Opinion Analysis Task of Simplified Chinese Monolingual at NTCIR-8

Yuan Kuang Bo Zhang Wenbin Pan Yujie Zhang Bin Zhang
Rongjun Li Yu Mao Yanquan Zhou
Research Center of Intelligence Science and Technology of Beijing University of Posts and Telecommunications, Beijing, 100876, China
zhouyanquan@bupt.cn

Abstract

In this paper, we present our work for Simplified Chinese Monolingual opinion analysis task at NTCIR-8 by BUPT. We participated in four of all tasks except opinion target detection and answerness judgment, and submitted two runs for each task. For opinion sentence detection, we propose some features both semantic-level and grammar-level, and also summarizes some syntactic structure templates to achieve a more satisfactory classification results based on TSVM. For opinion holder detection, we firstly use CRF including six corresponding features to detect, then we propose two syntactic rules based on opinionated trigger words from syntax trees taken as additional features for the CRF to train our model. By introducing Statistical Language Models with expansion of topic words we train a relevance judgment model. To judge polarity, we compute the value of the text by our algorithm with a large-scale emotional dictionary, and set a threshold, to classify the sentiment polarity of sentence in each text.

Key words:

NTCIR-8, Simplified Chinese Monolingual, opinion, relevance, opinion holder, polarity

1. Introduction

Nowadays, with the rapid development of the information technology, a large number of users express their views on the products in the forum, blog and other platforms. On one hand, opinion analysis can provide us an effective way to help understand and study sentiment, and on the other, it can be used on Human-Computer interaction, question-answering system, watchdog to public opinion, personalized retrieval, etc. So opinion analysis based on natural language technology is of great value. For the Third Multilingual Opinion Analysis Task (MOAT) NTCIR-8 workshop explore tasks including opinion judgment (required), relevance judgment, answerness judgment, opinion holder extraction, opinion target detection and polarity judgment. The dataset about Simplified Chinese Monolingual coves many topics. For more information about the tasks and dataset can be found in the overview paper [1].

At NTCIR-8, we participated in the four subtasks for simplified Chinese monolingual, including opinion and relevance judgment, opinion holder extraction and polarity judgment. For identifying opinion sentences, our training data comes from NTCIR-7 simplified Chinese data and COAE (Chinese Opinion Analysis Evaluation) that released forty thousand documents consists of subjective and objective sentences, then after manually labeling, we extracted 2606 opinion and 3496 non-opinion sentences as our training set. We take the opinion judgment as binary classification by using TSVM based on features both semantic-level and grammar-level and also propose some effective syntactic structure templates as opinion rules. For opinion holder identification, we view it as sequential labeling task via CRF, given a sentence of tokens, we classify the tokens into 3 categories, B(beginning of an opinion holder), I(inside an opinion holder) and O(outside an opinion holder). Then a chunk BI or B is thought as an opinion holder. So identifying opinion holder via CRF is to label B, I, O for each token in a sentence. We firstly explore Conditional Random Field(CRF) and define corresponding feature templates based on six features including contextual, opinionated trigger words, POS tags, named entity, dependency that we adjust to be better helpful for containing contextual dependency information and sentence structure feature that we propose. Then we propose two syntactic rules with opinionated trigger words from the syntax trees and we take the rules as additional features for the CRF combined with the mentioned six features to extract opinion holder. For relevance judgment, we regarded it as retrieving sentences that are relevant to the given topic for each document by introducing Statistical Language Models in the mean while expanding topic words and we score each topic based on PMI. For polarity judgment, we set up a dynamic adjustment model for the sentiment value of the words based on sentiment dictionary.

The rest of this paper is organized as follows: Section 2 gives an overview of the related works. Section 3 describes features we used for TSVM to judge opinionated and propose some syntactic structure templates. In Section 4, we show our proposed syntactic rules combined with CRF to identify opinion holder. And in Section 5 and Section 6 we give a brief introduction for relevance and polarity judgment. The evaluation results will be listed in Section 7. And finally Section 8 concludes our paper with future works.

