

2017 年以来 python 在 AI 项目上的应用前沿论文最新进展

2018.11.06 方建勇

提示: 采用手机 safari 微软翻译技术

1. 建议: 1810.13310[[pdf](#),[其他](#)] cse

web 应用程序 python 包中漏洞的实证分析

作者:[jukka ruohonen](#)

摘要: 本文研究了用于 web 开发的常用 python 包中的软件漏洞。经验数据集基于 pypi 包存储库和所谓的安全数据库, 用于跟踪存储库中选定包中的漏洞。方法方法的基础是对包的释放易受攻击的条件概率进行基于发布的时间序列分析。根据研究结果, 观察到的许多 python 漏洞似乎只是轻微地严重; 输入验证和跨站点脚本一直是最典型的漏洞。就基于释放历史的时间序列分析而言, 只观察到最近的过去与统计预测有关; 古典马尔科夫物产举行。少

2018 年 10 月 31 日提交; 最初宣布 2018 年 10 月。

评论: 即将出版: 第九届经验软件工程实践国际研讨会论文集 (iwesep 2018), 奈良县

2. 建议: 1810.12297[[pdf](#),[其他](#)] Cs。直流

可分割性注释: 优化现有库中的黑盒功能组合

作者:[shoumik palkar](#), [matei zaharia](#)

摘要: 数据移动是并行数据密集型应用程序中的主要瓶颈。针对这一问题, 研究人员提出了新的运行时和中间表示 (ir), 这些方法在现有库 api 下应用了循环融合等优化。尽管这些运行时通常不需要更改用户代码, 但它们需要对库本身进行侵入性更改: 通常, 所有库函数都需要为新的 ir 或虚拟机重写。在本文中, 我们提出了一种新的抽象称为可拆分注释 (sa), 它支持对黑盒库函数进行关键数据移动优化。sa 只要求用户为现有的、未修改的函数添加批注, 并实现一个小 api 来拆分库中的数据值。此接口一起描述如何对函数之间传递的值进行分区, 以启用数据流水线和自动并行化, 同时尊重每个库的正确性约束。我们在一个叫做莫扎特的系统中实现 sa。在 python 中使用 numpy 和 pandas 的工作负载上, 如果不修改任何库功能, moztart 就能提供性能竞争力, 而侵入式解决方案在许多情况下需要重写库, 有时可以改进在过去的系统上的性能高达 2 倍, 并将工作负载加速了 30x。少

2018 年 10 月 29 日提交; 最初宣布 2018 年 10 月。

3. 第 1810.11530[[pdf](#),[其他](#)] Cs。Lg

ml 中的自动区分: 我们在哪里, 我们应该去哪里

作者:[bart van merriënboer](#), [olivier breuleux](#), [amaud bergeron](#), [pascal lamblin](#)

摘要: 我们回顾了机器学习 (ml) 中阵列编程的自动微分 (ad) 的现状, 包括用于 ad、基于图形的中间体的运算符重载 (oo) 和源转换 (st) 等不同的方法程序和源语言的表示形式。基于这些见解, 我们引入了一种新的基于图形的中间表示 (ir), 其具体目的是为阵列编程有效地支持完全通用的 ad。与 ml 框架中的现有数据流编程表示不同, 我们的 ir 自然支持函数调用、高阶函数和递归, 使 ml 模型更易于实现。表示闭包的能力使我们能够在没有

磁带的情况下使用 `st` 执行 `ad`, 使生成的导数 (伴随) 程序适合使用功能语言编译器中的工具进行提前优化, 并启用高阶导数。最后, 我们引入了一个名为 `myia` 的概念编译器工具链的证明, 它使用 `python` 的子集作为前端。少

2018 年 10 月 26 日提交;最初宣布 2018 年 10 月。

4. 第 1810.1268[[pdf](#),其他] [cs.PL](#)

自动并行: 用于自动并行化和仿射环嵌套分布式执行的 `python` 模块

作者:[克里斯蒂安·拉蒙-科尔特斯](#)、[拉蒙·阿米拉](#)、[豪尔赫·埃吉尔克](#)、[菲利普·克劳斯](#)、[罗莎·巴迪亚](#)

摘要: 编程语言、编程模型和框架的最后改进侧重于从许多编程问题中抽象用户。除其他外, 最近的编程框架包括更简单的语法、自动内存管理和垃圾收集, 它简化了通过库包重新使用代码的过程, 以及易于配置的部署工具。例如, `python` 由于其语法的简单性, 已经上升到编程语言列表的首位, 同时即使是一种解释语言, 它仍然取得了良好的性能。此外, 社区还帮助开发了大量的库和模块, 对它们进行了调整, 以获得出色的性能。但是, 在阻止用户直接处理分布式和并行计算问题方面仍有改进的余地。本文提出并评估了一个 `python` 模块 -- **-auto 并行**, 以自动找到合适的基于任务的仿射环嵌套并行化, 在分布式计算基础结构中并行执行它们。这种并行化还可以包括构建数据块, 以增加任务粒度, 从而实现良好的执行性能。此外, `auto 并行` 基于顺序编程, 并且仅包含 `python` 装饰器形式的一个小注释, 以便任何没有编程技能的人都可以将应用程序扩展到数百个内核。少

2018 年 10 月 26 日提交;最初宣布 2018 年 10 月。

评论:参加第 8 届 `python` 高性能和科学计算研讨会 (pyhpc 2018)

5. 第 1810.11190[[pdf](#),其他] [Cs. CI](#)

规模: 一种快速、高效的通用矢量嵌入实用程序包

作者:[ajay patel](#)、[亚历山大 sands](#)、[chris callison-burch](#)、[marianna apidianaki](#)

摘要: 矢量空间嵌入模型, 如 `word2vec`、`环球`、`快速文本`和 `elmo`, 是自然语言处理 (nlp) 应用中非常流行的表现形式。我们介绍了 "规模", 这是一种快速、轻量级的工具, 用于利用和处理嵌入。重要性是一个开源的 `python` 包, 具有紧凑的矢量存储文件格式, 允许高效地操作大量的嵌入。规模执行常见操作的速度比 `gensim` 快 60 到 6,000 倍。重要性引入了几个新功能, 以提高鲁棒性, 如词汇外查找。少

2018 年 10 月 26 日提交;最初宣布 2018 年 10 月。

6. 建议: 1810.09538[[pdf](#),其他] [Cs. Lg](#)

pyro: 深层通用概率编程

作者:[eli bingham](#)、[jonathan p. chen](#)、[martin jankowiak](#)、[fritz obermeyer](#)、[neeraj pradhan](#)、[theofanis karaletsos](#)、[rohit singh](#)、[paul szerlip](#)、[paul horsfall](#)、[诺亚·古德曼](#)

摘要: `pyro` 是一种基于 `python` 的概率编程语言, 它是在 `ai` 研究中开发高级概率模型的平台。为了扩展到大型数据集和高维模型, `pyro` 使用随机变分推理算法和概率分布建立在 `pytorch` 之上, 这是一个现代 `gpu` 加速的深度学习框架。为了适应复杂或特定于模型的算法行为, `pyro` 利用 `outline` (一个可组合构建块库) 来修改概率程序的行为。少

2018 年 10 月 18 日提交;最初宣布 2018 年 10 月。

评论:提交给 `jmlr mloss` 轨道

7. 建议: 1810.09311[[pdf](#),其他] [Cs. CI](#)

重新求解分布式通信索引: python 重新实现与新实验

作者:[alejandro moreo](#), [andrea esuli](#), [fabrizio sebastiani](#)

文摘: 本文介绍了 pydci, 一种用 **python** 编写的分布式通信索引 (dci) 的新实现。dci 是一种跨域和跨语言文本分类的转移学习方法, 我们为其提供了一个基于 jatecs (用于文本分类的 jatecs 框架) 的实现 (此处称为 jadci)。pydci 是一个独立版本的 dci, 利用科学学习和科学堆栈。我们在这里报告了我们为测试 pydci 而进行的新实验, 在这些实验中, 我们使用 dci 最初提出后出现的新的性能方法作为基线。这些实验表明, 由于我们改进了 dci 的一些微妙方法, pydci 的性能优于 jadci 和上述高性能的方法, 并在我们测试 dci 的两个流行基准上提供了最著名的结果, 即: 多域感知 (又名 mds----用于跨域适应) 和 webis-cls-10 (用于跨语言适应)。pydci 以及允许复制实验的代码可在 <https://github.com/AlexMoreo/pydci>。少

2018 年 10 月 19 日提交;最初宣布 2018 年 10 月。

8. 第: 1810.09065[[pdf](#),[其他](#)] Cs. 铬

高级加密抽象

作者:[克里斯托弗·凯恩](#),[林波](#),[沙沙昌](#), [严红 a. 刘](#)

摘要: 常用的加密库公开的接口是笨拙的、复杂的, 并且假定对加密算法的理解。本文提出了由简单的加密基元和声明性配置组成的高级抽象。这些抽象可以在任何语言的任何加密库之上实现。我们在 **python** 中实现了这些抽象, 并使用它们编写了各种众所周知的安全协议, 包括 signal、kerberos 和 tls。我们表明, 与使用低级库相比, 使用我们抽象的程序要小得多, 编写起来也容易得多, 并且可以安全地抵御文献中报告的绝大多数加密滥用行为。当使用我们的抽象编写时, 安全协议实现的大小平均减少了三分之一左右。我们还表明, 我们的实现产生的开销很小, 共享密钥操作的开销小于 5 微秒, 公钥操作的开销小于 341 微秒 (& lt; 1%)。少

2018 年 10 月 21 日提交;最初宣布 2018 年 10 月。

评论:23 页, 3 个数字

9. 第 1810.08061[[pdf](#),[其他](#)] cs.PL

自动图: 基于图形的性能的固定式编码

作者:[dan moldovan](#), [james m decker](#), [fei wang](#), [andrew a johnson](#), [brian klee](#), [zachary nado](#), [dsculley](#), [tiark rompf](#),[亚历山大 b wiltschko](#)

摘要: 在易于编写的机器学习代码和可扩展或快速执行的机器学习代码之间存在着明显的权衡。在机器学习中, 像 autograd 和 pytorch 这样的命令式样式库很容易编写, 但具有较高的解释开销, 并且不容易在生产或移动设置中部署。像 tensorflow 和 theano 这样的基于图形的库受益于整个程序优化, 可以进行广泛的部署, 但这使得表达复杂模型变得更加繁琐。我们描述了在 **python** 中使用分阶段编程如何通过源代码转换在这两种库设计模式之间提供一个中点, 抓住了这两种模式的好处。一个关键的洞察是通过动态调度将所有依赖于类型的决策推迟到运行时。我们在 autograph 中实例化了这些原则, 这是一个软件系统, 它改进了 tensorflow 库的编程体验, 并演示了可用性改进, 与本机的 tensorflow 图相比, 性能没有损失。我们还展示了我们的系统是后端不可知论的, 并演示了针对具有拉伸流图中找不到的特性的备用红外。少

2018 年 10 月 16 日提交;最初宣布 2018 年 10 月。

10. 建议: 1810.07791[[pdf](#),[其他](#)] cs. cy

maasim: 提高城市生活质量的宜居性仿真

作者:[dominika woszczyk](#), [gerasimos spanakis](#)

摘要: 由于拥有强大的模型和 3d 仿真平台, 城市化不再是纸上谈兵。然而, 目前的工作并不对公众开放, 缺乏有助于决策的优化机构。本文介绍了基于现有荷兰人的宜居性评分和内置 ai 模块的开源模拟的创建过程。使用要素工程和随机林选择特征。然后, 在以前宜居性类的基础上, 建立了一个修正的评分函数。使用随机森林进行回归预测, 并通过 10 倍交叉验证实现了 0.83 的召回。然后, 应用探索性因子分析来选择模型中存在的动作。由此产生的指标分为 5 组, 并产生 12 项行动。比较了 n 化二、paes、spea2 和 eps-moea 这四种优化算法在三个既定质量标准上的性能: 基数、解的传播、间距以及由此产生的分数和转数。虽然这四种算法都表现出不同的优势, 但 eps-moea 被选择为最适合这个问题的算法。最后, 模拟将模型和选定的 ai 模块集成到 python 的 kfy 框架中编写的 gui 中。对用户进行的测试显示了积极的反应, 并鼓励采取进一步的举措, 将技术和公共应用程序结合起来。少

2018 年 10 月 13 日提交;最初宣布 2018 年 10 月。

评论:16 页

11. [建议: 1810.0791](#)[pdf,其他] cs.PL

基于任务的运行时系统上 python 代码的异步执行

作者:[r. tohid](#), [bibek wagle](#), [shahrazad shirzad](#), [patrick diehl](#), [adrian serio](#), [alireza kheirhahan](#), [parsa amini](#), [k 佳 williams](#), [kate isaacs](#), [kevinhuck](#), [steven brbrt](#), [Hartmut kaiser](#)

摘要: 尽管在并行计算和分布式计算领域取得了进展, 但在高性能计算 (hpc) 资源上编程的复杂性使许多领域的专家望而却步, 尤其是在机器学习和人工智能 (ai) 领域, 从利用这些系统的性能优势。研究人员和科学家倾向于高生产率语言, 以避免低级别语言编程带来的不便, 以及获得在这一级编程所需的必要技能的成本。近年来, python 在 numpy 等线性代数库的支持下, 尽管面临着阻止分布式运行此代码的限制, 但仍越来越受欢迎。在这里, 我们提出了一个既保持高级编程抽象又保持并行和分布式效率的解决方案。phylanx 是一个异步数组处理工具包, 它将 python 和 numpe 操作转换为代码, 通过将 python 和 numpy 函数和变量映射到依赖关系树中, 可以在 hpc 资源上并行执行由 hpx 执行, hpx 是一个通用的、并行的、基于任务的、基于任务的运行时系统, 用 c++ 编写。phylanx 还提供了用于调试和性能分析的内省和可视化功能。我们通过将广泛使用的机器学习算法的实现与公认的 numpy 标准进行比较, 测试了我们方法的基础。少

2018 年 10 月 22 日提交;v1 于 2018 年 10 月 17 日提交;最初宣布 2018 年 10 月。

12. [建议: 1810.07277](#)[pdf,其他] Cs. Ce

基于分子动力学的冷喷涂过程闭环图像辅助优化

作者:[程振兴](#),[王虎](#),[刘桂荣](#)

文摘: 为了提高冷喷涂过程中的沉积质量, 提出了一种闭环图像辅助优化 (cliao) 方法。最近的一些研究表明, 粘结颗粒的扁平比测量的沉积质量与撞击速度、角度和粒径有关。因此, cliao 的初衷是通过获得直接从分子动力学 (md) 模拟快照中提取的最大扁平比来提高沉积质量。若要完成此策略, 建议使用 python 脚本自动从结果文件生成所需的快照, 并使用图像处理技术来评估快照的展平比率。此外, 还采用了代理优化 (高效全局优化) 和启发式算法 (粒子群优化、不同进化算法) 三种优化方法。然后利用反向传

播神经网络 (bpnn) 加速优化过程, 利用 bpnn 建立元模型而不是正向计算。优化结果表明, 上述方法都能得到可接受的解决方案。还给出了这些方法之间的比较, 其选择应通过效率和准确性之间的权衡来确定。少

2018 年 10 月 16 日提交;最初宣布 2018 年 10 月。

评论:15 页, 4 张表格, 17 个数字

13. **建议: 1810.06599**[pdf, ps,其他] cse

多伊 [10.114/3283812.3283822](#)

使用 cmg 从源代码生成注释

作者:sergey matskevich, colin s. gordon

摘要: 好的注释可以帮助开发人员更快地理解软件并提供更好的维护。然而, 评论往往缺失, 一般不准确, 或过时。其中许多问题可以通过自动生成注释来避免。本文提出了一种利用自然语言处理中的通用技术直接从源代码生成信息性注释的方法。我们使用现有的自然语言模型生成注释, 该模型将单词与其各自的逻辑含义和语法规则结合起来, 允许通过从程序文本的声明性描述中进行注释生成。我们在 python 中实现的几个经典算法上评估我们的算法。少

2018 年 10 月 15 日提交;最初宣布 2018 年 10 月。

评论:nl4se 2018 预印

14. **第 1810.05281**[pdf,其他] cs. ne

iohiler: 迭代优化启发式的基准和分析工具

作者:caro a doerr, hao wang, furong ye, sander van rij, thomas bäck

文摘: iohprofiler 是分析和比较迭代优化启发式的一种新工具。作为用 c 或 python 编写的输入算法和问题, 它通过在固定目标运行时间和固定预算函数值上的分布, 提供了对算法性能的统计评估。此外, iohprofiler 还允许跟踪算法参数的演变, 使我们的工具特别适用于 (自适应) 算法的分析、比较和设计。iohprofiler 是一种随时可用的软件。它由两个部分组成: 一个是生成运行时数据的实验部分, 另一部分是后处理部分, 用于进行汇总比较和统计评价。实验部分建立在 coco 软件的基础上, 并对其进行了调整, 以应对作为函数的优化问题 $f: \mathbb{N} \rightarrow \mathbb{R}$ 与 s 是一个整数的离散字母表。后处理部分是我们自己的工作。它可作为评估任意基准问题运行时间数据的独立工具。它不仅接受 IOHprofiler 的输出文件作为输入文件, 而且还接受原始的 coco 数据文件。后处理工具是为交互式评估而设计的, 允许用户根据自己的需要选择所显示数据的范围和精度。iohprofiler 可在 github 上查阅, 网址是 [//github.com/IOHprofiler](https://github.com/IOHprofiler)。少

2018 年 10 月 11 日提交;最初宣布 2018 年 10 月。

15. **第: 1810.4825**[pdf] cse

多伊 [10.1109/ACCESS.2018.2872669](#)

基于 github 存储库的器官移植初探

作者:王尚文,毛晓光,俞觉敏

摘要: 器官移植是利用与某些特定功能直接相关的代码来完成自己的程序, 它比传统的组件重用为开发人员提供了更多的便利。然而, 最近的技术面临着缺乏移植器官的挑战。因此, 我们对从 github 存储库中提取器官进行了实证研究, 以探索基于大规模数据集的移植。我们分析了 12 个具有代表性的 github 项目的统计数据, 得出的结论是:

1) 在提交中存在大量的实际器官,并在评论中添加一个关键词;2) 本资料库中的机关主要有四种内容;3) 大约 70% 的器官易于移植。实施不同类型器官的移植策略,我们用三种不同的编程语言 (java、**python** 和 c) 手动提取 30 个器官,并利用四种测试工具 (两个用于 java, 一个用于 java, 一个 **python**, 一个为 c)。最后,我们将三个 java 器官移植到一个特定的平台上进行性能检查,以验证它们是否能在新系统中正常工作。我们的策略提取的 30 个器官在单元测试中都有良好的性能,最高合格率达到 97%,最低的器官仍然通过 80%,三个 java 器官在新系统中工作良好,为主机提供了三个新的功能。结果表明,基于开源存储库的器官移植是可行的,为代码重用带来了新的思路。少

2018 年 10 月 10 日提交;最初宣布 2018 年 10 月。

评论:14 页

16. [建议: 1810.04317\[pdf,其他\]](#) lo c

[多伊](#) 10.4204/EPTCS.280.11

smtlink 2. 0

作者:[严鹏](#),[马克 r. 绿街](#)

摘要: smtlink 是 acl2 的扩展,具有令人满意的模态理论 (smt) 求解器。我们在 acl22015 上推出了早期版本。smtlink 2.0 在健全性、可扩展性、易用性以及支持的类型和相关理论求解器的范围方面与初始版本相比有了重大改进。大多数需要证明使用 smt 求解器的定理必须首先被翻译为只使用 smt 求解器支持的原始操作--这种转换包括函数扩展和类型推断。smtlink 2.0 使用由经过验证的子句处理器和计算提示执行的一系列步骤来执行此转换。这些步骤是健全的。从 acl2 到 **z3 python** 接口的最终转写需要一个受信任的子句处理器。与最初的 smtlink 相比,这在健全性和可扩展性方面有了很大的改进,而 smtlink 是作为一个单一的、整体的、受信任的子句处理器实现的。smtlink 2.0 通过使用 z3 的数组和用户定义的数据类型,为 fty defprod、平减、反犹太主义和 defoption 类型提供支持。我们已经确定了常见的使用模式,并简化了使用 smtlink 所需的配置和提示信息。少

2018 年 10 月 9 日提交;最初宣布 2018 年 10 月。

评论:在记录 acl2 2018, arxiv:1810.3762

日记本参考:ettcs 280, 2018, 143-160 页

17. [第 1810.02835\[pdf\]](#) Cs. 简历

公路交通视频车辆分割背景建模方法的比较

作者:[l. a. marcomini](#), [a. l. cunha](#)

文摘: 本文的目的是比较三种背景建模算法在公路交通视频车辆分割和检测中的性能。所有算法都在 opencv 中可用,并且都在 **python** 中进行了编码。我们分析了七个视频,总共录制了 2 个小时。为了比较这些算法,我们创建了 35 张地面真相图像,每个视频中创建了 5 个,我们使用了三个不同的指标:准确率、准确率和处理时间。通过使用准确性和精确性,我们的目标是确定算法在检测和分割方面的性能,同时利用处理时间来评估对计算系统的影响。结果表明,这三种算法的准确率均在 90% 以上,而平均准确率为 80%。该算法对处理时间的影响最小,允许每秒计算 60 帧。少

2018 年 10 月 5 日提交;最初宣布 2018 年 10 月。

评论:12 页, 11 个数字, 1 个表

18. 第 1810.2561[[pdf](#), [ps](#),其他] cs. ms

gpdoemd: 用于设计模型判别实验的巨蟒包

作者:[simon olofsson](#), [ruth misener](#)

