学校编码：10384
学号：31520131153279

厦门大学

硕 士 学 位 论 文

面向社交媒体内容的多模态情感特征学习研究

Research on multi-modal sentiment feature learning of social media content

李凌霄

指导教师：李绍滋

专业名称：模式识别与智能系统

答辩日期：2016年5月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外，该学位论文为(   )课题(组)的研究成果，获得(   )课题(组)经费或实验室的资助，在(   )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于:

(　　)1.经厦门大学保密委员会审查核定的保密学位论文，于　年　月　日解密，解密后适用上述授权。

(　　)2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名):

年　　月　　日

摘 要

社交媒体已成为现代社会舆论交流和信息传递的主要平台。针对社交媒体的情感分析对于舆论监控、商业产品导向和股市预测等都具有重大应用价值。但社交媒体内容的多模态性（文本、图片等）让传统的单模态情感分析方法面临许多局限，多模态情感分析技术对跨媒体内容的理解与分析具有重大的理论价值。

多模态情感分析区别于单模态方法的关键问题在于，如何综合利用形态各异的多模态情感信息，来获取整体的情感倾向性，同时考虑单个模态本身在情感表达上的性质。针对该问题，利用社交媒体上的多模态内容在情感表达上所具有的关联性、抽象层级性的特点，提出了一套面向社交媒体的多模态情感特征学习与融合方法，实现多模态情感分析，主要内容和创新点如下:

1.针对社交媒体多模态信息在情感表达上具有关联性的问题，提出了一种基于贝叶斯网络的跨模态情感关联模型（Multi-modal Correlation Model， MCM），来对社交媒体内容的情感表达进行建模。利用网络中不同模态间的先验和后验概率，综合考虑了1）各模态自身对情感的贡献；2）不同模态在情感表达上的相关性。实现了对文本、图片和表情符号等不同模态情感特征的跨模态融合。实验结果验证了模态关联性对多模态情感分类具有积极效果。

2.针对模态和情感间存在的语义鸿沟问题，结合深度学习算法和人类的先验知识，提出了一种层级性的情感特征学习策略（Hierarchical Fine-tune Learning Strategy，HFLS）。从特征学习数据和学习模型的层次结构两方面入手，利用微调（fine-tune）机制，逐层地学习模态的低层特征、中层情感特征和高层情感特征。实验结果说明，相比于现有的情感特征学习方法，HFLS能够很好地减小模态与情感间的语义鸿沟。

3.针对多模态在情感表达上的关联性和层级性问题，结合MCM的分析结论和HFLS学习策略，提出了一种通用的多模态情感特征学习方法。利用社交媒体中存在的大规模情感信号，以无监督的方式学习具有强情感关联性，和高度情感抽象层级的多模态情感特征。实验结果说明了该方法产生的情感特征优于现有情感特征学习方法、并且具有很好的泛化能力。

关键词：多模态关联；层级特征；情感信号

Abstract

Social media has become a main platform of public communication and information transmission. Therefore, social media sentiment analysis has great application values in many fields, such as public opinion monitoring, production marking, stock forecasting and so on. But the multi-modal characteristic of social media content (e.g. texts and images) significantly challenges traditional text-based sentiment analysis approaches, multi-modal sentiment analysis gets great theoretical value for understanding and analysis of multi-modal contents.

The key challenge of multi-modal sentiment analysis is how to recognize the integral sentiment with diverse modalities, simultaneously consider the sentiment presen- tation of each modal. In this paper, we propose a set of novel multi-modal sentiment analysis methods for scial media content, which synthetically utilize the correlation and hierarchical characteristics of multi-modal sentiment expressions. The main contents and innovations are as follows:

1.To solve the problems of sentiment correlations among different modalities, we propose a Multi-modal Correlation Model (MCM). Compared with other multi-modal methods, MCM models hierarchical correlations among modalities, as well as between modalities and sentiments. Specifically, a probabilistic graph ical model is subsequently built upon the proposed MCM model, which considers the hierarchical correlations and preserves the classification ability of each modality. Experimental results demonstrate the importance of mulit-modal hierarchical correlations to sentiment analysis

2.To solve the problem of the semantic gap between modalities and sentiment, we propose a Hierarchical Fine-tune Learning Strategy (HFLS), which takes advantages of deep learning technologies and human experiences to learn hierarchical sentiment features. HFLS employs layer-wise fine-tunes on deep network with stacked datasets, and successively learns low-level, mid-level and

high-level sentiment features for sentiment analysis. Experimental results show the hierarchical features can improve sentiment classification and bridge the semantic gap.

