

基于二重SVM模型的移动互联网用户情绪预测方法

姜作振, 章旭*, 明晨宇, 孟婷, 汪乐

安徽工程大学经济与管理学院, 安徽 芜湖
Email: *2857461189@qq.com

收稿日期: 2020年11月15日; 录用日期: 2020年11月30日; 发布日期: 2020年12月7日

摘要

针对移动互联网用户情绪预测问题, 收集了40名被试观看偏好以及不偏好视频的主观情绪以及实时心率波动的数据。独立样本T检验结果表明, 不同情绪下心率变化幅度存在显著差异, 验证了心率在鉴别用户情绪的可行性。在此基础上, 提出采用机器学习构建面向移动互联网用户的情绪预测模型。所构建的二重SVM模型对三类情绪(开心、中性和难过)的鉴别成功率均在75%以上。研究结果还表明实时心率数据能够较好地反映出移动互联网用户的情绪变化。

关键词

情绪测量, 心率波动, 独立样本T检验, 二重SVM

A Mobile Internet User Emotion Prediction Method Based on Dual SVM Model

Zuozhen Jiang, Xu Zhang*, Chenyu Ming, Ting Meng, Le Wang

School of Economics and Management, Anhui Polytechnic University, Wuhu Anhui
Email: *2857461189@qq.com

Received: Nov. 15th, 2020; accepted: Nov. 30th, 2020; published: Dec. 7th, 2020

Abstract

To solve the problem of mood prediction of mobile Internet users, the subjective mood and real-time heart rate fluctuation data of 40 subjects were collected. Independent sample t-test re-

*通讯作者。

文章引用: 姜作振, 章旭, 明晨宇, 孟婷, 汪乐. 基于二重 SVM 模型的移动互联网用户情绪预测方法[J]. 统计学与应用, 2020, 9(6): 958-963. DOI: 10.12677/sa.2020.96100

sults show that there are significant differences in the range of heart rate changes under different emotions, which verifies the feasibility of heart rate in identifying users' emotions. On the basis of this, the paper proposes to use machine learning to build an emotion prediction model for mobile Internet users. The success rate of the three kinds of emotions (happy, neutral and sad) was more than 75%. The results also show that real-time heart rate data can better reflect the emotional changes of mobile Internet users.

Keywords

Emotion Measurement, Heart Rate Fluctuation, Independent Sample T-Test, Dual SVM

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着移动互联网时代的发展, 互联网用户的数量急剧增加。互联网行业竞争亦十分激烈。移动终端要想在众多的竞争对手中脱颖而出, 获得用户的喜爱, 必须要对其用户的使用感受有一个深入的了解。这就使得网站设计和运营人员要能够及时了解用户情感需求, 根据用户情感动态调整网站设计和运营方案。在不影响用户浏览和操作体验的情况下, 如何测量并预测移动互联网用户情感状态, 是目前情感测量领域的热点与难点问题[1]。

2. 文献综述

Picard 在 1995 年率先采用了血容量脉搏、呼吸信号、皮肤电信号和肌电信号 4 种生理信号, 并证明了通过提取生理信号特征进行情绪识别的可行性。其中, 脉搏信号是一种较弱的生物电信号, 蕴含了人体的多种生理及病理信息。相比于其他生理信号, 脉搏信号的噪声较少且易采集[2]。目前, 在情绪识别研究领域, 用脉搏信号识别情感的研究较少, 因此研究脉搏信号的情感识别具有重要意义。

于目前用于情感识别的其它分类器, 支持向量机因其优秀的小样本分类效果, 被广泛使用[4]。朱菊霞, 吴小培[3]等人利用支持向量机作为分类器, 以语音信号特征参数作为依据, 对高兴、愤怒和悲伤 4 种情感进行识别, 识别率比人工神经网络等分类算法均高了 7% 以上。实验表明了支持向量机算法具有高效性和更高的准确性。杜昭慧, 司玉娟在解决随着情感种类增加而增加的支持向量机的个数时, 运用层次 SVM 分类在保证分类准确率的同时, 提高了分类速度。本文构建的二重 SVM 采用了偏二叉树[5]构造从而达到了分类三大情感的设计目的, 分类速度快且容易实现。

