



南京航空航天大学  
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

---

# 基于大模型的智能软件开发

---

计算机科学与技术学院

智周万物·道济天下

---

# 目录



- 基于大模型的智能软件开发框架
- 智能软件开发中的大模型及预训练范式
  - 常用大模型
  - 预训练范式
- 智能软件研发的下游任务
  - 程序语言PL相关任务
  - 自然语言NL相关任务
  - 程序语言与自然语言交互任务
- 常用数据集
  - 预训练数据集
  - 下游任务数据集
- 思考



- 基于大模型的智能软件开发框架
- 智能软件开发中的大模型及预训练范式
  - 常用大模型
  - 预训练范式
- 智能软件研发的下游任务
  - 程序语言PL相关任务
  - 自然语言NL相关任务
  - 程序语言与自然语言交互任务
- 常用数据集
  - 预训练数据集
  - 下游任务数据集
- 思考

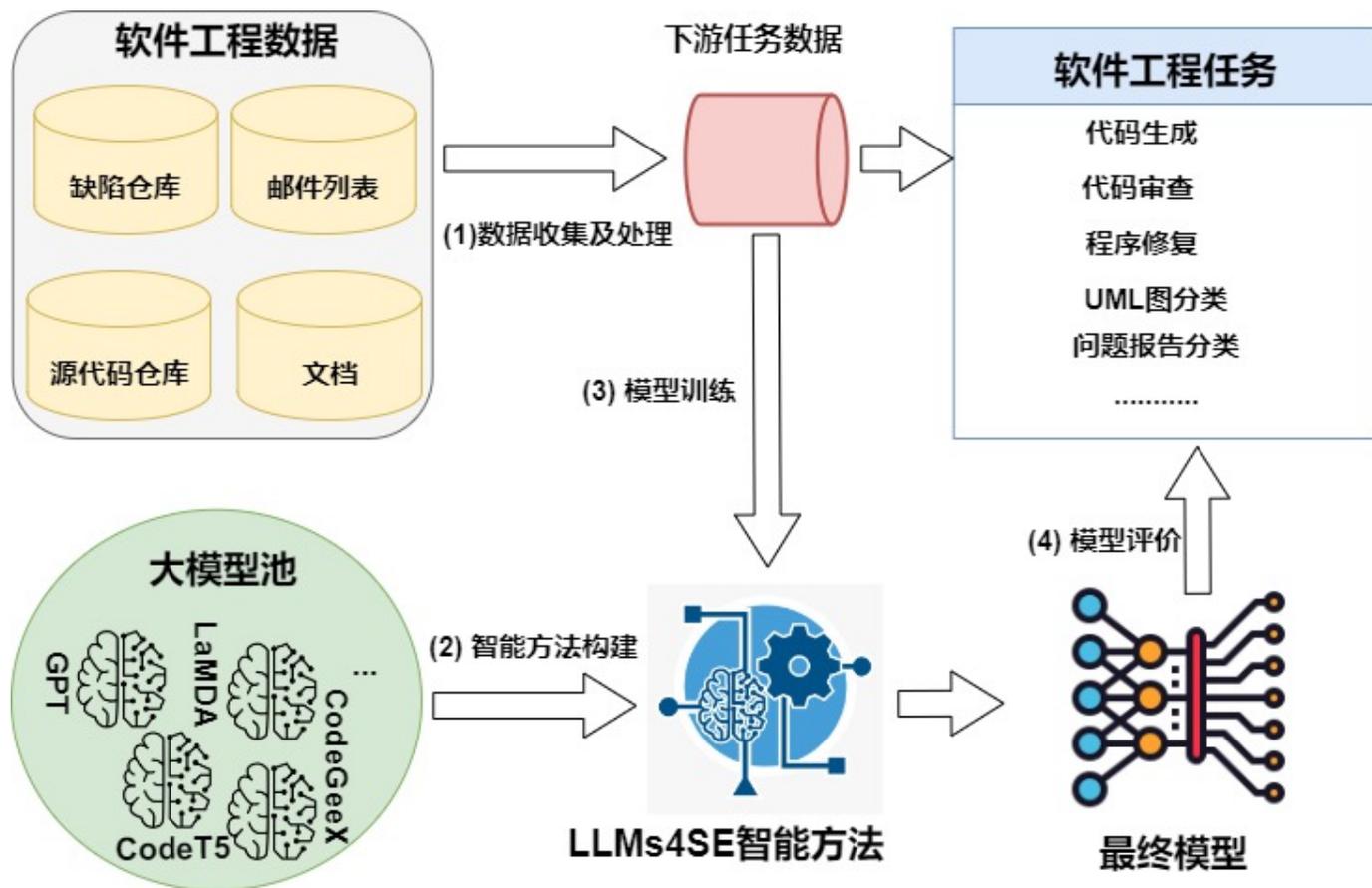
# 基于大模型的智能软件开发框架



南京航空航天大学  
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

- **LLM4SE ( Large Language Model for Software Engineering )** : 使用具有少量标签的软件工程下游任务数据，通过微调已有大模型构建的针对软件开发中各种编程任务的智能模型
- 具体的构建过程主要包括：
  - **软件开发任务数据集收集及处理**
