post of Facebook users, and question is a function of Facebook fans groups. Testing data were labeled by four annotators and results are shown in Table 2. From Table 2, we can also see that the unsupportive documents are less than one tenth of supportive documents.

| Type | Photo | Link | Video | Status | Question | Total |
|---|---|---|---|---|---|---|
| Testing | 1,264 | 1,806 | 336 | 594 | 0 | 4,000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Training | 12,286 | 14,024 | 3,755 | 7,836 | 1 | 37,902 |

**Table 1.** Five types of Facebook documents

| Sentiment | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Number | 652 | 3,297 | 51 |

**Table 2.** Testing dataset

<table>
<tr><th>Sentiment</th><th>Post time</th><th>Type</th><th>Title</th><th>Likes</th><th>Shares</th><th>Comments</th></tr>
<tr><td>Supportive</td><td>2014/6/25 08:52:12 PM</td><td>Photo</td><td>台灣能源短缺，…要面臨電價成本提高及…的問題。<br>(There is a shortage of energy in Taiwan… will face the problems of higher electricity rate and some other problems.)</td><td>146</td><td>0</td><td>0</td></tr>
<tr><td colspan="7">廢話+笨蛋！問題是：經長無腦，不知「能源金字塔」，不知再生能源不需燃料費、成本低、而且持續降低。<br>(Rubbish + Idiot ! The problem is that they know nothing about Energy Pyramid and don't know renewable energy needs no fuel cost and the cost is decreasing continuously. )</td></tr>
<tr><td>Un-supportive</td><td>2014/5/17 10:33:30 AM</td><td>Status</td><td>No content</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td colspan="7">請問你們能提出比核能發電廠更適合當基載的發電廠嗎？<br>(Could you propose other base load power plant which is better than the nuclear power plant?)</td></tr>
<tr><td>Neutral</td><td>2014/6/23 12:07:44 AM</td><td>Video</td><td>No content</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td colspan="7">搶救新營最美好的 溼地<br>(Saving the most beautiful wetlands in Hsinying.)</td></tr>
</table>

**Table 3.** Examplar documents

We select some example testing documents shown in Table 3 to give a brief view of the material. The *title* field could be a website address of in a *link* type post or the context of the shared (*status* type) post. The *message* written by the author is then listed, which may show a complete different sentiment to the context in the title (e.g., the supportive one, the author argued that the *title*'s author, who supported nuclear power, is an idiot). For the neutral ones, it could be a piece of news that reported the event related to the nuclear power without showing the stance, a political event that prompted to these fans groups to ask for supporting, or something just unrelated to the nuclear power (the listed neutral one). As a result, it is hard to determine the stance only by surface information while documents of the same stance may contain similar context and sentiment. Moreover, even though these documents were collected from related fans groups, many unrelated posts from social movement groups asking for support still bring a lot of noise.

# 4 Method

From all documents, keyterms are extracted first and utilized to find relevant, or even supportive documents. Then the SVM models or CopeOpi will report supportive, neutral, and unsupportive documents from the relevant ones. All related modules are introduced in this section.

## 4.1 Keyterm Extraction

Keyterms and their ranks are given by InterestFinder [6], a system which proposes terms to indicate the focus of interest by exploiting words' semantic features (e.g., content sources and parts of speech). The approach adopted by InterestFinder involves estimating topical interest preferences and determining the informativity of articles. InterestFinder estimates the topical interest preferences by TFIDF, a traditional yet powerful measure shown in formula (1). Then semantic aware PageRank in formula (2) is used on candidates to find keyterms. In this paper, we hope to find keyterms related to certain seed words. Therefore, we only keep terms which are within the window of 6 words to the seed words in the article. Then we let InterestFinder propose keyterms of each article for us.

$$tfidf\left(art,w\right)= freq\left(art,w\right)/\,artFreq(w) \tag{1}$$

$$\mathbf{IN'}[1,j]=\lambda\times\begin{pmatrix}\alpha\times\sum_{i\in v}\mathbf{IN}[1,i]\times\mathbf{EW}[i,j]+\\ \left(1-\alpha\right)\times\sum_{k\in v}\mathbf{IN}[1,k]\times\mathbf{EW}[k,j]\end{pmatrix}+\left(1-\lambda\right)\times\mathbf{IP}[1,j] \tag{2}$$

## 4.2 Co-Training Process

As mentioned, our dataset is highly skewed and lacking of labeled instance dataset. In addition, as experimental documents are collected from Facebook, each document contains two types of information: content and metadata. Content information contains post title and the post itself, while the metadata information contains the number of comments, likes and shares. With all these properties of the dataset, we propose a co-training approach which can build two classifiers, the content classifier and the metadata classifier, and train each other by starting from a small set of labeled data. The co-training process then iterates to finish the labeling of the whole dataset.