3.To solve the problem of the correlations and hierarchy of multi-modal sentiment expressions, we propose a general multi-modal sentiment feature leraning method based on the conclusions of MCM and HFLS. The method is unsupervised, which use the large scale sentiment signals in social media as hint informations to learn high sentiment-correlated and hierarchical sentiment features. Experi- mental results show the features outperform the state-of-the-art methods and have good generalization ability.

Keywords: Multi-modal Correlation;Hierarchical Feature Learning;Sentiment Signal

参考资料

[1]R. W. Picard. Affective computing[M]. Cambridge: MIT press, 1997.
[2]S. Das and M. Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards[C]. Proceedings of the Asia Pacific Finance Association Annual Conference (APFA), 2001.
[3]R. M. Tong. An operational system for detecting and tracking opinions in on-line discussion. Proceedings of the Workshop on Operational Text Classification (OTC), 2001, 35: 43.
[4]C. Cardie, J. Wiebe, T. Wilson, and D. Litman. Combining low-level and summary representations of opinions for multi-perspective question answering[C]. Proceedings of the AAAI Spring Symposium on New Directions in Question Answering, 2003: 20-27.
[5]K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews[C]. Proceedings of WWW, 2003: 519-528.
[6]L. Dini and G. Mazzini. Opinion classification through information extraction[C]. Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields (Data Mining), 2002: 299-310.
[7]H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge[C]. Proceedings of Intelligent User Interfaces (IUI), 2003: 125-132.
[8]S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the Web[C]. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2002: 341-349.
[9]T. Nasukawa and J. Yi. Sentiment analysis: Capturing favorability using natural language processing[C]. Proceedings of the Conference on Knowledge Capture (K-CAP), 2003.
[10]B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002: 79-86.
[11]K. Tateishi, Y. Ishiguro, and T. Fukushima. Opinion information retrieval from the internet[C]. Information Processing Society of Japan (IPSJ) SIG Notes, 2001, vol. 69(7): 75-82.
[12]B. O'Connor, R. Balasubramanyan, B R. Routledge, N A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series[J]. ICWSM, 2010, 11(122-129): 1-2.
[13]P. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews[C]. Proceedings of the Association for Computational Linguistics (ACL), 2002: 417-424.
[14]J. Wiebe, E. Breck, C. Buckley, C. Cardie, P. Davis, B. Fraser, D. Litman, D. Pierce, E. Riloff, T. Wilson, D. Day, and M. Maybury. Recognizing and organizing opinions expressed in the world press[C]. Proceedings of the AAAI Spring Symposium on New Directions in Question Answering, 2003.
[15]H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.
[16]JC. Na, H. Sui, C. Khoo, S. Chan, Y. Zhou. Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews[J]. Advances in Knowledge Organization, 2004, 9: 49-54.
[17]E. M. Airoldi, X. Bai, and R. Padman. Markov blankets and meta-heuristic search: Sentiment extraction from unstructured text[M]. Advances in Web Mining and Web Usage Analysis. Springer Berlin Heidelberg, 2004: 167-187.
[18]M. Gamon. Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis[C]. Proceedings of the International Conference on Computational Linguistics (COLING), 2004.
[19]S. Matsumoto, H. Takamura, and M. Okumura. Sentiment classification using word sub-sequences and dependency sub-trees[C]. Proceedings of PAKDD ’05, the 9th Pacific-Asia Conference on Advances in

