**社会计算 2024 Final Project**

本学期《社会计算》课程由同学们自由组队完成Final Project，每队2-4人（最多每组4人，最少每组两人；组内所有同学得分相同）。本门课程给出**部分选题参考方向**，每组队伍可以选择已给出的参考方向或者**社会计算课程涵盖**（包括AI与数据工程方向、社会信息系统的系统软件支持、数据安全、人机交互与情感计算，自然语言处理等）且自己感兴趣的**任意**主题。

**重要：如希望获得“优秀”（即4.0绩点），需提出单独申请，并参与项目答辩。答辩安排在17、18周（根据参与答辩的人数具体决定）。不申请优秀的组，最高只能获得“良好”（即B）的成绩，同时无需参加项目答辩。**

**验收形式：**

**1. 书面报告文档（必要）**

**2. 原型实现代码（可选；如项目不涉及代码，可不提交）**

**3. 答辩 （不申请优秀可不参加答辩）**

**答辩要求：**每组12分钟讲述，5分钟老师提问**+答辩PPT** （答辩后助教收集）

**提交内容及要求：**

提交地址：**水杉系统**（提交窗口标题为Final Project）

提交内容：**报告文档，如有代码需提交对应的实现代码**（打包压缩成一个文件提交）

**报告文档内容及其格式要求：**

1. 报告文档格式需为PDF形式
2. 报告文档基础内容要求（基础意味着必须含有，各组可基于项目本身增添其他章节）
3. 对关键的分析过程或结果的模式，进行直观可视化展示。项目背景描述（基于什么领域，解决什么样的问题）
4. 问题定义（明确定义解决问题，说明前提假设和问题是否/如何简化，）
5. 实现系统架构/算法框架描述（技术方案，算法流程描述。对关键部分给出示意图、伪代码或公式补充理论依据）
6. 实验设置（详述数据集、评价指标、实现细节、参数设置等）
7. 结果分析（通过表格、图像等直观展示。客观分析结果，解释现象）
8. 团队成员贡献（组内成员简单说明自己做了什么即可）

**截止日期：2024年7月3日23时59分59秒**

**参考方向：**

**方向一**

**基于自然语言的中国人心理行为分析**

任务描述：

本项任务旨在探索中国人的心理和行为模式，研究主题广泛，可包括但不限于松-紧文化、集体主义-个人主义等文化心理维度，以及大五人格等心理特质，亦或其他心理和行为方面等。同学们需运用一种或几种的自然语言处理技术，比如传统的词频分析、主题分析等，或前沿的词嵌入技术等，对所选定的心理和行为方面进行解析。研究包括但不限于：心理和行为特征的提取和分析，在不同地域或时间尺度上的差异性模式，与历史-社会生态环境因素的相互作用和联系等。

同学们可充分利用自然语言处理技术捕捉语义细节的优势，刻画中国人群的心理和行为特性。通过此任务研究，我们期待加深对中国人心理和行为复杂性的认识，并为预测未来发展趋势提供一定科学基础。

任务与测评目标：

(a) 应用恰当的自然语言处理技术，从自然语言文本中抽取能够映射中国人心理和行为特征的语义信息，并检验其信度和效度。

(b) 分析不同心理和行为特征指标和历史-社会生态环境指标之间的关系，尝试提出相应的理论解释与实践价值

**方向二**

**面孔知觉的可解释性深度学习算法**

任务描述：

面部知觉（facial perception）是社会心理学、进化心理学与人格心理学交叉融合的经典研究领域。在过去的三十多年里，传统心理学实验法积累了大量研究成果和理论洞察；而近十年来，基于机器学习和深度学习技术的研究同样取得了突破，但仍存在一些待解决的问题，尤其是在计算领域内面部知觉研究的算法可解释性，以及对面孔知觉领域理论贡献的探讨。

