基于深度学习的情绪轮盘建模及应用

摘要：目前已有的情绪分类模型都是基于表情分类进行一维结果的单输出，这种输出并不能很形象、丰富地传达情绪，只是浮于表情这一概念的表面。因此将认知心理学的理论与与机器学习模型相结合以产生更加丰富、能够满足实际需求的表情分类结果非常重要，这也是本文提出将普洛特契克的情绪轮盘模型通过深度学习的方法运用到情绪识别领域的原因。本文设计的情绪识别模型的输出结果为一个四元向量，能够识别总计 81 种情绪状态，对应于几百个情绪词，几乎覆盖了人类全部的情绪。该模型相比传统情绪识别模型，识别结果更准确、覆盖面更广，能够满足于更广泛的应用场景，包括但不限于文本挖掘、演技评估、智能推荐、心理检测等等。

关键词：情绪轮盘 情绪识别 深度学习 人脸识别

Abstract: At present, the existing models of emotion classification are based on the single output of the one-dimensional result of facial expression classification, which can not convey the emotion vividly and widely. Therefore, it is of paramount importance to combine the theory of cognitive psychology with machine learning model to produce more abundant and practical expression classification results This is also the reason why we hope to apply the Emotion Wheel Model (Plutchik,1980) to the field of emotional recognition through the method of deep learning. The output result is a quaternion vector, covering 81 kinds of emotional states, corresponding to hundreds of emotional words, covering almost all human emotions. Compared with the return value of a single result, it has a wider coverage and more accurate result, which is satisfied for application in a wider region. The application scenarios include not limited to text mining, performance evaluation, smart recommendation, psychological examination, etc.

第一章 绪论

1.1 研究背景

1.1.1 普洛特契克情绪轮盘模型

1980 年，罗伯特·普洛特契克绘制并提出了情绪轮盘模型，把情绪定义为 4 个维度、8 种两两对立的基本情绪（快乐对悲伤、惊讶对预期、恐惧对生气、信任对厌恶）的组合[1]。

8 种基本情绪经过两两组合可以得到一级组合情绪、二级组合情绪和三级组合情绪。相邻的基本情绪两两组合可以构成一级组合情绪，相隔一位的基本情绪可以组成二级组合情绪，相隔两位的基本情绪可以组成三级组合情绪，对立的两种情绪不能相互组合（组合情绪的具体情况见见表 1.1，其中细体词表示基本情绪，黑体词表示组合情绪）。其中一级组合情绪出现的频率最高，人们产生三级组合情绪的情况最少。

表 1.1 基本情绪的组合

一级情绪	二级情绪	三级情绪	相反情绪
快乐 信任 **爱**	快乐 害怕 **愧疚**	快乐 惊奇 **欣喜**	快乐 悲伤 二者冲突
信任 害怕 **服从**	信任 惊奇 **好奇**	信任 悲伤 **多愁善感**	信任 厌恶 二者冲突
害怕 惊奇 **惊觉**	害怕 悲伤 **绝望**	害怕 厌恶 **羞耻**	害怕 生气 二者冲突
惊奇 悲伤 **失望**	惊奇 厌恶 **怀疑**	惊奇 生气 **愤慨**	惊奇 期待 二者冲突
悲伤 厌恶 **悔恨**	悲伤 生气 **嫉妒**	悲伤 期待 **悲观**	
厌恶 生气 **蔑视**	厌恶 期待 **愤世嫉俗**	厌恶 快乐 **病态**	
生气 期待 **有攻击性**	生气 快乐 **骄傲**	生气 信任 **支配**	
期待 快乐 **乐观**	期待 信任 **希望**	期待 害怕 **焦虑**	

与此同时，8 个基本情绪也分别有更加强烈的形式和更加平和的形式。例如，“快乐”是“狂喜”与“平静”的中间状态，“生气”是“狂怒”和“烦恼”的中间状态。

表 1.2 同一类情绪的不同形式

平和形式	基本形式	强烈形式	平和形式	基本形式	强烈形式
平静	快乐	狂喜	悲观	悲伤	悲痛
分心	惊奇	惊讶	专注	预感	警觉
烦恼	生气	狂怒	理解	害怕	恐惧
接受	信任	崇拜	无聊	厌恶	痛恨

情绪轮盘模型分为平面和立体两种形式，平面图形为一个中间为八个扇形组成的圆形、沿着八个角度展开的二维空间，立体模型是一个倒立的圆锥体，原理与平面图形相同。平面模型中，8 种强烈情绪组成了情绪轮盘最中间的圆形，不能参与情绪的组合，8 个扇形各自独立。相比内层的强烈情绪和外层的平和情绪，中间的基本情绪在人们的日常生活中出现的概率是最大的。每个角度之间的区域为相邻基本情绪组合成的一级组合情绪。

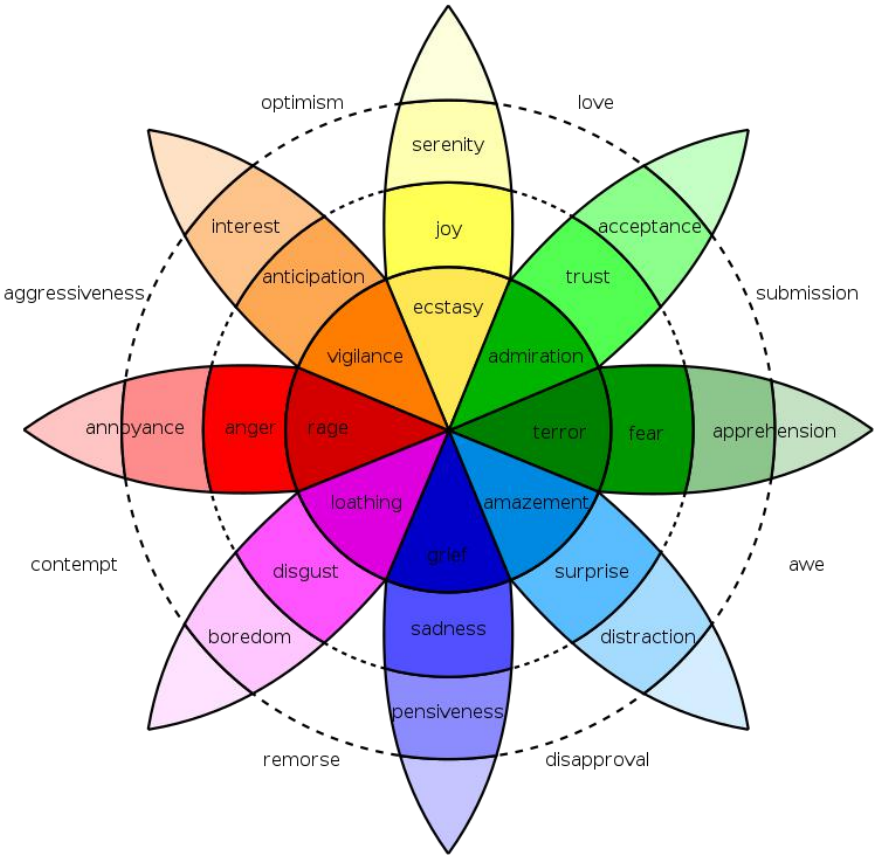

图 1.1 情绪轮盘的二维平面图示[1]

1.1.2 情绪轮盘模型的提出准则

普洛特契克在 1980 年提出的基于心理进化理论的基本情绪轮盘框架有十个假设前提[2]。

1. 情绪的概念适用于所有进化阶段，不仅适用于人类，也适用于动物。
2. 情绪同样拥有进化的历史，在不同的物种中进化出不同的表达形式。
3. 当生物遇到由环境引起的关键生存问题时，情绪帮助生物适应环境。
4. 尽管不同物种有不同的情绪表达形式，仍存在一些确定的共同的元素或原型模式。
5. 有少量的基本的、主要的或原型的情感。
6. 所有其他情绪都是混合或衍生的状态，即它们以原始情绪的组合、混合或化合的形式出现。
7. 基本情绪是一种假设的结构或理想化的状态，其属性和特征只能从各种证据中推断出来。
8. 基本情绪可以用两极对立的一对来概念化。
9. 所有的情感都有不同程度的相似之处。
10. 每种情绪都可以存在于不同程度的觉醒强度或程度。

1.1.3 情绪轮盘模型的详细解读

对于第三个假设提到的关键生存问题，普洛特契克指出：所有进化水平的生物体都面临着某些共同的功能性生存问题，这也是情绪产生的前提条件。

将基本情绪的产生汇总一下，得到以下对于情绪产生原因的总结：

表 1.3 情绪产生原因

刺激事件	认知评估	主观反应	动作反应	希望获得的功能
某一事件中有所收获	“获取”	快乐	保持	收获资源
某一事件中有所失去	“抛弃”	悲伤	哭	重新得到失去的事物
来到新的领域	“仔细检查”	预感	了解信息	获取知识

没想到的事件	“这是什么”	惊奇	停止	得到时间来适应
阻碍	“敌人”	生气	攻击	打破阻碍
威胁	“危险”	害怕	逃避	安全
某组织中成员	“朋友”	信任	照顾	共同支持
难吃的东西	“毒药”	厌恶	呕吐	排出毒药

而中文环境下每个情绪的词汇含义和英文词汇也有着轻微的差别。我们将“快乐-悲伤”代表人的得志与失落；“预感-惊奇”代表人对即将发生的事情有无心理准备；“生气-害怕”代表人基于其自我认知是否处于有能力影响周围环境的状态；“信任-厌恶”代表人是否接受现在所处的环境。

将基本情绪分为正负两极，生气、期待、快乐和信任是正的，而恐惧、惊奇、悲伤和厌恶在相对是负的。生气(anger)被归类为一种“正的”情绪，因为主人公认为自己有能力改变周围环境，这个概念涉及朝着一个目标前进的积极性。而惊讶是负面的，因为它是对某人熟知的领域的一种侵犯。

1.2 本文研究的思路和价值

1.2.1 本文的研究动机

目前各情绪识别的 AI 平台（例如 Face++） 主要将情绪识别限制于愤怒、厌恶、恐惧、高兴、平静、伤心、惊讶等七类情绪，识别结果为单一的情绪类别和相应的置信度。由于情绪识别结果不够丰富、缺少深度，因而应用场景受限，应用价值不高。

因此，本文基于深度学习模型设计了一个情绪轮盘模型，将心理学的知识与深度学习的方法结合起来，可以针对输入的图片或视频识别出当中人物的情绪状态，产生的识别结果（返回值）为 4 维向量，覆盖了 81 种情绪状态，对应于几百个情绪词，几乎囊括了人类全部的情绪。本文提出的情绪模型识别结果相比单一的识别结果不仅更加准确，而且覆盖面更广，因此有更加广泛的应用场景。

1.2.2 本文的研究思路和主要贡献

本文首先介绍了普洛特契克的情绪轮盘体系和情绪词表的构建，并回顾了人脸识别的技术流程框架和常用的深度学习模型。而后进行了实验探索，实验探索主要围绕着“建模”和“应用”两个方面。

“建模”方面采取了深度学习的方法构建了情绪轮盘结构，其中主要用了 CNN 的网络构造。从影视作品中获取训练数据，并对于网络上流行的人脸数据集

作为测试集进行检测。通过深度学习得到的情绪轮盘模型，可以通过输入图片或视频得到人物的情绪状态，返回值为 4 维向量，识别覆盖了 81 种情绪状态，对应于几百个情绪词，覆盖了人类几乎全部的情绪，比单一结果的返回值的覆盖面更广、结果也更准确，因此有着非常广泛的应用。

在完成了模型的构建后，将模型应用于不同场景，相比于一维的表情识别，情绪轮盘模型可以帮助我们更好地理解不同情绪之间的联系和差异，对人物情绪的识别在影视作品的理解、监控驾驶员情绪保障交通、推荐系统、心理状态检测等方面有着非常重要的应用价值。