Knowledge Discovery and Data Mining, 2005.
[20]T. Mullen and N. Collier. Sentiment analysis using support vector machines with diverse information sources[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2004: 412-418.
[21]A. B. Goldberg and X. Zhu. Seeing stars when there aren ' t many stars: Graph-based semi-supervised learning for sentiment categorization[C]. TextGraphs: HLT/NAACL Workshop on Graph-based Algorithms for Natural Language Processing, 2006.
[22]Y. Mao and G. Lebanon. Isotonic conditional random fields and local sentiment flow[C]. Advances in Neural Information Processing Systems, 2007.
[23]B. Pang and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales[C]. Proceedings of the Association for Computational Linguistics (ACL), 2005: 115-124.
[24]T. Wilson, J. Wiebe, and R. Hwa. Just how mad are you? Finding strong and weak opinion clauses[C]. Proceedings of AAAI, 2004: 761-769.
[25]Pang B, Lee L. Opinion Mining and Sentiment Analysis[J]. Foundations and trends in information retrieval, 2008, 2(1-2): 1-135.
[26]V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity[C]. Proceedings of the International Conference on Computational Linguistics (COLING), 2000.
[27]E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.
[28]N. Kaji and M. Kitsuregawa. Automatic construction of polarity-tagged corpus from HTML documents[C]. Proceedings of the COLING/ACL Main Conference Poster Sessions, 2006.
[29]J. M. Wiebe and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts [M]. Computational Linguistics and Intelligent Text Processing. Springer Berlin Heidelberg, 2005: 486-497.
[30]E. Riloff, J. Wiebe, and T. Wilson. Learning subjective nouns using extraction pattern bootstrapping[C]. Proceedings of the Conference on Natural Language Learning (CoNLL), 2003: 25-32.
[31]T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis[C]. Proceedings of HLT and EMNLP , 2005
[32]A. Hanjalic. Extracting moods from pictures and sounds: towards truly personalized TV [J]. Signal processing magazine, IEEE, 2006, 23(2): 90-100.
[33]W. Wang and Q. He. A survey on emotional semantic image retrieval. ICIP. 2008: 117-120.
[34]N. Bianchi-Berthouze. K-dime: an affective image filtering system [J]. MultiMedia, IEEE, 2003, 10(3): 103-106.
[35]R. Datta, D. Joshi, J. Li, and J. Wang. Studying Aesthetics in Photographic Images using a Computational Approach [M]. Computer Vision – ECCV 2006. Springer Berlin Heidelberg, 2006: 288-301.
[36]L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the Aesthetic Quality of Photographs using Generic Image Descriptors[C]. ICCV, 2011.
[37]P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an Image Memorable[C]? CVPR, 2011: 145-152.
[38]V. Vonikakis and S. Winkler. Emotion-based Sequence of Family Photos. ACM MM, 2012: 1371-1372.
[39]J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. Can we understand van Gogh's Mood?: Learning to infer Affects from Images in Social Networks[C]. ACM MM, 2012: 857-860.
[40]J. Machajdik and A. Hanbury. Affective Image Classification using Features inspired by Psychology and Art Theory[C]. ACM MM, 2010: 83-92.
[41]V. Yanulevskaya et al. Emotional Valence Categorization using Holistic Image Features[C]. ICIP, 2008: 101-104.
[42]V. Yanulevskaya et al. In the Eye of the Beholder: Employing Statistical Analysis and Eye Tracking for Analyzing Abstract Paintings[C]. ACM MM, 2012.
[43]B Li, W Xiong, W Hu, and X Ding. Context-aware affective images classification based on bilayer sparse representation[C]. ACM MM, 2012: 721-724.

[44]P Lee, Y Teng, and TC Hsiao. Xcsf for prediction on emotion induced by image based on dimensional theory of emotion[C]. ACM GECCO, 2012: 375-382.
[45]Borth, D., Ji, R., Chen, T., Breuel, T., & Chang, S. F. Large-scale visual sentiment ontology and detectors using adjective noun pairs[C]. ACM MM, 2013: 223-232.
[46]Yuan J, Mcdonough S, You Q, et al. Sentribute: image sentiment analysis from a mid-level perspective[C]. Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining. ACM, 2013: 10.
[47]D Cao, R Ji, D Lin, and S Li. Visual sentiment topic model based micro-blog image sentiment analysis[J]. Multimedia Tools and Applications, 2014: 1-14.
[48]T Chen, D Borth, T Darrell, and SF Chang. Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks [J]. arXiv preprint arXiv:1410.8586, 2014.
[49]Q You, J Luo, H Jin, and J Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks[C]. AAAI, 2015.
[50]Plutchik R. Emotion: A psychoevolutionary synthesis[M]. Harpercollins College Division, 1980.
[51]Poria S, Cambria E, Howard N, et al. Fusing audio, visual and textual clues for sentiment analysis from multimodal content[J]. Neurocomputing, 2016, 174: 50-59.
[52]LP Morency, R Mihalcea, and P Doshi. Towards multimodal sentiment analysis: harvesting opinions from the web. ACM ICMI, 2011: 169-176.
[53]M Wang, D Cao, L Li, S Li, and R Ji. Microblog sentiment analysis based on cross-media bag-of-words model[C]. ACM ICIMCS, 2014: 76.
[54]You Q, Luo J, Jin H, et al. Cross-modality Consistent Regression for Joint Visual-Textual Sentiment Analysis of Social Multimedia[C]. Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. ACM, 2016: 13-22.
[55]D Cao, R Ji, D Lin, and S Li. A cross-media public sentiment analysis system for microblog[J]. Multimedia Systems, 2014: 1-8.
[56]Q You and J Luo. Towards social imagematics: sentiment analysis in social multimedia[C]. ACM MDMM, 2013: 3.
[57]Chen C, Chen F, Cao D, et al. A Cross-media Sentiment Analytics Platform For Microblog[C]. ACM MM, 2015: 767-769.
[58]N Srivastava and RR Salakhutdinov. Multimodal learning with deep boltzmann machines[C]. NIPS, 2012: 2222-2230.
[59]Zhou D, Luo J, Silenzio V M B, et al. Tackling Mental Health by Integrating Unobtrusive Multimodal Sensing[C]. AAAI, 2015: 1401-1409.
[60]Ma L, Lu Z, Shang L, et al. Multimodal convolutional neural networks for matching image and sentence[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 2623-2631.
[61]Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3128-3137.
[62]Srivastava N, Salakhutdinov R R. Multimodal learning with deep boltzmann machines[C]. Advances in neural information processing systems. 2012: 2222-2230.
[63]Rasiwasia N, Costa Pereira J, Coviello E, et al. A new approach to cross-modal multimedia retrieval[C]. Proceedings of the 18th ACM international conference on Multimedia. ACM, 2010: 251-260.
[64]黄伯荣, 廖序东. 现代汉语[M]. 5版. 北京: 高等教育出版社, 2011: 356.
[65]Sun W. A stacked sub-word model for joint Chinese word segmentation and part-of-speech tagging[C]. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011: 1385-1394.
[66]Huang C, Zhao H. Chinese word segmentation: A decade review[J]. Journal of Chinese Information Processing, 2007, 21(3): 8-20.
[67]龙树全, 赵正文, 唐华. 中文分词算法概述[J]. 电脑知识与技术: 学术交流, 2009, 5(4): 2605-2607.

