

面向脑电数据的知识建模和情感识别

宿云^{①②}, 胡斌^{①*}, 徐立新^①, 张晓炜^①, 陈婧^①

① 兰州大学信息科学与工程学院, 兰州 730000;

② 西北师范大学计算机科学与工程学院, 兰州 730070

* 联系人, E-mail: bh@lzu.edu.cn

2014-10-13 收稿, 2015-03-05 接受, 2015-04-01 网络版发表

国家重点基础研究发展计划(2014CB744600)、国家国际科技合作专项(2013DFA11140)、国家自然科学基金(61210010, 61402211)和甘肃省青年科技基金计划(1308RJYA085, 1208RJYA015)资助

摘要 心理科学研究依赖于对生理、心理数据的分析, 情感是心理研究的重要内容. 近年来随着认知神经科学研究技术的成熟, 研究者利用脑电(electroencephalogram, EEG)等可以反映脑功能活动的生理信号, 直接研究情感问题, 如情感识别、情绪脑等. 但是, 生理信号将会产生TB级甚至PB级的数据量, 认知研究和临床神经科学在过去几十年中已产生大量生理数据, 对这些大数据的表示和情感知识挖掘需要更高级的工具. 构建能够表示数据含义和情感相关知识的模型, 能够给心理研究者提供一个知识共享平台, 以便使用这些大数据进行情感方面的科学研究. 本文构建一个可以表示EEG数据语义和被试者上下文信息的本体模型, 并基于该模型使用推理引擎进行基于EEG生理信号数据的自动情感识别. 实验结果表明, 模型在eINTERFACE 2006数据集上能够以99.11%的平均准确率识别被试者的情感状态, 并从实验结果分析发现基于EEG数据情感识别最关键的特征是Beta波与Theta波的绝对功率比.

关键词

EEG
本体
情感识别
基于规则推理
随机森林

自1897年冯特(Wilhelm Wundt)建立第一个心理实验室以来, 心理学的研究就与大量心理、生理数据紧密联系. 情感作为心理学研究的重要内容, 更是依赖大量生理信号数据. 如今, 这些生理数据的获取已经从传统的小样本数据采集到互联网上的大数据共享. 与情感相关的生理数据正在以指数速度迅速增长, 如单兵生命体征检测系统(War-fighter Physiologic Status Monitoring, WPSM)、基于生物反馈的心理障碍干预系统等, 这些系统通过无线传感器实时采集包括人体体温、血压、脑电、皮肤电、心电等生理信号, 检测士兵和患者的情感、压力等心理生理状态. 这些系统将产生GB级别甚至PB级别的数据量, 并且这些数据具有超高数据维度、高度计算复杂性及不确定性等特征^[1]. 对于这些生理大数据的表示和挖掘需要更高级的工具.

对于情感生理大数据, 不同领域的研究人员要对其进行分析, 如计算机、心理学领域研究者和医疗工作者等. 这些来自不同领域的研究人员需要这些大数据以分析研究对象的情感状态, 但他们不了解学科的术语、数据的含义、处理这些数据的计算方法, 研究人员该如何使用这些大数据? Hendler^[2]提到语义网技术将对这一问题的解决起重要作用. 语义网技术可以实现将非结构化数据(如生理信号、语音、视频等)转化为机器可识别的, 具有明确语义的信息, 如果人和机器能够知道数据的含义和领域的知识, 那便有利于不同领域的交流与合作, 实现语义的一致性和可判性, 扩展数据库之间的互操作性, 从而易于这些大数据的分析和挖掘. 语义技术正在医疗保健和生命科学等领域展现出强大的优势^[3], 如目前公开发布的基因本体、蛋白质本体等. 如何使用语义的

引用格式: 宿云, 胡斌, 徐立新, 等. 面向脑电数据的知识建模和情感识别. 科学通报, 2015, 60: 1002–1009

Su Y, Hu B, Xu L X, et al. EEG-data-oriented knowledge modeling and emotion recognition (in Chinese). Chin Sci Bull, 2015, 60: 1002–1009, doi: 10.1360/N972014-00829

方法和工具对生理大数据和相关知识进行建模,使得这些大数据和知识可以被不同领域的研究人员使用具有重要意义。

生理大数据背后隐藏着许多重要信息。如脑电 EEG 是由置于头皮上的传感器采集到的、由大脑产生的、由皮层传导的电信号^[4]。由于 EEG 可以反映大脑功能活动,并且具有无创、便携、低成本和高时间分辨率等特点,现已广泛用于生理心理学研究^[5,6]。很多研究表明,情感状态不同,EEG 模式也会随之不同,使用 EEG 信号进行人类情感识别是可行的^[7-10],基于 EEG 信号进行情感识别的研究受到越来越多的关注^[11-15]。

