**第四章 GIF 情感分析时序模型**

上文讲到， 要实现GIF 情感分析，主要有两个问题需要解决。一是确定GIF 视频中出现概念语义的范围，针对这个问题，我们在第三章中首先提出“情感对”作为 GIF 视频概念语义的表示，然后在情感对的基础上提出了一个概念语义体系。概念语义体系为 GIF 视频情感分析中出现的概念语义划定了范围。解决了第一个问题。

GIF 情感分析面临的第二个问题是如何处理GIF 视频的时序关系。在本章中，我们首先比较了常见的时序模型构建方法，接下来提出了一个 GIF 情感分析时序模型—“情感对序列”模型。之后，通过实验评估了该时序模型的效率，并比较了其他研究成果中不同序列构建方式对最终情感判断的影响。试图解决第二个问题。

**4.1 情感分析领域常见时间序列分析方法**

在图像情感分析领域，由于图像没有时序信息，没有相关工作可以参考。而现有的GIF 情感分析工作，要么没有考虑到时序信息，例如[32]，要么只是将每一帧的信息简单堆积，例如[1]。简单堆积又有两种形式：

第一： 将 GIF 各个视频帧的概念语义信息直接拼成一个长向量

第二： 使用Bag of Words 模型，忽略不同视频帧之间的时序关系，得到一个词袋特征。

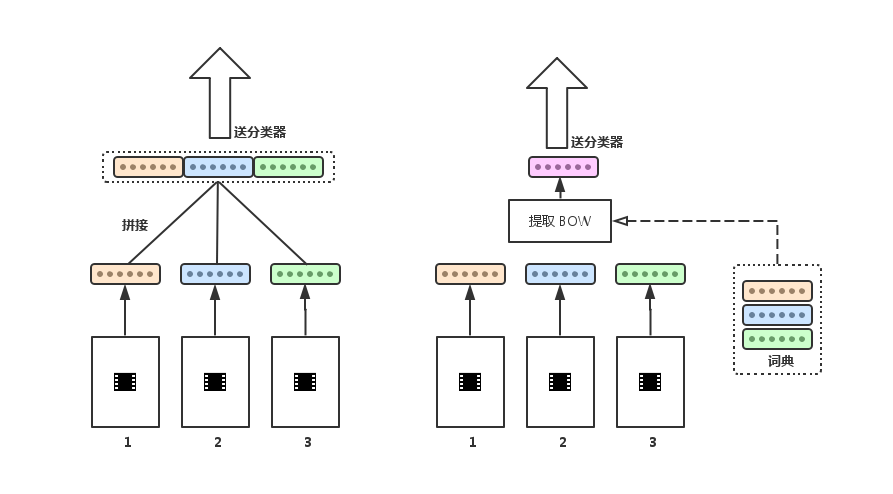


图4.1， 处理时序信息的两种简单形式

如上图所示，左图展示的是[1]文中将各个帧的特征简单拼接构成视频特征的处理流程。在该工作中，先是对各个帧分别提取底层特征，再将各个帧的底层特征直接拼接为一个长向量作为GIF视频的特征。由于不同视频的帧数不一样，需要统一从视频中截取指定长度的帧，例如等间距截取六帧或八帧。由于底层特征每一个维度之间没有确定的关系，按照这种方式拼接成的特征也没有可以解释的现实意义。右图展示的是视频词袋特征的提取过程。首先提取每一帧的底层特征，然后通过大量的数据构建底层特征词典。最后从词典中取出各个词的出现情况和位置作为整个视频的底层特征。特征维度就是词典的长度。这种形式实际上忽略了各个帧的信息在视频中出现的原始位置，仅仅是统计了视频中各个词的出现情况。

但是，将每一帧的信息按照上面介绍的两种方式进行简单堆积真的合理么？来看下面的例子。

【一个时序关系导致情感判断出现误差的对比图，最好是同一张图的不同顺序放置导致判断结果不一】

图4.2 时序关系对 GIF 情感判断的影响

如上图所示，GIF 视频帧出现顺序的不同会令情感的极性出现完全相反的表现。同时，很多概念之间的联系是随着时间依次变化的。

【又一张图说明】

图 4.3 不同帧序对情感判断的影响

**4.2 情感分析时序模型提出—情感对序列**

为了解决上述问题，我们设计了一个 GIF 情感分析时序模型—情感对序列。顾名思义，一个情感对序列是一个情感对的向量。为了评价不同长度的视频，向量的维度是不确定的。向量中的每个值代表一个情感对。而情感对则是从 GIF 视频帧中检测得到的。下图显示了一个典型的情感对序列：