2. Related Work

For extracting opinion sentence as a relatively new field is still at the exploratory stage and the current classification algorithm is still relatively simple. In contrast, research of English carried out earlier. Wiebe et al. [2] proposed the following classification features: pronouns, adjectives, cardinal numeral, modal verbs beside will, adverbs beside not, they also believe that the location of sentences in paragraphs also implies important information. Hatzivassiloglou, Vasileios, and Wiebe [3] do further analysis on characteristics of adjectives. They consider the influence of dynamic adjectives, adjectives with semantic information and graded adjectives to the classification. Riloff and Wiebe [4] proposed the bootstrapping method based on a large number of unlabeled corpus to identify opinion sentences in 2003. By 2005, Wiebe and Riloff joined the extraction of the objective models based on bootstrapping method and the F-measure had further improve.

For identifying opinion holder, Yejin Choi and Claire Cardie [6] adopted a hybrid approach that combined Conditional Random Field (Lafferty et al.,2001) and a variation of AutoSlog(Riloff,1996a). In Soo-Min Kim and Eduard Hovy's work [7][8], they firstly extracted the syntactic paths between opinion holder candidates and opinion_expressions then they treated paths as features added to Maximum Entropy to train their model to select the most probable opinion holder; their second work used FrameNet and Semantic Role Labeling as an intermediate step to identify opinion holder. In Yohei Seki's work [9], his approach of opinion holder extraction was based on the discrimination between author and authority viewpoints in opinionated sentences, while Steven Bethard [10] used an extension of semantic parsing techniques coupled with additional lexical and syntactic features, Ruifeng Xu [11] thought the nearest named entities before an opinion word as an opinion_holder. Ruihong Huang [12] and Kang Liu [13] both used CRF model to identify opinion holder.

For relevance judgment, Evans [14] used standard vector space model combined with TFIDF method to calculate weight obtained good results. Li et al [15] determined the relevance by calculating the eigenvector of theme feature with inner product of the sentence, while Youngho Kim [16] used language model.

For polarity judgment, Bo Pang and Lillian Lee [17] used Naive Bayes, ME and SVM to classify sentences. Peter D. Turney [18] judged polarity based on unsupervised method. Soo-Min Kim and Eduard Hovy[19] set up a sentimental classifier through finding synonym based on WordNet.

3. Opinion sentence extraction

In my opinion, as long as the sentence contains a tendentious opinion, whether it is declared by the first person or quoted by the third person, is judged to be an opinion sentence. In this task, we firstly use the syntactic structure templates to tag sentences as opinion sentence and then input the remaining data to a TSVM classifier. We think this task as binary classification using TSVM classifier to label the samples, each sentence is expressed in the form of eigenvectors, and weight of each dimensional characteristic is the number of times this characteristic appears in this sentence. In addition, the opinion training set is less than the non-opinion training set which resulted in imbalance of training data as section 1 mentioned, so we train a TSVM classifier based on strict result of NTCIR-7 to obtain more opinion sentences to balance the training set at first.

3.1 Feature Selection

Reference the opinion and non-opinion sentences, we think that both of them imply their own unique characteristics on the semantic level and the grammatical structure.

1) Sentiment

Sentiment means adjectives with some kind of emotional tendencies. Usually, when someone says a few words with emotional tendencies, we will consider them as being of opinions.

2) Indicative verb

Indicative verbs are usually considered as signs of expressing words of views, such as "express", "believe," "think," "forecast," "In my opinion," and so on. We finally summed up 64 verbs, which are used as indicative verb vocabularies for experiment.

3) Indicative adverb

Indicative adverbs, of which total number is 365, are mainly extracted from the experimental corpus. They include two kinds, one for adverbs of degree, such as "very(非常)", "most(最)", "somewhat(稍微)", "probably(大概)", "extremely(极其)", "especially(格外)", etc.; the other for adverbs pointing with mood or attitude, such as "anyway(反正)", "rather(未免)", "precisely(恰恰)", " Admittedly(固然) " and so on.

4) Interjection & Punctuation

Interjections, such as "Ah, La, Oh, Gosh, Well (“啊/哪/啦/呀/吧/呢”)", Punctuation, such as "?","!", they are used when people usually express their own particular feelings, like praise, complaints, or doubt.