3. 实验设计

本实验研究目的是测量被实验者在观测不同内容时心率波动数据以及主观评测情绪变化, 通过对数据的分析处理, 探讨情绪变化对心率波动的影响。

3.1. 实验对象

本研究采取随机抽样法, 抽取某高校在校大学生 40 名, 男性、女性各 20 名, 被试者均为健康状态, 拥有充足的睡眠, 无疲劳症状, 在进行实验前 24 小时内没有进行药物治疗、没有进行高强度的训练, 自愿参加实验。实验环境: 室内照明 125 LX, 声环境 30 dB, 温度 24℃, 空气湿度 60%。

3.2. 实验过程

首先,开始试验前,记录被试者的第一次心率数据。试验过程,根据个人喜好,给予不同被试者观看事先收藏且尚未观看的不同内容的短片,试验过程为30秒,将每个实验过程按照总持续时间截取相应片段:0~10 s、10~20 s、20~30 s,记录三个时间段各自的平均心率数据。试验结束后,调查被试者的主观情绪感知,记录主观数据。

4. 数据分析

为了便于接下来的数据收集与分析工作,把40个被试分别以{P1, P2, P3, ..., P40}表示,将每个实验过程按照总持续时间截取相应片段:0~10 s、10~20 s、20~30 s共三段,分别用{T1, T2, T3}表示,实验前与实验后的静息时间用T0和T4表示。在剔除部分因人为因素导致心率波动较大以及缺少部分片段平均心率的数据后,收集到15份主观情绪为开心,10份难过和5份中性的问卷以及对应的心率波动数据。

通过对开心情绪的心率变化幅度进行描述分析可以得出实验前T0的静息心率与实验开始后T1的心率差距较小,变化幅度约为1%。而在实验开始后的T1与T2有较大的变化幅度,总体呈现上涨的态势,上涨幅度差超过3%,T2与T3的平均心率也变化较大,总体表现出下降的趋势,降低幅度达到3.5%。中性情感的心率变化幅度较低,保持稳定,这与实际相符,同时再次验证了情感会影响心率变化的假设。T3与T4波动幅度约降低了1%,本文认为是实验后平静状态下心率自发的轻微降低。难过情感的心率变化幅度总体呈现降低趋势。第一次心率波动表现轻微的上涨,上涨幅度约为1%,第二次心率波动幅度为-5.5%,可以看出实验开始后的10秒与10到20秒之间心率明显降低。第三次心率降低幅度几乎为0,本文认为这是因为第二次心率降低幅度较大使第三次心率降低的空间大大减少导致的。

由于三种情感第一次心率波动情况无明显差异,故本文猜测情感的变化与心率对应发生变化存在延迟,为验证这一猜测,选取3名被试在心率平静的情况下立刻做急速的俯卧撑,记录心率仪检测到心率上升所需时间,取3名被试心率上升所需时间平均值,初步发现剧烈运动后的12秒左右才会被心率仪检测到心率上升。初步认为该原因是由于运动开始与心率开始上升之间存在时间差,且此次实验采用的心率仪是基于光电容积脉搏波描记法原理的测量仪器,实时检测数据与心率的变化存在一定的时间差[6][7]。故可以初步认为实验开始的T1内被试虽然受到不同的刺激,但心率依然处于静息心率,在12秒后心率的变化才被体现出来。实验前T0与实验开始的T1心率都处于静息的状态,故第一次心率波动表现出整体变化幅度较小且与情感无关的现象,后文的研究将跳过实验前10秒的心率测量和第一次心率变化结果。

采用独立样本T检验[8]分析比较心率波动是否存在显著性差异。结果表明中性与开心的心率波动不存在显著差异($p = 0.14$);难过与中性、难过与开心的第二次心率波动幅度存在显著性差异($p = 0.044$)。虽然第三次心率波动开心与难过、中性与难过的p值大于0.05,但心率波动情况总体均值存在较大差距,实验结果为后文二重SVM情绪识别分类的构建提供了依据。