摘要: gpdoemd 是一个开源巨蟒包, 用于设计模型判别实验, 使用高斯过程替代模型来近似和最大化边际预测之间的差异竞争对手机械模型的分布。gpdoemd 使用发散预测来建议一个最大信息的下一个实验。

2018 年 10 月 5 日提交;最初宣布 2018 年 10 月。

19. 建议: 1810.043[[pdf](#),其他] Cs. 艾

ai 公平 360: 用于检测、理解和减轻不需要的算法偏差的可扩展工具包

作者:[rachel k. e. bellamy](#), [kuntal dey](#), [michael hind](#), [samuel c.hoffman](#), [stephanie houde](#), [kalapriya kannan](#), [pranay lohia](#), [jjelyn martino](#), [sameep Lohia](#), [Aleksandra mojsilovic](#), [seema nar](#), [karthikeyan natesan ramamurthy](#), [john richards](#), [Diptikalyansaha](#), [prasanna sattigeri](#), [moninder singh](#), [kush r. varshney](#), [yufeng zhang](#)

摘要: 公平是一个越来越重要的问题, 因为机器学习模式被用来支持抵押贷款、招聘和监狱判决等高风险应用的决策。本文介绍了一个新的开源 **python** 算法公平工具包 **ai fairness 360** (aif360), 在 **apachv2.0** 许可证 (<https://github.com/ibm/aif360>) 下发布。该工具包的主要目标是帮助促进公平研究算法过渡到工业环境中使用, 并为公平研究人员共享和评估算法提供一个共同框架。该软件包包括一组针对数据集和模型的全面公平指标、这些指标的解释以及减轻数据集和模型偏差的算法。它还包括交互式 **web** 体验 (<https://aif360.mybluemix.net>), 为业务线用户提供了概念和功能的温和介绍, 以及广泛的文档、使用指南和特定行业教程, 使数据科学家和从业人员能够将最适合他们的问题的工具纳入他们的工作产品。该包的体系结构经过设计, 符合数据科学中使用的标准范式, 从而进一步提高了从业人员的可用性。这样的架构设计和抽象使研究人员和开发人员能够使用他们的新算法和改进扩展工具包, 并将其用于性能基准测试。内置的测试基础结构维护代码质量。少

2018 年 10 月 3 日提交;最初宣布 2018 年 10 月。

评论:20 页

20. 新建: 1810.00905[[pdf](#),其他] cs.PL

codestischer: 程序间基本块布局优化

作者:[rahman lavaee](#), [john criswell](#), [chen ding](#)

摘要: 现代软件执行大量的代码。以前的代码布局优化技术是在一、二十年前开发的, 已经不足以应对新类型应用程序 (如编译器、浏览器、解释器、语言虚拟机和共享库) 的规模和复杂性。本文介绍了 **cod** 着想彻, 它是一个跨过程的基本块代码布局优化器, 它重新排序可执行文件中的基本块, 以便从更好的缓存和 **tlb** 性能中获益。

codetitcher 提供了一个层次结构框架, 可用于改善内存层次结构各层中的局部性。我们的评估显示, **Codestitcher** 在具有大代码大小的 5 个广泛使用的应用程序上将原始程序的性能提高了 3% 至 25% (平均为 10%): **mysql**、**clang**、**firefox**、**apache** 和 **python**。它提供了比 **llvm** 的 **pgo** 增加 4% 和超过 3% 的 **pgo** 结合最佳的功能重新排序技术。少

2018 年 10 月 1 日提交;最初宣布 2018 年 10 月。

评论:24 页, 6 个数字, 预打印

21. 第: 1810.00674[[pdf](#),其他] cs. ms

使用 sfepi 的 python 中的多尺度有限元计算

作者:robert cimrman, v 首屈一指 Lukeš, edward rohan

摘要: spf 无穷无尽 (python 中的简单有限元) 是用有限元法求解一个、两个或三个空间维度中的偏微分方程所描述的各种问题的软件。它的源代码大多 (85%)python 和依赖于 num 居包提供的快速矢量化操作。对于特定的问题, 可以使用两个接口: 一个声明性应用程序编程接口 (api), 其中问题描述/定义文件 (python 模块) 用于定义计算, 以及一个命令式 api, 可用于交互式命令, 或在脚本和库中。在概述了 sfep 封装开发的基础上, 介绍了其实现、结构和一般特点。利用一个简单的热传导问题的例子描述了定义偏微分方程的分量。具体来说, 在示例中显示了 sfep 的声明性 api。为了说明 sfep 的主要资产之一, 提出了基于均质化理论的复杂多尺度模型实现框架, 并给出了一个二尺度压电弹性模型的实例, 既显示了问题的数学描述, 又展示了该问题的数学描述。相应的代码。少

2018 年 10 月 1 日提交;最初宣布 2018 年 10 月。

评论:此手稿版本可在 cc-by-nc-nd 4.0 许可证下提供

msc 类: 35qxx;65n30;65m60;65y05;74s05

22. 第 1809.10024[[pdf](#)] cs. cy

神经成像中可重复数据分析的计算与信息学研究进展

作者:russell a. poldrack, krzysztof j. gorgolewski, gael varoquaux

摘要: 科学研究的重现性已成为一个值得关注的问题。我们认为, 公开和透明对于重现性至关重要, 我们概述了人类神经成像社区中出现的开放透明科学生态系统。我们讨论了为神经成像数据开发的开放数据共享资源的范围, 以及数据标准 (特别是大脑成像数据结构) 在实现大型神经成像的自动共享、处理和重用方面的作用数据。我们概述了开源 python 语言如何为数据科学平台提供基础, 该平台可实现可重复的数据分析和可视化。我们还讨论了软件工程的新进展 (如容器化) 如何为数据分析的更高重现性提供基础。这一新生态系统的出现为许多科学领域提供了一个例子, 这些领域目前正在努力实现可重复性。少

2018 年 9 月 24 日提交;最初宣布 2018 年 9 月。

23. 第 1809. 09569[[pdf](#),其他] Cs. Lg

切线: 使用源代码转换进行动态类型化数组编程的自动微分

作者:bart van merriënboer, dan moldovan, 亚历山大 b wiltschko

摘要: 在 python 中表示的日益复杂的模型的一阶和高阶导数的需要强调或超过了可用工具的能力。在这项工作中, 我们从自动分化 (ad) 领域探索了能够给研究人员提供表现力、性能和强大可用性的技术。其中包括源代码转换 (sct)、灵活梯度手术、高效的就地阵列运算、高阶导数以及正向和反向模式 ad 的混合。我们在 python 的切线软件库中实现并演示了这些想法, 这是使用 sct 的动态语言的第一个 ad 框架。少

2018 年 9 月 26 日提交;v1 于 2018 年 9 月 25 日提交;最初宣布 2018 年 9 月。

24. 第 1809. 08311[[pdf](#),其他] Cs. pf

范围: c3sr 系统表征和基准框架

作者:carl pearson, abdul dakkak, cheng li , sarah hashash, jinjun xong, wen-meu

摘要: 本报告介绍了可扩展和可移植基准测试的 scope 基础结构的设计。高性能计算系统的改进依赖于不同级别的系统抽象之间的协调。在系统层次结构的所有级别开发和定义准确的性能度量是必要的, 并且对于不同背景的开发人员应尽可能方便访问。scope 项目旨在通过提供允许独立开发基准的软件体系结构、提供有用的 csc++ 抽象和实用程序以及提供 python 包, 用于生成生成的测量的发布质量图。少

2018 年 9 月 18 日提交;最初宣布 2018 年 9 月。

评论:8 页, 草稿

25. 第: 189.07258[pdf, ps,其他] Cs. Lg

dppy: 使用 python 采样确定点进程

作者:guillaume gautier, rémi bardenet , mumal valko

摘要: 确定点进程 (dpp) 是点云上的特定概率分布, 这些点被用作物理、概率、统计和最近的机器学习的模型和计算工具。从 dpp 进行采样是一项挑战, 因此我们提出了 dppy, 这是一个 python 工具箱, 收集已知的精确和近似采样算法。该项目设在 github 上, 并配备了大量的文件。本文档采用对 dpp 进行简短调查的形式, 并将每个数学属性与 dppy 对象联系起来。少

2018 年 9 月 19 日提交;最初宣布 2018 年 9 月。

评论:<http://github.com/guilgautier/DPPy/> 文档中的代码在

<http://dppy.readthedocs.io/>

26. 第 xiv: 1809.06520[pdf, ps,其他] cs. ms

r 的随机问题

作者:kellie ottoboni, phillip b. stark

摘要: r (版本 3.5.1 已修补) 的随机抽样功能存在两个问题。首先, 它使用了已知有种子问题的默森-扭曲的一个版本, 2002 年《默森-特维斯特》的作者纠正了这一版本。更新后的 c 源代码可在

<http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/MT2002/CODES/mt19937ar.c>。

第二, r 在 1 和米通过将随机浮动乘以米, 并采取的发言, 并添加 1 到结果。在这种方

法中众所周知的量化效应导致了一个不均匀的分布{1,那是什么? ,米}.差异, 这取

决于米, 可以是实质性的。由于 r 中的采样函数依赖于生成随机整数, 因此 r 中的随机抽样是偏置的。有一个简单的解决方法: 直接从随机位构造随机整数, 而不是将随机

浮点乘以米.这就是 `python numpy.random.randint ()` 函数等中采取的策略。python

中的示例源代码可在

<https://github.com/statlab/cryptorandom/blob/master/cryptorandom/cryptorandom.py> (请参阅函数 `getrandbits ()` 和 `randbelow _` 来自 `_randbelow ()`)。少

2018 年 9 月 19 日提交;v1 于 2018 年 9 月 17 日提交;最初宣布 2018 年 9 月。

27. 第 1809.0574[pdf] Cs. 简历

面部分析的实时系统

作者: [janne tommola](#), [pedram ghazi](#), [biswo adhikari](#), [heikki hhtunen](#)

摘要: 本文描述了实时面部分析系统的解剖结构。该系统识别年龄, 性别和面部表情从用户出现在镜头前。所有组件都是基于卷积神经网络的, 我们研究了卷积神经网络在常用训练和评价集上的准确性。这项工作的一个关键贡献是描述了帧抓取、人脸检测和三种类型识别的处理线程之间的相互作用。执行该系统的巨蟒代码使用普通的库 --kerasensorloox、opencv 和 dlib--and 可供下载。少

2018 年 9 月 14 日提交;最初宣布 2018 年 9 月。

评论:提交给 euvip2018 会议

28. 第 [xiv:1809.04166](#)[pdf,其他] cs. ne

leabra7: 用于建模反复的、生物现实的神经网络的 python 包

作者: [c. daniel greenidge](#), [noam miller](#), [kerneth a. norman](#)

摘要: 急诊是一个软件包, 它使用 adex 神经动力学模型和 leabra 学习算法, 以生物现实的方式模拟和训练任意复发的神经网络架构。我们介绍了 leabra7, 这是一个实现这些相同算法的互补 python 库。leabra7 是使用现代软件开发原则开发和分发的, 并与 python 的科学堆栈紧密集成。我们演示了使用传统模式关联任务和标准机器学习任务的经常性 leabra7 网络, 对 iris 数据集进行了分类。少

2018 年 9 月 19 日提交;v1 于 2018 年 9 月 11 日提交;最初宣布 2018 年 9 月。

评论:修复小拼写错误

29. 第 [xiv:1809.03609](#)[pdf] Cs. 简历

urban-i: 从城市场景到利用深度学习和计算机视觉绘制城市贫民窟、交通方式和行人的地图

作者: [mohamed r. ibrahim](#), [james haworth](#), [tao cheng](#)

摘要: 在不同科学领域深度学习和计算机视觉的迅速扩展中, 在城市发展方面, 深度学习和计算机视觉应用仍然局限于智能城市和自主的概念车辆。事实上, 在欠发达国家的城市和城市地区, 知识差距很大, 在这些国家, 非正式的混乱是主要的计划。深度学习和人工智能 (ai) 如何解开非正式的复杂性, 以推进城市建模和我们理解城市? 在大赦国际和计算机视觉的范式中, 可以就北方和南方城市的未来提出各种问题和辩论。本文介绍了一种利用深度学习和计算机视觉的多功能现实动态城市建模方法, 利用深卷神经网络 (cn) 从空中和空中感知和检测城市场景中的信息和贫民窟。除探测行人和交通方式外, 还可拍摄街景图像。该模型已在全球城市的城市场景图像上进行了培训。该模型很好地验证了对包括非正规和贫民窟地区在内的有计划地区和无计划地区之间各种细微差别的理解。我们试图推进城市建模, 以更好地了解城市发展的动态。我们还旨在举例说明 ai 在城市中的重大影响, 而不仅仅是在主流中如何讨论和感知智慧城市。urban-i 模型的算法在 python 编程中完全编码, 并采用预先训练的深度学习模型, 作为全球各个角落, 包括非正规住区和贫民窟地区的制图和城市建模工具。少

2018 年 9 月 10 日提交;最初宣布 2018 年 9 月。

评论:12 页, 9 个数字

30. 第 [1809.001913](#)[pdf,其他] Cs. Lg

高斯过程 (gp) 的实际经验: 在 python-i 中实现全科医生

作者: [kshitij tiwari](#)

摘要: 这份文件是对我们网站的补充, 我们的网站是为了让学生接触高斯过程 (gp) 而开发的。gp 是非参数贝叶斯回归模型, 主要用于统计人员和地理空间数据科学家对空间数据建模。来自 matlab [1]、python [2]、r [3] 等的多个开源库已经可用于简单的插件和使用。这个讲义的目的, 反过来是允许用户依靠最小的外部依赖性在 python 中开发独立的全科医生。为此, 我们只使用默认的巨蟒模块, 并协助用户从零开始开发自己的全科医生, 让他们深入了解引擎盖下发生的事情。该模块涵盖了使用最大似然估计 (mle) 的 gp 推理, 并给出了一维 (虚拟) 空间数据的示例。少

2018 年 9 月 6 日提交;最初宣布 2018 年 9 月。

评论:34 页, 7 个数字

31. 第 09iv:1800.00976[pdf] cs.PL

控制网络编程中的语言互操作性

作者:kostadin kratchanov, efe ergün

摘要: 控制网络编程 (cnp) 是一种编程范式, 用 "原语 + 控制网络 = 控制网络程序" 来描述。它是一种图形编程。控制网络是一个递归系统的图形;它可以是对正在解决的问题的纯粹描述性的规范。显然, "绘制" 控制网络不包括任何编程。原语是基本的、易于理解的、明确指定的操作。最终, 它们必须被编程。从历史上看, 它们通常是用自由 pascal 编码的。基元的实际代码从未被认为是重要的。"算法" 的本质是由其控制网络来表示的。cnp 一直是一种简单、快速的软件应用程序开发方法, 实际上几乎不涉及实际编程。语言互操作性 (在同一软件项目中使用不同的语言) 是当前软件开发的一个显著趋势。在 cnp 的情况下, 它比其他编程范式更重要、更自然。在这里, 互操作性实际上意味着可以使用各种编程语言编写的基元。本报告介绍了我们使用多语言基元集创建应用程序的第一步。大多数流行和有趣的编程语言已经得到了解决: python、java 和 c。我们展示了如何使用这些 "非本机" 语言编写的基元创建应用程序。我们考虑了同时使用这四种编程语言 (多语言 cnp) 中的原语的示例。我们还讨论了不编程的 cnp 编程 (无语言 cnp)。少

2018 年 8 月 28 日提交;最初宣布 2018 年 9 月。

评论:12 页

日记本参考:《科学和工程调查国际图》, 第 7 卷, 第 78 期, 2018 年 7 月, 第 79-90 年

32. 第 xiv:1800.00974[pdf] cs. cy

以透明的方式进行能源建模工作, 处理发电厂数据库中的数据质量问题

作者:fabian Gotzens, heidi heheichs, jonas hörsch, fabian hofmann

摘要: 在能源建模中, 开放数据和开放源代码有助于提高模型练习的可追溯性和重现性, 这有助于促进有争议的辩论和改进政策建议。虽然近年来开放发电厂数据库的可用性有所增加, 但它们之间往往有很大差异, 其数据质量尚未与专有来源进行系统的比较。在这里, 我们介绍了基于巨蟒的 "动力匹配" (ppm), 这是一个用于清理、标准化和组合多个电厂数据库的开源工具集。我们只在开放数据库中应用一次, 一次应用于额外的专有数据库, 以便通过分析常规的能力、国家、燃料类型、地理坐标和调试年份来讨论和阐述数据质量问题。电厂。我们发现, 纯粹基于开放数据的派生数据集还没有与将专有数据库添加到匹配中的数据集相当, 尽管容量的统计值与这两个数据集在很大程度上匹配。当需要调试年来对最终数据集中进行建模时, 专有数据库将极大地帮助提高派生数据集的质量。少

2018 年 8 月 29 日提交;最初宣布 2018 年 9 月。

33. 第 1808. 810[[pdf](#),其他] Cs。 铭

利用图形数据结构和并行挖掘提高区块链性能

作者:[贾干](#),[陈尚哲](#),[黄新新](#)

摘要: 区块链技术迎来了又一个突破年, 区块链的挑战还有待解决。分析了基于区块链的比特币和比特币-ng 系统的特点, 提出了一种改进的区块链系统实现方法, 用图形数据结构取代了原有链的结构。它被命名为 "图形链"。每个块表示一个交易记录, 并包含交易者的余额状态。此外, 众所周知, 比特币系统中的所有交易都将由一个矿工进行, 这将导致大量的浪费, 因此, 提高资源利用率的另一种方法是改变最初竞争矿工的方式, 让他们成为选举和平行的人挖掘。研究人员利用图结构和通过巨蟒进行并行挖掘来模拟区块链, 并提出了概念上的新图形模型, 可以提高容量和性能。少

2018 年 10 月 30 日提交;v1 于 2018 年 8 月 31 日提交;最初宣布 2018 年 8 月。

评论:6 页, 2 个数字, [ieee hoticn 2018](#)

34. 第 1808. 10703[[pdf](#)] 反渗透委员会

python 机器人: 机器人算法的 python 代码集合

作者:[atsushi sakai](#), [daniel ingram](#), [joseph dinius](#), [karan chawla](#),[安东南·拉芬](#),[亚历克西斯·帕克斯](#)

文摘: 本文介绍了一个开源软件 (oss) 项目: pythonrobotics。这是在 python 编程语言中实现的机器人算法的集合。该项目的重点是自主导航, 目标是让机器人初学者了解每个算法背后的基本思想。在本项目中, 选择了实用且在学术界和产业界广泛使用的算法。每个示例代码都是用 python3 编写的, 并且仅依赖于一些标准模块的可读性和易用性。它包括直观的动画, 以了解模拟的行为。少

2018 年 9 月 7 日提交;v1 于 2018 年 8 月 31 日提交;最初宣布 2018 年 8 月。

评论:8 页, 6 个数字, 这是代码存储库:

<https://github.com/AtsushiSakai/PythonRobotics>

35. 第 1808. 10692[[pdf](#),其他] Cs。 Lg

apes: 用于模拟强化学习环境的 python 工具箱

作者:[aqeel labash](#), [ardi tampuu](#), [tambet matiisen](#), [jaanaru](#),[劳尔·维森特](#)

抽象: 在神经网络的帮助下, 增强学习代理在过去几年中能够解决日益复杂的任务。代理交互的模拟环境是任何强化学习问题的重要组成部分。环境模拟代理世界的动态, 从而在状态观察和外部奖励方面为其行为提供反馈。为了便于设计和模拟此类环境, 本文介绍了猿, python 中的一个高度可自定义和开源的包, 用于创建 2d 网格世界环境,

以加强学习问题。猿为代理提供模拟任何视野的算法, 允许根据用户定义的规则创建

和定位项目和奖励, 并支持多个代理的交互。少

2018 年 8 月 31 日提交;最初宣布 2018 年 8 月。

评论:6 页, 8 个数字

36. 第 xiv:1808. 09094[[pdf](#)] Cs。 Hc

pydraw: 一种基于 tkinter 的 gui 绘图生成器及其设计思想

作者:[林金伟](#),[周爱敏](#)

摘要: gui 的出现是计算机科学和软件设计史上的一大进步。gui 使人机交互更加简单和有趣。**python** 作为近年来流行的编程语言, 在 gui 设计中还没有实现。tkinter 的优点是支持 **python** 的本机支持, 但支持 tkinter 的视觉 gui 生成器太少。本文提出了一种基于 tkinter 框架的 gui 生成器 pydraw。详细介绍了 pydraw 的设计原理及其背后的强大设计理念。根据 pydraw 的 gui 设计理念, 它可以轻松地任何具有画布功能的 gui 框架或具有屏幕显示控制的编程语言设计可视 gui 呈现生成器。本文致力于传达 pydraw 的免费 gui 设计概念。通过实验证明了 pydraw 的实用性和有效性。为了更好地传达 pydraw 的设计理念, 让更多的爱好者加入 pydraw 的更新和演变, 我们有 pydraw 的源代码。在文章的最后, 我们总结了我们的经验, 并表达了我们未来 gui 设计的愿景。我们相信, 未来的 gui 将在图形软件编程中发挥重要作用, 未来少代码甚至没有代码编程软件的设计方法必须成为焦点和热点, 免费, 如绘图 gui 将是值得追求的。少

2018 年 8 月 27 日提交;最初宣布 2018 年 8 月。

37. 第 [xiv.1808.09008](#)[pdf,其他] cse

这就像 **python**, 但: 支持编程语言知识的传输

作者:[nischal shrestha](#), [titus barik](#), [chris parnin](#)