  - **基于大模型的智能方法构建**
  - **模型训练**
  - **模型评价**

# 基于大模型的智能软件开发框架



## □ 数据收集及处理

- 从软件工程仓库中收集所需的数据集，如源代码仓库中的代码数据、缺陷跟踪系统的缺陷报告等，对数据集进行标记及预处理，使其符合模型输入的要求。

## □ 基于大模型的智能方法构建

- 根据软件工程具体任务，从大模型池中选取合适的大模型，如CodeGeeX, CodeBERT、GPT-C, 基于该大模型构建符合需求的智能方法。

## □ 模型微调

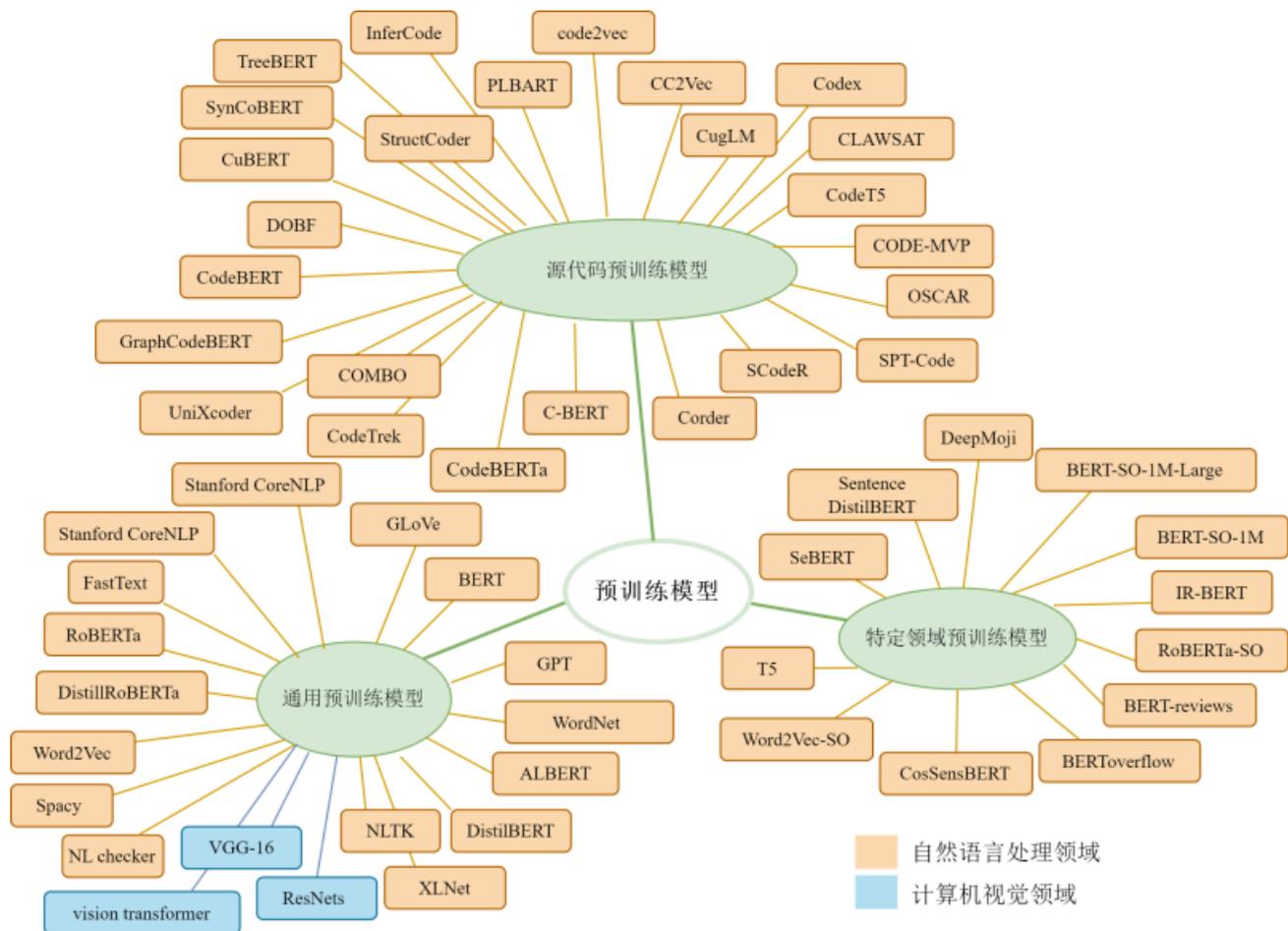
- 使用少量标签数据，选择合适的微调模式来微调大模型，从而构建高性能的LLMs4SE模型。

