一种用于社交网络 GIF 视频情感分析的时域空域视觉中层表示概念语义体系

# 摘 要

近年来随着社交网络的快速发展，人们日益倾向于在网络上表达自己对热点事件的态度、网购产品的评价、热门影视作品的观感、 政治问题的立场等等。在日益增多的富含情绪的网络内容中，GIF动画正在占据日益重要的地位。据我们从新浪微博中获取到的数据，有15.1%的微博中含有GIF动画等动态多媒体内容。而在这些多媒体内容中，只有不到10%的内容是描述一些客观存在的物体或事件（例如堵车，雾霾）。也就是说， 90%以上的 GIF 动画都被用来表达博主的主观情感。

对GIF动画进行情感倾向性分析，可以帮助政府或公司了解网民对舆论热点事件的态度，从而提供决策支持。

不幸的是，对GIF动画的情感倾向性分析是一项很有挑战性的工作。GIF动画的倾向性取决于时间和空间上的图像语义关系。 同时， 这些时间空间上的图像语义关系和最终的情感倾向之间的联系是不明确的。

在本文中， 为了探索时间空间上的图像语义关系和最终情感倾向之间的联系， 我们提出了一个被称为“情感对序列” 的模型。在这个模型的基础上， 我们提出了一套时间空间上的 GIF 图像情感理论体系。这套理论体系构成了我们的 GIF 动画中层表示。“情感对”的构造过程包括两个步骤：首先， 我们构建了一个语义森林来表示 语义项的层次关系。然后， 从语义森林中选择合适的语义项来构成“情感对”，将得到的“情感对”按照时间顺序排列成一个序列， 也就得到了我们的中层表示。

我们的实验结果表明， 在 GIF 动画情感分析领域拥，使用“情感对”模型的分类器拥有比使用其他中层模型的分类器更高的分类准确率。我们也同时向科学界发布了我们人工标注的数据集（GSO-2015）。 该数据集包含从40000张 GIF 动画中选出的6000张人工标注的 GIF 动画和相应的“情感对”标注和情感倾向。

**关键词**：情感倾向性分析；语义概念检测；本体树

# Abstract

With the progress of social medias, an increasing number of dynamic multimedia, such as video clips (GIF), were used in Internet. Many of them not only express the objective facts of event, but also show the subjective sentiment of users. GIF sentiment analysis can be used to predict political elections, evaluates economic indicators and so on. Unfortunately, such task is quite challenging, the sentiment judgment hinges on spatial-temporal visual concepts and the relationship between such concepts and final sentiment remains unknown. In this paper, dedicated to build a bridge for exploring such relationship, we proposed a SentiPair Sequence based spatial-temporal visual sentiment ontology, which forms the mid-level representations of GIF sentiment. The establishment process of SentiPair contains two steps. First, we construct a Synset Forest to define the semantic tree structure of visual sentiment label elements. Then, through the Synset Forest, we organically select and combine sentiment label elements to form a mid-level visual sentiment representation. Our experiments indicate that SentiPair outperforms other competing mid-level attributes. We also released our dataset (GSO-2015) to the research community. GSO-2015 contains more than 6,000 manually annotated GIFs out of more than 40,000 candidates. Each is labeled with both sentiment and SentiPair Sequence.

***Keywords—Sentiment Analysis; Concept Detection; Ontolog***

第一章 绪论

1.1 研究意义及问题背景

近年来，日益增多的互联网用户正在将社交网络变成一个文本、图像、视频剪辑（GIF 动画）的巨大集合。用户在社交网络上分享他们的生活点滴和各种观点。例如对政府政策的态度，对股票投资的选择和对流行文化的认可程度等（如图1.1所示）。在社交网络中存在的海量数据为各种各样的大数据分析工作提供了便利。通过在社交网络中引入针对图片、文本、GIF 动画的情感倾向性分析，政府和掌握数据的大公司未来将可以根据海量用户的行为数据评估政策的执行效率，考核官员的清廉程度或业绩，预测未来的经济走势，根据海量用户的使用习惯修改产品等等。

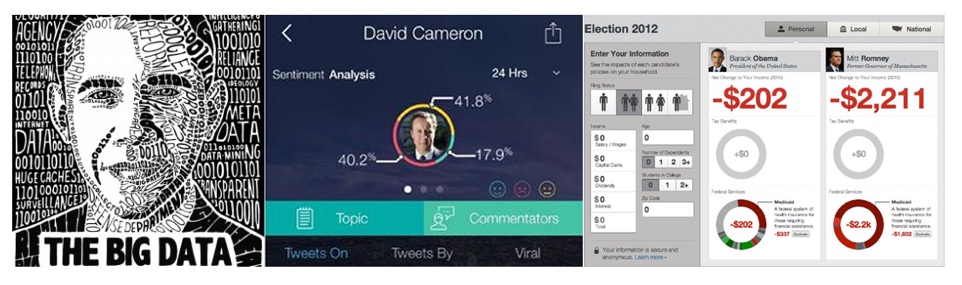


图1.1 情感计算运用于社交网络

与静态图片相比，GIF 动画更加形象生动，一幅合适的 GIF 动画可以让微博内容增色不少。过去，高昂的带宽价格限制了 GIF 动画在社交媒体的使用,而如今随着带宽的提高， 相比单幅的静态图片，人们更愿意上传形象生动的 GIF 动画。根据一项最近的调查 [2]，美国社交网络 Twitter 上共享的链接中有36%含有视觉信息。然而，真正投入实际应用的情感倾向性分析系统都是基于文本的。对 GIF 动画的情感倾向性分析研究依然处于其初级阶段。GIF 视频广泛的分布和 GIF 情感分析研究缓慢的进展是社交媒体情感分析领域的一大矛盾现象。

这个矛盾主要是由 GIF 情感分析的难度和挑战性导致的，总的来说 GIF 动画倾向性分析有以下几个挑战：

**倾向性位置不确定：** GIF 动画是一个图片帧的序列。在全部的帧中可能只是少数几帧包含了倾向性。而对于不同的 GIF 视频，包含情感倾向的这些“关键帧”所处的位置又不一样。如何在一个 GIF 视频中定位出这些富含情感信息的“关键帧”是一个较难的研究性问题。

**抽象性：**在GIF 动画中，情感倾向需要经过高层抽象才能得出。这种抽象性存在的根本原因是在人们观察到的感知层内容和判断情感倾向的认知层内容之间存在着巨大的“语义鸿沟”。感知层内容可以理解为人的视觉系统可以直接从GIF视频中获取到的信息，例如色彩、形状、纹理、运动情况等等；认知层内容则是通过大脑的识别得到的对视频内容的客观认知判断，例如行人、蓝天、跌倒、嘲讽等等。如何将该“语义鸿沟”桥接起来是一个巨大挑战。

**噪声的广泛存在：**GIF 图片中经常混杂着说明性文字或者字幕。有些文字会影响情感倾向。有些文字则不会。如图1.2所示，左右两幅图片显示的实际上是同一个动作，然而，由于图片中说明性文字的不同，V 字型手势可以理解为两个不同的含义。对于这种图文混合的 GIF 视频，情感倾向性的预测是一个巨大挑战。



图1.2 GIF 视频中文字噪声对情感倾向性的影响

**象征性：**在一个GIF视频中有很多概念语义，而这些概念语义往往是隐含在感知层表观之后的。例如一个 GIF 动画中出现了史泰龙的形象，对最终情感倾向产生影响的应该是“强壮的男人”而不应该是史泰龙形象本身。史泰龙只能说是“强壮的男人”的一个象征物。如何处理感知层表观之后的象征意义是一个较难的问题。

1.2 研究现状

情感倾向性分析有被称为“情感计算”（affective computing）,最初是由 MIT 媒体实验室的 Picard在《情感计算》一文中提出的。Picard 将情感计算定义为“与情感有关、由情感引发或可以影响情感的计算”。其目的是通过赋予计算机识别、理解、表达和适应人的情感的能力来强化计算机的智能。目前比较成熟的情感计算领域有文本情感分析、静态图像情感分析，随着新媒体和社交网络的兴起，针对 Twitter、微博等的情感分析也逐渐流行，越来越多的研究者开始目光投向社交媒体。而GIF 视频的情感分析由于上文介绍的困难和挑战，目前的研究较少。

1.2.1 文本情感分析

文本情感倾向分析的任务是从文本内容得出用户的情感倾向性（正面、负面或中性）。目前效果最好的文本情感分析方法大致分为两类：基于情感知识的无监督方法以及基于机器学习的有监督方法。

基于情感知识的无监督方法

基于情感知识的无监督方法主要依赖领域词典和情感词典的组合，使用情感知识来评价计算文本的情感极性[3][4]。无监督方法的优势在于节省了大量的人工标注，减少重复劳动。但情感分类的准确率严重依赖于情感词典的准确率。同时，情感词典的覆盖性若不够全面将影响最终情感分类，在新语料中缺少灵活性。构建情感词典的方法主要有基于语料和基于单词网络两种：

**基于语料的方法：**Riloff等人[5]用统计方法在语料数据库的基础上构建了语义词典；Hatzivassiloglou等人[6]从海量语料中提取同现的形容词对，Hatzivassiloglou认为整个语料库中的同现形容词之间具有相近的语义关系。他们使用对数线性回归和聚类将形容词分为褒义和贬义两个情感倾向。

**基于单词网络的方法：**基于单词网络的方法是根据现有的单词网络或通用词典，从词典中取出情感词。研究中为大家所公认的单词网络主要有国外的WordNet和国内的HowNet。

此外，在构建情感词典时，也可以使用同义词词林[7][8]来构建。

在基于情感知识的分类方法中，一个典型的方法是Turney[9]提出的。首先利用事先定义好的词性组合从语料中提取情感短语，然后通过点互信息计算出情感短语的极性值，最终整合句子中各个情感短语的极性值得到整个句子的情感倾向性。在股票指数预测领域，一个成熟的方法 [10]是 构建特殊领域情感词典。