本文的主要贡献有：

1. 基于情绪轮盘模型构建情绪词表，词表共有 81 种情绪状态，每一种状态对应于一个 4 维向量，共对应于几百个情绪词，几乎囊括了人类全部的情绪。

2. 利用深度学习的方法设计并构造了一个情绪轮盘模型，可以针对输入的图片或视频识别出当中人物的情绪状态。

3. 总结了情绪轮盘模型的应用场景，并对于影视语言方面进行了详细的应用探索，包括视频推荐语的文本生成、演员演技评估等等方面。

第二章 相关工作综述

2.1 情绪词表的构建

本文将人类的 8 种基本情绪分为四个独立的维度，每个维度的正向情绪为+1，负向情绪记为-1。同时将每个维度处于中间的情绪状态记为 0。

表 2.1 情绪维度划分

	正情绪(+1)	负情绪(-1)
维度 1	快乐	悲伤
维度 2	预感	惊讶
维度 3	生气	害怕
维度 4	信任	厌恶

基于情绪轮盘中基本情绪可以组合的特点，本文将四个维度的结果组合成一个情绪状态，即每个情绪是一个四维向量。如“快乐+信任=爱/友善”，快乐和信任分别是维度 1、4 的正极，所以(1,0,0,1)代表的情绪即为一种爱慕或者友好亲和的情绪。

除了情绪轮盘上面已经展示的情绪双元对，我们认为除了对立的情绪外的所有基本情绪之间都是可以组合的。因此，四个维度上+1、-1、0 三种极性，进行排列组合后可以得到 81（3^4）种情绪。将 81 种情绪状态整理罗列一下构建情绪词表，有的情绪状态对应的词汇不只一种类型的情绪，例如：(0,0,-1,-1)可以同时代表羞耻/羞怯两种情绪。

表 2.2 情绪词表示例

I	II	III	IV	情绪状态	中文情绪词	英文情绪词	替换的词
1	0	0	0	(1,0,0,0)	joy	快乐	开心、欢喜、兴奋、喜悦、欢乐、欣喜、高兴、愉快
1	0	0	-1	(1,0,0,-1)	morbidness /dersivene -ss	病态/嘲弄	揶揄、取笑、嘲笑、冷笑、嘲讽、讽刺、奚落
1	0	0	1	(1,0,0,1)	love/frien -dliness	爱/友善	爱慕、相恋、喜欢、亲近、深爱、迷恋、宠爱、爱恋、友好、亲和
1	0	1	0	(1,0,1,0)	pride	自豪	骄傲、引以为傲、得意、自信
1	0	1	-1	(1,0,1,-1)	arrogant	傲慢	颐指气使、嚣张、盛气凌人

情绪三元对、与情绪四元对是情绪二元对的融合，例如：“ (1,0,1,-1)傲慢”可以看做“(1,0,1,0)自豪”、“(1,0,0,-1)嘲弄”、 “(0,0,1,-1)鄙视”的混合。然而，情绪三元对与情绪四元对是很难用单独的一个词形容的，这是本

文任需解决的地方。

我们希望赋予情绪识别更高的应用价值，首先可以做的便是将词汇转换为文本。根据每个情绪词的词性、语义等特点，将其套入主谓宾等结构的语句中。如下图是部分情绪词的句式示例。

表 2.3 情绪词的句式示例

情绪状态	英文情绪词	中文情绪词	文本句式
(1,0,-1,0)	excitement	激动	A（因为 B）感到激动
(0,0,0,1)	trust	信任	A（对 B）感到信任
(0,0,1,0)	anger	气愤	A（因为 B）感到气愤
(1,0,1,-1)	arrogant	傲慢	A（对 B）傲慢
(-1,-1,0,0)	disapprove	不赞同	A 不赞同 B

2.2 人脸识别的技术流程框架

本文提出的人物情绪识别模型和人脸识别高度相关，所以在此简要介绍一下人脸识别技术。目前现有的人脸识别模型基本都是表情识别，缺少对于情绪的准确划分和表情到情绪的过渡。

人脸表情识别的一般步骤为获得图片、预处理图片、特征提取及表情分类四个步骤[8]。

1. 图片预处理：包括人脸检测与定位、人脸对齐、数据增强、图片归一化等方法。归一化包含几何归一化与灰度归一化两种方法。

① 人脸对齐是指通过仿射变换的方法将人脸图片显示到统一的预定义模板上，从而减轻旋转和面部变形带来的变化。

② 数据增强有线上和线下两种，线上增强是比较常用的随机中心裁剪或是随机水平翻转；线下增强包含了添加噪声、改变曝光与饱和度、随机扰动变换等等。

③ 图片归一化是将直方图均衡和光照归一化互相结合的方法，由于存在于三维现实生活空间中的人脸往往会受到光照而产生灰度变化，故基于各向同性扩散归一化、或是离散余弦变换归一化（DCT）、高斯差分（DoG）等等来减弱光照的影响。

2. 人脸特征提取的主要方法有：主成分分析法（PCA）、卷积神经网络（CNN）、局部二值模式（LBP）、循环神经网络（RNN）等。

3. 表情特征分类有两大类方法，传统的分类器有 SVM、随机森林（Random Forest）、 AdaBoost, Fisher 线性判别（FDA）、k-最近邻学习法，隐马尔科夫算法等等分类模型；或是直接基于深度学习学习图片特征预测概率：通过正向传播和反向传播两种模式优化网络减少损失，直接在网络末端输出每个样本的预测概率。

2.3 基于静态图像的人脸表情识别模型综述

2.3.1 人脸表情识别常用 CNN 模型

人脸识别可以利用的经典 CNN 模型有：AlexNet[9] 、GoogLeNet[10]、VGGNet[11]、ResNet[12]等。他们有很多相同的地方，例如都含有 DropOut 层，都用到了数据增强等特点。他们之间的区别如下：

表 2.4 人脸识别的经典 CNN 模型举例

深度学习模型	层数	卷积核大小	是否有 Inception	是否有 BatchNorm	备注
AlexNet	8	11,5,3	无	无	首次使用了新的激活函数整型线性单元和 dropout 机制
VGGNet	16/19	3	无	无	由许多具有 3*3 小滤波器的卷积层彼此堆叠来模仿出更大的感受野的效果
GoogLeNet	22	7,1,3,5	有	无	将 Inception 层应用于多个数据库
ResNet	152	7,1,3,5	无	有	将层定义为参照层输入的学习残差函数

2.3.2 优化方法

（1）预训练和微调

此方法分为两个阶段：第一阶段在已经预训练好的网络模型上进行微调，网络模型如 AlexNet[9] 和 GoogLeNet[10]等等；第二阶段利用目标训练库（即 target dataset）中的训练数据进行微调，作为额外的任务导向[13]。

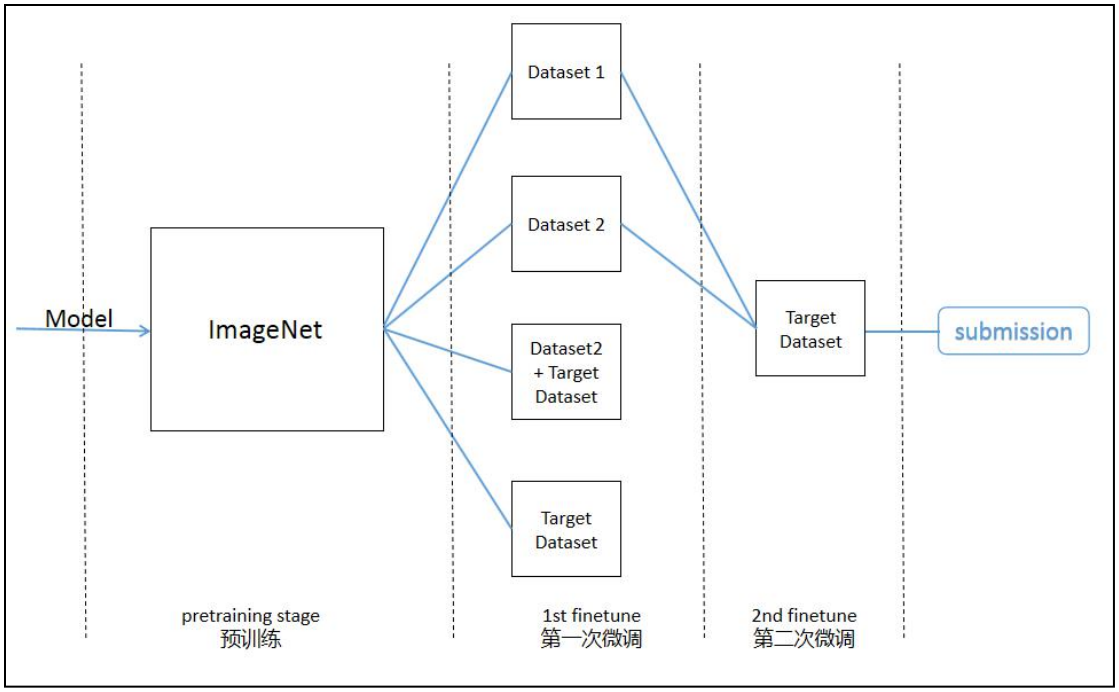

图 2.1 预训练和微调结构示例

（2） 多样化的网络输入

由于深度学习的网络中输入为整张的图片（通常输入值为像素），会失去一些纹理特征。对于这种情况，传统的非深度学习方法可以用来提取人体图像的特征，然后将新的特征融合到网络模型中。

可以编码给定 RGB 图像中的小区域特征，然后聚类并融合这些特征，使得它们对小的配准误差和光照变化具有鲁棒性。尺度不变特征变换（SIFT）对图像缩放和旋转具有鲁棒性的特征被用于多视图的人脸识别任务。将轮廓、纹理、角度和颜色中的不同描述符组合为输入数据也有助于提高深度网络性能。

图 2.2 三维度量空间输入模拟[14]

如上图示例是将图像强度（左图）和 LBP 代码（中图）中的值映射到三维度量空间（右图）作为卷积神经网络的输入时的模拟[14]。

（3） 辅助块或层的增加

由于表情的类间区分度较低，因此传统 CNN 中的 softmax 归一化输出在表情识别领域的表现并不理想。针对表情分类层的改进，可以对特征与相应的类距离加了惩罚项，这分为两各部分：其一是增加类间距离的 island loss，其二是减小类间距离的 LP loss。这两部分同时有着对 CNN 的训练进行监督的作用，旨在扩大类间差异并减小类内变化，并使同一类的局部相邻特征得以更多地结合在一起[15]。

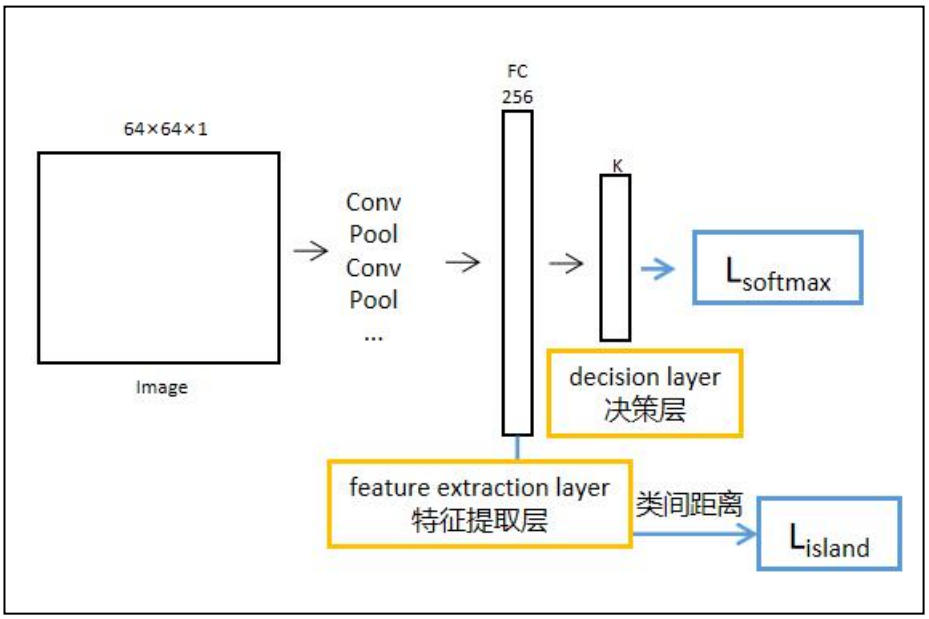

图 2.3 类间距离的 island loss 与 LP loss

（4） 网络集成

一般情况下，多个网络的集合可以表现出比单个网络更好的效果。网络集成分为两类，一种是针对特征的集成，另一种是对于输出决策的集成。

最常见的特征集成的方法是直接连接不同网络模型的特征，组成一个新的特征矢量来表示图像。例如：2016 年，Bargal 等人提出将三种不同的特征在归一化后连接在一起，生成一个单一的特征向量（FV），然后用它来描述输入帧[17]。

