

总能量、 E_α/E_θ 、 E_β/E_θ 、 E_γ/E_θ 、 E_β/E_α 、 E_γ/E_α 、 E_γ/E_β 、 C_0 复杂度^[15]、盒维数^[16-17]、信息熵^[18]。

2.2.1 C_0 复杂度

C_0 复杂度是一种适用于 EEG 这样的非平稳、非线性连续信号的复杂度指标，计算量较小、易于推广到高维，在实际应用中效果较好^[15]，能够识别癫痫发作和衡量睡眠程度。它的直观想法是将数据分为规则和非规则两部分， C_0 复杂度定义为信号非规则部分的能量与信号总能量的比值。 C_0 复杂度的数学定义为：

设 $f(k)$ ($k = 0, 1, 2, \dots, N - 1$) 是一个长度为 N 的时间序列，

$$F_N(j) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-2\pi i \frac{kj}{N}}, j = 0, 1, 2, \dots, N - 1$$

构成相应的 Fourier 变换序列，其中 $i = \sqrt{-1}$ 是虚数单位。

设 $F_N(j)$ 的均方值为

$$G_N = \frac{1}{N} \sum_{j=0}^{N-1} |F_N(j)|^2.$$

记

$$\tilde{F}_N(j) = \begin{cases} F_N(j), & \text{当 } |F_N(j)|^2 > G_N \text{ 时,} \\ 0, & \text{当 } |F_N(j)|^2 \leq G_N \text{ 时.} \end{cases}$$

对 $\tilde{F}_N(j)$ ($j = 0, 1, 2, \dots, N - 1$) 作 Fourier 逆变换

$$\tilde{f}(k) = \sum_{j=0}^{N-1} \tilde{F}_N(j) e^{2\pi i \frac{kj}{N}}, k = 0, 1, 2, \dots, N - 1,$$

定义 C_0 复杂度为

$$C_0 = \frac{\sum_{k=0}^{N-1} |f(k) - \tilde{f}(k)|^2}{\sum_{k=0}^{N-1} |f(k)|^2}.$$

2.2.2 盒维数

盒维数 (称闵可夫斯基维数^[17]) 是分形维数的一种，用于表征复杂对象的分形特性^[14]。

分形 S 的盒维数被定义为：

$$D_{\text{box}}(S) = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)},$$

其中格子边长为 ε ， $N(\varepsilon)$ 表示覆盖分形 S 最少所需的边长为 ε 的格子数。通过格子的逐步精细化，统计覆盖分形 S 所需的格子数目的变化，从而计算出盒维数。

本研究利用相空间重构和经典的 Takens 时间嵌入定理^[19]，将 EEG 信号嵌入二维平面构成二维轨迹，令该二维轨迹的盒维数代表该段 EEG 信号的盒维数。

2.2.3 信息熵



信息熵研究的是一段信号的不确定程度,反应信号分布均匀度,信息熵方法在过去十几年中被越来越多地证明可以有效地分析复杂的 EEG 信号^[5, 15, 20]. 经典信息熵的数学定义为:

$$H(X) = -\sum_i P(x_i)\log_2 P(x_i),$$

其中, X 是符号 x_i 组成的集合, $P(x_i)$ 是符号 x_i 出现的概率.

3 数据挖掘及结果

数据挖掘是从大量数据中获取隐含的有用信息的技术,可划分为分类、回归、聚类、关联规则、特征选择等几类. 为了找到禅修分类性能较好的算法,本研究使用开源机器学习和数据挖掘软件 Weka^[21]尝试了多种分类和回归算法,包括随机森林(Random Forest, RF)、随机树(Random Tree, RT)、C4.8 决策树(J48)、局部决策树(Partial decision tree, PART)^[22]、支持向量机(Support Vector Machine, SVM)、贝叶斯网络 (BayesNet)、多层感知机 (Multilayer Perceptron, MP)、 k 近邻 (k -Nearest Neighbor, k NN), 将 15 维 EEG 特征和禅修评分作为输入, 并尝试对每位被试的数据进行个性化校准, 以解决个体差异问题, 提高分类准确率.