有监督方法

有监督的情感倾向性分类方法，通常包含三个步骤，第一个步骤是特征的设计和提取，第二个步骤是分类器的训练，第三个步骤是分类器的应用。

在第一个步骤中，需要根据实际问题的需要，设计合适的特征，常见的有 n-gram[11]（n 元文法中 n 的取值一般是1，2，3等等），词性标注[12]等等，也有使用多种特征组合[12]的。在第二个步骤中，需要选用合适的分类器模型，常用的模型有支持向量机(SVM)，朴素贝叶斯(Naïve Bayes)和最大熵(Maximum Entropy)等等。在第三个步骤中，应用前面学习到的分类器模型，对结果进行分类。Pang[12]采用三种分类算法对电影评论进行极性分类。他比较各种特征和分类器后发现基于一元文法的特征在SVM下可以取得最佳的分类效果。此后，便有很多人在他的基础上做了很多特征表示、特征选取和分类算法的研究，以提高情感分类的准确率。在特征表示和特征选择的研究中，文献[13][14][15]比较上下文、依存关系以及文档子成分信息等特征。Deng等[16]对特征加权策略进行了研究。唐慧丰等[17]研究了对器学习分类技术和特征选择。他选择名词、动词以及形容词作为文本特征，对比文档频率、KL散度等特征选择方法以及SVM、Navie Bayes等机器学习算法。Zheng[18]采用了N元文法配合特征选择算法。陈俊杰等[19]在传统的方法上引入了模糊理论。有监督的情感倾向性分类方法中的一个最新方法是使用带长短期记忆单元（LSTM） 的循环神经网络来判断整个句子的情感倾向，Kai Sheng Tai[45]等人设计了两种树状（Tree Structure）的LSTM网络，这两种树状网络分别对应了自然语言处理领域两种不同的句法树。

1.2.2 社交网络情感分析现状

在移动互联网时代，人们越发集中地在社交网络中发布新内容。传统的文本情感分析技术主要聚焦在商品评价、电影评论等较长的文本上。而在社交网络中，文本内容往往呈现碎片化的趋势，例如推特和微博在早期都有140个字数的长度限制。此外社交网络上的内容构成也与传统长文本存在很大区别。传统的针对长文本的情感分类技术已经不能适应新媒体的发展。目前面向微博的情感分析在近几年来成为了研究热点。由于Twitter相对国内的微博发展较早，且针对中文的情感分析与针对英文的情感分析相比有诸多挑战。目前的研究还是主要集中在英文语料中。Davidiv等人[20]使用Tweets(微博消息)中的标签和表情符号作为粗标注，利用KNN训练分类模型，从而对微博消息进行情感极性分类。Barbosa[21]等人利用已有的面向Twitter的情感分析网站和工具(Twitter Sentiment，TweetFeel)对微博进行极性分类，它们中有的使用简单投票的方法得出训练集的情感标注信息；然后选取表情、情感短语、动词等作为特征，采用SVM进行分类。Go[22]通过设定的词语检索Twitter中含有笑脸和哭泣表情的微博，根据表情对消息进行正负极性标注，选取1-gram(一元文法)、2-gram(二元文法)以及POS(词性标注)的组合作为特征，比较 SVM、Navie-Beyes和ME分类算法训练分类器，最终在数据集中筛选出了0.8以上准确率的分类算法。Bravo-Marquez等[23]将流行的情感分类系统进行结合，加入表情等其他特征，对微博进行极性分类。

关于中文微博的研究在这几年迅速增加。较为典型的研究是谢丽星等[24]提出的基于级联支持向量机方法，他们采用了主题无关和主题相关的情感词典作为特征。 对文本进行主客观分类和正负极性主观分类，并将提出的方法与基于表情符号和基于情感词典的方法进行了对比。最终实验结果表明基于主题相关的特征和SVM，在测试集中准确率为67.28%。除了微博本身的消息外，也有研究者将评论信息加入进行分析[25]。Lu[26]等人在进行情感分类时还考虑了微博用户间的关系。

1.2.3 静态图像情感分析

静态图像情感分析方法主要有三种，一是直接使用底层特征进行情感分类，而是提炼出中间层表示进行情感分类，三是使用深度学习的方法直接学习情感分类。Borth**[28]**等人提出了一个视觉情感分类器的集合 SentiBank。这个分类器集合构成了一个视觉情感的中层表示。简单的说，中层表示中包含了1200个线性 SVM 分类器的分类结果，这些分类器的分类依据是 “形容词名词对”。每一个“形容词名词对”包含一个容易被检测到的名词和一个修饰形容词。例如“可爱的小狗”，“美丽的夕阳”，“难吃的食物”，“可怕的事故”等等。这些“形容词名词对”是从图片社交网站 Flickr 中爬取的。在**[27]**中，作者研究了图像的“元数据”，使用文本语言模型分析静态图片中用户体验到的情感。在视觉情感计算中，**[30]**文的作者提出了一组由颜色，纹理以及从美学理论中引出的复合特征，采用朴素贝叶斯模型来计算静态图像的情感特征。在**[31]**文中，一个全局的基于颜色的特征被设计出来，这个特征被用于因子图的构建。所有以上这些方法采用的都是特征的早期融合手段。并且需要为每一个情感类别训练一个分类器。

图1.3 早期融合和晚期融合的比较（图片来自文献**[29]**）

1.2.4 动态视频情感分析

在动态视频情感特征领域，目前的研究方式主要是在视频底层特征与高层语义之间建立映射，桥接语义鸿沟，通过视频底层特征实现对高程信息的映射。传统的 GIF 动画情感分析问题，更多的是考虑少数几种底层特征[1]，例如图像纹理，对称性，颜色直方图等等。GIF 动画情感分析的另一个思路[32]是构建深度神经网络，期望通过大量的带监督数据和训练让神经网络自主学习上述特征。常用的底层特征包括：颜色特征、纹理特征、形状特征、场景亮度、颜色能量、场景节奏等等。下面介绍三种常见的底层特征。

颜色直方图：GIF 动画本身含有大量的视频帧，每一帧都是一副静态图像。心理学的研究表明，人类的一些情绪的产生与日常接触到的颜色有关，例如红色可以代表热情和暴力，蓝色可以代表纯洁与忧郁。在动态视频情感分析任务中，经常使用图像颜色在红、绿、蓝三个通道上的分量的平均值来构建特征向量，希望以此建立颜色向量与情感之间的对应关系。

纹理特征：静态图像中有图像的表面纹理，GIF 视频中对应的不同概念语义有其不同的组织结构和平滑程度。以人类皮肤为例，不同年龄层次的皮肤会给观看者带来不同的情感，婴儿的皮肤光洁嫩滑，给人愉悦的感受，老人的皮肤干瘪多褶皱，给人留下岁月的痕迹。 GIF 动画中的视频纹理也经常被用于构建底层特征。

亮度特征：在 GIF 视频中，可以在 HSV 模式下通过色度、亮度、饱和度等来表示颜色，而亮度特征可以看做是视频色彩特征的另一种表达方式。

1.3 本文的研究内容

本文的研究内容围绕 GIF 动画的情感识别展开。针对目前 GIF 动画标注数据少，GIF情感分析问题缺乏完整的概念语义体系， GIF动画情感分析缺乏有效的时序表示等问题，分别提出了解决方案，这些解决方案构成了本文的三个贡献：针对第一个问题，我们从社交网络中爬取了大量数据，并进行了人工标注，这就是我们提出的 GSO-2015数据集；针对第二个问题，我们提出了 GIF 视频情感分析概念语义体系（GIF Sentiment Ontology）；针对第三个问题，我们提出了一个时序特征表示—“情感对序列”(SentiPair Sequence)。

本文系统研究了 GIF 动画情感分析问题。我们首先提出了 GIF 情感分析中需要解决的两个核心问题：第一个核心问题是如何确定哪些 GIF 视频中的概念语义需要被处理。第二个核心问题是如何处理 GIF 视频中概念语义之间的时序关系。并对解决问题的思路进行了规划。

在这个研究思想的指导下， 针对GIF 动画情感分析领域缺乏完整的概念语义体系问题， 我们提出了GIF 情感分析的概念语义体系结构 GSO（GIF Sentiment Onotology）。在构建出 GSO 之前，我们首先提出了一个 GIF 视频中出现的概念语义的表示方法--“情感对”。之后，我们从一个成熟的单词网络（WordNet）中筛选概念语义来构成我们的概念语义体系。为了完成筛选工作，我们提出了两个权值，一个是 SentiWeight 情感权值，负责考察概念语义的情感倾向程度，另一个是 GiphyWeight 情感权值，负责考察概念语义在实际 GIF 视频中出现的概率。构建好概念语义体系后，我们需要一种方法来从 GIF 视频中检测出概念语义体系内包含的所有概念语义。传统的 GIF 动画情感分析问题，更多的是考虑少数几种底层特征[1]，例如图像纹理，对称性，颜色直方图等等。 受到卷积神经网络在计算机视觉领域巨大成功的启发，我们选择了加入情感相关性挖掘和多任务学习目标函数的卷积神经网络作为检测模型。该模型在我们采集并标注的 GSO-2015 数据集中进行训练。概念语义体系结构和对应的检测器的提出解决了 GIF 情感分析的第一个核心问题。

针对GIF 情感分析缺乏好的时序特征表示问题，我们在分析传统 GIF 视频情感分析方法之后，提出了被称为“情感对序列”的 GIF 视频情感分析时序表达。“情感对序列”相比以往的时序分析方法具有很多先天优势，例如，可以描述连续的动作信息，可以描述不同状态的转变等等。为了利用“情感对序列”来预测最终情感倾向，我们使用了带长短期记忆单元（LSTM） 的循环神经网络（RNN） 作为分类模型，该模型同样在 GSO-2015数据集中进行训练。实验结果表明，无论是与统计学习方法（如 SVM,SMO,Logistic Regression，Navie Bayes）还是与传统循环神经网络相比，我们的模型都有着更优秀的性能。“情感对序列”和对应最终情感检测器的提出解决了 GIF 情感分析的第二个核心问题。

此外，为解决目前带标注的 GIF 动画数据少的问题，我们从新浪微博中爬取了超过四万张 GIF 动画，并人工标注了其中6000张，解决了GIF 数据集缺失的问题。同时我们将人工标注的 GIF 动画整理公开发表，使得后续的研究者可以使用我们的数据集进行测试。

1.4 本文的主要创新点

本文的一个创新点是提出了 GIF 情感分析的一个中层特征体系。该中层特征体系作为从底层特征到最终情感判断之间的桥梁。现有的情感分析方法要么依赖底层特征，要么依赖深度学习自动学习出中层特征。与底层特征相比我们的中层特征概括力更强。与自动学习出的中层特征相比，我们基于单词网络和概念语义体系的可解释更好。同时由于深度学习中网络层数和训练数据量的限制，目前深度学习想要达到我们中层特征体系的语义覆盖程度是不现实的。