决策集成往往采用投票机制，对于不同的网络可以尝试不同的权重，有简单平均、加权平均、多数投票几种方法[18]。

2.4 现有方法的局限性

目前现有的情绪识别模型通常是将人的表情做一个纯粹的多分类问题。如：Face++模型将情绪分为高兴、平静、惊讶、伤心、厌恶、愤怒、恐惧七种，返回人脸在各类不同情绪上的置信度分数。还例如百度人脸检测与属性分析 API 中将表情分为愤怒、厌恶、恐惧、高兴、伤心、惊讶、无表情、撅嘴、鬼脸几种，其中“噘嘴”和“鬼脸”并不属于情绪的范畴，其他的表情分类与 Face++大同小异。

然而，人类表情并非只局限于 7 或 8 种基本表情，目前已有的这种一维单输出的分类问题并不能很形象地传达情绪，也不能表达情绪的丰富层次，只适用于简单的人脸识别、人脸属性分析的应用场景，浮于表情这个概念表面，不能满足于更深层次（如文本等领域）的应用。因此，将认知心理学的理论与模型搭建相结合客观重要，这也是提出将情绪轮盘模型通过深度学习的方法运用到情绪识别领域的原因。

第三章 基于深度学习的情绪建模

3.1 算法介绍

3.1.1 模型的架构

（1）沿着情绪轮盘模型对角线切割，将情绪轮盘模型的架构问题转换为四个三分类问题，每个维度有+1、0、-1 三个标签。

（2）人脸检测

利用 dlib 提供的开源模型文件 shape_predictor_68_face_landmarks 检测人脸，可以精准识别到 68 个人脸特征点。一般来说，68 个特征点的分布为：

表 3.1 人脸 68 个特征点的分布

位置	编号	具体
脸部轮廓	1-17	左脸最外 0　下巴 8 右脸最外 16
眉毛	18-27	左眉尾 18 左眉心 22 右眉尾 23 右眉心 27
鼻子	28-36	鼻根 28　鼻尖 31
眼	37-48	左眼外角 37　左眼内角 40　右眼外角 46　右眼内角 43
嘴巴	49-68	嘴中心 68　嘴左角 49 嘴右角　55

以上左右为照片中显示的左右。

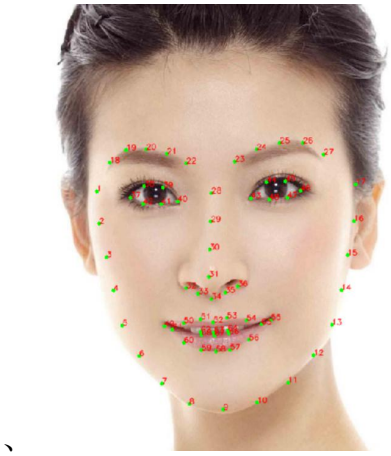

图 3.1 人脸 68 个特征点

（3）利用其裁剪图片得到人脸，去掉背景和非面部区域，并完成将侧脸旋转一定角度得到正脸实现人脸对齐的操作。

（4）将每张图片提取特征。

特征主要分为两类，第一类是 68 个坐标点的二维坐标，我们将其记为 Face landmarks。

第二类是人脸图片自身的特征，我们将每张图片通过 tensor 转换为 64*64 的大小，而后通过 CNN 训练。

为了将数据转换为神经网络便于学习的类型，还需要将数据归一化，同时还可以使用灰度归一化增加图片的亮度、调整对比度，减弱光线的影响。

（5）CNN 训练的模型建构

输入为 64*64*1 的灰度图片或者 68 个坐标位置，输出为 3 个值，对应于 +1\0\-1 的置信度。模型连接三层卷积层，而后加入全连接层。其中也加入了 Dropout 层，即在深度学习网络的训练过程中，为了防止过拟合的发生，按照一定的概率将其暂时从网络中丢弃神经网络单元。

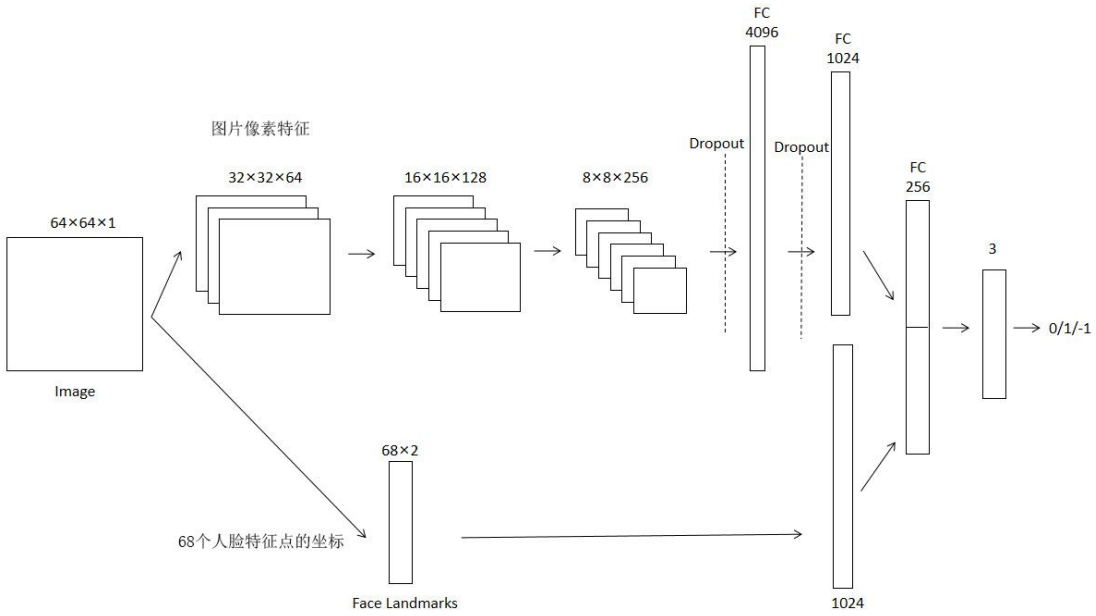

图 3.2 本实验所用的 CNN 模型架构

总体的实验框架如下：

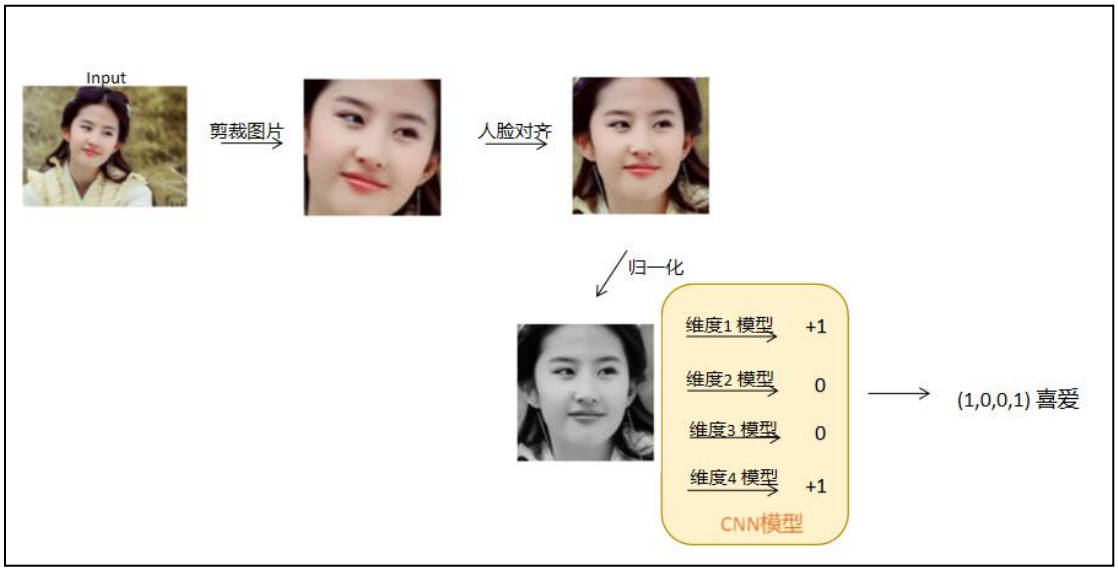

图 3.3 实验总体框架流程

3.1.2 目标函数

CNN 三分类问题输出的 3*1 向量结果经过 softmax 处理为 0-1 之间后，用交叉熵函数计算损失。

（1）softmax

softmax 函数的作用是将 N 分类的问题，转化为一个 N × 1 维的向量，概率和为 1 ，通常位于 CNN 网络的最后一层。

$$S_j = \frac{e^{a_j}}{\sum_{k=1}^{3} e^{a_k}} \tag{1}$$

每个类别的输出结果是正数，而且范围是(0,1)，所有类别的输出结果之和为 1，可以认为该输出结果为该类别的可能性概率。概率最大的类别即可认为是图片的分类结果。

（2）交叉熵

利用交叉熵计算预测结果的损失，预测结果与 label 值越接近，损失越小，反之则越大。交叉熵刻画了两个概率分布之间的距离，公式如下：

$$L = -\sum_{j=1}^{3} y_j \log s_j \tag{2}$$

其中：L 是损失，S_j是 softmax 的输出向量 S 的第 j 个值，表示的是这个样本属于第 j 个类别的概率。j 的范围也是 1 到类别数 T，因此 y 是一个 1*T 的向量，里面的 T 个值，而且真实标签对应位置的值是 1，其他 T-1 个值都是 0。这个公式更简单的形式为：

$$L = -\log s_j \tag{3}$$

3.2 实验方法及结果

3.2.1 数据集的获取

虽然网络上有很多数据集提供了“快乐”、“悲伤”、“生气”、“害怕”等分类的人脸图片数据（例如：fer2013 数据集[19]），但是其中往往缺失了“预感”、“信任”的维度。贸然使用网络上已经构建好的数据集的个别维度也会导致训练数据分布不均匀，导致四个维度的分类模型不统一，使得训练结果变差。

由于网络上并没有已经准确地分好类的数据集，因此先构建部分的数据集，然后与 fer2013 数据集融合，避免过度拟合基准数据集的测试集。fer2013 数据集是用谷歌的人脸识别 API 获得的，是一个大型人脸识别项目的一部分，进行了各种边界处理、去重和裁剪的 48*48 灰度图。

3.2.2 数据集的构建

在调查中发现，在搜索引擎中直接搜索 8 个基本情绪得到的结果比较杂乱，其中很少有图片是合格的人脸图片，包含人脸的图片也大部分是卡通头像，或是分辨率较低的表情包等。同时，在搜索引擎中搜索“预感”、“信任”得到人脸图片的概率是很低的。因此这个方法并不可行。

在日常生活中，反映最全面的情绪的数据资源便是影视作品。每部作品每个角色都在不同的场合下反映出不同的情绪，这也反向映射了情绪轮盘中情绪的丰富多样。

因此在观看影视作品的时候，将精彩的视频片段的选取出来，通过 ffmpeg 等手段解帧成图片（一般情况下每个视频每秒的帧数 FPS 随分辨率变化，600P

的视频一般 1 秒 8 帧，1080P 的视频一般 1 秒 25 帧）。

而后通过计算人脸相似性，可以从每个视频提取出的所有图片找出含有主人公面部特征的图片，这一步筛选同时过滤出了人脸质量较差的图片，因为不含人脸的图片、人脸质量过低的图片（如：侧脸的图片、含有人脸但因运动而模糊的图片、人脸框在整个画面中占比很小的图等）无法识别出人脸特征。

在得到图片后进行数据的分类——分析视频中主人公的情绪，例如，紧张是一种的预感到不好的事将要发生的情绪，即 紧张=预感（维度 2 的+1 情绪）+害怕（维度 3 的-1 情绪）。因此将这一片段的图片移到维度 2 的+1，和维度 3 的-1 的集合中。

