

Opinion Sentences Extraction and Polarity Classification Using Automatically Generated Templates

Wan-Chi Huang, Meng-Chun Lin, Shih-Hung Wu*
 Department of Computer Science and Information Engineering
 Chaoyang University of Technology
 Taichung County 41349, Taiwan (R.O.C)
 *Contact author: shwu@cyut.edu.tw

ABSTRACT

The paper reports the approach of cyut system in NTCIR-8 MOAT subtask. We submitted the results of opinion judgment and polarity judgment in Traditional Chinese. Our study focused on automatically generated templates as the only features of classifier. The templates combining words with Part-of-speech or named-entity (POS/NE) tags are acquired from the training set. Experiment results show that, the template generation technology can get the same result without human edited knowledge.

Keywords

Opinion mining, automatically generated templates

1. INTRODUCTION

In this paper, we show how using template help with opinion sentence extraction and polarity identification in the NTCIR-8 MOAT subtasks. The extraction of opinion is to identify the sentences with opinion in an article. The polarity identification is a classification of a sentence into one of the three classes: positive, negative, or neutral. We treat the two sub-task as a two stage of classification. The flowchart is shown in Figure 1.

1.1 Related Work

Ku et al. [1] used various linguistic features for opinion sentences extraction and polarity identification. This approach requires a human edited dictionary NTUSD [2] as domain knowledge. Our approach is to generalize the approach, i.e. using domain independent features for classification.

2. OUR APPROACH

Our system adopts Support Vector Machine (SVM) [6, 7, 8, 9] as our classifier to decide whether a sentence is opinioned or not, and if a sentence is opinioned, another classifier is used to decide whether the input sentence is positive, negative, or neutral. The first classifier reports the result of opinioned or not, and the second classifier reports the polarity. The opinioned sentences in NTCIR6 and NTCIR7 corpus are used to train the classifier.

The features used in our system are templates combining words with the Part-of-speech or named-entity (POS/NE) tags acquired from training set and a manually edited knowledge base, the NTUSD [2]. Since it is costly to build the knowledge and the knowledge is not portable to other languages, we will try to use only the automatically generated templates instead manually edited lexicon resource.

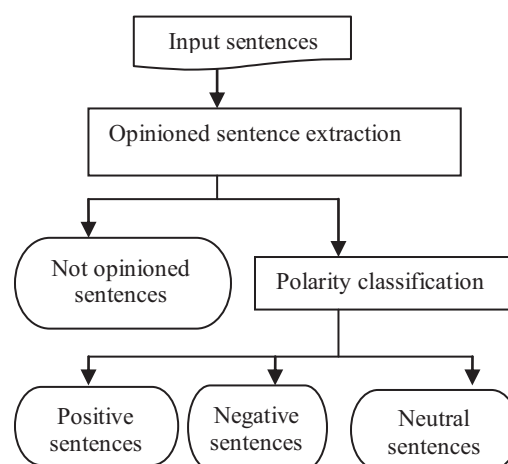


Figure 1. System judgment flowchart

2.1 Template Generation

To construct the templates form training set, we select the opinioned sentences in NTCIR6 and NTCIR7 MOAT corpus. The sentences were tagged as positive, negative, and neutral. Our system adopts the regular expression learning technology [3] to generate templates that consist of words, POS or NE tag. The Part-of-speech (POS) is tagged by the Academia Sinica CKIP toolkit (<http://ckipsvr.iis.sinica.edu.tw/>). The named-entity (NE) extraction tag is tagged by Mencius system [5], which will give person name, location name, and organization name.

The generation process is as follows. Consider an opinioned sentence “大陸/大部分的/地區/降水/偏/多”, which will be tagged with POS and NE tags, see Table 1.

The first column shows the segmented words of the original sentence. The second column shows the corresponding POS tags. The third column shows the NE tags if any. To avoid the explosion of combination, we use NE to replace POS if there is one NE tag for a text. Therefore, the number of possible template is reduced from 3^n to 2^n . This replacement is valid, since the NE tag (Location) is a more detail example of the POS tag (Nc), which also indicates a name of a place.

