

# 疾病危险因素跨时空超图构建工具软件

## 1. 工具软件背景介绍

### 1.1 开发背景

随着全球人口的不断增长和生活方式的变化，疾病的流行模式也在不断演变。疾病的爆发和传播不仅受到遗传因素的影响，还受到环境、社会和行为因素的影响。因此，了解和预测疾病危险因素的变化越来越显得至关重要。然而，要构建准确的疾病危险因素模型，需要整合跨不同时间和空间尺度的大量数据，这对于传统的数据分析方法来说是一项巨大的挑战。因此，有必要开发一种能够帮助研究人员整合各种数据，构建跨时空超图，并能通过可视化结果进行高效的分析和预测的软件工具，可以实现使用医疗单病种特征-样本数据集对多个病种的危险因素的跨时空超图的构建，通过可视化工具展示构建结果，从而表征疾病危险因素的时间累积效应和空间层次效应以及危险因素间复杂的关联关系，软件旨在为医学领域的科研工作者提供辅助和支持。

### 1.2 工具软件目标

#### 1. 时间累积效应表征

有效的时间累积效应表征方法能够捕捉疾病发展过程中关键因素的动态变化和累积效应。例如，患者在不同时间点的病理特征、医疗记录等数据，通过跨时空超图构建工具可以展示其随时间的变化和影响，以及所产生的新特征，有助于更准确地分析和预测疾病的进展。

#### 2. 空间层次效应表征

空间层次效应表征不仅包括医学上的生理指标、人口学和行为学等层次，还涉及到患者样本所在的地理空间。通过跨时空超图构建工具，能够将这些不同层次的信息有机结合，展现出不同空间层次因素对于疾病风险的影响。这种综合表征方法有助于识别出特定地区或特定人群中的高危因素，为医疗决策提供有力支持。

#### 3. 超图可视化

通过直观的图形展示，可以更清晰地理解复杂的疾病危险因素及其相互关系。超图可视化不仅能够展示不同特征之间的关联，还能通过多维度的交互展示，帮助医疗专业人员和研

究人员快速洞察危险因素与疾病背后的规律，从而更有效地进行疾病预防、诊断和治疗。

## 2. 工具软件功能模块介绍

疾病危险因素跨时空超图构建工具软件主要包括了五个模块，分别是首页、登录与注册、任务记录、数据管理和系统管理，其中系统管理为管理员功能。

### 1. 首页

首页包括快捷入口和统计信息。

### 2. 登录与注册

登录与注册完成用户的登录与注册功能，系统分管理员和普通用户，通过注册的用户默认状态为普通用户。

### 3. 超图构建

超图构建模块主要包括对两种超图构建任务的新建，危险因素超图和空间多层网络，新建任务需要输入任务信息、选择目标疾病和数据集、选择算法、手动选择特征列和标签列，运行任务后展示结果。

### 4. 任务记录

任务记录查看实现对相关任务的查询、删除功能。用户通过相关的筛选条件选择相应的任务查看详情，并选择回溯任务，跳转到可视化结果页面，对任务结果进行查看；用户可以删除相应的任务。

### 5. 数据管理

数据管理模块实现对医疗数据的数据上传、数据存储、数据查询、数据筛选功能。其中数据上传是指用户的数据文件传到数据库中，通过前端页面可以上传数据；数据存储功能将医疗数据存储到 PG 数据库中，可以根据上传的文件进行动态生成表来进行数据的存储；数据查询功能可以根据用户输入的特定的条件查询想要的数据库；数据筛选功能用户可以通过相应的筛选条件来构造专病数据集并保存在数据库中。其中数据可以分三类，私有数据集，共享数据集以及公共数据集。私有数据集和共享数据集都可以是通过用户自己上传，用户自己所上传的数据集可以进行转为私有，转为共享以及删除操作。其中公共数据集普通用户无法进行操作。

### 6. 系统管理

系统管理为管理员专属功能，分为用户管理、信息发布、系统数据管理、日志查看和病种设置。

用户管理完成用户的增、删、改、查操作；信息发布完成系统信息的发布，在登录界面滚动播放；系统数据管理完成对单病种任务的数据管理；日志查看可以查看所有用户的操作日志；病种设置对系统中的病种进行增、删、改、查操作，由管理员按照 icd10 编码统一管理。