3.2.3 数据集的清理

通过上面的方法可以得到 4 个维度+1/-1 的图片，再将除了每个维度不属于+1/-1 的图片全部归于每个维度 0 的集合中，得到初步的训练集。

这时需要整理数据集，我们希望得到的数据集有以下特点：

① 每一维度的 3 个极性的训练图片尽量数量相等。

② 每个维度的每一极性尽可能覆盖更多角色的人脸图片，不能让某一角色的图片过多，否则训练的特征将不是表情特征，而是该演员自身脸的特征。

③ 训练图片的人物尽可能多，同时性别比例尽量相等，并涵盖各年龄层。

④ 训练图片需要尽可能清晰，及人脸占画面的比例要大，同时尽可能是正脸。

根据以上规则进行数据集的清理。由于每个维度中 0 的图片数目要比+1/-1 多很多，可以规定数目来随机挑选图片。

3.2.4 训练数据的特点

本文构建的训练数据集收集了 632 个影视片段，涵盖了 144 位演员的 161 集作品，其中有 71 位女演员和 73 位男演员。以演员在影视作品中扮演的角色年龄来看，其中以年轻角色居多，但也尽量覆盖了各个年龄层，年轻：中年：老年的比例为：108:38:7，每个年龄段中男女比例都接近相等。

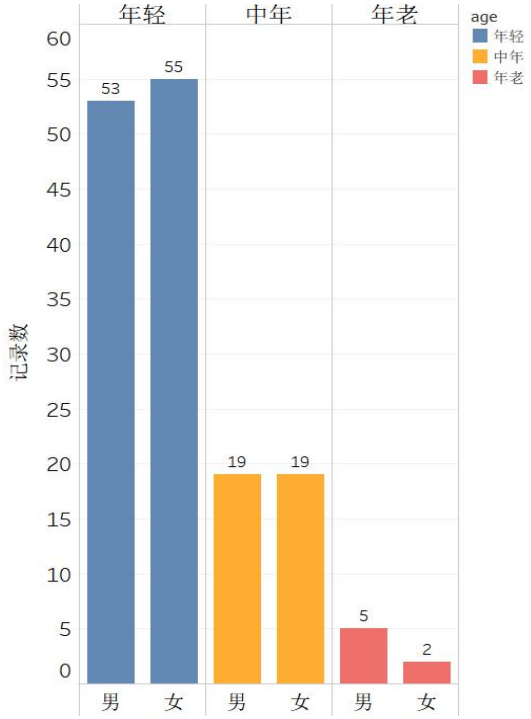

图 3.4 影视作品的训练数据中男女角色的比例和年龄比例

训练图片还覆盖了 175 种情绪，其中以开心、生气、难过等基本情绪最多，组合情绪的数目也十分突出。

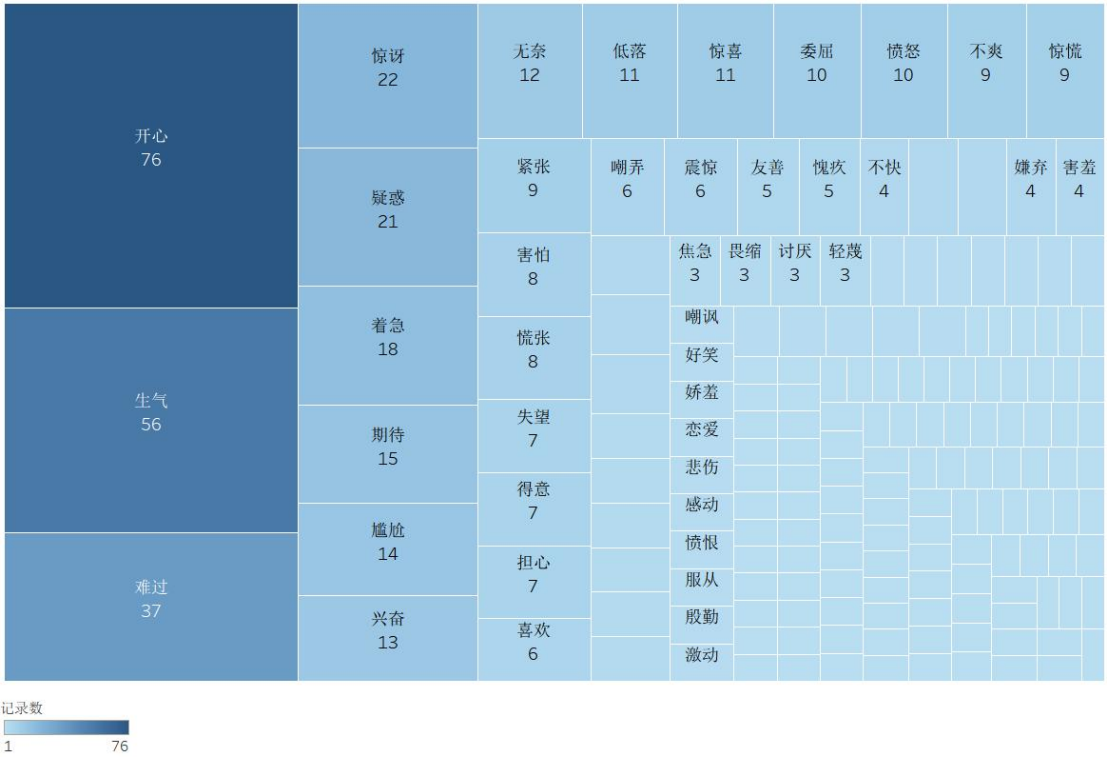

图 3.5 影视作品的训练数据中情绪词统计

对于每个视频片段，还统计了片段主人公与之对手戏的角色之间的关系，其中，“朋友”这一身份出现的是最多的，大部分角色之间的关系都是友好和谐的，例如：喜欢、夫妻。但也有“对手”、“敌视”、“讨厌”这样的负面关系，这也增添了数据的多样性。

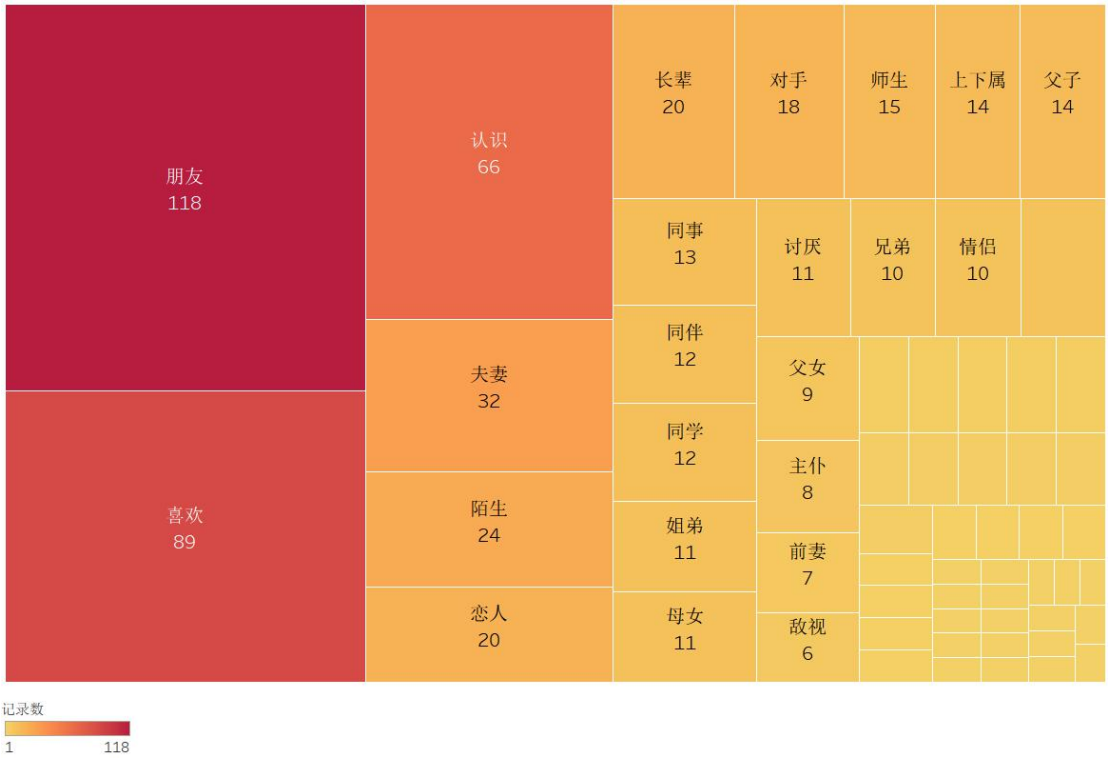

图 3.6 影视作品的训练数据中人物关系统计

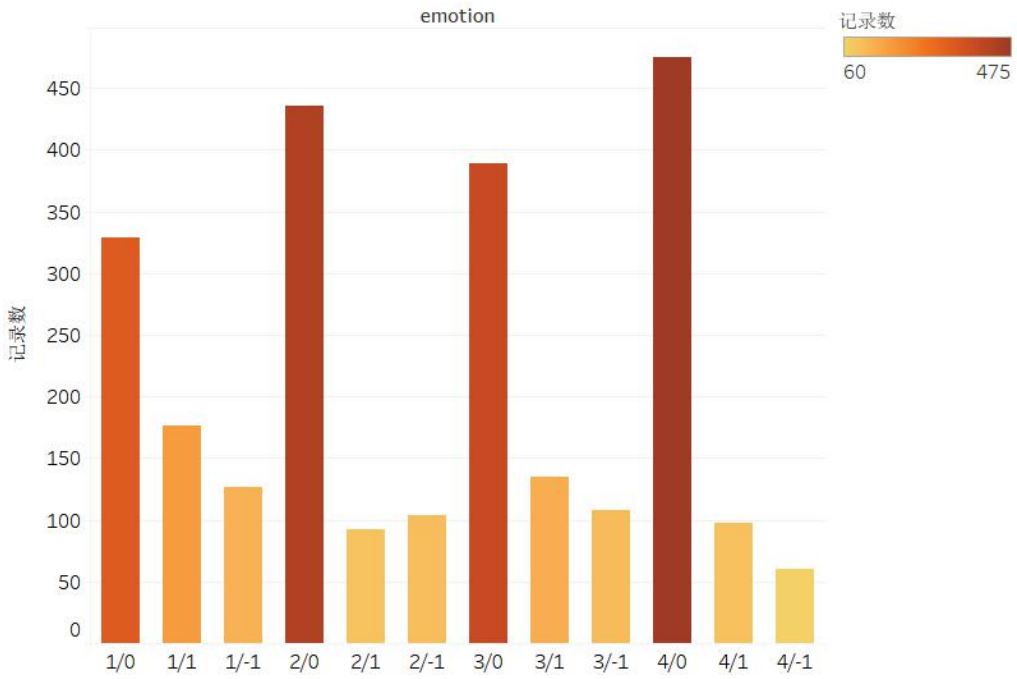

图 3.7 影视作品的训练数据中各维度图片数量统计

视频段落的情绪维度的数目也如下图所示，每个维度的三个极性的视频数量总和应都相等。可以看到正负一维度的视频段落数目都接近相等，由于每个视频提取的是一个人物的人脸图片，即每个极性分类包含的人物数目接近相等。