Table1. The data to generate template

TEXT	POS	NE
大陸	Nc	Location

大部分	Neqa	Other
的	DE	Other
地區	Nc	Other
降水	VA	Other
偏	D	Other
多	VH	Other

An n-gram template is a sequence of length n . The terms in the sequence can be text or POS/NE tag. For example, a tri-gram templates in Table 1 can be:

“大陸/大部分/的”, “location/大部分/的”, “大陸/Neqa/的”, “location/Neqa/的”, “大陸/大部分/DE”, or “location/大部分/DE”.

And the bi-gram templates can be:

“大陸/大部分”, “location/大部分”, “大陸/Neqa”, “location/Neqa”, or “局部/地區”.

Since the POS/NE tag is too general, they do not suite to be used as features for classifier, unigram features used in our system are the text.

After all the template candidates are generated, our system counts the frequency of each candidate in each of the three training corpus. Templates with low frequency will be deleted. The templates with relative high frequency will be used as features for classifiers.

2.2 Feature sizes

We submitted three official runs. At run 1, we adopt NTUSD and templates up to length 3 as features of SVM. The total dimension is 11. Two are from NTUSD positive and negative. Nine from our N-gram templates: they are unigram, bigram and trigram templates for positive, negative, and neutral.

Since the numbers of templates are very large, we divide the templates into several sub-set, and treat each sub-set (size 6000 or 3000) as a feature. The number of feature thus increases to 57 at Run 2, and 114 at Run 3. And in Run 2 and Run 3, we omitted the NTUSD to see if we can use domain independent features only. The setting is summarized in Table 2.

Table2. Setting of official runs of cyut

	NTUSD	N-gram	Feature size
Run 1	Yes	1,2,3	11
Run 2	NO	1,2,3	57
Run 3	NO	1,2,3	114

2.3 Setting of additional runs

In additional runs, we compare three factors in our official

runs. First is the number of N-gram. Second is with or without NTUSD. Third is the number of features.

In the official runs, we test only the templates with length less than 3. In the additional runs we test the templates with length less than 4. Table 3 listed the experiment settings of comparable runs. The sizes of features increase with different number, since we divide the 4-gram templates into sub-sets with various sizes. We merge 3000 templates into one feature in Run 6, 7, 10, and 11. We merge 6000 templates into one feature in Run 4, 5, 8, and 9. Originally, in Run 1, 2, 3, 4, 6, 8, an 10, the feature size of NTUSD is 2. Since the number of words in NTUSD is also above ten thousand, we divide it into several sub-set (size 300), therefore, the feature size of NTUSD is 37 for Run 5, 7.

Then we compare the features with or without NTUSD. Note that, without NTUSD, the feature size will decrease by 2. Table 4 listed the experiment settings comparable runs.

We also change the sizes of sub-sets in different settings, the feature sizes are also changed accordingly. Table 5 listed the experiment settings of comparable runs.

Table3. Comparable runs with different N-gram

	NTUSD	N-gram	Size
Run 4	Yes	1,2,3	59
Run 5	Yes	1,2,3	116
Run 6	Yes	1,2,3	94
Run 7	Yes	1,2,3	151
Run 8	Yes	1,2,3,4	106
Run 9	Yes	1,2,3,4	141
Run 10	Yes	1,2,3,4	163
Run 11	Yes	1,2,3,4	198

Table 4. Comparable runs with or without NTUSD

	NTUSD	N-gram	Size
Run 2	No	1,2,3	57
Run 3	No	1,2,3	114
Run 12	No	1,2,3,4	104
Run 15	No	1,2,3,4	161
Run 4	Yes	1,2,3	59
Run 6	Yes	1,2,3	116
Run 8	Yes	1,2,3,4	106
Run 10	Yes	1,2,3,4	163

Table 5. Comparable runs with more or less features

	NTUSD	N-gram	Size
Run 2	No	1,2,3	57
Run 12	No	1,2,3,4	104
Run 15	No	1,2,3,4	161
Run 13	No	1,2,3	29
Run 14	No	1,2,3,4	52
Run 16	No	1,2,3,4	81

3. EXPERIMENT RESULTS

Table 6 shows the opinion sentence extraction and polarity identification results of our system. Where Run 1 to 3 [10] are official runs of NTCIR-8 MOAT subtask in traditional Chinese; and Run 4 to 16 are additional runs. The P, R, and F are the

precision, recall, and F-score respectively. The best results are in bold font.