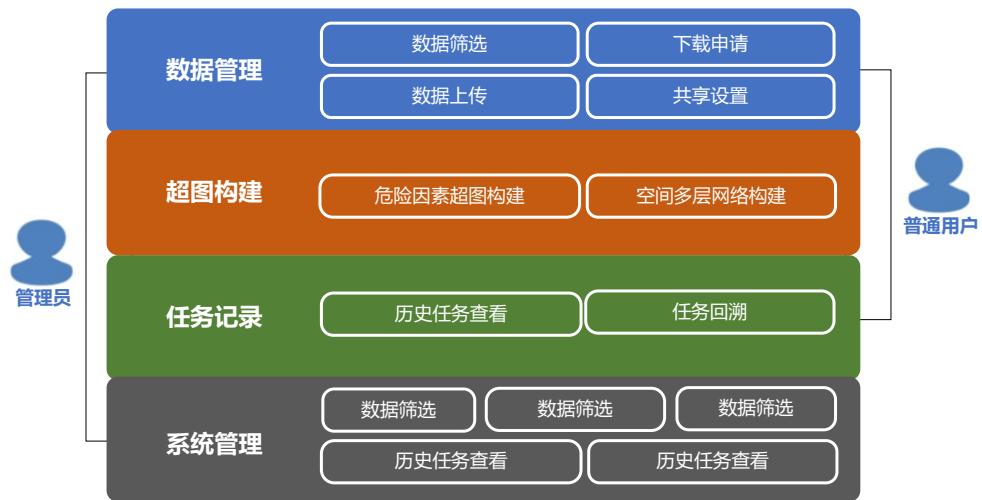


图 1 功能模块图

### 2.1 工具软件整体架构

该项目结合了网络技术、数据库技术，采用了 B/S 的浏览器/服务器架构模式，浏览器通过 http 向服务器发起请求进行操作，服务器接收请求后从数据库获得数据返回给浏览器，浏览器通过 html 将数据渲染给用户使用。前端技术：Vue+ElementUI，后端技术：SpringBoot+MyBatis，数据库：Postgre SQL。

### 2.2 功能 1：首页

首页上部为数据库综合统计信息，以及四个主要模块的入口，下部包含了病种数据集的数量统计信息以及样本地理空间分布信息。

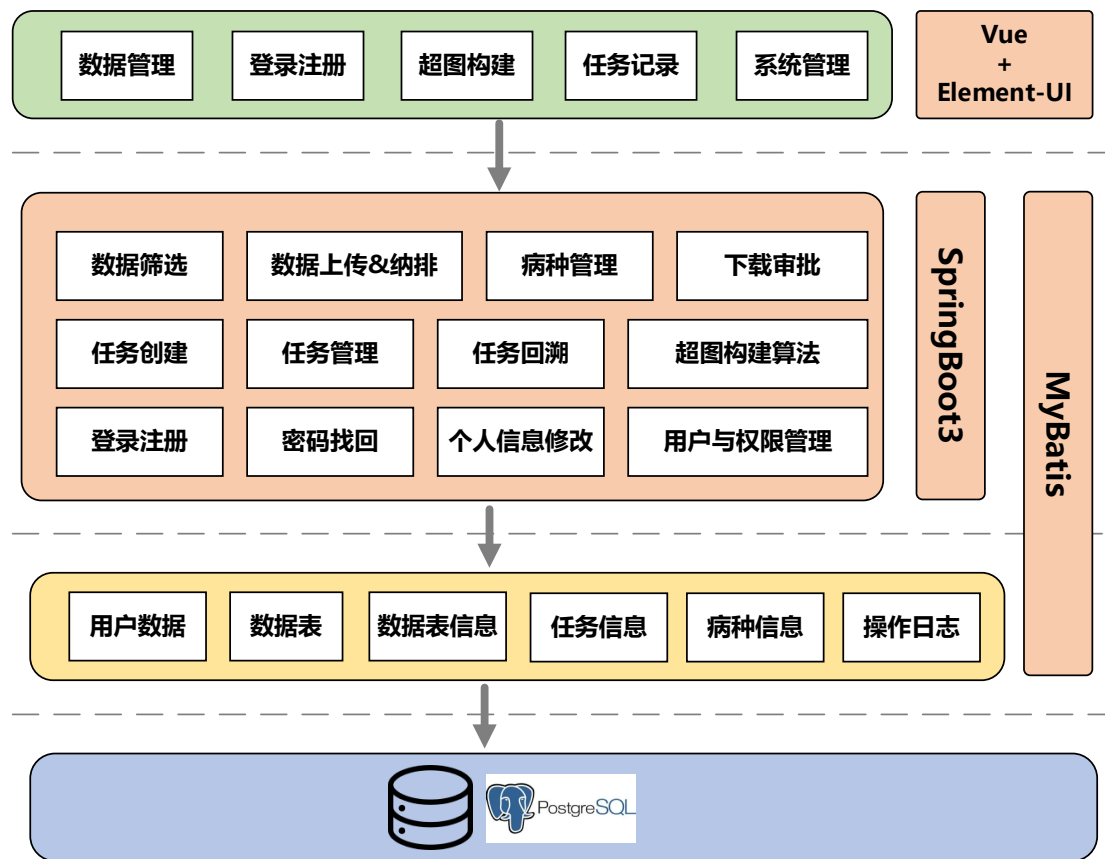


图 2 软件架构图

### 2.3 功能 2：登录与注册

登录与注册完成用户的登录与注册功能，系统分管理员和普通用户，通过注册的用户都是普通用户。同时登陆页面还会展示相关性的系统发布的信息。具备用户忘记密码验证后找回密码的功能。

### 2.4 功能 3：超图构建

超图构建模块主要包括对两种超图构建任务的新建，危险因素超图和空间多层网络，新建任务需要输入任务信息、选择目标疾病和数据集、选择算法、手动选择特征列和标签列，运行任务后展示结果。

