# 第3章
# 情感计算模型

为了准确刻画情感状态，需要基于有效的情感表示模型对不同粒度的情感进行描述，在此基础上分析情感相关的属性。现有研究方法对于情感表示尚缺乏统一的定量测量评价标准，需要结合具体的目标设定。人类的情感可以通过行为表现、生理唤醒和主观体验进行体现，但是对于计算机而言，并不具有人类的上述特殊功能，难以对情感状态进行有效区分；情感的分类粒度、精确度以及覆盖程度很大程度上影响着情感分析的性能，它是情感计算领域急需解决的关键难题。本章重点阐述了当前主流的离散情感计算模型和连续情感计算模型，并进一步拓展介绍了基于个性化的情感模型。

## 3.1 离散情感计算模型

本节主要介绍离散情感计算模型，具体包括基本情感论、离散情感数据库、离散情感评价标准。

### 3.1.1 基本情感论

根据情感的纯度和原始度，情感可以分为两大类：基本情感和复合情感，这就是基本情感论。基本情感论认为人们与生俱来的情感在发生上有原型形式，即存在多种基本情感类型，每种情感类型都有其独特的体验特性、生理唤醒模式和外显模式。通常情况下，悲伤与丧失的知觉相关，恐惧与受到惊吓和身体受到伤害的知觉相关，生气与侮辱或不公平的知觉相关。不同形式的组合形成了人类的所有情感。国内外研究者对情感状态的分类有很长时间的争论，有 4 种情感得到了最为普遍的认同，它们是恐惧、愤怒、悲伤和高兴。近年来，很多研究者提出的基本情感类别存在差异，从 2 种到二十几种不等。情感表示如图 3-1 所示。

伴随人类认识客观世界的过程，基本情感可以通过多种方式定义。1962 年，汤姆金斯提出有 8 种基本情感：恐惧、愤怒、痛苦、高兴、厌恶、惊奇、关心、羞愧。谢弗等学者认为情感有 6 种基本类别，分别是爱、喜悦、惊奇、愤怒、悲伤和恐惧。进一步将基本情感和对应的面部表情与其他属性相关联。表 3-1 列举了不同学者对基本情感的划分，其中，美国心理学家艾克曼提出的 6 大基本情感(生气、厌恶、恐惧、高兴、悲伤和惊讶)在当今情感相关研究领域使用较为广泛。

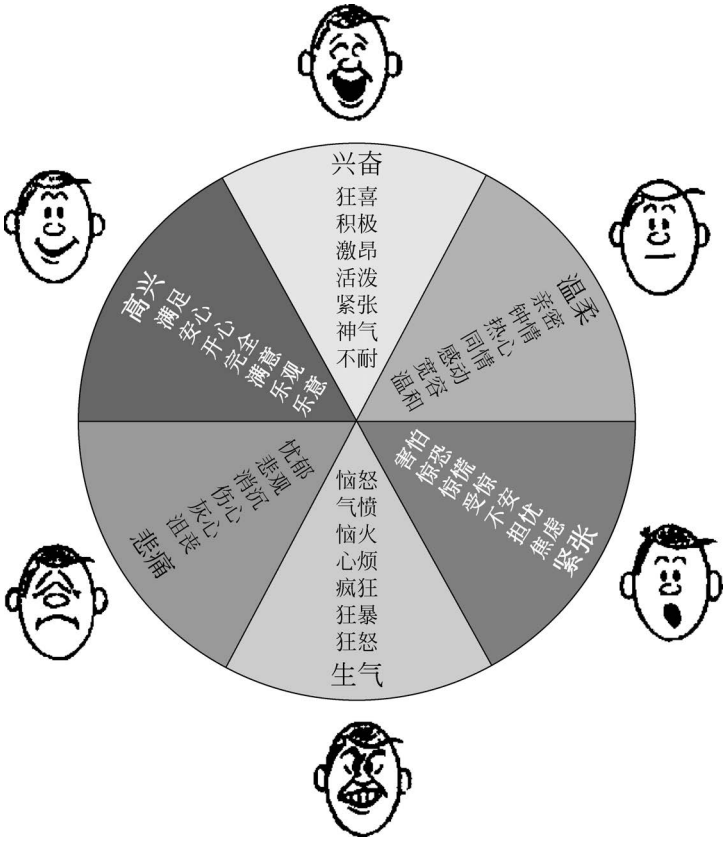



图 3-1　情感表示

**表 3-1　基本情感的定义**

| 编号 | 基本情感定义 |
|---|---|
| 1 | 满意、不满意 |
| 2 | 正面情感、负面情感 |
| 3 | 接纳、愤怒、期待、厌恶、喜悦、恐惧、悲伤、惊奇 |
| 4 | 愤怒、厌恶、勇气、心情低落、欲望、绝望、恐惧、仇恨、希望、爱情、悲伤 |
| 5 | 愤怒、厌恶、恐惧、快乐、悲伤、惊奇 |
| 6 | 欲望、快乐、兴趣、惊奇、悲伤 |
| 7 | 愤怒、恐惧、焦虑、喜悦 |
| 8 | 愤怒、轻蔑、厌恶、痛苦、恐惧、内疚、兴趣、快乐、羞愧、惊喜 |
| 9 | 恐惧、悲伤、爱情、愤怒 |
| 10 | 愤怒、厌恶、得意、恐惧、屈从、柔情、难怪 |
| 11 | 痛苦、快乐 |
| 12 | 愤怒、厌恶、焦虑、悲伤、幸福 |

除了前面介绍的将情感分为基本情感和复合情感，还有其他一些分类方法。福克斯提出的三级情感模型，按照情感中表现的主动和被动的程度不同将情感分成不同等级，如表 3-2 所示。等级越低，分类越粗糙，等级越高，分类越精细。过细的情感分类不一定对情感计算的研究有很大的意义。情感分得越细，情感特征会更加模糊，从而影响到情感分析的

性能。当前的离散情感模型，多采用 4～6 种情感类别。

表 3-2 情感三级分类模型

| 层 级 | 情 感 | | | | | |
|---|---|---|---|---|---|---|
| 1 级 | 接近 | | | 退回 | | |
| 2 级 | 高兴 | 兴趣 | 愤怒 | 难受 | 厌恶 | 害怕 |
| 3 级 | 骄傲 | 关心 | 敌意 | 痛苦 | 藐视 | 恐惧 |
| | 祝福 | 责任 | 嫉妒 | 烦躁 | 怨恨 | 焦虑 |

