

基于最大熵模型和最小割模型的中文词与句褒贬极性分析*

dongxishuang@gmail.com zouqibo2009@163.com guanyi@hit.edu.cn hngaoxiang@gmail.com
mingitouch@gmail.com

摘 要

关键词

Positive and Negative Polarity Analysis on Chinese Words and Sentences Based on Maximum Entropy Model and Min-Cut Model

Dong Xi-Shuang, Zou Qi-Bo, Guan Yi, Gao Xiang, Yan Ming

Harbin Institute of Technology Harbin 150001

dongxishuang@gmail.com zouqibo2009@163.com guanyi@hit.edu.cn hngaoxiang@gmail.com
mingitouch@gmail.com

Ab In this paper, Maximum Entropy Model and Min-Cut Model were adopted to predict the positive and negative polarities of Chinese words and sentences. First, we built a domain sentiment lexicon. Then, the candidate sentiment words were recognized by the lexicon, and the polarity was predicted by Maximum Entropy model. Finally, we used Min-Cut model to optimize the polarity results. On the side of sentiment analysis on sentences, opinion sentences were obtained by the lexicon. These opinion sentences were split up into short sentences, and sentiment features were extracted by rules. Then the polarity of short sentences was predicted by Maximum Entropy model. Finally, polarities of opinion sentences were predicted according to polarities of short sentences.

摘 要

关键词

1 引言

*

2 相关研究

WordNet Kamps, Marx, Mokken Rijke

NEAR Peter D. Turney Michael L. Littman
Yu Hatzivassiloglou

Soo-Min Kim Eduard Hovy

Hong Yu Vasileios Hatzivassiloglou

3 相关理论和模型

3.1

[11] $G = (V, E)$ (S, T) v s
 $T = V - S$ $s \in S, t \in T$ (S, T)
 $f(S, T)$ (S, T) $c(S, T)$ [11]

$G = (V, E)$ $c(S, T)$ $|f|$ [11]
[11] 1965 Ford Fulkerson Ford-Fulkerson [11]

