

以下是该项目的简要文档说明：

项目目标：

该项目旨在实现知识图谱异常检测，通过结合文本信息和结构信息来发现知识图谱中的异常三元组。知识图谱中的异常三元组可能表示错误的事实或不常见的模式，检测这些异常对于提高知识图谱的质量和可靠性具有重要意义。

主要思路：

数据加载与预处理：使用 `KGCDDataModule` 类加载和预处理知识图谱数据，包括文本描述和结构信息。将数据划分为训练集、验证集和测试集，并对文本进行编码，将结构信息转换为模型可处理的格式。

模型构建：

文本模型：使用预训练的 `BertForMaskedLM` 模型作为基础，构建 `NBert` 类来处理文本信息，提取文本中的语义特征。

结构模型：定义 `Knowformer` 类来处理知识图谱的结构信息，捕获实体和关系之间的结构模式。`NFormer` 类则负责将结构信息与文本信息进行融合。

交互分类器：通过 `Inter_Classifier` 类对文本模型和结构模型的输出进行综合分析，学习两种模型输出之间的关联，以确定三元组的异常程度。

训练策略：

联合训练：采用联合训练的方式，同时优化文本模型、结构模型和交互分类器。通过定义合适的损失函数，使模型在学习文本和结构信息的同时，能够有效地识别异常三元组。

软标签更新：在训练过程中，根据样本的损失动态更新软标签。这有助于模型更加关注那些难以分类或容易被误分类的样本，从而提高模型的检测性能。

验证与评估：在每个训练周期后，使用验证集对模型进行评估。通过计算不同阈值下的精确率和召回率，评估模型对异常三元组的检测效果。