

# 基于最大熵模型的中文词与句情感分析研究\*

150001

[dongxishuang@gmail.com](mailto:dongxishuang@gmail.com), [guanyi@hit.edu.cn](mailto:guanyi@hit.edu.cn), [libenyang012566@163.com](mailto:libenyang012566@163.com), [ruoyu\\_928@126.com](mailto:ruoyu_928@126.com),  
[lisheng@hit.edu.cn](mailto:lisheng@hit.edu.cn)

摘 要

COAE2009

关键词

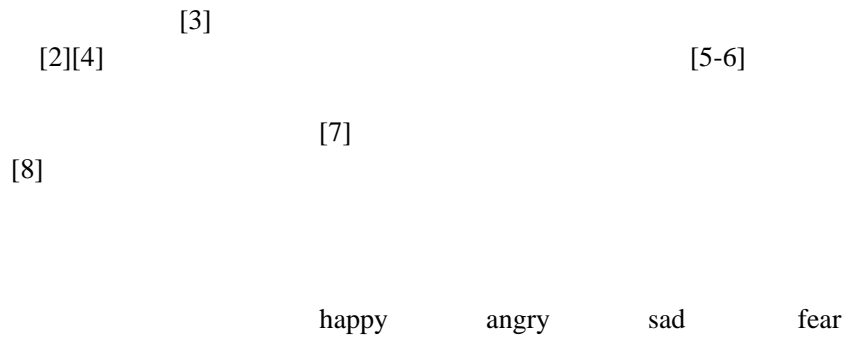
## Sentiment Analysis on Chinese Words and Sentences Based on Maximum Entropy Model

Dong Xi-Shuang, Guan Yi, Li Ben-Yang, Chen Zhi-Jie, Li Sheng

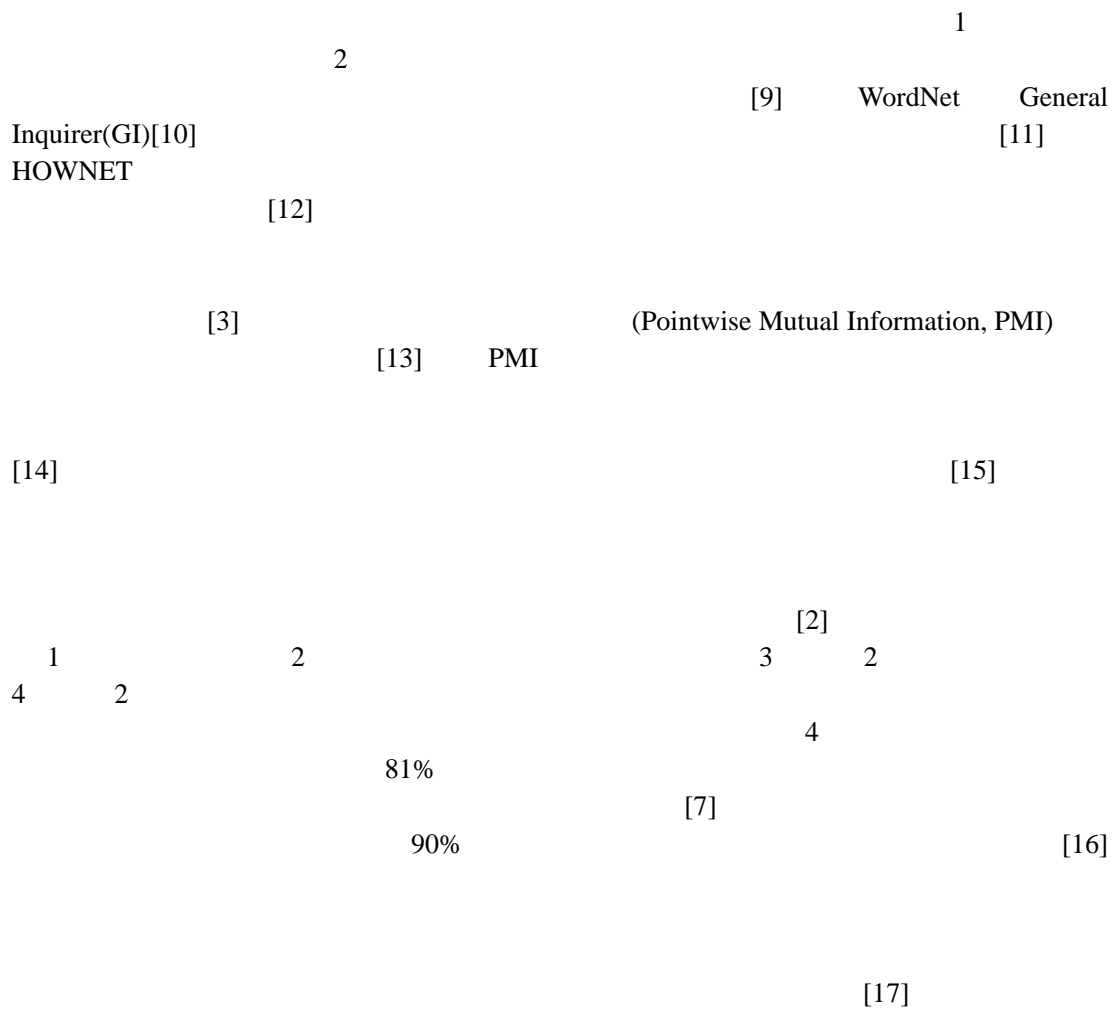
Harbin Institute of Technology, Harbin 150001