同时，也收集 fer2013 数据集的训练图片。将训练图片分配到各个维度的各个极性。

最终，自己构建的数据集和 kaggle 数据集的训练图片数量如图所示，可以看到维度一的图片数量最多，理论上维度一的训练效果应该是最优的。

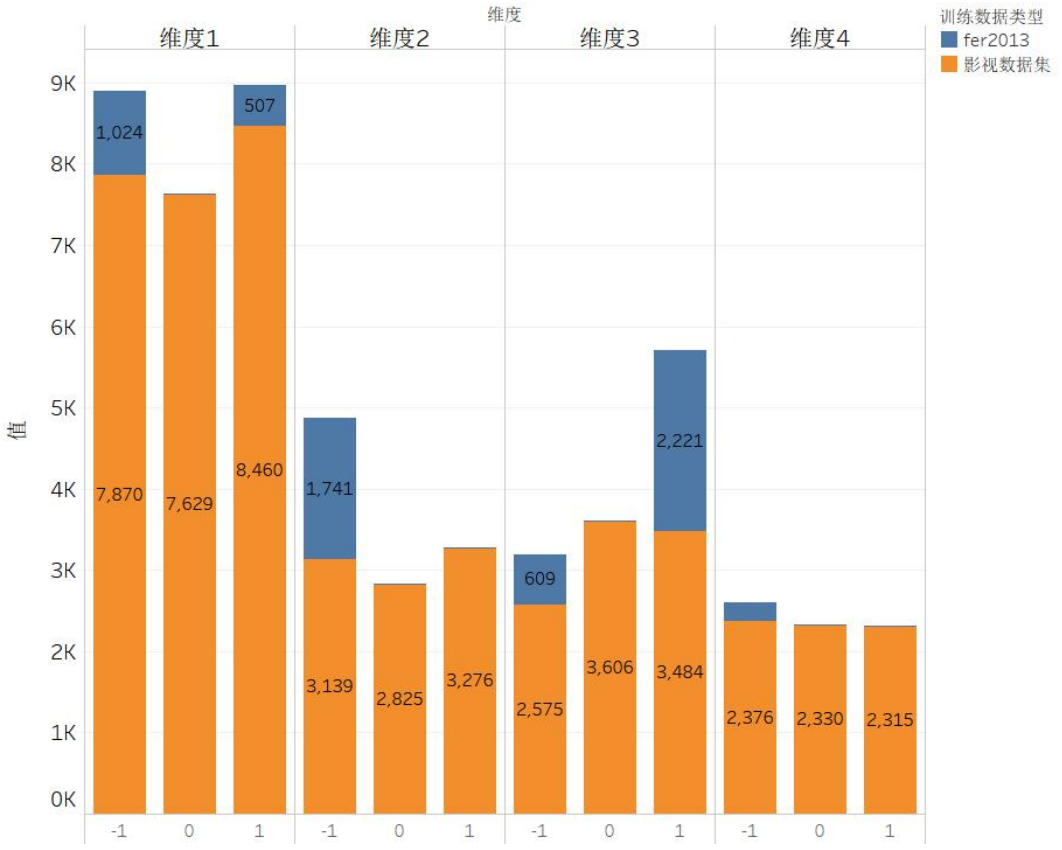

图 3.8 训练数据中各情绪维度的图片数量统计

3.2.5 训练结果

在 pytorch1.0.0 环境中构建 CNN 模型并进行训练，采用随机梯度下降的优化方法。

将收集好的数据随机抽出 1/10 作为测试集，其余作为训练集。经过 100 个

循环的训练，4 个三分类模型的训练集的准确度都可以达到 98%以上，测试集的准确率都可以达到 90%以上。保存每个维度在测试集上测试准确度最高的模型。

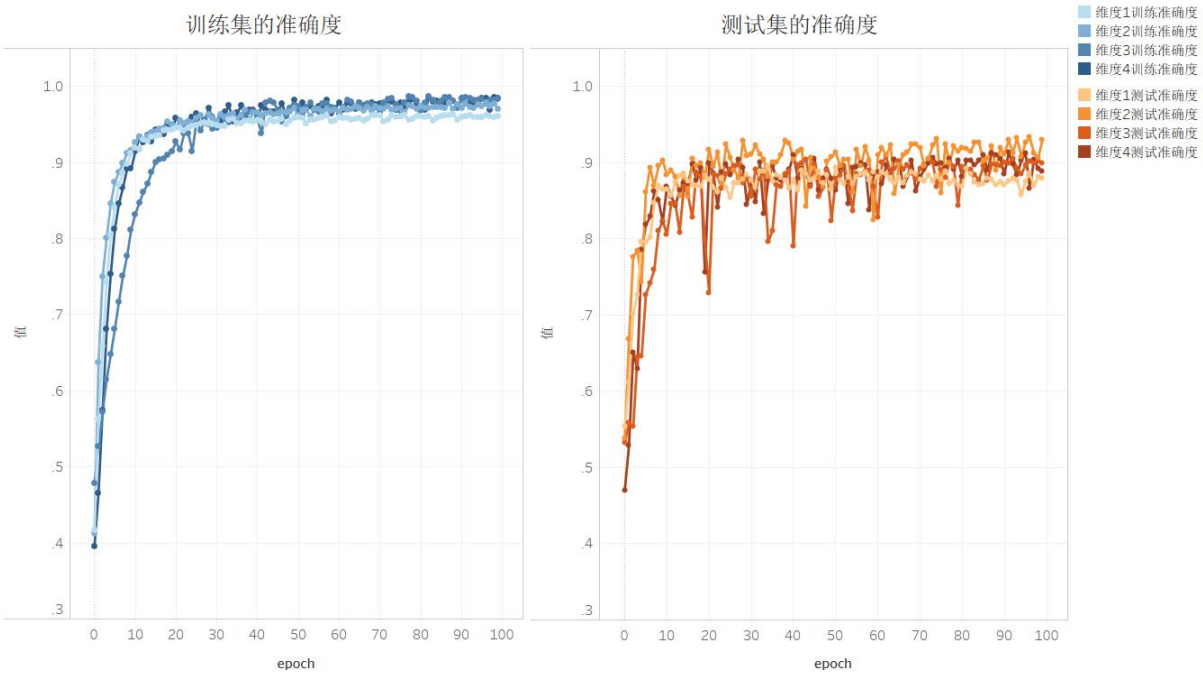

图 3.9 各维度训练结果和测试结果

3.2.6 模型对情绪识别的结果展现

经过之前的训练得到了四个三分类模型。输入一张含有人脸的图片，裁剪出人脸提取 64*64 的图片特征，经过四个模型后，每个模型输出一个+1/0/-1 的值，组合成一个四元向量。如果检测不到人脸，则输出结果为(0, 0, 0, 0)。

即输出结果为（E_1, E_2, E_3, E_4） $E_i \in \{1,0,-1\}, i \in \{1,2,3,4\}$ 。

3.2.7 其他数据集上的测试结果

选用 JAFFE database 数据集[20]作为新的测试集做测试。该数据库收集了照相机拍摄获取的 10 位日本女性根据实验命令做出的各种表情图像。整个数据库一共有 213 张图像，10 个女性每个人做出 7 种表情，分别是： 悲伤 sad， 快乐 happy， 生气 angry， 厌恶 disgust， 惊奇 surprise， 害怕 fear， 平和 neutral。每个人提供的每种表情大概有 3, 4 张样图。

利用四个分类模型测试后，得到 200 多张图片的输出结果。只看基本情绪所

在维度的结果，例如表情为“surprised 惊奇”，则只关注维度 2 的结果。

可以看出，“happy 快乐”的分类准确率最高，31 张显示快乐的图片只有 1 张存在没有检测出快乐情绪的情况，这也与维度 1 的训练数据图片数量最多有关，验证了之前的猜测。同时，维度 4 的“disgust 厌恶”正确分类率最低，这与没有收集到足够多的训练数据也有关系，影视作品中直接表现出“厌恶”和“信任”的情绪片段不够丰富，FER2013 数据集中“disgust 厌恶”的情绪图片也是最少的。

与此同时我们观察到，发生错分的情况大多数是 1 和 0 之间以及-1 和 0 之间的，然而和“表情识别”不同的是，“情绪识别”中有无情绪这个分界的标准本来就比较模糊。因此我们可以把重点放在 1 和-1 之间的错误率上，把其他的结果都当做是“acceptable 可接受的”，把 1 和-1 之间的错分当作是“unacceptable 不可接受的”，这样“不可接受”的结果占全部结果的比例只有 8% 左右。

另外，我们还可以采取打分制度，将正确的分类结果记为 1 分，1 和-1 之间的错分记为 0 分，而没有识别出情绪的结果记为 0.5 分，这样计算每个基本情绪测试情况的平均分：

$$\frac{\#(\text{正确分类})\times 1+\#(\text{分类为}0)\times 0.5}{\text{该类别图片总数}}$$

例如：31 张“sad 悲伤”图片中，有 17 张分类正确，有 4 张被分为了“快乐”，10 张没有识别出情绪，则平均分为 $(17\times 1+10\times 0.5)/31=0.7097$ 。

表 3.2 JAFFE database 测试结果统计

	维度 1		维度 2	维度 3		维度 4
	快乐	悲伤	惊奇	生气	害怕	厌恶
1	96.77%	12.90%		53.33%	15.63%	6.90%
0	3.22%	32.26%	26.67%	26.67%	28.13%	48.28%
-1		54.84%	73.33%	20%	56.25%	44.83%
准确率	0.9677	0.5484	0.7333	0.5333	0.5625	0.4483
得分	0.9839	0.7097	0.8667	0.6667	0.7031	0.6897

可以看出每个基本情绪的得分都在 0.6 以上，是可以接受的水平，说明训练成功。

第四章 情绪轮盘模型的应用价值及前景蓝图

4.1 情绪识别的应用总述

众所周知，人类总是说不清自己的感受，情绪更是一个抽象的概念，人们很难描述刻画自己的情绪。借助深度学习模型，机器不仅可以识别人们的部分感受，理解发生的场景，甚至还可以更进一步地预测出他们接下来可能采取的行动。可以把这些应用归于“情感计算”的领域。情绪轮盘模型所覆盖的情绪识别可以应用于不同领域，其中以影视行业和推荐领域为代表，如下图 4.1 为不同应用场景的总结。