5) N-POS

N-POS means a combination of sequence of N continuous parts of speech. We think that a combination of a single part of speech or several continuous parts of speech, implies information of opinion or non-opinion. For example, there are more adjectives and adverbs in opinion sentences than in non-point sentences. Specifically, such as adjective + "de(的)", adverb + adjective and "de(的)" + noun, are of more obvious opinions. "

In this experiment, we mainly take “N=1, N=2, N=3” these three cases into consideration. Extracting 1-POS, 2-POS, 3-POS respectively from the training text, then sequencing them by calculating the CHI, to select some forward models of combination of parts of speech as the characteristic. Formula of CHI is as follows: [5]

$$CHI(p,c_j)=\frac{N\times(AD-CB)^2}{(A+C)\times(B+D)\times(A+B)\times(C+D)} \qquad (1)$$

The Formula calculates the CHI of the N-POS pattern p and the category c_j. In this task, there are only two categories, opinion sentence and non-opinion sentence. A is the number of times p and c_j, B is the number of times p and $\overline{c_j}$, C is the number of times c_j without p, D is the number of times $\overline{c_j}$ without p.

6) N-Word

N-Word means a combination of sequence of N consecutive words. We think that a combination of a single word or several consecutive words also implies information of opinions, such as, "I" + "think", "someone" + "indicates", ","+" however", "say" + “colon” + “quotation”, are usually signs of opinion sentences. We observe a unigram, bigram, tri-gram and then still sequence them by calculating the CHI values to select some forward models of combination of consecutive words as the characteristic.

3.2 Syntactic Structure Templates

By studying the large-scale data sets and the regular opinion expression method, we find some kind of way often used to express views, opinions, forecasts or comments. This kind of way is summarized as following syntactic structure templates:

$$SS_1=\langle Subject\rangle NP+\langle Predicate\rangle Indicative\ Verb+...ADJP...$$

//This template mains the subject of sentence is NP and the predicate of sentence is one of indicative verb mentioned in section 3.1 and other child nodes which have the same father node with the predicate verb contain adjective phrases.

$SS_2 = \langle Subject \rangle NP + \langle Predicate \rangle Indicative\ Verb + ...ADVP...$

//This template is similar to SS_1 but other child nodes which have the same father node with the predicate verb should contain adverb phrases.

$SS_3 = \langle Subject \rangle NP + \langle Predicate \rangle Indicative\ Verb + ...VC是...$

//The difference between SS_3 and SS_1 is that other child nodes which have the same father node with the predicate verb contain 'VC(是)' instead of adjective phrases.

4. Identifying opinion holder

4.1 Identifying opinion holder via Conditional Random Field

Conditional Random Field model is defined as the joint probability distribution of a particular label sentence Y given the observation sentence X, not the distribution of the following state given the current state. CRF-based machine learning model can be arbitrarily added an effective feature vector to make full use of contextual information, which can overcome the HMM independence assumptions and the inherent label bias of MEMM models.

In our work, we adopted Yet Another CRF toolkit implemented by C++, whose nonlinear optimization parameters are trained by Limited-memory BFGS(LBFGS) algorithm.

4.2 Feature Selection

With the definition of opinion holder that it must be an entity and it must be preceded by an opinionated trigger word, we firstly explore the following features:

1) Contextual features

For the opinion holder maybe a chunk BI or B, we must possibly contain contextual information of opinion holder.

2) Opinionated trigger words features

Since almost all opinion holders are triggered by opinionated trigger words as is same to our definition, we extracted all words that are labeled as <OPINION_OPR> from the corpus of NTCIR-6. We treated opinionated trigger words features as binary to indicate whether or not each token belongs to.

3) Part-of-speech features

Since the opinion holders are noun phrases, we use POS tags for each token as features including noun, verb, adverb, adjective, pronoun, punctuation, etc.

4) Named entity features

The opinion holder may be a named entity, so we treat named entities as candidates of opinion holder. We use the module of HIT_IR Lab language information retrieval technology platform to gain named entity.

5) Dependency features

For almost all opinion holders are triggered by opinionated words, between them it must be the dependency relationship. We use the module of HIT_IR Lab language information retrieval technology platform to analyze dependency relationship.

6) Sentence structure features

The opinion holders are always depended on opinionated trigger words, so it's necessary to analyze the structure of each sentence. We obtain rough structure of each sub-sentence, which is visualized by a triplet <before, center, after>.