## □ 模型评价

- 通过性能评价指标，评价基于大模型的智能方法在软件研发具体任务中的性能。



- 基于大模型的智能软件开发框架
- 智能软件开发中的大模型及预训练范式
  - 常用大模型
  - 预训练范式
- 智能软件研发的下游任务
  - 程序语言PL相关任务
  - 自然语言NL相关任务
  - 程序语言与自然语言交互任务
- 常用数据集
  - 预训练数据集
  - 下游任务数据集
- 思考



## 通用大模型

- **一般领域数据集上训练**出的大模型，如自然语言处理领域中使用英文维基百科或普通新闻数据集训练出的GPT模型等，以及计算机视觉中使用ImageNet等数据集预训练的ResNet和VGG模型等。

## 特定领域大模型

- **软件工程领域特定数据集上训练**出的大模型。一般领域数据集训练出的大模型不能很好地适应软件工程领域的文本特性。因此，研究者收集大量的软件工程领域数据集从零训练深度学习模型形成软件工程特定领域大模型，如SeBERT、T5，进而解决软工领域任务。

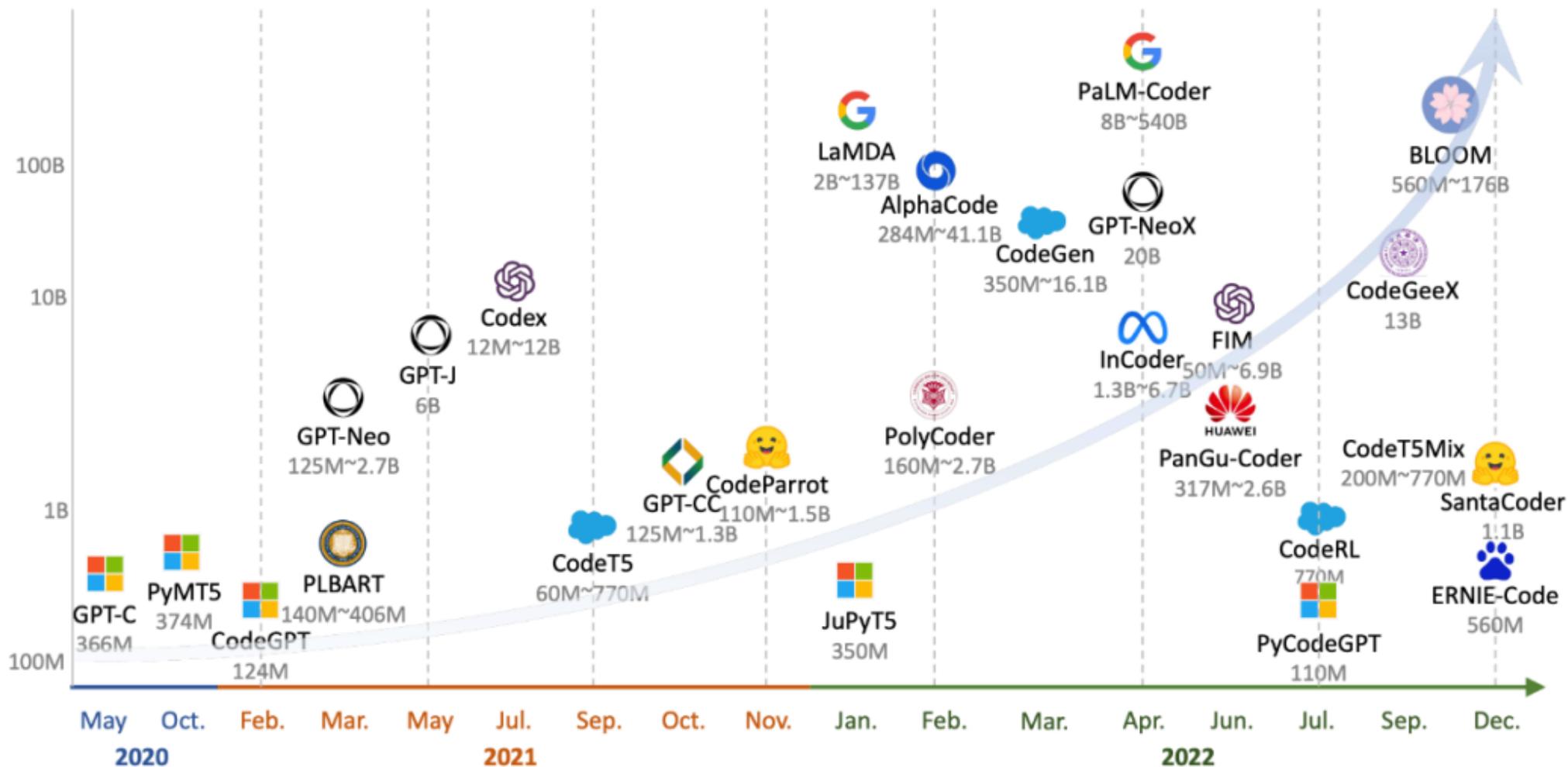
## 代码大模型

- **在源代码数据集上训练**出的大模型。为了更好地捕获源代码数据中的语法和语义信息，研究者们收集大量的源代码数据集，重新训练深度学习模型形成软工领域的代码大模型，如CodeGPT、Starcoder、CodeT5等。

