

情感分析模型综述*

董喜双,邹启波,关毅*

150001

摘要: 对情感分析模型的研究现状与进展进行了总结,包括三个部分:首先介绍情感分析相关概念,包括情感分析形式化定义以及情感极性和情感强度;然后将情感分析归纳为3个主要层级模型,即词级情感分析模型、句级情感分析模型和篇章级情感分析模型。针对不同层级,重点介绍基于无监督、有监督和半监督学习方法的情感分析模型,并对模型的优缺点进行了详细总结;最后介绍了情感分析模型在不同领域的应用,包括问答系统、产品推荐和舆情分析等。

关键词: 情感分析;情感极性;情感强度;机器学习

A Survey on Sentiment Analysis Models

Xishuang Dong, Qibo Zou, Yi Guan

Harbin Institute of Technology, Harbin, China, 150001

Email: dongxishuang@gmail.com; zouqibo2009@163.com; guanyi@hit.edu.cn

Abstract This paper presents an overview of sentiment analysis models. It includes three parts: related concepts including the definition of sentiment analysis and introduction on sentiment polarity and strength; a summary of Word-level Sentiment Analysis Models, Sentence-level Sentiment Analysis Models, and Document-level Sentiment Analysis Models, which contain unsupervised based models, supervised based models, semi-supervised based models and their advantages and disadvantages; model applications such as the question and answer system, the production recommendation system, and the public opinion analysis system, respectively.

Key words Sentiment Analysis; Sentiment Polarity; Sentiment Strength; Machine Learning

引言

情感分析是自然语言处理(NLP)的一个重要分支,旨在从文本中识别和提取情感信息。随着互联网的普及和社交媒体的兴起,情感分析在舆情分析、产品推荐、客户服务等领域得到了广泛的应用。本文旨在对情感分析模型的研究现状与进展进行综述,包括情感分析的形式化定义、情感极性和情感强度的定义,以及情感分析模型的分类和评价。本文将情感分析模型分为词级、句级和篇章级三个层级,分别介绍其模型结构、优缺点和应用。最后,我们将介绍情感分析模型在不同领域的应用,包括问答系统、产品推荐和舆情分析等。

情感分析的研究始于20世纪60年代,最初的研究主要集中在基于规则的方法上。随着机器学习技术的发展,基于统计的方法和基于神经网络的方法逐渐成为情感分析的主流。近年来,深度学习在情感分析领域取得了显著进展,特别是在处理长文本和上下文信息方面。然而,情感分析仍然面临着许多挑战,如情感极性模糊、情感强度量化、跨领域情感迁移等问题。本文将对情感分析模型的研究现状与进展进行总结,并对未来的研究方向进行展望。

本文的结构如下:首先介绍情感分析相关概念,包括情感分析形式化定义以及情感极性和情感强度;然后将情感分析归纳为3个主要层级模型,即词级情感分析模型、句级情感分析模型和篇章级情感分析模型。针对不同层级,重点介绍基于无监督、有监督和半监督学习方法的情感分析模型,并对模型的优缺点进行了详细总结;最后介绍了情感分析模型在不同领域的应用,包括问答系统、产品推荐和舆情分析等。

关键词: 情感分析;情感极性;情感强度;机器学习

*基金项目: 国家自然科学基金“面向语句间相似度计算基于词主体自治学习的强化学习机制研究”,项目编号: 60975077; 国家自然科学基金“非常规突发事件网络舆情分析方法和预警机制研究”,项目编号 90924015。

《 且 (Support Vector Machine SVM) 《 (Maximum Entropy ME) 《 (Conditional Random Fields CRFs) [2] 《

专 《 专 《 专 《 互 乏 《 专 于 二 《

1 相关概念

D 《 $C = \{c_1, c_2, c_3 \dots c_n\}$ 1 $f : D \rightarrow C$ 1 Sentiment Labels n f

余 个 乏 (Positive) 《 乏 (Negative) (Neutral) 余 二 二 乏

2 情感分析模型

2.1 词级情感分析模型

借 乏 借 乏 《 乏 《 乏 借 专 借 克

2.1.1 基于无监督的词级情感分析模型

《 [5] 个 乏 借 且 乏 互 乏 借 乏

乏 借 专 乏
 Y. Wu M. We^[6]
 Pointwise Mutual Information PMI
 且 PMI 且 PMI 二
 专 专 A. Hassan D.
 Radev^[7] 二 乏 借
 借 二 乏
 业 乏 WordNet 借