GIF 动画情感分析的另一个难点在于时间序列，情感分析不能孤立地看待每一帧图像，因此一个好的中层特征也需要时间维度的信息。例如一幅先出现笑脸再出现哭脸的GIF和另一幅先出现哭脸后出现笑脸的 GIF 所蕴含的情感倾向是不一样的。为了解决这个问题，我们提出了一个GIF 的时序中层特征表示--情感对序列。我们把GIF 动画抽象成一串情感对。

有了中层特征，和时序表示，我们需要训练两个检测器，一个检测器用于检测中层特征，另一个检测器用于检测最终情感倾向。

针对情感对的检测，我们使用了多种方法，受到卷积神经网络在计算机视觉领域取得巨大成功的启发，我们选择卷积神经网络作为“情感对”检测器的基础模型。同时，根据“情感对”的特点，我们选用“多任务学习”构造卷积神经网络的损失函数。根据观察，我们发现很多概念语义之间存在上下位等关联关系，我们还在“情感对”检测器中加入情感相关性挖掘，扩展了分类结果。得到了良好的效果。此外我们发现在全部标注信息中，人脸的表情占得比例很大，而统一的卷积神经网络在检测表情方面有所欠缺，因此我们选择将表情部分提取出来单独训练，取得了良好的效果。

针对最终情感倾向的检测， 我们使用了带“长短期记忆单元”（LSTM） 的循环神经网络作为检测模型。在实验中我们比较了分别对每一帧抽取不同个数的情感对情况下，检测效率的差异，通过分析实验结果我们找到了实验条件下最优的参数组合。

1.5 本文的组织结构

第一章绪论介绍了本课题的研究背景和意义，对本文情感分析、图像情感分析和 GIF 动画情感分析的现状进行了总结，在此基础上介绍了本文的主要研究工作、主要贡献、主要创新点和组织结构。

第二章GIF 动画相关特性和情感分析方法对GIF 动画的相关特性进行概括，说明研究存在的挑战和待解决的问题，提出了 GIF 视频情感分析需要解决的两个核心问题：第一个核心问题是如何确定哪些 GIF 视频中的概念语义需要被处理。第二个核心问题是如何处理 GIF 视频中概念语义之间的时序关系。并对解决问题的思路进行了规划。

第三章GIF 情感分析概念语义体系。通过提出 GIF 情感分析概念语义体系，我们为GIF 情感分析中的概念语义划定了范围。该体系的完整度由已经被广泛验证过的 WordNet 保证。同时，由于 GIF 视频内容的特色，直接照搬 WordNet 体系结构显然是不合适的。需要对 WordNet 进行筛选。为此我们提出了两条筛选原则，一是概念语义要有足够的情感丰富性，二是概念语义在 GIF 视频中出现的频率要足够高。为了满足这两条原则，我们提出了 SentiWeight 情感权值和 GiphyWeight 情感权值。在概念语义体系的基础之上，我们使用深度神经网络训练了“情感对”检测器，实验表明，使用多任务学习并利用概念语义体系结构挖掘情感相关性后，检测性的准确率得到了大幅度提升。

第四章是GIF 情感分析时序模型。我们在分析传统 GIF 视频情感分析方法之后，提出了被称为“情感对序列”的 GIF 视频情感分析时序表达。“情感对序列”相比以往的时序分析方法具有很多先天优势，例如，可以描述连续的动作信息，可以描述不同状态的转变等等。为了利用“情感对序列”来预测最终情感倾向，我们使用了带长短期记忆单元（LSTM） 的循环神经网络（RNN） 作为分类模型，该模型同样在 GSO-2015数据集中进行训练。实验结果表明，无论是与统计学习方法（如 SVM,SMO,Logistic Regression，Navie Bayes）还是与传统循环神经网络相比，我们的模型都有着更优秀的性能。“情感对序列”和对应最终情感检测器的提出解决了 GIF 情感分析的第二个核心问题。

第五章是微博 GIF 情感分析数据集和人脸表情检测器。在本章中我们主要介绍了 GSO-2015数据集以及第三章中介绍的“情感对”检测器的补充部分—人脸表情识别检测器 。在介绍 GSO-2015时，我们介绍了数据集的组成部分，数据集使用“情感对序列”标注的原理以及数据集的采集方法。GSO-2015数据集有4万余张 GIF 动画组成，其中带标注的视频数有6千余张。全部通过人工标注了情感极性和情感对序列。作为第三章中介绍的“情感对”检测器的补充部分，人脸表情识别检测器具有与原始模型相似的结构，通过在 Kaggle 数据集中进行训练，达到了较好的效果。人脸表情识别检测器的加入，丰富了“情感对”检测器的内容，也体现了我们提出的情感分析框架的可扩展性。随着未来各项识别技术的进步，我们的识别准确率也可以不断提高。

第六章是总结和展望，总结全文的工作，分析存在的不足，对未来的研究提出建议和设想。

第二章 GIF相关特性及现有的情感分析技术

为了分析GIF 视频的情感倾向，首先需要分析 GIF 视频有哪些特性。通过从在社交网络中的 GIF图像有以下特点：

1. 语义信息覆盖范围广， GIF 图像之所以含有情感倾向是因为 GIF 的内容是多样的。例如我们看到美丽的夕阳会产生积极的情感，看见可爱的婴儿会产生积极的情感，看到可口的食物会产生积极的情感。产生情感的因素很多，我们需要一种方法将他们识别出来才能完成最终的情感判断。如下图是我们在在线 GIF 图像网站 Giphy 中爬取到的标注动词数量分布，可见GIF 的语义信息分布十分广泛。

图2.1 GSO-2015数据集动词频率表（部分） 按照出现次数降序排列

1. GIF 是由一串时间上连续的视频帧组成的，情感的检测不是对孤立的视频帧进行的，仅仅考虑视频帧中存在哪些语义信息是不够的，还需要考虑视频帧之间的时序关系。

2.1 GIF 视频情感分析需要解决的两个核心问题

由此，我们提出GIF 视频情感分析需要解决的**两个核心问题**：第一个核心问题是如何确定哪些 GIF 视频中的概念语义需要被处理。第二个核心问题是如何处理 GIF 视频中概念语义之间的时序关系。

2.2 静态图像情感分析技术

目前现有的图像情感分析技术主要有三种，第一种是使用复杂的模型（通常是深度神经网络）加上大量训练数据，期待机器学习到情感分析问题所需要的全部中层表示，第二种是使用简单的模型加上少量的训练数据，通过人工定义的中层表示来实现情感分类。第三种是直接使用底层特征，套上传统的分类器直接得到分类结果，因为情感分析问题十分复杂，直接使用这种方法效果不好。而第一第二种思路各有优缺点，第一种的优点是中层特征的抽取过程自动进行，无需人工干预，缺点在于情感分析是一个系统工程，需要大量的要想得到好的效果，需要大量的带标注训练数据和复杂的模型[32]。第二种思路是使用人工设计的中层特征，先在待分类的图像上检测出中层特征的存在，再根据中层特征得到最后的情感分类。这种思路的优势是模型简单，只需要单独训练每一个中层特征的检测器。训练过程简单并且可以并行化。这种思路的劣势在于需要设计一个覆盖面足够广并且有效的中层特征体系需要一些领域知识。

2.2.1 自动学习中层特征体系的图像情感分析技术

[32]文给出了一种自动学习图像情感的框架。 如下图所示，他们设计了一种渐进式优化的图像情感分析框架。

图2.2 渐进式优化情感分析结构

这种框架在五十万张粗标注的 Flickr数据集上进行了训练，之后为了训练中层特征，使用了少量人工标注的 Twitter 数据。

如下图所示， 传统的用于图像情感分析的卷积神经网络结构

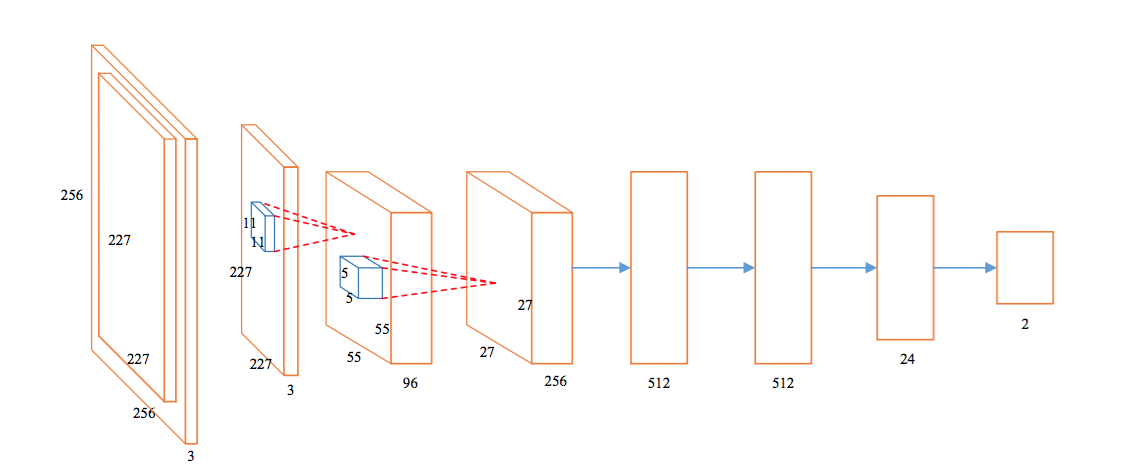


图2.3 常见的采用5个卷积层的卷积神经网络结构

在该结构中，每张待分类的图片被缩放为256\*256大小，然后中间227\*227的区域被裁剪出来输入到网络中。首先是两个卷积层，每个卷积层后都有一个 max-pooling 层。第一个卷积层有96个卷积核，每个卷积核的尺寸是11\*11\*3.卷积层进行卷积的步长是4.第二个卷积层有256个卷积核组成，每个卷积核的尺寸是5\*5，步长为2.

