

CovScript 编程语言 在学术研究中的应用

李登淳 四川大学计算机学院



讲者简介

- 李登淳，2000年生，山东济宁人，常用网络ID为 mikecovlee
- CovScript 编程语言创始人，川大智锐科创计算机协会创始人
- 活跃开源社区贡献者，资深 C++ 跨平台开发者
- 四川大学计算机学院2022级全日制硕士研究生
- 四川大学吴玉章学院2018级双特生(本科学位)
- 研究方向为编译语言、AIGC 和数据库



四川大学
SICHUAN UNIVERSITY

目录

CONTENTS



01

编程语言在学术研究中的应用现状

02

CovScript 编程语言简介

03

CovScript 为前沿研究赋能

04

CovScript 应用案例



编程语言在学术研究中的应用现状

Chapter 1



学术研究 vs. 工程开发

- 软件工程开发的目标一般是使用成熟技术满足工程需求，对于新技术往往不是特别欢迎，因为新技术会带来变量。对于商业公司来说，这首先会增加工作量，但最重要的还是会导致产品不稳定。除非一项新技术能明显提升效率或降低成本，否则很难在市场中取得优势地位；
- 计算机的学术研究一般围绕一个课题进行，每个课题的技术选型可能是完全不同的。若是一个新方法能对实验结果产生哪怕百分之零点几的提升，就算方法不完善、Bug多，只要不影响最终实验，都无关紧要。而且与工程开发不同，往往完成一个课题后就不需要提供“售后服务”了，极大降低了维护成本。这为新技术的应用提供了广阔的舞台。

编程语言在学术研究中的应用现状： 以低门槛语言为主，部分语言在特定领域大放异彩

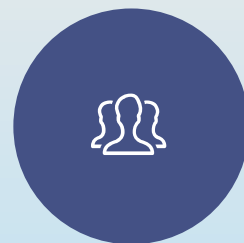
- 在大多数学术研究中，以 Python、R 为首的低门槛语言是最常用的：在计算机的学术研究中，受益于神经网络和 AIGC 的发展，Python 已经形成了某种程度的垄断；在一些需求更加简单的专业中，如商业分析和医学数据分析，SPSS、Stata 甚至内嵌于 Excel 中的 VBA 是更流行的选择；
- 形成这种现象的主要原因是大多数学术研究往往更专注于理论层面的改进以及论证一项工作的必要性和先进性，代码实现越简单越好。类似 C++、Java 等工程上常用的编程语言往往会带来不必要的复杂度（研究牲生们写论文已经够难了，为什么还要折磨我们☹）；
- 但对于一些专业性非常强的领域，如神经网络，很多 DSL（Domain-Specific Language，领域专属语言）大放异彩。比如大名鼎鼎的 PyTorch 深度学习框架就基于 Python 和 Triton 执行引擎设计了 TorchScript。

结论： 新兴编程语言在科研赛道中仍有很多发展机遇

- 实际上，大家已经很久没有在视野中看到新兴的编程语言了。我们认为比较新的 TypeScript、Go、Rust 等语言实际上都已经发展了十多年（就连 CovScript 都已经快六年了）；主要原因是编程范式已经被发掘殆尽，现有流行的编程语言如 C++、Java 也在不停进化以适应发展趋势；
- 在国内，因为编程语言不像其它软件如操作系统、EDA 等有很严峻的“卡脖子”风险，其发展也不受重视。CovScript 作为国内最早一批投入应用的编程语言，至今生态依旧不算全面，应用也仅限于高校和科研赛道；
- 其他的一些编程语言，如 KCL 等，通过解决某个领域中的特定问题，得到了很好的发展。因此，若要发现机遇，就需要寻找和解决这种领域问题。而科研赛道中有数不清的领域问题，新兴编程语言可以通过解决这些问题，从而在科研赛道中取得不错的成绩。

CovScript 编程语言简介

Chapter 2



Covariant Script

开源、开放的跨平台动态编程语言

- Covariant Script 编程语言，简称 CovScript 或智锐编程语言，诞生于2017年，是一门动态强类型的编程语言
- CovScript 是一门以过程式为主、函数式和 OOP 为辅的通用编程语言。简单易用、符合直觉的语法配合完善的自建生态使 CovScript 除了非常适合快速开发，也能承担复杂的系统性开发任务
- CovScript 是一门开放的编程语言，其语言标准、DBC 标准、SDK 标准由社区主导，核心工具链基于 Apache 2.0 License 全部开放源代码，甚至可以在其基础上快速派生出一门全新的编程语言¹
- 主页：<https://covscript.org.cn/> (English) <https://unicov.cn/covscript/> (中文)
- GitHub：<https://github.com/covscript/>

1: 基于 ParserGen 框架，详见 <https://zh-lang.osanswer.net/t/topic/138/30>

Covariant Script 助力前沿领域应用

- Data Analysis
 - CovScript 支持 Unicode 字符处理，内嵌高效 JSON 解析库；CovScript 具有完善的数据分析框架 CovAnalysis，性能比肩 Pandas；CSDBC 兼容大多数主流 RDBMS
- DevOps
 - CovScript 在现有生产系统中已被用于容器调度、系统监控、应用部署等关键任务，其可信、可靠、可裁剪的特性使其适用于广泛的系统级应用中
- AIGC
 - CovScript 在 AIGC 领域中已被用于训练数据处理、训练任务监控等应用
- Web
 - CovScript 原生支持 Stackful Coroutine，上下文切换性能可达 36 GOPS²