2.1.2 基于有监督的词级情感分析模型

1

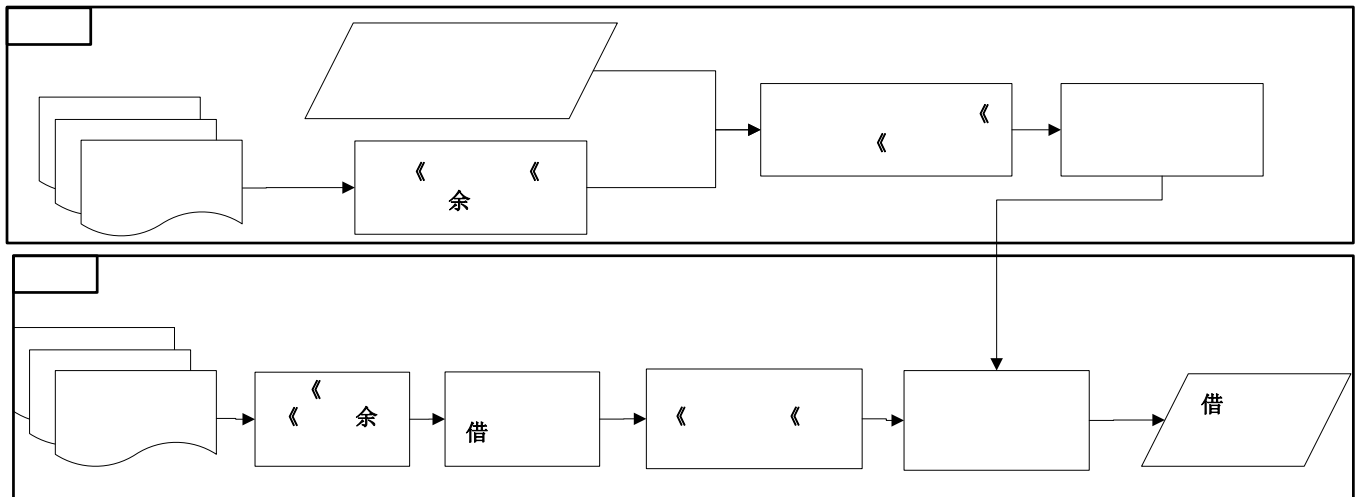


图 1 识别情感词流程图

借 《 余 借 借
 [8] HowNet 乏
 专 D. Das S. Bandyopadhyay^[9] CRFs
 《 《 《 《 《

2.1.3 基于半监督的词级情感分析模型

D. Rao D. Ravichandran^[10] 二 《
 [11]《 (Label Propagation) [12]
 二 乏 体 二 二

2.2 句级情感分析模型

借

《

借

2.2.1 基于无监督的句级情感分析模型

G. Fu X. Wang^[13]

Fine-to-coarse Strategy

余

于

《

A. Meena T. Prabhakar G. Amati^[14]

传 专 专

二

业

专

2.2.2 基于有监督的句级情感分析模型

2

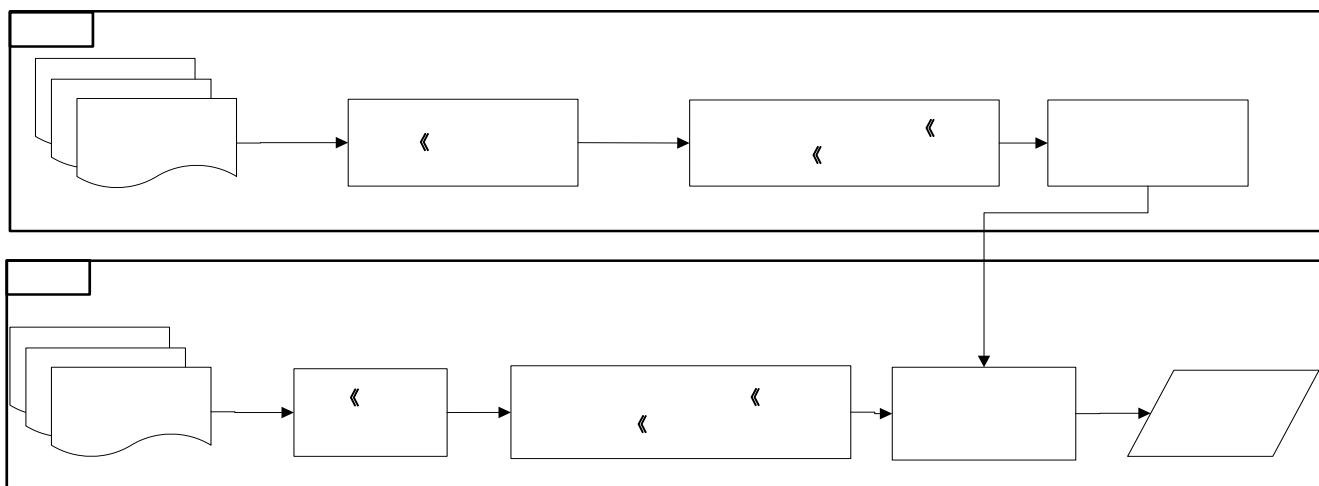


图2 情感句分析流程图

《 《 余

H. Guo H. Zhu Z. Guo Z. Su^[15]

1

2

3

4

《 《 余

《 《 《

R. Xia C. Zong^[16]

(Fast Mutual Information, FMI)