5. 基于二重SVM的情感识别

根据前文的研究结果,开心与难过、难过与中性均存在显著性差异,故本文构建“开心鉴定”的SVM和“难过鉴定”的SVM。鉴定开心、中性与难过情感的过程如图1。

与之前实验类似(不同的是省去实验前10秒静息心率的测量与处理,理由见前文解释),再次收集6组开心、4组难过和两组中性心率每10秒的平均值及两次心率波动幅度数值。首先根据前文收集到的30组数据作为训练样本,依次构建能最优划分开心与非开心、难过与中性的超平面[9]。然后将新收集的12组数据作为测试样本检验SVM分类的精确度。从30组已分类样本的分类结果可以看出“开心鉴

定” SVM 对原始样本的分类精度为 76.67%。分类结果可以发现“开心鉴定” SVM 在对开心的判断上较主观判断更为严苛，使得它将被试主观认为开心的结果分到非开心类别中去，相对的，对于非开心的情感分类准确度就非常高(准确度 86.67%)。“开心鉴定” SVM 测试样本的分类结果如图 2。

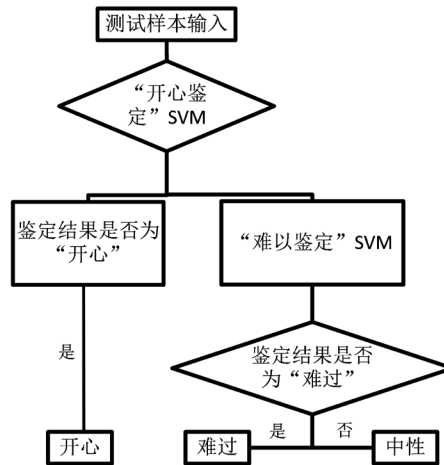


Figure 1. Dual SVM emotion recognition logic diagram

图 1. 二重 SVM 情感识别逻辑图

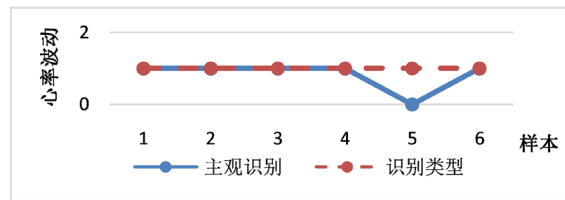


Figure 2. The SVM classification results of “happy”

图 2. “开心” SVM 分类结果对比图

可以看出验证样本分类准确率为 83.33%。对于非开心类别的样本(与主观分类不同的是将样本 1、2、3、4、6 停止分类，将样本 11 进一步分类)，再代入“难过鉴定” SVM，分类情况如图 3。

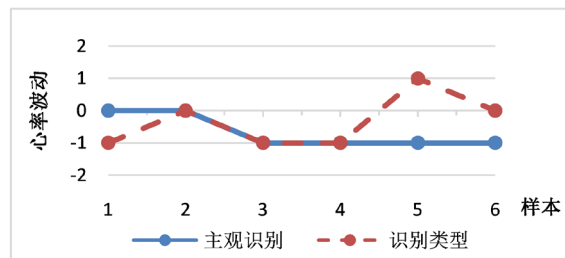


Figure 3. The SVM classification results of “sad”

图 3. “难过” SVM 分类结果对比图

结合原始的 15 组非开心样本的主观分类结果相统计，“难过鉴定” SVM 精度为 80%，12 个测试样本基于二重 SVM 对三种情感的分类准确度高达 75% (图 4)。

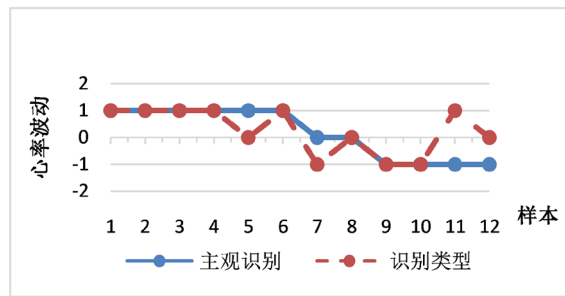


Figure 4. The SVM classification results of “double”

