

# 情感计算中的实验设计和情感度量方法研究

张迎辉 林学闇

(清华大学计算机科学与技术系, 北京 100084)

**摘要** 情感计算的目标是赋予计算机感知、理解和表达情感的能力,而情感数据的获取和情感真值的度量是情感计算的研究基础。为了更好地进行情感度量的研究,以 FIFA 足球电子游戏为场景,进行了精心设计的实验,即首先通过游戏场景诱发游戏者自然的情绪反应;接着提出了基于情绪脚本的情感编码方法,并结合范畴法和维度法对情感数据进行了自评和他评;最后对观察者间的一致性进行了统计分析。实验结果表明,该方法能够获得有效一致的情感度量结果。

**关键词** 情感计算 实验设计 情感度量 情感编码

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2009)02-0292-11

## Investigation on Experiment Design and Affect Measure in Affective Computing

ZHANG Ying-hui, LIN Xue-yin

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

**Abstract** The purpose of affective computing is to enable computers to perceive, understand and express emotions. Affect data acquisition and affect measure are essential to the research of affective computing. To explore the problem of affect coding, experiments were elaborately performed in the scenario of FIFA football electronic game. Firstly natural affect of game players were elicited in the game environment. Then a scheme of affect coding based on emotion prototypes was proposed, and both categorical emotions and levels within the affective space were assessed by game players as well as third-party observers. Finally an analysis was performed on the inter-observer agreement. The experiment data showed that valid and consistent affect coding results can be attained with the proposed method.

**Keywords** affective computing, experiment design, affect measure, affect coding

## 1 引言

“情感计算”一词最初由美国麻省理工学院(MIT)的 Picard 提出,她将其定义为“关于情感、由情感引发以及意图影响情感的计算”<sup>[1]</sup>。情感计算的目标是赋予计算机感知、理解和表达情感的能力。情感计算技术将使计算机更加主动、友好地为人类服务。情感计算的研究内容主要包括情感数据获

取、情感识别、情感反馈、情感表达等。其中情感数据获取是情感计算的研究基础。要获得真实自然的情感数据,就要精心设计实验情景,使被试者自发地产生情感并自然地流露出来。此外,由于情感的真值不是显而易见的,因此为获得情感的真值,有必要对情感数据进行心理学度量。本文结合 FIFA 足球电子游戏(以下简称 FIFA 游戏)的场景,讨论了情感计算中的实验设计和情感度量方法。

基金项目: 国家自然科学基金项目(60433030)

收稿日期: 2007-05-22; 改回日期: 2007-08-31

第一作者简介: 张迎辉(1975 ~ ),女。清华大学计算机科学与技术系博士研究生。主要研究方向为计算机视觉、模式识别、情感计算等。E-mail: allyzhang@gmail.com

## 2 研究现状

### 2.1 情感数据获取

“情感数据”包括反映人的情绪变化的表情、语音、生理指标等多种信息。在情感计算的研究中,对情感数据获取的要求主要体现在以下4个方面:

(1)数据的典型性。数据的典型性要求采集最能体现某种情感特征的情感数据,例如面部表情、语音韵律以及生理特征等都具有特定情感的显著特征,以明显地区别于别种情感。高典型性的情感数据不仅可以为情感建模提供良好的基础,而且有助于提高情感识别的准确性和鲁棒性。

(2)数据的自然度。数据的自然度是指被试者的情绪表现与其在真实生活中的情绪表现相似的程度。其自然度越高,情感数据越接近真实的情绪表现,情感计算技术也越容易从实验室过渡到实际应用中。

(3)数据的完整性。人类情感千变万化,不仅具有面部表情、肢体动作、语音韵律、生理反应等多种表现,而且情绪本身也具有产生、发展、高潮、消退的动态变化过程。情感数据应尽可能完整地保留情绪的各种表现和动态特征。此外,情感的类别也应丰富多样。除快乐、悲伤、惊讶、恐惧、愤怒、厌恶等6种基本的人类情绪外,还应包含特定情境下的各种典型情绪。例如,学习情境下的兴趣、困惑、厌倦、挫折等情绪对理解学习者的心理状态是非常有价值的;又如游戏情境下的投入、紧张、失望、得意、轻松等情绪经常出现;而困倦情绪则对检测疲劳驾驶具有特别重要的价值。因此研究特定情境下的若干种典型情绪逐渐成为情感计算研究的主流。

(4)实验环境的可控性。实验的优越性在于可以通过控制实验条件来观察变量的变化,并可排除无关因素的干扰。在对情感进行的实验研究中,既要严格控制实验环境,以诱发特定的情绪反应,又要尽量减少实验环境对被试者情绪的干扰,为情感的自然流露创造轻松的氛围。

获取情感数据的方法目前主要有以下4种:演员表演的情感,以刺激物诱发的情感、在实验场景中被试者自然流露的情感,以及记录的真实生活中人们的情感。

演员表演情感的方法,最大的优点是数据典型性高、实验环境容易控制。最初,情感计算的研究重

点是快乐、悲伤、惊讶、恐惧、厌恶、愤怒等6种基本情绪。这些由遗传获得的人类情感,全世界的人群都是一样的,与种族、性别、年龄等基本无关。正如达尔文在《人与动物的情感》中提到的:“天生的盲人也能像有良好视力的人一样,表现出自己的情感动作”<sup>[2]</sup>。演员能够以典型的、夸张的表演手法突出情感的这种共性,对后续的情感识别等研究非常有利。因此,早期情感数据获取所用的方法一般是表演情感。但由于演员的内心未必体验到这种情绪,因此数据的自然度较低。此外,用这种方法获得的情感数据往往是单一的情绪表现。由于情感的产生缺少前因后果,且采集的情感数据往往是静态表情图像,缺少动态的情感变化,因此情感数据不够完整。

