

Modeling Pollyanna Phenomena in Chinese Sentiment Analysis

Ting-Hao (Kenneth) Huang¹ Ho-Cheng Yu² Hsin-Hsi Chen²

(1) Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, U.S.

(2) National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan (R.O.C)

tinghaoh@cs.cmu.edu, p98922004@csie.ntu.edu.tw, hhchen@ntu.edu.tw

ABSTRACT

This paper proposes a method to enhance sentiment classification by utilizing the Pollyanna phenomena. The Pollyanna phenomena describe the human tendency to use positive words more frequently than negative words. This word-level linguistic bias can be demonstrated to be strong and universal in many languages. We perform detailed analyses of the Pollyanna phenomena in four Chinese corpora. Quantitative analyses show that for documents with few positive words, the word usages in documents from either the positive or the negative polarities become similar. Qualitative analyses indicate that this increase of similarity of word usage could be caused by the concentration of topics. By taking advantage of these results, we propose a partitioning strategy for sentiment classification and significantly improve the F1-score.

應用於中文情緒分析之波莉安娜效應研究

摘要

本研究提出一以波莉安娜效應改善情緒分類之方法。波莉安娜效應指於人類語用中，正面詞詞頻高於負面詞詞頻之語言現象。此一詞彙層次偏斜現象非僅存在許多語言中，且強度更十分顯著。本研究首先於四個中文語料庫中分析波莉安娜效應。定量分析顯示，含有較少正面詞彙的特定情緒傾向文件間，相較於正負面詞比例正常的文件間，具有較高的文字相似度；定性分析則指出，該現象乃由於主題集中化所造成。基於上述分析結果，本研究繼而提出一切分資料之策略以改進情緒分類效能，並於實驗中有效提升F1-score。

KEYWORDS : Sentiment Classification, Pollyanna Phenomena

KEYWORDS IN MANDARIN : 情緒分析, 情緒分類, 波莉安娜效應

1 Introduction

The human tendency to use positive words more frequently than negative words was originally called “the Pollyanna Hypothesis,” named after a fictional young girl with infectious optimism (Porter, 1913), by Boucher and Osgood (1969). This word-level linguistic positivity bias had been not only discussed by early studies, e.g., (Johnson et al., 1960) and (Zajonc 1968), but also explored by contemporary scholars. Zou (2004) analyzed the frequencies of the positive and negative Chinese words based on the Modern Chinese Frequency Dictionary (Wang 1986), and concluded their ratio as 7:3. Similar supporting evidences were also found in various other languages, e.g., English (Augustine et al., 2011; Garcia et al, 2012; Kloumann et al, 2012), Italian (Suitner and Maass, 2008), German and Spanish (Garcia et al, 2012), and even across 20 different languages (Rozin et al., 2010). In contrast, only a few works addressed this particular issue in opinion mining and sentiment analysis. These papers (Bolasco and della Ratta-Rinaldi, 2004; Brooke, 2009; Mohammad et al., 2009) demonstrated supporting evidences of the Pollyanna Hypothesis. Taboada et al. (2009) and Brooke et al. (2009) claimed the positivity bias could affect lexicon-based sentiment analysis systems like those of Kennedy and Diana (2006), and proposed an adjusting strategy (Taboada et al., 2011).

The contribution of this paper is three-fold: (1) To the best of our knowledge, we conduct the first detailed survey of the Pollyanna phenomena in various modern Chinese corpora. (2) Through quantitative and qualitative analyses, we discover that for the documents with relatively fewer positive words, the intra-polarity document similarity, of either positive or negative opinion polarity, significantly increases. (3) Based on our findings, we propose a strategy for sentiment classification and improve performance significantly.

2 Word-Level Linguistic Positivity Bias

In this section, we aim to explore details of the Pollyanna phenomena in Chinese. Our work focuses on the word-level linguistic bias, i.e., the unbalanced distributions of positive/negative words’ occurrences.