摘要: 编程方面的专业知识传统上假定二进制的副专家分歧。学习资源通常针对首次学习编程的程序员或该语言的专家程序员。一个代表性不足但很重要的程序员群体是那些在一种编程语言中有经验的程序员, 但他们希望用另一种语言编写代码。对于这种情况, 我们假设一种有效的反馈形式是作为从第一语言中的概念到第二种语言的转换而呈现的。当前的编程环境不支持这种形式的反馈。在这项研究中, 我们应用学习迁移理论来教授一种语言, 程序员不太熟悉的语言--比如 **r**--在他们已经知道的编程语言方面--比如 **python**。我们使用一种名为 "转移导师" 的新工具来研究学习迁移, 该工具根据等效的 **python** 代码对 **r** 代码进行解释。我们的研究发现, 即使在没有提示的情况下, 参与者也会将学习转移作为一种认知策略。参与者发现, 在一些功能方面, "转移导师" 是有用的, 例如逐步了解和突出可能被忽略或误解的事实。但是, 参与者不愿意在没有代码执行的情况下接受事实, 或者有时难以阅读冗长或复杂的解释。这些结果为未来的设计和研究方向提供了指导, 这些设计和方向可以支持学习新编程语言时的学习转移。少

2018 年 8 月 27 日提交;最初宣布 2018 年 8 月。

评论:参加 [ieee 视觉语言与以人为本计算研讨会 \(vl/hcc\)](#)

38. 第 [1808.08357](#)[pdf] Cs. 红外

tux 博士: ubuntu 用户的问答系统

作者:[bijil abraham phillip](#), [manas jog](#), [aprv milind upasani](#)

摘要: 各种论坛和问答 (问答) 网站可在线使用, 使 ubuntu 用户能够找到与他们的查询类似的结果。但是, 搜索结果通常非常耗时, 因为它要求用户从大量问题中查找与他/她的查询相关的特定问题实例。在本文中, 我们为 ubuntu 用户提供了一个名为 tux 博士的自动问答系统, 该系统旨在通过从在线数据库中选择最相似的问题来回答用户的问题。该原型在 **python** 中实现, 并将 **nltk** 和 **corenlp** 工具用于自然语言处理。原型的数据来自 askubuntu 网站, 其中包含大约 150k 的问题。从对原型的人工评价中获得的结果很有希望, 同时也提供了一些有趣的改进机会。少

2018 年 8 月 25 日提交;最初宣布 2018 年 8 月。

39. 第 188.07827[[pdf](#),其他] cs.PL

一种基于自动的字符串操作语言的抽象语义

作者:[vincenzo arceri](#), [isabella mastroeni](#)

摘要: 近年来, 动态语言, 如 javascript 或 **python**, 在广泛的领域和应用中面临着重要的使用增加。它们的棘手和误解行为对这些编程语言的静态分析构成了严峻的挑战。任何动态语言程序的一个关键方面都是字符串的多重用法, 因为它们可以隐式转换为另一个类型值, 通过字符串到代码的基元转换或用于访问对象属性。遗憾的是, 动态语言的字符串分析仍然缺乏准确性, 并且没有考虑到一些重要的字符串功能。此外, 字符串混淆在动态语言恶意代码的上下文中非常流行, 例如, 隐藏字符串中的代码信息, 然后将字符串动态转换为可执行代码。在这种情况下, 更精确的字符串分析成为必要。本文提出了一种新的字符串分析语义, 为处理动态语言字符串功能提供了第一步。少
2018 年 8 月 17 日提交;最初宣布 2018 年 8 月。

40. 建议: 1808.07432[[pdf](#),其他] Cs. 铭

多伊 [10.114/3229565.3229567](#)

智能家居物联网的友好型图书馆

作者:[trisha datta](#), [noah apthorpe](#), [nick feamster](#)

摘要: 在过去几年中, 互联网连接设备的数量和种类大幅增长, 给安全和隐私带来了新的挑战。研究表明, 网络对手可以使用来自消费者物联网设备的流量速率元数据来推断敏感的用户活动。将流量调整为适合独立于用户活动的发行版可以保护隐私, 但由于开发人员的工作量和开销带宽成本, 这种方法几乎没有被采用。在这里, 我们为物联网开发人员提供了一个 **python** 库, 以便轻松地将隐私保护流量整形集成到他们的产品中。该库将标准网络功能替换为通过有效负载填充、碎片和随机覆盖流量的组合自动混淆设备流量模式的版本。我们的库成功地保护了用户隐私, 并要求大约 4 kbn 的架空带宽为物联网设备具有较低的发送率或较高的延迟容忍度。考虑到美国家庭的正常互联网速度, 这种开销是合理的, 是对现有解决方案带宽要求的一种改进。少

2018 年 8 月 22 日提交;最初宣布 2018 年 8 月。

评论:6 页, 6 个数字

日记本参考:2018 年物联网安全与隐私研讨会论文集, 43-48 页, 2018 年 8 月

41. 第 1808. 06226[[pdf](#),其他] Cs. CI

句子: 一种简单而语言独立的神经文本处理用子词和解毒器

作者:[taku kudo](#), [john richardson](#)

摘要: 本文介绍了一种独立于语言的子词标记器和解毒器, 用于基于神经机器翻译的文本处理。它为子字单元提供开源 **c++** 和 **python** 实现。虽然现有的子词分割工具假定输入是预先标记为单词序列, 但 **sentencesi** 分为直接从原始句子训练子词模型, 这使我们能够建立一个纯粹的端到端和语言无关的系统。我们对英日机器翻译进行了 **nmt** 的验证实验, 发现有可能实现与原始句子中的直接子词训练相当的准确性。我们还比较了子词训练和分割与各种配置的性能。

<https://github.com/google/sentencepiece> 的 **apache 2** 许可证下可以使用句子。少

2018 年 8 月 19 日提交;最初宣布 2018 年 8 月。

评论:在 **emnlp2018** 被接受为演示论文

42. 第 **xiv:1808. 04919**[[pdf](#), **ps**,其他] cse

giststgithb 上 python 代码段的可执行性的评估

作者:eric horton, chris parnin

摘要: 软件开发人员在线创建和共享代码, 以演示编程语言概念和编程任务。代码段可以是解释和演示编程概念的有用方法, 但可能并不总是可以直接执行的。代码段可能包含分析错误, 或者在环境包含未满足的依赖项时无法执行。本文对通过 github gist 系统共享的 **python** 代码段的可执行状态以及熟悉软件配置的开发人员正确配置和运行它们的能力进行了实证分析。我们发现, 75.6 的专家需要非平凡的配置来克服缺少的依赖项、配置文件、对特定操作系统的依赖或其他一些环境配置。我们的研究还表明, 当解决配置错误不到一半的时间时, 开发人员对资源名称的自然假设是正确的。我们还提出了 gistable, 一个数据库和可扩展的框架, 建立在 github 的 gist 系统上, 它提供了可执行的代码片段, 以便能够在软件工程中进行可重复的研究。Gistable 包含 10, 259 个代码段, 大约 5, 000 个代码段, 其中包含一个 dockerfile 文件, 用于配置和执行它们, 而不会出现导入错误。Gistable 于 <https://github.com/gistable/gistable> 公开提供。少

2018 年 8 月 14 日提交;最初宣布 2018 年 8 月。

43. 第: 1808. 4849[pdf] Cs. 直流

用于医疗保健和精密医学研究的可扩展数据科学平台

作者:jacob mcpadden, thomas js durant, dustin r bunch, andreas coppi, nathan price, kris rodgerson, charles j torre jr, william byron , h patrick young, allen l xiao, harlan m krumholz, wade l schulz

摘要: 目的: (1) 演示在大型学术医疗系统中构建基于开源技术的数据科学平台的实施情况, (2) 描述在该平台上构建的两个计算医疗应用。材料与方法: 部署了基于多种开源技术的数据科学平台, 以支持实时大数据工作负载。apache storm 和 nifi 的数据采集工作流程是在 java 和 **python** 中开发的, 用于捕获患者监控和实验室数据, 以便进行下游分析。结果: 使用新兴的数据管理方法以及 hadoop 等开源技术可用于创建集成的数据湖, 以存储大型实时数据集。该基础架构还提供了一个强大的分析平台, 可在近实时的情况下分析医疗保健和生物医学研究数据, 以满足精密医学和计算医疗保健用例的需要。讨论: 综合数据科学平台的实施和使用使各组织有机会将包括电子健康记录数据在内的传统数据集与新出现的大数据源 (如连续性病人) 结合起来监测和实时实验室结果。这些平台可以为信息提供经济高效且可扩展的分析, 而这些信息将是提供精密医学举措的关键。结论: 能够利用数据科学平台中发现的技术进步的组织将有机会为计算医疗和精密医学研究提供全面的医疗数据访问。少

2018 年 8 月 14 日提交;最初宣布 2018 年 8 月。

评论:8 页, 4 个数字, 1 个表

44. 第 xiv:1808. 04790[pdf,其他] cs et

基于 dna 的化学编译器

作者:沙林·沙阿,马尼什·古普塔

摘要: 1997 年, 马尔切洛正式证明, 化学动力学可以成为一台通用计算机, 即可以复制任何数字电路。最近, soloveichik 等人表明, 化学动力学可以进行快速可靠的图灵宇宙计算。为了模拟化学反应的行为, sean 等人开发了一个名为 cain 的软件, 它以 xml 格式表示化学反应。在这项工作中, 我们试图创建跨编译器, 它可以将巨蟒代码作为输入, 并将 cain 支持的化学反应文件作为输出。这可以与从高级编程语言生成程序集代

码相比较。此外, soloveichik 等人还展示了 dna 作为通用入门实施 cm 的和安德鲁斯菲利普斯开发的可视化 dsd 编程语言, 以模拟所有可能的 dsd 反应。crn2dsd 是 manish gupta 团队开发的软件, 它已经可以将 cain 文件转换为 microsoft 的 visual dsd 代码, 即程序集级别转换为计算机级别。因此, 我们尝试将高级代码转换为程序集代码, 这使我们离完成创建化学编译器的梦想又近了一步。少

2018 年 8 月 14 日提交;最初宣布 2018 年 8 月。

评论:这项工作是作为沙林 2015 年学士学位论文的一部分完成的

45. 第 xiv:1808.03753[[pdf](#),其他] Cs. Lg

marvin: 一种用于自动机器学习原始注释和执行的开放式机器学习语料库和环境

作者:[chris a. mattmann](#), [sujen shah](#), [brian wilson](#)

摘要: 在本演示本文中, 我们介绍了用于自动机器学习 (ml) 的 darpa d3m 程序和 jpl 的 marvin 工具, 它提供了一个用于 ml 管道的机器学习原语的定位、注释和执行环境。marvin 是一个以 python 编写的基于 web 的应用程序和关联的后端接口, 它允许从科学-学习、keras、dl4j 和其他广泛使用的库的数百个基元组成 ml 管道。marvin 允许在 darpa 内的 kubernetes 集群上创建 docker 容器, 以便为自动机器学习提供执行环境。marvin 目前包含 400 多个数据集, 并在各种 ml 领域面临挑战, 包括常规分类和回归到高级视频图像分类和遥感。少

2018 年 8 月 11 日提交;最初宣布 2018 年 8 月。

46. 建议: 1808.03292[[pdf](#),其他] Cs. 马

nl4py:用于并行网络徽标工作区的 python 中基于代理的建模

作者:[chathika gunaratne](#), [ivan garibay](#)

摘要: nl4py 是 python 的 netlogo 控制器软件, 用于快速、并行地执行 netlogo 模型。nl4py 通过 python 提供无头 (无图形用户界面) 和 gui netlogo 工作区控制。在 python 包索引上越来越多的开源计算和机器学习库的推动下, 对通过 python 快速、并行地执行基于代理的模型的需求越来越大。netlogo 是大多数基于代理的建模驱动研究项目的首选语言, 它需要集成到 python 中, 以便研究人员希望使用这些库对基于代理的模型输出进行统计分析。不幸的是, 直到最近推出的 pynetlogo, 现在 nl4py, 这样的控制器是不可用的。本文详细介绍了 nl4py 的使用情况, 并解释了其客户端-服务器软件体系结构, 强调了 pynetlogo 的体系结构差异。然后通过 nl4py 对狼羊捕食模型进行全球灵敏度分析和参数校准的逐步演示。最后, NL4Py's 的性能以 pynetlogo 及其与 IPyParallel 的组合为基准, 并证明在这两种配置上都能显著节省执行时间。少

2018 年 8 月 20 日提交;v1 于 2018 年 8 月 9 日提交;最初宣布 2018 年 8 月。

47. 第 xiv:1808.02860[[pdf](#),其他] Cs. Gr

多分辨率数据的电影可视化: 胡迪尼自适应网格细化的伊提尼

作者:[kalina borkiewicz](#), [j. p. naiman](#), [ha 定位 m lai](#)

摘要: 我们已经进入了以日益复杂的数据结构为代表的大型多维数据集的时代。目前用于科学可视化的工具没有进行优化, 无法高效、直观地创建电影制作质量、时间变化的数字数据表示, 以便通过电影、媒体或新闻传播进行广泛的影响科学传播。要在电影环境中显示此类数据, 最好开发将这些复杂的数据结构集成到行业标准视觉效果软件包中的方法, 这些软件包提供了无数其他信息在传统的科学可视化软件。本文介绍了将嵌套多分辨率数据集导入和可视化到市售视觉效果软件中的一般方法。我们进一步提供

了一个将自适应网格细化数据导入到软件 houdini 中的具体示例。本文以我们以前的工作为基础,介绍了一种利用胡迪尼可视化统一笛卡尔数据集的方法。我们总结了 www.ytini.com 网站上提供的教程,其中包括示例数据下载、python 代码和各种其他资源,以简化导入和呈现多分辨率数据的过程。少

2018 年 11 月 1 日提交;v1 于 2018 年 8 月 8 日提交;最初宣布 2018 年 8 月。

评论:24 页, 14 位数字

48. 第 [xiv:1808.02731](#)[pdf,其他] cs. ms

热解谱延迟校正

作者:[罗伯特·斯派克](#)

文摘: 本文提出了用光谱递延校正方法 (sdc) 及其时间并行变种 pfastst (空间和时间的并行全逼近方案) 求解搭配问题的 python 框架 pysdc。pysdc 具有许多 sdc 和 pfastst 的实现,从简单的隐式时间步进到高阶隐式或多隐式分裂和多级谱延迟校正。它有许多不同的、预先实现的例子,有七个教程来帮助新用户的第一步。时间并行性是以模拟的方式实现的,既用于调试和原型设计,也可以使用 mpi 进行基准测试。代码经过全面记录,并使用持续集成进行测试,包括以前发布的大多数结果。在这里,我们通过从两个不同的角度来描述代码的结构:用户的和开发者的视角。虽然第一个描述了前端、示例和教程,但第二个用于描述基础实现和数据结构。我们展示了三个不同的例子,以突出实现的各个方面,能力和使用的 pysdc。此外,还描述了与 fenics 框架和 petsc 的耦合,后者包括与 mpi 的空间并行性。少

2018 年 8 月 8 日提交;最初宣布 2018 年 8 月。

49. 第: [1808.1995](#)[pdf,其他] Cs. Dm

devito: 一种用于有限差异和地球物理探索的嵌入式域特定语言

作者:[mathias louboutin](#), [michael lange](#), [fabio luporini](#), [navjot kukreja](#), [phillipa .witte](#) , [felix j.herrmann](#), [paulius velesko](#) , [gerard j. gorman](#)

摘要: 我们介绍了设在特定于域的一种新语言--实现高性能有限差分偏微分方程求解器。激励的应用是探索地震学,使用全波形反转和反向时间偏移等方法来反转 tb 的地震数据,以创建地球地下的图像。即使使用现代超级计算机,也可能需要数周的时间来处理一次地震调查,并创建一个有用的地下图像。计算成本主要是由波动方程及其相应的伴随点的数值解决定。因此,在为不同的计算机体系结构中,积极优化这些波方程传播器的性能投入了大量的精力。此外,随着物理学日益逼真的表示方法的发展,正在求解的偏微分方程的实际集合及其数值离散化也在不断创新,进一步提高了实际的成本。解决。通过在 python 中嵌入特定于域的语言并大量使用 sympy 这一符号数学库,我们可以使用与数学非常相似的语法快速开发有限差分模拟器。德维托编译器读取此代码,并应用广泛的分析来生成高度优化和并行的代码。这种方法可以将经过验证和优化的求解器的开发时间从几个月缩短到几天。少

2018 年 8 月 6 日提交;最初宣布 2018 年 8 月。

评论:提交给地球科学模型开发 (gmd)

50. 第 [xiv:1808.01743](#)[pdf,其他] Cs. Lg

尼玛卡: 非负矩阵分解的 python 库

作者:[marinka zitnik](#), [braz zupan](#)

摘要: nimfa 是一个开源的 **python** 库, 它为非负矩阵分解算法提供了一个统一的接口。它包括最先进的分解方法、初始化方法和质量评分的实现。它支持密集和稀疏矩阵表示。nimfa 基于组件的实现和分层设计应有助于用户采用已经实现的技术, 或为矩阵分解任务设计和编码新的策略。少

2018 年 8 月 6 日提交;最初宣布 2018 年 8 月。

日记本参考:机器学习研究杂志 13 (2012) 849-853

51. 第 1808. 01100[[pdf](#), [ps](#),其他] [cs. cy](#)

代码 shrew: 通过绘图和动画进行编程教学的软件平台

作者:[l 宽 wik trammer](#), [jamie nunez](#)

摘要: 在本文中, 我们提出了一个新的软件平台, 伴随着一个互动的编程课程代码。其目的是通过使用户能够创建自己的绘图和动画来教授计算机编程的基础知识。编程语言有一个基于 **python** 的简单语法, 它增加了一些功能, 可以使用面向对象的代码轻松绘制和进行动画处理。编辑器可无缝、即时地做出反应, 为实验和测试想法提供一个引人入胜的交互式环境。编程课程包括涵盖基本编程原则的课程, 以及在用户在整个课程过程中测试其技能所面临的挑战。课程和挑战都利用了编辑器的即时反馈, 使人们能够专注于边干边学。我们描述了软件和内容, 它们背后的动机, 以及它们与建构主义的联系。少

2018 年 8 月 8 日提交;v1 于 2018 年 8 月 3 日提交;最初宣布 2018 年 8 月。

评论:7 页

52. 第 [xiv:1808.00823](#)[[pdf](#),其他] [cse](#)

多伊 [10.115/3237009.23237017](#)

调试动态语言的本机扩展

作者:[jacob kreindl](#), [manuel rigger](#), [hanspeter mössenböck](#)

摘要: 许多动态编程语言 (如 **ruby** 和 **python**) 使开发人员能够使用所谓的本机扩展, 这些代码是在通常静态编译的语言 (如 **c** 和 **c++**) 中实现的。但是, 这些动态语言的调试器通常不支持也调试这些本机扩展。graalvm 可以执行各种动态编程语言实现的程序, 并通过使用 **llvm-ir** 解释器 **sulong**, 也可以执行它们的本机扩展。通过将 **llvm-ir** 级别的运行时调试信息与原始程序代码相关联, 我们在 **graalvm** 调试框架的基础上为苏龙添加了对源级别调试的支持。因此, 开发人员现在可以使用 **graalvm** 调试以多种基于 **lvm** 的编程语言编写的源代码, 以及在公共调试器前端调用的各种动态语言实现的程序。少

2018 年 8 月 2 日提交;最初宣布 2018 年 8 月。

评论:在第十五届国际管理语言与运行时大会 (曼朗 18) 上接受 7 页, 7 位数

53. 第 [xiv:1808.00773](#)[[pdf](#),其他] [Cs. Sd](#)

dcase 2018 挑战苏瑞交叉任务卷积神经网络基线

作者:[孔秋强](#), [turab iqbal](#), [yong xu](#), [wwwu wang](#), [mark d. plumbley](#)

摘要: 声场景和事件的检测和分类 (dcase) 由五个音频分类和声音事件检测任务组成: 1) 声学场景分类, 2) 通用音频标记的免费, 3) 鸟音频检测, 4) 弱标记的半监督声音事件检测和 5) 多声音频分类。在本文中, 我们建立了一个基于卷积神经网络 (cnn) 的所有五个任务的交叉任务基线系统: 一个 "cnn 基线" 系统。我们实施了 4 层和 8 层的 cnn, 来自于亚历克网和 vgg 从计算机视觉。我们研究了使用相同的神经网络配置的不同任

务的性能变化。实验表明, 较深的美国有线电视新闻网与 8 层执行更好的比美国有线电视新闻网与 4 层的所有任务, 除了任务 1。使用 8 层 cnn, 我们在任务 1 上实现了 0.680 的精度, 在任务 2 上实现了 0.89 95 的平均精度 (map), 在任务 2 上的精度为 0.680, 在任务 3 上的曲线下的面积为 0.680, 声音事件检测 f1 分数为 20.8%4, 任务 5 的 f1 分数为 87.75。我们发布了麻省理工学院许可下基线系统的 **python** 源代码, 以便进一步研究。少

2018 年 9 月 29 日提交;v1 于 2018 年 8 月 2 日提交;最初宣布 2018 年 8 月。

评论:接受 dcase 2018 讲习班。可用的源代码

54. 第 [xiv:1808.00532](#)[pdf,其他] cs. ms

guetnet: 张量网络的图形用户界面

作者:[lisa Sahlmann](#), [christian b. mendl](#)

摘要: 我们引入了一个图形用户界面, 用于构造任意张量网络并指定常见的操作, 如收缩或拆分, 表示 guetnet。张量表示为带有连接腿的节点, 对应于张量的有序尺寸。guetnet 可视化当前网络, 并立即为迄今为止的用户操作序列生成 **python**/numpy 源代码。计划今后支持其他编程语言。我们讨论了 guitenet 使用的张量网络的基本操作, 以及高级优化策略。该软件直接在 web 浏览器中运行, 可在线 <http://guitenet.org>。少