本项任务致力于推动基于面孔局部及整体物理特性对心理特征知觉评价的影响研究。此任务期望构建或优化基于面孔图像库及相关心理特质知觉评价信息的预测模型，该模型可融合深度学习技术，如卷积神经网络（CNN）、Transformer算法等，以实现对面孔特性和心理特征知觉间关联的准确预测。

同学们应专注于提升模型预测的精确度，力图超越已有研究，致力于增强算法的透明度和可解释性。为此，可探索和整合可解释性机制，如注意力机制、可视化技术等，以便深入理解模型的决策过程和预测效果。通过这些努力，本任务在追求算法性能提升的同时，也致力于阐明面孔局部和整体特性如何映射到心理特征知觉评价上，及其背后可能的心理学或认知神经科学原理。本任务期待为面孔知觉心理学领域贡献新洞见，并为未来心理学研究及其在社会工程、人机交互等领域的应用提供一定科学技术基础。

任务与测评目标：

(a) 模型训练：设计、改进或应用一种或几种深度学习算法模型，处理面孔局部及整体物理特性，预测相关心理特征评价；可使用但不限于卷积神经网络（CNN）、Transformer等算法作为模型基础架构；对模型进行训练，调整参数以优化模型性能。

(b) 模型评估：相较于以往基于同一或类似图片数据库的研究基准，衡量现有模型的预测性能水平。

(c) 可解释性：采用适当的可解释性技术，直观描述或可视化模型决策过程中的关键特征。

(d) 贡献阐述：说明模型预测结果背后的逻辑和原因，与现有面孔知觉领域发现或理论之间的关联，提供可能的新洞察。

**方向三**

**人格特质的社会生态因素亲和性分析**

任务描述：

本项任务旨在探讨大五人格等个性心理特征，与自尊、生活满意度、主观幸福感、身体健康等生存发展适应性指标之间的历史-社会生态因素动态关联性。既有研究发现，在不同的微观和中观情境下，人格与适应性指标间关系呈现出不同程度的波动性，这表明人格特征的适应性功能并不发生在“真空”中。人格的各个维度与其所处的微观或中观情境之间存在着特定的匹配适应模式。亲和性假说进一步拓展了这一分析框架，将研究视野扩展到更宏观、更广泛的历史-社会生态因素中，如经济、政治、社会、文化、气候、人口和卫生等；并初步发现，在宏观层面上，人群的人格特征与贫富差距、集体主义-个人主义文化等社会生态因素间存在显著的交互效应，影响宏观人群的整体适应性水平，如国民幸福感、人均寿命等区域健康指标。

本项任务要求利用中国或美国或全球的人格数据集，以及较为典型的历史-社会生态宏观指标数据，分析个体或区域群体水平上的人格维度，与不同历史-社会生态环境因素的匹配适应性模式（选择一种或几种适应性指标皆可），并对其进行可能的理论提炼，以形成关于人格-宏观环境-适应性的综合考察。

任务与测评目标：

(a) 提供基于传统心理测量法或大数据分析所得关键研究指标的信效度信息。

(b)收集和整理典型的历史-社会生态宏观指标数据，并对其进行必要的降维简化处理。

(c) 对特定的亲和-匹配适应性模式进行较深入分析，并做出归纳总结。

(d)对关键的分析过程或结果的模式，进行直观可视化展示。

**方向四**

**基于图分析的真实场景应用**

在大数据时代,图形数据正以前所未有的速度增长。社交网络、电商平台、物联网等领域产生了海量的图数据,蕴含着丰富的关联信息和隐藏模式。高效地管理和分析超大规模的图数据,尤其是包含复杂关系和网络结构的图数据,已成为学术界和工业界共同关注的热点问题。