### 4.2.1 Detailed Steps

The main idea of the co-training is to use the classifier trained by labeled data of one aspect to predict unlabeled data of another aspect. Instances predicted with confidence are then added into the labeled data and this updated labeled data are used to train the new classifier. The whole process is described as follows. We apply several context features to be the first independent feature set, including bag of words and combinations of word vector, detailed description of the features will be introduced in next section. Numbers of likes, shares and comments are included in the second independent feature set.

*All*: The whole dataset
Feature *F1*: Context Features
Feature *F2*: Numbers of likes, shares and comments (LSC)

**Step 1**: Find an initial labeled dataset, which contains supportive and unsupportive documents. It is usually a small set. Two labeled datasets, *L1* and *L2*, are both set as the initial labeled dataset. Two unlabeled dataset, *U1* and *U2* are both set as the complement of the initial labeled dataset, i.e., *U1=All-L1*, *U2=All-L2*.

**Step 2**: Build the content classifier *C1* using features *F1* extracted from the content information of *L1*. Here the first independent feature set is utilized. Use *C1* to label *U2*.

**Step 3**: Build the metadata classifier *C2* using features *F2* extracted from the metadata information of *L2*. Here, the second independent feature set is utilized. Use *C2* to label *U1*.

**Step 4**: Move highly confident labeled instances from *U2* to *L2* and set *U2* to All-L2; Move highly confident labeled instances from *U1* to *L1* and set *U1* to *All-L1*.

**Step 5**: Iterate from Step 2 to 4 until no confident labeled instances can be added into *L1* or *L2*, or the number of iteration exceeds a threshold *k*. Here *k* is set to 100.

### 4.2.2 The context features

For the context features used in the co-training process, we considered the following three types:

**Bag of Words.** This feature consider the words using in a post, which is widely used in the information retrieval domain and they are often the baseline in related work [14], [17]. Post representation by BOW is:

$$PR_{BOW} = \left[x_1, x_2, ..x_L\right],\ PR_{BOW} \in R^L \tag{3}$$

, where $\mathrm{x}_i \in \{0,1\}$ and *L* is the size of vocabulary in the corpus.

**BOW with word vector.** As the dimension of BOW feature is usually very large, and the feature is very sparse, we combine the idea of using BOW and the word vector. The word vector is able to represent every word as a feature vector with user-defined, reasonable number of feature dimensions. We use Glove [11] to generate word vectors for all words in the vocabulary, noted as W where $\mathrm{W} \in \mathrm{R}^{d \times L}$, *d* is the feature dimension and *L* is the size of vocabulary. Then we calculate the average of feature vectors of post words. Finally, the post representation by BOW with word vector then is defined as:

$$PR_{WV} = \frac{W \cdot PR_{BOW}}{|N_{WV}|},\ PR_{WV} \in R^{d} \qquad (4)$$

, where $N_{WV}$ is the set of words of one post. The idea of averaging the word vectors was introduced in Maas's work [9].

**Dependency tree with word vector.** Some words (like verbs) carry more information than other words (such as *a/an, the)* in one sentence. To capture this, we extract words that strongly related to the "root word" reported by the Stanford dependency parser. Then we extract all words that directly depended on the root word in the dependency tree and the root word to form a dependency vector, $dep_s$:

$$dep_s = \left[ x_{s,1}, x_{s,2}, .. x_{s,L} \right] \qquad (5)$$

, where $\mathrm{x}_{s,i} \in \{0,1\}$ and $s$ is the sentence index in a post. For example, in the sentence "*My dog also likes eating sausage*.", the root word is *likes*, and other extracted words are *dog*, *also*, *likes* and *eating*. Then the vectors of these words in the *s-th* sentence are averaged to get a sentence representation, $SR_{dep,s}$, defined as:

$$SR_{dep,s} = \frac{W \cdot dep_s}{|N_{dep_s}|},\ SR_{dep,s} \in R^{d} \qquad (6)$$