### 3.1.2 离散情感数据库

**1. 语音情感数据库**

目前，语音情感数据库的建立尚缺乏统一标准，语音情感数据库可分为表演型、引导型、自发型。表演型情感数据库通常是让职业演员以模仿的方式表现出相应的情感状态，虽然表演者被要求尽量表达出自然的情感，但刻意模仿的情感还是显得有些夸大，使得不同情感类别之间的差异性比较明显。表演型的语音情感数据库有柏林 EMO-DB 德语情感语音库和 CASIA 汉语情感语料库等。早期对语音情感识别的研究都是基于表演型语料库，随着人们意识到引导型情感具有更加自然的情感表达之后，研究者们开始基于引导型情感数据库进行研究，比如 eNTERFACE。随着对自然场景下真实情感状态的分析不断深入，迫切需要一些自发的语音情感数据库，包含语音信息的自发型情感数据库包括 FAU Aibo 数据库、TUM AVIC 数据库、SUSAS 数据库、VAM 数据库、DES 数据库。常用的几个语音情感数据库如表 3-3 所示，表中描述了不同数据库在年龄、语言、情感、样本个数、记录环境和采样率之间的差异。

表 3-3 不同语音情感库之间的差异

| 语料库 | 年龄 | 语言 | 情感 | 样本个数 | 记录环境 | 采样率/kHz |
|---|---|---|---|---|---|---|
| FAU Aibo | 小孩 | 德语 | 自发型 | 18216 | 正常 | 16 |
| CAISIA | 成人 | 汉语 | 表演型 | 9600 | 工作室 | 16 |
| Emo-DB | 成人 | 德语 | 表演型 | 494 | 工作室 | 16 |
| eNTERFACE | 成人 | 英语 | 引导型 | 1277 | 正常 | 16 |
| SUSAS | 成人 | 英语 | 自发型 | 3593 | 噪声 | 8 |
| VAM | 成人 | 德语 | 自发型 | 947 | 噪声 | 16 |
| TUM AVIC | 成人 | 英语 | 自发型 | 3002 | 工作室 | 44 |
| DES | 成人 | 丹麦语 | 表演型 | 419 | 工作室 | 48 |

下面详细介绍 3 种较常用的语音情感数据库。

(1) FAU Aibo 录制了 51 名儿童(10～13 岁，21 男 30 女)在与索尼公司生产的电子宠物 AIBO 游戏过程中的自然语音，并且只保留了情感信息明显的语料，总时长为 9.2h(不包括停顿)，包括 48401 个单词。使用一个无线高保真麦克风收集语音，由 DAT-recorder 工具录制，语音格式为 48kHz 采样率，16bit 量化。为了记录真实情感的语音，工作人员让孩子们相信 AIBO 能够对他们的口头命令加以反应和执行；实际上，AIBO 则是由工作人员暗中人为操控的。标注工作由 5 名语言学专业的大学生共同完成，并通过投票方式决定最终标注结果，标注涵盖包括高兴、愤怒、生气、中性等在内的 11 个情感标签。

(2) CAISIA 汉语情感语料库是由中国科学院自动化研究所录制的。语料设计包含 6

类不同情感：高兴、悲哀、生气、惊吓、难过、中性。每种情感有 50 句语料，由 4 位录音人(2 男 2 女)在纯净录音环境中对 50 句语料赋予不同的情感演绎而得到。语音信号采用 16kHz 采样率以及 16bit 量化。经过听辨筛选，最终保留其中 1200 句语音样本。

(3) Emo-DB 是由柏林工业大学录制的德语情感语音库，由 10 名演员(5 男 5 女)对 10 个语句(5 长 5 短)进行 7 种情感(高兴、生气、焦虑、害怕、无聊、厌恶和中性)的演绎而得到，共包含 535 句语料。语音信号同样采用 16kHz 采样以及 16bit 量化。语料文本的选取遵从语义中性、无情感倾向的原则，且为日常口语化风格，无过多的书面语修饰。语音录制在专业录音室中完成，要求演员在演绎某个特定情感前通过回忆自身真实经历或体验进行情感诱发，以增强情感的真实性。经过 20 个参与者(10 男 10 女)的听辨实验，得到 84.3%的听辨正确率。

**2. 视频数据库**

目前主要的视频情感语料包括：① HUMAINE 数据库；② SEMAINE 数据库；③IEMOCAP 数据库；④CHEAVD 2.0 数据库。主要的人脸表情数据库包括：①JAFFE 数据库；②Cohn-Kanada Facial Expression 数据库；③BHU 人脸表情数据库；④FEEDTUM 数据库；⑤MMI 面部表情数据库；⑥Oulu-CASIA 面部表情数据库。下面具体介绍上述的 10 个数据库。

(1) HUMAINE 是从诱导性数据中提取 50 个片段，内容包括身体姿势、面部、声音、言语内容等，说话者的性别以及文化背景都呈现多样化。

(2) SEMAINE 是一个相对较大规模的视频数据库，其中有 150 个参与者、959 个对话。有 6～8 位标注者对数据进行情感标注，共包括 27 个情感类别。

(3) IEMOCAP 是由南加州大学录制的情感数据库，包含约 12h 的视听数据。10 名专业演员(5 男 5 女)在有台词或即兴的场景下，诱发出情感表达。人工将每段对话切分成单句，每个单句至少由 3 个标注者进行类别标注。为了平衡不同情感类别的数据量，通常将高兴和兴奋合并成高兴类别。由高兴、生气、悲伤和中性最终构成了 4 类情感识别数据库。

(4) CHEAVD 2.0 数据库由中国科学院自动化研究所建立，从电影、电视和综艺节目中剪辑出情感片段，共计 474min。依据以下几个原则选取数据：①待剪辑的视频面向日常生活场景；②避免选取有浓厚口音的片段；③避免选取表演痕迹过重的片段。数据库由 4 位标注者进行标注，共覆盖 11 种出现较多的情感类别以及出现较少的 11 种情感，具体分布如表 3-4 所示，数据库中几种典型的情感片段如图 3-2 所示。

**表 3-4　出现较多的 11 种情感**

| 出现较多的情感类别 | 片段数 | 出现较多的情感类别 | 片段数 |
|---|---|---|---|
| 高兴 | 400 | 生气 | 596 |
| 中性 | 679 | 焦虑 | 119 |
| 悲哀 | 351 | 害怕 | 35 |
| 厌恶 | 92 | 紧张 | 14 |
| 惊讶 | 74 | 无助 | 11 |
| 担心 | 62 | | |

(5) JAFFE 数据库由日本 Kyushu 大学建立，该数据库是由 10 位日本女性在实验环境下根据提示做出各种表情，再由照相机拍摄获取人脸表情图像。整个数据库共有 213 张图


图 3-2 中文多模态情感语料库几种典型的情感

像，每人做出 7 种表情，分别是悲伤、高兴、生气、厌恶、惊奇、害怕、中性。每组都含有上述 7 种表情，每种表情包括 3～4 张图像。每组有约 20 张图像，图片大小为 256×256，如图 3-3 所示。