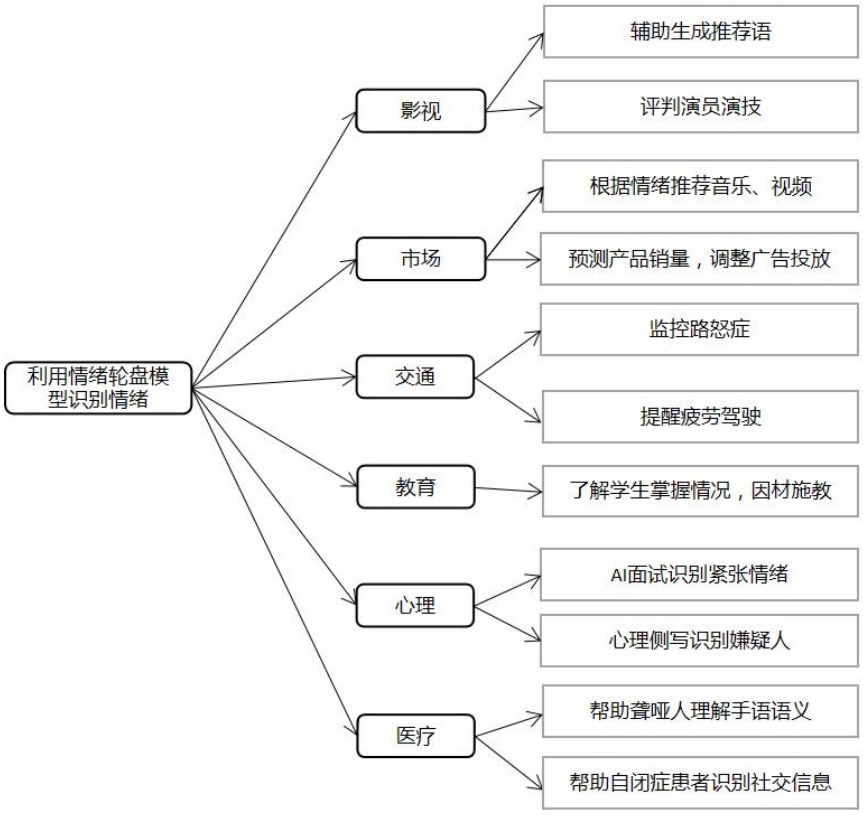

图 4.1 情绪轮盘的应用总结

4.2 不同领域中的应用前景

4.2.1 影视方面的应用——影视作品的内容理解、演员演技的评估

单单通过一张图片准确地描述人物的情绪是很难的，即使是人像清晰的正脸，由于人的微表情等误差的存在，单独的图片也很难表达人的情绪。而通过一

段完整的视频判断人的情绪则更为准确。而这使得情绪轮盘模型在影视领域的应用就十分重要。

这一领域的应用主要分为两个方面，第一个方面是结合文本方面的应用，第二个方面是关于情绪变化的应用。主要总结如下图所示：

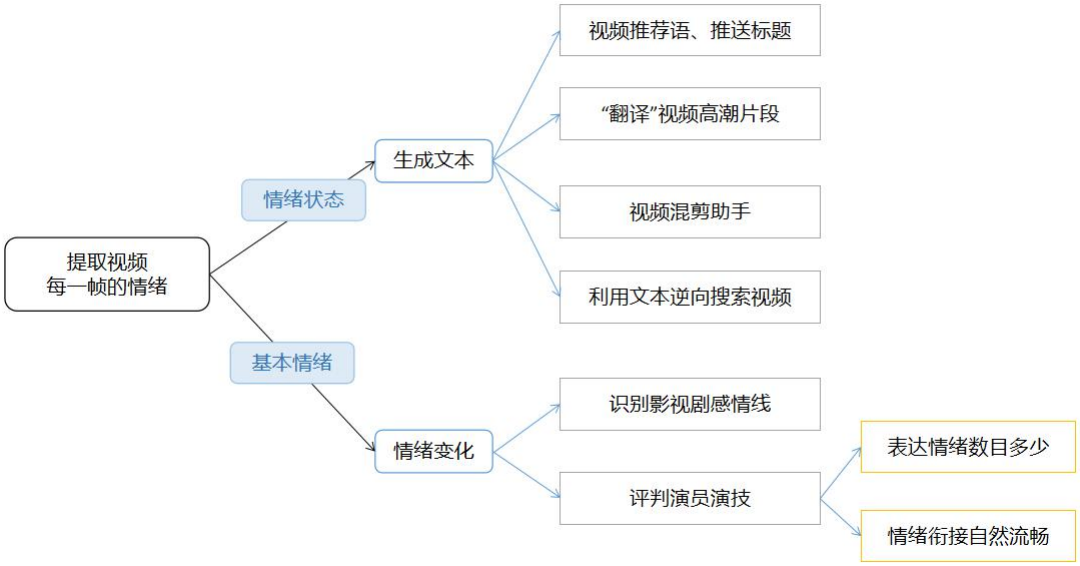

图 4.2 影视方面的情绪轮盘应用总结

1. 提取情绪的步骤

我们可以选取与之前训练集相似的方法，通过一段影视视频来提取角色的情绪，主要步骤如下：

① 将一个角色的某段视频（例如是电视剧的一集或是一段），通过 ffmpeg 将这段视频按帧截取成图片。

② 通过人脸相似度的计算选取其中我们想要的人物的图片。

③ 调用情绪识别模型，每张图片都会有四个维度的结果，对应成一个四维向量——情绪状态。

表 4.1 情绪轮盘模型的视频输出结果示例

图片帧数	维度 1	维度 2	维度 3	维度 4	单张图片的情绪状态
318	1	0	-1	0	(1, 0, -1, 0)
319	1	0	-1	0	(1, 0, -1, 0)

320	1	0	-1	0	(1,0,-1,0)
321	1	1	-1	1	(1,1,-1,1)
322	1	1	-1	1	(1,1,-1,1)
323	1	0	-1	1	(1,0,-1,1)

④ 区间密度的计算

接下来的步骤需要区分情绪维度、情绪状态、情绪词几个概念，情绪维度指的是 8 个基本情绪所划分的四个维度，每个维度里面有 1、0、-1 三种极性；情绪状态指的是四个情绪维度结合起来的一个四元向量，例如(1,0,-1,0)；情绪词指的是情绪状态对应的词，每种情绪状态可能有很多情绪词，我们通常选取其中比较典型的一个或两个情绪词对应一个情绪状态。

其次提出了区间密度的概念。如果人物的情绪达到顶点，每一帧都可以检测出这种情绪（理论上是不可能达到的），即区间密度为满分 1。

区间密度=情绪出现的次数/区间的总帧数

将每 1000 帧作为一个区间片段，我们可以计算每种情绪状态和每个基本情绪的在这 1000 帧中出现的频数和区间密度。

2.计算每种情绪状态的区间密度

（1）计算过程

我们将 81 种情绪状态转换为 one-hot 向量，可以求出每 1000 帧中每个情绪状态的数目。

我们以 2019 年的现象国产偶像剧《亲爱的，热爱的》中第 38 集的片段作为例子。在第 478 帧到 1477 帧这 1000 帧的区间中，’(0,0,-1,0)fear 害怕’这一情绪出现了 83 次、’(1,0,-1,0)guilt/excitement 有罪/激动’这一情绪出现了 52 次（这也正好对应着女主角在带男主角回家拜访父母时的害怕紧张又有些激动的心情）。

对于区间的全部情绪状态结果进行一个阈值分割，选取阈值为 50（即区间密度=0.05），保留出现次数大于 50 的情绪。即可得到每个区间片段的情绪结果。将单张图片的结果转换为一个区间的结果可以削弱单张图片的分类错误带来的影响。

选择 1000 帧为一个区间片段的原因是，当 FPS=8 时，1000 帧对应 125 秒、两分钟左右的视频，这差不多是一场对手戏的长度。

而选择 50 帧为阈值（即区间密度=0.05）也经过了一系列的实验，发现当阈值=0.05 时，筛选出的情绪既不会过多而过于琐碎，又不会过少，同时情绪的表露和体现又十分典型。

表 4.2 每 1000 帧的区间片段中达到阈值的情绪状态示例

起始帧	结束帧	情绪状态	*情绪状态的*记录数
474	1474	[(0,0,-1,0)fear 害怕]	[83.0]
475	1475	[(0,0,-1,0)fear 害怕]	[83.0]
476	1476	[(1,0,-1,0)guilt/excitement 愧疚/激动, (0,0,-1,0)fear 害怕]	[51.0, 83.0]
477	1477	[(1,0,-1,0)guilt/excitement 愧疚/激动, (0,0,-1,0)fear 害怕]	[51.0, 83.0]
478	1478	[(1,0,-1,0)guilt/excitement 愧疚/激动, (0,0,-1,0)fear 害怕]	[52.0, 83.0]

经过“滑窗”处理（即每隔一帧选取片段 1-1000、2-1001、...）后，即可得到一个角色该视频中所有的区间片段的情绪。每个区间可能含有不只一种情绪状态，也会存在大段的区间由于角色没有出现或出现的密度较低而不表露出情绪状态。如，以下示例中男女主可以检测到情绪区间占区间总数的比例分别为 28.5%和 41.6%。

将相同情绪的区间片段按顺序合并，就可得到这一个视频的情绪变化图表。

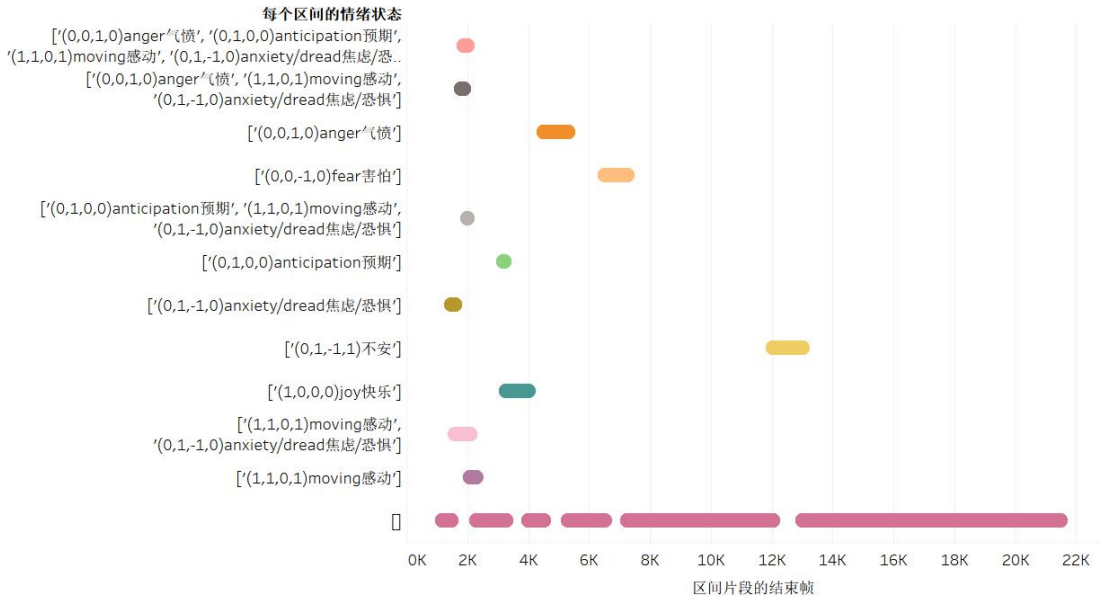

图 4.3 一集电视剧中主角的情绪变化图表

将情绪做直方图统计，可以得到这一个视频中的情绪出现次数。

表 4.3 一集电视剧中主角的情绪状态统计

区间的情绪状态	计数
[]	11890
['(0,0,-1,0)fear 害怕', '(0,1,-1,0)anxiety/dread 焦虑/恐惧']	2125
['(1,0,-1,0)guilt/excitement 有罪/激动', '(0,0,-1,0)fear 害怕']	1465
['(0,0,-1,0)fear 害怕']	923
['(-1,1,-1,0)worry 担心', '(0,1,-1,0)anxiety/dread 焦虑/恐惧']	878
['(1,0,-1,0)guilt/excitement 有罪/激动', '(0,0,-1,0)fear 害怕', '(0,1,-1,0)anxiety/dread 焦虑/恐惧']	679
['(0,1,-1,0)anxiety/dread 焦虑/恐惧']	679
['(0,0,-1,0)fear 害怕', '(0,0,-1,-1)submission/modesty 服从/谦虚']	604

（2）主要应用：文本生成

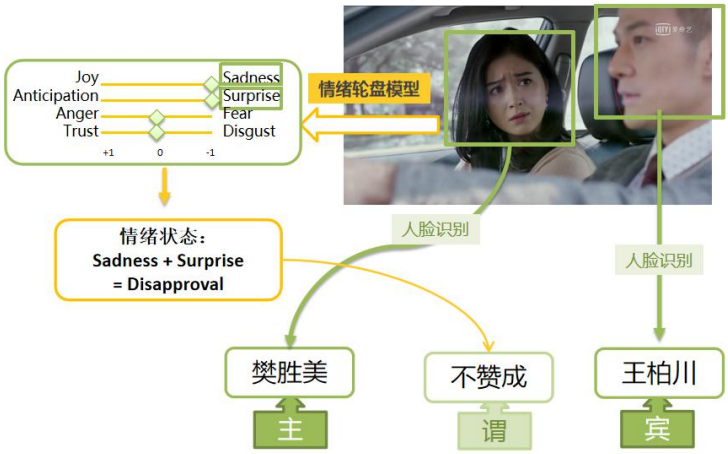

图 4.4 文本生成过程介绍

把情绪词带入句式关系中，可以将词转换为一句文本。情绪词从词性上分为动词和形容词两种，

动词是 AXXB 的句式，例如 A 信任 B、A 不赞成 B。这类的情绪词通常不能省略宾语，否则句义会不连贯。

形容词通常需要“感到”“表示”等动词的连接，以及“因为”、“对于”等连词的辅助，例如：A 因为 B 而感到激动，A 对于 B 感到惊讶。这类的情绪词通常省略宾语也可以得到连贯的句义，直接把人名作为宾语对象有时会比较生硬，需要表示原因的名词的帮助，例如：A 因为 B 的成就感到激动，A 对于 B 的行为感到惊讶。

由于每个段落也许会有不只一个情绪状态，我们需要“既”、“又”等句式，例如：A 对 B 表示担忧又焦虑，A 既信任又喜欢 B。

如此，每输入一段 1000 帧的视频，都可以得到一句文本。而输入一段较长的视频，对于同一主人公，将相同情绪的区间片段按顺序合并，也就得到了几句文本，这几句文本的宾语可能因为对手戏的不同而发生变化。