3.2

4 词句情感分析

4.1

4.1.1 领域情感词典构建

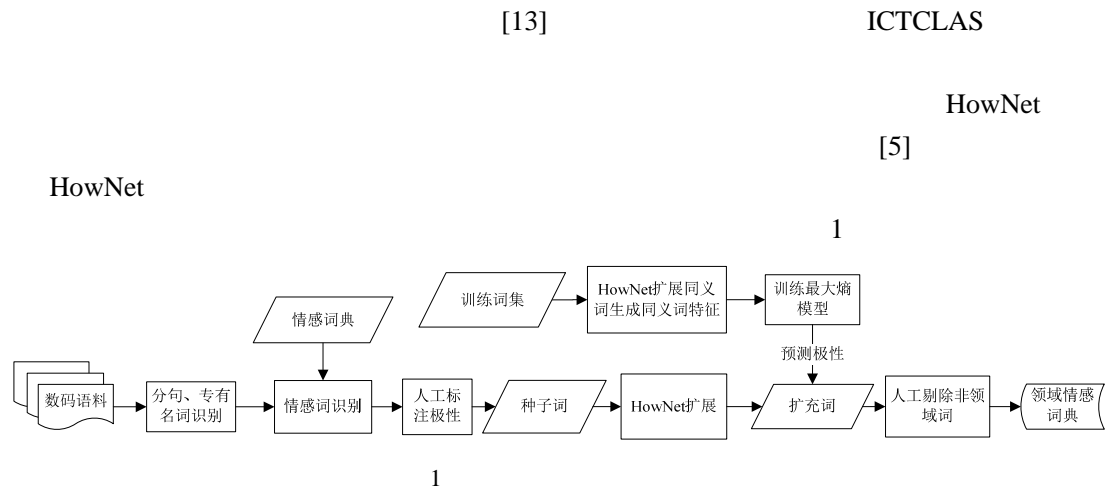


图 4.1.1.1 the Flow Chart of Building Domain Sentiment Lexicon

4.1.2 观点词抽取

ICTCLAS

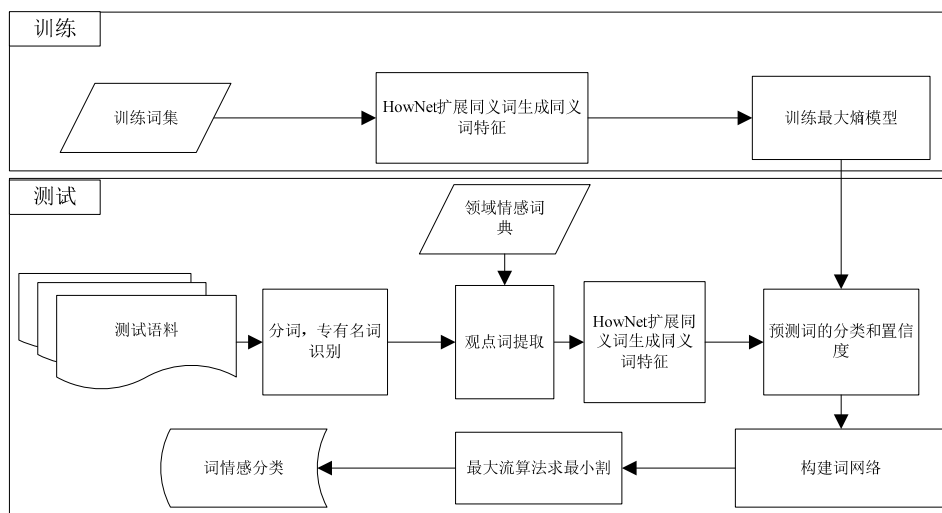
” ” { } []

[18]

4.1.3 观点词情感分类

HowNet

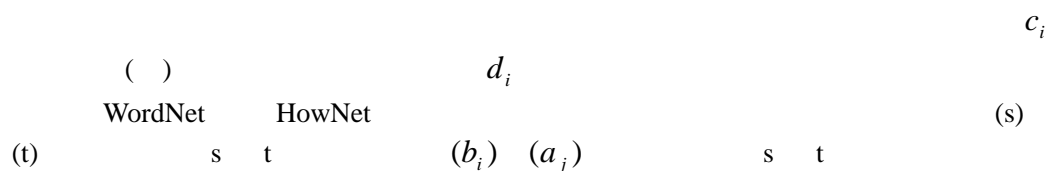
2



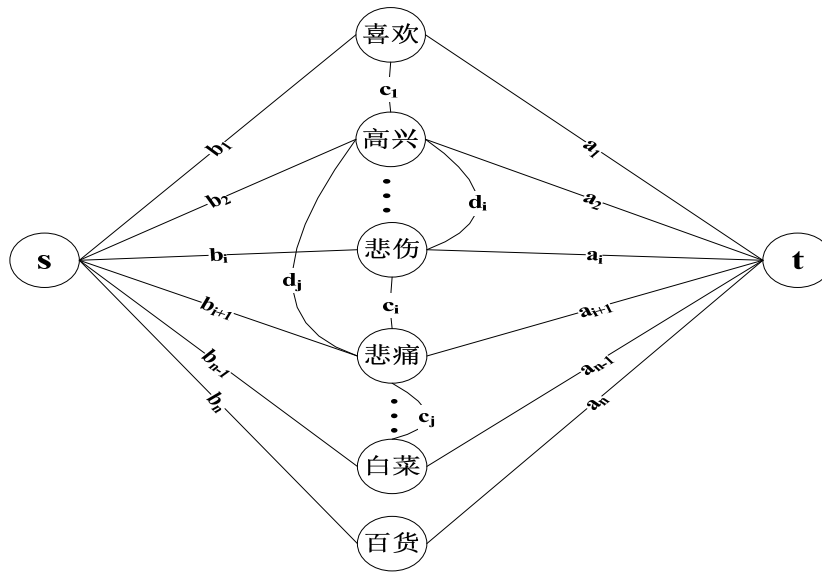
2

the Flow Chart of Classification of Sentiment word

4.1.4



3



BA

b_i a_j

d_i c_i

HowNet

4.1.5 基于最小割模型的词褒贬极性分析

[12]

S T $s \in S, t \in T$

$$W(S, T) = \sum_{u \in S, v \in T} w(u, v)$$

1

$w(u, v)$ u v

$\arg \min W(S, T)$

$W(S, T)$

s t

3 4 4

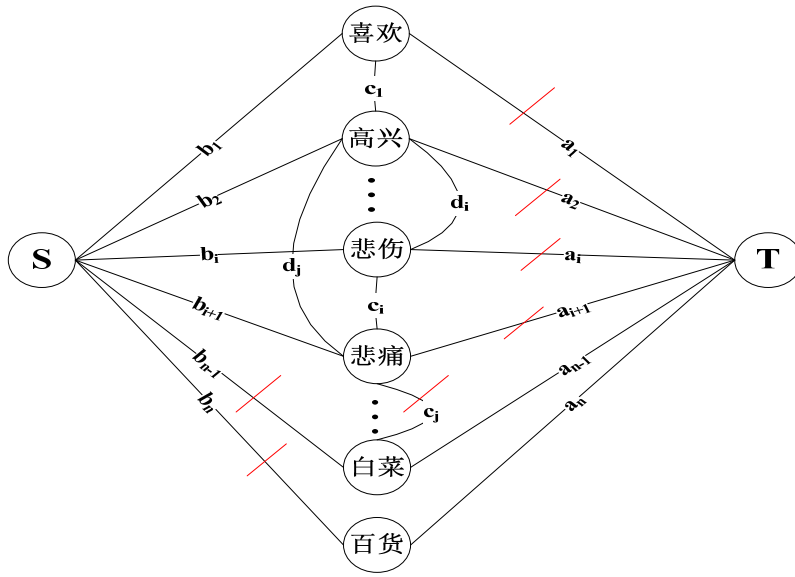


图 4.2.1

4.2

2000

[13]