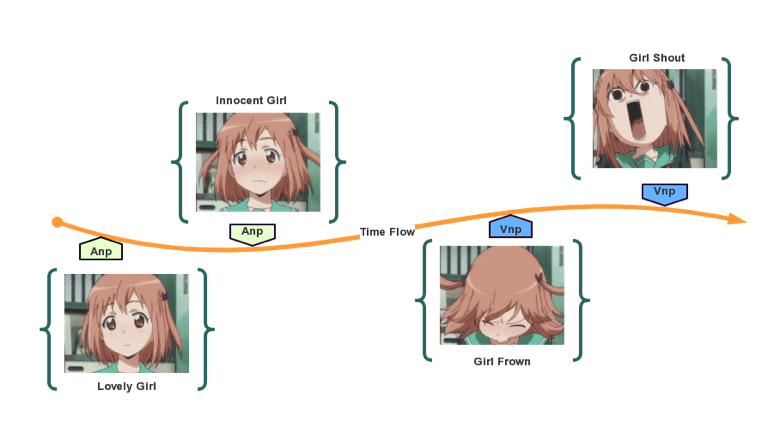


图4.4 情感对序列示意图

如图4.4所示，GIF 视频的内容为一个小姑娘的情绪变化过程。在第一帧中呈现的是一个可爱的女孩，因此，我们的情感对检测器检测到 Lovely Girl， 第二帧中呈现的是一个比较中性的小女孩脸庞，我们的情感对检测器检测到的是 Innocent Girl无辜的女孩，在第三帧中，小女孩皱起了眉毛，我们的情感对检测器检测到了小女孩皱眉这个动作，并使用动词名词对（ VNP） 的形式描述为Girl Frown， 在第四帧中，小女孩呈现出一个抓狂的咆哮动作，该动作被情感对检测器捕捉到并使用 Girl Shout进行描述。由此可见，使用情感对序列，我们可以准确地对 GIF 视频中出现的概念语义进行建模。在描述视频概念语义的过程中，我们使用了情感对作为基本的概念语义描述单元。情感对序列中各个情感对由第三章中介绍的检测器负责检测。

**4.3 评估情感对序列在 GIF 情感识别中的有效性**

上文提到，为了解决如何处理 GIF 视频中概念语义之间的时序关系问题，我们设计了 GIF 情感分析时序模型。为了评估情感对序列的有效性，受到近期循环神经网络在众多自然语言处理[45]和视频动作识别[44]方面取得的巨大成功的启发，我们引入了带长短期记忆单元（LSTM）的循环神经网络（RNN）作为我们时序分析的模型。