在前文《亲爱的，热爱的》这一例子中，一集 46 分钟的电视剧里，以女主角为主语可以合并得到 12 个句子，以男主角为主语，可以得到 9 个句子。如以下表格选取了女主角 A 在一集中文本生成的部分结果示例。

表 4.4 一集电视剧文本生成部分结果示例

场景	片段起始时间	片段结束时间	区间数	宾语	情绪状态&情绪词	生成句子
A 和 B 在楼下亲热时被保安打断	11’44	13’54	34	B	[(0,0,-1,-1)submission 顺从]	A 对 B 很顺从
	11’49	15’9	604	B	[(0,0,-1,0)fear 担忧, (0,0,-1,-1)submission 顺从]	A 对 B 既担忧又顺从
妈妈向 A 表达了对 B 的认可，A 很感动	15’19	17’57	263	妈妈	[(0,1,-1,0)anxiety 紧张]	A 因为妈妈感到紧张
	15’52	18’56	474	妈妈	[(1,0,-1,0)excitement 激动, (0,1,-1,0)anxiety 紧张]	A 因为妈妈感到既紧张又激动
A 严肃地拒绝了 C 突然的表白	35’52	38’18	166	C	[(0,1,-1,0)anxiety 紧张]	A 因为 C 感到紧张
	36’13	40’8	878	C	[(-1,1,-1,0)worry 担心,(0,1,-1,0)dread 恐惧]	A 因为 C 感到既担心又恐惧

（3）生成文本的下一步计划——丰富句式

上面的应用只是“主谓宾”的简单构造，实际上这样的句式过于简单，只是一个文本的雏形。我们可以补充以下方面作为改进：

A. 补充人物关系等额外信息

在上面的文本中，A 和 B 只是人物的名字或是代称，并不能反映两者之间的关系。我们可以利用台词来获取故事发生的背景信息。

示例：A 对于 B 的表白感到不安

B. 生成 A 和 B 共同为主语的句子

我们可以利用多人情绪状态的组合来丰富词表，例如：A 和 B 的状态都是“生气”或是“激动”，我们可以推测他们在“争吵”。

示例：A 和 B 激动地争吵。

C. 补充复合词和流行用语，丰富情绪词汇

目前的词库只包含了 81 种情绪状态的词汇，实际上有很多情绪是难以表达的。所以希望丰富情绪词库，来构建情绪状态的新表达方式。

示例：A 觉得这个结果喜忧参半。B 对于这个结果十分佛系。

D. 补充肢体动作

我们可以继续训练肢体动作模型，通过剧照判断动作。

示例：A 气愤地指着 B。C 开心地转圈。

E. 补充面部表情

和上一点相似的是，我们可以继续训练面部表情模型，通过剧照判断表情。

示例：A 因为 B 开心地笑了。C 被 D 气得流泪。

（4）生成文本的后续应用

根据情绪词生成文本后有许多应用，大的方向有进一步加工生成电视剧、电影的推送标题、推荐语等，或是“翻译”影视中的高潮片段，或者逆向用文本对视频做检索、视频混剪助手等等。

A. 推荐语

具体而言，我们在得到全部的句子后，可以根据情绪状态的多少从中挑选最合理、最典型的句子，进一步加工成影视作品的推送标题或是推荐语。如上面的例子可以用问句表示：“为何 A 对 B 表示担忧又焦虑？”“点击收看 A 和 B 的感动片段”等等类似的推荐语。

B. “翻译”视频

利用以上方法可以“翻译”视频中的高潮片段，在出现重大情绪转折的时候给出弹幕的文本提示。

C. 视频混剪助手

在剪辑视频的时候可以根据文本提示智能剪辑，作为一个视频混剪助手。

D. 逆向用文本对视频检索

有时用户会希望根据自己的需求寻找片段，这时就可以利用以上模型输入主语、宾语、以及情绪词来定位到期望的片段。

3. 计算每种情绪维度的区间密度

（1）计算过程

我们将 4 个情绪维度的 8 种基本情绪转换为 one-hot 向量，进行滑窗处理后可以求出每 1000 帧中每个基本情绪的数量。

同样进行区间密度的阈值分割，由于一个特定情绪维度的存在会比一个情绪状态的存在更加广泛和普遍，因此每个基本情绪的区间密度阈值也会相较更大。我们选择 0.8 为阈值（FPS=8），即 1000 帧中某一基本情绪出现的次数超过 100 帧的时候我们认为这个维度的基本情绪是存在的。

如下表所示，除了“快乐 joy”、“预期 anticipation”、“害怕 fear”三个基本情绪，其余基本情绪在这几个区间的记录数都没有超过 100，因此省略。

表 4.5 文本生成结果示例

开始帧	结束帧	“快乐 joy” 记录数	“预期 anticipation” 记录数	“害怕 fear” 记录数	emotion
1667	2668	103	189	320	(1,1,-1,0)
1668	2669	102	189	319	(1,1,-1,0)
1669	2670	101	189	318	(1,1,-1,0)
1670	2671	100	189	317	(0,1,-1,0)
1671	2672	99	189	316	(0,1,-1,0)
1672	2673	98	189	315	(0,1,-1,0)

除了情绪轮盘中基本情绪可以组合的特点，还可以把情绪轮盘更多的特性囊括进来。例如，更高浓度的基本情绪应为层级更高、情绪轮盘中心一圈的情绪，例如密度更高的“快乐”情绪对应着狂喜，密度更高的“悲伤”对应着“悲痛”。我们可以划分一个更高的区间密度的阈值，来对应这一点。

（2）主要应用：情绪变化

如图所示是上例中一集时间内男主角和女主角每个维度的情绪变化，其中前半段是男女主角有对手戏的部分。

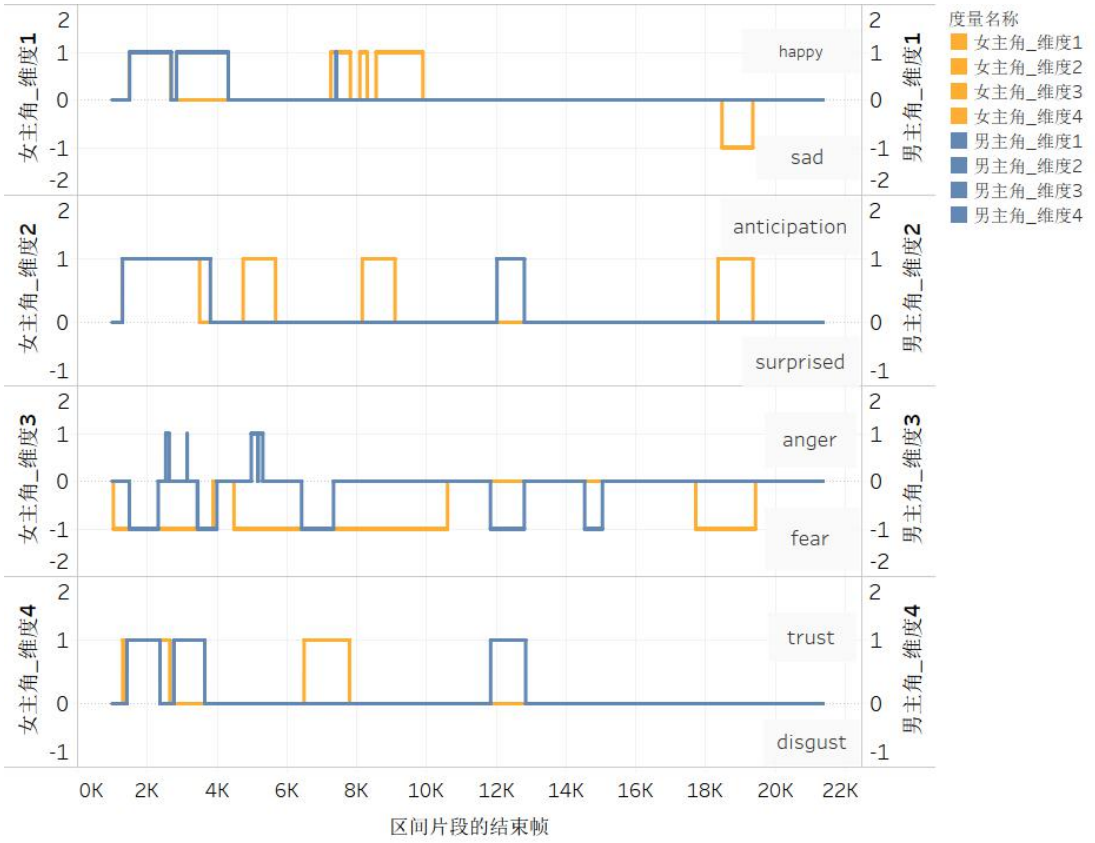

图 4.5 一集电视剧中两主角每个维度的情绪变化图表

可以看出两人对手戏时，情绪维度 1 都是“joy”，情绪维度 2 都为“anticipation”，情绪维度 4 都是“trust”，说明两人相处的过程是甜蜜的、互相信任的、对于未来有着共同期盼的。

而其中维度 3 又有着不同。女主角“fear”的比例更大，而相比之下男主角的情绪中含有很多“anger”的成分。在之前的定义中，我们把“生气-害怕”总结为人基于其自我认知是否处于有能力影响周围环境的状态，因此我们可以分析出两人的关系中，男主角更多地占在主导地位，比较强势地输出自己的观点，而女主角更多的时候是个“顺从”的角色。而这一分析也与本剧中两人“强势霸道”、“纯情乖巧”的人设相符合。

（3）情绪变化的应用

A. 识别影视剧中的感情线

在大众文化发展的泛娱乐化的今天以及各种媒介融合背景下，观看恋爱题材的影视作品满足了很多年轻观众放松减压的需求。

如果男女主之间的对手戏的情绪中“开心 joy”和“信任 trust”占比较多，则可以判断两人之间的关系更“甜”，满足观众对于恋爱向的期待。

由于每个演员和角色表达情绪的区别，为了控制变量，具体实现为探索同一角色对待不同亲密关系（家人，恋人，朋友）时的情绪占比区别。如果在对于“恋人”这一亲密关系中情绪流露出更多的部分，则可以看出男女主之间的“CP 感”更足，则该剧成为爆款的可能性也就越大，有着更高的商业价值。

通过情绪轮盘模型对于“CP 感”的识别，判断感情线中微妙的情感交流，进而可以在播出前就预测电视剧的播放量或电影票房。

B. 评判演员演技

演技的含义是指在舞台或是摄像机前，通过语言、姿态、动作来扮演某角色的艺术。演技是以正确的语调来诠释不同的情绪。情绪轮盘的模型可以很好地评判演员的演技，主要通过两个方面：表达情绪数目与情绪的衔接。

- 通过表达不同情绪的数目判断演员演技

通常情况下，我们认为优秀的演员能够表达的情绪种类更加丰富，则可以据此判断演员的演技。与此同时，也可以定性地评判演员演出的情绪与剧本中期望表达的情绪是否相同。

近几年诞生了许多演技竞演类综艺或年轻演员品训真人秀节目，即选手演绎经典剧目，可以通过比较选手的演绎和原版的演绎来评判演员演技的高低，从而让演艺圈内的年轻人正确认识演员的使命，珍惜自己的职业，为广大观众呈现更优秀的作品，弘扬正能量文化。在这些综艺节目中便可以应用情绪识别的技术来判断选手的情绪是否达到了剧本的标准，以及通过计算演绎相同片段、相同剧本时，不同演员表达情绪种类的多少来定量评判演技。

- 通过情绪的过渡和衔接判断演员的入戏程度

优秀的演员表达情绪的时候，一般过渡都会很自然；而在演出戏剧张力大的片段时在表达“吃惊”、“悲伤”等比较夸张的情绪上又很直接，情绪转换很快，

入戏的程度很深，这也都是演技的体现。

4. 与现有的表情识别模型的对比

如下图是某角色的四个视频片段每一帧的图片采取单一输出值的表情识别的结果，输出值为 8 个表情，对应的情绪也是“开心”、“悲伤”、“惊讶”、“厌恶”、“平静”等唯一值，四个视频片段实际代表的情绪是“开心”、“紧张”、“害怕”、“惊恐”。可以看出绝大多数的情绪体现的都是“平静”这一表情，由于输出值有 8 种，所以变化的幅度比较大，很难看出情绪变化的规律，应用于文本生成的方面也很困难。相比之下，情绪轮盘的模型输出四元向量的形式就有更多的应用场景。