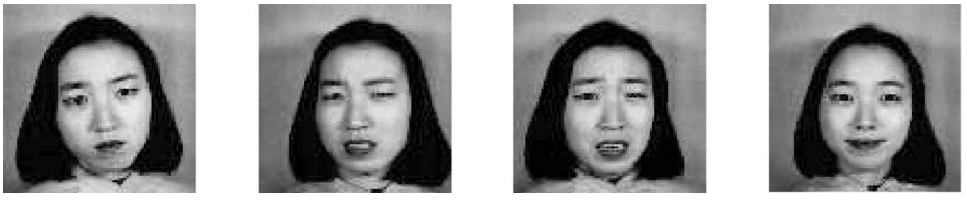


图 3-3 JAFFE 数据库部分表情

(6) Cohn-Kanade Facial Expression 数据库由美国卡耐基·梅隆大学机器人实验室建立。这个数据库包括 123 个被试个体、593 个图像序列，每个序列的最后一张图像都有活动单元的标记，在这 593 个图像序列中，有 327 个图像序列是有情感标签的，部分实例如图 3-4 所示。

图 3-4 Cohn-Kanade Facial Expression 数据库中不同表情的样本图像

(7) 北京航空航天大学毛峡教授团队建立的 BHU 人脸表情数据库是面向国内人群的数据库。数据库中包括 21～25 岁的 18 名女性和 14 名男性的 25 种面部表情，其中有 18 种单纯面部表情、3 种混合面部表情和 4 种复杂面部表情，相比较于其他数据库，该数据库中除了微笑、大笑、嘲笑、生气等 18 种单纯表情外，增加了惊奇-高兴、惊奇-悲伤、惊奇-生气等混合面部表情，部分实例如图 3-5 所示。

图 3-5 BHU 数据库中不同表情的样本图像

（8）FEEDTUM 数据库由慕尼黑大学的人机交互实验室建立，共包括 19 个被试个体，每个被试个体采集 21 张表情图像，数据库中包含 6 种基本表情，皆为 3 通道 RGB 格式，大小为 320×240，如图 3-6 所示。

图 3-6　FEEDTUM 数据库中不同表情的样本图像

（9）MMI 数据库由荷兰代尔夫特理工大学建立，提供 1500 多个表情正脸或侧脸的静态图像和图像序列，均为 3 通道 RGB 图像。参加采集的被试中 44%是女性，年龄为 19～62 岁，来自欧洲、亚洲或南美洲。图 3-7 所示为 MMI 数据库中的示例图像。

图 3-7　MMI 数据库中的不同表情的样本图像

（10）Oulu-CASIA 数据集由中国科学院自动化研究所和芬兰奥鲁大学联合建立，包含 6 种基本表情：生气、厌恶、害怕、高兴、悲伤和惊讶。数据库中受试年龄分布在 23～58 岁，共包括 80 个受试者。图 3-8 所示为在明亮、弱光、黑暗 3 种不同光照条件下采集的数据，每种光照条件下包括 480 个序列（80 个受试分别采集 6 种不同表情）。所有表情序列都是从中性开始，到表情强度最大时结束。

图 3-8　Oulu-CASIA 数据库中的不同表情的样本图像

**3. 生理信号数据库**

在国际上被学术界公认的生理信号情感数据库主要有德国奥格斯堡大学建立的情感数据库和上海交通大学建立的脑电情感数据集 SEED，这些数据库均对受试者在不同情感状态下的多种生理信号进行采集并用于情感识别的研究。

（1）德国奥格斯堡大学建立的情感数据库是利用音乐对受试者进行不同类型的情感诱发实验，进而采集受试者在不同情感状态下的多种生理信号。该数据库是对单一受试者进行实验，为了让音乐能够有效唤起受试者的情感状态，要求受试者选择对其有特殊意义的歌曲，这些歌曲能够唤起受试者的特殊回忆进而能够较容易唤起受试者的不同情感状态。该数据库同时对维度情感状态和离散情感状态进行标注，通过效价度-唤醒度定义二维情感表示模型标注受试者连续情感变化，通过高兴、愤怒、悲伤、愉悦 4 种具有一定差异性且较常见的情感状态进行离散情感标注。受试者可以自己挑选对其具有影响性的音乐，同时受试者

在听音乐过程中想象相应的场景，以便对受试者进行更加全面的情感诱发。当受试者聆听不同音乐进行情感诱发时，通过4通道的生物传感器对受试者的4种生理信号同时进行采集，分别为心电图(Electrocardiogram，ECG)信号、肌电(Electromyogram，EMG)信号、皮肤电(Skin Conductance，SC)信号及呼吸(Respiratory，RSP)信号。每种生理信号的数据采集时间均为2min。不同生理信号的采样率不同，其中ECG信号的采样率为256Hz，EMG信号、SC信号及RSP信号的采样率均为32Hz。为了保证受试样本充足，需要连续25天对受试者进行情感诱发实验及信号采集，因此该数据库的样本量为100，每种情感状态下各有25个样本。该数据库的详情见表3-5。

**表3-5 德国奥格斯堡大学情感数据库内容**

| 数据库内容 | 具体内容 |
|---|---|
| 情感诱发素材 | 音乐＋想象 |
| 情感类型 | 高兴、愤怒、悲伤、愉悦 |
| 受试人数 | 1人 |
| 采集天数 | 25天 |
| 生理信号类型 | 心电(ECG)、肌电(EMG)、皮肤电(SC)、呼吸(RSP) |
| 信号采样率 | ECG：256Hz；EMG、SC、RSP：32Hz |
| 样本容量 | 100(其中每种情感装备样本均为25个) |

(2) 脑电情感数据库SEED使用电影片段作为情感诱发材料，共分为3种情感类别：愉悦、平静、悲伤。每个电影片段时长大约4min，每次实验由15个电影片段组成。在每次实验中，3种情感状态的电影片段数量相等，均为5次。电影片段来自中文电影，在每段电影片段放映之前有5s的提示，放映之后有45s的反馈时间与15s的休息时间。有15名受试者(7名男性、8名女性，平均年龄为23.27岁)参加实验，受试者均具有正常的视听觉能力。在受试观看电影片段的同时，通过电极帽记录他们的脑电信号，脑电信号采样频率为1000Hz。按照国际10-20系统，实验使用62通道的电极帽。每名受试参加3次实验，每次实验间隔为一周左右。将原始EEG降采样至200Hz，然后经过0.5～70Hz带通滤波，得到预处理后的脑电数据。按照每秒信号划作一个样本，则每名被试共有3394个样本，且在3分类的情感识别任务中，每个类别的样本数目大致相等。