[68]Zhang H P, Yu H K, Xiong D Y, et al. HHMM-based Chinese lexical analyzer ICTCLAS[C]. Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17. Association for Computational Linguistics, 2003: 184-187.
[69]徐琳宏,林鸿飞,潘宇,等.情感词汇本体的构造[J]. 情报学报, 2008, 27(2): 180-185.
[70]谢丽星, 周明, 孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取[J]. 中文信息学报, 2012, 26(1): 73-83.
[71]Machajdik J, Hanbury A. Affective image classification using features inspired by psychology and art theory[C]. Proceedings of the international conference on Multimedia. ACM, 2010: 83-92.
[72]Smeulders A W M, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000, 22(12): 1349-1380.
[73]Gudivada V N, Raghavan V V. Content based image retrieval systems[J]. Computer, 1995, 28(9): 18-22.
[74]Zhao R, Grosky W I. Narrowing the semantic gap-improved text-based web document retrieval using visual features[J]. Multimedia, IEEE Transactions on, 2002, 4(2): 189-200.
[75]A. Hanjalic. Extracting moods from pictures and sounds: Towards truly personalized TV[J]. Signal processing magazine, IEEE, 2006, 23(2): 90-100.
[76]Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]. NIPS, 2013: 3111-3119.
[77]Tang D, Wei F, Yang N, et al. Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification[C]. ACL, 2014: 1555-1565.
[78]Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.
[79]Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. CVPR, 2005, 1: 886-893.
[80]Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. International journal of computer vision, 2001, 42(3): 145-175.
[81]Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems, 2012: 1097-1105.
[82]Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(1): 221-231.
[83]Oquab M, Bottou L, Laptev I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]. CVPR, 2014: 1717-1724.
[84]Donahue J, Jia Y, Vinyals O, et al. Decaf: A deep convolutional activation feature for generic visual recognition[J]. arXiv preprint, 2013: 1310.1531.
[85]Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. CVPR, 2014: 580-587.
[86]Esuli A, Sebastiani F. Sentiwordnet: A publicly available lexical resource for opinion mining[C]. Proceedings of LREC. 2006, 6: 417-422.
[87]Thelwall M, Buckley K, Paltoglou G, et al. Sentiment strength detection in short informal text[J]. Journal of the American Society for Information Science and Technology, 2010, 61(12): 2544-2558.
[88]Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
[89]Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J]. Eprint Arxiv, 2014: 675-678.
[90]Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]. ICLR, 2015.
[91]Rifai S, Bengio Y, Courville A, et al. Disentangling factors of variation for facial expression recognition[M]. ECCV, 2012: 808-822.
[92]Bengio Y. Learning deep architectures for AI[J]. Foundations and trends in Machine Learning, 2009, 2(1): 1-127.

[93]Hu X, Tang J, Gao H, et al. Unsupervised sentiment analysis with emotional signals[C]. Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013: 607-618.