本文构建了一个 EEG 数据语义表示和情感识别模型。模型能够为心理学和其他领域研究者提供一个知识共享平台,并且可以在这个共享平台上实现基于 EEG 数据的自动情感识别,可用于监测高心理压力人群的情感状态。模型首先构建一个 EEG 与情感本体,该本体对被试者的上下文信息和 EEG 信号数据的相关知识进行建模;再使用机器学习算法——随机森林构建基于 EEG 数据进行情感识别的规则集;最后将建立的规则集写入推理引擎,该推理引擎就可对写入本体的被试者数据进行情感状态识别。我们之前的工作^[16,17]给出了这一研究的初步成果。本

文工作的主要贡献有以下两点:首先,使用了机器学习方法——随机森林来建立情感推理规则。实验结果表明,模型在 eINTERFACE 2006 数据集上能够以 99.11% 的平均分类准确率识别被试者的情感状态,改进了之前的 97.8% 的研究结果。其次,我们对随机森林建立的规则集进行分析,发现情感分类的关键特征是 Theta 波与 Beta 波的绝对功率比。

1 数据与方法

EEG 数据语义表示和情感识别模型框架如图 1 所示。该框架由数据处理、知识建模和情感推理 3 个模块构成。首先对数据进行预处理、特征提取;之后,为了共享和复用这些数据,提供对这些数据语义的共同理解,我们建立一个 EEG 与情感本体,该本体表达了 EEG 与情感领域相关的概念、属性及概念间的关系等陈述性知识。为了使用这些生理数据进行情感状态的识别,我们使用机器学习方法——随机森林从每名被试者的 EEG 数据中分别构建各个被试者的情感推理规则等程序性知识,从而实现对陈述性知识和程序性知识的建模^[18];最后推理引擎对本体中建模的被试者的 EEG 数据使用该被试者的推理规则集进行情感状态的自动识别。下面将详细介绍各个模块。

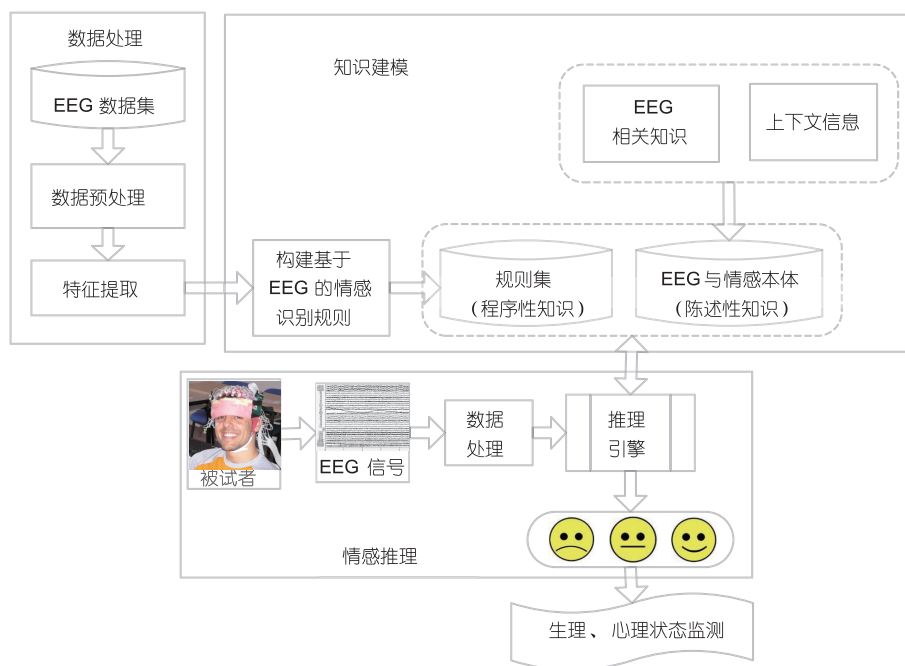


图 1 (网络版彩色) EEG 数据语义表示和情感识别模型框架

Figure 1 (Color online) Model framework for EEG semantic representation and emotion recognition