**4.3.1 传统循环神经网络**

一个典型的循环神经网络（RNN）的结构如图所示：

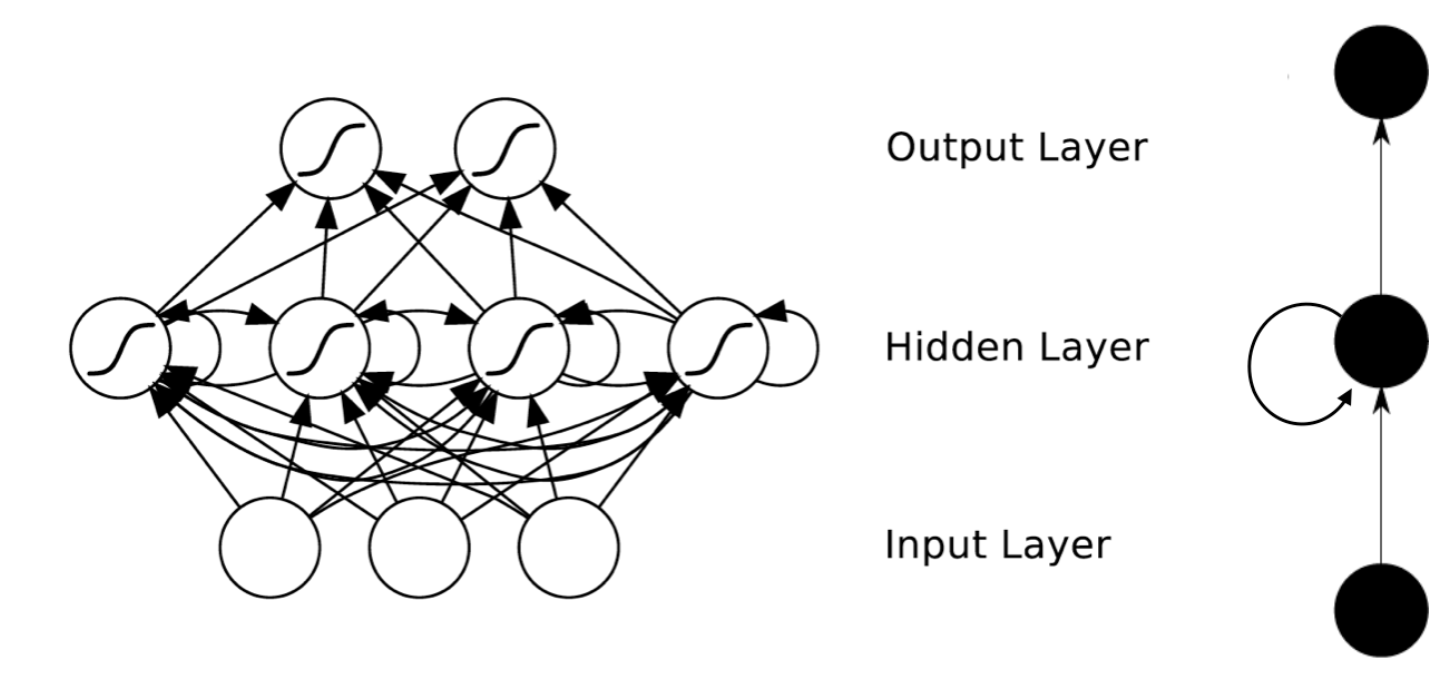


图 4.5 典型的循环神经网络（RNN）

如图4.5（右）所示，一个典型的循环神经网络的隐含层（Hidden Layer）结点中存在着连向自身的边。Input Layer 是输入层，Output Layer是输出层，Hidden Layer是隐含层。对于 RNN 的隐含层结点而言，需要处理的输入既包括输入层的输出，又包括上一个时刻的输出。由于这条边的存在，循环神经网络可以认为是有记忆能力的。那么，如何使用循环神经网络来处理时序信息呢？为了将时序信息输入网络，循环神经网络将神经元展开为若干个神经元的并行连接，如图4.6所示：

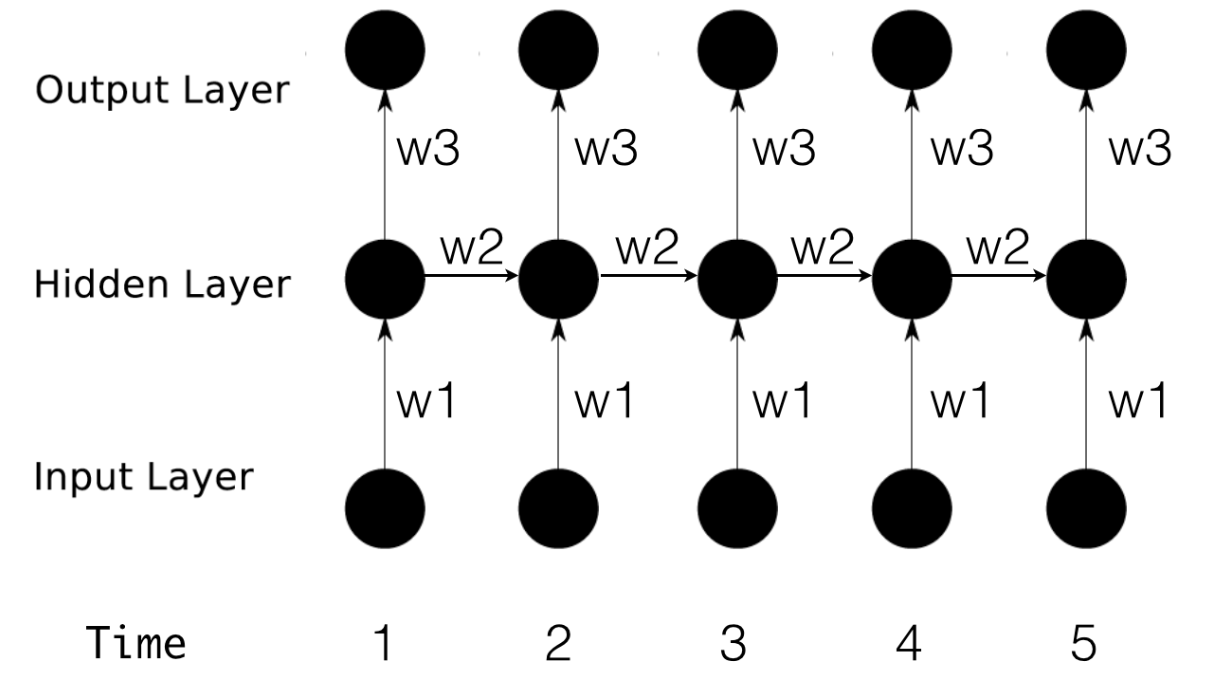


图4.6 展开连接后的循环神经网络

如图4.6所示，循环神经网络会对之前输入的结点信息进行记忆，并应用于当前输出的计算中。与传统的神经网络相比，隐藏层之间的节点不再无连接而是有连接的。各个隐藏层节点的输入不仅包括输入层的输出还包括上一时刻隐藏层的输出。理论意义上，循环神经网络能够处理任何长度的序列数据。但是在实践中，为了降低时间复杂度，往往假设当前的状态只与前面的几个状态相关。

**4.3.2 包含长短期记忆结构（LSTM）的循环神经网络**

传统循环神经网络的一个问题是，当输入序列很长时，神经网络学习时会出现“梯度消失”的现象。这是由于传统的循环神经网络将“过去”的所有节点都记住，导致参数过多，为了解决这个问题，Hochreiter 和 Schmidhuber[46]在1997年提出了长短期记忆的循环神经网络结构。在传统神经网络基础上引入了若干个门，使用这些门来控制网络的记忆范围和记忆深度。其中输入门（input gate）被用来控制输入信号的进入，输出门（output gate） 被用来控制神经细胞状态的输出，遗忘门（forget gate）被用来控制遗忘。一个典型的长短期记忆神经元如下图所示。