随着研究的深入,情感计算研究者借鉴了心理学的方法,以刺激物来诱发情感。实验中,被试者受到素材的刺激,产生某种情绪反应,例如,给被试者观看外科手术的录像,以诱发厌恶的情绪。这种方法获取的情感数据自然度有所提高,特别有利于研究被试者的生理反应与情绪变化的关系。但是,刺激物能够诱发的情绪种类有限,许多日常生活中常见的情绪,如失望、沮丧、得意、困惑等都不易通过刺激物诱发获得。此外,在日常生活中,人们的情感往往既不典型,也不单纯,而是微妙和富于变化的,无论其外在表现还是内在感受都有别于刺激物诱发的情感。情感计算要想真正走向应用,则必须研究这些真实细微的情感,这就为情感数据获取提出了新的挑战。

获取情感最自然的方式是在日常生活中捕捉这些细微的情感变化,不过目前还有不少技术上的困难。其中一个主要的问题就是采集情感数据的设备会对被试者造成干扰,这个问题的解决需依赖于可穿戴技术的发展;另一个问题是,由于人在一天中的大部分时间里处于平静的情感状态中,致使日常生活中直接采集的情感数据缺乏典型性。同时,由于生活场景复杂多变,要记录日常生活的情境信息,并对其进行建模也是一个颇为棘手的问题。

目前,在精心设计的实验情境中,记录被试者自然流露的情感变化,是比较理想的情感数据获取方法,例如观察孩子在游戏时的表现;让鼠标随机地发生故障,并观察用户遇到挫折时的困惑表情等。这种方法在控制实验条件,以获得典型的情感数据,与减少对被试者的干扰,以提高情感数据自然度之间

取得了较好的平衡,其既能获得典型的、完整的情感数据,又尽可能保留了情感的“原汁原味”。因此,在采集情感数据时,本文选择了这种方法。

4 种情感数据获取方法的对比如表 1 所示。

表 1 情感数据获取方法的对比

Tab. 1 Comparison of affect data acquisition

情感数据获取方法	数据典型性	数据自然度	数据完整性	实验环境可控性
演员表演情感	高	低	低	高
刺激物诱发情感	高	高	低	高
实验场景中被试者自然流露情感	高	高	高	中
真实生活中记录情感	低	高	低	低

## 2.2 情感度量方法

### 2.2.1 情感的量化

对事物的描述,往往既可以分成不同的类别,又可以用多种属性的组合描述其性质。例如,颜色既可以分成红、橙、黄、绿等类别,也可以用 RGB、HIS 等 3 维度的组合来描述。与之类似,在心理学中,对情感的描述也有范畴和维度两种体系。

从范畴法的观点看,情感可以分为若干性质不同的种类。范畴法通常用一个或一组词语来描述某一类情感的性质。前文中提到的 6 种基本情绪就是一种范畴的描述方法。在对婚姻关系的研究中,研究者往往采用范畴法,即选择若干种与婚姻关系最相关的情绪进行研究。SPAFF 编码系统<sup>[3]</sup>包含了厌恶、愤怒、抱怨、悲伤、紧张、喜爱、兴趣、幽默等婚姻关系中常见的情绪。

从维度法的观点看,被表达的情感可看作是愉快、激活等成分的混合物。冯特在 1896 年提出情绪的 3 维学说<sup>[4]</sup>。冯特认为情感过程是由 3 对情感元素构成的。每一对情感元素都具有处于两极之间的程度变化。它们是愉快-不愉快、兴奋-沉静、紧张-松弛这 3 个维度。每一种具体情绪发生时,都按照这 3 个维度分别处于它们两极的不同位置上。冯特的情感 3 维理论虽然建立在主观推测的基础上,但它至今仍有理论和实际的意义。后人在冯特学说的基础上,提出了许多种情绪维度量表。情绪心理学家一直试图寻找情感的基本维度。在面部或言语表情、情绪词的语义分析,以及情绪体验的自评研究中,有两个主导维度被重复发现<sup>[5]</sup>。这两个维度通常被称为效价(valence)及唤醒(arousal)。其中第 1

个维度是效价,指涉及情感感受的正性或负性程度差异,其代表情绪状态的总体性质。第 2 个维度是唤醒,它反映了情感体验的力量或强度。

事实上,范畴和维度这两种描述方法不是互相对立的,而是互相补充的,其中范畴法强调事物的性质或其占主导地位的成分,而维度法则要定量地描述事物,它们分别从不同的角度描述情绪。某一类情感往往对应多维空间中的特定区域,而情感多维空间中的点则可以用某一类情感来描述。因此,若想反映情感的全貌,则应当尽可能使用两种描述方式,并研究二者之间的对应关系。

### 2.2.2 情感编码方法

在情感计算的研究中,不可避免地面临如何度量情感真值的问题。这正是情感编码要解决的问题,即确定某一时刻被试者的情感类别(范畴描述体系)或各情感维度的成分(维度描述体系)。与演员表演的、单一的、典型的情绪不同,被试者在实验中自然流露的情绪,由于其反映的是内在的情感体验,因此其真值并不是显而易见的。同时由于对情感的认知存在较大的个体差异,因此情感的度量很难得到一致的结果。这也正是情感编码的困难所在。