1.1 数据处理

(i) 数据描述. 本研究使用的数据集来自 eINTERFACE 2006 Project #7(<http://www.interface.net/results/>)^[19]. 由 Savran 等人^[20]发布的这个情感数据库记录了5名男性被试者在使用国际情感图片系统(International affective picture system, IAPS)中的图片作为诱发材料唤起消极(negative)、平静(calm)和积极(positive) 3种情感下采集到的EEG数据、功能近红外光谱(functional near infrared spectroscopy, fNIRS)数据及面部表情数据. 实验方案为: 每名被试者参与的实验分为3节, 每一节做30次试验, 每一次试验播放5张能诱发同一种情感的图片, 每张图片播放2.5 s. 播放完毕后, 被试者对唤起的情感状态进行自评. 该实验使用Russell效价-唤醒度(valence-arousal)二维情感模型^[21], 在该情感模型下定义消极、平静和积极3种情感.

(ii) 数据预处理和特征提取. eINTERFACE 2006数据库提供了BDF(Biosemi Data Format)格式的原始EEG数据、txt格式的原始fNIRS数据和被试者面部表情数据. 本文只处理EEG生理信号数据. 为方便情感识别, 需要对数据预处理和特征提取. 首先将被试者的自评与诱发图片的效价和唤醒度数值作比较, 发现有些诱发图片并没有唤起被试者期望的情感状态. 我们将未唤起期望情感状态, 即不满足下式的试验从数据集中移除.

$$Valence = |\overline{selfassessmentvalence} - E(valence)| \leq 1,$$

$$Arousal = |\overline{selfassessmentarousal} - E(arousal)| \leq 1.$$

最终, 被试者1和被试者2分别有40个有效试验, 而其余每名被试者分别有63个有效试验. 每一试验数据采集时间为12.5 s. 接着对这些筛选过的EEG数据做处理. 首先使用带通滤波平滑信号, 同时消除EEG漂移信号和肌电干扰信号, 接着使用小波算法去除眼电干扰, 最后对54导电极点的每1导采集的EEG信号在2 s重叠, 4 s滑动窗口提取线性和非线性特征, 如绝对功率、Shannon熵、C₀复杂度等.

1.2 知识建模

(i) 本体构建陈述性知识. 本体(ontology)是一种形式化的、对于共享概念体系的、明确而又详细的说明^[22], 即对研究领域的概念、属性以及概念之间的关系进行形式化表达的工具. 生活中, 图书分类法

可以看作是一个本体, 图书分类法一方面限定了术语集合, 另一方面定义了术语之间的关系, 如: 情绪与情感隶属于心理过程与心理状态, 心理过程与心理状态隶属于心理学等. 大家按照它来编排和索引书目, 将会便于书目的管理. 本体具有很强的知识表达能力, 可以较容易地表示出复杂数据和复杂的数据关系. 本体模型可以对领域知识进行有组织、按层次的管理, 赋予语义的数据可以方便地共享和复用, 并可实现知识表达的可判性和一致性.

为了表达eINTERFACE 2006数据库中的EEG数据和实验上下文的相关陈述性知识, 即它们的语义信息, 我们使用本体编辑工具Protégé软件(<http://protege.stanford.edu/>)构建了一个最基本的EEG与情感本体, 本体结构如图2所示. 从EEG相关知识和实验细节2个方面, 采用从上至下(top-down)策略定义类及类的层次, 首先定义最一般的父类, 即最抽象的实体概念, 然后对类进行细化, 定义更具体或范围更小的子类. 父类与子类之间的继承关系在图2的类层次框中以方框之间的连线表示, 如EEG特征类“EEG_Feature”下定义了2个子类: 线性特征子类“Linear_Statistical_Characteristics_Analysis”和非线性特征子类“Nonlinear_Characteristics_Analysis”. 结构图中底层圆弧框中的项是本体中构建的对象, 即类的实例. 如“calm”和“positive”是情感“Emotion”类的对象. 属性用来描述类具有的特性, 子类可以继承父类的属性, 类的对象可以继承该类的属性. 类与类之间的关系用对象属性表示, 部分对象属性如表1所示, 如被试者“Subject”类与EEG特征“EEG_Feature”类之间具有“hasEEGFeature”关系, 表示被试者有EEG特征. 类与数值之间的关系用数据属性表示, 部分数据属性如表2所示, 如类“EEG_Feature”具有数据属性“hasValue”, 这个属性的数据类型是实型“double”, 数值表示某一EEG特征值, 如类“EEG_Feature”的对象“AF3_Absolute_Power_Beta/Absolute_Power_Theta”的特征值, 即电极点AF3的EEG信号的Beta波与Theta波的绝对功率比特征值.