4.3 Identifying opinion holder via Syntactic Rules

On the basis of our definition about opinion holder and intuitively, there is a structural relationship between an opinion holder and an opinionated trigger word. So after analyzing parse trees we propose two syntactic rules given the opinionated trigger words to extract opinion holder directly, using the Stanford Parser, Figure 1 shows an example of a parsing sentence.

Before setting the rules, we must locate an opinionated trigger word(*) node signified by (VP(VV* in the parsing sentence. The two syntactic rules are the followings:

1) If there is a node (VP whose child is the node (VP(VV*, then we treat the first node (NP as the opinion holder whose height is equal to that of the node (VP or we treat the first node (NP that is contained in the node whose height is equivalent to that of the node (VP.
2) If there is not a node (VP whose child is the node (VP(VV*, then we treat the first node (NP as the opinion holder whose height is equal to that of

the node (VP(VV* or we treat the first node (NP that is contained in the node whose height is equivalent to that of the node (VP(VV*.

As the syntactic rules indicate, in Figure1, there is a node (VP whose child node is (VP(VV 认为(think), then we regard the node (NP whose height is equal to that of the node (VP that contains the node (NN 科学家(the scientists) as the opinion holder.

The syntactic rules take two features (SRFs) to the CRF, which are showed in the following:

SRF1: Whether or not each token is the child of the located node (NP.

SRF2: if the frequency of the path from the located node (NP to opinionated trigger word node is only one, then we treat this path as null, and we also nullify the tokens activated by this node (NP.

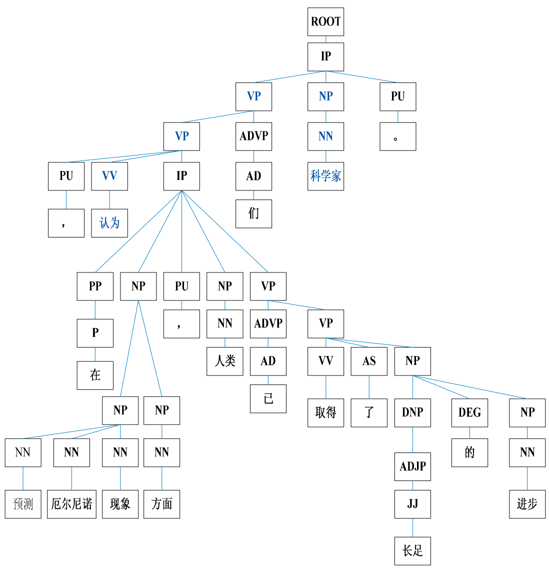

Figure1. A parsing example.

5. Relevance sentence Judgment

For the sentence relevance subtask we applied language model method. For the given topic, we evaluate the relevance between the topic and the sentence as formula (2),

$$p(Q/d)=\sum_{\forall Q'} p(Q/Q',d)p(Q'/d) \approx \sum_{\forall Q'} p(Q/d)p(Q'/d) \quad (2)$$

Q is for the terms extract from the given topic and Q' is for the expansion terms. The key point of this model is query expansion so we expand the query as follows:

We firstly find the relevant terms from Wikipedia. We put the given topic into Wikipedia and find the terms as the candidate expansion from the returned results. Then compute PMI value between the topic and candidate expansion as formula (3):

$$PMI = \frac{C(ext,top)}{C(ext) * C(top)} \quad (3)$$

Where top are the terms extract from the given topic and ext are the expansion terms. $C(ext,top)$, $C(ext)$, $C(top)$ are the number of the search results from Google for searching $ext \& top$, ext, top. The term with a higher value of PMI is more relevant to the topic. We sort the results according to the PMI value and select 20 of them.

6. Polarity Judgment

As mentioned in section 1, we set up a model of dynamic adjustment for the sentiment value of the words based on sentiment dictionary (SVDA), as Figure 2 shows.

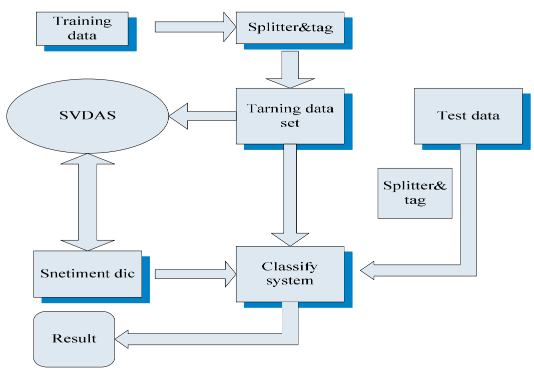

Figure2. Dynamic adjustment model for the sentiment value of the words.