回归分析将禅修评分视作数值型数据, 对评分进行拟合. 这相当于直接构造一个打分模型, 将高维特征映射为一个数值型得分. 采用 10 折交叉验证, 结果以相关系数的形式给出, 相关系数越高意味着拟合结果与测试集越接近.

多分类将禅修导师打分视作离散型数据, 2 分至 8 分共 7 类, 采用 10 折交叉验证.

进一步地, 考虑到准确识别出高水平学员的 EEG 特征是问题关键, 对低分样例进行详细划分和挖掘的意义不大, 因此将高分样例 (分数 ≥ 7 , 前 20%) 作为正例, 其余样例作为反例进行二分类, 采用 10 折交叉验证.

未经校准过程的数据挖掘实验结果如表 2 所示.

EEG 的个体差异往往很大, 既与被试生理差异有关, 又和被试当时的意念状态、情绪、思维、性格等心理因素有关. 本研究尝试利用静虑冥想的基准数据对每名被试进行个性化校准, 假设被试静虑时的 EEG 包含该被试的个体特征, 在其进行专注、正念和坐忘任务时的 EEG 既包含特定冥想任务特征, 也包含相同的个体特征, 两者相减就能够得到表征特定冥想任务的 EEG 特征.

定义绝对校准数据 x'_{ijk} 和相对校准数据 x''_{ijk} 分别为:

$$x'_{ijk} = x_{ijk} - \overline{x_{0j}}$$

$$x''_{ijk} = \frac{x'_{ijk}}{\overline{x_{0j}}}$$

式中, i 表示冥想类别, $i \in \{1, 2, 3\}$, 分别代表专注、正念和坐忘, j 表示被试编号, k 表示样



例编号, \bar{x}_{0j} 表示被试j的基准数据的平均值.

本研究试验了绝对校准和相对校准两种情况, 数据挖掘结果分别如表 3 和表 4 所示.

表 2 未经校准过程的数据挖掘结果

Tab. 2 Data mining result (without calibration)

数据挖掘方法	相关系数 (回归)			准确率 (多分类) /%			准确率 (二分类) /%		
	专注	正念	坐忘	专注	正念	坐忘	专注	正念	坐忘
RF	0.607	0.634	0.862	68.4	70.9	80.6	82.7	81.1	93.7
RT	0.340	0.388	0.695	54.7	60.6	66.6	79.9	77.6	91.2
J48	-	-	-	61.6	64.3	76.3	79.6	78.5	90.0
PART	-	-	-	57.2	63.4	70.9	78.6	79.5	90.9
SVM	0.520	0.504	0.745	61.9	62.5	64.7	79.6	79.8	86.2
BayesNet	-	-	-	45.0	48.6	66.6	77.0	76.7	85.3
MP	0.546	0.359	0.357	59.1	53.8	65.6	76.7	77.0	84.6
kNN(k=1)	0.360	0.340	0.652	54.1	52.2	61.3	78.3	79.2	87.5
kNN(k=5)	0.478	0.482	0.740	55.9	59.0	63.1	78.9	78.9	86.2

表 3 绝对校准后数据挖掘结果

Tab. 3 Data mining result (with absolute calibration)

数据挖掘方法	相关系数 (回归)			准确率 (多分类) /%			准确率 (二分类) /%		
	专注	正念	坐忘	专注	正念	坐忘	专注	正念	坐忘
RF	0.829	0.875	0.933	85.6	85.6	88.8	93.1	93.8	96.9
RT	0.668	0.686	0.837	71.9	73.1	78.8	89.1	87.2	95.6
J48	-	-	-	78.8	76.9	82.5	88.8	87.5	97.2
PART	-	-	-	76.6	79.1	83.4	86.2	92.2	95.9
SVM	0.695	0.596	0.755	61.9	56.3	67.5	85.0	82.8	90.3
BayesNet	-	-	-	74.7	69.1	83.8	88.4	87.2	96.6
MP	0.547	0.492	0.636	66.9	58.4	69.4	86.6	85.0	93.8
kNN(k=1)	0.555	0.400	0.678	60.9	55.6	62.8	82.5	81.0	90.0
kNN(k=5)	0.645	0.567	0.773	63.8	59.4	69.4	85.3	83.8	91.3