(ii) 随机森林构建程序性知识. 随机森林是一种组合分类器算法, 由多棵决策树组成. 对每一棵树, 从容量为 N 的原始样本集中采取有放回抽样的方法抽取相同容量的样本集, 从所有 M 个特征中随机选取 m 个特征, 根据节点不纯度最小原则^[23]从中选出一个特征进行分支生长, 直到节点上样本来自

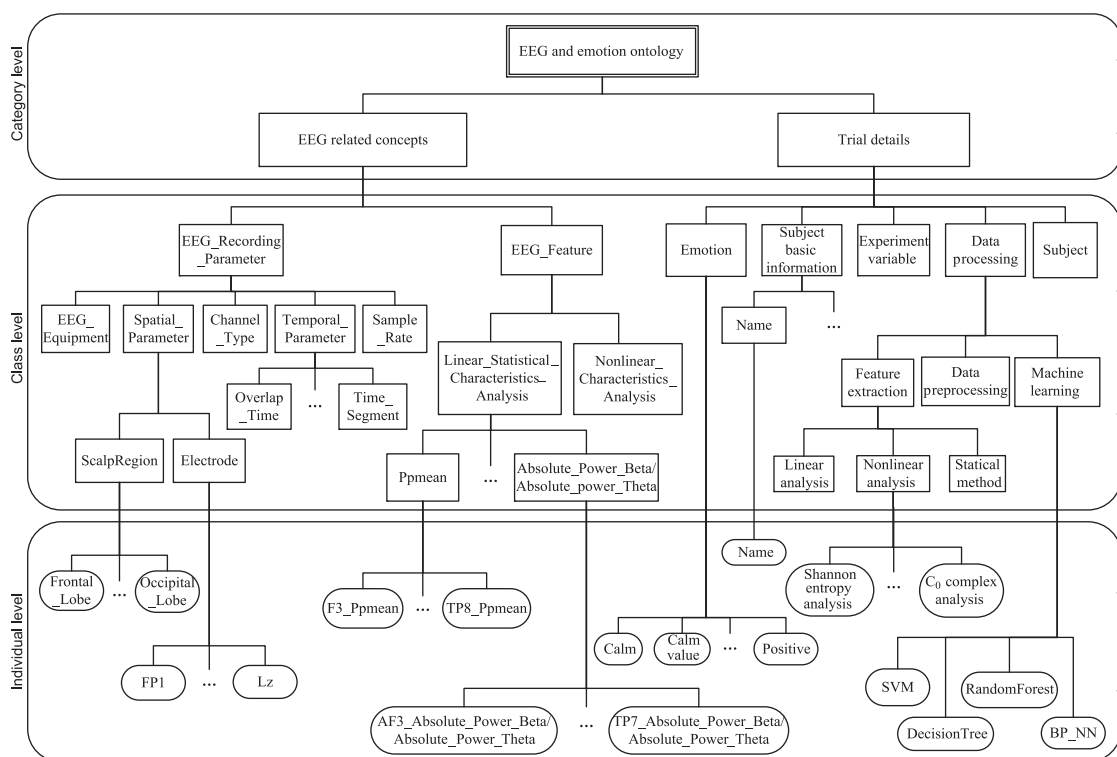


图2 EEG与情感本体结构
Figure 2 The structure of EEG and emotion ontology

表1 EEG与情感本体中的部分对象属性

对象属性	域	值域
hasEEGFeature	Subject	EEG_Feature
reflect	EEG_Feature	Emotion
onElectrode	EEG_Feature	Electrode

表2 EEG与情感本体中的部分数据属性

数据属性	域	数据类型
hasValue	EEG_Feature	double
deviceName	EEG_Equipment	string
hasName	Subject	string

同一类别或不纯度为零时停止。对于测试样本，每一棵树给出一个分类结果，然后进行投票确定最终分类结果^[24]。随机森林算法在具有很高分类准确率的同时，由它产生的分类规则还易于理解和分析，这些规则可以很方便地写成IF-THEN形式的产生式。因此，为了推理情感状态，我们使用随机森林算法从每

名被试者的EEG数据中分别构建各个被试者的情感推理规则等程序性知识。使用WEKA^[25] (版本3.7.11)中的随机森林算法构造的部分决策树如图3所示。从根节点到叶节点的一条路径就形成了一条推理规则。