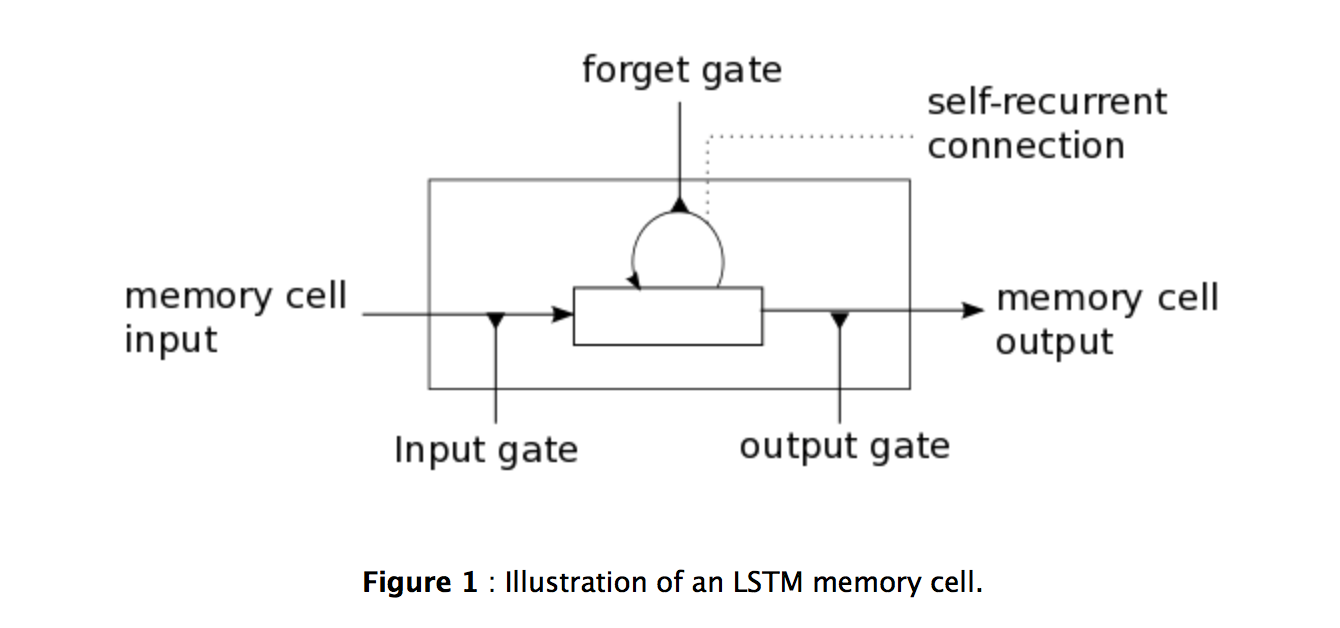


图4.7 LSTM结构示意图

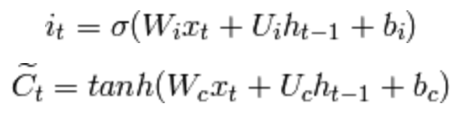
**4.3.3 长短期记忆神经元的学习方法**

设xt 是 t 时刻神经细胞的输入

Wi,Wf,Wc,Wo,Ui,Uf,Uc,Uo和Vo 是权值矩阵

bi,bf,bc和bo是偏置向量

在 LSTM 的学习过程中首先计算输入门在 t 时刻的激励 it 以及 t 时刻神经细胞的备选状态 :