2: 在 CovScript 3.4.3.12 @ AMD R7-5800X, Linux 5.15.90.1-WSL2 中测试

GOPS 表示每秒进行的十亿次操作数 (Giga Operation Per Second)

Covariant Script 践行中国智造

- 可信、可靠
 - 语言核心生态 100% 自主知识产权，周边生态 100% 开源、可信
 - 已被四川大学信息化办采用，支持着上百间教室的智慧教学工作
- 支持国产生态
 - 针对龙芯架构和国产操作系统专门优化、测试
 - CovScript 每个发行版都会有对应的龙芯版 (UOS@3A4000)
 - 兼容华为鲲鹏处理器和 openEuler 操作系统
 - CSDBC 兼容华为 openGauss
- 兼容存量系统
 - CovScript 主流版本兼容 Windows 7 64bit
 - 可定制兼容 Windows XP SP2

CovScript 为前沿研究赋能

Chapter 3



CovScript 为前沿研究赋能：Motivation 帮助科研人员专注于专业知识研究

- CovScript 本身是一门门槛较低的编程语言，面向的市场与 Python 是比较重合的
- Python 靠生态取胜，语言设计本身一言难尽
- CovScript 不仅有丰富的应用框架（CovAnalysis、NetUtils、CSDBC、PicassoUI 等），还能轻松的在其基础上派生出一门解决特定需求的 DSL（ParserGen 语法规则基于 EBNF 派生）
- 近期 CovScript 的更新和改进很多都基于我个人在科研工作中的痛点和需求

CovScript 为前沿研究赋能：Biomedical Information 生物学信息整体解决方案（四川大学华西医院）

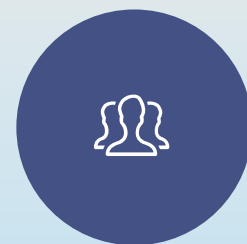
- 网页前端：基于 Vue.js 搭建
- 网页后端：基于 Apache2 CGI 引擎 + CovScript 3 + CSDBC
- 容器服务：
 - 基于 CovScript 实现了容器调度和负载均衡
 - CovScript 实现 Container Daemon，监控容器运行状态、收集运行结果
- 数据库：
 - 基于 CovScript 实现了数据库的多机备份和 Data Auto-Retriever
- 数据分析：
 - 基于 CovAnalysis 实现了 CSV 的高效解析

CovScript 为前沿研究赋能：AIGC 大语言模型 Evaluator (KDDE Lab)

- GitHub: <https://github.com/scukdde-llm/llm-evaluators>
- 基于 CovScript 4 实现了训练数据的转化和随机抽取
- 基于 CovAnalysis 实现了评估结果的收集、处理和分析

CovScript 应用案例

Chapter 4



Showcase 1: 使用 CovScript Netutils 框架搭建简单的网页

- 请先注册、登录 GitHub 后在浏览器中打开：
 - <https://codespaces.new/covscript/codespace?quickstart=1>
 - 请确保网络连接顺畅（必要时使用专线）
- 待 Codespace 加载完毕后，网页将在 5000 端口呈现
- 修改 hello.ecs 后，在终端中先使用 Ctrl + C 关闭正在运行的应用，重新键入 ecs hello.ecs 即可查看修改后的效果
- 若需使用中文，请在文件开头插入一行 @charset:utf8
- session.send_response(响应代码, 响应数据, 数据类型)
 - 一般只需要修改响应数据，如遇问题可以开启一个新的 Codespace

Showcase 2: 使用 CovAnalysis 框架实现数据合并以及缺失值补全

- 请先注册、登录 GitHub 后在浏览器中打开：
 - <https://codespaces.new/covscript/data-analysis-showcase?quickstart=1>
 - 请确保网络连接顺畅（必要时使用专线）
- 待 Codespace 加载完毕后，系统将自动运行并输出结果到 output 文件夹中，可以直接使用 Codespace 可视化查看
- 不建议修改程序代码，可以修改 run.sh 中的参数，对比不同缺失值补全方法的差别
 - `ecs fix_abnormal.ecs 5(Top-k) cosine(距离度量)`
 - Top-k 必须大于 1，一般不超过 10；距离度量还可以是 euclidean
 - 在终端中运行 run.sh 重新生成：`bash run.sh`

受限于演讲时间，很遗憾无法直接演示
欢迎大家明天(10.29)下午2点到开源百宝箱
分论坛中参与我们组织的 CovScript 工坊活
动，届时我会详细介绍这些应用案例
也可以到开源市集中我们的摊位上与我们
进一步沟通



THANK YOU

后生拙见，望各位前辈多多指导

主页: <https://covscript.org.cn/>

GitHub: mikecovlee

WeChat: mikecovlee

E-Mail: lee@unicov.cn



扫码关注开源社公众号



扫码添加讲师联系方式