表 4 相对校准后数据挖掘结果



Tab. 4 Data mining result (with relative calibration)

数据挖掘方法	相关系数 (回归)			准确率 (多分类) /%			准确率 (二分类) /%		
	专注	正念	坐忘	专注	正念	坐忘	专注	正念	坐忘
RF	0.692	0.741	0.869	66.6	71.3	80.9	87.2	88.1	94.7
RT	0.397	0.498	0.737	58.1	61.9	66.9	80.3	80.3	91.2
J48	-	-	-	56.6	64.1	75.6	83.8	84.4	89.4
PART	-	-	-	57.8	68.4	71.6	81.3	84.1	91.3
SVM	0.658	0.640	0.783	61.6	57.8	66.3	81.9	85.9	93.1
BayesNet	-	-	-	61.3	58.8	70.6	82.5	91.9	93.4
MP	0.416	0.462	0.511	63.8	55.9	70.0	78.8	85.3	93.1
kNN(k=1)	0.442	0.501	0.765	57.2	53.1	64.4	78.1	81.6	93.1
kNN(k=5)	0.582	0.600	0.792	61.3	59.4	68.4	81.0	86.6	91.9

经个性化校准后,多个回归和分类算法结果均有大幅度提升.对比表3和表4可以看出,绝对校准方案比相对校准方案对准确率的提升更有效.利用绝对校准数据 x'_{ijk} ,随机森林算法在回归任务中的平均相关系数达到0.879,比表2的结果提高了0.175;多分类任务的平均准确率达到86.7%,二分类任务的平均准确率达到94.6%,比表2的结果分别提高了13.4和8.8个百分点.

为了更精确衡量绝对校准方案对二分类问题准确率的提升作用,本研究进一步对比了绝对校准前后随机森林、支持向量机和贝叶斯网络算法的准确率和受试者工作特征曲线下面积 (Area Under Receiver Operating Characteristic Curve, AUC)^[23],结果如表5所示.

表5 绝对校准前后二分类准确率结果

Tab. 5 Accuracy result (without/with absolute calibration)

冥想类别	方法	Accuracy/%	PRC Area	ROC Area
专注	RF	82.7 → 93.1	0.891 → 0.968	0.865 → 0.962
	BayesNet	77.0 → 88.4	0.715 → 0.945	0.654 → 0.928
	SVM	79.6 → 85.0	0.689 → 0.763	0.612 → 0.700
正念	RF	81.1 → 93.8	0.886 → 0.982	0.862 → 0.979
	BayesNet	76.7 → 87.2	0.713 → 0.943	0.611 → 0.930
	SVM	79.8 → 82.8	0.691 → 0.733	0.603 → 0.665



坐忘	RF	93.7 → 96.9	0.972 → 0.997	0.971 → 0.997
	BayesNet	85.3 → 96.6	0.854 → 0.993	0.822 → 0.993
	SVM	86.2 → 90.3	0.781 → 0.849	0.726 → 0.831

注：校准前结果 → 校准后结果

4 讨 论

本研究专门针对乐易心法打坐冥想流程设计实验, 被试皆为乐易心法七日禅学员, 采集到的 EEG 数据对乐易心法冥想具有较高参考价值, 比一般的正念冥想研究^[3-4, 11, 24]和类似“开心禅”和“禅定花园”等应用更具中华传统文化内涵, 也更具前沿性和难度. 乐易心法禅修按中华传统禅修方式进行, 所有禅修者与导师一起封闭培训, 这样的禅修班难以大批量复制, 使得本研究被试数量较少. Ali 等^[11]研究人员对 6 000 多名各个年龄段的被试进行了研究, 发现了 EEG 特征在年龄上的变化规律. 但该研究采用的冥想为“呼吸计数法”, 只是让用户通过冥想来放松心情, 无禅修和禅悟的内容. 相比于 MeditAid 系统^[10]利用双耳节奏神经相关效应 (binaural beats neurocorrelate) 通过播放音频激发特定 α 波的做法, 本研究尝试将 NFB 技术与传统禅修方式融合, 被试的禅修水平受禅修导师评估, 能够更好地为禅修者服务. 本研究广泛利用了多种 EEG 指标和分析方法, 分类准确性能够随着数据集的增大有更好的表现.