[dongxishuang@gmail.com](mailto:dongxishuang@gmail.com), [guanyi@hit.edu.cn](mailto:guanyi@hit.edu.cn), [libenyang012566@163.com](mailto:libenyang012566@163.com), [ruoyu\\_928@126.com](mailto:ruoyu_928@126.com),  
[lisheng@hit.edu.cn](mailto:lisheng@hit.edu.cn)

**Abstract:**



## 2 相关研究





(3)

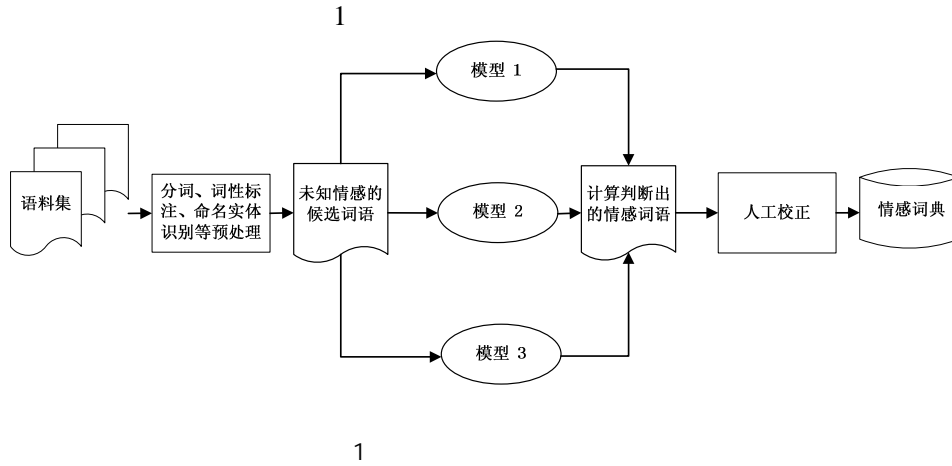


Fig. 1 the FlowChart of Constructing the Chinese Sentiment Lexicon

1 1 2

3

(1)

(2)

(3)

[17] 3

(1) LMR [17] LMR L M R

"

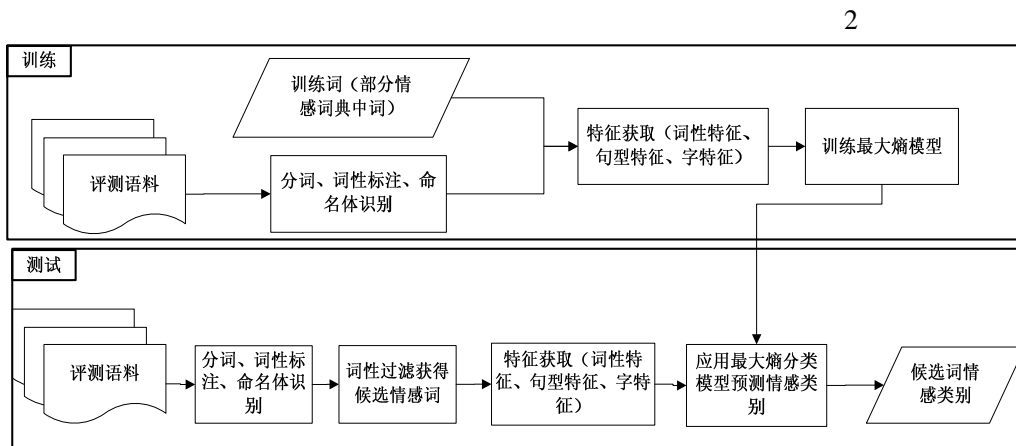
/n /d /v /n /w

n\_d\_v

(2)