In the model of SVDA, we record the times of the sentiment word that occurs in the training data. T_{pos} is the times of the word in dictionary occur in the positive train data, and T_{neg} is in the negative train data. Then we calculate the $Sentiment\_val$ (the sentiment value of the word which can be found in the dictionary) by the following formulas:

$$val = (T_{pos} - T_{neg}) / (T_{pos} + T_{neg}) \quad (4)$$

If ($val > 0$)

$$Sentiment\_val = T_{pos} \times val \quad (5)$$

If ($val < 0$)

$$Sentiment\_val = T_{neg} \times val \qquad (6)$$

If (val = 0 && word is in positive dictionary)

$$Sentiment\_val = 1 \qquad (7)$$

Otherwise:

$$Sentiment\_val = -1 \qquad (8)$$

So each word in the sentiment dictionary will be given a sentiment value, and we accumulate the value of each sentiment word in the sentence, and set the result as the sentiment value of the sentence, denoted as $Sentence\_val$.

If the $Sentence\_val$ is greater than a threshold, the sentence will be classified as positive, otherwise, if the $Sentence\_val$ is less than another threshold, the sentence will be classified as negative, others will be charged as neutral.

7. Experiments and Evaluation

As mentioned in section 1, we submitted two runs for each task. The lenient evaluation results for simplified Chinese monolingual are listed in the following Tables. Table 1 shows the results of opinion sentence judgment, our system ranked second among all participants, while the F-value of run2 is a little higher than run1, the reason is that in run1 we only use TSVM to classify the testing set, but in run 2 we firstly use the syntactic structure templates to tag sentences as opinion sentence and then we input the remaining testing data to the TSVM classifier. So we obtain a higher recall with a little lower precision. Also, the results indicate the features we extract for TSVM are effective for judging. In the following experiments including opinion holder identification, polarity and relevance judgment, we firstly filter testing set by using our opinion sentence extraction system, so the recall of all of these three subtasks is a little lower, which can not completely indicate the performance of our systems. Though, we still got good performance.

Table 2, 3 shows the results of opinion holder identification for opinionated sentences and for all sentences submitted, as can be seen, the precision of the system is good and stable, which demonstrates effectiveness of the features we used for CRF and syntactic rules we proposed, as well as the additional features that are extracted from the syntactic rules. However, there still exists some problems, like in one sentence there may have two opinion who put forward an opinion, it is hard for us to identify; another problem we need to research is how we should limit the length of opinion holder, some results are wrong since the opinion holder may only contain one word after segmentation or may contain more than one word.

Table 4 displays the results of relevance judgment, and shows the effectiveness of expansion terms given topics, the problem is in both runs we the sentences we proposed are not enough, we should focus on improving the recall effectively.

Table1. The lenient results of opinion sentence judgment

Lenient	Precision	Recall	F-value
Run1	38.43	67.53	48.98
Run2	35.02	87.81	50.07

Table2. The results for identifying opinion holder for opinionated sentences

	Precision	Recall	F-value
Run1	0.929	0.473	0.627
Run2	0.894	0.496	0.638

Table3. The results for identifying opinion holder for all sentences

	Precision	Recall	F-value
Run1	0.31	0.473	0.375
Run2	0.242	0.496	0.325

Table4. The lenient results of relevance sentence judgment

Lenient	Precision	Recall	F-value
Run1	98.23	39.64	56.49
Run2	97.9	51.55	67.55

Table5. The lenient results of polarity judgment

Lenient	Precision	Recall	F-value
Run1	60.41	27.18	37.49
Run2	58.13	31.13	40.55

8. Conclusion and Future works

This paper shows the methods for Multilingual Opinion Analysis Task (MOAT) by BUPT to Simplified Chinese monolingual, and submitted four tasks including opinionated, opinion holder identification, relevance and polarity judgments. For opinionated judgment, our results ranked two, which indicates the effectiveness of the syntactic structure templates we proposed and the TSVM classifier based on features both semantic-level and grammar-level. To extract opinion holder, we propose a novel syntactic rules to identify opinion holder combined with CRF. As to judge relevance, we considered the expansion of terms based on Statistical Language Models. For polarity judgment, we set up a dynamic adjustment model for the sentiment value of the words based on sentiment dictionary.