		CTB	RECI	iPeen	MOAT
Basic Information	Data Type	Generic Corpus	News Opinion Summary	Restaurant Review	Evaluation Data
	Data Instance Type	Document			Sentence
	#Instance (Inst.)	892	2,389	19,986	4,652
	Avg. Inst. Length (#Word)	532.25	60.36	331.84	16.06
Sentimental Information	Opinion Polarity Label	Untagged	POS, NEG	POS, NEG, NEU	POS, NEG, NEU
	% Pos. Instance	-	59.36%	44.16%	8.88%
	% Neg. Instance	-	40.64%	7.88%	11.11%
	Avg. Pos. WF	0.10	0.11	0.08	0.11
	Avg. Neg. WF	0.02	0.03	0.03	0.06
	Avg. Bias	0.66	0.47	0.48	0.15

TABLE 1: Statistics of the four corpora

2.1 Experimental Datasets

To realize the Pollyanna phenomena in Chinese, we analyze four modern corpora of different genres: opinionative real estate news (RECI), users' comments on restaurants (iPeen), a dataset for a multilingual opinion analysis task (MOAT), and a generic corpus (Chinese Treebank).

The Quarterly Report of Taiwan Real Estate Cycle Indicators (RECI) (Chin, 2010) has been collecting and analyzing Taiwan's opinionative real estate news, and releasing reports to the public every three months since 2002. Each RECI report contains 50 to 70 opinionative news excerpts labeled with opinion polarities. In this study, we select excerpts with "Positive" and "Negative" labels from 2002 to 2010 to set up a RECI corpus; iPeen (<http://www.ipeen.com.tw/>) is a restaurant review website. Each registered user can post his/her comments and rating points from 0, 5, ..., 55, to 60 toward any restaurants. We randomly collect about 20,000 non-empty posts from iPeen and tri-polarize the opinion polarities of posts. The posts with rating points lower than 20 are labeled as "Negative", those higher than 40 are labeled as "Positive", and the remaining posts are labeled as "Neutral"; The NTCIR Multilingual Opinion Analysis Task (MOAT) (Seki et al., 2008) provided a dataset for evaluating opinion mining technologies. We use the Traditional Chinese test set with the "strict" annotating standard. Each sentence in the set is annotated with opinion polarity by three assessors. All three assessors must provide the same label for a sentence, otherwise its label will be "Neutral"; Finally, the Chinese Treebank 5.1 (CTB) (Palmer et al., 2005) is also adopted for comparison. The statistics of these corpora are summarized in TABLE 1.

2.2 Deep Analysis

In document/sentence level, TABLE 1 demonstrates that the linguistic bias is not always positive. RECI and iPeen have different degrees of preference for positive document, while MOAT corpus has a slight higher percentage for negative sentences.

In word level, we tag all four corpora by an extended version of the NTU Sentiment Dictionary© (NTUSD) (Ku and Chen, 2007), which contains 9,365 positive words and 11,230 negative words. Every word in NTUSD is annotated by multiple human annotators and is examined by one or more experts. We then define positive word frequency (WF) in a document/sentence to be the total number of occurrences of positive words divided by the length of the document/sentence. (The negative word frequency is defined in the same way.) TABLE 1 shows that the average positive word frequencies in these four corpora are 1.83 to 5 times of those of negative words. The ratio Zou (2004) concluded between positive and negative words, i.e., 7:3, also fell within this range. We further propose an indicator Bias(d) in Equation (1) to measure the degree of word-level linguistic bias in a given document/sentence d.

$$\text{Bias}(d) = \frac{c_p(d) - c_n(d)}{c_p(d) + c_n(d)}, \quad \begin{cases} C_p(d) = (\text{Number of positive words in } d) + 1 \\ C_n(d) = (\text{Number of negative words in } d) + 1 \end{cases} \quad (1)$$

Bias(d) is a smoothed and normalized version of the positive-negative word count ratio. The absolute value of Bias(d) denotes the magnitude of bias, and the sign shows the direction of bias. The last row of TABLE 1 shows that the average biases in the document-based corpora (i.e., CTB, RECI and iPeen) are 0.66, 0.47, and 0.48, respectively; and 0.15 in the sentence-based corpus (i.e., MOAT). These values reflect that the word-level positivity bias is not only strong, but also universal. These four corpora all have different characteristics in many aspects, but all of them show strong agreements with the Pollyanna Hypothesis.

3 Intra- and Inter-Polarity Similarity Analysis

The above analyses raise an interesting question: What happens in those outlier documents whose positive-negative word ratios are “not that positive”? One intuitive guess is that those documents mostly represent negative opinions. However, when we look at the lower positively biased set, the number of negative documents is not always the largest; Another guess is that people tend to use fewer positive words only in some specific occasions or to describe some certain content, i.e., the use of words in those less positively biased documents could be possibly similar to each other. Our preliminary study suggests that this rise of similarity does not happen uniformly in all documents with lower bias values, but only in certain opinion polarity. In this section, we aim to quantitatively examine this observation. If this guess turns out to be true, it could be potentially beneficial for sentiment classification.

