基于情感倾向的互动策略：实现直觉式人机交互

Thorsten Hempel [a], Laslo Dinges [b] 和 Ayoub Al-Hamadi [c] 神经信息技术，奥托·冯·格里克大学电气工程与信息技术学院，德国马格德堡{*thorsten.hempel, laslo.dinges, ayoub.al-hamadi*}*@ovgu.de*

关键词: 人机交互、接近策略、情感估计、情绪检测、预测人类行为、接近人类

摘要: 情绪表达是重要的交际信号，也是人类之间直觉式互动的关键线索。因此，在与人互动时，将这些基本要素纳入机器人行为策略至关重要，这有助于促进相互理解并减少误判。为应对这一挑战，我们通过检测并利用潜在人类互动对象的情绪状态和注意力进行情感分析，从而选择恰当的互动策略。这样一来，我们便能为更直觉式的人机交互铺平道路，因为机器人的行动会符合人的情绪和预期。我们提出了四种不同的互动策略，采用隐式和显式沟通技术，并将其部署在移动机器人平台上开展初步实验。

arXiv:2301.03867v1 [cs.RO] 10 Jan 2023

1 引言

机器人曾被视为乌托邦式的设想，如今正逐渐从工业和实验室环境走向现实世界，以协助人们完成日常任务。在这一过程中，人机交互成为机器人研究的核心关注点，研究者致力于探索执行交互与协作任务时面临的诸多挑战。一般而言，机器人需要掌握必要技能，以便与任意人类互动对象进行直观的互动，而不受具体任务、人类意图和沟通行为的影响。为此，一项基本目标是根据具体情境，预测适当的策略，主动接近或避开人，例如考虑对方的情绪、态度（M¨uller-Abdelrazeq 等，2019年；Elprama 等，2016年）以及对机器人的怨恨（Naneva 等，2020年；Miller 等，2021年）。首先，需要判断人是否有交互意愿（Satake 等，2009年；Finke 等，2005年）。随后，机器人必须找到合适的策略，以便在适当时与人互动并建立联系，或避免发生对抗。如果希望避免对抗，为了做出这些基础决策，机器人必须处理一系列子任务：估计人的注意力焦点（Dini 等，2017），预测其心理状态，并执行参与或回避策略（Avelino 等，2021年）。

尤其是，估计互动意愿是一项极具挑战性的任务，因为这种心理状态往往只会发出模糊的社交信号。作为替代方法，许多方法（Oertel et al., 2020）会检测参与情况，以确定某人是否已经走近机器人并等待机器人的回应。这种状态的表现更加明显，例如语音指令（Foster et al., 2017）、手势和人际距离学；但这会削弱主动式方法，也忽略了人类伙伴需要帮助、却过于不确定或没有意识到机器人存在的情形。

在这项工作中，我们通过引入一种基于情感的参与策略，弥合主动行为不足与高级心理状态预测之间的差距。

通过结合注意力焦点与情绪，我们得出一种情感分析方法，使我们能够选择比当前二元参与和不参与策略更细致的行为模式。我们在移动机器人平台上实现了该方法，并通过显式和隐式

a https://orcid.org/0000-0002-3621-7194 b
https://orcid.org/0000-0003-0364-8067 c
https://orcid.org/0000-0002-3632-2402

具身交互来执行这些策略。我们的模型采用最小规模的网络进行训练，以便在计算资源有限的移动系统上运行该方法。

2 相关工作

近年来，人际距离规则已成为一种常用工具，可用于推断人机交互的不同状态（Mumm 和 Mutlu，2011），并了解如何利用空间区域来改善交互（Syrdal 等，2007）。（Repiso 等，2018年）预测合适的相遇点，以便与人群自然地参与互动。类似地，（Satake 等，2009）会主动接近检测到的人并提供帮助。（Fischer 等，2021年）会以口头方式向人打招呼，并根据与目标人物的距离动态调整音量。遵循（Kendon，1990）的研究，（Carvalho 等，2021年）应用Kendon问候模型，在接近人时追踪互动中的人类心理状态，所用特征包括注视、面部表情和手势。然而，所有这些方法都忽略了人类互动对象是否确实感兴趣并准备好发起互动。这会带来问题，因为如果对方持消极态度，机器人的互动努力可能会被视为粗鲁和烦人（Br¨ohl 等，2019年）。针对这一问题，（Kato 等，2015）提出了初步方案：只有检测到对方表现出注意机器人的迹象时，才会接近并提供帮助。但这一方案仍未纳入人类互动对象对互动本身的实际心理态度。

深度学习领域的最新进展为人类视觉分析开辟了新的可能性，涵盖从一般检测任务到面部微表情估计等方面。这使得我们能够汇总情绪等其他相关信息（Chuah and Yu, 2021年；Spezialetti et al., 2020），从而更好地理解人类互动对象的行为和意图。据我们所知，我们是首个通过基于视觉情绪估计预测情感状态来改进机器人交互策略的研究团队。我们的主要贡献总结如下：

- 我们提出了一种基于图像的情感分析方法，用于评估潜在人类互动对象当前的交互偏好。
- 我们提出了四种不同的行为策略（参与、吸引、忽略、回避），其中包含多种不同的显式和隐式交流沟通模态。



图 1：TIAGo 机器人平台及本方法使用的关键操作。

- 我们设计了超轻量级模型来实现该方法，并将其集成到移动机器人平台上。
- 我们在实验室环境中开展了初步实验，以进行早期评估。

3 方法

本节将介绍我们提出的基于情感的交互行为方法的各个步骤。

情感分析基于两个主要特征：头部姿态和情绪估计。头部姿态估计是衡量人员当前视觉注意焦点的重要指标。情绪估计用于预测当前的情绪状态 o f 此人。综合来看，这使我们能够推断此人*a)* 寻求互动，*b)* 对此犹豫不决，或*c)* 完全不想与机器人参与互动。判断每种情感倾向状态，有助于机器人执行与此人预期反应相对应的特定行为，从而提升互动的直观性。图 2 展示了我们提出系统的全景视野。在下文的小节中，我们将详细介绍系统的各个组件及其相互作用。

3.1 机器人平台

我们使用 PAL Robotics 的 TIAGo 机器人，这是一款用于室内环境的移动服务机器人。它的头部内装有一台 RGB-D相机，可通过偏航和俯仰调整，动态感知环境。为了抓取和移动物体，

机器人配备了一条带有五指夹爪的机械臂。躯干可以升降，以调整机器人的高度；集成的扬声器则可用于播放语音指令。这些能力使我们能够结合显式和隐式沟通模态，并将其用于我们的方法中。

- 头部跟随：机器人头部朝向并跟随移动的人，以此作为表示专注的隐性信号，而不会主动靠近。
- 身体跟随：如果某人超出机器人移动头部的范围，机器人底座便会整体转动，以持续追踪此人。这种行为传达了注意力承诺，因此相比头部跟随，它是更强烈的信号。
- 躯干抬升：抬升躯干是一种相当含蓄的沟通信号，因为人们会将其感知为一种社会惯例，类似于问候他人时起身。
- 语音：通过语音接近人类是一种明确的沟通策略，会更强烈地促使对方作出回应。因此，只有在高度预期被接近者乐于互动时，才应使用语音。

3.2 视觉注意力

衡量人的视觉注意力时，典型指标是注视方向，以及较为粗略的头部姿态。由于注视方向取决于眼睛，而眼睛在面部图像中所占面积很小，因此容易出现误差。因此，我们基于头部姿态预测算法来估计周围人的视觉注意力。首先，我们使用轻量级SSD人脸检测器在图像流中定位人脸。随后，将裁剪出的人脸图像输入6DRepNet（Hempel等，2022年）进一步处理；该模型采用旋转表示，直接回归人脸的偏航角、俯仰角和滚转角。为确保具备实时处理能力，我们将原始6DRepNet骨干网络替换为高效的MobileNetv3-Small（Howard等，2019年）。新的头部姿态预测模型能够保持原始模型90%的准确率，同时参数数量缩减至原来的十分之一。

3.3 情绪估计

为了评估人们对机器人的态度，以及厌恶或合作意愿等细分的主观状态，可以将基本情绪作为指示性特征。这些情绪通常通过面部表情来传达。因此，自动面部表情识别是为人机协作和注意力预测生成特征的一种重要方法。

以往的面部表情识别方法遵循既定流程：人脸检测、关键点提取和动作单元（AU）识别。随后，使用这些动作单元作为特征向量对情绪进行分类（Wegrzyn et al., 2017; Vinkemeier et al., 2018; Werner et al., 2017）。然而，使用单一但更全面、泛化能力更强的数据库进行端到端学习，往往优于这类传统方法。例如，AffectNet数据库（Mollahosseini et al., 2017）包含约 50 万个野外样本，涵盖中性以及快乐、厌恶、恐惧、惊讶、愤怒和悲伤等基本情绪。此外，它是唯一一个还包含效价与唤醒度（VA）真实标注的数据库。与基本情绪类别不同，效价与唤醒度是连续标签 $\in [-1, 1]$，可用于回归任务。效价与唤醒度对人类而言不如基本情绪那样直观易懂，但能区分程度各异的面部表情。这对于捕捉情绪表达细微但长期的变化尤其有用。此外，多任务网络（同时进行分类和回归训练）也提高了分类准确率。效价还将多种情绪的重要信息概括为一个参数（负面、中性、正面），这可能有助于针对某些场景制定启发式行为规则。

在我们之前的工作中，我们提出了一种基于 YOLOv3 的多任务网络（Dinges 等，2021）。在本研究中，我们比较了 EfficientNetV2-S、ResNet18 和 MobileNetV3-Large。尽管 MobileNet 的训练损失收敛表现较差，但它针对硬件资源有限的情况进行了优化，在测试集上的准确率几乎与 EfficientNet 相同（偏差 $< 1\%$）。此外，我们使用加权随机采样器处理类别不平衡数据集，并在训练集和测试集上重新训练网络，从而生成一个对未见数据具有更强鲁棒性的应用模型。

3.4 情感分析与处理策略

检测情绪表达并结合视觉注意焦点，有助于感知重要的交际信号。就像人与人互动时一样，陌生人对你微笑，你可能更愿意向对方打招呼；但如果对方怒目而视，你可能会避免目光接触。我们将这些原则迁移到.....

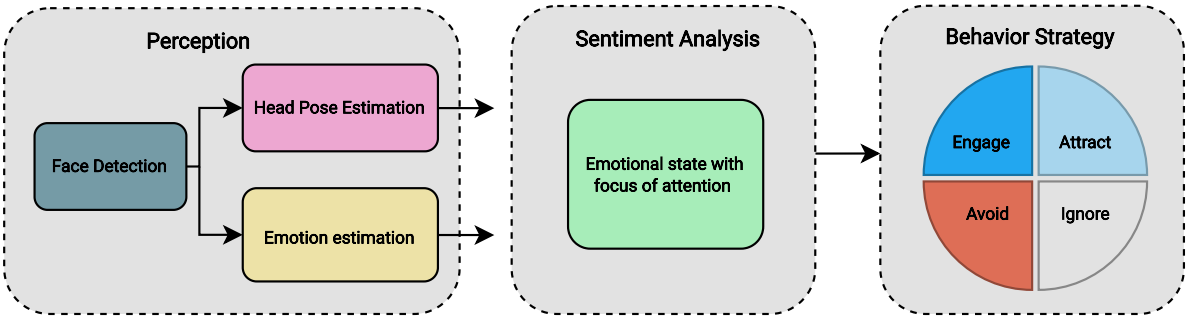

图2：所提系统的方法全景视野。检测到的人脸用于估计头部姿态，并根据面部表情判断相应的情绪。这两项信息用于确定情绪状态，以及与此人注意力相对应的视觉焦点；视觉焦点可反映其情感倾向，并有助于制定精细化的行为策略。

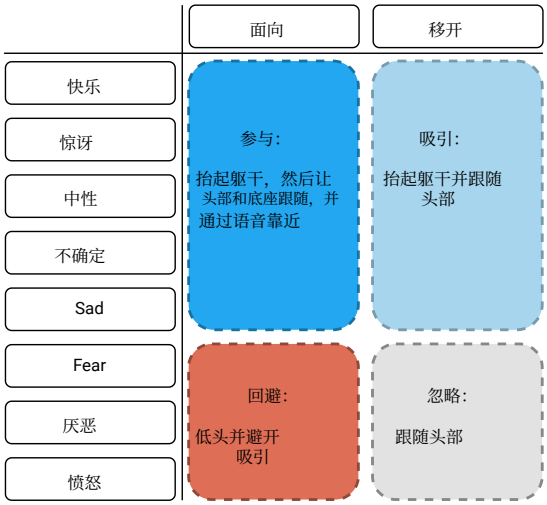

图3：基于情绪和注意力状态的行为策略。

.....根据不同情绪与视觉注意力的组合来判断情感倾向状态。如果一个人表现出恐惧或愤怒的视觉迹象，同时注意着机器人，这表明他对机器人持负面态度。在这种情况下，机器人采取被动行为，避免剧烈动作，有助于缓和局势、安抚这个人。如果一个人感到恐惧，却没有注意机器人，那么恐惧的原因可能不是机器人。因此，机器人采取更主动的行为来表明自己有所察觉，或许是更合适的策略。由此，我们的情感分析能够将常见的非黑即白的互动策略细分为柔性行为。我们不再只是在参与和不参与之间做选择，而是选择四种不同的行为：参与、吸引、回避、忽略。

- 参与：参与后，机器人的躯干会伸展，以示其察觉到人的存在。此外，机器人会通过语音指令向人问候，主动发起互动。当注意力集中在机器人身上，且态度明显积极时，这种策略称为.....
- 吸引：积极的情绪态度但未将注意力投向机器人的人，会引发较为被动的行为。其目标是示意愿意提供帮助并进行互动，但不会主动投入互动。具体做法是伸展躯干以表明已察觉对方，并面向对方，但不主动打招呼。这样便能含蓄地表达提供支持的意愿（尤其是在对方感到不确定或悲伤时），对方可轻松接受或忽略。
- 回避：当机器人面对强烈的负面情绪（恐惧、厌恶、愤怒）时，会采取回避策略。在这种情况下，机器人会避免目光接触（让面部避开摄像头中心），并采取被动姿态。尤其是在对方感到恐惧时，目标是通过避免意外动作来增强其安全感。
- 忽略：如果负面情绪并非针对机器人，机器人通常会忽略此人，只用头部跟随此人，以捕捉其情绪或注意状态的变化。

图3展示了情绪和注意状态的所有组合及相应的行为策略。

4 实现

我们使用 ROS[1] 中间件，在机器人平台上实现了所提出的解决方案，这也简化了与其他基于 ROS 的

[1]https://www.ros.org/

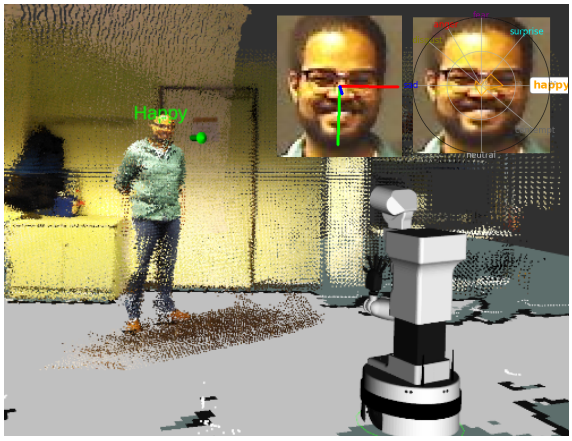

图4：实验室环境中的情感分析。系统使用人脸检测器跟踪人物，随后进行头部姿态估计和情绪估计。当人物表现出专注和快乐时，预计其对机器人的情感倾向为正面，并选择主动发起策略。

机器人平台的集成。人脸检测、头部姿态和情绪估计分别在独立节点中执行，这些节点会转发进行情感分析和动作控制所需的信息。后者在机器人本体上执行，而神经网络的推理则交由安装在机器人肩部的独立笔记本电脑完成。这样既能确保机器人完全移动自如，又能利用额外的计算资源。该笔记本电脑配备了 Quadro RTX 4000，可并行运行所有模型，其推理时间如下：

- 人脸检测：6.7ms
- 头部姿态估计：1.4 ms
- 情绪估计：6.3 ms

因此，视觉感知能够实时进行，并能够持续跟踪周围环境中的人。

5 实验

我们在实验室环境中开展了一项初步实验，在 TIAGo 机器人平台上测试所实现的方法。实验旨在了解该模型在实际应用场景中的表现。图4展示了我们一项测试的快照。机器人平台检测到环境中的一张人脸，并预测其头部姿态和情绪状态。注意焦点在机器人身上（由绿色箭头指示），同时检测到快乐情绪，因此机器人按照参与策略发起互动。

尽管这些模型能够提供非常鲁棒的预测，但目前的实现完全基于单帧检测。然而，在人脸跟踪过程中，面部表情会不断变化，短暂的注意力分散也会使头部姿态偶尔发生振荡。这会导致状态在短时间内突然变化。因此，为了提高情感分析的鲁棒性，对当前状态的判断应结合多个时间帧的预测结果。

6 结论

本文提出了一种基于情感倾向的方法，以提升人机交互的直观性和推理能力。我们的方法侧重于对心理状态和注意焦点进行视觉感知，从而开展情感分析，为选择合适的互动发起策略提供依据。我们也考虑了人们完全不愿参与互动的情况，而近期文献大多忽略了这一点。为实现该方法，我们训练了轻量级感知模型，并将其集成到移动机器人平台上。最后，我们使用移动机器人平台开展了初步实验，以分析模型及整个系统的性能。

未来，我们将把情感分析纳入时序概率框架，以提升其在真实场景中的鲁棒性。我们还将纳入效价与唤醒度的预测结果，用于评估情绪的强度以及情绪状态之间的转换。进一步的改进包括加入机器人的移动能力，使其能够主动与人互动或结束互动；也可以增加其他互动方式，例如按指令跟随人、避让经过的人，或前往能够更好地纵览周围环境的位置。

最后，我们将开展一项测试对象研究，从直观性和亲和力方面评估我们的方法。

致谢

本研究由德国联邦教育与研究部（BMBF）资助并支持（AutoKoWAT-3DMAt项目，资助编号 13N16336），并由德国研究联合会（DFG）通过资助编号Al 638/13-1、Al 638/14-1和Al 638/15-1 提供支持。

参考文献

Avelino, J., Garcia-Marques, L., Ventura, R., and Bernardino, A.（2021）。打破隔阂：关于人机初次相遇中社会感知参与的综述。《国际社会机器人学杂志》，13:1851–1877。Br¨ohl, C., Nelles, J., Brandl, C., Mertens, A., and Nitsch, V.（2019）。人机协作接受度模型：在德国、日本、中国和美国的开发与比较。《国际社会机器人学杂志》，11:709–726。Carvalho, M., Avelino, J., Bernardino, A., Ventura, R. M. M., and Moreno, P.（2021）。人类问候机器人：追踪人类问候心理状态并据此采取行动。*2021*年第*12*届图像与信号处理及分析国际研讨会，第1935–1941页。Chuah, S. H.-W. and Yu, J.（2021）。服务的未来：情绪在人机交互中的力量。《零售与消费者服务杂志》，61:102551。Dinges, L., Al-Ha madi, A., Hempel, T., and Al Aghbari, Z.（2021）。使用面部动作识别评估高难度人机协作场景中的用户感知。载于*2021*年第*12*届图像与信号处理及分析国际研讨会，第195–199页。Dini, A., Murko, C., Yahyanejad, S., Augsd¨orfer, U., Hofbaur, M., and Paletta, L.（2017）。基于概率注意力框架测量和预测人机交互中的情境意识。载于*2017*年*IEEE/RSJ*国际智能机器人与系统会议，第4354–4361页。Elprama, S., El Makrini, I., Vanderborght, B., and Jacobs, A.（2016）。工厂工人对协作机器人的接受度：关于拟人化机器人社会线索作用的试点研究。Finke, M., Koay, K. L., Dautenhahn, K., Nehaniv, C., Walters, M., and Saunders, J.（2005）。嘿，我在这儿——机器人怎样才能吸引人们的注意力？载于*RO-MAN 2005*。*IEEE*机器人与人类交互通信国际研讨会，*2005*年。，第7–12页。Fischer, K., Naik, L., Langedijk, R. M., Baumann, T., Jel´ınek, M., and Palinko, O.（2021）。使用渐进式语音适应发起人机交互。载于*2021*年*ACM/IEEE*人机交互国际会议伴随会议，HRI '21伴随会议，第421–425页，纽约州纽约，美国。计算机协会。Foster, M. E., Gaschler, A., and Giuliani, M.（2017）。自动分类动态多方人机交互中的用户参与度。《国际社会机器人学杂志》，9。Hempel, T., Abdelrahman, A. A., and Al-Hamadi, A.（2022）。用于无约束头部姿态估计的六维旋转表示。载于*2022*年*IEEE*国际图像处理会议，第2496–2500页。Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., and Adam, H.（2019）。搜索 MobileNetV3。载于*IEEE/CVF*国际计算机视觉会议（*ICCV*）论文集。Kato, Y., Kanda, T., and Is higuro, H.（2015）。我能帮您吗？——类人礼貌接近行为的设计。载于*2015*年第十届*ACM/IEEE*人机交互国际会议（*HRI*），第35–42页。Kendon, A.（1990）。互动的开展：聚焦交往中的行为模式。剑桥大学出版社，剑桥，英国。Miller, L., Kraus, J., Babel, F., and Baumann, M.（2021）。不止于感受——人机交互中信任层的相互关系，以及用户倾向和状态焦虑的作用。《心理学前沿》，12。Mollahosseini, A., Hasani, B., and Mahoor, M. H.（2017）。AffectNet：用于自然环境下面部表情、效价和唤醒度计算的数据库。《IEEE情感计算汇刊》，10：18–31。M¨uller-Abdelrazeq, S. L., Sch¨onefeld, K., Haberstroh, M., and Hees, F.（2019）。与协作机器人互动——工业环境中的态度与接受度研究，第101–117页。Springer International Publishing，查姆。Mumm, J. and Mutlu, B.（2011）。人机空间距离：人机交互中的物理与心理距离。载于*2011*年第六届*ACM/IEEE*人机交互国际会议（*HRI*），第331–338页。Naneva, S., Gou, M. S., Webb, T. L., and Prescott, T. J.（2020）。关于人们对社交机器人的态度、焦虑、接受度和信任的系统综述。国际社交机器人杂志，第1–23页。Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., and Peters, C.（2020）。人机交互中的参与：全景视野。《机器人学与人工智能前沿》，7。Repiso, E., Garrell, A., and Sanfeliu, A.（2018）。在人机伴侣框架中接近并与人互动的机器人。载于*2018*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*），第8200–8205页。Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., and Hagita, N.（2009）。如何接近人类？——社交机器人发起互动的策略。载于*2009*年第四届*ACM/IEEE*人机交互国际会议（*HRI*），第109–116页。Spezialetti, M., Placidi, G., and Rossi, S.（2020）。人机交互中的情绪识别：最新进展与未来展望。《机器人学与人工智能前沿》，7。Syrdal, D. S., Lee Koay, K., Walters, M. L., and Dauten-hahn, K.（2007）。个性化机器人伴侣？——个体差异对人机交互场景中空间偏好的影响。载于*RO-MAN 2007*——第*16*届*IEEE*机器人与人类交互通信国际研讨会，第1143–1148页。

Vinkemeier, D., Valstar, M. F., and Gratch, J.（2018）。使用动作单元检测器和决策树预测扑克中的弃牌。载于*FG*，第504–511页。Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., and Kissler, J.（2017）。绘制情绪面孔图谱：各个面部部位如何共同促进情绪识别成功。《PLOS ONE》，12(5)：e0177239。Werner, P., Handrich, S., and Al-Hamadi, A.（2017）。使用随机回归森林估计面部动作单元强度并可视化特征相关性。载于*2017*年第七届情感计算与智能交互国际会议（*ACII*），第2018年1月卷，第401–406页。IEEE。