### 3.1.3 离散情感评价标准

离散情感识别所使用的评价指标主要有分类准确率、召回率、精确率、$F$值等。设共有A和B两种类别，$n_{\mathrm{TP}}$是A类样本正确分类的样本数，$n_{\mathrm{FN}}$是A类样本错误分类的样本数，$n_{\mathrm{TP}}$是B类样本错误分类的样本数，$n_{\mathrm{TN}}$是B类样本正确分类的样本数，则整体分类准确率定义为

$$P_{\mathrm{acc}}=\frac{n_{\mathrm{TN}}+n_{\mathrm{TP}}}{n_{\mathrm{TN}}+n_{\mathrm{FN}}+n_{\mathrm{TP}}+n_{\mathrm{FP}}} \tag{3-1}$$

A类样本的分类准确率或召回率定义为

$$P_{\mathrm{re}}=\frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}}+n_{\mathrm{FN}}} \tag{3-2}$$

A类样本的分类精确率为

$$P_{\text{pre}}=\frac{n_{\text{TP}}}{n_{\text{TP}}+n_{\text{FP}}} \tag{3-3}$$

A 类样本的分类 $F$ 值定义为

$$P_{\text{f}}=\frac{2P_{\text{pre}}P_{\text{re}}}{P_{\text{pre}}+P_{\text{re}}} \tag{3-4}$$

## 3.2　维度情感计算模型

### 3.2.1　维度情感模型

离散情感模型将情感状态标注为离散的形容词标签，只能表示出有限种类的、单一明确的情感类型。离散情感模型具有简单直观的优点，在情感计算领域得到了广泛的应用。但是它存在着以下缺点：①离散情感模型能够表示的情感范围有限，例如悲喜交加、喜极而泣等复合情感并不属于某一基本情感类别，从而限制了这类模型进一步应用的普适性；②不同情感类别之间存在着高度的相关性，离散情感模型难以对这种相关性进行度量；③情感的生成、演化与消失是一个连续化的过程，而离散情感模型无法描述细微情感的连续变化。

为了克服离散情感模型的缺点，研究者建立了维度情感模型。冯特最早在 1896 年提出情感维度的观点，为维度情感模型的提出奠定了基础。维度情感模型认为情感是一种高度相关的连续体，从情感的多个维度量化了复杂情感的隐含状态，运用各种取值连续的基本维度将情感状态描述为多维空间中的某一个坐标，每个维度是对情感某一方面的度量。维度情感模型的本质是将不同的情感状态映射为多维情感空间中的点，典型的情感维度包括激活度、效价度、控制维等，每个维度都是动态连续的，不同强度的情感在维度情感空间中被映射为不同的坐标点。与离散情感模型相比，维度情感模型对情感的描述方式与人类的自然情感状态更接近，在情感计算中同样得到了广泛的研究与应用。二维情感模型和三维情感模型是最常用的维度情感模型。

二维情感模型可以使用激活度和效价度两个维度来对情感状态进行描述，它们分别表示情感的激烈程度和情感的正负性。如图 3-9 所示，在这种二维情感模型中，以情感的激活度和效价度为坐标轴，坐标原点表示没有任何情感状态的中性情感。激活度是指与情感状态相关联的机体能量激活的程度，是对情感内在能量的一种度量，表征个体对于各种活动的参与性；效价度主要体现为情感主体的感受，表征情感的积极或消极程度、喜欢或不喜欢程度、正面或负面程度。这种二维情感模型可以表示出多数基本情感状态，很多维度情感分析的研究都是在这两个维度

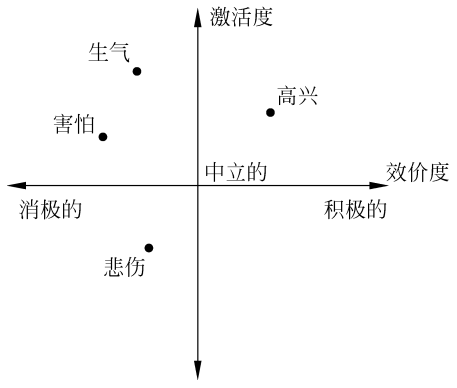



图 3-9　激活度-效价度二维情感模型

上进行的。

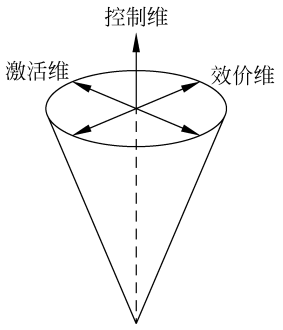



图 3-10 三维情感模型

在上述二维情感模型基础上增加一个控制维，从而构成了以激活度、效价度和控制维组成的一种典型的三维情感模型，控制维体现的是个体对情感状态的主观控制程度，用以区分情感状态是由主体主观发出的还是受客观环境影响产生的，比如轻蔑和恐惧，就处于控制维原点的两侧。如图 3-10 所示，以中性情感所在的原点为中心，沿着三条坐标轴方向延展，离原点越远说明情感的激烈程度越强，人对情感的控制能力越强，个体对情感的感受越强。

对于情感具有哪些维度，心理学家并没有统一的认识，其中认同度最高的一种模型为愉悦-唤醒-支配（Pleasure-Arousal-Dominance，PAD）模型，如图 3-11 所示。该模型将情感分为愉悦度、唤醒度和支配度三个维度。愉悦度也称作效价度，是对个体愉悦程度的度量；唤醒度也称作激活度，是对生理活动和心理警觉水平的度量，如睡眠、厌倦等为低唤醒，清醒、紧张等为高唤醒；支配度也称作注意度或能量度，是指影响周围环境及他人或反过来受其影响的一种感受，高的支配度是一种有力、主宰感，而低的支配度是一种退缩、软弱感。

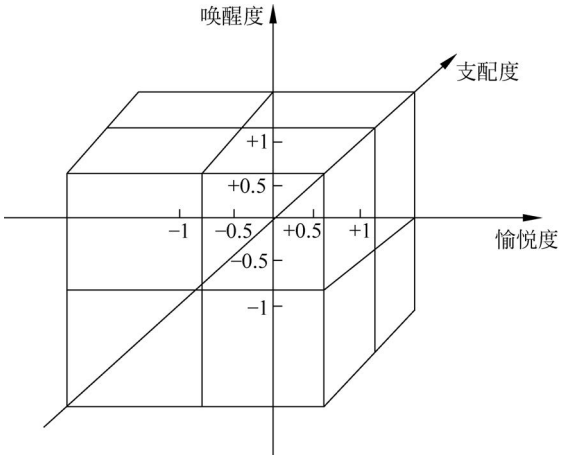



图 3-11 PAD 情感模型

PAD 情感模型中，每种情感状态都可以看作以 PAD 三维坐标系中的一个点。许多研究者通过实践总结出常见情感状态和对应的 PAD 坐标之间的映射关系，如梅拉宾研究了 12 种情感状态的 PAD 映射关系，德国的比勒菲尔德大学学者研究了 9 种情感状态的 PAD 映射，虚拟人则对 OCC 模型中的 24 种基本情感状态实现了基于 PAD 模型的情感表示。可以通过计算预测不同样本的 PAD 值与各种基本情感状态的 PAD 值的距离，判别待测样本的情感状态并分析其内在的情感组成，实现对情感状态的有效测量。中国科学院心理学研究所提出了汉化后的适用于评定汉语情感的 PAD 参考值，如表 3-6 所示。在 PAD 情感模型中，每种情感状态都可以与 PAD 空间的位置相对应，能够用唯一的三维坐标表示。