(3)

" " " " " "



2

Fig. 2 the FlowChart of Recognition of Sentiment Words

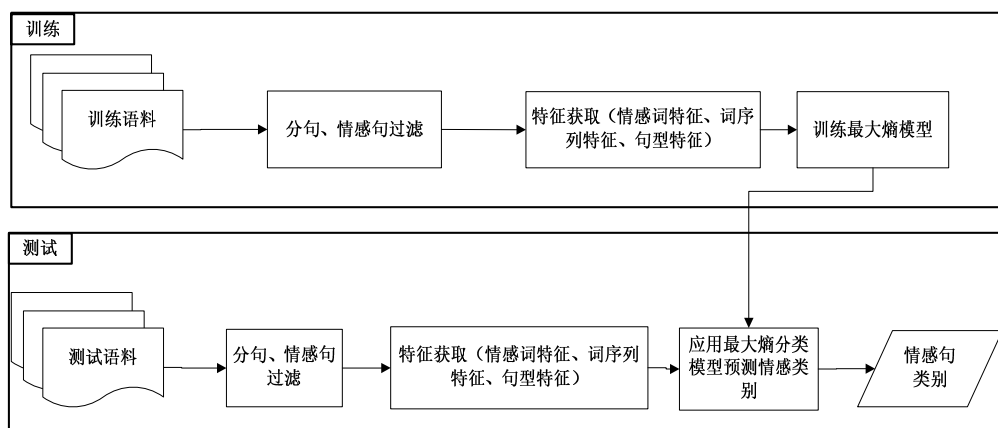
#### 4.2 句级情感分析

[3]

F

1%-2%

3



3

Fig. 3 the FlowChart of Recognition of Sentiment Sentences

1.

2.

(1)

(2)

6

3

$$C(x) \begin{matrix} 0 & x & 0 \\ 1 & 4 & x & 0 \\ 2 & 7 & x & 4 \\ 3 & 10 & x & 7 \\ 4 & x & 11 \end{matrix} \quad (3)$$

(3)  $x$

4

$$F(t) \begin{matrix} 1 & t & 24 \\ 2 & 100 & t & 24 \\ 3 & 300 & t & 100 \\ 4 & t & 300 \end{matrix} \quad (4)$$

$t$

## 5 实验与分析

### 5.1 数据

COAE2009  
dataset1 4

dataset1

txt  
20%

## 5.1 实验结果与分析

1

1

Tab.1 Sentiment Analysis on Words

	Preci si on	P@1000	Recal l	F1	R-accuracy
H TLTRTask1-angry-1. txt	0. 486239	0. 106	0. 114101	0. 18483	0. 114101
H TLTRTask1-fear-1. txt	0. 560284	0. 079	0. 116691	0. 193154	0. 116691
H TLTRTask1-happy-1. txt	0. 569892	0. 159	0. 102913	0. 174342	0. 102913
H TLTRTask1-sad-1. txt	0. 465565	0. 169	0. 137287	0. 212045	0. 137287
AVG	0. 520495	0. 128	0. 117748	0. 1910928	0. 117748
MEDI AN	0. 479425	0. 14263	0. 1254997	0. 1775775	0. 1254997
MAX	0. 619184	0. 192	0. 1795165	0. 2078715	0. 166954

F1

P@1000

R-accuracy

2 3

2

1

Tab.2 Sentiment Analysis on Sentences 1

	Preci si on	P@1000	Recal l	F1	R-accuracy
H TLTRTask2-angry-1. txt	0. 374618	0. 245	0. 149299	0. 213508	0. 149299
H TLTRTask2-fear-1. txt	0. 404124	0. 392	0. 238298	0. 299809	0. 238298
H TLTRTask2-happy-1. txt	0. 212308	0. 414	0. 201853	0. 206948	0. 182838
H TLTRTask2-sad-1. txt	0. 154557	0. 351	0. 182243	0. 167262	0. 136033
AVG	0. 28640175	0. 3505	0. 19292325	0. 22188175	0. 176617
MEDI AN	0. 158407	0. 366	0. 201223	0. 1246522	0. 15746
MAX	0. 370662	0. 6805	0. 374071	0. 232405	0. 189788