2018 年 7 月 30 日提交;最初宣布 2018 年 8 月。

评论:7 页, 4 个数字

55. 第 [xiv:1808.00106](#)[pdf,其他] cse

sourcerer 的学徒和代码段迁移的研究

作者:[stephen romansky](#), [cheng chen](#), [baljeet malhotra](#), [abram hindle](#)

摘要: 在全球网络上, 不仅网页已连接, 源代码也是如此。软件开发对每个人来说都变得越来越容易获得, 软件的许可仍然很复杂。我们需要知道软件许可证在整个重用和进化过程中是否得到了适当的维护。这推动了 sourcerer 学徒的开发, 这是一项帮助跟踪克隆重新注册的网络服务, 因为软件通常使用软件许可证来描述如何使用和调整他们的软件。但大多数开发人员不具备解决许可证冲突的法律专业知识。在本文中, 我们将学徒放在实证研究上, 证明 stackoverflow 代码和 **python** 模块以及 **python** 文档之间有很多共享, 这违反了原始 **python** 的许可模块和文档: 通过 stackoverflow 共享的软件片段通常被不适当地重新分配为 cc-by-sa 3.0, 而不保持适当的归属。我们展示了 stackoverflow 上的许多代码段被 stackoverflow 用户不适当地重新设置, 从而危及使用 stackoverflow 代码段的公司和开发人员构建的软件的状态。少

2018 年 7 月 31 日提交;最初宣布 2018 年 8 月。

评论:10 页, 2 个数字

56. 第 [xiv:1807.10581](#)[pdf,其他] Cs. 简历

多尺度逐步整合 cnn 在肺结节检测中的假阳性减少

作者:[bum-chae kim](#), [jun-sik choi](#), [heung-il suk](#)

摘要: 肺癌是一种全球性的危险疾病, 其早期发现对于降低死亡风险至关重要。在这方面, 它一直非常感兴趣的开发一个计算机辅助系统的肺结节检测尽早在胸部 ct 扫描。一般而言, 结核检测系统涉及两个步骤: (一) 候选结核检测具有高度敏感性, 可捕获许多误报; (二) 候选项的假阳性减少。然而, 由于结核形态特征的变化很大, 而且有可能

将其误认为邻近器官, 候选结核的检测仍然是一个挑战。在本研究中, 我们提出了一种新的多尺度渐进式融合卷积神经网络 (mgi-cnn), 设计有三个主要策略: (1) 使用不同层次的上下文信息的多尺度输入, (2) 使用抽象信息固有于不同的输入尺度与逐步集成, (3) 学习多流功能集成的端到端的方式。为了验证所建议网络的有效性, 我们通过将所提出的方法的性能与文献中最先进的方法进行比较, 对 luna16 挑战数据集进行了详尽的实验。在 luna16 数据集的两个候选子集上, 即 v1 和 v2, 我们的方法平均 cpm 为 0.908 (v1) 和 0.908 (v2), 在很大程度上超过了可比方法。我们的 mgi-cnn 是使用滕索流在 **python** 中实现的, 源代码可从 "https://github.com/ku-milab/MGICNN" 中获得。少

2018 年 7 月 24 日提交;最初宣布 2018 年 7 月。

评论:11 页, 6 个数字, 5 个表

57. 第 1807. 10129[[pdf](#),[其他](#)] **cs. ms**

计算机视觉和机器学习中几个问题的选定算法微分工具的基准

作者:[filipšrajter](#), [zuzana kkkelova](#), [andrew fitzgibbon](#)

摘要: 算法微分 (ad) 允许精确计算导数只给目标函数的一个实现。虽然有许多 ad 工具可用, 但正确和有效地实现 ad 方法并不简单。现有的工具往往太不一样, 无法使用通用的测试套件。本文比较了 15 种计算导数的方法, 包括 11 种自动微分工具, 实现各种方法, 并以各种语言 (c++、f#、matlab、julia 和 **python**) 编写, 这两种符号化差异化工具, 有限的差异, 和手工派生的计算。我们从计算机视觉和机器学习三个客观功能来研究。这些目标在很大程度上很简单, 也就是说不涉及迭代循环, 并且条件语句封装在函数中, 如 {tt abs} 或 {结节 logsumexp}。然而, 有效地处理这种 "简单" 的目标函数对于算法微分的成功是很重要的, 因为计算机视觉和机器学习中的许多问题都是这种形式的。当然, 我们的结果取决于程序员的技能, 以及对工具的熟悉程度。然而, 我们认为, 本文提出了一个重要的数据点: 一个熟练的程序员花了大约一个星期的时间来每个工具产生的时间, 我们提出的。我们已将我们的实现作为开源提供, 以允许社区复制和更新这些基准。少

2018 年 7 月 26 日提交;最初宣布 2018 年 7 月。

评论:本篇文章的早期版本出现在 ad2016-- 第七届算法差异化国际会议上, 以及在优化方法和软件中, 泰勒和弗朗西斯, 2018 年 2 月 (在线)

58. 第 xiv:1807. 09688[[pdf](#),[其他](#)] **Cs. Ce**

湍流: 流体流动模拟后处理的 **python** 包

作者:[timofey mukha](#)

文摘: 提出了一种用于计算流体力学模拟平面二维数据后处理的 **python** 包。该软件包称为 "湍流", 它为大型模拟活动的脚本化、可重现性分析提供了手段, 并包括用于数据提取和可视化的例程。对于前者, 使用可视化工具包 (vtk), 允许对在非结构化网格上执行的模拟进行后处理。对于可视化, 提供了几个基于 matplotlib 的函数, 用于创建高度可定制的、公共质量的绘图。为了演示湍流的功能, 它在这里应用于后处理, 模拟向后一步上的流。还讨论了包的实现和体系结构, 以及它的重用潜力。少

2018 年 7 月 25 日提交;最初宣布 2018 年 7 月。

59. 第 xiv:1807. 09632[[pdf](#),[其他](#)] **Cs. 直流**

多伊 [10.114/32194.3229289](#)

帕帕斯: 用于并行参数研究的便携式、轻量级和通用框架

作者: [爱德华多·庞塞](#), [布列塔尼·斯蒂芬森](#), [苏珊娜·伦哈特](#), [朱迪·戴](#), [格雷戈里·彼得森](#)

摘要: 目前的科学研究环境广泛地建立在建模和仿真的基础上, 通常具有复杂的仿真执行流和参数化特性。执行流不一定简单, 因为它们可能需要多个处理任务和迭代。此外, 参数和性能研究是用于描述模拟的常用方法, 通常需要遍历较大的参数空间。高性能计算机提供实用的资源, 而牺牲了处理作业的设置、提交和管理的用户的费用。本工作介绍了 papa 的设计, 这是一个可移植、轻量级和通用的工作流框架, 用于进行并行参数和性能研究。工作流是使用基于关键字值对语法的参数文件定义的, 从而从用户身上消除了创建复杂脚本以管理工作流的开销。参数集由环境变量、文件、部分文件内容和命令行参数的任意组合组成。papa 是在 **python 3** 中开发的, 支持使用 ssh、批处理系统和 c++ mpi 进行分布式并行化。papa 框架将作为用户进程运行, 并可用于单节点和多租户计算系统。分别将 netlogo 的行为空间工具和使用 openmp 的矩阵乘法进行了实例仿真, 作为参数和性能研究。结果表明, papa 框架为定义和管理参数研究提供了一种简单的方法, 同时提高了资源利用率。少

2018 年 7 月 25 日提交;最初宣布 2018 年 7 月。

评论: 8 页, 6 位数字, pearc ' 18: 高级研究计算的实践和经验, 2018 年 7 月 22-26 日, 匹兹堡, 宾夕法尼亚州, 美国

60. 第 1807. 09224[[pdf](#),其他] Cs。哦

FluidDyn: 用于流体动力学研究和教学的 python 开源框架

作者: [pierre augier](#), [ashwin vishnu mohan](#), [cyrille bonamy](#)

摘要: FluidDyn 是一个在流体动力学社区中培养开放科学和开源的项目。它被认为是一个研究项目, 将开源动态、方法和工具引导到做科学上。我们提出了一套 **python** 包, 形成了一个框架, 以研究流体动力学与不同的方法, 特别是实验室实验 (包流体实验室), 模拟 (包流体, 流体和流体泡沫) 和数据处理 (包)。在本文中, 我们概述了项目的专业包, 然后重点介绍了称为 fluiddyn 的基本包, 其中包含专门包中使用的通用代码。在 mohan 等人 (2018a. b) 的两份配套论文中更详细地介绍了流式流体和流体。通过 FluidDyn 项目, 我们证明了专门的科学代码可以使用开源社区的方法和良好做法来编写。在 bitbb (<https://bitbucket.org/fluiddyn/>) 中提供了汞存储库。所有代码都使用 spinx 和 read 文档进行记录, 并通过在 bit 存储桶、管道和 travis 上连续集成进行测试。为了提高重用潜力, 利用 **python** 的简单面向对象编程模型, 代码尽可能模块化。所有代码也都是高效编写的, 使用 c++、cython 和 pythran 来提高关键函数的性能。少

2018 年 7 月 3 日提交;最初宣布 2018 年 7 月。

61. 第: 1807. 07693[[pdf](#)] Cs。Ce

通过模型抽象实现高效仿真的挑战

作者: [hessam s. Sarjoughian](#), [william a. boyd](#), [miguel f. acevedo](#)

摘要: 耦合的自然系统通常在多个抽象级别上建模。这些模型的结构尺度和行为复杂性都是可以提出和回答的各种问题的决定因素。随着模型规模和复杂性的增加, 仿真效率必须提高, 以解决模型分辨率和仿真时间之间的权衡。从这一有利的角度出发, 我们将以植物景观模型为例, 展示一些问题和解决方案, 在这个模型中, 属于不同物种的个别植物被代表为经历了跨越数百种生长和衰退周期的集体。年。集体工厂实体被分配给静态二维网格的单元格。这个粗粒度模型, 在同态建模思想的指导下, 是从一个将植

物表示为单个对象的细粒模型中推导出来的。这些模型是使用 **python** 和 **grass** 工具开发的。设计了一组实验来揭示此类系统建模和仿真中的一些障碍。少

2018 年 7 月 19 日提交;最初宣布 2018 年 7 月。

评论:美国亚利桑那州坦佩亚利桑那州立大学计算、信息学和决策系统工程学院内部报告

62. 特别报告: 1807. 07643[[pdf](#)] [cs. ms](#)

科学 **python** 中的物理类型正确性

作者:[marcus foster](#), [sean Tregagle](#)

摘要: 在信息系统中, 单位和尺寸的表示几乎没有编纂成文, 而且往往被忽视。例如, 科学计算中使用的主要语言 (**fortran**、**c** 和 **python**) 没有维度或单位的类型, 因此物理数量在程序中由实际类型的变量表示, 从而有可能尺寸误差。针对这种危险, 许多作者提出了单位检查和转换的语言方案。但是, 由于许多物理数量具有相同的单位, 因此代码块可能与单位兼容, 但在物理上仍然毫无意义。我们演示了三个 **python** 单元库的局限性, 并提出了检验数量类型的理由和方法。少

2018 年 8 月 3 日提交;v1 于 2018 年 7 月 17 日提交;最初宣布 2018 年 7 月。

评论:6 页. v2 扩展仅摘要. v2 增加了参考到代码的描述为最近单位库 `unyt` `ary:1806.02417`; 固定拼写错误

63. 第 1807. 05317[[pdf](#)] [Cs. Lg](#)

leflow: 实现张力深部神经网络的灵活 **fpga** 高级合成

作者:[daniel h. noronha](#), [bahar salehpour](#), [steven j. e. wilton](#)

摘要: 最近的研究表明, 现场可编程门阵列 (**fpga**) 在加速机器学习应用中发挥着重要作用。机器学习应用程序的初始规范通常使用高级 **python** 导向框架 (如 **tensorflow**) 完成, 然后手动转换为 **c** 或 **rtl**, 以便使用供应商工具进行合成。这个手动翻译步骤非常耗时, 需要专业知识来限制 **fpga** 在这一重要领域的适用性。本文提出了一种开源工具流, 将用 **tensorflow** 编写的数值计算模型映射到可合成硬件。与其他工具 (通常受到少量不灵活模板的限制) 不同, 我们的流程使用 **google xla** 编译器, 该编译器直接从 **tensorflow** 规范中发出 **llvm** 代码。然后, 此 **llvm** 代码可与高级合成工具一起使用, 以自动生成硬件。我们展示了我们的流程允许用户使用很少的 **python** 代码行生成深度神经网络。少

2018 年 7 月 13 日提交;最初宣布 2018 年 7 月。

评论:将在软件程序员的 **fpga** 上发布 (`fsp 2018`)

64. 第 1807. 05252 —[[pdf](#),其他] [cs. ms](#)

沙丘 **python** 模块

作者:[andreas dedner](#), [martin nolte](#)

文摘: 在本文中, 我们提出了新的 **dune-python** 模块, 它为沙丘核心提供 **python** 绑定, 它是一个用于求解偏微分方程的 **c++** 环境。这个新模块的目的是首先提供基于实时编译导出静态多态接口实现的一般基础结构, 其次是为沙丘核心模块的中心接口提供绑定。在第一个版本中, 我们将重点放在网格接口上。我们的目标是在将对象传递到 **python** 时引入一个薄层, 当对象被传递回 **c++** 算法时, 可以将其删除。因此, 不会损失任何效率, 也不会产生什么额外的代码维护成本。为了使沙丘用户过渡到 **python** 环境变得简单, **python** 类提供了一个与 **c++** 对应方非常相似的接口。此外, 许多接口的

矢量化版本允许在 **python** 端实现更高效的代码。本文详细介绍了导出这些接口的基础结构以及对沙丘网格的绑定, 这些基础结构适用于有经验的沙丘用户和其他对灵活的 **python** 环境感兴趣的用户, 用于实现基于网格的**网格**求解偏微分方程的方案。少

2018 年 7 月 13 日提交;最初宣布 2018 年 7 月。

65. 第: 1807.04662[[pdf](#), [ps](#),其他] Cs. Lg

多流: 多输出流框架

作者:[jacob montiel](#), [jesse read](#), [albert bifet](#), [talel abdessalem](#)

摘要: 科学多流是 **python** 编程语言的多输出/多标签和流数据挖掘框架。它被认为是鼓励河流学习研究民主化的一个平台, 它为河流学习、流式生成器和评价者提供了多种最先进的方法。科学多流建立在流行的开源框架之上, 包括科学学习、moa 和 meka。开发遵循自由和开放源码软件的原则, 通过遵守 pep8 准则并使用持续集成和自动测试来执行质量。源代码可在 <https://github.com/scikit-multiflow/scikit-multiflow> 公开使用。少

2018 年 7 月 12 日提交;最初宣布 2018 年 7 月。

评论:5 页, 开源软件

66. 特别报告: 1807.0 04053[[pdf](#),其他] Cs. Cl

unicsease:ajol 个基于图形的解析工具包

作者:[daniel varab](#), [natalie schluter](#)

文摘: 本文介绍了作为开源巨蟒软件包发布的基于图形的解析框架和工具包 unispase 的设计和使用。unispase 作为一个框架, 简化了基于图形的依赖关系解析体系结构的研究原型、开发和评估。unispase 通过为所有依赖解析器组件启用高效、充分独立、易于阅读和易于扩展的实现来实现这一点。我们分发带有现成配置的工具包, 作为当前所有最先进的基于一阶图形的解析器的再实现, 包括更高效的编码器和解码器的 cython 实现, 以及所需的专门的损失功能。少

2018 年 7 月 11 日提交;最初宣布 2018 年 7 月。

67. 建议: 1807. 03662[[pdf](#)] Cs. 直流

trialchain: 一种基于区块链的平台, 用于在大型生物医学研究中验证数据完整性

作者:[郝 dai](#), [h patrick young](#), [thomas js durant](#), [guannan gong](#), [minming kang](#), [harlan m krumholz](#), [wade l schulz](#), [lixin jiang](#)

摘要: 用于生物医学研究和临床试验的数据治理是产生准确结果的重要要求。为了提高数据质量和分析的可见性, 我们开发了 trialch 之上, 这是一个基于区块链的平台, 可用于验证大型生物医学研究的数据完整性。我们使用多链平台实施了一个私有区块链, 并将其与部署在大型研究中心内的数据科学平台集成。使用 **python** 构建了一个管理 web 应用程序来管理该平台, 该平台是使用 docker 构建的微服务体系结构。在数据采集过程中, TrialChain 平台被集成到我们现有的数据科学平台中。使用 nifi, 在本地区块链基础结构中对数据进行哈希处理和记录。为了提供公共验证, 局部区块链状态定期与公共 ethereum 网络同步。使用组合的隐私/公共区块链平台可以对结果进行公开验证, 同时保持额外的安全性和更低的区块链交易成本。由于下游分析而产生的原始数据和修改可以在 trialchain 中记录, 并且可以使用对平台的 api 调用在需要时快速验证数据资产或结果。trialchain 平台为审计生物医学研究数据的获取和分析提供了数据治理解决方案。该平台提供了数据真实性的加密保证, 也可用于记录数据分析。少

2018 年 7 月 10 日提交;最初宣布 2018 年 7 月。

68. 第: 1807. 03043[[pdf](#),[其他](#)] Cs。简历

用于血糖预测的卷积递归神经网络

作者:[李克志](#)、[约翰·丹尼尔斯](#)、刘成元、[保罗·赫雷罗](#)、[潘特利斯·格奥尔基乌](#)

摘要: 控制血糖对糖尿病管理至关重要。目前的数字治疗方法的 1 型糖尿病 (t1dm), 如人工胰腺和胰岛素泵计算器利用机器学习技术预测皮下葡萄糖, 以改善控制。深度学习最近被应用于医疗和医学研究, 以在疾病诊断和患者状态预测等一系列任务中取得最先进的成果。在这项工作中, 我们提出了一个深度学习模型, 能够预测 30 分钟地平线上的葡萄糖水平, 并在模拟患者病例中具有领先的准确性 ($\text{rmse} = 9.38 \pm 0.71$ [mg/dl] 和 $\text{mard} = 5.50 \pm 0.62$) 和真正的病人病例 ($\text{rmse} = 21.13 \pm 1.23$ [mg/dl] 和 $\text{mard} = 10.08 \pm 0.83\%$)。此外, 该模型还在预测模拟患者数据集中的不良血糖事件时具有竞争力, 且时间滞后最小 ($\text{mcc}_{\text{HYper}} = 0.83 \pm 0.05$ 和 $\text{mcc}_{\text{HYp 不, 不, 不}} = 0.80 \pm 0.10$) 和在真实的患者数据集中 ($\text{mcc}_{\text{HYper}} = 0.79 \pm 0.04$ 和 $\text{mcc}_{\text{HYp 不, 不, 不}} = 0.38 \pm 0.10$)。对这一方法的评估是根据从 `uwa/padova` 模拟器生成的 10 个模拟病例的数据集和 5 个真实病例的临床数据集进行的, 每个病例包含葡萄糖读数、胰岛素泵和膳食 (碳水化合物) 数据。递归卷积神经网络的性能以四种算法为基准。该预测算法在 android 手机上实现, 执行时间为 6 毫秒在电话中的执行时间比 786 在 `python` 的笔记本电脑上。少 2018 年 8 月 16 日提交;v1 于 2018 年 7 月 9 日提交;最初宣布 2018 年 7 月。

评论:11 页, 7 个数字

69. 第 1807. 03032[[pdf](#),[其他](#)] cs. ms

用于自动模具计算的设在设计系统--德维托的体系结构和性能

作者:[fabio luporini](#), [michael lange](#), [mathias louboutin](#), [navjot kukreja](#), [jan hückelheim](#), [charles yourt](#), [philip witte](#), [paul h. j. kelly](#), [gerard j. goman](#), [费利克斯·赫尔曼](#)

文摘: 模具计算是许多高性能计算应用的关键部分, 如图像处理、卷积神经网络和偏微分方程的有限差分求解器。`devito` 是一个能够生成高度优化的代码的框架, 给定了用 `python` 表示的符号方程, 专门研究但不限于仿射 (模具) 代码。降低过程--从数学方程到 `c++` 代码--由德维托编译器通过一系列中间表示来执行。介绍了几种性能优化, 包括高级常用子表达式消除、平铺和并行化。其中一些是通过集成在 `devito` 编译器后端的成熟的模具优化器获得的。介绍了 `devito` 编译器的体系结构以及生成代码时应用的性能优化。利用从地震成像应用中提取的运算符证明了此类性能优化的有效性。少 2018 年 7 月 9 日提交;最初宣布 2018 年 7 月。

评论:提交给 `siam` 科学计算杂志

msc 类: 65n06;68n20

70. 第 `xiv:1807.01775`[[pdf](#),[其他](#)] cs. ms

`fluidfft`: 用于快速傅立叶变换 `hpc` 库的通用 `api` (`c++` 和 `python`)

作者:[ashwin vishnu mohan](#), [cyrille bonamy](#), [pierre augier](#)

摘要: `python` 包 `fluidfft` 提供了一个通用的 `python` `api`, 用于连续、并行地执行快速傅立叶变换 (fft) 和 `gpu`, 使用不同的 `fft` 库 (`fftw`、`p3dfft`、`pfft`、`cufft`)。 `fluidfft` 是一个全面的 `fft` 框架, 它允许 `python` 用户轻松高效地执行 `fft` 和相关任务, 例如计算线性运算符和能量谱。我们描述了由 `c++` 和 `cython` `fft` 类、`python` "运算符" 类和

pythran 函数组成的包的体系结构。该软件包提供实用程序,可轻松测试自身,并针对特定案例和特定机器上的不同 fft 解决方案进行基准测试。我们对三种不同的计算集群进行了性能扩展分析,并提出了一个微基准测试,表明 fluidfft 是使用 fft 编写高效 python 应用程序的一个有趣的解决方案。少