# 常用大模型



南京航空航天大学  
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS



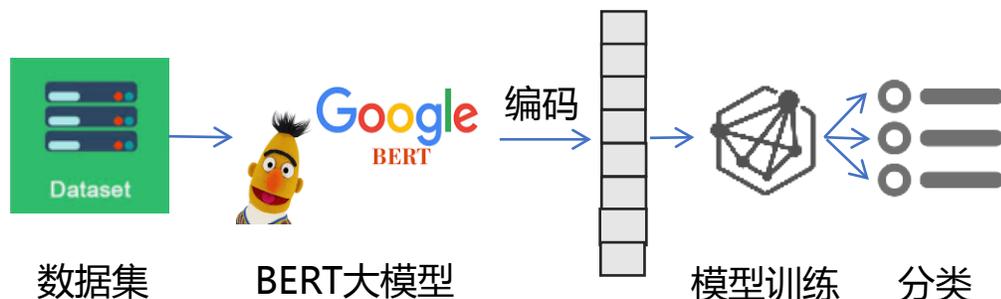
目前流行的27个源代码大模型



- 基于大模型的智能软件开发框架
- 智能软件开发中的大模型及预训练范式
  - 常用大模型
  - 预训练范式
- 智能软件研发的下游任务
  - 程序语言PL相关任务
  - 自然语言NL相关任务
  - 程序语言与自然语言交互任务
- 常用数据集
  - 预训练数据集
  - 下游任务数据集
- 思考

## 预训练—特征表示

- 软件工程领域研究者直接使用在大规模数据集上训练的大模型，对软工领域的任务数据集进行编码特征表征，然后通过编码的特征表征，构建较好的分类器等模型，以实现软件工程领域任务。



## 预训练—微调

- 软件工程领域研究者以在大规模数据集上训练的大模型为基础，构建包含下游任务的智能模型，然后通过少量下游任务数据集对构建的智能模型进行微调训练，最终构建适应下游任务的智能模型。这种模式避免了针对不同任务需要大量数据集从头训练模型。
- 预训练-微调范式中根据微调的范围又可分为只微调任务层和微调整个智能模型两类。
  - 微调任务层是**冻结大模型的特征表示层**，通过少量的下游任务数据，**只微调模型的任务层参数**。
  - 微调整个智能模型是使用下游任务数据集**微调基于大模型的智能模型**，包括大模型的参数，得到最终的智能模型。



- 基于大模型的智能软件开发框架
- 智能软件开发中的大模型及预训练范式
  - 常用大模型
  - 预训练范式
- 智能软件研发的下游任务
  - 程序语言PL相关任务
  - 自然语言NL相关任务
  - 程序语言与自然语言交互任务
- 常用数据集
  - 预训练数据集
  - 下游任务数据集
- 思考