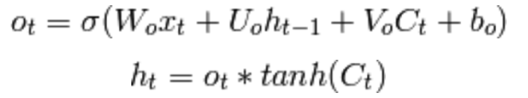
接下来，计算遗忘门在 t 时刻的激励ft：



在得到了遗忘门，输入门的激励和 t 时刻的备选状态后，我们可以得到t时刻的神经细胞状态：



最后，计算t时刻的输出：



ht就是t 时刻隐含层的状态。训练的目标函数是 softmax的损失函数

**4.3.4 LSTM和 RNN 用于情感对序列分类**

为了对情感对序列进行情感分类，我们采用了如下图所示的网络结构：

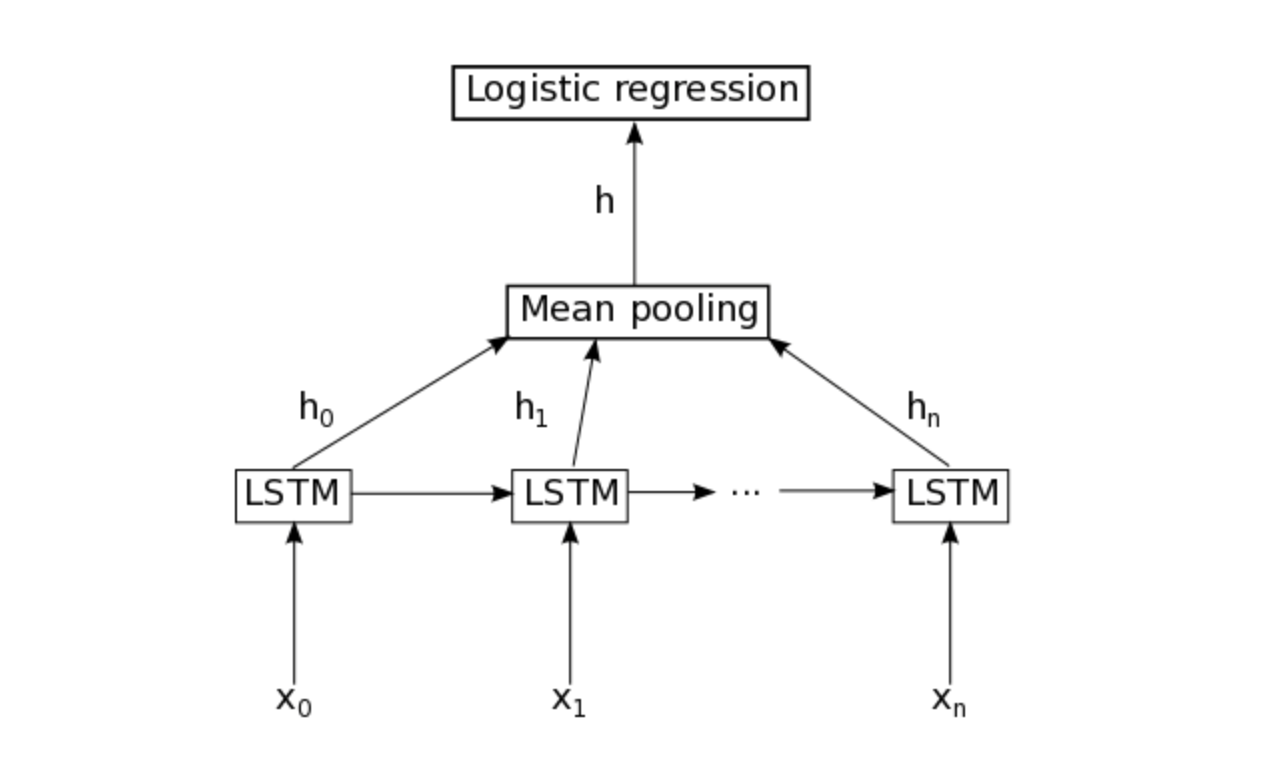


图4.8 情感对序列分类中使用到的网络结构

其中，隐含层包含500个 lstm 节点，也即图4.8中 n 值为500，训练的初始学习速率为0.0001，学习算法使用 adadelta, 训练过程采用五折交叉检验。训练数据是我们提出的 GSO-2015数据集。

**4.4 实验结果**

**4.4.1 使用传统方法训练情感极性分类器**

我们使用了多组评估 GIF 情感分析时序模型有效性的基准, 它们是基于词袋特征的 SMO 模型、基于词袋特征的逻辑回归模型、基于词袋特征的朴素贝叶斯模型、基于词袋特征的 AdaBoost 模型、降维后的 SMO 模型、降维后的逻辑回归模型、降维后的朴素贝叶斯模型、降维后的 AdaBoost模型。

使用这些基准的原因是这些方法在过去的情感分析问题中被经常使用，下面以使用词袋特征的 SMO 模型为例，首先将 GIF 视频按照等间距分为六个视频帧，对每一个视频帧提取 “情感对” 中层特征。具体的提取方法详见第三章，之后对于提取出来的“情感对”建立词典。这样可以将每一个视频帧表示为长度为词典大小的向量。在这个向量中绝大多数维度的值都是0，只有检测出的“情感对”对应的维度的值才为1。