2018 年 7 月 3 日提交;最初宣布 2018 年 7 月。

71. 第 xiv:180. 7.01769[[pdf](#),[其他](#)] Cs. Ce

fluidsim: 模块化、面向对象的 python 包,用于高性能 cfd 模拟

作者:[ashwin vishnu mohan](#), [cyrille bonamy](#), [miguel calpe linares](#), [pierre augier](#)

文摘: 本文介绍了 python 包流体作为计算流体力学 (cfd) 求解器的可扩展框架。它是作为 FluidDyn 项目 (augier 等人, 2018 年) 的一部分开发的, 该项目旨在促进流体力学社区内的开源和开放科学合作, 既用于教育目的, 也用于研究目的。fluidsim 中的求解器是可扩展的高性能计算 (hpc) 代码, 由丰富、科学的 python 生态系统和 fluidsim 和 fluidfft 包 (mohan) 提供的应用程序编程接口 (api) 提供引擎盖下供电等人, 2018 年)。本文介绍了流体的设计方面, 即使用 python 作为主要语言;专注于代码的易用性、重用性和维护性, 而不会影响性能。简要介绍了实现的细节, 包括优化方法、功能的模块化组织和使用类实现求解器的面向对象方法。目前, 流体包括使用不同数值方法 (包括有限差分法) 解决各种物理问题的解算器。但是, 此元底纸应只涉及其伪光谱求解器的实现和性能, 特别是二维和三维纳维尔-斯托克斯求解器。我们研究了流体在最先进的 hpc 集群中的性能和可扩展性。提出了三种基于 python (dedalus, 光谱 dns) 和 fortran (ns3d) 的类似伪光谱 cfd 码, 并与流体进行了定性和定量。源代码作为一个永存储库 bitbucket.org/fluidyn-fluidsim 托管在 bitbb 上, 使用 spinx 生成的文档可以在 fluidsim.readthedocs.io 在线阅读。少

2018 年 7 月 3 日提交;最初宣布 2018 年 7 月。

72. 第: 1807. 01069[[pdf](#), [ps](#),[其他](#)] Cs. Lg

v0.3.0 的对抗鲁棒性工具箱

作者:[maria-irina nicolae](#), [mathieu sinn](#), [minh ngoc tran](#), [ambrish rawat](#), [martin wistuba](#), [valentina zantedeschi](#), [nathalie baracaldo](#), [bryant chen](#), [heikoludwig](#), [ian m. molloy](#), [ben edwards](#)

摘要: 基于深度神经网络 (dnn) 的现代人工智能系统的安全性已成为无可争辩的威胁。对抗性鲁棒性 (art) 是一个 python 库, 旨在支持研究人员和开发人员创建新的防御技术, 以及部署真实世界 ai 系统的实际防御。研究人员可以使用 art 来衡量针对最先进的技术的新型防御措施。对于开发人员, 该库提供的接口支持使用单个方法作为构建块的综合防御系统的组合。对抗鲁棒性工具箱支持机器学习模型 (以及特定于深度神经网络 (dnn)), 这些模型是在任何最流行的深度学习框架 (tensorflow、keras、pytorch 和 mxnet) 中实现的。目前, 该库的主要目的是提高视觉识别系统的对抗性鲁棒性, 但是, 设想未来的版本将包括对其他数据模式 (如语音、文本或时间序列) 的调整。art 源代码是根据 mit 许可证发布 (<https://github.com/IBM/adversarial-robustness-toolbox>) 的。该版本包括代码示例和广泛的文档

(<http://adversarial-robustness-toolbox.readthedocs.io>), 以帮助研究人员和开发人员快速入门。少

2018 年 8 月 8 日提交;v1 于 2018 年 7 月 3 日提交;最初宣布 2018 年 7 月。

评论:33 页

73. 第 1807. 0017[[pdf](#),[其他](#)] Cs. Cl

使用 wombat 嵌入透明、高效和可靠的 word

作者:[Mark-Christoph müller](#), [michael strube](#)

摘要: 我们介绍了 wobat, 这是一个 **python** 工具, 支持 nlp 实践者从代码访问单词嵌入。wombat 解决了常见的研究问题, 包括统一访问、缩放以及可靠且可重现的预处理。使用 wombat 访问单词嵌入的代码不仅更干净、更易读、更易于重用, 而且比使用标准内存中方法的代码效率要高得多: 使用 wombat 的 **python** 脚本用于评估七个大的单词嵌入集合 (总共 87m 嵌入向量) 在一个简单的 semeval 句子相似任务上, 涉及 250 个原始句子对, 在标准笔记本电脑上端到端不到 10 秒完成。少

2018 年 7 月 2 日提交;最初宣布 2018 年 7 月。

74. 第 [xiv:866.11532](#)[[pdf](#),[其他](#)] Cs. Lg

文本世界: 基于文本的游戏的学习环境

作者:[marc-亚历山大 cété](#), [Ákos kádár](#), [xghdi yuan](#), [ben kybartas](#), [tavia barnes](#), [emeravia barnes](#), [emeryfine](#), [james moore](#), [matthew hausknecht](#), [layla el asri](#), [mahmoud adada](#), [wendy tay](#), [adam trischler](#)

摘要: 我们介绍了 textworld, 这是一个沙盒学习环境, 用于在基于文本的游戏的对 rl 代理进行培训和评估。textworld 是一个 **python** 库, 用于处理文本游戏的交互式播放, 以及状态跟踪和奖励分配等后端功能。它配备了一个由我们分析过的功能和挑战的游戏策划列表。更重要的是, 它使用户能够手工制作或自动生成新游戏。它的生成机制可以精确地控制构建游戏的难度、范围和语言, 并可用于缓解商业文本游戏固有的挑战, 如部分可观察性和稀疏奖励。通过生成各种但相似的游戏集, textworld 也可以用来研究泛化和转移学习。我们在强化学习形式主义中设计了基于文本的游戏, 使用我们的框架来开发一组基准游戏, 并评估此集合和计划列表上的几个基线代理。少

2018 年 6 月 29 日提交;最初宣布 2018 年 6 月。

评论:29 页, 将出席在斯德哥尔摩 icml/ijcai 2018 年举行的计算机游戏讲习班

75. 第 [xiv:1806. 09997](#)[[pdf](#), [ps](#),[其他](#)] Cs. 艾

使用生成器的概率推理--雕像算法

作者:[皮埃尔·丹尼斯](#)

摘要: 我们提出了一种新的概率推理算法, 该算法在离散概率分布领域给出了精确的结果。该算法称为状况算法, 计算定义为直接无环图的概率模型上的边际概率分布。这些模型由定义明确的基元组成, 特别允许表达联合概率分布、贝叶斯网络、离散马尔可夫链、条件和概率算法。雕像算法依赖于基于生成器构造的变量绑定机制, 这是一种特殊形式的 coroutine;与枚举算法相关的是, 该算法在效率方面带来了重要的改进, 这使得它对其他精确的边缘化算法具有价值。在介绍了几种定义、原语和组合规则之后, 我们详细介绍了雕像算法。然后, 我们简要讨论了该算法与其他算法相比的兴趣, 并提出了可能的扩展。最后, 我们将介绍 lea 和 microlea 这两个实现雕像算法的 **python** 库, 以及几个用例。附录中提供了该算法正确性的证明。少

2018 年 8 月 2 日提交;v1 于 2018 年 6 月 24 日提交;最初宣布 2018 年 6 月。

评论:50 页, 包括 3 个附录 (v2:mods 和小更正, 增加附录 c, 并证明正确)

76. 第 [09iv:1806. 09907](#)[[pdf](#),[其他](#)] Cs. 简历

航空实验室: 自动毕业影像注册实验室

作者:robin sandkühler, christoph jud, simon andermatt, 菲利普 c. cattin

摘要: 医学图像配准是一个活跃的研究课题, 是许多医学图像分析任务的基础。虽然图像配准是一个相当一般的概念, 但通常需要专门的方法来解决特定的配准问题。就必须计算目标的梯度而言, 这些方法的制定和实施一直很困难。此外, 对于更大的图像和更复杂的转换模型和正则化项, 必须最好在 gpu 上进行评估。这阻碍了研究人员快速原型设计, 并给复制研究成果带来障碍。显然需要一个隐藏这种复杂性的环境, 将注册方法的建模和实验探索放在前景的视野中。通过 "autograd 图像注册实验室" (airlab), 我们引入了一个用于图像注册任务的开放式实验室, 在该实验室中, 目标函数的分析梯度会自动计算, 并且在该设备上执行计算。cpu 或 gpu 是透明的。它是一个实验室的研究人员和开发商, 使他们能够迅速尝试新的想法, 登记图像, 并复制注册结果已经公布。airlab 在 **python** 中实现, 使用 pytorch 作为张量和优化库, 并使用 simpleitk 进行基本图像 io。因此, 它受益于机器学习界在优化和深度神经网络模型方面取得的最新进展。本文的本草案用第一个代码段和性能分析大致概述了 airlab。随后将作为最终版本进行更详尽的介绍。少

2018 年 6 月 26 日提交;最初宣布 2018 年 6 月。

评论:记者: christoph jud, 电子邮件: christoph.jud@unibas.ch

77. 第 xiv:1806.09046[[pdf](#),[其他](#)] Cs. 简历

二维嵌入深度学习在 metagenomics 中的疾病分类

作者:thanh hai nguyen, edi prifti, yann chevaleyre, nataliya sokolovska, jean-daniel zucker

摘要: 深度学习 (dl) 技术在应用于图像、波形和文本时已显示出前所未有的成功。一般情况下, 当样品尺寸 (n) 比功能数量大得多 (D), dl 通常优于其他机器学习 (ml) 技术, 通常通过使用卷积神经网络 (cnn)。然而, 在许多生物信息学领域 (包括元基因组学), 我们遇到了相反的情况, D 是显著大于 n。在这些情况下, 应用 dl 技术将导致严重的过度拟合。在这里, 我们的目标是通过使用 cn 来改进各种疾病的分类与元基因组数据。为此, 我们建议将元基因组数据表示为图像。提出的 met2img 方法依靠分类和 t-SNE 嵌入将大量数据转换为 "合成图像"。我们将我们的方法应用于 12 个基准数据集, 包括 1400 多个元基因组样本。我们的结果显示, 与最先进的算法 (随机林 (rf), 支持向量机 (svm)) 相比有了显著的改进。我们观察到, 系统发育信息与丰度数据的整合提高了分类。该方法不仅在分类设置中很重要, 而且还可以可视化复杂的元基因组数据。met2img 是在 **python** 中实现的。少

2018 年 6 月 23 日提交;最初宣布 2018 年 6 月。

评论:一年一度的法国机器学习会议 (cap 2018)----la conférence sur l' apprentissage automatisé (cap), 2018 年 6 月, 法国鲁昂。口头陈述 c12。

<http://cap2018.litislabs.fr/Programme-en.html> 10 页、7 个数字、4 个表格

78. 第 xiv:1806.08126[[pdf](#),[其他](#)] Cs. Dm

利用拓扑工具包简化拓扑数据分析

作者:guillaume favelier, charles gueunet, attila gyulassy, Gyulassy kitware, joshua levine, jonas lukasczyk, daisuke sakurai, maxime soler, Gyulassy tierny, 威尔·厄谢尔, 齐武

摘要: 本教程介绍了从用户的角度分析和可视化科学数据的拓扑方法, 以及最近发布的用于拓扑数据分析的开源库拓扑工具包 (ttk)。在过去的二十年里, 拓扑方法的普及

和成熟程度有了相当大的提高, 既定方法的成功案例也被记录在广泛的应用中 (燃烧、化学、天体物理学、材料科学等)。在事后和现场的情况下获取和模拟数据。虽然已经出版了关于这个主题的参考教科书, 但 [ieee vis](#) 的任何教程近年来都没有涉及这一领域, 从来没有从软件层面和用户的角度来看。本教程通过为从业者、研究人员、学生和讲师提供对拓扑方法的初学者介绍来填补这一空白。特别是, 本教程没有关注理论方面和算法细节, 而是重点介绍拓扑方法如何在实际中对具体的数据分析任务 (如分割、特征提取或跟踪) 有用。本教程详细介绍了如何使用 `ttk` 完成这些任务。首先, 在介绍了拓扑方法及其在数据分析中的应用之后, 将简要概述 `ttk` 为最终用户提供的主要切入点, 即 `pareview`。其次, 将概述 `ttk` 的主要特点。将详细介绍一个正在运行的示例, 演示如何通过 `pareview`、`python`、`vtksc++` 和 `c++` 访问 `ttk` 的功能。第三, 实践课程将具体展示如何在 `pareview` 中使用 `ttk` 来执行多个具有代表性的数据分析任务。第四, `ttk` 的使用将为开发人员提供, 特别是通过描述在 `ttk` 之上构建的可视化和数据分析项目的几个示例。最后, 对 `ttk` 作为拓扑分析教学平台的使用提出了一些反馈意见。本教程的主讲人包括拓扑方法方面的专家、`ttk` 的核心作者以及来自学术界、实验室或行业的活跃用户。本教程的很大一部分将专门用于动手练习, 并将向参与者提供丰富的材料包 (包括在虚拟机中预装 `ttk`、代码、数据、演示、视频教程等)。本教程主要针对的是不是拓扑方法专家但有兴趣在日常工作中使用这些方法的学生、从业者和研究人员。我们还针对已经熟悉拓扑方法的研究人员, 他们有兴趣使用或为 `ttk` 做出贡献。

少

2018 年 6 月 21 日提交;最初宣布 2018 年 6 月。

79. [xiv:866.07832\[pdf,其他\]](#) Cs. Cl

结构 `vae`: 半监督语义解析的树状潜在变量模型

作者:[尹彭成](#), [周春婷](#), [何俊贤](#), [格雷厄姆·诺伊比希](#)

摘要: 语义解析是将自然语言 (nl) 话语转换为形式意义表示 (mr) 的任务, 通常表示为树状结构。用相应的 `mr` 注释 `nl` 话语是昂贵且耗时的, 因此标记数据的有限可用性往往成为数据驱动、监督模型的瓶颈。我们介绍了结构 `vae`, 这是一种用于半继承语义解析的变分自动编码模型, 它既从有限的并行数据中学习, 又从现成的未标记 `nl` 话语中学习。结构 `vae` 将未标记数据中未观察到的潜在 `mr` 建模为树状结构的潜在变量。在 `atis` 域和 `python` 代码生成上进行语义分析的实验表明, 使用额外的无标记数据, 结构 `vae` 优于强监督模型。少

2018 年 6 月 20 日提交;最初宣布 2018 年 6 月。

评论:[acl 2018](#)

80. [xiv:1806.06465\[pdf,其他\]](#) 反渗透委员会

铰接式对象与相互作用的 `rbo` 数据集

作者:[罗伯托·马丁马丁](#)、[克莱门斯·埃普纳](#)、[奥利弗·布洛克](#)

摘要: 我们提出了一个数据集, 其中包含人类环境中常见的 14 个铰接式物体的模型, 以及 `rgb-d` 视频序列和记录人类与它们相互作用的扳手。358 相互作用序列在不同的实验条件下 (相互作用类型、照明、透视和背景) 共 67 分钟的人类操作。与物体的每一次相互作用都用其刚性部分的地面真值和运动捕获系统获得的运动状态进行注释。对于 78 个序列的子集 (25 分钟), 我们还测量了相互作用扳手。对象模型包含每个链路的纹理三维三角形网格及其运动约束。我们提供 `python` 脚本来下载和可视化数据。这

些数据可在 <https://tu-rbo.github.io/articulated-objects/> 获得, 并在 <https://zenodo.org/record/1036660/> 托管。少

2018 年 6 月 17 日提交;最初宣布 2018 年 6 月。

评论:6 页;提交给《国际机器人研究杂志》(数据文件), sage;前两位作者的同等贡献

81. 第 xiv:1806.04973[[pdf](#),[其他](#)] Cs. CI

openedgar: 用于 sec edgar 分析的开源软件

作者:[michael j bommarito ii](#), [daniel martin katz](#), [eric m detterman](#)

摘要: openedgar 是一个开源 **python** 框架, 旨在基于美国证券交易委员会 (sec) 运营的电子数据收集、分析和检索 (edgar) 系统快速构建研究数据库。openedgar 构建在 django 应用程序框架上, 支持跨一个或多个服务器的分布式计算, 包括以下功能: (i) 检索和分析索引并从 edgar 中归档数据, (ii) 为窗体类型和文件管理等关键元数据生成表, (iii) 检索、分析和更新 cik 到股票代码和行业映射, (iv) 从归档文档中提取内容和元数据, 以及 (v) 搜索归档文档内容。openedgar 设计用于学术研究和工业应用, 并根据 mit 许可证在 <https://github.com/LexPredict/openedgar> 分发。少

2018 年 6 月 13 日提交;最初宣布 2018 年 6 月。

评论:12 页, 3 个数字, 2 个表

类:I.2.7;F.2.2;H.3.1;H.3.3;i。7

82. 第 xiv:1806.03688[[pdf](#),[其他](#)] Cs. CI

lexnlp: 法律和法规文本的自然语言处理和信息提取

作者:[michael j bommarito ii](#), [daniel martin katz](#), [eric m detterman](#)

摘要: lexnlp 是一个开源 **python** 包, 专注于自然语言处理和机器学习, 用于法律和监管文本。该软件包包括以下功能: (一) 分部文件, (二) 确定标题和章节标题等关键文本, (三) 提取超过 18 种类型的结构化信息, 如距离和日期; (四) 提取指定实体, 如公司和地缘政治实体, (v) 将文本转换为模型培训的功能, (vi) 建立无监督和监督的模型, 如单词嵌入或标记模型。lexnlp 包括基于从 sec edgar 数据库提供的真实文档以及各种司法和监管程序中提取的数千个单元测试的预先培训模型。lexnlp 设计用于学术研究和工业应用, 并以 <https://github.com/LexPredict/lexpredict-lexnlp> 分发。少

2018 年 6 月 10 日提交;最初宣布 2018 年 6 月。

评论:9 页, 0 个数字; 另见 <https://github.com/LexPredict/lexpredict-lexnlp>

类:I.2.7;F.2.2;H.3.1;H.3.3;i。7

83. 第 xiv:1806.01423[[pdf](#),[其他](#)] cs. ne

bindsnet: python 中面向机器学习的尖峰神经网络库

作者:[hananel hazan](#), [daniel j. saunders](#), [hasan khan](#), [darpan t. sanghavi](#), [hava t. siegelmann](#), [robert kozma](#)

摘要: 尖峰神经网络仿真软件的开发是一个关键组件, 能够对神经系统进行建模和开发具有生物学启发的算法。现有的软件框架支持广泛的神经功能、软件抽象级别和硬件设备, 但通常不适合快速原型设计或应用于机器学习领域的问题。在本文中, 我们描述了一个新的 **python** 包模拟尖峰神经网络, 专门针对机器学习和强化学习。我们的软件称为 bindsnet, 可快速构建和模拟尖峰网络和功能, 采用用户友好、简洁的语法。bindsnet 是建立在 pytorch 深层神经网络库之上, 可为大型尖峰网络实现快速的

cpu 和 gpu 计算。bindsnet 框架可以进行调整,以满足其他现有计算和硬件环境(例如 tensorflow) 的需要。我们还提供了一个接口到 openai 健身房图书馆,允许培训和评估尖峰网络的强化学习问题。我们认为,这个软件包促进了在大规模机器学习实验中使用尖峰网络,并展示了一些简单的例子,我们设想如何在实践中使用 bindsnet。bindsnet 代码可在 <https://github.com/Hananel-Hazan/bindsnet>

2018 年 6 月 4 日提交;最初宣布 2018 年 6 月。

84. 第 xiv.006-00875[[pdf](#),其他] Cs. Lg

在机器学习应用程序中部署自定义数据表示和近似计算

作者:mahdi nazemi, masoud pe 德拉姆

摘要: 在构建通用和定制硬件方面取得的重大进展是机器学习模型(如深度神经网络)的通用性和普遍性的关键推动因素之一。为了维持机器学习模型的这种无处不在的部署并应对其计算和存储复杂性,提供了几种解决方案,例如使用定点表示和部署近似值的模型参数的低精度表示算术运算已经被采用。研究此类解决方案在不同应用中的有效性需要将其集成到现有的机器学习框架中进行高级仿真,并在硬件中实现,以分析它们对功耗的影响,和芯片面积。洛普是一个设计空间探索的图书馆,它弥合了机器学习和高效硬件实现之间的差距。它包括一个 python 模块,它可以与一些现有的机器学习框架集成,并实现各种可定制的数据表示,包括定点和浮点以及近似算法操作。此外,它还包括一个高度参数化的 sc 级模块,它允许基于所述数据表示和算术运算合成硬件。lep 允许研究人员和设计人员使用 python 中的各种数据表示和算术运算快速比较其模型的质量,并通过在目标上对其进行综合来对比可行表示的硬件成本平台(例如, fpga 或 asic)。据我们所知,lop 是第一个允许使用自定义数据表示和近似计算技术进行软件模拟和硬件实现的库。少