我给大家一个例子：国际互联网服务提供商联盟(ICISP)预计,物联网(IoT)的快速增长将从根本上改变互联网的结构。据估计,互联网已经增加了60亿个节点,并且还在以每天数百万的速度继续增长。然而,IoT设备的大规模部署也带来了新的安全隐患。网络犯罪分子和恶意企业可能会利用IoT设备的脆弱性,构建大规模的僵尸网络,发动网络攻击,危害互联网的安全和稳定。我们能否实施一个全球互联网流量监测系统。该系统需要能够实时分析海量的网络连接数据,及时发现潜在的僵尸网络和恶意节点。图分析技术通过将网络连接数据建模为图,利用图算法进行高效的结构分析和异常检测,有望实现对互联网安全威胁的早期预警和快速应对。

本次任务要求各位能够将图分析领域的核心方向子图同构匹配和社区检测应用在“真实”世界的数据集中，比如通过三角形检测和k-truss算法在多项式运行时间内对互联网上的节点进行快速表征，同时在网络中分区，缩短网络安全团队的响应时延，不需要等待完整的数据摄取和处理完成。

任务与测评目标：

（a）查阅选择一种图分析应用场景，例如,在社交网络中实现实时的社群检测和演化分析,帮助理解群体行为模式;在金融领域构建风控知识图谱,助力智能反欺诈;在生物医疗领域对蛋白质相互作用网络进行高效分析,加速新药研发等等。基于应用场景，对真实世界数据进行数据处理与图构建。

（b）基于现有的图分析算法应用于已构建好的图数据中，学有余力可以基于基准算法进行算法优化创新。例如，设计用于检测紧密连接社区(潜在僵尸网络)的图分析算法;对算法本身进行基本理论分析。

（c）借助各类图处理框架实现（b）中的图分析算法，学有余力可以优化算法实现。例如，提升图分析的并行度和可扩展性。对实现的图分析引擎进行性能测试,分析其吞吐量、延迟和资源消耗等指标

（d）利用实现的图分析引擎,应用于真实世界数据。评估运行结果的各项指标，对关键的分析过程或结果的模式，可进行可视化展示和交互式分析。

数据来源参考：

·<https://graphchallenge.s3.amazonaws.com>

·真实世界数据 - <https://snap.stanford.edu/data/>

·扩展 - <http://www.graph500.org/specifications#sec-3_3>

**案例 -- 基于参考方向一的小例子：**

情绪是人们内心世界的真实映射，社交平台中情绪的产生，对于情绪产生原因的分析是社会计算中重要的分支之一。通过对社交媒体帖子、评论、等数据进行分析建模，能够快速定位人们对一特定主题或事件的看法和感受。

**目标：**实时文本情绪分析器—中国主流社交媒体

**实现内容：**

1. 利用提供的包含情绪标注的推文数据集，对数据集进行数据预处理->特征工程->模型构建->分析模型的准确性和性能
2. 采用网络爬虫技术对任一事件、话题、现象或商品等议题的评价，通过模型分析后用可视化图表来表示爬取评价数据的情绪分布并根据时序阐述公众情绪变化趋势。

**注：**

1. 网络爬虫技术的实现方式不限（如request，Selenium，Scrapy等python库）,对于爬取后数据提取解析方法不限（推荐自行定义正则表达式）。爬取目标平台可以为外网或是国内媒体数据爬取后翻译为英文。
2. 模型构建不需要用复杂的模型嵌套（如果你想也可），使用简单的机器学习内容即可（如：朴素贝叶斯，支持向量机等），使用机器学习算法需自己实现不允许掉包。

**参考数据集情况：**

数据集包含了英文评论以及它们对应的情绪标签。数据集包含了六种基本情绪：悲伤 (sadness)、喜悦 (joy)、爱 (love)、愤怒 (anger)、恐惧 (fear)、惊讶 (surprise)。每条语句都对应着一个情绪标签，标签的取值范围为 0 到 5，分别代表六种情绪（如：0 代表悲伤，1 代表喜悦，以此类推）。数据集来源：<https://www.kaggle.com/datasets/nelgiriyewithana/emotions>