[3]

2000

4.2.1 观点句识别

”” { }

4.2.2 长句切分

A
B

C
D

4.2.3 特征获取

[5] 3

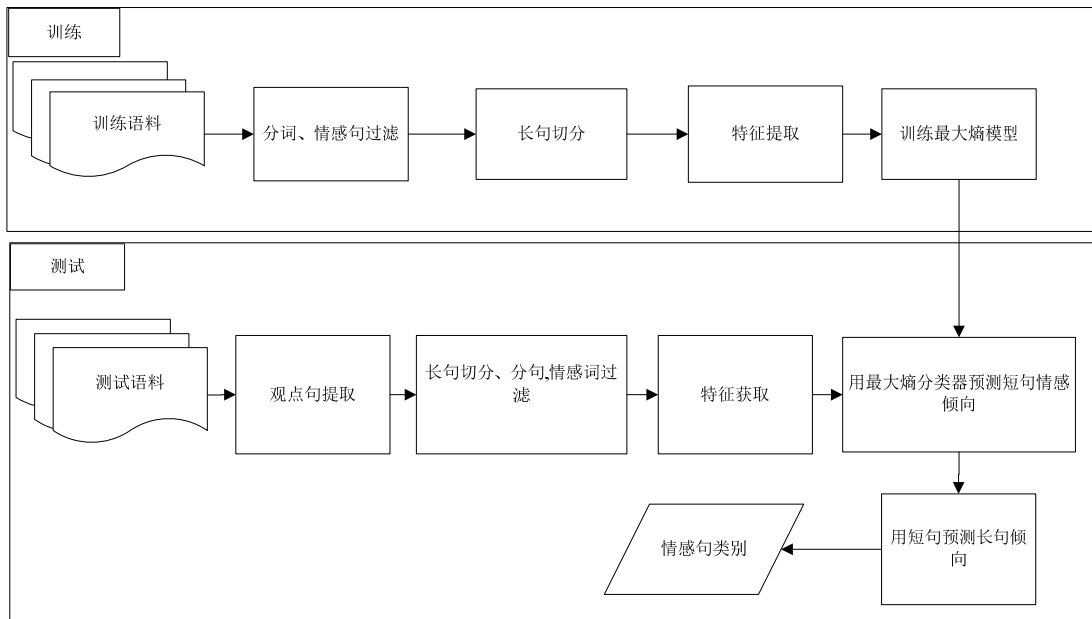
COAE2009

4.2.4 长句情感倾向性计算

1
2
3

$$p = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n}$$

5



5

the Flow Chart of Positive and Negative Polarities Analysis on Sentiment Sentences

5 实验与分析

Tab.1 Positive and Negative Polarities Analysis on Words

Tab.2 Positive and Negative Polarities Analysis on Sentences

6 结果与展望

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.
- 2009 1-8
- [6] J. Kamps, M. Marx, R. J. Mokken and M. D. Rijke. Using WordNet to measure semantic orientation of adjectives. In: Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation, Lisbon, 2004, 1115-1118.
- 2006,20(1):
- 14-20.
- ,2007,17-23.
- [9] Peter D. Turney and Michael L. Littman. Measuring praise and criticism: Inference of semantic orientation from association [J]. ACM Transactions on Information Systems. 2003. 21 (4): 315 – 346.
- [10] H. Yu and V. Hatzivassiloglou. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences [A], In: M. Collins and M. Steedman(eds). Sapporo. Japan: 2003.129-136
- [11] Thomas H. Cormen Charles E. Ierserson Ronald L. Rivest Clifford Stein. Introduction of Algorithms. Second Edition. China Machine Press.2006.396-419
- Fangzhong Su; Katja Markert. Subjectivity Recognition on Word Senses via Semi-supervised Mincuts. Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL, pages 1–9, Boulder, Colorado, June 2009.
- <http://www.searchforum.org.cn/tansongbo/corpus-senti.htm>
- [14] 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.
- [15] T. Jaynes. Information Theory and Statistical Mechanics. Physics Reviews. 1957(106): 620-630
- [16] Adam L. Berger, 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.
- [18] <http://list.video.baidu.com/manhotlist/>