3.1 Methodology

Our goal is to measure the average degree of similarity of word use between documents of the same and different opinion polarities. The analyses are setup in the following way: First, we explore a bias value threshold β from -1 to 1 in steps of 0.1. For each β , those documents with bias values smaller than β form a lower set of β , and the remainder is called an upper set. In a lower set, we would then put documents of the same “target polarity” (positive or negative) together in a set P_T and place all remaining documents -- including all neutral documents if any -- into a “non-target polarity” set P_{NT} . Note that RECI only has two sentiment polarities (see TABLE 1), so its target and non-target polarities are interchangeable. Second, we use the TF-IDF vectors to represent documents, and calculate four different average cosine similarities S of all document pairs (x, y) in a lower set as follows.

- $S_T : x, y \in P_T, x \neq y$
- $S_{NT} : x, y \in P_{NT}, x \neq y$
- $S_{Inter} : x \in P_T, y \in P_{NT}$
- $S_{All} : x, y \in \text{lower set}, x \neq y$

S_T , S_{NT} and S_{Inter} are then normalized by S_{All} . In this paper, we use the asterisk ($*$) to indicate the normalized similarity. Finally, for every lower set with different β , we have three normalized similarities, i.e., S_T^* , S_{NT}^* and S_{Inter}^* . The former two respectively represent the intra-polarity similarity of documents with target and non-target polarities, and the latter one represents the inter-polarity similarity of documents from different opinion polarities.

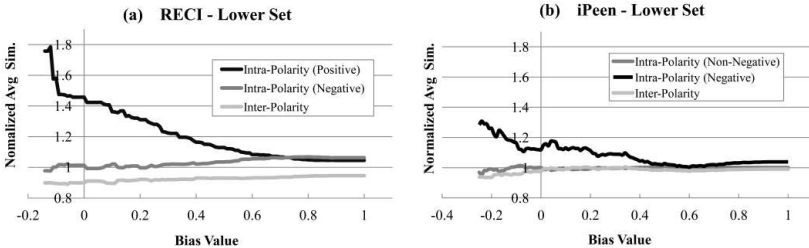


FIGURE 1: The curves of S_T^* , S_{NT}^* and S_{Inter}^* of the lower sets in (a) RECI ($P_T = \text{Positive}$) and (b) iPeen ($P_T = \text{Negative}$). The S_T^* obviously rise up when bias value decreases.

3.2 Results and Discussion

In each of RECI, iPeen and MOAT, the S_T^* , S_{NT}^* and S_{Inter}^* are drawn with respect to the different bias values. FIGURE 1(a) and 1(b) are the resulting curves of RECI and iPeen, where positive and negative polarities are respectively explored as targets. These two figures confirm our guess: Within the target opinion polarity, the average cosine similarity among documents (S_T^*) obviously rises up in the portion of data which has less positivity bias, while the S_{NT}^* and S_{Inter}^* still remain stable. In other words, when we look at those outlier documents with lower bias values, for certain target opinion polarity, people actually tend to use more similar words with each other. Incidentally, we also analyze the same target polarities of the upper sets respectively in RECI and iPeen. But the curves do not display any obvious consistent trends. Another issue is the choice of the target polarity. We run the analysis in iPeen corpus when targeting at the other opinion polarity, i.e., positive. However, neither in the lower set nor in the upper set of iPeen corpus the similarities demonstrate any obvious trends. Meanwhile, we find that the S_T^* in MOAT is always apparently higher than both S_{NT}^* and S_{Inter}^* , regardless of bias values, lower/upper sets, and target polarity. It could be caused by its strict annotating standard which only accepts the labels with perfect inter-annotator agreement.

4 Qualitative Analysis

In the less positively biased portion of data, we have quantitatively observed the increase of cosine similarity among documents. Then the next question should naturally be, when using less positive words, what do people actually talk about? In this section, we adopt quantitative analyses and try to give an insightful interpretation.