实验中采集的情感数据往往是一段时间内的数据,而被试者的情感则是随时间发生变化的。情感编码面临的另一个问题就是如何在编码中反映情感的动态变化。情感编码采用的第 1 种方法是把情感看作随时间变化的量,观察者在时间轴上实时标出情感的变化。FeelTrace<sup>[6]</sup>就是一个实时情感编码软件,这种方法的优点是操作直观方便,缺点是难以捕捉在极短时间内发生的情感变化,本文 3.2.1 节中对此有详细的分析;第 2 种方法是把原始数据均分为等时长的片段,观察者对每个片段进行编码<sup>[7]</sup>,这种方法的优点是便于进行时间序列分析,缺点在于等时长的片段划分难以在持续时间较长的情绪(如感兴趣的)和短时间内突然变化的情绪(如愤怒)之间取得平衡;第 3 种方法则是直接对连续的视频片段进行编码,并由观察者指定情感变化的时间点。例如, Jones 等人设计的实验中<sup>[3]</sup>,由观察者指定情绪变化的起始点。由 Aune 设计的实验中,被试者两人一组对某个问题进行讨论,随后两人被分开独立观看其讨论问题时的录像,当回忆起某种情感体验时,被试者暂停录像播放,并编码情感<sup>[8]</sup>。这种方法的优点是能够对连续的时间片段进行逐帧的细致分析,不会错过情绪瞬间变化的时刻,同时又

能够在情绪持续时间较长时避免大量重复编码,而缺点则是不同的观察者独立编码时,由于情感变化的时间点可能不同,从而降低了一致性。

按照情感编码的主体区分,情感编码方法可以分为:自评和他评。自评即请被试者评价自己的情感体验;他评则是请其他人评价被试者的情感表现。一个复杂的问题是,人们的主观情感体验与他人的客观感受有时并不一致。主观体验是人们内心感受到的情感,客观感受则是观察者接收到的情感信息。从某种意义来讲,二者都是真实的。因此,在对情感进行心理学度量时,自评和他评都是必要的。

情感主要由主观体验、外显表情以及生理反应等成分组成,这些也提供了识别情感的线索。此外,情绪心理学的研究表明,情感具有强烈的情境依赖性,且情感同人的切身需要和主观态度联系着。当情境发生变化时,个体通过对情境的认知把这种变化与自己的愿望和需求联系起来,由此产生了对事物的评价,而这种评价可以发动、转移或者改变情感反应和体验,反过来情感反应和体验又可以通过发动、干涉、组织以及破坏认知过程和意志行为来发挥个体的适应性和主观能动性,以适应或改变环境<sup>[4]</sup>。因此情境信息也是识别情绪的重要线索。在情感编码过程中,往往可综合利用多种线索。例如,Noller提出了利用注视、凝视、开口笑、抿嘴笑、皱眉、撇嘴、低头、抬头、点头、摇头、身体前倾、手势等16种非言语动作来区分夫妻在交谈时的正性、负性和中性情感<sup>[9]</sup>。Jones等人综合了面部肌肉运动(面部表情)、语音的韵律特征(语速、重音等副语言特征)和谈话内容(情境信息)用来识别夫妻在发生冲突时表现出的各种情感<sup>[3]</sup>。Roberts等人在提供给观察者的指南中,除了提供每种情感的字典解释及同义词外,还列出了可能引起该情感反应的话语或行为、该情感的典型面部表情、该情感对应的语音韵律特征以及帮助识别该情感的其他非语言现象<sup>[10]</sup>。但是,各线索的简单组合缺乏内在联系,不便于理解和记忆。此外,由于观察者的注意力有限,因此同时观察几种线索时,容易顾此失彼。

本文提出了基于情绪脚本的编码方法。这种方法采纳了情绪概念的“原型”观点<sup>[5]</sup>。其核心是一个范畴里的成员是与其相似程度而判定的,即成员与原型越接近,越确凿地属于该范畴,但成员与非成员之间没有截然划分的界限。“脚本”可以看作是情绪概念的原型。情绪被看作是一系列事件的组

合,用于构成描述情绪的以下特征:原因、信念、生理变化、欲望、外显行为和语言、面部表情。这些事件形成一个因果序列,其情形即如剧本中的情节被有机地组织在一起。下面是Russell提供的一个怒的脚本:

- (1) 此人被冒犯了。冒犯是有意的、有害的。  
——原因
- (2) 此人是无辜的,这对他不公平。 ——信念
- (3) 此人瞪眼注视着冒犯者。 ——表情
- (4) 此人感到内心紧张、不安,仿佛热力和压力在内心快速升高,心脏猛跳,肌肉拉紧。  
——生理反应
- (5) 此人想报复、惩罚对方。此人失去控制,出手伤害对方。  
——行为

从情绪脚本的观点来看,被试者在一段时间内的情绪表现可以划分为一系列相对独立的情绪脚本。由于被试者在同一个情绪脚本中的情绪是一致的,因此每个情绪脚本编码一次即可。在情绪脚本的框架下,情境、面部表情、语音、行为等线索被有机地组织在了一起,这有助于观察者理解和识别各种情绪。基于情绪脚本的编码方法是先将连续的情感数据切分成独立的情绪脚本,再对各情绪脚本进行编码。这种方法既可以对情绪变化进行细致入微的分析,又便于理解和操作,不失为描述情感动态变化的理想工具。在第3节中将结合具体实验介绍基于情绪脚本的编码方案。

## 3 实验

### 3.1 情感数据获取

由于本项目的研究目标是识别自然的、细微的情感,因此采用了在实验中诱发情感的方法。

在实验室环境中,人的情绪很容易受到干扰。因为当被试者意识到自己的行为被观察时会感到不自在,甚至无法自然地表达情感。“当正在观察中的行为发生某种程度的改变时,其测量结果就会出现反应性”<sup>[11]</sup>。为了预防反应性问题,本实验选择了对抗激烈的FIFA游戏。因为FIFA游戏的游戏者要迅速做出防守、进攻、射门等决策,并熟练运用断球、加速奔跑、过人、传球、射门等技术动作,其注意力很快就被吸引,而无暇顾及及其他。为了增强游戏的对抗性,在实验中还采用了双循环的积分赛,即每两人之间赛4场,最后根据总积分排定名次,并给