1.3 情感推理

推理引擎将陈述性知识与程序性知识相关联以实现从陈述性知识中推理得出新的知识。我们使用Jena(<http://jena.apache.org/>，是一个创建语义网应用系统的Java框架)作为推理引擎，将用Protégé创建的EEG与情感本体和随机森林构建的被试者的推理规则集导入Jena，它包含的本体子系统和推理子系统允许处理本体数据并做基于规则的推理分析，并允许将数据存储到本体或者关系数据库中。当其中一名被试者的EEG数据输入推理引擎，推理引擎将会基于被试者的上下文信息找到对应的该名被试者的情感推理规则集，调用推理规则集识别出该名被试者的情感状态，之后，推理引擎还可将EEG数据和情感状态写入本体和数据库。图3中随机森林构建的一条规则(箭头所示)，Jena中的格式如下：

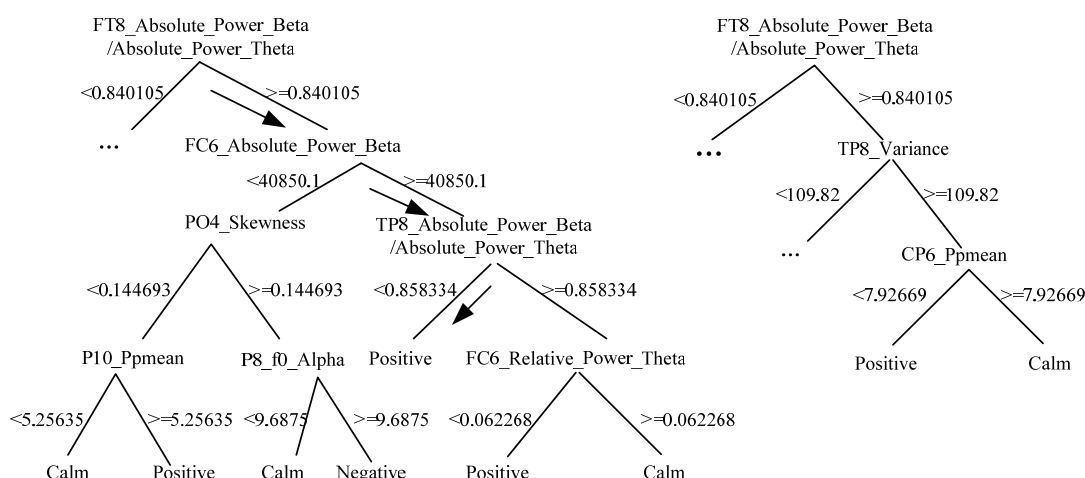


图3 随机森林生成的部分决策树

Figure 3 Part of decision trees constructed by random forest

```
[rule1: (?subject rdf:type base:Subject)
  (?f1 rdf:type base:Absolute_Power_Beta/Absolute_Power_Theta) (?f1 base:hasValue ?value1)
  greaterThan(?value1, 0.840105) (?f1 base:onElectrode ?electrode1) (?electrode1 rdfs:label "FT8")
  (?f2 rdf:type base:Absolute_Power_Beta) (?f2 base:hasValue ?value2) greaterThan(?value2, 40850.1)
  (?f2 base:onElectrode ?electrode2) (?electrode2 rdfs:label "FC6")
  (?f3 rdf:type base:Absolute_Power_Beta/Absolute_Power_Theta) (?f3 base:hasValue ?value3)
  lessThan(?value3, 0.858334) (?f3 base:onElectrode ?electrode3) (?electrode3 rdfs:label "TP8")
  (?positiveValue rdf:type base:Emotion) (?positiveValue base:hasEmotionValue ?value4)
  -> addOne(?value4, ?PositiveValue)]
```

其中?subject, ?f1, ?f2, ?f3, ?electrode1, ?electrode2, ?electrode3, ?value1, ?value2, ?value3 和 ?positive Value是自定义变量; base是EEG与情感本体的URL (Uniform Resource Locator, 互联网上标准资源地址); Subject, Absolute_Power_Beta, Emotion和 Absolute_Power_Beta/Absolute_Power_Theta是EEG与情感本体中定义的类; hasValue, onElectrode 和 hasEmotion Value是本体中定义的属性; greaterThan, lessThan和 addOne是Jena中的函数. 该条推理规则表示如果被试者?subject在FT8电极点上的EEG信号的Beta波与Theta波绝对功率比特征值大于或等于0.840105, 并且在FC6电极点上的EEG信号的Beta波绝对功率特征值大于或等于40850.1, 并且在TP8电极上的EEG信号的Beta波与Theta波绝对功率比特征值小于0.858334, 则该名被试者当前处于积极情感状态的可能增加1.