2.2.2 使用底层特征直接学习的情感分析技术

好的特征应具有不变性（旋转，尺寸，光照），和分类精度。在[27]文中，作者使用四种图像特征表示，它们分别是：

**颜色直方图：** 作者在 HSV 颜色空间内对每帧的颜色直方图进行计算。采用颜色直方图的原因是在以往的很多静态图片情感分析工作（例如[1][28]等等）中，颜色直方图都有应用。从宏观上来看一帧的颜色分布和情感倾向存在关联性。

**面部表情：** [1]文的作者认为，面部表情在 GIF 图片的情感分析中存在显著影响。作者使用的面部表情检测方法如下：构建一个输入层和隐含层都如图2.3所示的卷积神经网络，与图2.3不同之处在于，[1]文作者构建的卷积神经网络的输出层是一个具有平方合页损失函数的支持向量机模型。这个网络在28，708张48\*48的人脸图片训练集中进行训练，这些训练集分为七类，分别是：愤怒，恶心，恐惧，开心，沮丧，惊讶和中性。[1]文的作者使用 OpenCV提供的 基于Haar-like特征和AdaBoost方法的人脸检测工具进行人脸检测。用上文的方法进行面部表情特征的抽取结果是一个6维的向量。

**图片美学特征：** 前人的工作证明视觉材料的情感与图片美学特征之间存在着内在的联系[35]. [1] 文的作者提取了在[35]文中描述的图片美学特征的一个子集。每一个 GIF 帧被切分成3\*3的小单元。在每个小单元中提取暗通道，亮度，锐度，对称性，白平衡，色彩丰富度，色彩和谐度以及目光敏感度等特征。正规化后的帧中占主导地位的物体里图片中心的距离也被作为图片美学特征的一个方面提取.以上所有这些特征构成了一个149维的向量。GIF 图片的每帧需要提取一次。

**SentiBank中层特征：** SentiBank中层是由 Borth[28]等人提出的一个针对静态图像情感分析的中层特征，[1] 文的作者直接将静态图像情感分析的特征用于 GIF 情感分析。SentiBank 中层特征由1200维向量构成，该向量的每一个维度都是一个线性支持向量机的分类结果，表示的是该图像包含某一个“形容词名词对”概念的可能性。SentiBank 特征在[36][37]等情感分析任务中都表现的很好。根据[36]文，SentiBank 特征在受控测试集上识别形容词名词对的 F 值已经超过了0.6。

2.3 现有方法的主要问题

以上我们介绍了两种现有的视觉情感分析技术，

第一种技术基于深度神经网络（卷积神经网络）。希望能够通过深度神经网络自动学习特征的特点自动学习情感分析需要的中层特征。例如 GIF 视频中包含小孩摔倒，暴风雨天气等概念。该方法希望深度神经网络能通过训练数据学习到这些概念。然而，根据文献[38]如图2.4，图2.5，图2.6所示，目前五层卷积层能学到的还只是图片的局部，并不能学到类似于情感倾向等复杂的语义概念。从想要学习到情感倾向性特征需要更加复杂的模型和更多的训练数据。

然而，即使是如图2.3所示的常见的采用五个卷积层的卷积神经网络在训练时也会遇到标注训练数据不足的问题。如果应用到 GIF 情感分析问题上，增加网络中隐含结点的数量势必需要更多的训练数据。更复杂的模型训练过程也会更加容易出现问题。但是，虽然现有的卷积神经网络无法可靠学习到 GIF 视频的所有中层特征以及中层特征到情感倾向的关系。如果设计好了可靠的中层特征，我们依然可以使用卷积神经网络来从 GIF 视频中检测它们。具体细节我们将在第三章中探讨。此外，这种方法并没有考虑帧和帧之间的时序关系。

图2.4 第一第二卷积层学习到的特征表示示例

图2.5 第三卷积层学习到的特征表示示例

图2.6 第四第五卷积层学习到的特征表示示例

(以上三张图片来源：Matthew D. Zeiler and Rob Fergus, Visualizing and Understanding Convolutional Networks, ECCV 2014, Part I, LNCS 8689, pp. 818–833, 2014)

第二种技术基于底层特征（颜色直方图，面部表情，美学特征）或静态图像的中层特征（SentiBank）直接套用常见分类器进行情感预测。对于底层特征，这些人工选择的底层特征虽然与情感倾向存在一定关联。但是对于内容包罗万象的 GIF 视频来说，单纯依靠底层特征肯定是不够的。对于SentiBank中层特征，虽然在静态图像的情感分析中表现不错，但是GIF 动画是视频帧的时间序列。具有诸如摔倒，跳跃，舞蹈等动作信息，仅仅依靠形容词名词对不能很好的进行描述。同时使用底层特征和中层特征组合进行描述会由于不同特征的维度相差太大导致分类器对某些特征“忽略”的问题。

由于形容词和名词对不能很好的描述 GIF 视频中出现的中层概念，我们提出了动词名次对的表示方法。我们将形容词名词对和动词名次对合称为情感对。使用情感对来描述 GIF 视频中出现的中层概念具有描述准确，可检测性好等特点。具体内容将在第三章中进行阐述。和第一种方法一样，这种方法也没有考虑 GIF 动画前后帧之间的时序关系。

2.4本章小结

在本章中，我们首先分析了 GIF 情感分析任务的两个基本特点。即 GIF 动画中语义信息和概念的分布范围很广，同时GIF动画是一个视频帧的时间序列。在这两个基本特点的基础上提出了 GIF 情感分析任务的两个核心问题：第一个核心问题是如何确定哪些 GIF 视频中的概念语义需要被处理。第二个核心问题是如何处理 GIF 视频中概念语义之间的时序关系。

之后，介绍了两种目前常见的视觉情感分析方法，第一种方法使用深度学习直接学习情感分类，然而，由于情感分析问题比较复杂，使用7层网络的效果并不能直接学到。同时卷积神经网络也不能处理时序信息。第二种方法采用颜色直方图，面部表情，美学特征等底层特征联合图像情感分析领域的“形容词名词对”中层特征。然而，这些特征并不足以描述 GIF 视频中分布广泛的语义和概念。不满足第一点要求，该方法也没有考虑时序。在这个基础上，我们将在第三章介绍我们的方法，通过构建 GIF 情感语义体系结构来解决第一个问题，通过构建情感对序列来解决第二个问题。

第三章 GIF 情感分析概念语义体系

前文讲到，GIF 视频有两个特点， 第一，视频内容包罗万象。没有弄懂这些内容的语义之前无法分析 GIF 情感。第二，视频中各个概念的语义信息存在时序关系，概念的顺序会影响最终的情感判断。而从第二章的内容我们知道现有的 GIF 情感分析方法没有处理好这两个特点，导致分类效果不佳。因此，一个好的GIF 情感分析方法应该既能够考虑GIF 视频中包罗万象的概念语义，例如美丽的夕阳，跌倒的儿童，又能对概念之间的时序关系进行建模，例如哭泣的儿童在前，欢笑的儿童在后。由此，我们引出了本文的主要工作:

1. 构建了GIF 情感分析概念语义体系。
2. 设计了GIF 情感分析时序模型。

3.1 GIF情感分析概念语义体系的构建

在这部分工作中，我们的目标是尽可能全面的概括在 GIF 视频中可能出现的导致情感倾向变化的概念语义。为了达到这个目标，我们首先要定义概念语义体系中概念的标准。

3.1.1 概念语义体系的评价标准

**情感相关性：**概念语义体系中的概念将被用于 GIF视频情感分析，因此只有和情感相关的概念才应该被纳入概念体系。

**广泛性：**前文提到，GIF 视频中的概念是包罗万象的，因此概念语义体系中的概念来源也应该来自于生活的方方面面。这样才能保证对任意的含有情感倾向的 GIF 视频，都能找到合适的概念来描述它。

**可检测性：**我们构建概念语义体系的目的是划定 GIF 情感识别中所需要的概念框架。因此概念语义体系中概念应该是具体的可检测的**。**

3.1.2 概念语义体系的构建

在定义了上面三个概念标准之后，受到 Borth.等人工作[37]的启发，在静态图像情感分析领域，使用“形容词名词对”可以同时符合以上三个标准。但是，在动态视频情感分析领域，“形容词名词对”在广泛性上有所欠缺，这主要体现在 GIF视频中有很多动作信息，使用形容词名词对不能很好描述。如下图所示，图的左边部分是我们提出的情感对的检测结果，右侧是使用现有的形容词名词对的检测结果。绿色方框中的图片用情感对可以准确描述，却不能用形容词名词对准确描述。例如第二排第三张图片中“狂奔的蜗牛”。需要动作信息才能表现视频的内容。红色圆圈中的图片为我们分类器错误分类的样本，第三行第一列那个样本对应的“情感对”为“许多人”，“行走的人”。第三行第五列中只用“形容词名词对”描述时，只有“许多人”的信息。因为信息的缺失导致分类出错。



图3.1 ANP与 SentiPair 的比较

3.1.3 形容词名词对与动词名词对

为此，我们提出了动词名词对的概念。与[37]文中的形容词名词对类似，动词名词对包含一个动词和一个名词，它描述的是一个具体的动作和动作的对象。为了表述方便我们将形容词名词对和动词名次对合称为“情感对”。

图3.2 情感对的构成

至此，我们有了比较好的表示 GIF 情感分析中概念语义的方法。但是还缺少层次体系结构。在自然语言处理领域，我们参考了[39][40]的工作.我们引入了WordNet作为我们概念语义体系结构的基础。WordNet是由普林斯顿大学提出的一个词典项目。与常见的词典项目不同，WordNet不仅把单词以字母顺序排列，而且按照单词的意义组成一个“单词的网络”。普林斯顿大学的心理学家，语言学家和计算机工程师们将 WordNet设计成一个覆盖范围宽广的英语词汇语义网。在这个语义网中包含了大量的概念语义。具有作为 GIF 情感分析概念语义体系的先决条件。

但是，仅仅使用 WordNet 的语义网络不能满足我们的要求。主要原因是在 GIF 情感分析的过程中，WordNet 中的很多词表示的概念是很少在 GIF 动画中出现的。因此，我们需要从 WordNet 的网络中筛选出在 GIF 动画中经常出现的概念。同时，构造“形容词名词对”，“动词名词对”的过程中，我们只需要 WordNet 中的形容词，动词和名词三种类型的词。其他的例如副词，介词，助词均需要删除。

图3.3 构建出的概念语义体系示意图

3.2 概念语义的筛选

如上文所述，我们选择了 WordNet 作为概念语义体系的基础。在 WordNet 的基础之上，我们删除了数词，介词，副词等和“情感对”无关的概念语义词。同时，我们还需要从 WordNet 中筛选在 GIF 动画中经常出现的概念语义项。

筛选的目标有两个：

1. 找出情感倾向性强的概念语义。在概念语义体系中不同的概念语义所含有的情感倾向是不同的，因此，我们需要筛选出那些情感倾向性强的概念语义。

2. 找出 GIF 视频中经常出现的概念语义。因为那些很少出现在 GIF 视频中的概念语义项目会增加概念语义的检测难度，并且这些项目对 GIF 情感分析没有什么贡献。