予优胜者物质奖励。选择 FIFA 游戏的另一个原因是游戏者能够自然地表现出多种情绪,例如快乐、轻松、投入、紧张、失望、惊讶等。

实验选择了 3 名男性作为被试者,年龄在 26 ~ 29 岁之间。他们都熟悉并喜爱 FIFA 游戏,且水平

比较接近。实验环境的布置如图 1 所示。实验中共采集了 3 路视频和 3 路音频,分别是两名游戏者各自的正面脸部视频、游戏画面视频、两名游戏者各自的语音和比赛现场的声音。由于实验时生理多导仪尚未准备好,故本实验未采集生理数据。

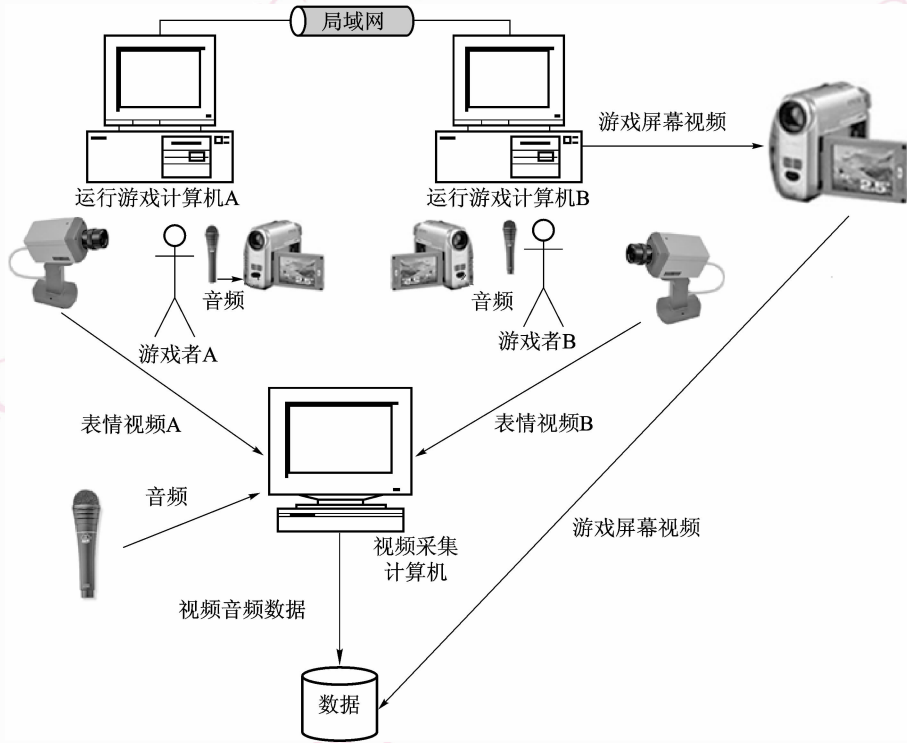


图 1 FIFA 实验环境示意图

Fig. 1 Environment configuration in FIFA experiment

### 3.2 情绪标注方案

#### 3.2.1 基于情绪脚本的切分

通过对情感数据的分析发现,被试者的情绪随着情境发生变化:当局势有利时,被试者表现出兴奋、投入等情感;当局势不利时,被试者表现出紧张、失望等情感。可见比赛情境成为诱发情感变化的重要原因。同时,被试者的情感自然地表现在面部表情、生理反应、语音、行为等方面。实验结束后,被试者观看自己的面部表情视频和比赛画面时,能够轻易地回忆起自己当时的情感体验。借助被试者的回忆,就可以重构出大部分情绪脚本的面貌。下面是一个典型的例子:

(1) 游戏者进球了。——原因

(2) 游戏者希望取得比赛的胜利,现在离胜利更近了。——信念

(3) 游戏者眉开眼笑,甚至手舞足蹈。——表情

(4) 游戏者感到血液循环加速,充满能量。——生理反应

(5) 游戏者津津有味地欣赏进球的回放,回味自己的胜利。——行为

由于整个比赛的情感数据就是一系列类似的脚本串在一起,而且每个脚本基本都具备了原因、信念、表情、生理反应、行为等要素,因此,本文提出了基于情绪脚本的编码方法,即先把连续情感数据切分为相对独立的情绪脚本,再对情绪脚本进行编码。切分的准则是:

(1) 每个片段内,游戏者的情感应当是一致的;

(2) 每个片段应当具有情绪脚本的基本要素。

值得一提的是,人的情绪在极短的时间内也会发生丰富的变化,例如,在对方射门之前,游戏者的情绪是紧张的;对方射门时,游戏者会有沮丧的情绪,但随之有些放松(因为现在局势已经无法控制了),接着又有些期待(怀着侥幸心理——也许对方踢不

进呢),然后是非常沮丧(对方真的踢进了)。在这短暂的几秒内,游戏者先后经历了紧张、沮丧、放松、期待等丰富而细微的情绪变化。由于每种情绪都有独立的原因、信念、表情、生理反应、行为等要素,因此都构成了独立的情绪脚本。这是本文没有选择 FeelTrace 软件编码情绪的主要原因。因为在情感瞬间变化时,观察者很可能来不及反应。而基于情绪脚本的编码方法则可以对情感变化进行细致的分析。

由于情感类别的不同,并非每个脚本都具备所有的要素。对人脑情绪中枢的研究发现,情绪可以不经意识层面而直接表达出来。在暴怒时做出不理智的举动即是明显的例子<sup>[12]</sup>。一般情况下,一个情绪脚本至少要具备原因、表情、生理反应3个要素。

由于本实验中未能采集生理指标,因此无法利用生理反应的线索。根据上述切分准则,本文设计了以表情、比赛情境、语音为线索的切分方案:

(1)以游戏者的表情变化作为新的情绪脚本的起始点。表情的变化是情感变化的一个最明显的外在标志。

(2)比赛情境发生转变时,需仔细观察游戏者的表情,若游戏者的表情发生了变化,则认为出现了一个新的情绪脚本。本实验中,情境变化是情绪变化最主要的推动力。由于游戏者的目标是取得胜利,因此比赛局势朝着有利或不利的方向转变时,游戏者的情绪都会随之朝着正性或负性的方向转变。对实验数据的分析表明,游戏者的情绪与比赛的情境具有很高的相关性。由于普通人对细微的面部表情变化不够敏锐,因此情境的变化可以作为分辨情绪变化的启发式线索。按照局势是否有利于游戏者,比赛情境被分为以下12种状态:中场争夺、前场进攻、射门进球、射门未进、后场防守、对方射门进球、对方射门未进、己方犯规、对方犯规、己方任意球、对方任意球、比赛暂停。整场比赛的情境可以用这12种状态组成的序列来表示。

(3)具有明确情感信息的语音是识别情绪的重要线索。尽管语音在游戏过程中出现的机会并不多,但由于笑声、感叹词等携带了非常明确的情感信息,因此也是不容忽视的一条线索。

为综合利用表情、情境、语音等线索,切分前还对数据做了预处理,即先利用视频编辑软件把游戏者的面部表情、语音和比赛画面合成为一个视频文件(如图2所示)。其中游戏者的面部表情居中,比赛画面出现在右上角。右下角标注内容依次为:比

赛场次(Game1)、双方所使用球队的缩写(Eng vs Fra)、游戏者编号(Player1)、游戏者所使用球队的名称(England)以及该队队员球衣的颜色(White)。此外,情境状态的标注也在情感切分前完成。



图2 用于情感编码的合成视频

Fig. 2 Video composed for affect coding

切分时,观察者通过观察游戏者表情、比赛情境和语音等线索来进行切分。当游戏者的表情或比赛情境发生变化时,则对照情绪脚本的准则来判断是否进入了一个新的情绪脚本。若发现了新的情绪脚本,则记录情绪脚本的起始和结束时间,并对表情、语音等要素进行标注。最理想的切分人选是游戏者本人,因为他最清楚自己情绪的变化。为了节省游戏者的时间,情感数据先由其他人进行预切,最后由游戏者本人对预切的结果进行调整,并保存为最终的切分结果。

### 3.2.2 基于情绪脚本的编码

情绪编码是在切分好的情绪脚本上完成的。为了尽可能全面地反映情绪,需对每个情绪脚本,先后进行自评和他评。在自评和他评中,同时采用了范畴和维度两种描述体系。

情绪编码的一个准备工作是情绪的量化,其在范畴法中就是要确定情绪的类别,维度法中就是要确定各维度的等级。科学的量化标准很重要,因为它可以提高编码数据的可靠性和准确性。

范畴法以情绪词描述情绪的类别。编码时使用的是情绪词组检表法,即观察者从给定的若干组情绪词中选择一组最能描述游戏者情绪的词。在设计情绪词组检表时,需遵循以下两个原则:一是全面,即情绪词组检表中的各组情绪词能够描述游戏者可能出现的所有情绪;二是能提高各组情绪词之间的

区分度,即将相同或相近的情绪词尽量归为一组。本实验选词参考了北大心理系左衍涛设计的情绪词自我评定表格<sup>[5]</sup>。左衍涛在权威词典条目和由大学生被试者提供的 697 个情绪词的基础上,经过筛选和归并,得到 147 组共 400 个情绪词。笔者在此基础上做了进一步的筛选和归并,只保留了与游戏者情绪相关的情绪词组,同时合并了一些相近的情绪词组。最终得到的情绪词组检表共有 23 组情绪词,各组的代表词依次为:得意的、喜悦的、兴奋的、自信的、满意的、轻松的、投入的、谨慎的、冷静的、平静的、无奈的、庆幸的、紧张的、埋怨的、失望的、无精打采的、吃惊的、犹豫的、轻蔑的、疲劳的、沮丧的、迟钝的、惊慌的。这 23 组情绪词被用于情绪的自评。范畴法的部分自评结果如图 3~图 7 所示(图中 Game1 代表第 1 局游戏,Eng vs Fra 代表英格兰对法国,Player1 England White 代表英格兰队队员,着白色球衣):



图 5 紧张的表情

Fig. 5 Expression of tension



图 6 失望的表情

Fig. 6 Expression of disappointment



图 3 平静的表情

Fig. 3 Expression of serenity



图 4 投入的表情

Fig. 4 Expression of concentration



图 7 兴奋的表情

Fig. 7 Expression of exciting

维度法一般以维度等级的组合来描述情绪,而编码时则使用维度等级量表:对每个维度,观察者从给定的若干等级中选择一个最能描述被试者情绪的等级。而且确定维度等级时需遵循以下两个原则:一是各个维度的等级意义明确,观察者容易达成一致意见;二是不同等级具有较高的区分度,观察者容易在不同等级间做出选择。本实验采用了效价-唤醒的2维情感空间,一方面是由于普通人对效价-唤醒的理解中,负唤醒的情绪(例如“静默的”)容易与负效价的情绪(例如“忧郁的”)混淆;另一方面是由于等级过多不易区分,因此,在维度等级中,唤醒不取负值,分为0、1、2等3个等级,分别对应无生理唤醒、一般生理唤醒水平、高生理唤醒水平;效价可取正、负、零,分为-2、-1、0、1、2等5个等级,分别对应高负性情绪、一般负性情绪、中性情绪、一般正性情绪、高正性情绪。该维度等级量表用于自评和他评。

通过对自评数据进行分析,发现有些情绪词组在编码结果中极少出现,这些情绪词组被从情绪词组表中删除了。还有一些情绪词组之间很容易发生混淆。为了进一步提高各组情绪词的区分度,还利用自评数据来对保留的各组情绪词进行了聚类。聚类的算法步骤如下:

(1) 设定阈值  $T$ ;

(2) 根据自评结果,计算每一组情绪词在(效价、唤醒)等级组合上的概率分布;

(3) 以步骤(2)中得到的概率分布作为特征向量,计算每两组情绪词特征向量的相关系数;