2 结果与讨论

本文构建了一个EEG数据语义表示和情感识别模型. 该模型使用本体建模数据库中EEG生理数据的语义信息及实验上下文信息; 使用随机森林算法从每名被试者已采集到的EEG数据样本分别建立各名被试者的情感推理规则集; 使用推理引擎推理情感状态. 推理引擎推理时, 首先利用本体中表示的被试者的上下文信息查询到该名被试者对应的推理规则集, 然后推理引擎调用这些情感推理规则集对本体中表示的EEG数据样本推理该名被试者的情感状态.

本研究实验使用的数据来自eINTERFACE 2006数据库中的EEG生理数据, 实验首先用Protégé软件建立EEG与情感本体, 接着使用WEKA工具箱中的随机森林分类器对每一位被试者构建情感状态推理规则集, 最后使用Jena推理机对本体中的知识进行情感状态推理.

为了评估由随机森林产生的规则集进行依赖于被试者的情感状态推理的准确率,本研究分别用不同的分类算法(包括支持向量机SVM, C4.5和神经网络BP)对处理后的eNTERFACE 2006中的EEG数据进行情感分类. 10折交叉验证分类准确率见表3. 其中C4.5算法采用WEKA工具箱中的J48分类器,默认参数. SVM使用LibSVM包,径向基核函数,参数 C 和 g 由网格寻优后得出. BP神经网络采用WEKA工具箱中的MultilayerPerceptron分类器,默认参数. 由于我们对54导EEG数据每1导都提取了时域、频域和统计特征,最终共提取了1300个特征,而每一名被试者的样本数最多只有315个,样本量远远小于特征维数,会影响SVM的分类准确率,也会使得神经网络需要很长的训练时间,因此在SVM和神经网络分类前都进行了特征选择.

由表3的数据可以看出随机森林可以获得相对较好的预测效果,5名被试者的平均分类准确率高达99.11%,高于之前报道的C4.5的准确率97.8%^[16]. 而BP和SVM只获得96.92%和83.17%的分类准确率. 虽然随机森林建模时间较长(平均建模时间为3.032 s, C4.5为0.912 s),但理论和实验研究都表明随机森林具有很高的准确率,对异常值和噪声具有很好的容忍度,且不容易过学习. 对于C4.5方法,由于构造树的过程是根据每个节点下局部的划分准则进行的,受样本随机性的影响明显,容易导致过学习. 同样,BP神经网络也有过学习问题. 而SVM在该样本集上准确率较低,可能原因是: SVM是一个2类分类器,而该数据集是3类情感数据集,如何提高SVM对多类问题的分类准确率和降低计算代价现在仍是一个在不断研究的课题^[26].

通过对随机森林得到的推理规则的分析得出,很多EEG电极点对应的特征对情感分类没有关系,

使用最多的EEG信号特征是Beta波与Theta波绝对功率比、Beta波绝对功率、Beta波最大功率、偏斜度、Kolmogorov熵、Alpha波绝对功率、方差、Shannon熵、峰峰平均值和峭度等特征,并且5名被试者Beta波与Theta波绝对功率比特征在规则集中出现的次数远远高于其他特征,如图4所示. 基于这一发现,在情感推理时,不需要提取所有电极点的所有特征,以减少预处理后的存储量.

本文提出的模型可以表达EEG信号的语义,并对其进行有组织、按层次的管理,进而赋予语义的EEG数据可以方便地共享和复用,因此可以为不同研究机构、不同领域专家提供一个数据共享平台;该模型可以实现与情感相关EEG知识表达的可判性和一致性;能够在表达了语义知识的脑电数据上进行情感状态推理. 由于本研究使用了随机森林算法,这使得被试者的情感推理具有很高的准确率;并且在训练规则集时,对样本的异常值和噪声具有很好的容忍度;在识别时,不容易过学习. 由于我们的模