表 3-6　汉化 11 种情感语音数据的 PAD 得分

| 情感 | 语音评价结果([−1,1]) | | | 词汇评价结果([−1,1]) | | |
|---|---|---|---|---|---|---|
| | P | A | D | P | A | D |
| 中性 | 0 | −0.31 | −0.06 | 0 | 0 | 0 |
| 放松 | −0.01 | −0.78 | 0.33 | 0.10 | −0.31 | 0.26 |
| 温顺 | 0.27 | −0.07 | −0.14 | 0.24 | −0.20 | −0.37 |
| 惊奇 | 0.23 | 0.64 | 0.01 | 0.43 | 0.43 | 0.05 |
| 喜悦 | 0.66 | 0.74 | 0.32 | 0.69 | 0.30 | 0.36 |
| 轻蔑 | −0.38 | −0.70 | 0.60 | −0.39 | 0.08 | 0.26 |
| 厌恶 | −0.44 | 0.22 | 0.36 | −0.45 | 0.10 | 0.17 |
| 恐惧 | −0.33 | 0.65 | −0.72 | −0.23 | 0.32 | −0.16 |
| 悲伤 | −0.28 | −0.36 | −0.78 | −0.22 | 0.04 | −0.17 |
| 焦虑 | −0.46 | 0.58 | −0.13 | −0.24 | 0.08 | −0.16 |
| 愤怒 | −0.86 | 0.66 | 0.91 | −0.49 | 0.28 | 0.28 |

拉塞尔在对 PAD 模型进行深入研究时发现，支配度主要与认知活动有关，愉悦度和唤醒度两个维度就可以表示绝大部分情感状态。因此，他采用环状结构模型来表示复杂的情感状态，认为情感分布在一个以中性情感为圆心（自然原点）的圆形结构上。每个维度的取值极限构成一个圆，圆的中心表示中性情感，并以此为中心向周围不同方向扩展，逐渐指向 8 种基本情感状态，在扩展的过程中呈现出不同强度的情感状态，情感点与自然原点之间的距离体现了情感强度。愉悦度和唤醒度是两个相互正交的维度。由于各种情感状态在自然原点的周围排成了一个圆形，所以这种对情感状态进行分类的方法叫作“情感轮”。对于任何一个情感样本，可以根据其情感强度和情感方向，在情感轮所组成的二维平面中用一维情感向量情感点和自然原点之间的情感向量 $\boldsymbol{E}$ 来表示。其中情感强度表现为这个情感向量的幅度值，而情感方向则表现为该情感向量的角度。

随后普拉奇克提出了以 8 种基本情感状态为原型的“情感轮”模型（图 3-12），8 种基本情感状态分别为喜悦、信任、害怕、惊讶、难过、厌恶、生气、期待。每种基本情感状态都有对应的颜色，并且根据颜色的深浅表示为 3 种强度，越邻近的情感状态越相似，距离越远则差异越大，互为对顶角的两个扇形中的情感状态则是相互对立的。圆形结构的中心为自然原点。在强度上延伸为三维椎体，强度越弱，情感的兴奋度越低，越消极；反之情感的兴奋度越高，越积极。普拉奇克的“情感轮”理论认为：①颜色的深浅代表情感状态的饱和度，如生气是愤怒的基本情感状态，暴怒是一种饱和状态，而烦躁则是一种不饱和状态；②相对的两种情感状态，在触发原因或者外在表现上通常是相反的，如喜与悲、爱与恨、怒与忧；③两种相邻的基本情感状态的结合，会产生一种复合的情感状态，而且这种情感状态的复杂性也体现了人类作为社会生物的高级状态。比如：一些外在刺激同时触发了个体的喜悦感和期待感，就会表现出乐观的倾向；而爱就是与喜悦和信任相关联的复合情感状态，这是符合人类认知心理的。

三维情感模型仍然不能表示人类所能体验的所有情感状态。为了更完整地描述情感，一些研究者将期望维作为第四个维度，强度作为第五个维度。期望维是对个体情感出现的突然性的度量，即个体缺乏预料和准备程度的度量；强度是指个体偏离冷静的程度。例如，在 PAD 模型的基础上引入期望维，能够将“惊讶”与其他情感类型区分开来，基本能够区分日常生活中的所有情感。

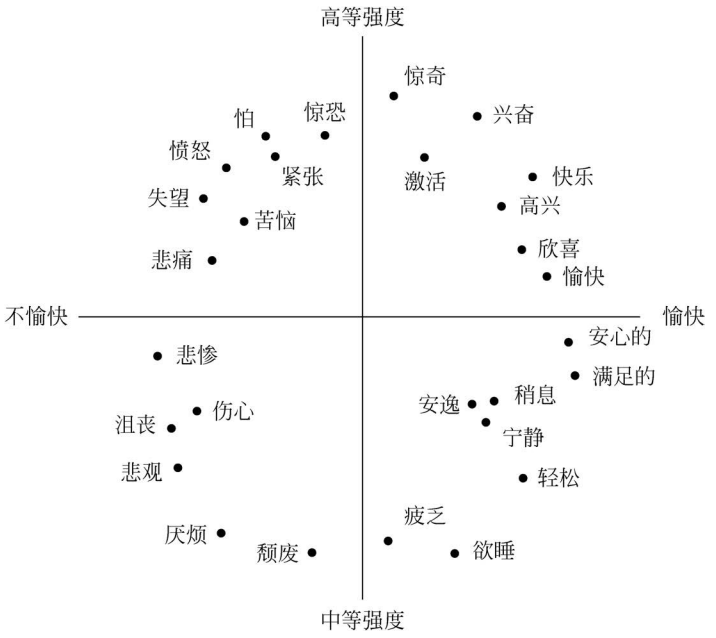



图 3-12 情感轮模型

近年来维度情感模型受到了越来越多研究者的关注，它的主要优势在于：①维度情感模型相比于离散情感模型具有更强的情感表达能力，尤其是在处理自然场景下的情感数据时有更加明显的优势，此时情感状态的范围非常广泛，难以用有限的几种情感类别来描述；②维度情感模型可以对不同个体的情感动态变化进行跟踪；③维度情感模型能够对情感的相似性与差异性进行精准度量；④心理学研究表明，人类的决策、推理、记忆、注意等认知都与维度情感模型存在着密切关系，例如在 PAD 模型中愉悦度决定着欲求动机系统和防御动机系统是否被情感刺激激活，唤醒度决定了每个动机系统被激活的程度。相比于离散情感模型，维度情感模型更加有利于促进机器充分地理解人的情感状态并做出合适的反馈。