目前条件下，我们无法对概念语义体系中所有的概念语义进行检测，我们的实验也证明，不同筛选方案会对情感对检测的准确率产生很大影响。

3.2.1 SentiWeight情感权值

为满足第一个筛选目标，我们设计了 SentiWeight情感权值。构建过程如下：在 SentiWordNet中，每个词的情感倾向性(SentiScore)分为若干个等级，如下图所示。正数表示积极情感倾向，负数表示消极倾向。SentiWordNet的情感权值绝对值越大表示情感倾向越强烈。

SentiScore 就是 SentiWordNet 中该概念语义的情感得分，显然，SentiWeight 的取值范围是[0,1]。

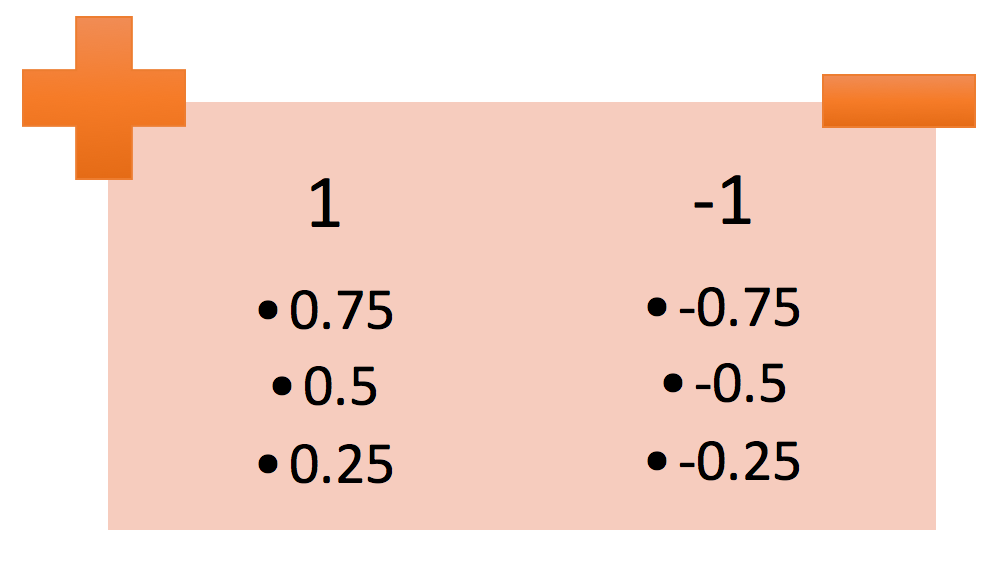


图3.4 SentiWordNet 中的情感权值

3.2.2 GiphyWeight 情感权值

为满足第二个筛选目标，我们设计了 GiphyWeight 情感权值。构建过程如下：我们在GIF 视频网站Giphy.com 中搜索情感词，统计 Giphy.com 搜索结果中 GIF 图像的数量Count, 每一个概念语义的GiphyWeight 权值依据下式得出：

在上式中，Counti 是第 i 个概念语义在 Giphy.com 中对应的 GIF 动画个数。分母则是全部概念语义对应的 GIF 动画个数的最大值。因此，GiphyWeight 情感权值的取值范围是[0,1]。为了得到 Counti 的数据，我们爬取了 GIF 收藏网站 Giphy.com 中的 GIF 标注（tag）.如下图所示：

图3.5 Giphy.com 网站

如上图所示， Giphy.com 是一个 GIF 动画的收藏网站。根据网站介绍， 它包含超过十万张 GIF 动画，同时，对每张 GIF 动画，会有一个标签描述 GIF 动画的内容。例如，第一列第二张 GIF 动画上一个小女生在玩呼啦圈，在转过几圈之后呼啦圈掉在了地上，而这张 GIF 动画对应的标签有三个，第一是失败，第二和第三分别是呼啦圈的缩写和呼啦圈。我们使用爬虫技术采集了 Giphy.com 中的英文标签。因为这些标签都是描述 GIF 内容的，我们使用这个集合对 WordNet 进行筛选，保留在 Giphy.com 的标签中出现数量多的词语。

图3.6 Giphy.com 标签中概念语义词分布（筛选出10000次以上的词）

由于搜索技术的限制，我们删除了其中词语长度小于3的词，因为这些词会匹配到其他无关的语义项。

3.2.3 筛选权值 FilterWeight

在得到 SentiWeight 情感权值和 GiphyWeight 情感权值后，我们根据下式计算出一个筛选权值 FilterWeight:

显然，由于 SentiWeight情感权值和 GiphyWeight 情感权值的取值范围均为[0,1], 筛选权值 FilterWeight 的取值范围为[0,1]. 由于 SentiWeight 情感权值描述的是概念语义的情感丰富程度，GiphyWeight 情感权值描述的是概念语义在 GIF 视频中分布的广泛程度，按照上式得出的筛选权值既能描述概念语义的丰富程度，又能描述概念语义在GIF 视频中出现的概率。我们对筛选权值定义了一组阈值，下表展示的是不同阈值下筛选出的 SentiPair 数。

表3.1 不同阈值下筛选出的“情感对”数

|  |  |
| --- | --- |
| 阈值 | 筛选出的“情感对”数 |
| > 0.7 | 31463 |
| > 0.8 | 5111 |
| > 0.9 | 1274 |

我们提出的GIF 情感分析感念语义体系是一个词的森林，该森林包括形容词树，名词树，动词树三棵树。在这三棵树的基础上，我们可以构建出形容词名词对和动词名词对的集合。

3.3 概念语义体系用于挖掘情感相关性

上文提到，我们提出的概念语义体系为 GIF 视频中出现的概念语义划定了范围。实际上，作为语义体系，概念语义体系的另一个功能是可以用于挖掘情感相关性。如下图所示。 GIF 视频内容是三只可爱的金毛犬，但是金毛犬作为犬类的一种，我们可以将“可爱的金毛犬”这个情感对扩展成概括性更好的“可爱的狗”和“可爱的动物”。这种高层抽象的能力是传统基于底层特征的情感分析方法所不具备的。我们在训练情感对检测器的时候使用了这种相关性。如果情感对检测器检测到了某一个概念语义，例如“可爱的金毛犬”，我们会对检测结果进行扩展，将“可爱的金毛犬”扩展成“可爱的动物”。



图3.7 概念语义体系用于挖掘情感相关性的示例

3.4 情感对检测器的训练

在构建好GIF 情感分析概念语义体系后， 我们要开始训练情感对的检测器。这些检测器的输入是 GIF 动画，输出是情感对的概率值. 而这些概率值将作为中层特征的特征值。

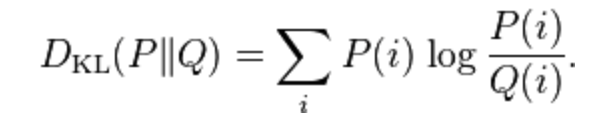
3.4.1 多任务情感回归

多任务学习又被称为多示例学习或多标签学习。传统的单标签分类学习是从一个只属于一个标签l的样本集合中学习，其中每一个标签属于一个互斥的标签集合L 。

在多标签分类中，每个样本属于一个L样本集合的一个子集。在过去，多标签分类由文本分类和医学分析而产生和推动的。现在，我们发现现代的许多应用对多标签分类方法需求持续增长，比如蛋白质分类，音乐归类，和语义场景分类。

传统单任务学习方法（SVM,LR 等等）忽略了每个分类项目之间的关联性。在 GIF 视频情感分析中, 检测到的概念语义存在关联性。例如“惊讶的男孩”实际上也可以是“惊讶的人”，“晴朗的天空”也可以是“美丽的景色”。因此，我们的情感对检测器应该是一个“多任务”回归问题。

我们在多任务情感分析中使用的损失函数是交叉熵损失函数。使用 KL 距离来计算标签与分类结果之间的相似度。



对于两个离散型分布 P,Q，KL 距离可以用上式进行计算。

3.4.2 模型选择

为了选择一个合适的模型进行情感对的检测，我们考察了多种方法。一种途径是类似于Borth [37]等人的方案，对每一个中层概念语义，也就是“情感对”训练一个支持向量机分类器。这种方法的优点是检测算法的鲁棒性取决于多个独立的检测器，某一个检测器性能不佳并不会造成全局的问题。但是缺点也很明显，由于我们的中层特征维数很高，需要训练大量的中层检测器，运行效率也会受影响。

受到深度学习尤其是卷积神经网络在视觉识别方面的巨大成功的启发[41]，我们选择了卷积神经网络作为我们情感对检测器的检测模型。正如前文所述，深度学习不适合用于直接学习 GIF 视频的情感倾向，因为他不能学习到情感分析所需要的全部中层特征，但是七层卷积网络学习视觉底层特征的能力已经被大量研究证实[41][42][43]。因此我们使用卷积神经网络来进行中层特征分类器的训练。所用网络结构如下图所示：

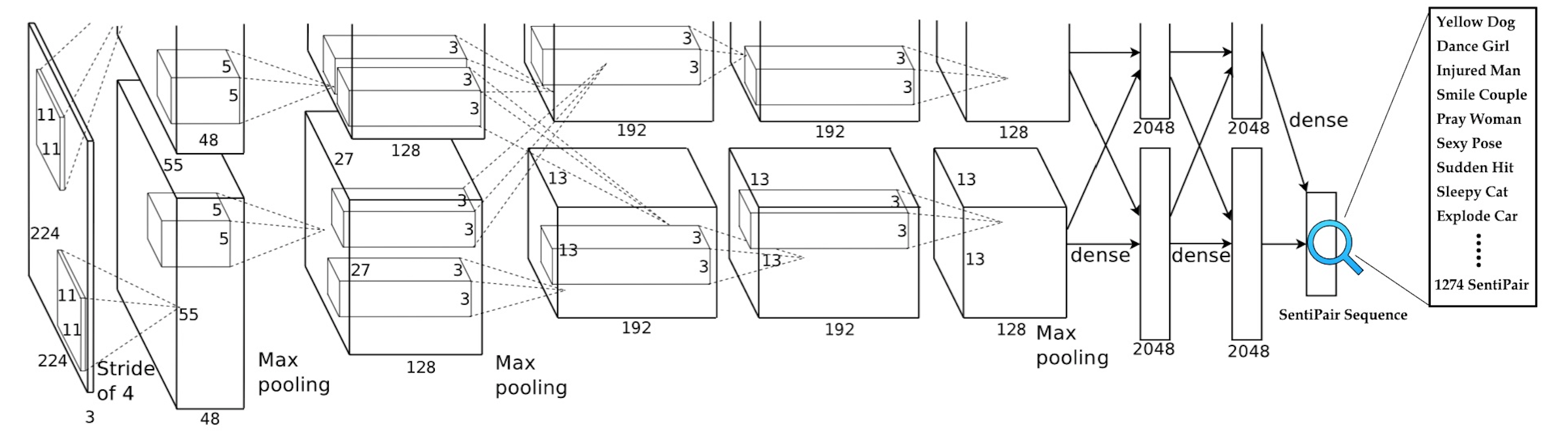


图3.8 我们所用的网络结构

第一个卷积层含有96个卷积核，每个卷积核的尺寸为11\*11\*3，扫描的步长为4像素。第二个卷积层的输入是第一层的输出，拥有256个5\*5\*48的卷积核，第三第四第五个卷积层都是互相直接连接，之间没有加入 Max Pooling 或正则化。 第三个卷积层有384个卷积核，每个卷积核的尺寸为3\*3\*256.第四个卷积层有384个卷积核，每个尺寸为3\*3\*192，第五个卷积层有256个卷积核，每个尺寸为3\*3\*192，两个全连接层拥有4096个神经元。