采用本文提出的个性化校准方案后, 多个回归和分类算法结果均有大幅度提升, 这可能与将“静虑”作为基准数据有关. 乐易心法中的静虑要求禅修者去除杂念, 平静内心, 是一种禅修语境下“更纯粹”的意念状态. 本研究假设静虑时的 EEG 包含被试个体差异成特征分, 在个性化校准时减去静虑特征值就能抵消个体差异和无关信息, 从而突出反映后续冥想阶段的特征, 提高分类算法的准确率. 多个算法对二分类问题的准确率均高于 93%, 证明了 NFB 系统能够较好地识别冥想好坏状态, 帮助使用者及时纠正错误.

受限于被试人数和实验时间, 本研究实际的被试数只有 16 人, 高分人数只有 4 人, 在训练和测试时虽然使用 10 折交叉验证, 但每个学员的样例以高概率参与了训练, 因此本研究的实验条件很接近测试集就是训练集的情况.

考察各个算法给出的分类模型后发现, 模型用到的特征数量一般在 5 个以上, 直接使用主成分分析对 15 维数据进行降维后给出的数据维度是 7, 这说明想要达到较好的识别准确度, 算法必须同时利用多个特征. 主成分分析会重构线性空间, 新坐标系的可解读性和可分



析性变差,不利于对 EEG 结果进行解读和分析,因此本研究不采用主成分分析进行维度压缩.由于人的注意力和意念控制能力有限,若 NFB 系统反馈的参数过多,很容易令使用者顾此失彼,达不到训练效果.优秀的 NFB 系统既要能准确识别目标状态,又要为使用者反馈精简有效的特征调整指令,因此后续研究应着眼于特征降维.另外,将禅修导师判断学员水平的一些依据纳入 NFB 系统也许是个可行办法.

5 结 论

本文采集和研究了乐易心法七日禅学员打坐冥想的 EEG 数据,采用数据挖掘的研究思路和方法,使用 15 维 EEG 特征和禅修导师评分作为数据集,以禅修学员水平作为分类目标,对数据采用个性化校准方案,测试了随机森林、支持向量机、 k 近邻、贝叶斯网络等多种分类和回归算法的准确率性能.实验结果表明,本文提出的 15 维 EEG 特征和个性化校准方案可以有效解决 EEG 研究中的个体差异问题,提高冥想水平分类准确率,为更智能的冥想 NFB 系统研发提供了支持,也为相关研究领域提供了一种解决 EEG 个体差异问题的思路.

参考文献

- [1] 余国良,董妍.我国心理健康研究的现状、热点与发展趋势[J].教育研究,2012,33(06):97-102.
- [2] 周昌乐.乐易心法介绍.博学切问[M].厦门:厦门大学出版社,2015:209-212.
- [3] TANG Y Y, HOLZEL K, POSNER M. The neuroscience of mindfulness meditation[J]. Nature Reviews Neuroscience, 2015, 16(4): 213-225.
- [4] ZHOU R L, LIU L. Eight-week mindfulness training enhances left frontal EEG asymmetry during emotional challenge: a randomized controlled trial[J]. Mindfulness, 2017, 1(8): 181-189.
- [5] 赵建强,周昌乐.短期禅修效果的脑电图多尺度排列熵分析[J].厦门大学学报(自然科学版),2016,55(3):420-425.
- [6] EVANS R J. Handbook of neurofeedback[M]. Binghamton: The Haworth Medical Press, 2007: 3-10.
- [7] 吴金华,张艳秋,唐毅.数据挖掘在生物信息学中的应用-文献计量学视角[J].生物信息学,2016,14(4):249-253.
- [8] LIU Y S, SOURINA O, HOU X Y. Neurofeedback Games to Improve Cognitive Abilities[C]//2014 International Conference on Cyberworlds, Santander: IEEE, 2014: 161-168.
- [9] BRANDMEYER T, DELORME A. Meditation and Neurofeedback[J]. Frontiers in Psychology, 2013, 4(1): 688-670.
- [10] SAS C, CHOPRA R. MeditAid: a wearable adaptive neurofeedback-based system for training mindfulness