#### （1）空间多层网络构建

空间多层网络构建主要通过关联系数算法计算疾病危险因素之间的关联关系，将不同层次的因素展示在不同层的网络中，层内以及层与层之间都包含危险因素之间的关联，并展示当前数据集的地理空间分布。

## （2）危险因素超图构建

危险因素超图构建主要通过使用 K-Means、KNN、基于稀疏表示的方法等构建关于疾病危险因素的超图，通过可交互可视化方法展示任务结果。

## 2.5 功能 4：任务记录

任务记录查看实现对相关任务的查询、删除功能。用户通过相关的筛选条件选择相应的任务查看详情，并选择回溯任务，跳转到可视化结果页面，对任务结果进行查看；用户可以删除相应的任务。

## 2.6 功能 5：数据管理

数据管理模块实现对医疗数据的数据上传、数据存储、数据查询、数据筛选功能。其中数据上传是指用户的数据文件传到数据库中，通过前端页面可以上传数据；数据存储功能将医疗数据存储到 PG 数据库中，可以根据上传的文件进行动态生成表来进行数据的存储；数据查询功能可以根据用户输入的特定的条件查询想要的数据库；数据筛选功能用户可以通过相应的筛选条件来构造专病数据集并保存在数据库中。其中数据可以分三类，私有数据集，共享数据集以及公共数据集。私有数据集和共享数据集都可以是通过用户自己上传，用户自己所上传的数据集可以进行转为私有，转为共享以及删除操作。其中公共数据集普通用户无法进行操作。

## 2.7 功能 6：系统管理

系统管理为管理员专属功能，分为用户管理、信息发布、系统数据管理、日志查看和病种设置。

用户管理完成用户的增、删、改、查操作；信息发布完成系统信息的发布，在登录界面滚动播放；系统数据管理完成对单病种任务的数据管理；日志查看可以查看所有用户的操作日志；病种设置对系统中的病种进行增、删、改、查操作，由管理员按照 icd10 编码统一管理。

### 3. 工具软件运行所需资源

#### 3.1 软件环境需求

工具	版本
CentOS	7
MySQL	5.7
Maven	3.9.2
jdk	1.8.0
Postgres	16.2
python	3.11

### 4. 工具软件所包含算法

#### 4.1 算法 1: K-Means 超图构建算法

**算法介绍:** K-Means 是一种无监督聚类算法，其目标是将数据集分成  $k$  个簇，使得同一簇内的点尽可能接近，而不同簇的点尽可能远离。通过对特征矩阵进行 K-Means 聚类，分成  $k$  个簇，每个簇对应一个超边，簇内的节点都连接到这个超边，并构建超图的邻接矩阵，表示节点和超边之间的连接关系。

**算法所需环境依赖包信息:**

python 包名	版本
pandas	2.0.1
numpy	1.24.3
psycogp2	2.9.9
cvxpy	1.4.1
scipy	1.10.1
scikit-learn	1.2.2
hyperg	0.0.3

### 4.2 算法 2：KNN 超图构建算法

**算法介绍：**K 最近邻（K-NearestNeighbors，KNN）是一种分类和回归算法，用于根据输入样本的最近 k 个邻居来预测其所属类别或值。通过计算特征矩阵中所有节点的距离矩阵。对每个节点，找到其最近的 k 个邻居。每个节点及其最近邻居形成一个超边。

**算法所需环境依赖包信息：**

python 包名	版本
pandas	2.0.1
numpy	1.24.3
psycogp2	2.9.9
cvxpy	1.4.1
scipy	1.10.1
scikit-learn	1.2.2
hyperg	0.0.3

### 4.3 算法 3：L1 正则化超图构建算法

**算法介绍：**L1 范数正则化利用稀疏表示的方法，将每一个节点表示为其最近邻居的线性组合，并使用 L1 范数正则化来保证表示的稀疏性。算法利用特征重构来确定顶点之间的关系。通过计算输入节点之间的距离矩阵，将每个节点表示为其邻居节点的稀疏加权组合，并通过优化问题最小化重构误差和 L1 范数来确定权重，从而生成一个稀疏的超图邻接矩阵。

**算法所需环境依赖包信息：**

python 包名	版本
pandas	2.0.1
numpy	1.24.3
psycogp2	2.9.9
cvxpy	1.4.1
scipy	1.10.1
scikit-learn	1.2.2
hyperg	0.0.3

#### 4.4 算法 4: Pearson 相关系数多层网络构建算法

**算法介绍:** Pearson 相关系数 (Pearson Correlation Coefficient) 是一种用于衡量两个变量之间线性关系强度和方向的统计指标, 取值范围从-1 到 1。其计算方法包括计算两个变量的协方差, 然后除以它们的标准差乘积。当 Pearson 相关系数接近 1 时, 表示两个变量之间有强正相关关系; 接近-1 时, 表示强负相关关系; 接近 0 时, 表示无线性相关关系。该系数广泛应用于统计分析和数据挖掘中, 用于检测变量之间的相关性。

算法所需环境依赖包信息:

python 包名	版本
pandas	2.0.1
numpy	1.24.3
matplotlib	3.7.1
networkx	2.8.8
psycogp2	2.9.9