(Fast Information Gain FIG)

$$GS(ws) \approx GS(w) + GS(s)$$

$$I(ws, c) \approx I(w, c) + I(s, c)$$

B. Wei C. Pal^[17]

Structural Correspondence Learning SCL

业

D. Davidov O. Tsur A. Rappoport^[18] K-

k-Nearest Neighbors (kNN)

《 n

《

W. Wei J. A. Gulla^[19]

二 余

余

二

H. Jang H. Shin^[20]

乏 《

乏 Q. Zhang Y. Wu T. Li ^[21]

《

《

《

SVM

二

二

[22]

《

3

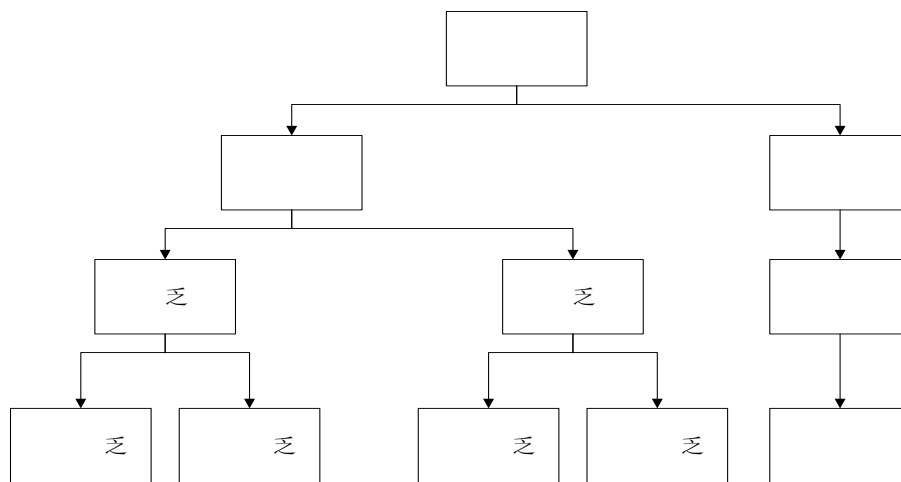


图 3 层级模型

传

2.2.3 基于半监督的句级情感分析模型

R. Socher J. Pennington E. H. Huang ^[23]

个个

专

专 互

业

《

2.3 篇章级情感分析模型

且

《

专

2.3.1 基于无监督的篇章级情感分析模型

M. Taboada J. Brooke M. Tofiloski ^[24]

二

业

专

A. L Maas R. E Daly P. T Pham ^[25]

乏

(专 互)

《 《 专 专传

2.3.2 基于有监督的篇章级情感分析模型

S. Tan Y. Wang^[26] 《 二 业 6 专

Y. Bengio^[27] 《 X. Gloro A. Bordes 22 二 Q. Wu S. Tan X. Cheng^[28] 《

《 《 《 《 《 专 《

A. Abbasi S. France Z. Zhang^[29] n 《 专 二 乏 业 二 n 3

A. Agarwal B. Xie I. Vovsha^[30] Twitter A. Balamurali 二 Tree Kernel A. Joshi P. Bhattacharyya^[31] WordNet 乏 乏 SVM

WordNet WordNet S. Li S. Y. M. Lee Y. Chen^[32] SVM 专 个 专 SVM 个 G. Paltoglou M. Thelwall^[33] SVM tf-idf S. Li R. Xia C. Zong^[34]

Wang X. Cheng^[35] 专 A. Abbasi H. Chen A. Salem^[36] SVM

SVM R. Mcdonald K. Hannan T. Neylon^[37] 二 专 个

且 《 [38] 乏 SVM S. Wang Y. Wei D. Li^[39]