(4) 取各相关系数的最大值与  $T$  进行比较,若大于  $T$ ,则将相关系数最大的两组情绪词合并为一类,转步骤(1)继续计算,否则结束退出。

反复调整  $T$ ,直至得到比较满意的聚类结果。最终,原有的23组情绪词组被删减合并为6组。这6组情绪词被用于他评实验。为帮助观察者更好地理解情绪词组检表,还编制了一张情绪词组术语表(附录1)。表中不仅列出了每组中的情绪词,还注明了各情绪词的语义和近义词。

自评由游戏者本人完成,他评则选择完全不认识游戏者的大学生作为观察者。他评实验开始前,每名观察者都得到一份情绪词组术语表,主持人逐项解释情绪词组术语表。在确认观察者对情绪词组术语表无疑问后,即开始他评实验。他评实验分两组进行:第1组有16名观察者参加,共编码了113

个片段的情感数据;第2组有15名观察者参加,共编码了173个片段的情感数据。

## 4 实验数据

### 4.1 观察者间一致性的度量方法

统计观察者之间的他评一致性,最常用的度量指标是Kappa系数及其扩展。Kappa系数最初只能度量两名观察者之间的一致性。Kappa系数的多种扩展则可以度量多名观察者间评价的一致性,Fleiss'Kappa系数是其中的一种<sup>[13]</sup>。本实验即采用Fleiss'Kappa系数来度量观察者间评价的一致性。

假定有  $N$  个对象,由  $n$  名观察者对其进行分类,每个对象被分为  $k$  类中的一类,  $n_{i,j}$  为把第  $i$  个对象分为第  $j$  类的观察者人数,则将  $P_i$  定义如下:

$$P_i = \frac{1}{n(n-1)} \sum_{j=1}^k n_{i,j}(n_{i,j}-1) \quad (1)$$

$P_i$  值反映了观察者对第  $i$  个对象进行评价的一致性。 $P_i$  的平均值  $\bar{P}$  的定义如下:

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i \quad (2)$$

$\bar{P}$  反映了观察者对所有对象评价的一致性。

此外,  $R_j$  定义为第  $j$  类的对象所占的比例,即

$$R_j = \frac{1}{Nn} \sum_{i=1}^N n_{i,j} \quad (3)$$

进一步,将  $\hat{R}$  定义如下:

$$\hat{R} = \sum_{j=1}^k R_j^2 \quad (4)$$

$\hat{R}$  反映了对象在  $k$  类上分布的均匀程度。

最终,将 Fleiss'Kappa 系数定义如下:

$$\kappa = \frac{\bar{P} - \hat{R}}{1 - \hat{R}} \quad (5)$$

由

$$\sum_{j=1}^k n_{i,j} = n \quad (6)$$

可得

$$\begin{aligned} P_i &= \frac{1}{n(n-1)} \left( \sum_{j=1}^k n_{i,j}^2 - \sum_{j=1}^k n_{i,j} \right) \\ &= \frac{1}{n(n-1)} \left( \sum_{j=1}^k n_{i,j}^2 - n \right) \end{aligned} \quad (7)$$

分析  $P_i$  的取值范围,可以帮助人们理解  $P_i$  的意义。分析如下:

(1)用 Lagrange 乘数法求  $P_i$  的最小值。令  $f = \frac{1}{n(n-1)} \left( \sum_{j=1}^k n_{i,j}^2 - n \right)$ ,  $g = \sum_{j=1}^k n_{i,j} - n$ 。令  $f + \lambda g$  对  $n_{i,j}(j=1, \dots, k)$  的偏导数为零可得

$$\begin{aligned} \frac{2n_{i,1}}{n(n-1)} + \lambda &= 0 \\ &\dots \\ \frac{2n_{i,k}}{n(n-1)} + \lambda &= 0 \end{aligned} \quad (8)$$

式(8)与式(6)联合即可得到以下一组解:

$n_{i,j} = \frac{n}{k}(j=1, \dots, k)$ ,  $\lambda = \frac{2}{k(1-n)}$ , 此时  $P_i$  取得最小值,即

$$\min P_i = \frac{n-k}{k(n-1)} \quad (9)$$

由此可以看出,当  $n$  位观察者的分类结果平均分布时,则  $P_i$  取得最小值。此时,观察者之间的一致性最差。

(2)由  $\sum_{j=1}^k n_{i,j}^2 \leq \sum_{j=1}^k (n_{i,j}n) = n \sum_{j=1}^k n_{i,j} = n^2$  可得

$$\max P_i = \frac{1}{n(n-1)}(n^2 - n) = 1 \quad (10)$$

此时,  $n$  位观察者的评价完全相同,一致性最高。

推广到所有对象,可得

$$\frac{n-k}{k(n-1)} \leq \bar{P} \leq 1 \quad (11)$$

类似的方法可得  $\hat{R}$  的如下取值范围:

$$\frac{1}{k} \leq \hat{R} \leq 1 \quad (12)$$

当  $N$  个对象在  $k$  类上均匀分布时,  $\hat{R}$  取得最小值;当所有对象同属一类时,  $\hat{R}$  取得最大值。

同时式(5)可写为

$$\kappa = \frac{\bar{P} - \hat{R}}{1 - \hat{R}} = 1 - \frac{1 - \bar{P}}{1 - \hat{R}} \quad (13)$$

由于  $\frac{1 - \bar{P}}{1 - \hat{R}} \geq 0$ , 所以  $\kappa \leq 1$ 。当且仅当  $\bar{P}$  取 1 时,

$\kappa$  取得最大值 1。此时,所有观察者对所有对象的评价都一致。当  $\hat{R}$  取 1 时,  $\kappa$  为负无穷大。此时所有对象同属一类,问题严重退化。