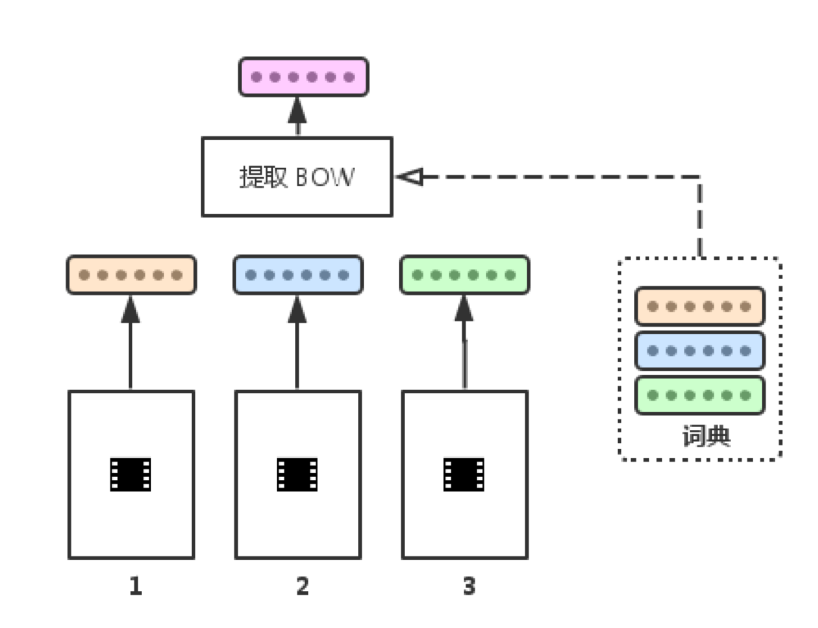


图 基于词袋特征的 SMO 模型

我们使用的训练数据是 GSO-2015 数据集中的人工标注数据，数据集的详细介绍在第五章中可以找到。在训练过程中为了数据集中偶然因素的作用，采用10折交叉检验，下图中的数据是10折交叉检验的平均值。另外一组基准是采用主成分分析方法将词袋模型的维度降到447维之后的分类结果。

表 训练集信息

|  |  |
| --- | --- |
| 样本分布 |  |
| 正极性 | 1239 |
| 中性 | 630 |
| 负极性 | 207 |

训练时所用的训练集信息如上表所示。共有1239个正极性 GIF 样本，630个中性样本和207个负极性样本。

表 GIF 情感分析时序模型基准

|  |  |  |  |
| --- | --- | --- | --- |
| 分类方法 | 情感对 | ANP | 情感对+PCA |
| SMO | 76.6 | 72.2 | 75.1 |
| NavieBay | 68.4 | 67.7 | 68.4 |
| AdaBoost | 64.8 | 64.4 | 64.8 |
| LogiRegr | 69.2 | 72.1 | 74.4 |

由上表可以看出，SMO 要显著优于其他的统计学习方法，仅仅使用情感对的词袋模型配合 SMO，就可以达到76.6%的分类准确率。单纯使用 ANP，在 GIF 情感分析中的效果均逊色于包含了动词名词对的“情感对”，而使用 PCA 对情感对进行降维后，分类效果没有明显提升。这个实验证明，如果仅仅配合词袋模型使用，SMO 是一个比较好的分类模型，降维带来的提升并不明显，而我们提出的情感对要显著优于单纯使用形容词名词对。

图 情感分析时序模型基准图，验证第二阶段

**4.4.2 使用带长短期记忆单元的循环神经网络训练情感极性分类器**

在该部分中，我们实现了4.3.4节中的带长短期记忆网络的循环神经网络。初始学习速率0.0001， 优化方法是 adadelta, batch\_size 为16，隐含层元素节点个数10000， 最大训练周期为1000。

图 **使用带长短期记忆单元的循环神经网络训练情感极性分类器**

如上图所示，使用上述参数训练出的情感极性分类器准确率相比基准中的分类器有很大的提升。提升幅度在（5%~20%之间）。经过分析，我们认为这主要是由于采用带长短期记忆单元的循环神经网络后，情感对之间的时序信息被保留下来，而这种时序信息是对照方法所不具有的。由于 GIF 视频的特性，前后时序关系对情感内容的影响无法忽略，使用可以处理时序信息的模型能更好地完成情感分析的任务。

**4.4.3 完整实验**

前文提到，在第三章，我们训练完成了情感对检测器，并评估了效果。在4.4.2节，我们训练完成了情感倾向检测器，并评估了效果。在本节中，我们综合运用这两个预先训练好的检测器，实现从抽帧，提取中层特征，完成最终情感分类的整个流程。本节的对照实验是 Brendan[1]等人的方法，该方法使用颜色，美学等底层特征来直接推断视频倾向性，实验结果如下:

图 完整实验结果

如上图所示，蓝色橙色和红色分别代表不同的训练周期数。第一列数据是对照试验的方法，因为该方法没有训练周期的概念，故准确率一致。第二列至第四列分别为不同的“情感对”提取方式的结果。其中6\*6指的是每帧抽取6个 “情感对”同时每个视频抽取6帧，10\*6指的是每帧抽取10个“情感对”每个视频抽取6帧，100\*6指的是每个视频抽取100个情感对每个视频抽取6帧。我们发现每帧抽取的情感对数目从6个提升到10个时，检验的准确率有小幅度提升（1%~%2），但是随着训练周期达到1000，反而导致准确率下降，这可能是过拟合的现象，模型在500个训练周期中已经充分训练。当每帧提取的“情感对”数量达到100时，模型的准确率反而下降，这可能是由于 RNN 网络的输入序列过长，欠拟合导致的。