2018 年 6 月 3 日提交;最初宣布 2018 年 6 月。

85. 第 1805.5.11557[[pdf](#),其他] Cs. 简历

coconet: 一种将像素坐标映射到颜色值的深层神经网络

作者:paul andrii bricman, radu tudor ionescu

摘要: 本文提出了一种深度神经网络方法,用于将图像中的二维像素坐标映射到相应的红绿蓝 (rgb) 颜色值。神经网络被称为 coconet,即协调到颜色的网络。在训练过程中,神经网络学习对其图层内的输入图像进行编码。更具体地说,网络学习了一个连续函数,它近似于在离散 2d 像素位置采样的离散 rgb 值。在测试时,给定二维像素坐标,神经网络将输出相应像素的近似 rgb 值。通过考虑每个二维像素位置,网络实际上可以重建整个学习的图像。需要注意的是,我们必须为每个输入图像训练一个单独的神经网络,即一个网络只对单个图像进行编码。据我们所知,我们首先提出了一种单独编码图像的神经方法,方法是通过学习从 2d 像素坐标空间到 rgb 颜色空间的映射。我们的神经图像编码方法具有各种低级图像处理应用,从图像编码、图像压缩和图像去噪到图像重采样和图像完成。我们进行的实验包括定量和定性结果,展示了我们的方法的效用及其优于标准基线的优势,例如双边滤波或双插值。我们的代码可在 <https://github.com/paubric/> 巨蟒-----少

2018 年 8 月 31 日提交;v1 于 2018 年 5 月 29 日提交;最初宣布 2018 年 5 月。

评论:2018 年神经信息处理国际会议接受

86. 第 1805.5.941[[pdf](#),其他] Cs. Ds

间隔中的快速随机整数生成

作者:[daniel lemire](#)

摘要: 在模拟、概率算法和统计测试中,我们通常在一个时间间隔(例如, $[0, s)$)中生成随机整数。例如,间隔内的随机整数对于费舍尔-耶茨随机洗牌是必不可少的。因此,流行的语言,如 java、python、c++、swift 和 go,将远程随机整生成函数作为其运行时库的一部分。伪随机值通常使用线性同余生成器等算法以固定位数(例如 32 位、64 位)的单词生成。我们需要函数在不引入统计偏差的情况下,在一个区间 $[0, s)$ 中将随机单词转换为随机整数。java 等编程语言中的标准函数涉及整数划分。不幸的是,部门的指示相对昂贵。我们回顾了一个无偏函数,从随机单词的来源生成远程整数,避免了概率高的整数除法。为了确定该方法的实用性,我们证明了该算法可以在 x64 处理器上实现无偏随机洗牌的速度。我们提出的方法已被 go 语言采用,以实现洗牌函数。少

2018 年 5 月 29 日提交;v1 于 2018 年 5 月 28 日提交;最初宣布 2018 年 5 月。

评论:出现在关于建模和计算机模拟的事务中

87. 第 1805.509110[[pdf](#),[其他](#)] Cs. Gr

拓扑工具包

作者:[julen tiemy](#), [guillaume favelier](#), [joshua a.levine](#), [charles gueunet](#), [michael richaux](#)

摘要: 本系统介绍了拓扑工具包 (ttk), 这是一个专为科学可视化拓扑数据分析而设计的软件平台。ttk 为标量数据的拓扑分析提供了统一、通用、高效和稳健的关键算法实现,包括:临界点、积分线、持久性图、持久性曲线、合并树、等高线树、摩斯梅尔复合物、纤维表面、连续散点图、雅可比集、reeb 空间等。由于与 paraview 的紧密集成,最终用户可以轻松访问 ttk。开发人员还可以通过各种绑定 (python、vtksc++) 轻松访问它,以便快速进行原型设计,也可以通过直接的、无依赖的 c++ 轻松集成到预先存在的复杂系统中。在开发 ttk 的同时,我们面临着一些算法和软件工程的挑战,我们在本文中进行了记录。特别是,我们提出了一种算法,用于构造一个符合分段线性设置中提取的临界点的离散梯度。该算法保证了 ttk 支持的拓扑抽象之间的组合一致性,重要的是,统一实现了多尺度探索和分析的拓扑数据简化。我们还提出了一个缓存的三角测量数据结构,它支持时间效率和泛型遍历,它根据输入简单网格的需求自动调整其内存使用情况,并隐式模拟没有内存的常规网格的三角测量开销。最后,我们描述了一个原始的软件体系结构,它保证了对 ttk 功能的内存高效和直接访问,同时仍然允许研究人员强大而简单的绑定和扩展。ttk 是开源的 (bsd 许可证),其代码、在线文档和视频教程可在 ttk 的网站上查阅。少

2018 年 5 月 22 日提交;最初宣布 2018 年 5 月。

88. 第 1805.08949[[pdf](#),[其他](#)] Cs. Cl

从堆栈溢出中学习挖掘对齐代码和自然语言对

作者:[尹鹏成](#),[邓博文](#),[陈爱德嘉](#), [博格丹·瓦西列斯库](#),[格雷厄姆·诺伊比希](#)

摘要: 对于自然语言的代码合成、代码检索和代码摘要等任务,数据驱动模型显示出了巨大的希望。但是,创建这些模型需要自然语言 (nl) 和具有细粒度对齐的代码之间的并行数据。堆栈溢出 (so) 是创建这样一个数据集的一个很有前途的源代码:问题是多种多样的,大多数问题都有相应的答案与高质量的代码段。但是,现有的启发式方法(例如,将帖子的标题与接受答案中的代码配对)的覆盖范围和所获得的 nl 代码对的正确性都是有限的。在本文中,我们提出了一种新的方法来挖掘高质量的对齐数据从

so 使用两组特征: 手工制作的特征考虑到提取的片段的结构, 和通过训练概率模型获得的对应特征利用神经网络捕获 nl 和代码之间的相关性。这些特征被输入到一个分类器中, 该分类器确定挖掘的 nl 代码对的质量。使用 **python** 和 **java** 作为测试台的实验表明, 与现有的挖掘方法相比, 该方法大大扩展了覆盖范围和准确性, 即使只使用少量标记的示例也是如此。此外, 我们发现, 即使在一种语言上训练分类器并在另一种语言上进行测试时, 也能获得合理的结果, 这表明我们有希望将 nl 代码挖掘扩展到超出我们能够注释数据的各种编程语言的编程语言。少

2018 年 5 月 22 日提交;最初宣布 2018 年 5 月。

评论:msr ' 18

89. 第: 1805.08899[[pdf](#),[其他](#)] Cs. Lg

ecorn:faststm mn 实现与数据布局优化

作者:陈博健,阿克沙伊·奈尔,吴琼思, 南迪塔·维贾伊库马尔,根能季·佩希门科

摘要: 长短时记忆递归神经网络 (lstm mn) 是一种用于分析序列数据的最先进的 (sota) 模型。当前在机器学习框架中实现的 lstm mn 通常缺乏性能或灵活性。例如, tensorflow 和 mxnet 中的默认实现调用许多微小的 gpu 内核, 从而导致启动 gpu 线程的开销过高。尽管 nvidia 的深度学习库 cudnn 可以将性能提高 2 倍左右, 但它是闭源和不灵活的, 阻碍了使用 cudnn 作为后端的框架 (如 pytorch) 的进一步研究和性能改进。在本文中, 我们引入了一种名为 ecomn 的新的 mn 实现, 它明显快于 mxnet 中的 sota 开源实现, 并且与闭源 cudnn 具有竞争力。我们表明, (1) 融合微小的 gpu 内核和 (2) 应用数据布局优化可以使我们的性能提升最大的 3 倍超过 mxnet 默认值和 1.5 倍超过 cudnn 实现。我们的优化也适用于其他 mn 单元类型, 如 lstm 变体和额定递归单位 (gru)。我们将 ecomn 集成到 mxnet **python** 库和开源库中, 以使机器学习实践者受益。少

2018 年 5 月 22 日提交;最初宣布 2018 年 5 月。

90. 第 1805.508612[[pdf](#),[其他](#)] Cs. Ds

试论时间排序的最坏情况和复杂性

作者:nicolas auger, vincent jugé, cyril nicaud, carine pivoteau

摘要: timsort 是 2002 年为 **python** 设计的一种有趣的排序算法, python 的最坏情况复杂性已经公布, 但直到我们最近的预打印才得到证实。事实上, 目前有两个略有不同的 timsort 版本分别在 **python** 和 **java** 中实现。我们提出了一个教学和有见地的证据, **python** 版本运行在 $O(n \log n)$ 。我们在分析中使用的方法也适用于 **java** 版本, 尽管并非没有非常涉及的技术细节。作为我们研究的副产品, 我们发现了 **java** 实现中的一个错误, 该错误可能会导致排序方法在执行过程中失败。我们还提供了一个证据, 证

明 python 的 timsort 运行时间是在 $O(n + n \log n)$ 日志), 其中, 好了, 好是运行的数量 (即最大单调序列), 这是一个相当自然的参数, 也是对 timsort 在部分排序的输入上的良好行为的部分解释。少

2018 年 5 月 31 日提交;v1 于 2018 年 5 月 22 日提交;最初宣布 2018 年 5 月。

91. 第 1805.08308[[pdf](#),[其他](#)] Cs. Lg

几何形状: 机器学习中黎曼几何的 python 包

作者:nina miolane, joan mathe, claire donnat, mikael j 若尔达, xavier pennec

摘要: 我们引入了地理统计, 一个巨蟒包, 对流形进行计算, 如双球、双曲空间、对称正定矩阵的空间和利群变换。我们为这些流形提供了高效且经过广泛单元测试的实现, 以及有用的黎曼度量和相关的指数和对数映射。相应的测地线距离提供了一系列直观的机器学习损失函数选择。我们还给出了相应的黎曼梯度。在地理统计中实现的操作可用于不同的计算后端, 如数字、张量流和角质。我们已经启用了 gpu 实现和集成到 keras 深度学习框架的几何形状的多个计算。本文还介绍了机器学习中的流形, 并概述了地像包, 并举例说明了它在高效和用户友好的黎曼几何中的用途。少

2018 年 5 月 21 日提交;最初宣布 2018 年 5 月。

评论:预印 nips2018

92. 第 1805.507613[[pdf](#),其他] Cs。镍

mac 地址去匿名

作者:[oisín kyne](#)

摘要: 该项目是对 wifi 探测器请求分析的一次探索, 该管理框架被描述为公开广播发件人 mac 地址的 ieee 802.11 协议的一部分。其目的是收集这些探测请求, 作为将人们的信息链接到其设备的 mac 的基础。这将使人们今后能够通过探测器请求探测器附近的设备的存在来识别身份。由于数据保护和隐私问题阻碍了人们姓名和地点的真实数据的获取, 该项目分为两部分。首先, 在 mac 地址和名称的仿真数据集中开发并测试了一种识别算法, 以证明 mac 地址识别是可行的。其次, 开发了一个分布式探针请求检测器系统, 并结合集中化的 mac 地址数据库, 证明了这些模拟 mac 地址在现实世界中是可以获得的。捕获软件最初是为 unix 系统开发的, 使用 django 供电的 web 服务器存储来自多个捕获设备的数据。利用 python 对识别算法进行了建模和测试。少

2018 年 5 月 19 日提交;最初宣布 2018 年 5 月。

评论:24 页, 12 个数字

93. 修订: 1805.06830[[pdf](#),其他] Cs。简历

差异滑动窗口: 来自差异图像的对象建议

作者:[julian müller](#), [andreas fregin](#), [klaus dietmayer](#)

摘要: 近年来, 滑窗方法在目标识别任务中得到了广泛的应用。它们保证了对要检测对象的整个输入图像的调查, 并允许对该对象进行本地化。尽管目前有深度神经网络的趋势, 滑动窗口方法仍然与卷积神经网络相结合。与基于形状、边缘或颜色检测对象的替代检测方法相比, 忽略对象的风险明显降低。然而, 滑动窗口技术大大增加了计算工作量, 因为分类器必须验证大量的对象候选项。本文提出了一种利用立体摄像机深度信息的滑动窗口方法。这导致了对象候选对象数量的大幅减少, 而不会显著降低检测精度。首先对传统的滑动窗口方法进行了理论研究。迄今的其他出版物只提到对计算费用的粗略估计。数学推导澄清了与图像和对象大小等参数相关的对象候选数。随后, 详细介绍了提出的视差滑动窗口方法。该方法利用 kititi 对象检测基准中的注释和图像对行人检测进行了评估。此外, 还与两种最先进的方法进行了比较。代码在 c++ 和 python 中可用 <https://github.com/julimueller/差异滑动窗口>中。少

2018 年 8 月 6 日提交;v1 于 2018 年 5 月 17 日提交;最初宣布 2018 年 5 月。

评论:提交给 2018 年国际能源组织

94. 第 1805.05121[[pdf](#), ps,其他] lo c

epfl 逻辑合成库

作者: [mathias soeken](#), [heinz riener](#), [winston haaswijk](#), [giovanni de micheli](#)

摘要: 我们提供了一个模块化开源 c++ 库的集合, 用于逻辑合成应用程序的开发。
alice 库是外壳接口的轻量级包装器, 是大多数逻辑合成和设计自动化应用程序的典型用户界面。它包括一个支持脚本编写的 **python** 接口。lorina 库是逻辑合成中常用的简单文件格式的解析库。它包括几种可自定义的解析算法和灵活的诊断引擎。kitty 库是一个真值表库, 用于显式表示和操作布尔函数。与二进制决策图等符号对应相比, 它所需的开销更少, 但受到要表示的布尔函数的变量数量的限制。最后, pery 是一个精确的综合库, 具有多个引擎, 以找到最佳的逻辑网络。所有库都有很好的文档记录和测试。此外, 作为仅针对源头的库, 可以很容易地用作复杂逻辑合成系统中的核心组件。
少

2018 年 5 月 14 日提交;最初宣布 2018 年 5 月。

评论:8 页, 被国际逻辑与综合研讨会接受 2018

95. 第 1805.04058[[pdf](#),[其他](#)] cs.PL

多伊 [10.114/3211366.3211349](#)

阿里阿德涅: 机器学习计划的分析

作者: [julian dolby](#), [avraham shinnar](#), [allison allain](#), [jenna reinen](#)

摘要: 机器学习已经改变了视觉和翻译等领域, 现在越来越多地用于科学领域, 在科学中, 这种代码的正确性至关重要。**python** 在机器学习中很受欢迎, 部分原因是它拥有丰富的机器学习库, 并被认为是使开发更快;但是, 与 eclipse 这样的工具相比, 这种动态语言对代码创建时的错误检测的支持较少。这对于机器学习来说尤其有问题: 考虑到它的统计性质, 带有细微错误的代码可能会运行并产生看起来可信但毫无意义的结果。这会破坏科学结果。我们报告 ariadne: 将静态框架 wala 应用于使用张量流的机器学习代码。我们为 **python** 创建了静态分析, 这是一个用于跟踪张量的类型系统--tensorflow 的核心数据结构--以及一个数据流分析来跟踪它们的使用情况。我们报告它是如何建立的, 并提出一些早期结果。少

2018 年 5 月 10 日提交;最初宣布 2018 年 5 月。

96. 第 1805.01597[[pdf](#),[其他](#)] Cs. 红外

多伊 [10.114/3209978.3210065](#)

pytrec _ eval: 一个非常快速的 **python** 接口

作者: [christophe van gysel](#), [maarten de rijke](#)

摘要: 我们向树 _ eval 信息检索评估工具包中引入了 **python** 接口 pytrec _ eval。pytrec _ eval 将 **python** 中的 trec _ eval 的参考实现作为本机扩展。我们表明, pytrec _ eval 比从 **python** 中调用 trec _ eval 作为子进程的速度快一个数量级左右。与 ndcg 的原生 **python** 实现相比, 对于实际规模的排名而言, pytrec _ eval 的速度是常人的两倍。最后, 我们在一个将 pytrec _ eval 与 pyndri 和 openai 健身房结合使用 q 学习学习查询扩展的应用程序中演示了它的有效性。少

2018 年 6 月 5 日提交;v1 于 2018 年 5 月 3 日提交;最初宣布 2018 年 5 月。

评论:信号 ' 18。第 41 届国际信息检索研究与开发研讨会

97. 第 1805.01041[[pdf](#),[其他](#)] Cs. Dm

实用的图双值化及其在近期量子计算中的应用

作者:timothy d. goodrich, eric horton, blair d. sullivan

摘要: 我们在近期量子计算硬件的最新进展和相关的嵌入问题的推动下, 对图形双值化 (奇数循环横向) 的实用技术进行了实验评估。我们从奇数循环横向 (oct) 和顶点覆盖率 (vc) 文献中组装了一个快速输入约简例程的预处理套件, 允许使用从量子文学。这些问题代表了广泛的属性, 导致 oct 实例比以前的基准更难。除了用于直接求解 oct 的组合分支算法外, 我们还研究了将各种重新配方转化为其他 np 难题 (如 vc 和整数线性规划), 从而能够使用求解器 (如 cplex)。我们发现, 对于具有时间约束的启发式解决方案, 在第二个时间限制下, 迭代压缩例程跳转从启发式解决方案开始, 执行效果最佳, 之后使用像 plex 这样的高度调谐求解器是值得的。精确的解也主要由 plex 的 ilp 配方控制, 通常由一个数量级决定。然后将这些结果与合成图语料库上的结果进行比较, 表明我们的结果是稳健的, 可以推广到其他域数据。最后, 我们在开源套件中提供所有代码和数据, 包括用于访问约简例程和分支算法的 **python** api, 以及用于完全复制结果的脚本。少

2018 年 10 月 26 日提交;v1 于 2018 年 5 月 2 日提交;最初宣布 2018 年 5 月。

评论:使用合成数据更新的研究

类:G.2.2;G.2。3

98. 第 09iv:1805. 00979[pdf,其他] Cs。Lg

模块化: python 的模块化主动学习框架

作者:tivadar danka, peter horvath

摘要: modal 是 **python** 的模块化主动学习框架, 旨在简化主动学习研究和实践。它的显著特点是 (i) 清晰的模块化面向对象的设计; ii) 与科学学习模型和工作流程完全兼容。这些功能使快速原型制作和易于扩展成为可能, 有助于开发现实生活中的主动学习管道和新颖的算法。modal 是完全开源的, 托管在

<https://github.com/cosmic-cortex/modAL> 的 github 上。为了保证代码质量, 提供了广泛的单元测试, 并应用了持续集成。此外, 还提供了包含多个教程的详细文档, 以方便使用。该框架以 ypi 版本提供, 并在麻省理工学院的许可证下分发。少

2018 年 5 月 2 日提交;最初宣布 2018 年 5 月。

评论:5 页, 1 个图

99. 第 6.0 5.00329[pdf,其他] Cs。简历

deepdiva: 一种用于可重复实验的高功能 python 框架

作者:mik 其 mare alberti, vinaychandran pondenkandath, marcel würsch, lof ingold, marcus liwicki

文摘: 我们介绍 deepdiva: 一种基础设施, 旨在通过大量有用的分析功能, 快速、直观地设置可重复的实验。复制科学结果可能是一种令人沮丧的经历, 不仅在文档图像分析方面如此, 在一般的机器学习中也是如此。使用 deepdiva, 研究人员可以用非常有限的信息复制给定的实验, 也可以与他人分享自己的实验。此外, 该框架还提供了大量的功能, 如样板代码、跟踪实验、超参数优化以及数据和结果的可视化。为了证明这个框架的有效性, 本文介绍了手写文档分析领域的案例研究, 研究人员从集成功能中受益。deepdiva 在 **python** 中实现, 并使用深入的学习框架 pytorch。它完全是开源的, 可通过 divaservices 作为 web 服务访问。少

2018 年 4 月 23 日提交;最初宣布 2018 年 5 月。

评论:在第十六届手写识别前沿国际会议 (icfhr) 上提交, 6 页, 6 位数字

100. 第 6.0 5.00290[[pdf](#),[其他](#)] Cs. Ce

多孔介质两相流的 hp 自适应不连续加勒金方法的 python 框架

作者:[andedas dedner](#), [birane kane](#), [robert klöfkom](#), [martin nolte](#)

摘要: 本文提出了一个求解多孔介质中两相流问题的框架。离散化是基于不连续加勒金方法, 包括局部网格自适应和多项式程度的局部选择。该方法是使用新的 python 前端 dune-ffapy 到开源框架 dune 实现的。用于模拟的代码可作为木星笔记本, 并可通过 docker 容器使用。我们提出了许多时间步进方法, 从经典的 impes 方法到完全耦合隐式方案。离散化的实施是非常灵活的, 允许测试两相流模型和适应策略的不同公式。少

2018 年 5 月 1 日提交;最初宣布 2018 年 5 月。

评论:关键词: dg, hp 适应, 两相流, impes, 完全隐式, 沙丘, python, porous 媒体. 28 页, 9 个数字, 各种代码段

msc 类: 65m08;65m60;65m50

101. 第 1804. 10869[[pdf](#),[其他](#)] q-fine. tr

概率图模型在原油价格预测中的应用

作者:丹麦语 [a. alvi](#)

摘要: 本文研究了概率图形模型 (pgm) 在原油价格预测中的应用。这项研究很重要, 因为原油在全球经济中发挥着非常关键的作用, 因此是工业增长的一个非常关键的宏观经济指标。鉴于影响原油价格的宏观经济因素很多, 如欧佩克国家的石油供应、经合组织国家的石油需求、地缘政治和地缘经济等许多变量----概率图形模型----之间的变化让我们通过学习图形结构来理解。本文提出将大量原油因子的数据浓缩到一个图形模型中, 试图对原油价格进行准确的预测。研究项目在 python 中使用不同的库, 以构建原油市场模型。本文的实验研究了金融市场石油交易中常见的三大挑战。它调查的第一个挑战是学习石油市场结构的过程;从而使原油交易员能够了解影响原油市场的不同物理市场因素和宏观经济指标, 以及它们与以下因素 (因果关系) 的关系。它解决的第二个挑战是勘探和利用现有数据和在预测石油市场行为方面学到的结构。它研究的第三个挑战是如何验证所构建模型的性能和可靠性, 以便将其部署到金融市场。作为研究的一部分, 还提出了预测原油价格的概率框架的设计和实现。少