The best F-score on opinion sentences extraction is in Run 7 and 11, which have the largest number of feature size. It suggests that, dividing templates into smaller groups might have better performance.

The best F-score on polarity classification is in Run 5. The setting is using up to tri-gram templates with NTUSD as the features.

4. CONCLUSIONS

By observing the result of the comparable runs in Table 4, we find that runs with NTUSD still get better results than runs without NTUSD. However, the performances are very close. The best f-score of opinion sentence extraction is 0.57 without NTUSD, compare to 0.59 with NTUSD. Thus, the domain independent features, automatically generated templates, can be useful on opinion mining.

We also find that the number of n-gram increase from 3 to 4 do not affect much on the performance. This might due to the 4-gram templates have less appearance in the test set. Changing the number of features by dividing templates into various sub-sets might increase or decrease the performance. Further study is necessary. In the future, designing a better feature selection strategy might be necessary.

Table 6. Experiment results

Run ID	Opinion			Polarity		
	P	R	F	P	R	F
1	0.42	0.87	0.57	0.40	0.35	0.37
2	0.41	0.82	0.54	0.31	0.25	0.28
3	0.47	0.44	0.45	0.36	0.16	0.22
4	0.44	0.85	0.58	0.39	0.34	0.36
5	0.45	0.82	0.58	0.44	0.37	0.40
6	0.45	0.83	0.58	0.40	0.33	0.36
7	0.46	0.81	0.59	0.43	0.35	0.39
8	0.44	0.83	0.58	0.39	0.33	0.36
9	0.45	0.82	0.58	0.44	0.36	0.39
10	0.45	0.83	0.58	0.40	0.33	0.36
11	0.46	0.80	0.59	0.43	0.34	0.38
12	0.44	0.82	0.57	0.39	0.33	0.36
13	0.43	0.81	0.56	0.41	0.33	0.37
14	0.43	0.83	0.57	0.41	0.34	0.37
15	0.44	0.81	0.57	0.40	0.32	0.36
16	0.44	0.80	0.56	0.40	0.33	0.36

5. ACKNOWLEDGEMENT

This research was partly supported by the National Science Council under Grant No. NSC98-2221-E-324 -025.

6. REFERENCES

[1] Ku, L.-W., Liu, I.-C., Lee, C.-Y., Chen, K.-h. and Chen, H.-H. (2008). Sentence-Level Opinion Analysis by CopeOpi in NTCIR-7. Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question

Answering, and Cross-Lingual Information Access, December 16-19, 2008, Tokyo, Japan, 260-267.

[2] Ku, L.-W., and Chen, H.-H. (2007). Mining Opinions from the Web: Beyond Relevance Retrieval. Journal of American Society for Information Science and Technology, Special Issue on Mining Web Resources for Enhancing Information Retrieval, 58(12), 1838-1850.

[3] Li, Yunyao, Rajasekar Krishnamurthy, Sriram Raghavan, and Shivakumar Vaithyanathan. "Regular Expression Learning for Information Extraction." Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. Honolulu, 2008. 21-30.

[4] C.-L. Sung, C.-W. Lee, H.-C. Yen and W.-L. Hsu, "An alignment-based surface pattern for a question answering system", in: IEEE International Conference on Information Reuse and Integration, 2008, pp. 172-177.

[5] Tzong-Han Tsai, Shih-Hung Wu, and Wen-Lian Hsu, "Mencius: A Chinese Named Entity Recognizer Based on a Maximum Entropy Framework," Computational Linguistics and Chinese Language Processing, Vol.9, No.1, pp.65-82, 2004.

[6] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, 1998.

[7] Christopher M. Bishop, "Pattern Recognition and Machine Learning", *Springer*, 2006.

[8] Rung-Ching Chen, Chung-Hsun Hsieh, "Web page classification based on a support vector machine using weighted vote schema", *Expert Systems with Applications*, Volume 31, Issue 2, August 2006, Pages 427-435.

[9] Michael Chau, Hsinchun Chen, "A machine learning approach to web page filtering using content and structure analysis", *Decision Support Systems*, Volume 44, Issue 2, January 2008, Pages 482-494.

[10] Yohei Seki, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, Noriko Kando, "Overview of Multilingual Opinion Analysis Task at NTCIR-8", In Proceedings of the 8th NTCIR Workshop, 2010.