**4.4.4 分类效果对比**

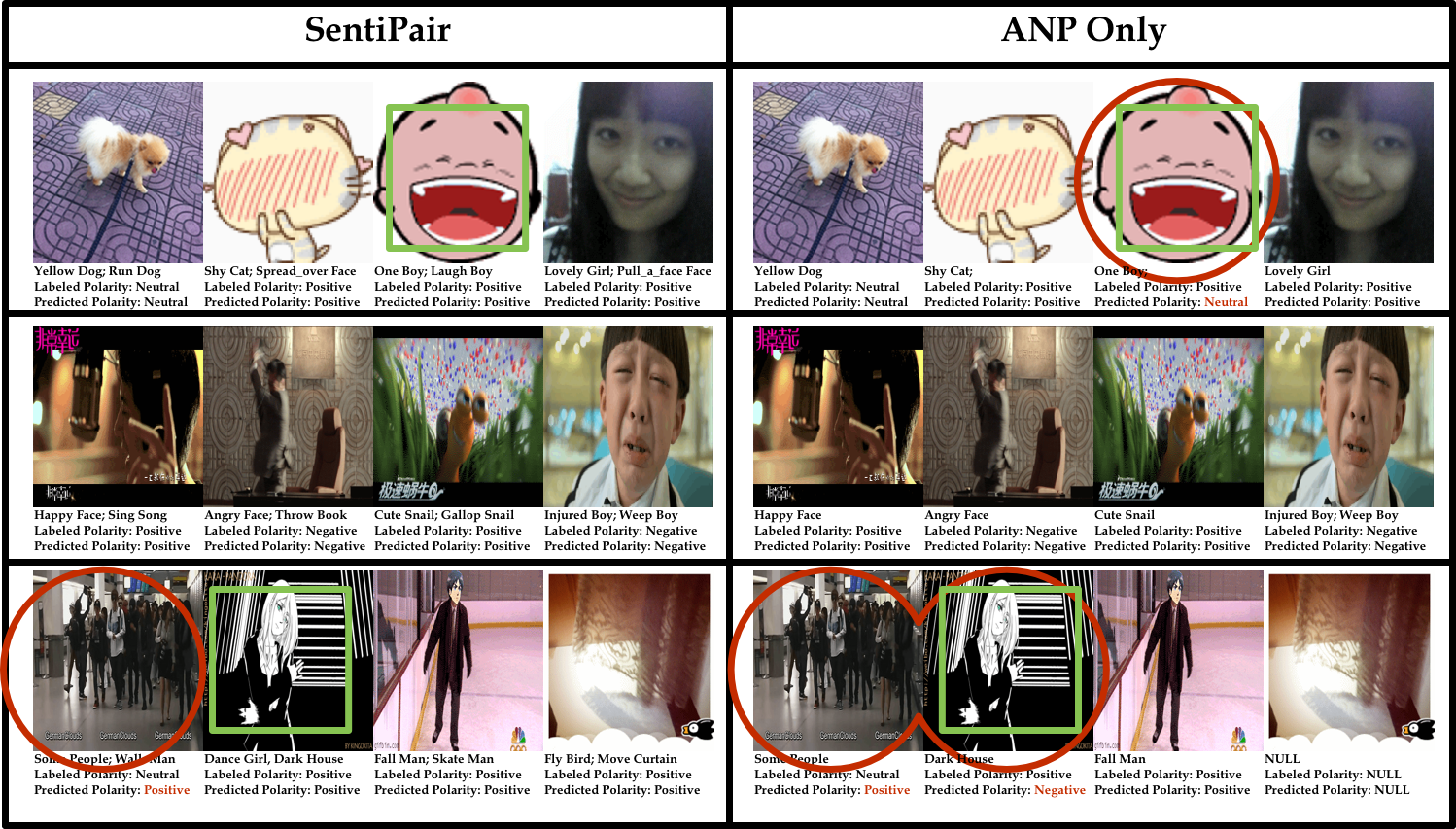


图 分类效果对比

如上图所示，红色方框中的是错误分类的样本。图的左侧展示的是“情感对”（SentiPair） 的分类结果和样本标签，右侧展示的是“形容词名词对”（ANP） 的分类结果和样本标签。在第一行右侧，有一个错误分类的样本。它的中间层分类是“one boy”,最终的情感分类是中性。而采用“情感对”检测器来检测时，中间层分类结果除了“one boy”以外还有“laugh boy”，由于“laugh boy”动作的存在，最终倾向性被分为正极性，左侧分类正确。可见，右侧的分类错误是由于形容词名词对的描述能力导致的。