从模型复杂度的角度分析，离散情感模型相对简洁易懂，而维度情感模型需要解决定性情感状态到定量空间坐标之间如何相互转换的问题；从情感描述能力的角度分析，离散情感模型的情感描述能力存在较大的局限性，仅能刻画出有限类别的情感状态，人们在日常生活中所描述的情感状态却是复杂多变的，维度情感模型能够从多个侧面连续地进行情感的描述，同时能够在很大程度上回避情感状态模糊性的问题。表 3-7 对两类情感模型之间的区别进行了总结。

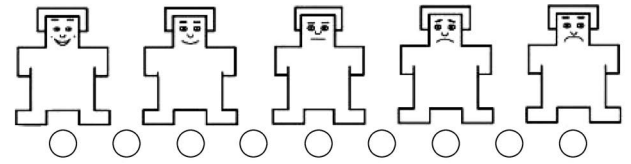
**表 3-7 两种情感模型的比较**

| 考 察 点 | 离散情感描述模型 | 连续情感描述模型 |
|---|---|---|
| 情感描述方式 | 形容词标签 | 笛卡儿空间中的坐标点 |
| 情感描述能力 | 有限的几个情感类别 | 任意情感类别 |
| 被应用到语音情感识别领域的时期 | 20 世纪 80 年代 | 21 世纪初 |
| 优点 | 简洁、易懂、容易着手 | 无限的情感描述能力 |
| 缺点 | 单一、有限的情感描述能力，无法满足对自发情感的描述 | 将主观情感量化为客观实数值的过程是一个繁重且难以保证质量的过程 |

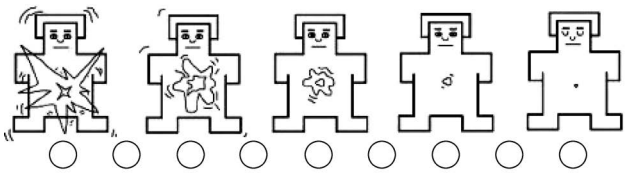
### 3.2.2　维度情感标注

维度情感标注不仅耗时耗力，而且是一个精细的过程，标注结果与标注者自身的偏好和经验都有着密切的关系。为了降低标注者自身因素对标注结果的影响，维度情感标注通常采用如下方法：①选择多个标注者共同完成标注任务；②选择与被标注对象具有相同母语的标注者；③在标注工作开始前对标注者进行培训，使其能够客观给出维度情感的标注，并且能够熟练使用维度情感标注工具；④对多个标注者的标注结果进行插值、标准化等一系列后处理，进一步减少标注偏差。

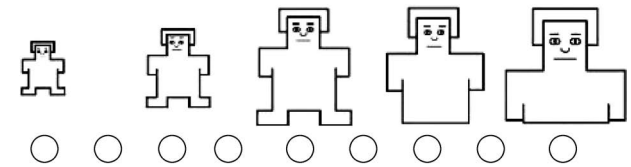
现有维度情感标注工作是基于情感量化理论实现的，目前并没有统一的方法。情感是一个不断变化的过程，为了对每个情感维度的取值进行实时跟踪，研究者开发了一系列标注工具。情感自我评估量表(Self-Assessment Manikin，SAM)系统是一种被大多数研究者所认可的维度情感量化方法，它基于 PAD 模型建立，使用卡通小人的形象来表示不同维度的情感取值。图 3-13(a)～(c)分别给出了效价度、唤醒度和支配度的取值分布，以卡通小人眉毛和嘴巴的变化来表示效价度的取值；以心脏位置出现的振动程度以及眼睛的注视程度来表示唤醒度的取值；以图片的大小来表示受控制的程度。在某个维度标注的过程中，标注者只需从对应的卡通小人中选出当前最符合的情感状态即可。使用的卡通小人数目由不同维度的量化数目决定，通常为 5 个或 9 个。每个小人对应的具体数值没有严格规定，当使用 9 个卡通小人时，对应的 9 个数字可以是 1～9 的整数，可以是－4～4 的整数，也可以是[－1，1]的 9 个等间隔的值。相比于其他情感量化方法，SAM 系统具有简单、快速、直观的优点；避免了不同标注者对同一样本的不同理解所造成的差异，从而所获得的标注结果方差较小、不同标注者间的一致性较高，因此 SAM 系统经常被用于维度情感的标注任务。在



(a) 效价度



(b) 唤醒度

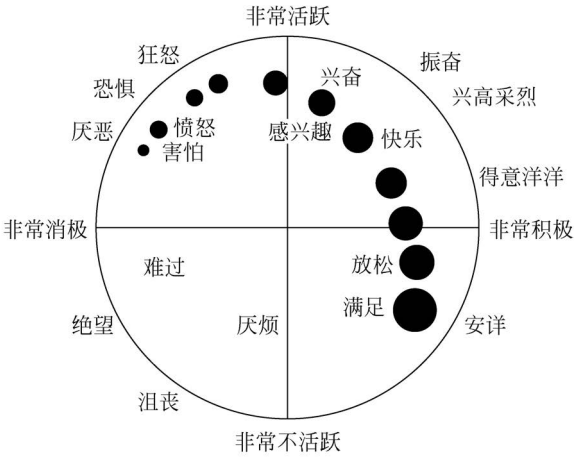


(c) 支配度

图 3-13　SAM 系统

每个卡通小人的下方标注数字并与卡通小人同时在屏幕上呈现，允许标注者点击两个数字之间的任意位置，便可以实现对目标维度的连续赋值。

Feeltrace 和 ANNEMO 是另外两个常用的标记工具。Feeltrace 是基于效价度-唤醒度模型建立的，如图 3-14(a)所示，将以效价度和唤醒度为主轴的圆在电脑屏幕上呈现，标注者只需根据自己所感知的情感用鼠标拖动圆形光标到合适的位置即可同时对效价度和唤醒度赋值。ANNEMO 是一种基于网页的维度情感标注工具，如图 3-14(b)所示，它将标注视频和标注光标同时在一个窗口显示，用户在观看视频的同时对视频中不同时间片段的情感维度进行连续标注。与 Feeltrace 相比，ANNEMO 使用更加方便，而且每次只对一个维度进行标记，所得到的结果更加精确。




(a) Feeltrace


(b) ANNEMO

图 3-14 Feeltrace 和 ANNEMO 标注示例

### 3.2.3 维度情感数据库

近年来研究者们在多个场景下构建了多模态情感数据，并在不同维度上进行情感标注，常用的多模态维度情感数据库包括 SEMAINE、RECOLA、IEMOCAP、CreativeIT、DEAP、VAM 等。表 3-8 总结了常用的维度情感数据库的数据采集场景、参与者数目、覆盖的模态、标注的情感维度、标注者人数、使用的标注工具和标注方法、标签的取值范围和取值类型。