综上所述,  $\kappa$  可以理解为观察者评价的一致性与观察者独自评价的对比。若观察者之间评价越一致,样本分布越均衡,则  $\kappa$  越大,度量的构造也越好,因此,  $\kappa$  适合作为观察者间评价一致性的度量。

#### 4.2 本实验中观察者间评价一致性的度量

本实验中的他评工作分两组进行。

第 1 组有 16 名观察者 ( $n=16$ ), 对 113 个情感编码片段进行情感度量 ( $N=113$ ), 每个编码片段的情感被分为 6 组情感中之一 ( $k=6$ )。然后按照 4.1 节中的方法统计观察者对各片段进行评价的一致性  $P_i$ , 以及反映整体一致性的指标  $\kappa_1$ 。

其中,  $\kappa_1 = 0.203$ ,  $P_i$  值的分布如图 8 所示。其中  $[0, 1]$  区间被均分为 5 段, 横坐标分别代表  $P_i$  的 5 段取值范围; 纵坐标为  $P_i$  落入相应取值范围内的片段数占总片段数的百分比。

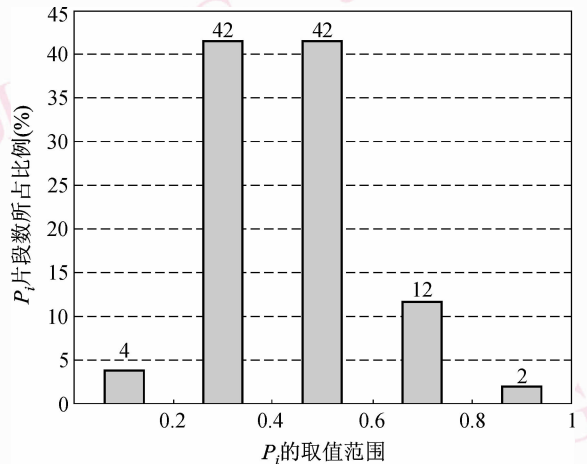


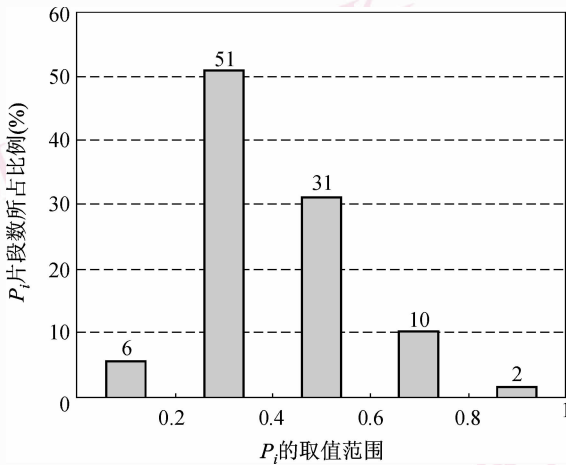
图 8 第 1 组  $P_i$  值的分布

Fig. 8 Distribution of  $P_i$  in group 1

第 2 组有 15 名观察者 ( $n=15$ ), 对 173 个情感编码片段进行情感度量 ( $N=173$ ), 每个编码片段的情感被分为 6 组情感中之一 ( $k=6$ )。  $P_i$  和  $\kappa_2$  的计算方法同第 1 组。

其中,  $\kappa_2 = 0.219$ ,  $P_i$  值的分布如图 9 所示。各坐标轴的含义同图 8。

从统计结果来看, 绝大部分  $P_i$  值都分布在  $[0.2, 0.6]$  区间内, 两组中该区间的情感编码片段数都占到了总片段数的 82%, 说明观察者对大部分

图9 第2组  $P_i$  值的分布Fig. 9 Distribution of  $P_i$  in group 2

编码片段的评价是比较一致的。从4.1节的分析可以看出,  $\kappa$  的取值范围与观察者的数量  $n$ 、类别数  $k$ , 以及对象在各类别的分布都有关, 不同的度量间难以进行横向比较。从目前的度量结果看,  $\kappa$  值还有一定的提升空间, 可以通过改进实验设计来进一步提高观察者间的一致性。

## 5 结论

获取真实自然的情感数据以及可靠地度量情感真值, 是情感计算的研究基础。本文以 FIFA 实验场景为例, 在介绍情感数据获取的实验设计和情绪度量方法的基础上, 提出了基于情绪脚本的编码方法, 并对观察者间的一致性进行了初步的统计分析。

实验数据表明, 在精心设计的实验情境中, 被试者的情感表现真实自然, 情绪类别丰富, 能够体现特定情感的典型特征和情感产生、发展、高潮、消退的动态变化。

基于情绪脚本的情感编码方法, 能够对情感数据进行细致全面的分析, 即既能清晰地刻画情绪随时间的动态变化, 又能细致地还原情绪的原因、外在表情、行为等要素。

本实验同时对情感数据进行了自评和他评, 并采用了范畴观和维度观两种体系来力图对情感数据进行多角度的、定性的和定量的描述。

对观察者之间的评价一致性的分析表明, 观察者为绝大多数片段的情绪编码是比较一致的, 可见编码方法基本是可行的。

本文的工作仍存在很多不足, 希望在未来的工作中加以改进, 例如:

(1) 由于实验条件所限, 本实验未能采集游戏者的生理指标, 未来的工作拟采用生理多导仪或可穿戴的设备来测量被试者的血压、脉搏、皮肤电等生理指标, 以分析生理反应与情绪之间的对应性。

(2) FIFA 是一个相当单纯的环境, 由于游戏者的情感变化与游戏情境存在着较强的因果联系, 游戏者一般无需要掩饰或压抑情感, 因此, 基于情绪脚本的编码方法取得了较好的效果。但是, 真实生活中的情感反应是复杂的, 情绪脚本中的要素(如原因)可能不易观察; 另外, 受礼貌或文化的约束, 人们还会修饰情感的表达。这些都给进行情绪脚本的分析带来了一定的困难, 也限制了基于情绪脚本的编码方法的应用。因此如何改进基于情绪脚本的编码方法, 以适应复杂的生活场景, 是下一步工作中要解决的问题。