3

2

Tab.3 Sentiment Analysis on Sentences 2

	Preci si on	P@1000	Recal l	F1	R-accuracy
H TLTRTask2-angry-2. txt	0. 301822	0. 265	0. 161487	0. 210401	0. 161487
H TLTRTask2-fear-2. txt	0. 422311	0. 424	0. 257751	0. 320121	0. 257751
H TLTRTask2-happy-2. txt	0. 242439	0. 497	0. 242321	0. 24238	0. 210629
H TLTRTask2-sad-2. txt	0. 140903	0. 34	0. 176532	0. 156718	0. 129283
AVG	0. 27686875	0. 3815	0. 20952275	0. 232405	0. 1897875
MEDI AN	0. 158407	0. 366	0. 201223	0. 1246522	0. 15746
MAX	0. 370662	0. 6805	0. 374071	0. 232405	0. 189788

2

P@1000

F1

R-accuracy

fear		10%	
happy	sad		3
2		F1	

## 6 结论与展望

## 感谢

## 参 考 文 献

- [1] Bo Pang, Lillian Lee. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In Proceedings of the ACL, 2004, pp.271-278.
- [2] Soo-Min Kim, Eduard Hovy. Determining the sentiment of opinions. In Proceedings of COLING, 2004, pp.1367-1373.
- [3] Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002, pp.417-424.
- [4] Y. Mao, G. Lebanon. Isotonic conditional random fields and local sentiment flow. In Proceedings of NIPS, 2006.
- [5] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of EMNLP, 2002, pp.79-86.
- [6] Zhang, Y., Li, Z., Ren, F., et al. Semi-automatic emotion recognition from textual input based on the constructed emotion thesaurus. In Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2005, pp.571-576.
- [7] Hong Yu, Vasileios Hatzivassiloglou. Towards Answering Opinion Questions: Separating Facts from Opinions. In Proceedings of EMNLP-03, 2003, pp.129-136
- [8] Cardie Claire, Janyce Wiebe, Theresa Wilson, et al. Combining low-level and summary representations of opinions for multi-perspective question answering. In AAAI Spring Symposium on New Directions in



- Question Answering, 2003, pp.20-27.
- [9] Vasileios Hatzivassiloglou, Kathleen R. McKeown. Predicting the Semantic Orientation of Adjectives. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL, 1997, pp.174-181.
- [10] General Inquirer. <http://wjh.harvard.edu/~inquirer>.
- [11] , . HowNet [J]. , 2006, Vol.20, No.1: 14-20.
- [12] , , , . . , 2007: 17-23.
- [13] , . [A]. - [C], 2007: 221-225.
- [14] Janyce Wiebe, Rebecca Brucey, Matthew Bell, et al. A Corpus Study of Evaluative and Speculative Language. In Proceedings of the Second SIGdial Workshop on Discourse and Dialogue, 2001, pp.1-10.
- [15] Qi Zhang, Xi-Peng Qiu, Xuan-Jing Hung, Li-De Wu. Learning Semantic Lexicons using Graph Mutual Reinforcement based Bootstrapping. Acta Automatica Sinica, Vol.34 (10), 2008: 1257-1261.
- [16] R. McDonald, K. Hannan, T. Neylon, et al. Structured models for fine-to-coarse sentiment analysis. In Proceedings of the 45th Association of Computational Linguistics, 2007, pp.435-439.
- [17] . PRIS . . 2008: 46-55.
- [18] T. Jaynes. Information Theory and Statistical Mechanics. Physics Reviews. 1957(106): 620-630
- [19] Adam L. Berget, Stephen A. Della Pietra, Vincent J.Della Pietra. A Maximum Entropy Approach Natural Language Processing. Computational Linguistics, Vol. 22, No. 1. (1996), pp. 39-71.
- [20] Sogou Labs. <http://www.sogou.com/labs/>.
- [21] Myspace.cn. <http://www.myspace.cn/>.
- [22] Lun-Wei Ku, Yu-Ting Liang, Hsin-Hsi Chen. Opinion Extraction, Summarization and Tracking in News and Blog Corpora. In Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs, 2006, pp.100-107.