如上图所示，我们主要对传统卷积神经网络的最后一层做了修改。将原先的Softmax 层改为了 Multi-Task 约束。网络的输入时150，528维，每一层的神经元个数分别为：253，440-186，624-64，896-64，896-43，264-4096-4096-1274。

我们训练的情感对检测器的性能如下图所示，横坐标是不同的实验配置，纵坐标是情感对检测器的准确率。横坐标中 TopK 指的是在分类的结果中，分类置信度最高的前 K 个结果用于计算准确率。横坐标中 Relative 指的是使用我们的概念语义体系进行关联性挖掘。横坐标中 MTL 指的是使用多任务损失函数。在图中的三种颜色代表三种不同筛选结果。从左至右分别使用0.7，0.8和0.9作为筛选阈值。从实验结果中我们发现，大体上筛选阈值越高，分类精度越高，当然，筛选过多会造成中层特征的表达能力下降。此外，TopK 中，K 值越大，准确率越高。值得注意的是，对于低阈值的筛选结果，加入相关性挖掘后，检测子的准确率反而下降了，这可能是由于低阈值的类别数过多，分类准确率太低，相关性挖掘后扩展的词往往和真实的“情感对”相差甚远。

也就是说，加入 Multi-Task 后，对所有的分类类别数，情感对检测器的检测效果均有提高。其中在1274个情感对的情况下，提高的更明显（3%）。加入情感相关性之后，大部分分类类别数都可以提升效果，但是在分类类别数较高的情况下分类准确率反而下降。这是因为加入关联性后，扩展的概念语义是和分类结果直接相关的概念语义，如果本身的分类准确率不高，加入相关性约束后会适得其反。此外，不同筛选阈值的筛选结果对分类结果的影响也是显著的。

图3.9 SentiPair 数目对检测器准确率的影响

3.5 本章小结

在本章中，我们致力于解决GIF视频情感分析中概念语义分布的广泛性问题。通过提出 GIF 情感分析概念语义体系，我们为GIF 情感分析中的概念语义划定了范围。该体系的完整度由已经被广泛验证过的 WordNet 保证。同时，由于 GIF 视频内容的特色，直接照搬 WordNet 体系结构显然是不合适的。需要对 WordNet 进行筛选。为此我们提出了两条筛选原则，一是概念语义要有足够的情感丰富性，二是概念语义在 GIF 视频中出现的频率要足够高。为了满足这两条原则，我们提出了 SentiWeight 情感权值和 GiphyWeight 情感权值。在概念语义体系的基础之上，我们使用深度神经网络训练了“情感对”检测器，实验表明，使用多任务学习并利用概念语义体系结构挖掘情感相关性后，检测性的准确率得到了大幅度提升。