SVM

X. Wan^[40]

《 个 个 业
专 业

2.3.3 基于半监督的篇章级情感分析模型

C. Tan L. Lee J. Tang^[41] 传

Twitter

(1) 专 传 SVM 业 (2) S. Li Z. Wang G. Zhou^[42]
专

S. Li C. Huang G. Zhou^[43]

二

个 专 SVM f_1 f_2 D SVM f_3 个 (Fixed Rule)
(Trained Rule) $f_1 \ll f_2 f_3$

3 情感分析模型应用

专 《

3.1 问答系统

M. Mohtarami H. Amiri M. Lan^[44] 专 二 乏
乏 乏 乏 余 (1) (3)

(2)

(4) 乏

(1)

(2) 乏 且

且

乏

且

乏

WordNet

(3)

专

(4) 乏

且

且

乏

3.2 产品推荐

Bing Qin Yanyan Zhao Leilei Gao^[45]

个 (1)

(2)

乏

50%

3.3 舆情分析

3.4 其他

《 《 二 QQ《 《 《

BBS

传

4 结论

且 二 , 二 《 乏 乏 二 专

参考文献

[1]2009.
[2] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," in Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, 2004, pp. 271-278.
[3] S. Kim and E. Hovy, "Determining the Sentiment of Opinions," In Proceedings of COLING, 2004, pp.1367-1373.
[4] Y. Mao and G. Lebanon, "Isotonic Conditional Random Fields and Local Sentiment Flow," In Proceedings of NIPS, 2006. pp: 961-968.
[5] , , . 乏2011.
[6] Y. Wu and M. Wen, "Disambiguating Dynamic Sentiment Ambiguous Adjectives," Proceeding COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics, Aug. 2010, pp. 1191-1199.
[7] A. Hassan and D. Radev, "Identifying Text Polarity Using Random Walks," Proceeding ACL '10 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Jul. 2010, pp. 395-403.
[8] , , . 且2011.
[9] D. Das and S. Bandyopadhyay, "Word to Sentence Level Emotion Tagging for Bengali Blogs," Proceeding ACL Short '09 Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Aug, 2009, pp. 149-152.
[10] D. Rao and D. Ravichandran, "Semi-Supervised Polarity Lexicon Induction," Proceedings of the 12th Conference of the European Chapter of the ACL, 2009, pp. 675-682.
[11] Blum, Lafferty, Rwebangira, et al. "Semi-supervised Learning Using Randomized Mincuts," In Proceedings of the ICML. 2004.
[12] X. Zhu and Z. Ghahramani, "Learning from Labeled and Unlabeled Data with Label Propagation," Technical Report CMU-CALD-02-107, Carnegie Mellon University. 2002.
[13] G. Fu and X. Wang, "Chinese Sentence-level Sentiment Classification based on Fuzzy Sets," Proceeding COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics, Aug, 2010, pp. 312-319.
[14] A. Meena, T. Prabhakar, G. Amati et al. "Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis," Advances in Information Retrieval, vol. 4425. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 573-580.
[15] H. Guo, H. Zhu, Z. Guo, and Z. Su, "Domain Customization for Aspect-oriented Opinion Analysis with Multi-level Latent Sentiment