2018 年 4 月 29 日提交;最初宣布 2018 年 4 月。

102. 第 1804. 10247[[pdf](#),[其他](#)] Cs. 艾

对机器人物流内部领域进行实验

作者:[martin gebser](#), [Philipp obermeier](#), [thomas otto](#), [torsten schaub](#), [orkunt sabunu](#), [van nguyen](#), [tran cao son](#)

摘要: 我们介绍了 asprilo [1] 框架, 以促进针对复杂动态应用的方法的实验研究。为此, 我们选择了机器人内部物流领域。这一领域不仅在当今第四次工业革命的背景下具有高度相关性, 而且还将众多具有挑战性的问题结合在一个统一的框架内。这包括多代理规划、行动推理、变更、资源、策略等。作为回报, asprilo 允许用户研究有效性和可扩展性方面的替代解决方案。尽管 asprilo 依赖于答案集编程和 python, 但任何符合其面向事实的接口格式的系统都可以随时使用它。这使得它对基准测试和教学具有吸引力, 远远超出了逻辑编程的范围。更确切地说, asprilo 由多功能基准生成器、解决方案

检查器和可视化工具以及一系列具有各种 asp 技术的参考编码组成。重要的是,可视化工具的动画功能对于复杂的场景(如物流内部)是不可或缺的,以便检查有效和无效的解决方案候选方案。此外,它还允许以图形方式编辑基准布局,可用作生成基准套件的基础。[1] asprilo 代表回答集合编程为机器人内部后勤少

2018 年 4 月 26 日提交;最初宣布 2018 年 4 月。

评论:在第 33 届国际逻辑编程会议 (iclp 2018) 上发表的论文, 英国牛津, 2018 年 7 月 14 日至 7 月 17 日, 18 页, latex, 8 个 pdf 数字 (arxiv/yymm)。nnnnn)

103. 第 xiv:804.0. 8032[pdf,其他] Cs. 艾

一种基于信道的贝叶斯网络精确推理算法

作者:巴特·雅各布斯

摘要: 本文基于最近提出的贝叶斯网络在信道方面的组合语义, 提出了一种新的精确贝叶斯推理算法。本文重点介绍了该算法背后的思想, 包括贝叶斯网络的线性化 ("拉伸"), 然后是正向状态变换和向后谓词变换的组合, 同时积累了证据的方式。将 python 中算法原型实现的性能与标准实现 (pgmpy) 进行了简要比较: 第一个结果显示了竞争性能。少

2018 年 4 月 21 日提交;最初宣布 2018 年 4 月。

msc 类: 62f15;18c50 类: F.3.2;I.2。3

104. xiv:1804. 05429[pdf,其他] cs. ne

gnowee: 约束、黑匣子、组合混合整数设计的混合元启发式优化算法

作者:james bevins, rachel Slaybaugh

文摘: 本文介绍了一种基于 python 的模块化、开源混合元启发式优化算法 (可从 <https://github.com/SlaybaughLab/Gnowee> 中获得)。gnowee 设计用于快速收敛到几乎全球范围内的最佳解决方案, 以解决复杂的、受约束的核工程问题, 这些问题具有混合整数和组合设计向量以及高成本、噪声、不连续、黑匣子目标函数评估。gnowee 的混合元启发式框架是一组多样化、稳健的启发式的新组合, 这些启发式方法能够在广泛的优化问题中适当地平衡多样化和强化策略。这种新的算法是专门为优化复杂的核设计问题而开发的;激励研究的问题是设计材料堆叠, 将中子能谱修改为特定的目标光谱, 用于核医学、核取证、核物理等领域。然而, 这种算法在核界内外都有更广泛的潜在应用。为了证明 gnowee 对各种问题类型的行为, gnowee 和几个成熟的元启发式算法之间的比较是针对一组 18 个连续的、混合整数和组合的基准。这些结果表明, gnowee 在广泛的设计空间中具有卓越的灵活性和收敛性。我们预计, 这种广泛的适用性将使这一算法在许多复杂的工程应用中成为理想的算法。少

2018 年 4 月 15 日提交;最初宣布 2018 年 4 月。

评论:43 页, 7 张桌子, 6 个数字

105. 第 1803: 09544[pdf,其他] cs.PL

一种基于路径的程序属性预测表示方法

作者:uri alon, meital zilberstein, omer levy, elan yahav

摘要: 预测程序属性 (如名称或表达式类型) 具有广泛的应用范围。它可以减轻编程任务, 提高程序员的工作效率。从课程中学习时的一个主要挑战是如何以促进有效学

习的方式来表示程序.我们提出了一个基于一般的基于路径的代表从课程中学习。我们的表示纯粹是句法的, 是自动提取的。主要的思想是使用抽象语法树 (ast) 中的路径来表示程序。这使得学习模型能够利用代码的结构化性质, 而不是将其视为一个平面序列的令牌。我们证明这种表示是通用的, 可以: (i) 涵盖不同的预测任务, (ii) 驱动不同的学习算法 (对于生成模型和判别模型), 以及 (iii) 跨不同编程语言工作。我们评估预测变量名、方法名称和完整类型任务的方法。我们使用我们的表示来驱动基于 crf 和基于 word2vec 的学习, 用于四种语言的程序: javascript、java、python 和 c#。我们的评估表明, 与针对特定任务的手工表示跨不同任务和编程语言相比, 我们的方法获得了更好的结果。少

2018 年 4 月 22 日提交;v1 于 2018 年 3 月 26 日提交;最初宣布 2018 年 3 月。

评论:将出现在 2018 年 pldi 中

106. 第 [xiv:1803.09371](#)[pdf,其他] Cs. CI

多伊 [10.114/3178876.3186081](#)

staqc: 一种系统挖掘的来自堆栈溢出的问题代码数据集

作者:[姚子宇](#),[焊接](#), [陈伟鹏](#),[孙欢](#)

摘要: 堆栈溢出 (so) 一直是自然语言问题及其代码解决方案 (即问题代码对) 的重要来源, 这些都是许多任务 (包括代码检索和注释) 的关键。在大多数现有的研究中, 问题代码对是以启发式收集的, 质量往往很低。本文研究了一个新的问题, 系统地挖掘问题代码对从堆栈溢出 (而不是启发式收集)。它被表述为预测代码段是否是问题的独立解决方案。我们提出了一种新的双视图分层神经网络, 它可以捕获代码段 (即两个视图) 的编程内容和文本上下文来进行预测。在 python 和 sql 域中的两个手动注释的数据集上, 我们的框架大大优于启发式方法, f1 和精度至少高出 15%。此外, 我们还介绍了 staqc (堆栈溢出问题代码对), 这是迄今为止最大的数据集, 约 148k python 和~ 120k sql 问题代码对, 使用我们的框架自动从 so 中挖掘。在各种案例研究中, 我们证明了 staqc 可以极大地帮助开发需要数据的模型, 以便将自然语言与编程语言相关联。少

2018 年 3 月 25 日提交;最初宣布 2018 年 3 月。

评论:接受 2018 年网络会议 (前 www 2018), 11 页, 6 位数字

107. 第 [1803.08790](#)[pdf] Cs. 红外

支持向量机对罗兴亚人运动评语的情绪分析

作者:[hemayet ahmed chowdhury](#), [tanvir alam nibir](#), [md. saiful isam](#)

摘要: 罗兴亚人运动和危机在孟加拉国的政治和经济状况中引起了巨大的轩然。难民运动是一个反复发生的事件, 脸谱等社交媒体上仍有大量意见形式的数据, 对这些数据的分析很少。为了分析所有罗兴亚人相关帖子的评论, 我们必须创建和修改一个基于支持向量机算法的分类器。该代码在巨蟒中实现, 并使用科学学习库。目前还没有关于罗兴亚人分析的数据集, 因此我们必须使用自己的数据集, 即 2500 条正面评论和 2500 条负面评论。我们专门使用了具有线性内核的支持向量机。我们以前使用天真的贝叶斯算法在同一数据集中进行了一次实验, 但没有产生令人印象深刻的结果。少

2018 年 3 月 22 日提交;最初宣布 2018 年 3 月。

108. 第 [xiv:1803.07764](#)[pdf,其他] cse

估计源代码的缺陷: 一种基于 github 内容的预测模型

作者:ritu kapur, balwinder sodhi

摘要: 本文提出的两个主要贡献是: i) 一种方法, 用于构建包含从开源软件 (oss) 和相关错误报告中提取的源文件中提取的源代码特征的数据集, 二) 估计缺陷的预测模型给定的源代码。这些工件可用于构建与多个自动化软件工程领域相关的工具和技术, 如错误本地化、代码评审、建议和程序修复。为了实现我们的目标, 我们首先提取 github 上存在的源代码文件的编码样式信息 (例如与源代码中使用的编程语言构造相关的信息)。然后提取与这些源代码文件关联的 bug 报告 (如果有) 中可用的信息。这样提取的非结构化信息就会被转换为结构化知识库。我们审议了来自 20 个不同 github 存储库的 30400 多个源代码文件, 其中包括 4 个 bug 跟踪门户中大约 14950 个相关的 bug 报告。所考虑的源代码文件是用四种编程语言 (即、c、c++、java 和 **python**) 编写的, 属于不同类型的应用程序。然后使用知识库训练用于估计给定输入源代码的缺陷的机器学习 (ml) 模型。为了选择最佳的 ml 模型, 我们评估了 8 种不同的 ml 算法, 如随机林、k 最近邻居和支持向量机, 它们具有大约 50 个参数配置, 以比较它们在我们任务中的性能。我们的研究结果之一表明, 最好的 k 折叠 (k=5) 交叉验证结果是通过 nusvm 技术, 给出了 f1 的平均分数 0.914。少

2018 年 3 月 21 日提交;最初宣布 2018 年 3 月。

评论:提交给 2018 年。关键词: 维护软件;源代码挖掘;软件缺陷识别;自动化软件工程;软件工程中的人工智能

109. 第 1803.06226[pdf,其他] cs. ms

字形: 符号回归工具

作者:markus quade, jien gout, markus abel

摘要: 我们提出了字形-一个 **python** 包, 用于基于遗传编程的符号回归。字形是为实际世界实验所允许的数值模拟而设计的。对于实验者来说, 字形遥控器提供了任务的分离: zeromq 接口将遗传编程优化任务与实验 (或数字) 运行的评估分开。可在 <http://github.com/ambrosys/glyph> 访问字形。领域专家能够轻松地在实验中使用符号回归, 即使他们不是专家程序员。通过通用的界面设计, 可保持较高的重用潜力。字形可在 pipi 和 github 上使用。少

2018 年 3 月 21 日提交;v1 于 2018 年 3 月 13 日提交;最初宣布 2018 年 3 月。

评论:提交给 josr. arxiv 管理说明: 文本与 arxiv:1612-015276 重叠

110. 第 18005536[pdf,其他] Cs. 简历

野外二维人脸图像密集三维重建的评价

作者:冯振华,帕特里克·胡伯,约瑟夫·基特勒,彼得 jb 汉考克,吴晓军,赵启军,保罗·科本,马蒂亚斯·雷茨奇

摘要: 本文研究了野外单个二维图像对密集三维人脸重建的评价。为此, 我们组织了一个竞赛, 提供了一个新的基准数据集, 其中包含 135 名受试者的 20002d 面部图像及其 3d 地面真实扫描。与以往的竞争或挑战不同, 这一新的基准数据集的目的是使用真实、准确和高分辨率的 3d 地面真实扫描来评估三维密集人脸重建算法的准确性。除了数据集之外, 我们还为评估提供了标准协议和 **python** 脚本。最后, 我们在新的基准数据集上报告了三个最先进的 3d 人脸重建系统所获得的结果。本次比赛与 2018 年第 13 届 ieee 自动面部和手势识别会议同时举办。少

2018 年 4 月 20 日提交;v1 于 2018 年 3 月 14 日提交;最初宣布 2018 年 3 月。

评论:8 页

111. 第 [xiv:18004884](#)[pdf,其他] Cs。Db

idel: 数据库内实体与神经嵌入链接

作者:[torsten kilias](#),[亚历山大·洛瑟](#), [felix a.gers](#), [richard koopmanschap](#), [ying zhang](#), [martin kersten](#)

摘要: 我们提出了一种新的体系结构,即数据库内实体链接 (idel),在该体系中,我们将分析优化的 rdbms monetdb 与神经文本挖掘能力集成在一起。我们的系统设计抽象了 monetdb 大多数神经实体链接系统的核心任务。据我们所知,这是第一个在数据库中集成实体链接的实际实现的系统。我们利用 monetdb 的能力,通过在 python 中实现的用户定义函数 (udf) 支持数据库内的分析。这些函数称为用于神经文本挖掘的机器学习库,如 tensorflow。该系统利用 monetdb 在数据库内核中嵌入 python 进程并在 numpy 数组中交换数据的能力,实现了数据传送和转换的零成本。idel 用神经嵌入表示联合向量空间中的文本和关系数据,并可以用模糊的实体表示来补偿错误。为了检测匹配实体,我们提出了一种新的基于关节神经嵌入的相似函数,该函数是通过最大限度地减少对对比排序损失来学习的。该函数利用高维索引结构快速检索匹配实体。我们首次使用 webnlg 语料库进行的实施和实验显示了 idel 的有效性和潜力。少

2018 年 3 月 13 日提交;最初宣布 2018 年 3 月。

评论:这份手稿是提交给 vldb2018 的一篇论文的预印

112. 第 [xiv:18004774](#)[pdf,其他] Cs。哦

多伊 [10.3389/fphys.2018.01046](#)

canalization: 用于量化布尔网络中的控制和 can 化的巨蟒包

作者:[rion brattig correia](#),[亚历山大 j.gates](#), [xuanwang](#), [luis m. rocha](#)

摘要: 逻辑模型提供了一个简单但强大的手段来理解生化调节的复杂动力学,而不需要估计动力学参数。然而,即使是简单的自动机组件也会导致在聚合到网络时在计算上难以解决的集体动态。在前面的工作中,我们证明了生化调节的自动机网络模型是高度规范化的,其中许多可变状态及其分组是冗余的 (马克-皮塔和罗查,2013 年)。这种运河化的精确制图和测量简化了这些模型,使即使是非常大的网络也易于分析。此外,运河化在布尔网络动力学的控制、鲁棒性、模块化和临界性方面发挥着重要作用,特别是用于模拟生化调节的网络动力学 (gates 和 rocha, 2016 年;盖茨等人,2016 年;manicka, 2017 年)。在这里,我们描述了一个新的公开提供的 python 包,它提供了必要的工具来提取、测量和可视化布尔网络模型中存在的冗余。它提取了在这些模型中最有效地控制动力学的途径,包括它们的有效图形和动力学规范化地图,以及其他工具来发现最小的控制变量集。少

2018 年 5 月 9 日提交;v1 于 2018 年 3 月 9 日提交;最初宣布 2018 年 3 月。

评论:提交给生理学前沿的系统生物学部分

msc 类: 94c (小学) 93;92c42 (中学)类: g.4;i.1;j。3

日记本参考:生理学前沿, 9: 1046, 2018

113. 第 [xiv:18002857](#)[pdf,其他] cs。cy

性别管理和紧凑性: 实施灵活性和滥用

作者:[richard bames](#), [justin solomon](#)

摘要: 选区的形态可能表明, 它是出于政治动机, 还是被篡改。因此, 量化地区的形态, 特别是其紧凑性, 是政治和公民权利的一项关键任务。越来越多的文献建议并从数学上分析了紧凑性的措施, 但在实践中如何计算这些分数却很少。在这里, 我们考虑在解释和实施一套流行的紧凑型分数时必须做出的一些决定的效果。我们表明, 在量化紧凑性方面做出的选择本身可能成为政治工具, 其看似无害的决定导致不同的分数。我们表明, 当使用全方位的实现灵活性时, 可以滥用它, 使明显杂乱无章的地区在数量上显得合理。这使得使用紧凑作为立法或司法标准来抵制不公平的重新划分做法变得更加复杂。本文伴随着 `c++`、`python` 和 `r` 中的包的发布, 这些软件包正确、高效且可重复地计算各种紧凑性分数。少

2018 年 3 月 7 日提交;最初宣布 2018 年 3 月。

评论:10 页, 17 个数字, 1 个表

114. 第 [xiv:18002337](#)[pdf] Cs. Sy

基于开源软件的电力系统广域测量控制的网络物理试验台

作者:[cui hantao](#), [fingim li](#), [kevin tomsovic](#), [siqi wang](#), [riyasat azim](#), [yidan lu](#), [haoyu yuan](#)

摘要: 电力系统是一个网络物理系统, 在物理系统中有动力流, 在网络中有信息流。仿真对于了解电力系统的动力学和控制至关重要, 但在这些研究中, 底层通信系统历来被忽略。本文旨在满足模拟实际电力系统运行的日益增加的需求, 包括物理系统、能源管理系统、通信系统以及新兴的基于广域测量的控制。本文提出了一种基于流式遥测和相量测量单元数据的网络物理试验台设计与实现, 用于验证和演示广域控制方法。所提出的解耦体系结构由基于微分代数方程的物理系统模拟器、软件定义的网络、用于设计 `ems` 系统的脚本语言环境和控制系统组成, 所有这些都是在集成的超过行业标准的通信协议。建议的测试平台是使用由 `python` 调度员管理的开源软件包实现的。最后, 在所实现的试验台中进行了演示, 说明了两种基于广域测量的控制--系统分离控制和分层电压控制。少

2018 年 3 月 6 日提交;最初宣布 2018 年 3 月。

评论:提交给 `iet cps`

115. 第 [xiv:180001094](#)[pdf,其他] Cs. Sd

多伊 [10.2110/joss.00749](#)

演讲-语音处理与识别图书馆

作者:[amirsina torfi](#)

摘要: `speechpy` 是一个开源 `python` 包, 包含语音预处理技术、语音功能和重要的后处理操作。它提供了最常用的语音功能, 包括 `mfcc` 和过滤库能量, 以及过滤库的日志能量。该包的目的是为研究人员提供一个简单的工具, 用于语音特征提取和处理目的, 在应用中, 如自动语音识别和扬声器验证。少

2018 年 5 月 25 日提交;v1 于 2018 年 3 月 2 日提交;最初宣布 2018 年 3 月。

日记本参考:开源软件杂志, 3 (27), 749, 2018

116. 第 [xiv:1803.00902](#)[pdf] Cs. Cl

萎缩, 德语形态学分析仪

作者:[duygu altinok](#)

摘要: 萎缩是一种用于德国人的形态学分析仪。它是建立在大的, 压缩词典从德国形态词典。还提供了一种基于德国退变后缀的猜测器。对于德语, 我们提供了最先进的形态分析仪。demorphy 是在 **python** 中实现的, 易于使用和附带的文档。该一揽子计划适用于学术和商业目的, 并提供许可许可。少

2018 年 3 月 2 日提交;最初宣布 2018 年 3 月。

评论:7 页, 2 个数字

117. 第 [xiv:1803.00105\[pdf\]](#) [cs. cy](#)

计算国际关系: 编程、编码和互联网研究能为学科做些什么?