, where $N_{dep_s}$ is number of dependency relations of the *s-th* $dep_s$. Finally, the post representation by dependency tree with word vector, $PR_{dep}$, is the average of $SR_{bin,s}$ of all sentences in a post as the equation (7), where $S$ is all sentences in a post and $|S|$ is the size of it.

$$PR_{dep} = \frac{\sum_{s \in S} SR_{dep,s}}{|S|} \qquad (7)$$

### 4.3 Using CopeOpi

CopeOpi [7] is selected as our sentiment analysis tool. CopeOpi can determine the sentiment score of Chinese words, sentences and documents without training. It not only includes dictionaries and statistic information, but also considers shallow parsing features such as negations to enhance the performance. The results it generates indicate sentiment polarities like many other similar tools. However, the polarities cannot directly be mapped to the standpoint. Therefore, we utilize CopeOpi together with the seed words to calculate the SUP, NEU, and UN_SUP to represent the sentiment comments on these seed words. First we categorize seed words into supportive, neutral and unsupportive classes. If we find any supportive seed word in a sentence, we calculate the score of this sentence. Then scores (could be positive, zero, or negative) of all sentences containing supportive seed words are added up to SUP, and neutral seed words and unsupportive seed words to NEU and UN_SUP, respectively. Note that it is important that the score of the seed word is not included to exclude its sentiment. As a result, having positive sentiment to the SUP and NEU seed words means supportiveness while having positive sentiment to the UN_SUP seed words means unsupportiveness. The final standpoint STD_PT of each document is calculated as formula (8). Since the polarity of UN_SUP seed words differ from SUP and NEU seed words, we reverse it by multiplying a negative sign to UN_SUP to contribute to the determination of the stance.

$$\text{STD\_PT} = \text{SUP-UN\_SUP+NEU} \qquad (8)$$

Generally, if STD_PT is greater than 1, the document is considered as supportive; if it is less than -1, the document is considered as unsupportive; otherwise the document is neutral. However, if the topic itself already bears an unsupportive concept (like in this paper, anti-construction), a STD_PT value greater than 1 identifies unsupportive documents, and vice versa.

## 5 Experiments and Results

We first test the performance of the relevance judgment. We generate relevant document sets using two seed word sets in four settings. The seed word set for setting **(Unsup)$_{all}$** only contains the word "擁核" (embracing nuclear power plant). We believe documents containing keyterms generated from it would not support the reconstruction of the nuclear power plant as the word "embracing" here is ironic pragmatically. The seed word set for setting **(Sup+Unsup)$_{all}$** includes five seed words "擁核" (embracing nuclear power plant), "廢核" (abandon nuclear power plant), "反核" (anti-nuclear), "核能" (nuclear power), "核電" (nuclear power electricity). Keyterms generated from this setting should contain most documents related to the nuclear power plant, including supportive, controversial, unsupportive and fact-describing ones. Setting **(Unsup)$_{contained}$** finds those documents found by setting **(Unsup)$_{all}$** but containing its one seed word, whereas setting **(Sup+Unsup)$_{contained}$** finds documents found by setting **(Sup+Unsup)$_{all}$** but containing one of its five seed words. According to the experimental result, the documents of the score greater than 15 found by setting **(Sup+Unsup)$_{all}$** are treated as relevant in the training and testing data, and then the support and unsupportive document detection is performed on them.

Next the performance of support detection is evaluated. Results of random selection are here reported as the baseline: the f-score for Unsupportive, Neutral and Supportive is 2.40%, 49.58%, and 22.23%, respectively. The results of adopting pure SVM as features are also reported as another baseline: the f-score for Unsupportive, Neutral and Supportive is 1.36%, 84.55%, and 28.76%, respectively. Here SUP, NEU, and UN_SUP together with BOW are utilized as features. Only 20 supportive and 20 unsupportive documents are used for training as generating the SVM model needs labeled data but we have only 51 documents labeled as unsupportive. Note that as 40 labeled documents are selected for training, only 3,960 documents are tested in this experiment. The low performance of the pure SVM experiment could be due to the small and highly skewed training set. Compared to the random selection results, only the performance of the supportive class is improved. We try to train by more instances to decrease the effect of a small training set. However, the problem of lacking labeled data remains. Therefore, the co-train process is adopted to utilize both labeled and unlabeled documents in the training phase.

