

SynSym：一种用于精神症状识别的合成数据生成框架

来源: arXiv

日期: 2026-03-23

DOI: -

链接: <https://arxiv.org/abs/2603.21529v1>

【中文标题】

SynSym：一种用于精神症状识别的合成数据生成框架

【研究背景】

社交媒体上的精神症状识别旨在从用户生成的内容中推断出精细的心理健康症状，以便更深入地了解用户的心理状态。然而，由于数据集仍然具有挑战性，这反过来限制了模型识别用户生成文本中多样化症状表达的可推广性。

【研究方法】

本研究提出了SynSym，一个用于构建可推广症状识别数据集的合成数据生成框架。SynSym利用大型语言模型（LLMs）通过以下方法增强生成表达的多样性，（1）产生反映不同语言风格的精神症状的合成表达，（2）根据临床共病模式创作现实的多症状表达。

【主要发现】

在涵盖不同抑郁症状表达风格的三个基准数据集上验证SynSym，实验结果表明，仅使用SynSym生成的合成数据训练的模型与使用真实数据训练的模型性能相当，且模型在未见数据上更具鲁棒性。这些发现强调了合成数据作为精神症状建模中现实世界标注的替代资源的潜力，SynSym作为一个实用的框架，用于生成多样化的精神症状表达。

【临床意义】

SynSym为精神症状建模提供了合成数据作为现实世界标注的替代资源，有助于提高模型识别多样化症状表达的能力，从而在临床应用中更具实用性。