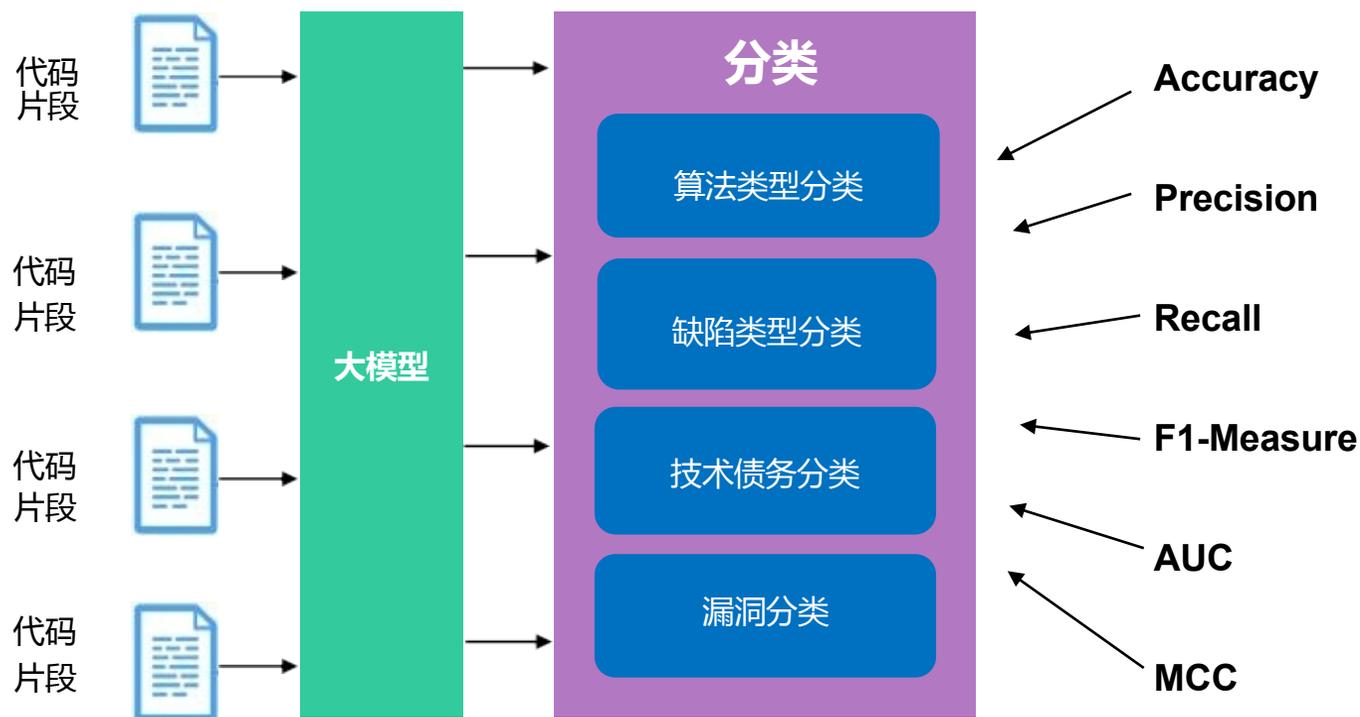


□ 根据输入数据类型，智能软件研发领域的下游任务可以分为**程序语言 (Program Language : PL) 相关任务**、软件工程领域**自然语言 (Natural Language : NL) 相关任务**、**程序语言与自然语言交互任务**、**软件工程领域图像相关任务**。

# 程序语言PL相关任务



- **代码片段分类(Code Snippets Classification):** 通过基于大模型的智能方法捕获代码片段中丰富的语法和语义信息，以预测代码片的类型，为开发者提供有用的分类信息，帮助开发者更好地理解代码。
- 该任务常用的性能评价指标有：Accuracy、Precision、Recall、F1-Measure、AUC、MCC(Matthews Correlation Coefficient score)



# 程序语言PL相关任务



- **程序修复(Program Repair):** 采用不同的技术自动生成修复缺陷代码片段。为了提高程序修复的性能，软件工程领域部分研究者利用大模型来学习成对的代码片段的语法与语义信息，并将该信息应用到程序修复任务中。
- 该任务常用的性能评价指标有：Exact Match(EM)、Number of fixed bugs、BLEU(Bilingual Evaluation Understudy)等。
  - EM用来衡量生成的修复代码是否与开发人员实际实现的修复代码完全相同。
  - Number of fixed bugs 是通过运行测试用例来查看生成的补丁是否通过测试, 进而得到修复的 bugs 的个数。
  - BLEU 用来计算预测和正确答案之间的 n-gram 相似度，为 n-gram 匹配精度分数的几何平均值。
  - EM、BLEU 和Number of fixed bugs 越大，模型越好。

# 程序语言PL相关任务



- **代码补全(Code Completion):** 基于上下文代码信息实时建议下一个可能的符号，例如类名、方法名等，用以补全代码片段，加速软件的开发。大模型技术可以通过上下文代码规律解决代码补全任务。
- 该任务常用的性能评价指标有：EditSIM(Edit Similarity)、Exact Match(EM)、Perplexity等。
  - Perplexity 是 token-level 代码补全的评估指标，用来度量模型预测样本的好坏程度，即下一个词时的平均可选择数量。
  - EM 和 EditSIM 是 Line-level代码补全的评估指标。
  - EditSIM 是两个单词之间 Levenshtein 距离，即将一个单词更改为另一个单词所需的最小单字符编辑次数，包括插入、删除或替换。
  - EM 用来评估模型预测中匹配到正确答案的百分比。
  - 通常是 Perplexity 和 EditSIM 越小，模型越好。Accuracy 和 EM越大，模型越好。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：  
<https://d.book118.com/156101231234011004>