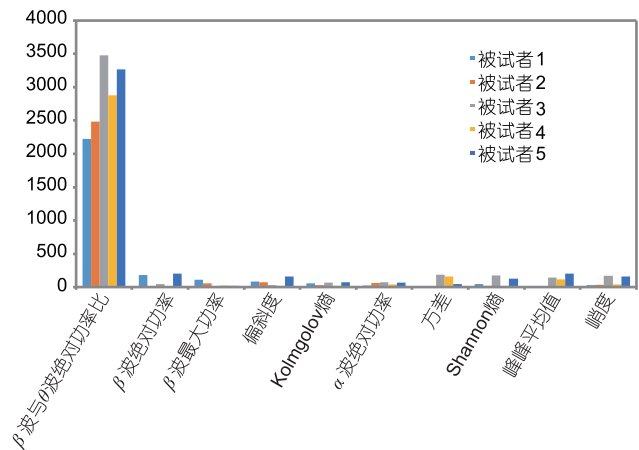


图4 情感推理规则中出现次数最多的特征

Figure 4 The most appeared features in emotion inference rules

表3 10折交叉验证以下几种分类算法的准确率

Table 3 Accuracy of several classifiers with 10-fold cross-validation

被试者	随机森林	C4.5	BP	SVM
被试者1	100.00% ($I = 50$ $K = 210$)	100.00%	94.50%	79.50% ($C = 11.3137$ $g = 0.011049$)
被试者2	100.00% ($I = 60$ $K = 144$)	98.50%	99.00%	86.50% ($C = 2.8284$ $g = 0.015625$)
被试者3	98.41% ($I = 50$ $K = 326$)	96.51%	99.68%	84.44% ($C = 5.6569$ $g = 0.015625$)
被试者4	100.00% ($I = 40$ $K = 210$)	100.00%	97.46%	79.68% ($C = 22.6274$ $g = 0.0055243$)
被试者5	97.14% ($I = 50$ $K = 196$)	93.97%	93.97%	85.71% ($C = 4$ $g = 0.03125$)
平均准确率	99.11%	97.80%	96.92%	83.17%

型是对每名被试者分别建立推理规则集,推理机进行推理时是对某被试者的一条样本用该被试者的推理规则集进行推理,因此,该模型可以实现生理大数据的语义表示和情感知识挖掘,这就为基于生理信号反馈的心理障碍(如抑郁)检测与干预提供了研究基础。

3 结论

为了应对生理大数据的发展趋势,更好地满足不同领域研究人员对生理数据进行情感分析的需求,亟需构建生理大数据平台,如本体知识平台。通过本体知识平台,研究者可以分析知识的来源、知识间的

因果关系,更好地理解 and 关联知识。本文构建了一个 EEG 生理数据与情感本体知识平台,并利用机器学习算法——随机森林构建情感推理规则,利用推理引擎将推理规则应用于本体知识平台,对其中建模的知识和数据推理被试者的情感状态,该方法不仅有效解决了 EEG 数据的语义表示问题,使得不同领域的研究者可以使用该数据开展不同的研究;而且可以在 eNTERFACE'06 数据库中的 EEG 数据集上以 99.11% 的平均准确率识别被试者的 3 种情感状态:消极、平静和积极。通过对实验结果分析得到对分类影响较大的特征是 Beta 波与 Theta 波的功率比,这为后续的研究提供有效的参考。