(3) 在本实验的情感编码中, 自评采用的是23组情绪词组检表, 他评采用的是6组情绪词组术语表, 二者并不相同, 这为分析自评和他评的一致性带来了困难。未来的工作拟采用6组情绪词组术语表重新进行自评, 以便在统一尺度下对自评和他评数据进行对比。

(4) PAD 模型<sup>[14]</sup>是一个3维情绪模型, PAD 分别代表愉悦度(pleasure)、激活度(arousal)和优势度(dominance)3个维度。PAD 量表以12对形容词组成的量表来间接度量 PAD 3个维度的等级。与直接的测量方法相比, PAD 量表更为科学。由于 PAD 量表的本地化工作尚未完成, 因此本实验没有采用 PAD 量表, 而是采用直接度量情绪维度等级的方法。未来的工作拟采用 PAD 量表对维度等级进行间接评价, 以提高度量的一致性和准确性。

## 附录 1: 情绪词组术语表

### 第1组

兴奋的: 心中激动难抑制

近义词: 亢奋的、狂喜的、激动的

喜悦的: 愉快、高兴, 从内心发出来的快乐

近义词: 欣喜的、开心的、快乐的、痛快的、畅快的

满意的: 满足自己的愿望, 符合自己的心意

近义词: 满足的

## 第 2 组

轻松的:负担不重或要求宽松

近义词:放松的、安心的、舒适的、闲适的

庆幸的:因可庆之事而欣喜

近义词:侥幸的

## 第 3 组

投入的:投到某种环境里去,指身心投入,倾注感情、精力等

近义词:专注的、陶醉的

冷静的:心中很沉静,常形容态度、头脑、思考等

近义词:镇定的

平静的:平稳安静

近义词:平和的

## 第 4 组

紧张的:精神处于高度兴奋状态

近义词:局促的

谨慎的:对外界事物与自己的言行都十分仔细小心,以免发生损害自己的不测之事

近义词:小心的、警惕的

## 第 5 组

失望的:不遂所望,因希望未实现而不愉快

近义词:扫兴的、遗憾的

无奈的:无可奈何,没有办法,无法可想

近义词:无助的

## 第 6 组

吃惊的:受惊,也指惊讶

近义词:惊讶的 惊恐的 震惊的

## 参考文献 (References)

- 1 Picard R. Affective Computing [M]. Cambridge, MA, USA: MIT Press, 1997.
- 2 Darwin C. The Expression of the Emotions in Man and Animals [M]. (Yu Ren Translate) London, United Kingdom: John Murray, 1872. [Darwin C 著. 余人译. 人与动物的情感 [M]. 成都: 四川人民出版社, 1999.]
- 3 Jones S, Carrere S, Gottman J. Specific affect coding system [A]. In;

Manusov V edi. The Sourcebook of Nonverbal Measures: Going Beyond Words [C], Mahwah, NJ, USA: Laurence Erlbaum Associates, 2005: 163-172.

- 4 Meng Zhao-lan. The Emotion of Mankind [M]. Shanghai: Shanghai Renmin Press, 1989. [孟昭兰. 人类情绪 [M]. 上海: 上海人民出版社, 1989.]
- 5 Zuo Yan-tao. Lexical Research on Chinese Personality [D]. Beijing: Peking University, 1995. [左衍涛. 汉语情绪词自评维度 [D]. 北京: 北京大学, 1995.]
- 6 Cowie R, Cornelius M. Describing the emotional states that are expressed in speech [A]. In: Proceedings of the ISCA Workshop on Speech and Emotion [C], Belfast, Northern Ireland, UK, 2000: 405-432.
- 7 Mota S A. Automatic Posture Analysis for Detecting Learner's Affective State [D]. Cambridge, MA, USA: Massachusetts Institute of Technology, 2002.
- 8 Aune K S. Assessing display rules in relationships [A]. In: Manusov V edi: The Sourcebook of Nonverbal Measures: Going Beyond Words [C], Mahwah, NJ, USA: Laurence Erlbaum Associates, 2005: 151-162.
- 9 Noller P. Behavioral coding of visual affect behavior [A]. In: Manusov V edi: The Sourcebook of Nonverbal Measures: Going Beyond Words [C], Mahwah, NJ, USA: Laurence Erlbaum Associates, 2005: 141-150.
- 10 Roberts N, Noller P. The affect measures of the couple communication scales [A]. In: Manusov V edit. The Sourcebook of Nonverbal Measures: Going Beyond Words [C], Mahwah, NJ, USA: Laurence Erlbaum Associates, 2005: 267-280.
- 11 Kantowitz B, Roediger H, Elmes D. Experimental Psychology Seventh Edition; Understanding Psychological Research [M]. (Guo Xiu-yan Translate) Belmont, CA, USA: Wadsworth/Thomson Learning, 2001. [Kantowitz B, Roediger H, Elmes D 著. 实验心理学——掌握心理学的研究 [M]. 郭秀艳译. 南京: 华东师范大学出版社, 2001.]
- 12 Goleman D. Emotional Intelligence [M]. (Geng Wen-xiu, Zha Bo, Translate) New York, USA: Bantam Books, 1997. [Goleman D 著. 情感智商 [M]. 耿文秀, 查波译. 上海: 上海科学技术出版社, 1997.]
- 13 Fleiss' kappa [EB/OL]. [http://en.wikipedia.org/wiki/Fleiss'\\_kappa](http://en.wikipedia.org/wiki/Fleiss'_kappa), 2007-08-21.
- 14 Mehrabian A. Framework for a comprehensive description and measurement of emotional states [J]. Genetic, Social and General Psychology Monographs, 1995, 121(3): 339-361.