

心理治疗中治疗工作联盟的深度标注

Baihan Lin¹, Guillermo Cecchi², Djallel Bouneffouf²

¹Columbia University

²IBM Thomas J. Watson Research Center

baihan.lin@columbia.edu, gcecchi@us.ibm.com, djallel.bouneffouf@ibm.com

Abstract

治疗联盟是心理治疗结果的重要预测因素。在实践中，工作联盟是从患者和治疗师填写的一系列评分问卷中估算出来的。在这项工作中，我们提出了一种分析框架，可以直接从心理治疗会话中的自然语言推断出治疗性工作联盟，并使用如 Doc2Vec 和 SentenceBERT 等深度嵌入模型以对话轮次级别进行解析。每次心理治疗会话的记录可以实时转录并生成，这些嵌入式对话与工作联盟清单中陈述的分布式表示形式进行比较。我们在一个包含超过 950 个焦虑、抑郁、精神分裂症和自杀倾向患者的心理治疗疗程的真实世界数据集中展示了这种方法在绘制患者-治疗师一致性轨迹的有效性及其能够为临床精神病学提供见解的解释能力。我们相信这样一个框架可以及时向治疗师反馈访谈过程中对话的质量。

索引项: 计算语言学, 计算精神病学, 自然语言处理

1. 介绍

心理治疗中的一个基本概念是治疗师与患者或更广泛地说，寻求帮助的客户之间的作业联盟 [1]。这种联盟涉及这两个主体之间关系的认知和情感成分，包括对需要实现的目标和执行的任务达成一致，以及在整个治疗过程中建立的信任、尊重等联系。定性方法量化治疗结果导致了这样一个结论：联盟的力量是预测成功的主要因素之一 [2]。量化联盟的操作方法依赖于患者和治疗师对整个会话的评估报告，通常仅限于量表估值 [3]。这种方法没有利用自然语言提供的细微差别，耗时且难以系统地在研究之外进行追踪；更不用说在整个会话过程中评估每个对话轮次了。在这里，我们提出了一种通过将治疗过程中的每一个轮次投射到临床建立的作业联盟清单表示上来量化患者与治疗师之间的联盟程度的方法，并使用语言模型对这些轮次和清单进行编码。这不仅使我们能够量化总体上的联盟程度，还能识别其在较短或较长的时间尺度上动态变化的细微模式。我们还讨论了我们的方法如何作为辅助工具提供反馈给治疗师并增加培训治疗师的学习机会。

2. 问题设定

2.1. 工作联盟分析

Algorithm 1 工作联盟分析 (WAA)

- 1: 对于 $i = 1, 2, \dots, T$ 做
- 2: 自动转录对话轮次对 (S_i^p, S_i^t)
- 3: 对于 $(I_j^p, I_j^t) \in \text{库存}(I^p, I^t)$ 做
- 4: 得分 $W_j^{p_i} = \text{相似度}(Emb(I_j^p), Emb(S_i^p))$
- 5: 得分 $W_j^{t_i} = \text{相似度}(Emb(I_j^t), Emb(S_i^t))$
- 6: 结束循环
- 7: 结束循环

上图是分析框架的概述。我们采用一位患者或同一病

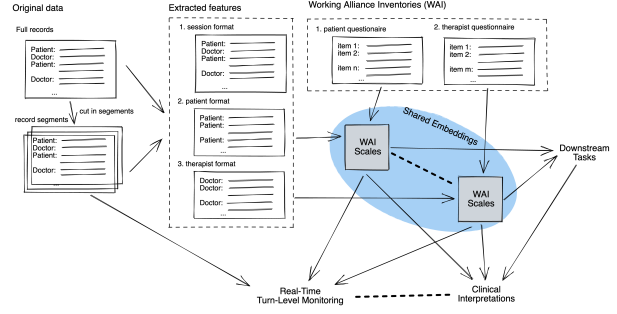


图 1: 工作联盟分析的理论框架

PATIENT : I don't know. I was laughing at myself for thinking it.
THERAPIST : Yeah.
PATIENT : Um, like I don't know, was I being a martyr? I don't know.
THERAPIST : No it sort of sounds like you feel like you're doing me a favor or something?
PATIENT : Um, I don't know.
THERAPIST : Like he's getting more out of this than I am. [00 : 03 : 07]
PATIENT : No, I don't think that. Naw, I'm just cranky.
THERAPIST : Yeah I'm hearing that, but I'm trying to explain what the, what that's about. I mean you ought to be glad I'm here at all.

图 2: 来自心理治疗记录的示例对话

情患者群体的完整记录。我们在特征提取前要么原样使用这些记录，要么根据时间戳或话题转变将它们截断成段落。如您所见，原始格式为对话对形式。我们可以以三种方式提取特征：第一种是使用完整的对话对；第二种只提取患者的发言；第三种只提取医生的发言。这三种特征格式各有优缺点。对话格式包含了所有信息，但句子中的意图来自两个个体，因此它们可能会混合在一起。患者格式包含患者完整的叙述，通常更连贯，但它只是故事的一部分。治疗师格式（计算精神病学领域的人也认为这是某种对患者感受的语义标签）可以提供信息，但也可能过于简单化。

当我们拥有特征时，我们将工作联盟清单与嵌入进行比较。算法 1 描述了该过程。在会话期间，患者和治疗师之间的对话被转录成一对轮次（如图 2 中的示例）。我们将每个患者的回应轮次标记为 S_i^p ，紧随其后的是治疗师的回应轮次 S_i^t 。它们被视为一个对话对。工作联盟问卷清单也成对出现： I^p 对应患者（或客户）， I^t 对应治疗师。每份清单包含 36 项陈述。我们将对话轮次和清单嵌入到深度句子或段落嵌入中，然后计算每个轮次的嵌入向量与其对应的清单向量之间的余弦相似度。通过这种方式，对于每一个轮次（无论是患者还是治疗师），我们获得了一个 36 维的工作联盟分数。我们将在第 3.1 节描述我们推断的工作联盟分数的具体量表，这将引入可解释的信息到我们的框架中。

这里有一些可以嵌入我们分析框架的下游任务和用户场景。我们可以使用这些提取的加权主题来判断治疗是否朝着正确的方向进行，患者是否进入某种不良的精神状态，或者治疗师是否应该调整其治疗策略。这可以构建为一个智能 AI 助手，提醒治疗师注意这些问题。

I felt uncomfortable with __.
 __ and I agreed about the things I will need to do in therapy to help improve my situation.
 I was worried about the outcome of the sessions.
 What I was doing in therapy gave me new ways of looking at my problem.
 __ and I understood each other.
 __ perceived accurately what my goals were.
 I find what I was doing in therapy confusing.
 I believe __ liked me.
 I wish __ and I could have clarified the purpose of our sessions.
 I disagreed with __ about what I ought to get out of therapy.
 I believe the time __ and I were spending together was not spent efficiently.
 __ did not understand what I was trying to accomplish in therapy.
 I was clear on what my responsibilities were in therapy.
 The goals of the sessions were important for me.
 I find what __ and I were doing in therapy was unrelated to my concerns.
 I feel that the things I did in therapy helped me to accomplish the changes that I wanted.

图 3: 工作联盟清单中的示例行陈述

TASK scale:	2,	4,	7,	11,	13,	15,	16,	18,	24,	31,	33,	35
Polarity	+	+	-	-	+	-	+	+	+	-	-	+
BOND scale:	1,	5,	8,	17,	19,	20,	21,	23,	26,	28,	29,	36
Polarity	-	+	+	+	+	-	+	+	+	+	-	+
GOAL scale:	3,	6,	9,	10,	12,	14,	22,	25,	27,	30,	32,	34
Polarity	-	+	-	-	-	+	+	+	-	+	+	-

图 4: 工作联盟量表的三个维度的关键点

2.2. 心理治疗转录数据集

Alex Street 咨询和心理治疗记录数据集¹ 包含超过 950 次匿名治疗师和患者之间真实疗法会话的转录记录。这一多部分集合包括从实际治疗会话中录制的语音翻译文本、4 万页客户叙述以及 2.5 万页参考作品。这些会话属于四种类型的精神状况：焦虑症、抑郁症、精神分裂症和自杀倾向。每位患者的回应轮次 S_i^p 之后跟着一位治疗师的回应轮次 S_i^t ，被视为一对对话。总计，这些材料包括患者和治疗师超过 20 万轮次的内容，并为我们的心理疗法过程的语言分析提供了最广泛范围客户的访问权限。

3. 方法

3.1. 工作联盟库存

工作联盟清单 (WAI) 是一组自我报告的测量问卷，量化了治疗关系、任务协议和目标协议 [3, 4, 5]。自最初的 12 项版本 [4] 以来，该清单为客户和治疗师使用平行版本，并具有良好的心理测量特性，帮助建立了治疗联盟在预测治疗结果中的重要性。现代版本的清单包含 36 个问题 (图 3)，参与者被要求对每个项目进行 7 分制评分 (1=从不, 7=总是) [5]。WAI 旨在 (1) 衡量所有类型治疗中的联盟因素，(2) 记录联盟测量与测量背后相应理论构建之间的关系，以及 (3) 将联盟测量与统一的治疗改变理论相关联 [6]。

操作上，目标是从这 36 个项目中得出三个联盟量表：任务量表、纽带量表和目标量表。它们衡量心理治疗结果的三大主题：(1) 患者-治疗师关系的合作性质；(2) 治疗师与患者之间的情感纽带；以及 (3) 治疗师和患者的同意能力，以确定相关的短期任务和长期目标。对应三个量表的得分来自一个关键表格 (图 4)，该表格指定了在最终求和时应用到问卷答案上的正性或符号权重。总评分只是这三个量表分数的简单相加。关键表格就像一个加权矩阵，规定了这些量表的方向性。

3.2. 句子嵌入

原则上，任何句子或段落嵌入都可以帮助我们表征对话回合和清单。在这项工作中，我们使用了两种深度嵌入。Doc2Vec 嵌入 [7] 是一种流行的无监督学习模型，它学习

¹<https://alexanderstreet.com/products/咨询与心理治疗记录系列>

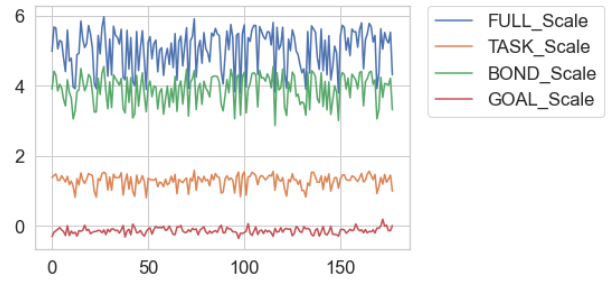


图 5: 工作联盟分数的示例轨迹

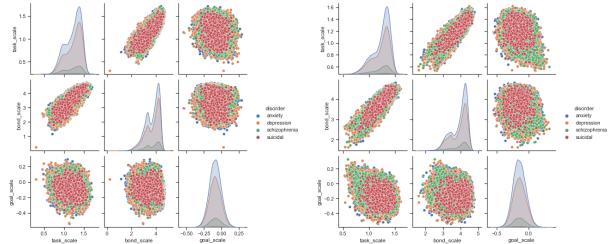


图 6: 工作联盟评分量表的关系图 (左: 患者版本; 右: 治疗师版本)

句子和文本文档的向量表示。它通过利用分布式内存改进传统的词袋表示法，记住当前上下文中缺少的内容。另一种评估的嵌入是 SentenceBERT [8]，它通过使用孪生网络和三元组网络结构修改预训练的 BERT 网络来推断语义有意义的句子嵌入。使用这两种深度嵌入，我们将回合级别的条目 (无论是转录中的对话回合还是工作联盟清单中的陈述项目) 嵌入到 300 或 384 维的向量中。然后计算特定回合向量与清单条目之间的余弦相似度。²

4. 结果

图 5 是一个焦虑心理治疗会话的时间序列示例。我们看到联盟分数在各个量表上有所不同。如果我们研究这些量表之间的关系，我们会观察到任务量表与纽带量表在两个版本中都呈正相关，而目标量表在治疗师版本中略微与任务量表呈负相关 (图 6)。

我们调查了患者与治疗师对联盟估计的一致性。总体而言，相较于患者的估计，我们观察到治疗师倾向于高估工作联盟。更具体地说，治疗师高估了任务和联系量表，但低估了目标量表。这些差异在统计上均具有显著性 ($p < 0.001$)。

在这些障碍之间，焦虑和抑郁之间的联盟分数以及焦虑与精神分裂症之间的联盟分数，在治疗师版本和患者版本中均有显著差异 ($p < 0.001$)。如图 8 所示，可以根据所有四个量表的工作联盟分数显著检测到自杀倾向。

我们还对联盟分数作为时间序列进行了四因素方差分析。图 9 展示了不同精神疾病条件下治疗联盟动态的差异。我们观察到它们在障碍和量表上都存在变化，并且在时间维度 (每个子图中的 x 轴) 上似乎有一定的趋势。这进一步得到了线性回归分析 (图 10) 的支持，即焦虑和抑郁患者有上升的联盟评价，而他们的治疗师往往持不同看法，自杀患者的治疗师倾向于比其患者拥有更高的联盟评价。

我们还可以绘制它们在三大尺度 (任务、纽带和目标) 的联盟空间中的轨迹。如图 11 所示，我们绘制了不同精神

² 鉴于空间限制，Doc2Vec 的结果将在后面展示，而 SentenceBERT 的结果将作为补充材料包含在扩展的在线版本中。

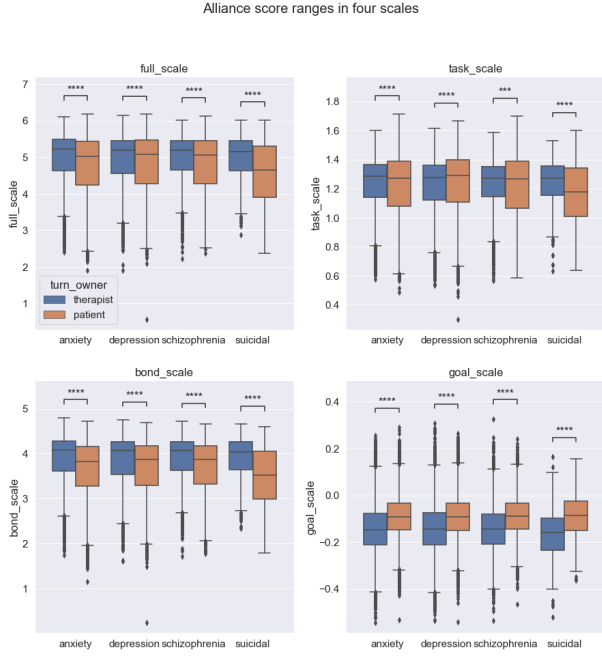


图 7: 工作联盟分数的箱形图

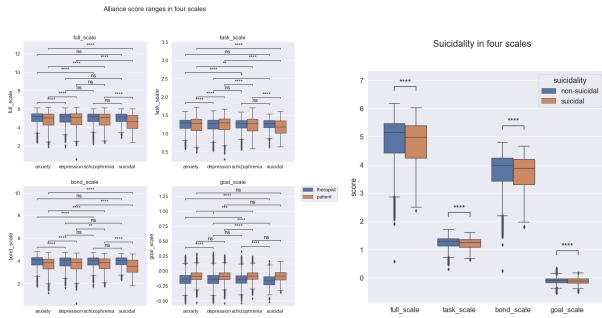


图 8: 各类疾病中的联盟得分

状况的平均轨迹，并注意到自杀倾向的轨迹在纽带和任务尺度上分布得更广（这与方差分析图表中的发现一致）。基于方向性，自杀倾向轨迹显示出显著的发散趋势。这是潜在的回合级分辨率时间分析的工作联盟的第一步。从某种意义上说，我们可以推广这种方法，即作为治疗师可以回顾你的疗程并事后分析其动态。

给定这些时间序列，我们可以使用诸如 t-SNE 等降维技术对其进行可视化。因为心理治疗会话的长度不同，我们计算了 36 维度联盟得分的会话轨迹之间的动态时间包裹距离，然后利用这个成对距离矩阵进行 t-SNE 无监督学习。图 12 展示了治疗师联盟轨迹空间与患者联盟空间流形之间的差异。我们注意到患者的轨迹有两大簇联盟，而治疗师只有一个。这与我们在关系图 6 中观察到的情况一致，即任务和联系量表中的患者联盟得分遵循双峰分布。

我们可以进一步通过将所有时间点平均到四个量表中来汇总联盟分数。为了在一个图中绘制各个量表之间的关系，我们将每个量表归一化为标准正态分布，并展示了患者和治疗师的量表特征雷达图（图 13）。我们观察到自杀倾向的患者相对来说最为不平衡，在目标量表上数值较大而在其他所有量表上较小。而治疗师版本则相反，这与在 3d 轨迹中所作的观察相一致。

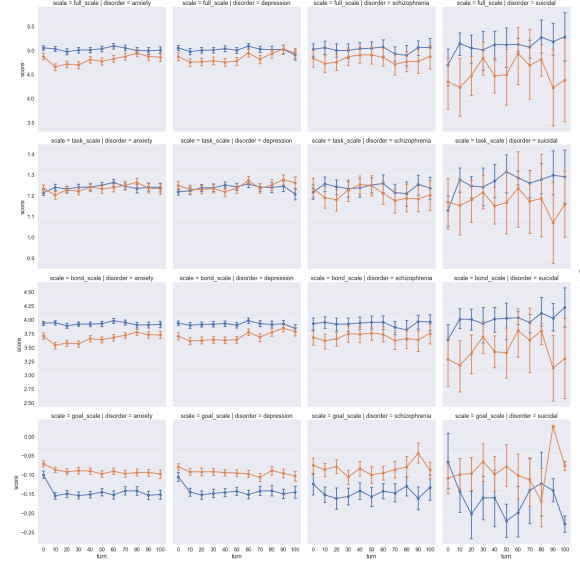


图 9: 四因素方差分析的联盟动态

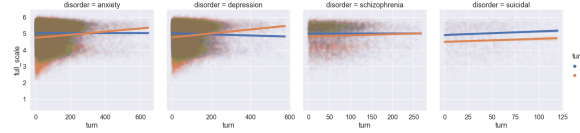


图 10: 工作联盟分数的时间进程回归分析

5. 讨论

我们的分析方法揭示了治疗关系的几个深刻特征。我们观察到患者和治疗师之间推断出的联盟得分存在系统性差异，而且这些差异也跨越不同的疾病。然而，在疗程中的推断分数变化提供了更有趣的观点。特别是，所有条件下患者的评分与治疗师之间的系统性不匹配，这种现象在自杀倾向中尤为显著，这可以在均值以及全量表和子量表的时间轨迹中观察到。相比之下，焦虑和抑郁呈现出随着疗程进展，全量表和联系量表向收敛发展的明确趋势，而在任务和目标量表、精神分裂症或自杀倾向中则没有这种现象。这些治疗对话的特点可以映射到精神病学中通常称为对齐的概念，并在多种神经精神疾病中发挥重要的症状性和诊断作用，例如与精神分裂症的“心智理论”假说 [9] 相关。通过分析过去的疗程，甚至实时会话，受过训练的治疗师可能能够识别出导致突破的关键部分，将他们的专业知识与进一步的因果/预测性分析模型相结合，而实习生则可以通过阅读或观看由专家进行的附注版本的会话来提高自己的直觉。不言而喻，结合生成语言模型和进一步的统计优化，有可能设计有限的聊天机器人以参与患者分类和紧急响应 [10]。

6. 结论

我们提出了一种方法，将最先进的语言建模与心理治疗中的知识和实践经验相结合，这些经验和知识体现在疗法评估量表中，以提供一种独特细致的表示形式，用于描述患者和治疗师互动演变的过程。在这里，我们特别关注工作联盟量表，但我们的方法是通用的，可以扩展到更广泛的评估工具范围。最后，可以通过提供标点评分手册作为推理锚点来改进和完善基于语言的工作联盟估计。下一步包括预测这些推理锚点作为状态（如 [11, 12]）以及根据这些状态训练聊天机器人作为强化学习代理（如 [13, 14, 15]）。

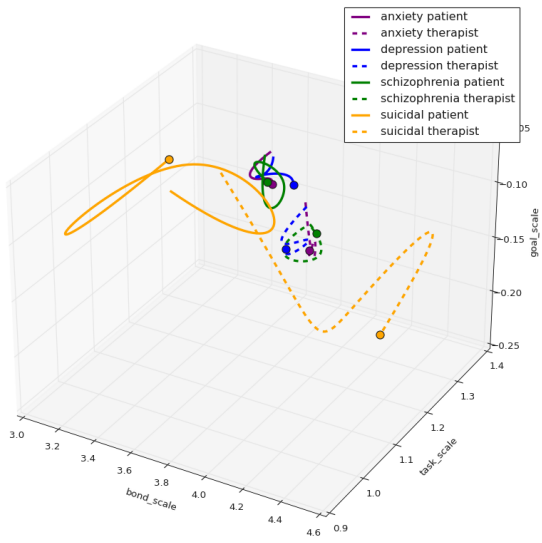


图 11: 不同类精神疾病条件在联盟空间中的平均 3d 轨迹 (点表示轨迹的终点)

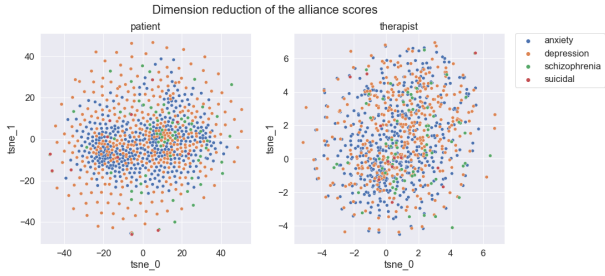


图 12: 联盟轨迹的降维

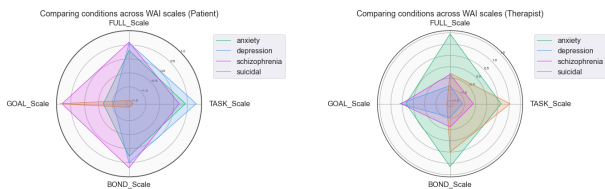


图 13: 工作联盟分数的雷达图

7. References

- [1] E. S. Bordin, “The generalizability of the psychoanalytic concept of the working alliance.” *Psychotherapy: Theory, research & practice*, vol. 16, no. 3, p. 252, 1979.
- [2] B. E. Wampold, “How important are the common factors in psychotherapy? an update,” *World Psychiatry*, vol. 14, no. 3, pp. 270–277, 2015.
- [3] A. O. Horvath, “An exploratory study of the working alliance: Its measurement and relationship to therapy outcome,” Ph.D. dissertation, University of British Columbia, 1981.
- [4] T. J. Tracey and A. M. Kokotovic, “Factor structure of the working alliance inventory.” *Psychological Assessment: A journal of consulting and clinical psychology*, vol. 1, no. 3, p. 207, 1989.
- [5] D. J. Martin, J. P. Garske, and M. K. Davis, “Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review.” *Journal of consulting and clinical psychology*, vol. 68, no. 3, p. 438, 2000.
- [6] A. O. Horvath and L. S. Greenberg, *The working alliance: Theory, research, and practice*. John Wiley & Sons, 1994, vol. 173.
- [7] Q. Le and T. Mikolov, “Distributed representations of sentences and documents,” in *International conference on machine learning*. PMLR, 2014, pp. 1188–1196.
- [8] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using siamese bert-networks,” *arXiv preprint arXiv:1908.10084*, 2019.
- [9] J. N. De Boer, S. G. Brederoo, A. E. Voppel, and I. E. Sommer, “Anomalies in language as a biomarker for schizophrenia,” *Current opinion in psychiatry*, vol. 33, no. 3, pp. 212–218, 2020.
- [10] S. Garg, I. Rish, G. Cecchi, P. Goyal, S. Ghazarian, S. Gao, G. Ver Steeg, and A. Galstyan, “Modeling dialogues with hashcode representations: A nonparametric approach,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 3970–3979.
- [11] B. Lin, D. Bouneffouf, G. Cecchi, and R. Tejwani, “Neural topic modeling of psychotherapy sessions,” *arXiv preprint*, 2022.
- [12] B. Lin, D. Bouneffouf, and G. Cecchi, “Predicting human decision making in psychological tasks with recurrent neural networks,” *arXiv preprint arXiv:2010.11413*, 2020.
- [13] B. Lin, G. Cecchi, D. Bouneffouf, J. Reinen, and I. Rish, “A story of two streams: Reinforcement learning models from human behavior and neuropsychiatry,” in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, 2020, pp. 744–752.
- [14] —, “Models of human behavioral agents in bandits, contextual bandits and rl,” in *International Workshop on Human Brain and Artificial Intelligence*. Springer, 2021, pp. 14–33.
- [15] —, “Unified models of human behavioral agents in bandits, contextual bandits and rl,” *arXiv preprint arXiv:2005.04544*, 2020.