- Clues,” in Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 2493-2496.
- [16] R. Xia and C. Zong, “Exploring the Use of Word Relation Features for Sentiment Classification,” Proceeding COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics, Aug, 2010, pp. 1336-1344.
- [17] B. Wei and C. Pal, “Cross Lingual Adaptation: An Experiment on Sentiment Classifications,” Proceeding ACL Short '10 Proceedings of the ACL 2010 Conference Short Papers, Jul, 2010, pp. 258-262.
- [18] D. Davidov, O. Tsur, and A. Rappoport. “Enhanced Sentiment Learning Using Twitter Hashtags and Smileys,” Proceedings of the 23rd International Conference on Computational Linguistics, 2010, pp.241-249.
- [19] W. Wei and J. A. Gulla, “Sentiment Learning on Product Reviews via Sentiment Ontology Tree,” Proceeding ACL '10 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Jul, 2010, pp. 404-413.
- [20] H. Jang and H. Shin, “Language-specific Sentiment Analysis in Morphologically Rich Languages,” Proceeding COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics, Aug, 2010, pp. 498-506.
- [21] Q. Zhang, Y. Wu, T. Li, et al, “Mining Product Reviews based on Shallow Dependency Parsing,” in Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, 2009, pp. 726-727.
- [22] , . CRF . 传 . 2007.
- [23] R. Socher, J. Pennington, E. H. Huang, et al, “Semi-supervised Recursive Auto-encoders for Predicting Sentiment Distributions,” Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Jul, 2011, pp. 151-161.
- [24] M. Taboada, J. Brooke, M. Tofiloski, et al, “Lexicon-Based Methods for Sentiment Analysis,” Computational Linguistics, vol. 37, no. 2, Jun, 2011, pp. 267-307.
- [25] A. L. Maas, R. E. Daly, P. T. Pham, et al, “Learning Word Vectors for Sentiment Analysis,” Computational Linguistics, vol. 31, no. 2, 2011, pp. 142-150.
- [26] S. Tan and Y. Wang, “Weighted SCL Model for Adaptation of Sentiment Classification,” Expert Systems with Applications, vol. 38, no. 8, Aug, 2011, pp. 10524-10531.
- [27] X. Glorot, A. Bordes, and Y. Bengio, “Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach,” Proceedings of the 28th International Conference on Machine Learning, 2011, pp.513-520.
- [28] Q. Wu, S. Tan, X. Cheng, et al, “MIEA: a Mutual Iterative Enhancement Approach for Cross-Domain Sentiment Classification,” Coling 2010 Posters, no. Aug, 2010, pp. 1327-1335.
- [29] A. Abbasi, S. France, Z. Zhang, et al, “Selecting Attributes for Sentiment Classification Using Feature Relation Networks,” IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 3, Mar, 2011, pp. 447-462.
- [30] A. Agarwal, B. Xie, I. Vovsha, et al, “Sentiment Analysis of Twitter Data,” Proceeding LSM '11 Proceedings of the Workshop on Languages in Social Media, Jun, 2011, pp. 30-38.
- [31] A. Balamurali, A. Joshi, and P. Bhattacharyya, “Harnessing WordNet Senses for Supervised Sentiment Classification,” EMNLPACL, 2011, pp. 1081-1091.
- [32] S. Li, S. Y. M. Lee, Y. Chen, et al, “Sentiment Classification and Polarity Shifting,” Proceeding COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics, Aug, 2010, pp. 635-643.
- [33] G. Paltoglou and M. Thelwall, “A Study of Information Retrieval Weighting Schemes for Sentiment Analysis,” Proceeding ACL '10 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Jul, 2010, pp. 1386-1395.
- [34] S. Li, R. Xia, C. Zong, et al, “A Framework of Feature Selection Methods for Text Categorization,” Proceeding ACL '09 Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Aug, 2009, pp. 692-700.
- [35] S. Tan, Y. Wang, and X. Cheng, “Combining Learn-based and Lexicon-based Techniques for Sentiment Detection without Using Labeled Examples,” in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, 2008, pp. 743-744.
- [36] A. Abbasi, H. Chen, and A. Salem, “Sentiment Analysis in Multiple Languages,” ACM Transactions on Information Systems, vol. 26, no. 3, Jun, 2008, pp.1-34.
- [37] R. McDonald, K. Hannan, T. Neylon, et al, “Structured Models for Fine-to-Coarse Sentiment Analysis,” Proceedings of the Association for

Computational Linguistics, 2007, pp. 432-439.

[38]2007.

[39] S. Wang, Y. Wei, D. Li, et al, "A Hybrid Method of Feature Selection for Chinese Text Sentiment Classification," in Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2007, pp. 435-439.

[40] X. Wan, "Bilingual Co-Training for Sentiment Classification of Chinese Product Reviews," Computational Linguistics, vol. 37, no. 3, Sep, 2011, pp. 587-616.

[41] C. Tan, L. Lee, J. Tang, et al, "User-level Sentiment Analysis Incorporating Social Networks," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 1397-1405.

[42] S. Li, Z. Wang, G. Zhou, et al, "Semi-Supervised Learning for Imbalanced Sentiment Classification", in Proc. IJCAI, 2011, pp.1826-1831.

[43] S. Li, C.-R. Huang, G. Zhou, et al, "Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification," Proceeding ACL '10 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Jul, 2010 pp. 414-423.

[44] M. Mohtarami, H. Amiri, M. Lan, et al, "Predicting the Uncertainty of Sentiment Adjectives in Indirect Answers," in Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 2485-2488.

[45] B. Qin, Y. Zhao, L. Gao, et al, "Recommended or Not? Give advice on Online Product," Fifth international conference on Fuzzy Systems and Knowledge Discovery, 2008, pp208-212.

[46] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in Twitter Events," Journal of the American Society for Information Science and Technology, vol. 62, no. 2, Feb. 2011, pp. 406-418.



1981-
且 2005 体 2005 2001 2005 2008 且
2008
体 BBS
且
dongxishuang@gmail.com



1993-
, 2010 2009
2011 2010
(COAE2011) (ACM/ICPC) 互 2011
2011
且 zouqibo2009@163.com



1970- 1988 1992
且 1992 6 体 1992 1995
体 1995 1995 1999
体 二 且
BOPOMOFO 且 1996 3 1997
3 1999 3
体 guanyi@hit.edu.cn