参考文献

1. Brendan J, Subhabrata B, Shih-Fu C. Predicting Viewer Perceived Emotions in Animated GIFs, ACM MM’14, November 3-7, 2014, Orlando, Florida, USA
2. Jianbo Yuan, Quanzeng You, Sean McDonough, and Jiebo Luo, Sentribute: Image Sentiment Analysis from a Mid-level Perspective, ACM SIGKDD, Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM), Pages: 10:1--10:8, 2013
3. 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.
4. Feldman R. Techniques and applications for sentiment analysis[J]. Communications of the ACM, 2013, 56(4): 82-89.
5. Riloff E, Shepherd J. A corpus-based approach for building semantic lexicons[J]. arXiv preprint cmp-lg/9706013, 1997.
6. Hatzivassiloglou V, McKeown K R. Predicting the semantic orientation of adjectives[C]. Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics. Association for Computational Linguistics, 1997: 174-181.
7. Esuli A, Sebastiani F. Sentiwordnet: A publicly available lexical resource for opinion mining[C]. Proceedings of LREC. 2006, 6: 417-422.
8. 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算 [J]. 中文信息学报, 2006, 20(1): 14-20.
9. Turney P D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews[C]. Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002: 417-424.
10. Yu E, Kim Y, Kim N, et al. Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary[J]. Journal of Intelligence and Information Systems, 2013, 19(1): 95-110.
11. Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[C]. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.
12. Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[C]. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.
13. Riloff E, Patwardhan S, Wiebe J. Feature subsumption for opinion analysis[C]. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2006: 440-448.
14. McDonald R, Hannan K, Neylon T, et al. Structured models for fine-to-coarse sentiment analysis[C]. Annual Meeting-Association For Computational Linguistics. 2007, 45(1): 432.
15. Nakagawa T, Inui K, Kurohashi S. Dependency tree-based sentiment classification using CRFs with hidden variables[C]. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010: 786-794.
16. Deng Z H, Luo K H, Yu H L. A study of supervised term weighting scheme for sentiment analysis[J]. Expert Systems with Applications, 2014, 41(7): 3506-3513.
17. 唐慧丰, 谭松波, 程学旗. 基于监督学习的中文情感分类技术比较研究[J]. 中文信息学报, 2007, 21(6): 88-94.
18. Zheng L, Wang H, Gao S. Sentimental feature selection for sentiment analysis of Chinese online reviews[J]. International Journal of Machine Learning and Cybernetics, 2015: 1-10.
19. 陈俊杰, 张大炜, 李海芳. 融入模糊理论的 SVM 在图像情感识别中的应用研究[J]. 计算机科学, 2009, 36(8): 288-290.
20. Davidov D, Tsur O, Rappoport A. Enhanced sentiment learning using twitter hashtags and smileys[C]. Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010: 241-249.
21. Barbosa L, Feng J. Robust sentiment detection on twitter from biased and noisy data[C]. Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010: 36-44.
22. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision[J]. CS224N Project Report, Stanford, 2009: 1-12.
23. Bravo-Marquez F, Mendoza M, Poblete B. Combining strengths, emotions and polarities for boosting twitter sentiment analysis[C]. Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining. ACM, 2013: 2.
24. 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.
25. Yan B, Zhang B, Su H Y, et al. Chinese microblog sentiment classification considering users’ reviews based on Naive Bayesian algorithm[J]. Future Communication, Information and Computer Science, 2015: 115.
26. Lu T J. Semi-supervised microblog sentiment analysis using social relation and text similarity[C]. Big Data and Smart Computing (BigComp), 2015 International Conference on. IEEE, 2015: 194-201.
27. Y.-Y.Chen,T.Chen,W.H.Hsu,H.-Y.M.Liao,andS.-F.Chang.Predictingviewer affective comments based on image content in social media. In ACM ICMR, 2014.
28. Borth, D.; Ji, R.; Chen, T.; Breuel, T.; and Chang, S.-F. 2013b. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In ACM MM, 223– 232. ACM.
29. Cees G.M. Snoek , etc. Early versus Late Fusion in Semantic Video Analysis
30. J.MachajdikandA.Hanbury.Affectiveimageclassificationusingfeaturesinspired by psychology and art theory. In ACM Multimedia, 2010.
31. X.Wang, J.Jia, P.Hu, S.Wu, J.Tang, and L.Cai. Understanding the emotional impact of images. In ACM Multimedia, 2012.
32. Quanzeng You, Jiebo Luo, Hailin Jin, and Jianchao Yang, Robust Image Sentiment Analysis using Progressively Trained and Domain Transferred Deep Networks, the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI), Austin, TX, January 25-30, 2015.
33. George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.
34. Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT
35. Bollen, J.; Mao, H.; and Pepe, A. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In ICWSM.
36. Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood pre- dicts the stock market. Journal of Computational Science 2(1):1–8.
37. Borth, D.; Ji, R.; Chen, T.; Breuel, T.; and Chang, S.-F. 2013b. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In ACM MM, 223–232. ACM.
38. Matthew D. Zeiler and Rob Fergus, Visualizing and Understanding Convolutional Networks, ECCV 2014, Part I, LNCS 8689, pp. 818–833, 2014.
39. George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.
40. Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.
41. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks
42. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Eric Tzeng. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition
43. Andrej Karpathy, George Toderici, Sanketh Shetty. Large-scale Video Classification with Convolutional Neural Networks
44. Joe Yue-Hei Ng, Matthew Hausknecht, etc. Beyond Short Snippets: Deep Networks for Video Classification. Computer Vision and Pattern Recognition 2015
45. Kai Sheng Tai, Richard Socher, Christopher D. Manning. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. Arxiv.org Doc ID: 1503.00075V3 30 May 2015
46. Hochreiter & Schmidhuber. Long Short-Term Memory. Neural Computation 9(8): 1735-1780, 1997
47. Andrea Esuli and Fabrizio Sebastiani. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining