

附录 3:

工 作 坊 简 介

眼动追踪在语言认知与测试领域的运用

主持人：王穗苹

语言是人类一项高级的认知功能，通过语言的使用，人类能够快速而准确地吸收外部信息，建构个人独特的知识背景，发展人类的潜能；与此同时，语言加工的认知基础又十分复杂，因而也吸引了心理学、语言学、神经科学、计算机科学等许多领域研究者的关注。与其他技术相比，眼动行为的测量为探索关键的语言认知机制及语言技能的发展提供了一种相对自然而又高度实时的研究方法。本次报告首先从认知加工的视角介绍语言认知关注的研究主题，随后将结合研究组实际工作的经验，从研究范式和推论的内在逻辑等两个方面介绍如何利用眼动技术开展语言认知和能力发展的研究，以期促进学员深入地思考语言认知与测试研究中理论与数据、方法和问题之间的关系等问题。

英语多文本阅读能力评估与智能应用研究

主持人：徐曼菲

本次工作坊以多文本阅读能力评估为切入点，探讨如何使用语料库与 ChatGPT 开发多文本阅读评估任务。工作坊分三部分：第一部分讲解多文本阅读（整合）的理论基础与构念框架，分享多文本阅读评估任务的特征属性与开发步骤；第二部分讲解多文本阅读评估的实施过程与操作方法，探讨如何基于语料库选取关联文本并设计多文本阅读任务；第三部分讲解如何使用 ChatGPT

生成多文本阅读评估任务，并在此基础上探讨人机协同式任务开发的可行性。每部分讲解均由“任务导入”“核心概念”与“解决方案”三个环节组成。

人工智能与认知诊断评估：理论与实践

主持人：陈锦 林江豪

随着人工智能技术的快速发展，智能化认知诊断评估成为重要的教育评价方法。深度学习具有较强的特征学习能力，模型构建适应性灵活等特点，对提高诊断评估的效率与精度具有重要意义。本次工作坊围绕深度学习方法，通过理论探讨与项目案例实际操作结合的方式，预期达到对智能化认知诊断全过程有一定的理论掌握和实操能力。

具体目标如下：（1）梳理人工智能技术在认知诊断领域的前沿研究，以期掌握未来的研究动态。首先，介绍基于深度学习的诊断模型及其优势——可面向课堂教学小样本、可融合多特征(如题目语义、多模态作答信息)构建模型、可提升诊断模型的精度等。其次，介绍 Q 矩阵自动标注，探讨如何利用深度学习解决人工标注高耗低效、一致性较低的困境。最后，介绍宏观因素与微技能的交互作用，旨在突破传统概率模型仅考虑微技能层面诊断的局限，探索如何利用深度学习将影响因素纳入诊断模型，从而提高模型的解释力，同时验证和丰富现有的理论框架。（2）学会 python 的环境搭建、使用和基本编码规则，以及如何用开放工具包去构建模型。在此基础上，进一步利用传统的 BP 神经网络和基于 LSTM、Bi-GRU 等深度神经网络构建诊断模型，通过指标对比模型效果。

（3）学会用预训练模型（如 BERT）提取文本的特征表示，学会基于斯坦福大学的 CoreNLP 提取句法信息等，探索宏观因素对微技能的影响。（4）学会基于自身的研究目的，将研究变量量化为计算机可提取的特征向量，构建量体裁衣的诊断模型，实现研究方法和理论模型的创新，助力学术论文撰写和发表。

工作坊难度适合零基础编程能力，学员只需要掌握语言测试知识，以及初步了解认知诊断研究。我们将以校本大学英语测试为案例，手把手教授 **python** 环境安装、诊断任务与模型构架之间的内涵关系定义、诊断数据的特征提取与表示学习方法、基于深度学习的认知诊断模型构建等问题。