表 3-8 常用维度情感数据库总结

| 数据库 | 场景 | 参与者数目 | 模态 | 情感维度 | 标注者人数 | 工具和方法 | 范围与类型 |
|---|---|---|---|---|---|---|---|
| SEMAINE | Solid SAL | 24 | Vi+Au | A、V、E、D、I | 2～8 人 | FEELtrace | [－1，1]的连续值 |
| RECOLA | 远程视频会议 | 46 | Vi+Au+Ph | A、V | 6 人 | ANNEMO | [－1，1]的连续值 |
| IEMOCAP | 双人对话表演 | 10 | Vi+Au | A、V、D | 至少 2 人 | SAM 系统 | 1～5 的整数值 |
| CreativeIT | 双人对话表演 | 16 | Vi+Au | A、V、D | 3～4 人 | FEELtrace | [－1，1]的连续值 |
| DEAP | 观看音乐视频 | 32 | Vi+Ph | A、V、D | 1 人 | SAM 系统 | [1，9]的连续值 |
| VAM | 电视脱口秀 | 47 | Vi+Au | A、V、D | 6～34 人 | SAM 系统 | [－1,1]的 5 点等间隔值 |

注：Vi——视觉模态，Au——听觉模态，Ph——生理信号，A——唤醒度，V——效价度，E——期望度，D——支配度，I——强度。

下面详细介绍 6 种较常用的维度情感数据库。

(1) SEMAINE 数据库是为了实现人机之间进行流畅富有情感的对话而建立的。数据库是在人机交互的场景下获取的，该场景模拟了人机对话的过程，由人扮演机器角色并与用户进行对话。机器角色根据用户的情感状态选择相应的词语与用户进行对话，并将用户向某个特定的情感状态引导。共有 24 个用户分别与 4 个不同性格的机器角色进行对话，每次对话都记录了用户和机器角色的正面视频，以及用户的侧面视频。标注人员按照视频的帧率逐帧标注用户在对话过程中的情感状态，从唤醒度、效价度、支配度、期望度和强度 5 个维度上取值。

(2) RECOLA 数据库中共采集了 46 个参与者的情感数据，这些参与者两人一组被分成 23 组，每组通过远程视频会议讨论某个灾难场景下的逃生方案，并达成一致意见。数据库中包含所有参与者在讨论过程中的面部视频和音频数据，以及其中 35 个参与者的 ECG 信号、皮肤电活动(Electrodermal Activity，EDA)信号数据。标注人员按照视频帧率逐帧给出参与者前 5min 讨论过程中的情感状态在效价度和唤醒度的值。

(3) IEMOCAP 数据库共采集了 10 名演员(5 男和 5 女)的情感数据，这些演员男女组合被分成 5 组，每组按照脚本或即兴进行对话表演。同一对话内容由相同的演员表演两次，每次使用运动捕获设备记录对话一方的面部表情、头部姿势和手部运动数据，同时记录对话双方的音视频数据。数据库中共有 174 段对话，每段对话都被分割成多个单句，每个单句所呈现的情感状态在效价度、唤醒度和支配度 3 个维度上用 1～5 的整数值进行标记。

(4) CreativeIT 数据库中共采集了 16 名演员的情感数据，这些演员两人一组被分成 8 组进行即兴表演，共进行 50 次表演。每次表演过程中记录了表演双方的音视频数据，以及使用动作捕获系统获取的演员全身动作数据。标注人员按照视频帧率逐帧给出了每个演员表演过程中的情感状态在效价度、唤醒度和支配度 3 个维度的取值。

(5) DEAP 数据库中共采集了 32 个参与者在观看音乐视频时的 EEG 信号、外周神经生理信号，以及其中 22 个参与者的正面视频。每个参与者都观看了 40 段音乐视频，并将自

己在观看音乐视频过程中所感受到的情感在唤醒度、效价度和支配度上给出 1～5 的整数值进行自我评估。

(6) VAM 数据库中的素材来自德国的电视脱口秀节目。数据分为三部分：VAM-video 数据集、VAM-audio 数据集、VAM-faces 数据集。VAM-video 数据集中的数据是从原始节目中分割出的 1421 条语句所对应的嘉宾视频。VAM-audio 数据集中的数据是从上述语句中选出的 1081 条语句所对应的语音信号。从 VAM-video 集中选取了大部分视频都是受试者正面人脸的视频，并从中提取出受试者的面部图像，构成了 VAM-faces 集，共包含 1867 张图片。由标注人员对每个样本所呈现的情感状态在唤醒度、效价度和支配度 3 个维度上用[－1，1]的 5 点等间隔值进行标注。

### 3.2.4 维度情感评价标准

早期的维度情感性能评价通常采用均方误差(Mean Square Error，MSE)。设 $\hat{\theta}$ 是估计的标签，$\theta$ 是真实的标签，$n$ 是样本数目，$\sigma_{\hat{\theta}}^2$ 和 $\sigma_\theta^2$ 分别是 $\hat{\theta}$ 和 $\theta$ 的方差，$\mu_{\hat{\theta}}$ 和 $\mu_\theta$ 分别是 $\hat{\theta}$ 和 $\theta$ 的期望，则 MSE 的定义为

$$\mathrm{MSE}=\frac{1}{n}\sum_{f=1}^{n}(\hat{\theta}(f)-\theta(f))^2 \tag{3-5}$$

MSE 描述了预测值与真实值的偏差，但 MSE 对于异常值敏感，且无法对 $\theta$ 和 $\hat{\theta}$ 的相对变化趋势进行描述，因此难以有效描述与真实值的吻合度。

鉴于 MSE 的缺点，Pearson 相关系数(Pearson Correlation Coefficient，PCC)被用于作为连续维度情感分析的评价指标，其定义为

$$\rho=\frac{\frac{1}{n}\sum_{f=1}^{n}[(\hat{\theta}(f)-\mu_{\hat{\theta}})(\theta(f)-\mu_\theta)]}{\sigma_{\hat{\theta}}\sigma_\theta}=\frac{E[(\hat{\theta}-\mu_{\hat{\theta}})(\theta-\mu_\theta)]}{\sigma_{\hat{\theta}}\sigma_\theta} \tag{3-6}$$