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。

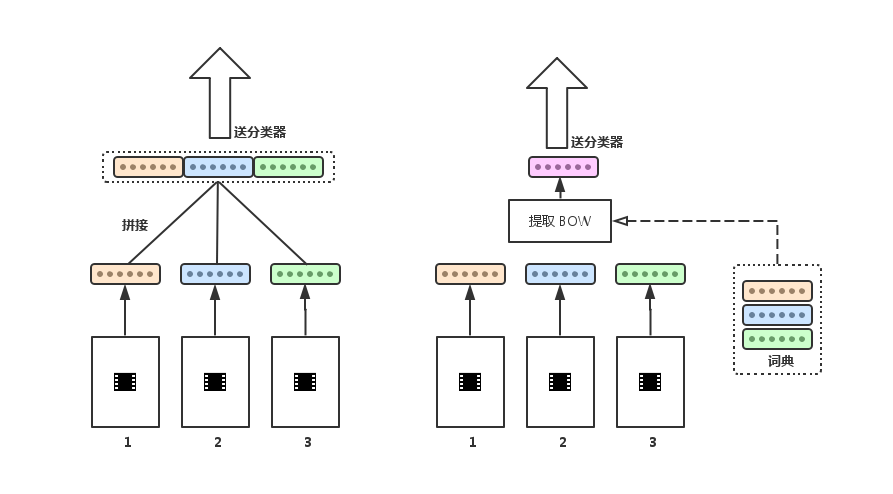


图4.1， 处理时序信息的两种简单形式

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。

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：

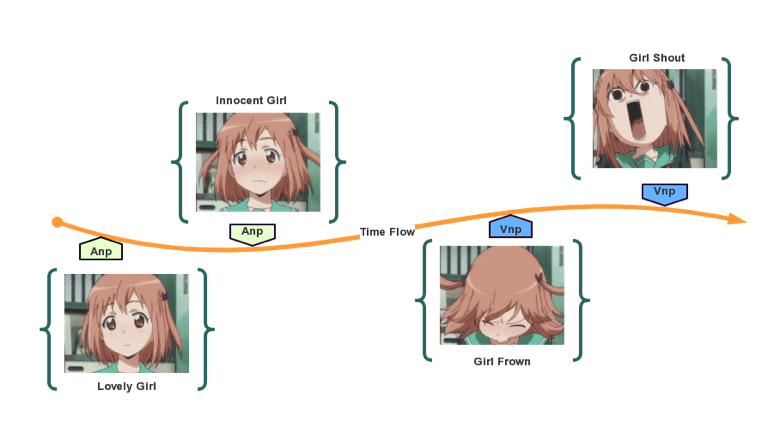


图4.2 情感对序列示意图

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：

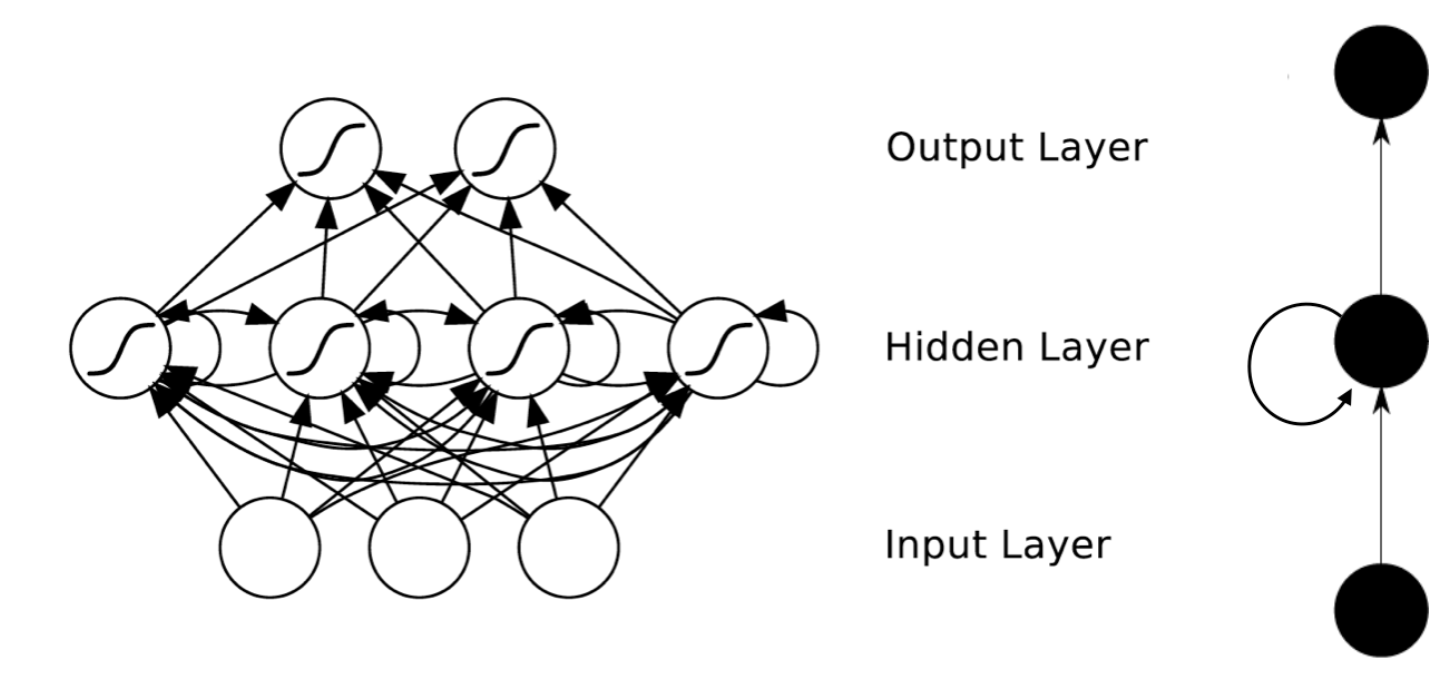


图 4.3 典型的循环神经网络（RNN）

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：

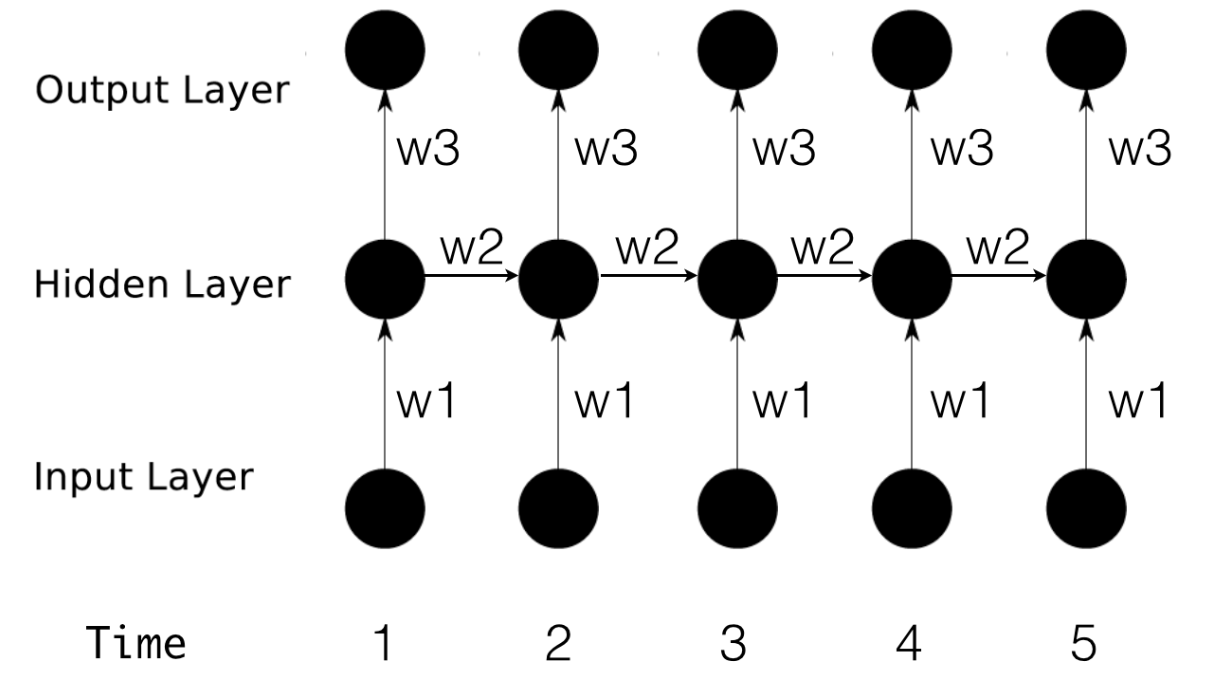


图4.4 展开连接后的循环神经网络

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。

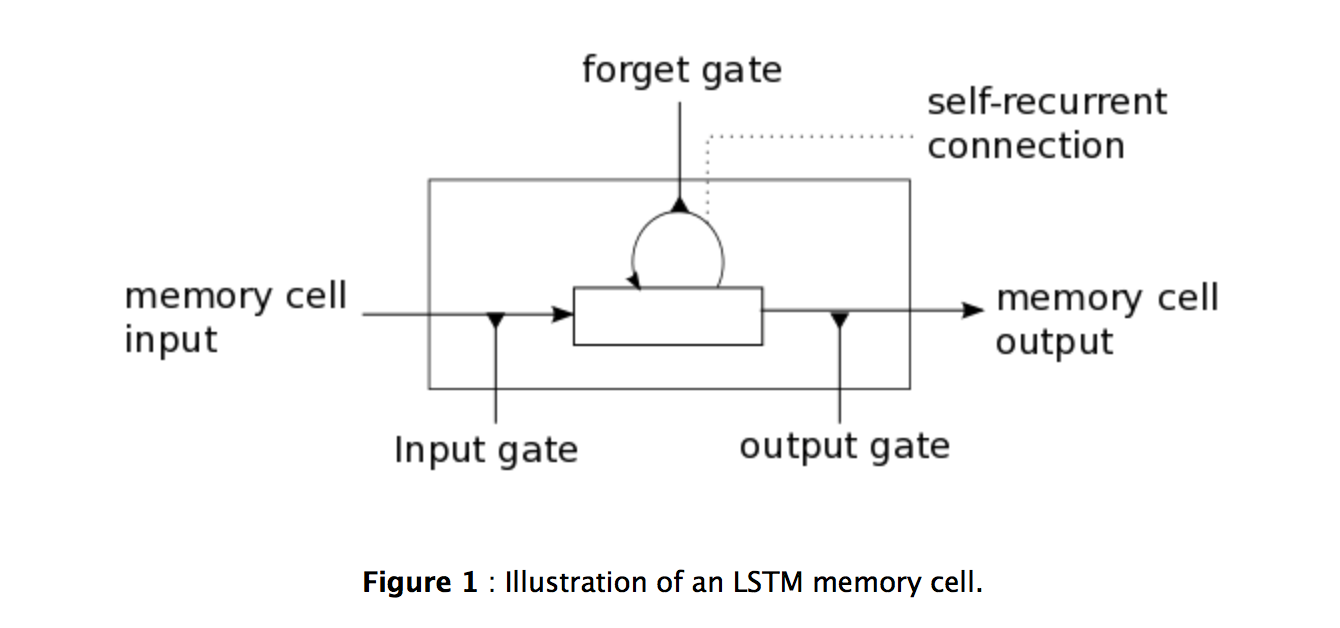


图4.5 LSTM结构示意图

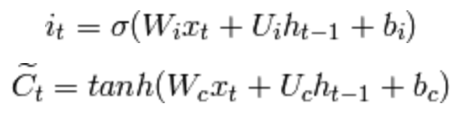
4.3.3 长短期记忆神经元的学习方法

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 :



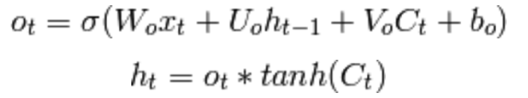
接下来，计算遗忘门在 t 时刻的激励ft：



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：

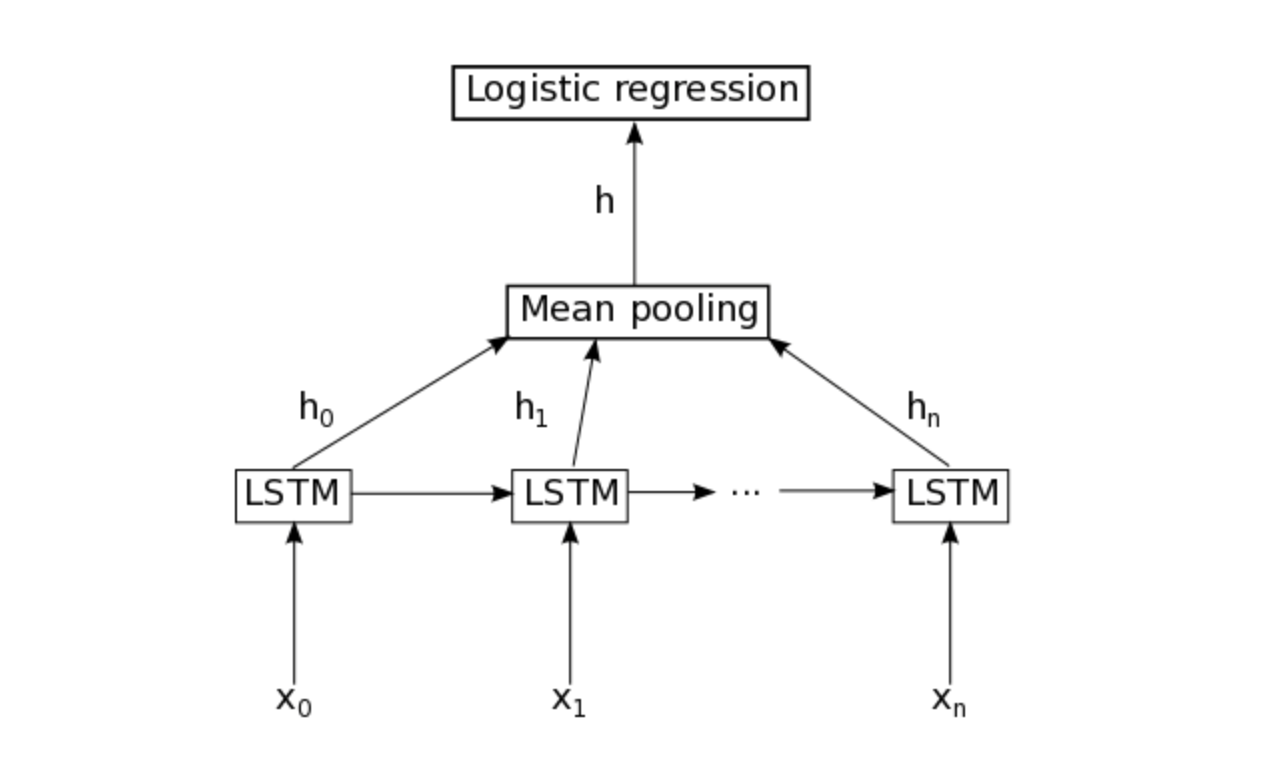


图4.6 情感对序列分类中使用到的网络结构

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。

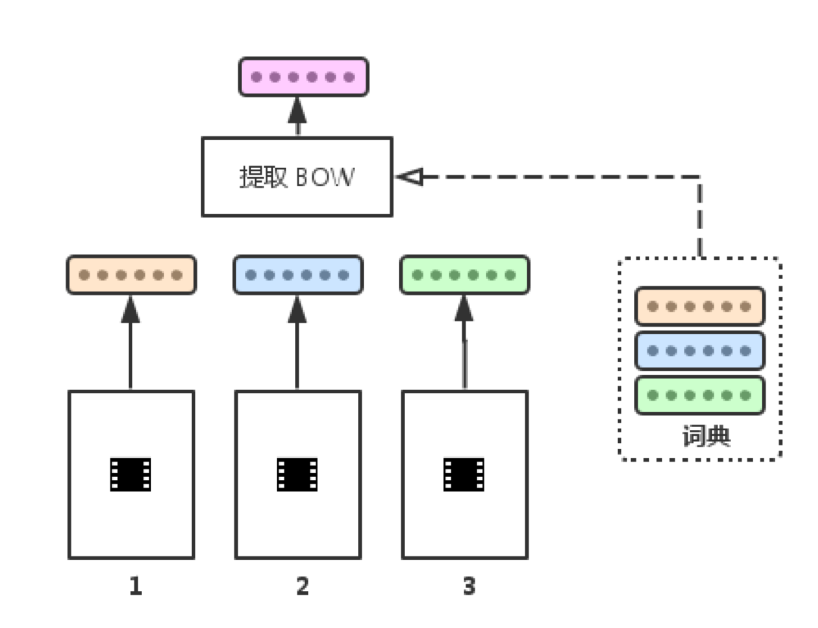


图4.7 基于词袋特征的 SMO 模型

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4.1 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4.2 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4.8 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4.9 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4.10 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4.11 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。