9. Acknowledgments

This work is supported by the research center of Intelligence Science and Technology of Beijing University of Posts and Telecommunications.

10. Reference

[1] Y.Seki, D.Evans, L. Ku, H.Chen, N. Kando, and C.Lin. Overview of Opinion Analysis Pilot Task at NTCIR-8. In Proceedings of the 8th NTCIR Workshop Meeting, Japan, 2010.

[2] Janyce Wiebe M, Bruce, Rebecca F, O'Hara, Thomas P. Development and use of a gold standard data set for subjectivity classifications. In Proc. 37th Annual Meeting of the Assoc. for Computational Linguistics (ACL-99). Association for Computational Linguistics, University of Maryland, June, pp. 246-253.

[3] Hatzivassiloglou, Vasileios, Janyce Wiebe. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. International Conference on Computational Linguistics (COLING-2000).

[4] Ellen Riloff, Janyce Wiebe. Learning Extraction Patterns for Subjective Expressions. Conference on Empirical Methods in Natural Language Processing (EMNLP-03).ACL SIGDAT. Pages 105-112.

[5] Yinming Yang, Jan O.Pedersen. A Comparative Study on Feature Selection in Text Categorization. Proceedings of the 14th International Conference on Machine Learing. Nashville ： Morgan Kaufmann, 1997：412～420.

[6] Yejin Choi , Claire Cardie , Ellen Riloff , Siddharth Patwardhan, Identifying sources of opinions with conditional random fields and extraction patterns, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp.355-362, October 06-08, 2005, Vancouver, British Columbia, Canada.

[7] Kim, Soo-Min and Eduard Hovy. 2005. Identifying Opinion Holders for Question Answering in Opinion Texts. Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains.

[8] S.-M. Kim and Eduard Hovy. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In Proceedings of workshop on sentiment and subjectivity in text at proceedings of the 21st international conference on computational linguistics/the 44th annual meeting of the association for computational linguistics (COLING/ACL 2006) (pp. 1-8). Sydney, Australia.

[9] Y. Seki, "Opinion holder extraction from author and authority viewpoints", Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2007), pp. 841-842.

[10] Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2004. Automatic Extraction of Opinion Propositions and their Holders, AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications.

[11] Ruifeng Xu, Kam-Fai Wong and Yunqing Xia. Opinmine-Opinion Analysis Sytem by CUHK for NTCIR-6 Pilot Task. In Proc of the six NTCIR Workshop. May.2007, Japan. 2007

[12] Ruihong Huang, Longxin Pan and Le Sun. ISCAS in Opinion Analysis Pilot Task: Experiment with sentimental dictionary based classifier and CRF model. In Proc of the six NTCIR Workshop. May.2007, Japan. 2007

[13] K. Liu, and J. Zhao, "NLPR at Multilingual Opinion Analysis Task in NTCIR7," pp. 226-231.

[14] D.Evans. A low-resources approach to Opinion Analysis: Machine Learning and Simple Approaches. In Proceedings of the 6th NTCIR Workshop Meeting, Japan, 2007.

[15] Y.Li, K. Bontcheva and H. Cunningham. Experiments of Opinion Analysis on the Corpora MPQA and NTCIR-6. In Proceedings of the 6th NTCIR Workshop Meeting, Japan, 2007.

[16] [14] Youngho Kim, Seongchan Kim, Sung-Hyon Myaeng. Extracting Topic-related Opinions and their Targets in NTCIR-7. In Proceedings of the 7th NTCIR Workshop Meeting, Japan,2008.

[17] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of EMNLP-2002.

[18] Turney, Peter D. 2002. Thumbs up or thumbsdown? Semantic orientation applied to unsupervised classification of reviews, Proceedings of ACL-02, Philadelphia, Pennsylvania,417-424.

[19] Kim, Soo-Min and Eduard Hovy. 2004. Determining the Sentiment of Opinions. Proceedings of COLING-04. pp. 1367-1373. Geneva, Switzerland.