In iPeen, we compare the negative documents in the upper set and the lower set (with partitioning bias value 0.1). In general, negative comments toward restaurants cover a wide range of topics. As expected, both the number of documents and the diversity of topics are much higher in the upper set. However, interestingly, we find that the negative comments mainly focus only on the "poor service" in the lower set. As a result, the topics of negative comments in the lower set become more focused than those in the upper set. This observation reasonably explains the increase of intra-class similarity of negative polarity in iPeen's lower set; In RECI, while the positive news in the upper set of RECI (with partitioning bias value 0.1) covers a wide range of topics, the positive news in the lower set of RECI mostly focus on the indicators and rates which would be better if reduced, e.g., unemployment rate, land value increment tax rate, lending rate, overdue loans ratio, inheritance tax, etc. As a result, the narrow focus of topics raises the intra-class similarity of positive polarity in RECI's lower set.

To conclude, the phenomena we find in Section 3 actually reflect the shrinkage of topics in the less positively biased portion of data. Most topics have strong preferences for positive words. However, few specific topics still relatively prefer negative words, e.g., the "poor service" of restaurant. These topics are emphasized when we isolate the lower positively biased data, and thus decide in which opinion polarity we can observed the rise of intra-polarity similarity.

5 Partitioning Strategy for Sentiment Classification

Our analyses above reveal the increase of intra-polarity similarity in certain part of data, and thus shed light on sentiment classification. In this section, we propose a strategy to partition the data

sets by bias value, and train another model for the data portion which has lower positivity bias. A set of experiments are run to determine how much better this strategy can achieve.

The goal of our sentiment classifier is to predict the opinion polarity of documents respectively in RECI and iPeen. The TF-IDF vector of each document is adopted as features, and the libSVM (Chang and Lin, 2011) with linear kernel is selected as our model. One fifth of data are randomly selected as the testing set, and the rest are the training set. Without loss of generality, we select a bias value β for each corpus based on the document distributions and the trend of S_T^* mentioned in Section 2. Both training and testing data are partitioned by this bias value. Two different models are trained on the upper and lower sets of training data, and then are evaluated on the corresponding subsets of testing data, respectively. For comparison, we also train a classifier with the whole training data. In the experiments, we explore all possible target polarities mentioned in the previous sections. The results are shown in TABLE 3(a) and 3(b). Note that besides evaluating our partition strategy in the upper and lower sets separately, we also merge the results of the two classifiers together (refer to the ‘‘Whole’’ column). The outputs of the original approach are also evaluated in the upper set, the lower set, and the whole sets, respectively. As a result, when classifying the target polarities which we found in FIGURE 1, our partition strategy significantly increases the F1-scores both of target and non-target polarities in the whole testing set of iPeen ($p < 0.01$) and RECI ($p < 0.01$). Note that each of our Partition classifier actually beats the Original classifier by using less training data. On the other hand, as expected, our Partition strategy does not outperform the Original approach when predicting the positive polarity in iPeen, which is not the opinion polarity we observed in Section 3.

(a) RECI	Polarity	Strategy	Lower	Upper	Whole	(b) iPeen	Polarity	Strategy	Lower	Upper	Whole
	Pos.	Original		.646	.846		.833	Neg.	Original		.333
Partition			.692	.874	.858**	Partition			.342	.342	.342**
Neg.	Original		.789	.692	.726	non-Neg.	Original		.811	.927	.921
	Partition		.869	.704	.761**		Partition		.870	.929	.926**
#Positive Docs.			191	1,227	1,418	#Negative Docs.			260	1,314	1,574
#Negative Docs.			348	623	971	#Non-negative Docs.			932	17,480	18,412

TABLE 3: The F1-Score of Sentiment Classification in (a) RECI, $\beta = 0.25$ (b) iPeen, $\beta = 0.15$, $P_T = \text{Negative}$ (**: $p < 0.01$).

Conclusion and perspectives

In this paper, we first provide a detailed study of the Pollyanna phenomena in various genres of Chinese. Then we focus on those documents which have less positivity bias. Through quantitative analysis, we reveal the obvious increase of average similarity in certain opinion polarity among these outlier documents; and through qualitative analyses, we draw insights to indicate that the increase could be caused by the concentration of topics. Finally, by taking advantage of the rise of intra-polarity similarity, we propose a partitioning strategy for sentiment classification and significantly improve the F1-score. Our goal is to build a robust automatic mechanism for sentiment modeling, and Pollyanna gives us a good clue about it.