本章小结

在本章中，我们的目的是解决 GIF 情感分析所必须解决的第二个问题—处理 GIF 视频中概念语义之间的时序关系。针对第二个问题，我们在分析传统 GIF 视频情感分析方法之后，提出了被称为“情感对序列”的 GIF 视频情感分析时序表达。“情感对序列”相比以往的时序分析方法具有很多先天优势，例如，可以描述连续的动作信息，可以描述不同状态的转变等等。为了利用“情感对序列”来预测最终情感倾向，我们使用了带长短期记忆单元（LSTM） 的循环神经网络（RNN） 作为分类模型，该模型同样在 GSO-2015数据集中进行训练。实验结果表明，无论是与统计学习方法（如 SVM,SMO,Logistic Regression，Navie Bayes）还是与传统循环神经网络相比，我们的模型都有着更优秀的性能。“情感对序列”和对应最终情感检测器的提出解决了 GIF 情感分析的第二个核心问题。

第五章 微博 GIF 情感分析数据集和人脸表情检测器

前文提到，为了解决 GIF 情感分析的概念语义体系缺失的问题，我们提出了“情感对”以及一套基于“情感对”的概念语义体系；为了解决 GIF 情感分析的时序表示问题，我们提出了“情感对序列”；针对情感对序列的检测，我们提出了基于卷积神经网络的一套检测方法；针对情感倾向性的检测，我们提出了基于带长短期记忆单元的循环神经网络的一套检测方法；在本章中，我们首先通过实验验证我们提出的概念语义体系在 GIF 情感分析中的合理性；然后通过实验验证我们提出的情感对序列在 GIF 情感分析问题中的合理性；之后，我们通过实验验证了引入人脸识别后的效果；最后与现在最好的 GIF 情感识别方法进行了对比。

首先，我们介绍实验中用到的数据集，GIF 情感分析问题目前没有统一的数据集可以参考，因此，我们采用爬虫的方法从社交网络（主要是新浪微博）中采集了

5.1 GSO-2015数据集

GSO-2015数据集是我们从新浪微博中爬取到的一个 GIF 视频数据集，包括四万余张 GIF 动画，在这些视频中，我们人工标注了六千余张，其中有1700余张是有明确情感倾向的，其余的4300余张要么情感倾向受 GIF 动画中的文字影响，要么无法判断倾向。我们的人工标注信息我们的实验主要在1700余张有明确情感倾向的数据中进行。下表列出了 GSO-2015数据集的详细信息：

|  |  |
| --- | --- |
| 表5.1 GSO-2015 数据集概要 | |
| 总GIF视频数 | 41092 |
| 人工标注过的 GIF 视频数 | 6177 |
| 有明确情感倾向的 GIF 视频数 | 1740 |
| 积极情感倾向 GIF 视频数 | 1059 |
| 中性情感倾向 GIF 视频数 | 551 |
| 消极情感倾向 GIF 视频数 | 130 |

图 5.1 GSO-2015数据集

在 GSO-2015数据集中，我们不仅标注了 GIF 视频的最终情感倾向，而且还标注了 GIF 视频中含有的情感对信息,按照 GIF 视频帧中出现情感对的先后顺序，我们标注了 GIF 视频帧的情感对序列。如表5.2所示，是GSO-2015数据集的标注示例。

表5.2 GSO-2015标注示例



5.2 GIF 视频情感分析在线标注系统

为了方便 GSO-2015数据集的人工标注工作，我们制作了一个在线标注网站。如下图所示。

网站可以从数据库中读取一张待标注 GIF 视频展示给标注者。同时，标注者可以从现有的概念语义中选择合适的词构成情感对序列，并确定最终GIF 视频的倾向性。

图5.2 GIF 视频情感分析在线标注系统示意图

图5.3 情感对序列标注模块示意图

图5.4 概念语义选择界面示意图

图5.5 最终情感倾向标注模块示意图

5.3 补充内容：使用人脸表情识别提升情感对检测器的性能

前文提到，为了表示 GIF 视频中出现的概念语义，我们提出了一个概念语义的中层表示—“情感对”。为了解决 GIF 视频中概念语义的分布问题，我们构建了一个概念语义体系。在 GIF 视频情感分析的过程中我们检测概念语义体系中的概念，得到一个基于“情感对”的中层特征。并进一步在这个中层特征的基础上训练最终情感极性的分类器，从而完成 GIF 视频情感分析的任务。

显然，在我们的流程中，“情感对”检测器的检测性能决定了中层特征的提取准确度。精确提取中层特征也能帮助后续的情感极性检测器更好的工作。考虑到卷积神经网络在视觉方面的巨大成功，我们采用了它作为我们“情感对”检测器的模型。虽然采用了诸如“情感相关性挖掘”，“多目标学习”等手段提高检测精度，总体的“情感对”检测器的检测性能也不够理想。尤其是对于人脸表情这种在 GIF 视频中高频率出现的概念语义，检测器经常会出现错判的情况。

受到其他研究者的启发[15]，针对某些经常出现的中层特征，使用额外的数据单独训练分类器要比“多目标学习”性能更优。因此，我们对人脸表情单独训练了分类器。我们的训练模型和第三章介绍的基本一致。人脸检测器采用的是经典的基于 Harr-like 特征的 AdaBoost 模型，训练数据采用的是 Kaggle 人脸表情识别竞赛的数据集，训练数据集共有28710张人脸，按照表情分为七类，分别是：Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral. 与第三张中“多目标学习”不同，由于类别数目少并且类别唯一，我们采用的是 Softmax 损失函数，五折交叉检验的识别精度达到63.1%。

图5.6 人脸表情检测器演示

上图是人脸表情识别的效果演示，当摄像头捕获一帧画面时，送人脸检测器检测出人脸位置，之后将人脸位置的截图送入卷积神经网络，网络的输出是人脸表情类别。从而完成人脸表情识别工作。在 GIF 情感分析中，原有的“情感对”检测器和人脸表情检测器同时工作。原有“情感对”检测器负责检测除“Angry Face, Happy Face, Neutral Face, Disgust Face, Fear Face，Sad Face, Surprise Face” 以外的概念语义， 两个检测器的检测结果拼接后得到 GIF 视频的检测结果，送入最终情感检测器进行情感检测。

5.4 本章小结

在本章中我们主要介绍了 GSO-2015数据集以及第三章中介绍的“情感对”检测器的补充部分—人脸表情识别检测器 。在介绍 GSO-2015时，我们介绍了数据集的组成部分，数据集使用“情感对序列”标注的原理以及数据集的采集方法。GSO-2015数据集有4万余张 GIF 动画组成，其中带标注的视频数有6千余张。全部通过人工标注了情感极性和情感对序列。作为第三章中介绍的“情感对”检测器的补充部分，人脸表情识别检测器具有与原始模型相似的结构，通过在 Kaggle 数据集中进行训练，达到了较好的效果。人脸表情识别检测器的加入，丰富了“情感对”检测器的内容，也体现了我们提出的情感分析框架的可扩展性。随着未来各项识别技术的进步，我们的识别准确率也可以不断提高。

第六章 总结与展望

6.1 全文总结

本文研究的问题是社交网络中 GIF 视频的情感分析问题，GIF 视频情感分析问题与静态图像情感分析问题有共性也有个性。共性体现在两个问题都需要处理视觉信息中隐含的概念语义，个性体现在 GIF 视频情感分析问题还需要处理帧与帧，概念语义与概念语义之间的时序关系。因此，为了处理好 GIF 视频情感分析问题，我们需要解决两个核心问题。第一个核心问题是如何确定哪些 GIF 视频中的概念语义需要被处理。第二个核心问题是如何处理 GIF 视频中概念语义之间的时序关系。

针对第一个问题，我们首先提出了一个 GIF 视频中出现的概念语义的表示方法--“情感对”。之后，我们建立了一个概念语义的体系结构。该体系结构中所有的概念语义都应该被情感分析问题所考虑。随着概念语义体系结构的建立，我们还要构建一系列的情感对检测器来检测可能出现的概念语义。我们选择了加入情感相关性挖掘和多任务学习目标函数的卷积神经网络作为检测模型。在我们采集并标注的 GSO-2015 数据集中进行训练。概念语义体系结构和对应的检测器的提出解决了 GIF 情感分析的第一个核心问题。

针对第二个问题，我们在分析传统 GIF 视频情感分析方法之后，提出了被称为“情感对序列”的 GIF 视频情感分析时序表达。“情感对序列”相比以往的时序分析方法具有很多先天优势，例如，可以描述连续的动作信息，可以描述不同状态的转变等等。为了利用“情感对序列”来预测最终情感倾向，我们使用了带长短期记忆单元（LSTM） 的循环神经网络（RNN） 作为分类模型，该模型同样在 GSO-2015数据集中进行训练。实验结果表明，无论是与统计学习方法（如 SVM,SMO,Logistic Regression，Navie Bayes）还是与传统循环神经网络相比，我们的模型都有着更优秀的性能。“情感对序列”和对应最终情感检测器的提出解决了 GIF 情感分析的第二个核心问题。

此外，由于情感对检测问题十分复杂，仅仅使用单一模型很难确保准确率，我们针对 GIF 视频中人脸表情出现数量多这一现状，对人脸表情单独训练分类器，实现了较好的分类效果。

6.2 未来工作展望

GIF 视频情感分析问题是一个很有挑战性也很有现实意义的问题，我们通过提出 GIF 视频情感分析中的两个核心问题将任务规划为了两个小问题，并分别进行了解决。然而，对于第一个问题的解决过程还存在很多不完美之处，多目标学习加上情感关联性挖掘虽然可以提高检测的准确率，但巨大的类别数还是导致准确率只有6%左右。虽然在后期我们加上了人脸表情检测器作为“情感对”检测器的辅助，但一定还有更多我们没有发现的“情感对”需要单独训练检测器。同时，使用卷积神经网络来检测多帧视频中的动作也不甚合理。不过，令人欣慰的是，我们的工作在 GIF 视频情感分析领域具有通用性，底层“情感对”检测器的检测精度可以通过“动作识别”，“视觉信息检索”，“行人检测”等计算机视觉领域的进步来不断完善。我们提出的“情感对序列” 这一中层特征表示也将在未来的 GIF 视频情感分析工作中发挥作用。