Before the co-train process starts, setting **(Sup+Unsup)$_{\mathbf{all}}$** is applied on the training data to first find the relevant document set, i.e., labeled and unlabeled data for co-training. Documents found by setting **(Unsup)$_{\mathbf{contained}}$** are the initial labeled supportive data and by setting **(Sup+Unsup)$_{\mathbf{contained}}$** are all initial labeled data for co-training. Table 4 shows the results of using these supportive and unsupportive documents as the two initial sets for SVM co-training with the BOW feature. Compared to the pure SVM, the performance improves a lot, but still it is not satisfactory. Table 5 and 6 show the results of using word vector features. However, the results of identifying unsupportive posts in Table 5 are all zero, which means the unsupportive ones are all classified as supportive. The results in Table 6 are slightly better than Table 4, which suggests when compositing word vector as features, considering dependency relations is better than adding up BOW.

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 32.78% | 86.74% | 0.71% |
| Recall | 30.37% | 81.95% | 3.92% |
| F-score | 31.53% | 84.28% | 1.20% |

**Table 4.** Co-training performance by BOW post representation

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 31.11% | 86.74% | 0.00% |
| Recall | 42.18% | 81.95% | 0.00% |
| F-score | 35.81% | 84.28% | 0.00% |

**Table 5.** Co-training performance by BOW with word vector post representation

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 30.32% | 86.74% | 1.14% |
| Recall | 32.98% | 81.95% | 3.92% |
| F-score | 31.59% | 84.28% | 1.76% |

**Table 6.** Co-training performance by dependency tree with word vector post representation

Next, we try to keep the same quantities of supportive and unsupportive labeled data during the co-training process to decrease the influence of the unbalance of data. In each iteration in step 4 of co-training, only a maximum number 10, an equal number of highly confidently labeled instances are moved to L1 or L2. The results are shown in Table 7 to Table 9, and the performance of finding the unsupportive evidence drops a lot compared to those reported in Table 4 to Table 6, respectively.

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 45.28% | 86.74% | 1.63% |
| Recall | 14.72% | 81.95% | 21.57% |
| F-score | 22.22% | 84.28% | 3.04% |

**Table 7.** Co-training performance by BOW post representation

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 31.11% | 86.74% | 0.00% |
| Recall | 42.18% | 81.95% | 0.00% |
| F-score | 35.81% | 84.28% | 0.00% |

**Table 8.** Co-training performance by BOW with word vector post representation

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 26.16% | 86.74% | 0.22% |
| Recall | 17.33% | 81.95% | 1.96% |
| F-score | 20.85% | 84.28% | 0.40% |

**Table 9.** Co-training performance by dependency tree with word vector post representation

So far we have tested the performance of using SVM and SVM co-training in different settings but achieved limited improvements. Experimental results show that learning from small and skewed data is challenging, and the precision is too low to fulfill our requirements. Therefore, we try the other proposed method involving the sentiment analysis tool. We keep the assumption that documents containing the keyterms generated from five seed words, setting **(Sup+Unsup)$_{\mathbf{all}}$**, are relevant and adopt CopeOpi to calculate the sentiment scores, which determine the sentiment of each document. If it is greater than the threshold, the document is positive; if it is less than the threshold multiplied by -1, the document is negative; otherwise the document is neutral. Negative documents are labeled as supportive (support anti-reconstruction), while positive documents are labeled as unsupportive. We then set the threshold to 1, which achieved the highest performance than other value and results are shown in Table 10.

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 18.68% | 86.32% | 1.60% |
| Recall | 45.71% | 51.50% | 13.73% |
| F-score | 26.56% | 64.51% | 2.86% |

**Table 10.** Using CopeOpi directly for supportive/unsupportive determination

The performance in Table 10 is worse than the results of the best SVM co-training model so far shown in Table 6. After analysis, we find that an additional transition process is necessary to tell the standpoint from the reported polarity. Again, the relevance judgment depends on keyterms generated from seed words. However in the transition process, seed words are also viewed as aspects to be commented on, i.e., people, organizations, events, etc. In addition, we categorize seed words into three aspect categories to calculate the SUP, NEU, and UN_SUP values as shown in Table 11. Results of adding this transition are shown in Table 12. The performance is boosted up and better than the SVM and SVM co-training models, especially the precision. However, the minority, the unsupportive class, is still difficult to be identified.

| Aspect | Value | Seed Word |
|---|---|---|
| Positive | SUP | embracing nuclear power plant |
| Neutral | NEU | nuclear power, nuclear power electricity |
| Negative | UN_SUP | anti-nuclear, abandon nuclear power plant |

**Table 11.** Seed words for polarity to standpoint transition

| Metric | Supportive | Neutral | Unsupportive |
|---|---|---|---|
| Precision | 94.62% | 88.86% | 10.47% |
| Recall | 18.87% | 99.67% | 35.29% |
| F-score | 31.46% | 93.95% | 16.14% |

**Table 12.** Performance of CopeOpi followed by a transition process for support determination

## 6 Conclusion

Finding supportive and unsupportive evidence usually encounters the issues of lacking labeled data and data skewness. In this paper, we have proposed two methods which can start from very few predefined seed words to find relevant supportive and unsupportive evidence. Results show that as the support determination module in the proposed methods, adopting the sentiment analysis tool together with a polarity to standpoint transition significantly outperforms using SVM or SVM co-training models.

Several aspects of our approach can be further improved. Sharing posts in Facebook may bring us many identical or very similar posts. Removing these redundant posts may give more reliable evaluation results. For a controversial topic that people pay attention to, data grow quickly in time. Some learning mechanisms can be injected into the CopeOpi-like sentiment analysis tool to enable the adaptation and improve the performance.

## 7 References

1. Blum, A., & Mitchell, T. (1998, July). Combining Labeled and Unlabeled Data with Co-training. In Proceedings of the eleventh annual conference on Computational learning theory, 92-100. ACM.
2. Bollen, J., Mao, H., & Pepe, A. 2011. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. In Proc. ICWSM 2011.
3. Gao, D., Wei, F., Li, W., Liu, X., & Zhou, M. 2013. Co-Training Based Bilingual Sentiment Lexicon Learning. In AAAI.

4. Hu, X.; Tang, J.; Gao, H.; and Liu, H. 2013. Unsupervised sentiment analysis with emotional signals. In Proceedings of the 22$^{nd}$ international conference on World Wide Web, WWW'13. ACM.
5. Hu, Y., Wang, F., & Kambhampati, S. (2013, August). Listening to the Crowd: Automated Analysis of Events via Aggregated Twitter Sentiment. In Proceedings of the Twenty-Third international joint conference on Artificial Intelligence (pp. 2640-2646). AAAI Press.
6. Huang, C. and Ku, L.-W. 2013. Interest Analysis using Semantic PageRank and Social Interaction Content. In Proceedings of the IEEE International Conference on Data Mining, SENTIRE Work-shop.
7. Ku, Lun-Wei, Ho, Xiu-Wei and Chen, Hsin-Hsi. 2009. Opinion Mining and Relationship Discovery Using CopeOpi Opinion Analysis System. Journal of American Society for Information Science and Technology. 60(7), 1486-1503.
8. Li, S., Wang, Z., Zhou, G., & Lee, S. Y. M. 2011. Semi-supervised Learning for Imbalanced Sentiment Classification. In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI 2011), Vol. 22, No. 3, p. 1826
9. Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 142-150). Association for Computational Linguistics.
10. Pang, Bo, Lee, Lillian and Vaithyanathan, Shivakumar. 2002. Thumbs Up? Sentiment Classification Using Machine Learning Techniques. In Proceedings of EMNLP. 79-86.
11. Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014), 12.
12. Ringsquandl, M. and Petkovic, D. 2013. Analyzing Political Sentiment on Twitter. In AAAI Spring Symposium: Analyzing Micro-text.
13. Seki, Yohei, Ku, Lun-Wei, Sun, Le, Chen, Hsin-Hsi and Kando, Noriko (2010). Overview of Multilingual Opinion Analysis Task at NTCIR-8: A Step Toward Cross Lingual Opinion Analysis. In Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access, June 15-18, 2010, Tokyo, Japan. 209-220.
14. Socher, Richard, Pennington, Jeffrey, Huang, Eric H., Ng, Andrew Y., Manning, Christopher D. 2011. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2011), 151-161.
15. Wang, X. 2009. Co-training for Cross-lingual Sentiment Classification. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Volume 1, 235-243. Association for Computational Linguistics.
16. Wang, Y., Clark, T., Agichtein, E., & Staton, J. 2014. Towards Tracking Political Sentiment through Microblog Data. ACL 2014, 88.
17. Yang, Y. H., & Liu, J. Y. 2013. Quantitative Study of Music Listening Behavior in a Social and Affective Context. IEEE Trans. Multimedia, 15(6).