图 4. 二重 SVM 分类结果对比图

6. 结论与展望

本研究提出利用心率变化构建移动互联网用户的二重 SVM 预测模型。研究发现，开心情绪、中性情绪和难过情绪的情绪变化与机体心率变化存在一定的相关关系。总体来看心率变化与开心情绪表现为增强关系，与难过情绪表现为抑制关系。这一结果符合情绪的生理机制[10]。由于主流心率测量仪是通过检测组织中血液容积变化来间接测量心率变化的，这种检测方法对于心率的变化在时间上受制于机体血液循环的影响[10]，即数据变化对于心率变化因机体不同存在 12 到 20 秒的时间差。所以测得心率数据变化对于情绪变化的反映上存在不可避免的误差。在以后的研究中可以从测量原理和机体血液循环时间上研究数据结果误差，在情绪测量上也要从更多的生理数据及角度来观察和研究。本研究的不足之处在于，中性静息心率的数值也不是一直保持不变的，在研究开心和难过的情感时没有成功消除静息心率波动的干扰；使用被试的主观情感感受带入到 SVM 进行监督式学习并检验分类准确度的过程无法发现并剔除被试的错误主观感受和异常心率波动；只凭借心率指标识别用户情感的准确度依然是有限的。

未来研究将收集更多的样本数据，同时采取多种非监督式的分类方法互相比较和验证，提取分类精确度最高的模型；后续的研究将不仅仅局限于心率指标，还将研究其他生理指标、微行为以及绩效，从而达到进一步提高检测器识别准确度的目的。情感的划分也不再是开心难过和中性简单的三类，将更加细分，做到更加全面的分类。

基金项目

安徽省大学生创新创业训练计划项目“基于多模式情绪测量的移动互联网用户情绪检测器研究”(201810363208)，教育部产学合作协同育人项目“基于 ErgoLAB 人机环境同步平台的人因工程课程设计”研究(201901024006)，国家级大学生创新创业训练计划项目“基于 kano 模型和机器学习的智能家居用户体验优化研究”(202010363072)。

致谢

感谢操雅琴老师对论文选题、实验设计、数据分析以及论文写作和修改提供的指导和帮助。感谢所有实验参与者的参与。

参考文献

- [1] Romeo, L., Cavallo, A., Pepa, L., et al. (2019) Multiple Instance Learning for Emotion Recognition using Physiological Signals. *IEEE Transactions on Affective Computing*, 1-18. <https://doi.org/10.1109/TAFFC.2019.2954118>
- [2] 黄晨晨, 索朗拉姆, 拉姆卓嘎, 群诺. 基于 SVM 的藏文微博文本情感分析研究与实现[J]. 高原科学研究, 2020, 4(1): 92-96.

-
- [3] 朱菊霞, 吴小培, 吕钊. 基于 SVM 的语音情感识别[J]. 计算机系统应用, 2011, 20(5): 87-91.
- [4] 马旭霞. SVM 理论及应用[J]. 科学技术创新, 2019(1): 15.
- [5] 杜昭慧, 司玉娟. 基于层次支持向量机的脉搏信号情感识别[J]. 吉林大学学报(信息科学版), 2017, 35(1): 37-42.
- [6] Wang, Y.-P., Kuo, T.B.J. and Yang, C.C.H. (2014) A Possible Explanation for the Effects of Respiration on Heart Rate and Blood Pressure Asymmetry. *International Journal of Cardiology*, **174**, 805-807.
<https://doi.org/10.1016/j.ijcard.2014.04.152>
- [7] 李敏. 基于光电容积脉搏波的抗运动心率及血氧提取算法研究[D]: [硕士学位论文]. 北京: 北京理工大学, 2016.
- [8] 汤明宏. 基于脑电信号的情感识别研究[D]: [硕士学位论文]. 南京: 南京邮电大学, 2018.
- [9] 马耀兰. 基于支持向量机方法的股票预测模型[J]. 鸡西大学学报, 2013, 13(4): 124-125+145.
- [10] 徐光. 基于光电容积脉搏波成像心率检测技术的研究[D]: [硕士学位论文]. 武汉: 武汉理工大学, 2014.