Acknowledgments

We would like to thank Carolyn P. Rosé, Brian MacWhinney, Shou-I Yu, Kuen-Bang Hou (Favonia), and the anonymous reviewers for their valuable comments.

References

- Augustine, A. A., Mehl, M. R., and Larsen, R. J.. (2011). A Positivity Bias in Written and Spoken English and Its Moderation by Personality and Gender. *Social Psychological and Personality Science*, 2(5): 508-515.
- Bolasco, S. and della Ratta-Rinaldi, F.. (2004). Experiments on Semantic Categorisation of Texts: Analysis of Positive and Negative Dimension. In the *7es Journées internationales d'Analyse statistique des Données Textuelles*, pages 202-210, Louvain La Neuve, Belgium.
- Boucher, J. and Osgood, C. E.. (1969). The Pollyanna Hypothesis. *Journal of Verbal Learning and Behavior*, 8(1): 1-8.
- Brooke, J.. (2009). A Semantic Approach to Automated Text Sentiment Analysis (Master's thesis). Simon Fraser University, British Columbia, Canada.
- Brooke, J., Tofiloski, M., and Taboada M.. (2009). Cross-Linguistic Sentiment Analysis: From English to Spanish. In *The International Conference of Recent Advances in Natural Language Processing (RANLP 2009)*, pages 50-54, Borovets, Bulgaria.
- Chang, C.-C. and Lin, C.-J.. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2(27): 1-27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Chin, Y.-L.. (2010). A Review and Discussion of Real Estate Cycle Indicators Analysis and Publication Method. Research Project Report. Architecture and Building Research Institute, Ministry of the Interior, Taiwan.
- Garcia, D., Garas, A., and Schweitzer, F.. (2012). Positive Words Carry Less Information than Negative Words. *EPJ Data Science* 2012, 1(3).
- Johnson, R. C., Thomson, C. W., and Frincke, G.. (1960). Word Values, Word Frequency, and Visual Duration Thresholds. *Psychological Review*, Vol. 67(5): 332-342.
- Kennedy, A. and Inkpen, D.. (2006). Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters. *Computational Intelligence*, 22(2): 110-125.
- Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., and Dodds, P. S.. (2012). Positivity of the English Language. *PLoS ONE*, 7(1): e29484.
- 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*, 58(12): 1838-1850.
- Mohammad, S., Dunne, C., and Dorr, B.. (2009). Generating High-Coverage Semantic Orientation Lexicons from Overtly Marked Words and a Thesaurus. In *The 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, pages 599-608, Singapore.
- Palmer, M., Chiou, F.-D., Xue, N., and Lee, T.-K.. (2005). Chinese Treebank 5.1. The LDC Corpus Catalog LDC2005T01U01.
- Porter, E. H. (1913). *Pollyanna*. Boston, MA: L. C. Page.
- Rozin, P., Berman, L., and Royzman, E.. (2010). Biases in Use of Positive and Negative Words Across Twenty Natural Languages. *Cognition and Emotion*, 24(3): 536-548.

- Seki, Y., Evans, D. K., Ku, L.-W., Sun, L., Chen, H.-H., and Kando, N.. (2008). Overview of Multilingual Opinion Analysis Task at NTCIR-7. In The 7th NTCIR Workshop, pages 185-203, Tokyo, Japan.
- Suitner, C. and Maass, A.. (2008). The Role of Valence in the Perception of Agency and Communion. *European Journal of Social Psychology*, 38(7): 1073-1082.
- Taboada, M., Brooke, J., and Stede M.. (2009). Genre-Based Paragraph Classification for Sentiment Analysis. In The 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue (SIGDIAL 2009), page 62-70, Queen Mary University of London, UK.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M.. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2): 267-307.
- Wang, H.. (1986). *Modern Chinese Frequency Dictionary*. Beijing, China: Beijing Language Institute Press.
- Zajonc, R. B.. (1968). Attitudinal Effects of Mere Exposure. *Journal of Personality and Social Psychology Monograph Supplement*, 9(2): 1-27.
- Zou, S.. (2004). On Positive and Negative Senses / 积极意义和消极意义二题. *Chasing the Truth: Some Thoughts on Chinese Grammar Issues / 求真集:对汉语语法问题的一些思索*. Beijing, Joint Publishing: 82-93.