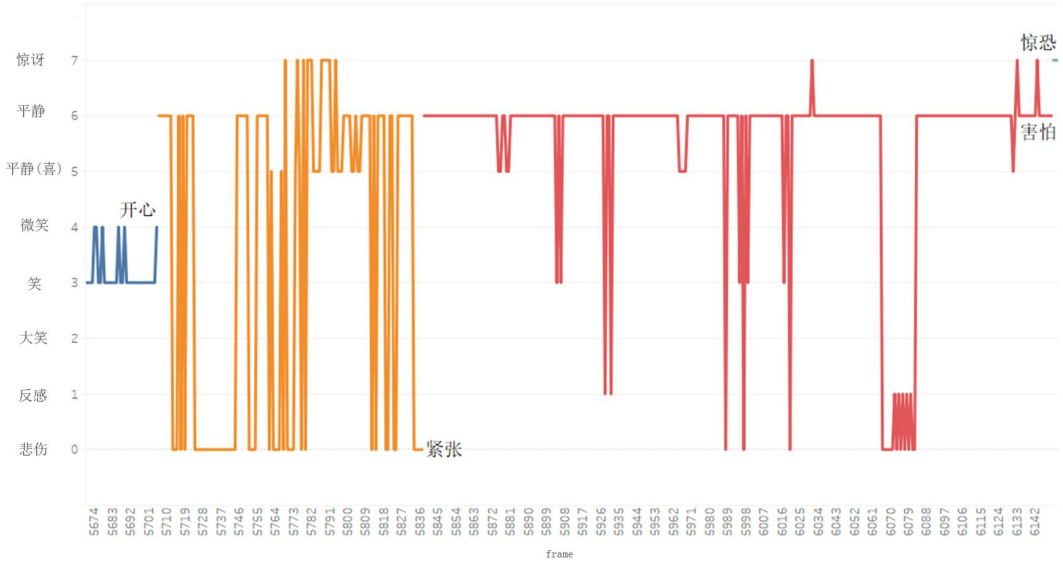

图 4.6 单一输出值的表情识别模型的情绪变化图表

4.2.2 推荐系统与市场营销

通过情绪识别推荐音乐或是视频，这是目前一项比较完善的工作。假设在工作或是驾驶的时候，摄像头实时识别面部表情根据情绪轮盘分析情绪，根据这些信息，音乐推荐系统播放不同风格的音乐。

在市场营销方面也有重大意义，例如收集志愿者在观看广告时的情绪，来预测产品销量，从而调整广告的投放量。

此外，基于面部编码的情绪识别在人们不愿意主动说出情感时具有更大的价

值。在一些思想、文化较为保守的地方，人们不便于直接说出自己的真实喜好，即使十分喜爱一件产品也不会流露出来，对其不满也不会直接表达出来，可以利用情绪识别来推测受访者内心的真实情绪，进行产品或是理念的推广。

4.2.3 交通方面——监控路怒症、提醒疲劳驾驶

Affectiva 公司推出的车载情绪识别系统可发出警告，安抚路怒司机情绪或者提醒心不在焉的司机专注于开车[21]。此情绪识别系统的主要元件是一个对焦在驾驶人面部的摄像头，基于对五百万种面部表情的深度学习，可识别 33 种面部特征，而后车载电脑的中枢结构网络将这些数据信息破译出来，识别 7 种情绪，包括喜悦、惊讶和恐惧。该系统不会将其录制的视频发送到中心服务器上，而是可以一次性地实时完成所有情绪识别过程。

如果利用情绪轮盘系统作为辅助，则可以更加准确地识别出驾驶者的情绪，更加及时地发出警报。

4.2.4 其他应用

现有的情绪识别大多都是辨别心情的好坏，基于情绪轮盘模型的情绪识别在更深层次的应用，如：焦虑或是平和、自卑或是自信。同时也需要注意在这些应用中需要保证人脸的隐私不被泄露、面部数据不被商业利用等问题。

1. 教育行业

如今的大环境下，远程教学时代也是一个使得情绪识别更广泛应用的契机。学生利用电子设备观看教学视频也便于收集面部情绪的数据，让教师更容易发现哪些学生更具有学习能力，以及哪些学生对于一些内容学习地很吃力，更有效地因材施教。

2. 心理行业

识别紧张、焦虑等情绪可以帮助进行心理检测，辅助于各行各业的人们缓解压力，以更加积极健康的心态生活和学习。

有些国家的公司，在招聘经常处理紧急状况、需要很大抗压能力的员工时（例如：记者、翻译等），会在面试中利用 AI 识别面试者的紧张情绪，做出心理层面的考核。

AI 识别嫌疑人的紧张情绪对于心理侧写、识奸除恶也对于有很重要的应用。

3. 医疗行业

情绪识别可以帮助很多有需要的人，例如：在手语中，有很多词的手语表示相同但词义大相径庭，这时区分就在于面部表情。智能地实时识别情绪可以帮助聋哑人理解语义。情绪识别还可以帮助自闭症患者识别社交信息，更好地融入社会。

以上这些领域中，“预感 Anticipation”和“信任 trust”这两维度尤为重要，Anticipation 的比重越高，代表对于情况有更多的掌握，即：学生理解知识更加透彻，人们更加放松等等，反之即表现出很紧张、焦虑的态度；trust 的比重更高，代表人们更相信自己说的话，也更加享受目前的环境，即更加坦然、自信。这也是目前的表情识别没有涉及到的领域。

第五章 工作总结和未来展望

5.1 未来展望和改进空间

本文基于情绪轮盘模型进行了深度学习模拟，该模型的效果虽然已经达到了基本要求，但仍然还有很多改进的空间。

1) 首先是训练数据集本身的问题，由于时间有限，不能仔细筛选影视作品的构建数据集的，很多微表情无法捕捉，导致情绪划分的不完全准确；fer2013数据集中有采集错误，以及分辨率较低，人类对于此数据集也只有65%+-5%的准确度。由于数据集本身存在一定的不完美的地方，同时数据数量并不很大导致了一定的过拟合现象。可以继续收集训练数据。

2) 由于情绪是很难捕捉与划分的，是否可以探测到情绪并没有一个完美统一的标准，故正负一与0之间的划分界限实际上很模糊。可以考虑在下一步使用逻辑回归的方法，将每个维度的输出值统一为0-1之间的一个数，而后划分阈值，大于α记为+1，小于β记为-1，位于α和β之间的情绪记为0。这样可以对于不同的维度划分不同的阈值，更可以保证情绪划分的合理性和准确性。

3) 由于没有足够的数据，很多应用没办法以更加形象具体的形式展现出来。如果可以收集到更多场景下的面部数据，则应用方面也可以更加完整。

5.2 工作总结

本文总结了普洛特契克的情绪轮盘模型的提出和相关概念，把情绪定义为4个维度、8种两两对立的基本情绪（快乐对悲伤、信任对厌恶、恐惧对生气、惊讶对期待）的组合，并总结了现有并流行的情绪识别的深度学习网络框架。

本文利用CNN网络构造采取了深度学习的方法构建了情绪轮盘结构，返回值为4维向量，识别覆盖了81种情绪状态，对应于几百个情绪词。从影视作品中获取训练数据，并对于网络上人脸数据集作为测试集进行检测，测试效果均令人信服。四维向量的输出结果比单一结果的返回值的覆盖面更广、结果也更准确，因此有着非常广泛的应用。

本文讨论了情绪轮盘模型的应用场景，相比于一维的表情识别，情绪轮盘模型可以帮助我们更好地理解不同情绪之间的联系和差异。本文主要探究了在影视

领域的应用，主要有文本生成和情绪变化两个大的模块应用，可以利用此生成视频的推荐语以及评判演员演技。除此之外，本文也讨论了情绪轮盘模型在推荐系统、心理状态检测等方面的重要应用价值。

参考文献

[1] Plutchik,Robert.The Emotions[M].University Press of America,1991

[2] Plutchik,Robert.A general psychoevolutionary theory of emotion[A].In: R. Plutchik & H. Kellerman. Emotion: Theory, research, and experience: Vol. 1. Theories of emotion[M].New York: Academic，1980:3-33

[3] Jonathan Turner. On the Origins of Human Emotions: A Sociological Inquiry Into the Evolution of Human Affect[M].Stanford University Press. 2000:76

[4] Jianhua Zhang,Zhong Yin,Peng Chen,Stefano Nichele. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review[J]. Information Fusion,2020,59:103-126.

[5] 陈子健,朱晓亮.基于面部表情的学习者情绪自动识别研究——适切性、现状、现存问题和提升路径[J].远程教育杂志,2019,37(04):64-72.

[6] 刘大诚.人工智能情绪识别应用研究[J].中国高新科技,2019(13):59-62.

[7] 李卓远,曾丹,张之江.基于协同过滤和音乐情绪的音乐推荐系统研究[J].工业控制计算机,2018,31(07):127-128+131.

[8] 陆嘉慧，张树美，赵俊莉. 基于深度学习的面部表情识别研究[J].计算机应用研究.2020,37(04):966-972.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks[A].In: Advances in neural information processing systems[C].2012:1097 - 1105.

[10] K. Simonyan and A. Zisserman. Very deep convolutional networks for large scale image recognition[A].In: 3rd International Conference on Learning Representations [C].2015:1-14.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions[A].In:the IEEE conference on computer vision and pattern recognition[C].IEEE,2015:1 - 9.

[12] K. He, X. Zhang, S. Ren, and J. Sun.Deep residual learning for image recognition [A].In:the IEEE conference on computer vision and pattern recognition [C].IEEE,2016:770 - 778.

[13] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler.Deep learning for emotion recognition on small datasets using transfer learning[A]. In: ACM on international conference on multimodal interaction[C].ACM,2015:443 - 449.

[14] G. Levi and T. Hassner.Emotion recognition in the wild via convolutional neural networks and mapped binary patterns[A]. In: ACM on international conference on multimodal interaction[C].ACM,2015: 503 - 510.

[15] J. Cai, Z. Meng, A. S. Khan, Z. Li, J. OReilly, and Y. Tong. Island loss for learning discriminative features in facial expression recognition [A]. In: 13th IEEE International Conference on Automatic Face & Gesture Recognition[C]. IEEE, 2018: 302 - 309.

[16] S. A. Bargal, E. Barsoum, C. C. Ferrer, and C. Zhang. Emotion recognition in the wild from videos using images[A] In: the 18th ACM International Conference on Multimodal Interaction[C]. ACM, 2016:433 - 436.

[17] S. Li and W. Deng. Deep Facial Expression Recognition: A Survey [A]. In: IEEE Transactions on Affective Computing[C]. IEEE, 2020:1-1.

[18] B.K. Kim, H. Lee, J. Roh, and S.Y. Lee. Hierarchical committee of deep cnns with exponentially-weighted decision fusion for static facial expression recognition[A], In: International Conference on Multimodal Interaction[C]. ACM, 2015:427 - 434.

[19] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.H. Lee et al. Challenges in representation learning: A report on three machine learning contests[A]. In: International Conference on Neural Information Processing[C]. 2013:117 - 124.

[20] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets[A]. In: Automatic Face and Gesture Recognition[C]. IEEE, 1998:200 - 205.

[21] Martin Magdin, F. Prikler. Real Time Facial Expression Recognition Using Webcam and SDK Affectiva[J]. International Journal of Interactive Multimedia and Artificial Intelligence, 2018, 5:7-15

[22] 戴维·塔尔博特. 情绪识别应用超乎想象[J]. 科技创业, 2014(Z1):11-12+14.

[23] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment[A] In: CVPR[C]. IEEE, 2013:532 - 539.

[24] B. Fasel. Robust face analysis using convolutional neural networks[A] In: Pattern Recognition[C]. IEEE, 2002:40 - 43.

[25] C. Darwin and P. Prodger. The expression of the emotions in man and animals[M]. Oxford University Press, USA, 1998.

[26] Y.I. Tian, T. Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis[A]. In: IEEE Transactions on pattern analysis and machine intelligence[C]. IEEE, 2001:97-115.