PCC 的取值范围是[－1，1]，它反映了预测值与真实值具有线性关系的紧密程度。PCC 能够有效反映预测值与真实值的协同变化关系。由于 PCC 对预测的幅值不敏感，无法对 $\theta$ 和 $\hat{\theta}$ 的偏差进行度量，因此仍不能有效描述预测值与真实值的吻合程度。为了更为有效地描述预测值与真实值的吻合程度，一致性相关系数(Concordance Correlation Coefficient，CCC)作为预测性能的评价指标，其定义为

$$\rho_c=\frac{2\rho\sigma_\theta\sigma_{\hat{\theta}}}{\sigma_\theta^2+\sigma_{\hat{\theta}}^2+(\mu_{\hat{\theta}}-\mu_\theta)^2} \tag{3-7}$$

CCC 结合了 PCC 与 MSE 的优点，既反映了预测值与真实值的协同变化关系，又反映了预测值与真实值的吻合程度，是目前广泛使用的连续维度情感分析性能评价指标，被多个国际情感识别竞赛所采用。

## 3.3 基于个性化的情感模型

个性与情感密不可分，个性主要影响情感状态的启动并控制情感强度。由于心理学中适合计算的个性化模型并不多见，直到最近几年，个性化情感模型的研究才得到发展。本节

重点介绍大五人格模型、Chittaro 行为模型、外向-担心-好斗（Extroversion-Fear-Aggression，EFA）性格空间模型、情绪-心情-性格模型等几种典型的基于个性化的情感模型。

### 3.3.1　大五人格模型

大五人格模型从开放性、认真性、外向性、宜人性和神经质 5 个维度描述人格特质。其中，开放性表示主体是否具有创造力、想象力，容易对事物产生兴趣；认真性表示主体是否具有责任心以及对事情关注的程度；外向性表示主体是否爱交谈、精力充沛；宜人性表示主体是否可信服、友好、具有合作精神；神经质表示主体是否缺乏安全感、情感易波动。这种通过不同维度值影响情感个性的建模方法能够有效赋予不同个体多样化的个性行为。目前在创建仿生代理时，常用的个性化建模方法是使用心理学家提出的基于维度的方法，其中每个维度对应个性的一个特质。大五人格模型如图 3-15 所示。


图 3-15　大五人格模型

### 3.3.2　Chittaro 行为模型

Chittaro 等构造了一个基于有限状态机的行为模型，它主要是通过个性选择来执行行为，体现出仿生代理不同的个性行为。在该模型中，每个状态表示主体的一个行为，个性表达采用大五人格模型。个性信息影响主体下一时刻以多大概率选择某个行为，而不是直接确定下一时刻要执行的行为。

### 3.3.3　EFA 性格空间模型

威尔逊在关于情感的论著中提出了一种 EFA 性格空间的构造方法。个性空间的三维分别是外向、担心和好斗。个性特征由它所处的个性空间的位置决定。例如，点坐标（$E-30, F+10, A-20$）表示和外向（$E$）相关程度有 30%的减弱，与担心（$F$）相关程度有 10%的加强，与好斗（$A$）相关程度有 20%的减弱。而原点（0，0，0）则代表一种中性平和的个性，或者代表个体没有个性特征。个性空间的三轴具有不同的含义：与 $E$ 轴的相关程度越大，则表明其个性越外向化，倾向于达到更大的积极情感；与 $F$ 轴的相关程度越大，则表示其更可能会倾向于达到更大的消极情感；与 $A$ 轴的相关程度则表示情感转移速度，$A$ 值越大则速度越快，$A$ 值越小则速度越慢。对于某个人而言，$F$ 在性格空间中的位置影响着个体的积极与消极情感的变化范围和变化率。一个时间步长内，情感变化的速度以及变化到何种程度，均是以性格为变量的函数。EFA 性格空间模型如图 3-16 所示。

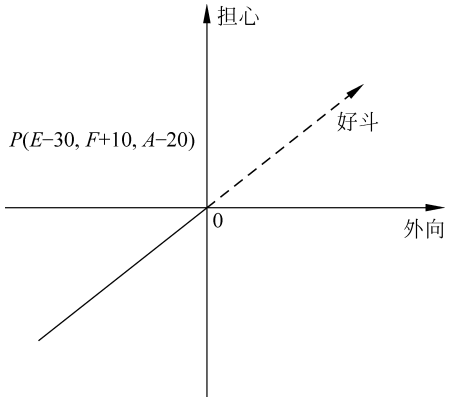



图 3-16　EFA 性格空间模型

### 3.3.4 情绪-心情-性格模型

多层情绪-心情-性格模型可以在人机对话系统中使用，该系统主要包括以下 4 个模块。

(1) 文本处理和响应生成模块：运用相关的自然语言理解理论提取用户输入的情绪信息。

(2) 性格模型：使用大五人格模型定义虚拟人的人格空间，针对每个因素构建贝叶斯信任网络，并通过调整 5 个因子的线性组合关系构造出具有不同性格的虚拟人，用来描述性格与心情之间的关系。使用贝叶斯推理规则，将当前的情绪状态(好、坏、一般)与前一个模块根据给定的人格类型生成的情绪信息进行组合，即可在下一刻获得情绪状态。

(3) 心情-情绪模块：情绪的转移主要取决于 3 个因素，即文本处理和响应产生模块产生的情绪信息，当前情绪状态和以前的情绪状态。为了连接这 3 个因素，该模块为每个情绪状态定义了一个情绪转移矩阵，采用 24 种情绪，并将这 24 种情绪简化为艾克曼所识别的 6 种基本情绪(高兴、悲伤、害怕、嫌恶、惊讶和生气)和中性情绪，由此得到的转移概率矩阵大小为 7×7，计算每种情绪的转移概率，并令其中最大者作为下一时刻的情绪状态。

(4) 同步模块：完成情绪状态与表情的映射。

情绪-心情-性格模型如图 3-17 所示。

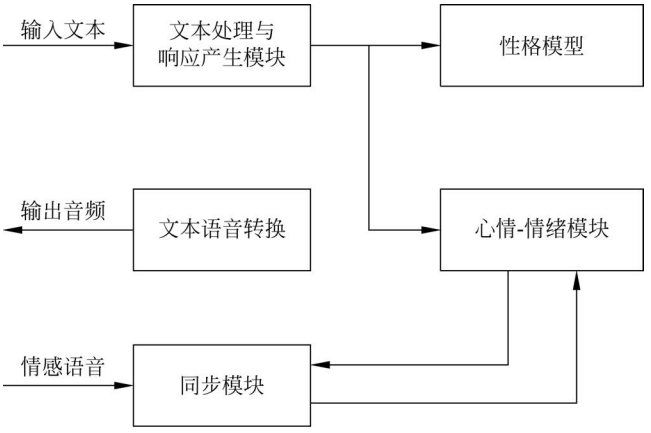



图 3-17 情绪-心情-性格模型

## 习题

1. 简述离散情感识别和维度情感识别的区别。
2. 列举 3 种典型的维度情感识别模型。
3. 什么是复合情感?
4. 离散情感和维度情感模型的评价标准有哪些?
5. 描述大五人格模型的特点。