- state[J]. *Pers Ubiquit Comput*, 2015, 7(19): 1169-1182.
- [11] HASHEMI A, PINO J LOU, MOFFAT G et al. Characterizing Population EEG Dynamics throughout Adulthood[J]. *eNeuro*, 2016, 3(6): 1-13.
- [12] 周昌乐. 从当代脑科学看禅定状态达成的可能性及其意义[J]. *杭州师范大学学报(社会科学版)*, 2015(3): 17-23.
- [13] Emotiv Inc. EMOTIV EPOC & TESTBENCH™ SPECIFICATIONS[EB/OL]. [2017-10-07].
<https://www.emotiv.com/files/Emotiv-EPOC-Product-Sheet-2014.pdf>.
- [14] FELL J, AXMACHER N, HAUPT S. From alpha to gamma: electrophysiological correlates of meditation-related states of consciousness[J]. *Medical Hypotheses*, 2010, 75(2): 218-224.
- [15] 汪云九. 神经信息学-神经系统的理论和模型[M]. 北京: 高等教育出版社, 2006: 403-407.
- [16] 佩特根, 于尔根斯, 绍柏. 混沌与分形[M]. 田逢喜, 译. 北京: 国防工业出版社, 2008: 300-305.
- [17] WEISSTEIN W E. "Minkowski-Bouligand Dimension." From *MathWorld*-A Wolfram Web Resource[EB/OL]. [2017-12-18]. <http://mathworld.wolfram.com/Minkowski-BouligandDimension.html>
- [18] SHANNON C. A mathematical theory of communication[J]. *Bell Systems Technical Journal*, 1948, 3(27): 379-423.
- [19] TAKENS F. Detecting strange attractors in turbulence[C]//RAND D A, YOUNG L S. *Dynamical Systems and Turbulence*. Heidelberg: Springer, 1981: 366-381.
- [20] 李颖洁, 邱意弘, 朱贻盛. 脑电信号分析方法及其应用[M]. 北京: 科学出版社, 2009: 1-3.
- [21] FRANK E, HALL A M, WITTEN H I. *The WEKA Workbench*[M]. San Francisco: Morgan Kaufmann, 2016: 1-8.
- [22] FRANK E, WITTEN H I. Generating Accurate Rule Sets Without Global Optimization[C]//JUDE W S. *ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning*, San Francisco: Morgan Kaufmann, 1998: 144-151.
- [23] 周志华. *机器学习*[M]. 北京: 清华大学出版社, 2016: 30-35.
- [24] CHIESA A, CALATI R, SERRETTI A. Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings[J]. *Clinical Psychology Review*, 2011, 31(3): 449-464.

Study on EEG Data Mining for Meditation Neurofeedback System



XU Hao¹, HUANG Min^{1,2}, ZHOU Changle^{1*}

(1. Fujian Key Lab of Brain-like Computation Technology and Application, School of Information Science and Engineering, Xiamen University,
2. College of Humanities, Xiamen University, Xiamen 361005, China)

Abstract: Meditation benefits the health of modern people. In order to develop smart neurofeedback system which is helpful for zen practice, several EEG indexes and 3 advanced analyze methods are used to classify good or bad new Zen practitioners based on the evaluation made by zen master. Individualized calibration process is conducted. The results indicate that 15-dimensional EEG feature and individualized calibration process could effectively solve problems of individual difference of EEG and the accuracy of the classification algorithms can get higher than 93% for identifying good meditation practice. The results support the development of a smarter meditation neurofeedback system.

Key words: electroencephalogram (EEG) ; meditation ; data mining ; machine learning ; neurofeedback ; individualized calibration ; zen

廈門大學學報 (自然科學版)