参考文献

- Guo H D, Wang L Z, Chen F, et al. Scientific big data and digital Earth (in Chinese). *Chin Sci Bull*, 2014, 59: 1047–1054 [郭华东, 王力哲, 陈方, 等. 科学大数据与数字地球. 科学通报, 2014, 59: 1047–1054]
- Hendler J. Science and the semantic web. *Science*, 2003, 299: 520–521
- Neumann E. A life science semantic web: Are you there yet? *Sci STKE*, 2005, 1–22
- Nunez P L, Srinivasan R. *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd ed. Oxford: Oxford University Press, 2006
- Liu J Z, Zhang D D, Luo Y J. Early development of the social and emotional brain in infancy (in Chinese). *Chin Sci Bull*, 2013, 58: 753–761 [柳俊哲, 张丹丹, 罗跃嘉. 婴儿社会和情绪脑机制的早期发展. 科学通报, 2013, 58: 753–761]
- Zhang J, Zhou R L. Frontal EEG laterality: An index of the capability of emotion regulation (in Chinese). *Adv Psychol Sci*, 2010, 18: 1679–1683 [张晶, 周仁来. 额叶 EEG 偏侧化: 情绪调节能力的指标. 心理科学发展, 2010, 18: 1679–1683]
- Ray W J, Cole H W. EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*, 1985, 228: 750–752
- Kostyunina M B, Kulikov M A. Frequency characteristics of EEG spectra in the emotions. *Neurosci Behav Physiol*, 1996, 26: 340–343
- Krause C M, Viemero V, Rosenqvist A, et al. Relative electroencephalographic desynchronization and synchronization in humans to emotional film content: An analysis of the 4–6, 6–8, 8–10 and 10–12 Hz frequency bands. *Neurosci Lett*, 2000, 286: 9–12
- Aftanas L I, Reva N V, Varlamov A, et al. Analysis of evoked EEG synchronization and desynchronization in conditions of emotional activation in humans: Temporal and topographic characteristics. *Neurosci Behav Physiol*, 2004, 34: 859–867
- Kim K H, Bang S W, Kim S R. Emotion recognition system using short-term monitoring of physiological signals. *Med Biol Eng Comput*, 2004, 42: 419–427
- Chanel G, Kronegg J, Grandjean D, et al. Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals. *Proceedings of the Conference on Multimedia Content Representation, Classification and Security*. Berlin Heidelberg: Springer Press, 2006. 530–537
- Schaaff K, Schultz T. Towards emotion recognition from electroencephalographic signals. *Proceedings of the 3rd International Conference Affective Computing and Intelligent Interaction and Workshops*. Piscataway: IEEE Press, 2009. 1–6
- Petrantonakis P, Hadjileontiadis L J. Emotion recognition from EEG using higher order crossings. *IEEE Trans Inf Technol B*, 2010, 14: 186–197
- Yazdani A, Lee J S, Vesin J M, et al. Affect recognition based on physiological changes during the watching of music videos. *ACM Trans Interact Intell Syst(TiiS)*, 2012, 7: 1–16
- Zhang X W, Hu B, Moore P, et al. Emotiono: An ontology with rule-based reasoning for emotion recognition. *Proceedings of the 18th International Conference, ICONIP 2011, Shanghai, China, Part II*. Berlin: Springer Berlin Heidelberg Press, 2011. 89–98
- Zhang X W, Hu B, Chen J, et al. Ontology-based context modeling for emotion recognition in an intelligent web. *World Wide Web*, 2012, 16: 497–513
- Esposito M, Pietro G D. An ontology-based fuzzy decision support system for multiple sclerosis. *Eng Appl Artif Intel*, 2011, 24: 1340–1354

- 19 Savran A, Ciftci K, Chanel G, et al. Emotion detection in the loop from brain signals and faial images. Proceedings of the eINTERFACE 2006 Workshop. Leuven: Presses Universitaires de Louvain, 2006. 1–10
- 20 Lang P J, Bradley M M, Cuthbert B N. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6, University of Florida, Gainesville. 2005, 1–56
- 21 Russell J A. A circumplex model of affect. *J Pers Soc Psychol*, 1980, 39: 1161–1178
- 22 Gruber T R. A translation approach to portable ontology specifications. *Knowl Acquis*, 1993, 5: 199–220
- 23 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Berlin: Springer Press, 2009
- 24 Breiman L. Random forests. *Mach Learn*, 2001, 45: 5–23
- 25 Garner S R. WEKA: The Waikato environment for knowledge analysis. Proceedings of the New Zealand Computer Science Research Students Conference, 1995. Alberta: GroupLab Press, 1995. 57–64
- 26 Hsu C W, Lin C J. A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw*, 2002, 13: 415–425

EEG-data-oriented knowledge modeling and emotion recognition

SU Yun^{1,2}, HU Bin¹, XU LiXin¹, ZHANG XiaoWei¹ & CHEN Jing¹

¹ School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China;

² College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China

Psychological research relies on physical and psychological data, and emotion has always been an important subject in the field of psychology. Recently, with the development of cognitive neuroscience technology, researchers can study topics such as emotion recognition and the emotional brain using electroencephalogram (EEG) and other physiological signals that reflect brain activity. These physiological signals can generate terabytes or even petabytes of data, and in fact, cognitive research and clinical neuroscience has already accumulated a wealth of data over the past several decades. Knowledge representation of this large amount of data and mining it for information regarding emotion requires more advanced tools. Making a data model that can clearly represent the meaning of data associated with emotion information would create a knowledge-sharing platform for psychological researchers to access the vast amount of data for further scientific research related to emotion. This paper provides such an ontology model that represents the semantics of EEG data with contextual information about the subjects. We used the model in conjunction with a reasoning engine to perform automatic emotion recognition based on EEG signals. Experimental results show that the ontology model reaches an average accuracy of 99.11% in identifying the emotional state of the subjects. Analysis of the results suggests that the most critical characteristic of EEG-based emotion recognition is the absolute power ratio between beta and theta waves.

EEG, ontology, emotion recognition, inference based on rules, random forest

doi: 10.1360/N972014-00829