作者:[h. akin unver](#)

摘要: 在过去几年里, 由于通信技术的大量进步和大量个人数据的日常生产, 计算社会科学成为一门技术性很强、很流行的学科。由于过去十年人均数据产量大幅增加, 无论是从规模、字节还是从细节、心率监测器、互联网连接的电器、智能手机、社会科学家提取有意义的社会政治能力来看, 都是如此数字数据中的人口统计信息也有所增加。计算国际关系 (comintt) 在方法上存在巨大差距, 该公司指的是使用数据挖掘、自然语言处理、自动文本分析、网络刮擦、地理空间分析和机器等工具的一种或多种组合学习提供更大、更有组织的数据来测试更先进的红外理论。在概述了计算 ir 的潜力以及红外学者如何建立计算机科学的技术熟练程度, 如从 **python**、r、qgis、arcgis 或 github 开始, 本文将重点介绍作者的一些作品。为红外学生提供了如何思考计算红外的想法。本文认为, 计算方法超越了定性方法和定量方法之间的方法论分歧, 为构建真正的多方法研究设计奠定了坚实的基础。少

2018 年 2 月 28 日提交;最初宣布 2018 年 3 月。

118. 建议: [1802.08960\[pdf,其他\]](#) 反渗透委员会

bonnet: 使用 cnn 的机器人语义分割的开源训练和部署框架

作者:[andres milioto](#), [cyrell stachniss](#)

摘要: 对于应该与其环境交互的机器人来说, 解释场景的能力是一个重要的能力。例如, 对机器人面前的东西的了解是导航、操作或规划的关键。语义分割用类标签标记图像的每个像素, 从而为机器人提供对周围环境的详细语义注释。卷积神经网络 (cnn) 成为解决这类问题的常用方法。然而, 可用的培训软件和 cnn 在真正机器人中的集成, 是相当分散的, 难以用于非专家, 尽管有几个高质量的开源框架用于神经网络的实现和培训。在本文中, 我们提出了一个新的框架称为 bonnet, 解决了这个碎片问题。它提供了一种模块化的方法来简化语义分割 cnn 的训练, 而不考虑使用的数据集和预期的任务。此外, 我们还在一个真正的机器人平台上处理部署问题。因此, 我们不建议在本文件中采用美国有线电视新闻网的新方法。相反, 我们提供了一个稳定且易于使用的工具, 使这一技术在自主系统的环境中更容易接近。从这个意义上说, 我们的目标是缩小计算机视觉研究与其在机器人研究中的应用之间的差距。我们为培训和部署提供开源框架。培训界面是使用 tensorflow 在 **python** 中实现的, 部署接口提供了一个 c++ 库, 可以轻松集成到现有的机器人代码库、ros 节点和两个独立的标签预测应用程序中。图像和视频。少

2018 年 2 月 25 日提交;最初宣布 2018 年 2 月。

评论:提交给 [ieee 机器人和自动化信函 \(ra-l\)](#) 2018 年

119. 建议: [1802.08522\[pdf,其他\]](#) [cs. it](#)

多伊 [10.1049/joe.2014.0055](https://doi.org/10.1049/joe.2014.0055)

simcommsys: 消除纠错代码模拟中的错误

作者: [john a. briffa](#), [stepfan wweemeyer](#)

摘要: 在本文中, 我们介绍了模拟, 一个通信系统的模拟器, 我们正在开源许可证下发布。该项目的核心是一组定义通信系统组件的 `c++` 库和一个分布式蒙特卡洛模拟器。主要感兴趣的是错误控制编码组件, 在该组件中实现了各种二进制和非二进制代码, 包括涡轮增压、ldpc、废除积累和 reed-solomon。该项目还包含许多实现通信系统各个阶段 (如编码器和解码器) 的现成生成二进制文件、一个完整的模拟器和一个系统基准。最后, simcommsys 还提供了许多 shell 和 **python** 脚本来封装常规用例。只要所需的组件在 simcommsys 中已经可用, 用户就可以模拟自己设计的完整通信系统, 而无需进行任何额外的编程。开发 (只需要实现新组件) 和使用 (模拟特定结构) 的严格分离鼓励了实验工作的重现性, 并降低了出错的可能性。在对框架进行概述之后, 我们提供了一些如何使用该框架的示例, 包括简单编解码器的实现、通信系统的规范及其模拟。少

2018 年 2 月 23 日提交;最初宣布 2018 年 2 月。

日记本参考: j. a. briffa 和 s. wsemeyer, "Simcommsys: 从纠错代码模拟中消除错误", iet 工程杂志, 2014 年 6 月

120. **建议: 1802.08 252**[pdf,其他] Cs。Ds

iisignature 库: 迭代积分签名和日志签名的有效计算

作者: [杰里米·雷森施泰因](#), [本杰明·格雷厄姆](#)

摘要: 在统计和机器学习中。我们提出了有效计算这些签名的算法, 并对其性能进行了基准测试。我们以 **python** 包的形式发布这些方法。

2018 年 2 月 22 日提交;最初宣布 2018 年 2 月。

评论: 18 页

121. **建议: 1802.06651**[pdf, ps,其他] cs.PL

计算列表: 一种具有强制性特征的扩展功能语言

作者: [domenico sacca](#) ', [angelo furfaro](#)

摘要: calculist (带有列表操作的计算器) 是一种用于教学功能编程的教育语言, 具有一些命令性和副作用的功能, 在程序员的明确要求下启用。除了字符串和列表外, 语言本机还支持 json 对象。该语言采用类似 **python** 的语法, 并允许通过 repl (读取评估打印循环) 外壳与用户进行交互式计算会话。编译生成的对象代码是最终将由 calculist 虚拟机 (clvm) 执行的程序。少

2018 年 2 月 19 日提交;最初宣布 2018 年 2 月。

122. **建议: 1802. 06 224**[pdf] cse

从对象 z 规范生成 python 代码

作者: [a. f. al atzawi](#), [m. bettaz](#), [h. m. al-refai](#)

摘要: 对象-z 是一种面向对象的规范语言, 它将 z 语言扩展为类、对象、继承和多态性, 可用于将复杂系统的规范表示为对象的集合。有许多现有的作品将 object-z 映射到 `c++` 和 `java` 编程语言。由于 **python** 和预料与 z 有许多相似之处, 两者都是面向对象的范式、支持集理论和谓词演算, 因此 **python** 是一种函数式编程语言, 自然更接近于形式化规范, 我们提出了一个从 obtt-z 规范到 **python** 代码的映射, 它涵盖了一些

Object-Z 构造, 并在 **python** 中表达了它的规范来验证这些规范。验证用于使用 **lambda** 函数和 **python** 的装饰器构建的包含的先决条件、后置条件和不变量的映射。这项工作发现, **python** 是开发将对象 z 规范映射到 **python** 的库的一种很好的语言。

2018 年 2 月 17 日提交;最初宣布 2018 年 2 月。

评论:12 页, 3 个数字

日记本参考:国际软件工程与应用杂志 (ijsea), 第 8 卷, 第 4 期, 2017 年 7 月

123. **建议: 180005022[pdf] cse**

pyfml-功能建模的文本语言

作者:[a. f. al-阿扎 awi](#)

摘要: 特征模型是在软件产品线设计和实现中捕获可变性的典型方法。为此, 大多数作品使用由命题逻辑表示并由 propositional 或 java 编程语言实现的有限图形表示法自动执行特征模型。这些工作不能正确地组合经典功能模型的扩展, 也不能提供可伸缩性来实现大型问题。在本文中, 我们提出了一种基于 **python** 编程语言 (pyfml) 的文本特征建模语言, 该语言将经典特征模型与实例特征的基数和属性进行了推广, 并以高光为扩展。复制和复杂的逻辑和数学交叉树约束。textx meta 语言用于构建 pyfml 来描述和组织特征模型依赖关系, 而 py 约束问题求解器则用于实现特征模型的可变性及其约束验证。本工作提供了一种文本人可读的语言来表示特征模型, 并将特征模型描述直接映射到面向对象的表示中, 供约束问题求解器用于计算。此外, 提出的 pyfml 使特征建模的表示更有表现力, 以处理复杂的软件产品线表示和使用金字塔约束问题解决者少

2018 年 2 月 17 日提交;v1 于 2018 年 2 月 14 日提交;最初宣布 2018 年 2 月。

评论:13 页, 13 图, 29 折射

日记本参考:国际软件工程与应用杂志 (isea), 第 9 卷, 第 1 期, 2018 年 1 月

124. **建议: 1802. 04967[pdf, ps,其他] Cs. Lg**

deslib: python 中的动态集成选择库

作者:[rafael m. o.cruz](#), [luiz g. Hafemann](#), [robert sabourin](#), [george d. c. cavalcanti](#)

摘要: deslib 是一个开源的巨蟒库, 提供了几种动态选择技术的实现。该库分为三个模块: (一) dcs, 包含动态分类器选择方法 (dcs) 的实现;(ii) des, 其中载有动态集成选择方法的实施;(iii) 静态, 采用静态集成技术。该库有完整的文档记录 (可在阅读文档上在线提供的文档), 具有较高的测试覆盖率 (覆盖率. io) 以及代码质量 (横向)。文档、代码和示例可在其 github 页面上找到: <https://github.com/Menelau/DESlib>。少

2018 年 2 月 15 日提交;v1 于 2018 年 2 月 14 日提交;最初宣布 2018 年 2 月。

评论:论文介绍 deslib: python 中的动态集成选择库

125. **建议: 18004766[pdf] Cs. 直流**

多伊 [10.1109/TCC.2017.2656088](#)

snc: 一种使用实时编译进行符号数字计算的云服务平台

作者:[张鹏](#),[刘月明](#),[邱美康](#)

摘要: 云服务已广泛应用于 it 行业和科学研究。通过使用云服务, 用户可以将计算任务和数据从本地计算机移动到远程数据中心。通过通过轻量级和移动设备访问基于

internet 的服务,用户可以在功能强大的计算机上部署多样化的云应用程序。科学计算领域这一范式的关键驱动因素包括强大的计算能力、按需配置和跨平台互操作性。但是,为了充分利用云服务进行科学计算,我们需要设计一个特定于应用程序的平台,以帮助用户高效迁移其应用程序。在此,我们提出了一个用于符号-数值计算的云服务平台--snc。snc 允许云用户通过 c++、python、java api 和 snc 脚本将任务描述为符号表达式。使用 llvmlvm 进行实时 (jit) 编译,用于将用户代码编译为计算机代码。我们实现了 snc 设计,并在几个流行的云平台 (包括线性最小化、蒙特卡罗积分、有限元组装和多体动力学) 上测试了广泛的符号数值计算应用 (包括非线性最小化、蒙特卡罗积分、有限元组装和多体动力学)。谷歌计算引擎、亚马逊 ec2、微软 azure、rackspace、hp helion 和 vmware vcloud)。这些结果表明,我们的方法可以跨多个云平台工作,支持不同的语言,并显著提高使用云平台进行符号数字计算的性能。这为激发在科学研究领域使用云计算进行符号数值计算的需要提供了一种途径。少

2018 年 2 月 9 日提交;最初宣布 2018 年 2 月。

评论:13 页, 23 个数字

日记本参考:ieee 云计算交易, 2017

126. 特别报告: 1802.04450[[pdf](#),其他] Cs. 直流

多伊 [10.1109/IPDPSW.2016.79](#)

cpu-gpu 平台频谱聚类的高性能实现

作者:yu jin, joseph f. jaja

摘要: 频谱聚类是最流行的图形聚类算法之一,它在许多科学和工程应用中实现了最佳的性能。但是,在常用的软件平台 (如 matlab 和 python) 中的现有实现不能很好地扩展到许多新兴的大数据应用程序。本文提出了一种在 cpu-gpu 异构平台上快速实现谱聚类算法的方法。我们的实现利用了多核 cpu 的计算能力以及 gpu 的大量多线程和 simd 功能。在将输入作为高维空间的数据点的情况下,我们提出了一种并行方案,以标准稀疏表示格式建立一个以稀疏表示格式表示的稀疏相似图。然后我们计算最小的 K 拉普拉斯矩阵的特征向量,利用 arpack 软件和 cusparse 库的反向通信接口,其中 K 通常非常大。此外,我们实现了一个非常快速的并行化 K-指 gpu 上的算法。与每个步骤最知名的 matlab 和 python 实现相比,我们的实现速度要快得多。此外,我们的算法还可扩展到大量集群的问题。少

2018 年 2 月 12 日提交;最初宣布 2018 年 2 月。

评论:2016 年 ieee 国际并行和分布式处理研讨会研讨会 (并行计算和优化 (pco) 研讨会)。 <https://github.com/yuj-umd/fastsc> 上提供了代码

127. 建议: 1802.03155[[pdf](#)] cs. ne

基于 web 的旅游销售人员问题的遗传算法实现

作者:aryo pinandito, novanto yudistira, fajar prprana

摘要: 世界是通过互联网连接的。随着大量互联网用户与网络的连接和云计算研究的普及,人工智能 (ai) 的需求越来越高。本研究采用遗传算法 (ga) 作为通过自然选择和遗传进化进行人工智能优化的方法。遗传算法有许多应用,如 web 挖掘、负载平衡、路由、调度或 web 服务选择等。因此,发现代码主要是服务器端和基于 web 的语言技术是否影响 ga 的性能是一项具有挑战性的任务。旅行销售人员问题 (tsp) 作为非多项式硬 (np-硬) 问题被认为是一个有问题的领域通过 ga 解决。虽然许多科学家更喜欢在 ga 实现中使用 python,但另一种流行的高级解释器编程语言,如 php (php

超文本预处理器) 和 ruby, 是有基准的。在这些编程语言中, 基于 ga 实现和运行时的代码、文件大小和性能行各不相同。在此基础上, 建议在 ga 实现中使用 ruby。少
2018 年 2 月 9 日提交;最初宣布 2018 年 2 月。

128. 建议: 18002914[[pdf](#)] Cs. CI

实践: 语音语料库研究的集成工具

作者:[乔治·赫里斯托杜利德斯](#)

摘要: 本文介绍了用于管理、注释、分析和可视化语音语料库的开源软件系统 praline。与言语语料库合作的研究人员通常面临多种工具和格式, 他们要以协作的方式处理越来越多的数据。praline 将现有的久经考验的口语语料库分析工具 (praat、sonic visualiser 和 r 统计包的桥梁) 集成到模块化系统中, 促进自动化和重用。用户可以使用一个集成的、用户友好的界面来访问多个工具。语料库元数据和批注可以存储在数据库中, 无论是本地还是远程, 用户可以定义元数据和批注结构。用户可以基于插件和脚本运行可自定义的分析步骤级联, 并使用结果更新数据库。可以查询语料库数据库, 以生成聚合数据集。praline 使用 **python** 或 **c++** 插件是可扩展的, 而 praat 和 r 脚本可能是针对语料库数据执行的。提供了一系列可视化、编辑器和插件。普拉林是自由软件, 根据 gpl 许可证发布。少

2018 年 2 月 8 日提交;最初宣布 2018 年 2 月。

日记本参考:第九届语言资源与评价国际会议 (lrec) 会议记录, 2014 年 5 月, 冰岛雷克雅未克

129. 建议: 1802.02474[[pdf](#),其他] cs. ms

用于反演问题中最佳检查点的高级巨蟒抽象

作者:[navjot kakreja](#), [jan hückelheim](#), [michael lange](#), [mathias louboutin](#), [andrea walthner](#), [simon w. funke](#), [gerard gorman](#)

摘要: 反演和 pde 约束优化问题往往依赖于解决伴随问题来计算目标函数的梯度。这需要存储大量的中间数据, 为使用给定的可用内存可能解决的最大问题设置限制。检查点是一种方法, 它可以减少通过重做部分计算而不是存储中间结果所需的内存量。旋转检查点算法是一个最佳计划, 用于计算较小的内存占用空间。将转换集成到现代巨蟒 hpc 代码中, 并将其与代码生成结合起来并不简单。我们提供了一个 api, 该 api 可从基于 dsl 的代码生成环境中访问检查点, 并提供一些以地震应用为重点的初始性能参数。少

2018 年 1 月 12 日提交;最初宣布 2018 年 2 月。

130. 建议: 1802.001502[[pdf](#),其他] Cs. Ce

考虑分布式发电的短路电流矢量计算--iec 60909 的开源实现

作者:[leon thurner](#), [martin braun](#)

摘要: 电网规划中的一项重要任务是确保检测网络中的故障, 并在任何网络元素中进行截断而不损坏。因此, 计算短路电流是电网规划应用的一项重要的网格分析功能。标准的 iec 60909 为短路计算提供了指导原则, 并定期应用于电网规划应用。本文提出了一种基于 iec 60909 的短路电流矢量化计算方法。分布式发电装置是根据标准的最新版本考虑的。该方法在基于巨蟒的开源工具 pandapower 中实现, 并根据商业软件和文献实例进行了验证。本文提出的实现是 iec 60909 标准的首次全面实施, 该标准可在开源许可下使用。它可用于高度自动化的电网研究中的故障电流评估, 并被证明可

以很好地扩展到大型网格中。它的实际适用性在一个真实的 mv 网格具有较高的 dg 渗透度的案例研究中得到了证明。少

2018 年 2 月 5 日提交;最初宣布 2018 年 2 月。

131. 建议: 18001115[pdf,其他] Cs。简历

end2you----通过端到端学习进行多式联运分析的帝国工具包

作者:panagiotis tzirakis, stefanos zafeiriou, bjorn w. schuller

摘要: 我们引入 end2you--伦敦帝国学院 (imperial college london) 工具包, 通过端到端深度学习进行多式联运分析。end2you 是在 **python** 中实现的开源工具包, 它基于 tensorflow。它提供了以端到端方式 (即使用原始输入) 对模型进行培训和评估的功能。它支持来自原始音频、视觉、生理或其他类型信息的输入或这些信息的组合, 输出可以是任意表示的, 用于分类或回归任务。据我们所知, 这是第一个提供通用端到端学习的工具包, 用于在单式或多式联运情况下的分析功能。为了测试我们的工具包, 我们使用了 avec 2016 挑战中使用的 remola 数据库。实验结果表明, end2 可以提供与最先进的方法相当的结果, 尽管不需要专家相似的特征表示, 但从数据 "端到端" 自学习这些。少

2018 年 2 月 4 日提交;最初宣布 2018 年 2 月。

132. 建议: 1802. 00565[pdf,其他] Cs。简历

多伊 10.528 半碘. 1189345

利用深层机器学习检测机场三维系统的安全区域和威胁

作者:abel ag rb guimaraes, ghassem tofighi

摘要: 在本研究中, 采用分割和分类的方法来识别机场安全的人体扫描仪图像中的威胁识别。美国国土安全部 (dhs) 有一个更高的假警报, 是用他们在机场使用今天的扫描仪的算法产生的。为了解决这个问题, 他们在 kaggle 网站开始了一场新的比赛, 要求科学界用新的算法改进他们的检测。本研究中使用的数据集来自

<https://www.kaggle.com/c/passenger-screening-algorithm-challenge/data> 的国土安全部根据国土安全部: "该数据集包含由新一代毫米波扫描仪获得的大量人体扫描。高定义-先进成像技术 (hd-ait) 系统。他们由穿着不同服装类型 (从轻装到厚重冬衣)、不同体重指数、不同性别、不同数量的威胁和不同类型的威胁的志愿者组成"。使用 **python** 作为主要语言, 对数据集图像进行预处理, 从 200 个物体中提取特征, 方法是: 强度、强度差异和要检测的局部邻域, 以生成分割区域并标记要检测的区域。作为培训和测试数据集中的真相使用。随后, 各区域将向美国有线电视新闻网深度学习分类器提供预测 17 个班级 (代表身体区域): 区域 1、zone1、. zone1 和总共 34 个区域中具有威胁的区域。分析结果表明, 该分类器的精度为 98.2863%, 损耗为 0.091319, 召回和精度平均为 100%。少

2018 年 2 月 9 日提交;v1 于 2018 年 2 月 2 日提交;最初宣布 2018 年 2 月。

评论:7 页, 17 位数字, 本文被明星会议、数据科学和大数据分析所接受。加拿大多伦多

133. 建议: 1802. 00036[pdf,其他] Cs。简历

在经典图像处理的防御中: cpu 的快速深度完成

作者:jason ku, ali harakeh, steven l. waslander

摘要: 随着数据驱动的深层神经网络作为通用函数逼近器实现的兴起, 对计算机视觉问题的研究大多已经脱离了手工制作的经典图像处理算法。本文表明, 采用设计良好的

算法,能够在深度完成任务上优于基于神经网络的方法。该算法简单快速,在cpu上运行,只依靠基本的图像处理操作来执行稀疏激光雷达深度数据的深度完成。我们在具有挑战性的 kitti 深度完成基准上评估我们的算法,在提交时,我们的方法在 kitti 测试服务器上的所有已发布方法中排名第一。此外,我们的算法是独立于数据的,不需要训练数据来执行手头的任务。用 **python** 编写的代码将在 https://github.com/kujason/ip_basic 上公开。少

2018 年 1 月 31 日提交;最初宣布 2018 年 2 月。

134. 建议: 1801. 10117[[pdf](#),其他] Cs。铭

私有 **py**: 实现可扩展和通用的优化机器学习

作者:[李毅](#),[关一涛](#), [于](#),[赵书耀](#),[徐伟](#)

摘要: 我们介绍了 **privy**, 这是一个实用的隐私保护协作计算框架,特别是针对机器学习任务进行了优化。**privy** 提供了一个易于使用和高度兼容的 **python** 编程前端,它支持高级阵列操作和不同的安全计算引擎,以允许安全假设和性能权衡。使用 **privy**, 程序员可以在 **python** 中方便、高效地编写现代机器学习算法。我们还设计和实现了一个新的高效计算引擎,人们可以使用竞争的云提供商有效地执行比实数更有效的一般算术。我们使用常见的机器学习模型(如逻辑回归和卷积神经网络)和现实世界数据集(包括 5000 比 100 万矩阵)演示 **privy** 的可用性和可扩展性。少

2018 年 5 月 24 日提交;v1 于 2018 年 1 月 30 日提交;最初宣布 2018 年 1 月。

135. 建议: 1801. 09847[[pdf](#),其他] Cs。简历

open3d: 用于 3d 数据处理的现代库

作者:[周千一](#),[杰西克公园](#),[弗拉德伦·科尔通](#)

摘要: **open3d** 是一个开源库,支持处理 3d 数据的软件快速开发。**open3d** 前端在 **c++** 和 **python** 中都公开了一组精心挑选的数据结构和算法。后端经过高度优化,并设置为并行化。**open3d** 是从一个干净的石板开发的,它有一组经过仔细考虑的小依赖关系。它可以在不同的平台上设置,并以最小的努力从源编译。代码是干净的、一致的样式,并通过清晰的代码审查机制进行维护。**open3d** 已在许多已发布的研究项目中使用,并被积极部署在云中。我们欢迎开源界的贡献。少

2018 年 1 月 29 日提交;最初宣布 2018 年 1 月。

评论:<http://www.open3d.org>

136. [xiv:1801. 09771](#)[[pdf](#)] cse

利用 **python** 开源基础结构降低电子表格模型风险

作者:[奥利弗·比弗斯](#)

摘要: 在 **eus** 学会演示文稿的聚合中,有两个格言是正确的:电子表格模型类似于软件,但电子表格开发人员并不是软件工程师。因此,由于缺乏传统的软件工程工具和协议,最终结果的误差率较高。本文为电子表格建模专业人员的基础工作奠定了基础,利用使用 **python** 编程语言构建的免费开源包开发可重现的审计工具,使利益相关者能够开发明确定义的数据包模型 "神谕",以测试和审计电子表格计算。少

2018 年 1 月 29 日提交;最初宣布 2018 年 1 月。

评论:14 页, 15 页彩色图

日记本参考:[esprig 2017 大会 "电子表格风险管理" 会议论文集](#), 伦敦帝国学院, pp91-105 国际标准书号: [978-1-905404-54-4](#)

137. [xiv:1801.06027\[pdf,其他\]](#) Cs. Db

多伊 [10.14788/23236187.236188](#)

rdbms 硬件加速高级分析

作者:[divya mahajan](#), [joon kyung kim](#), [jacob sacks](#), [adel ardlan](#), [arun kumar](#), [hadi esmaeilzadeh](#)

摘要: 机器学习、数据库和硬件设计的进步推动了数据革命。可编程加速器正在独立进入这些区域中的每一个区域。因此, 在这些不相交场的交叉点上, 存在着硬件加速的解决方案。本文阐述了实现高级分析 (dana) 数据库内加速统一解决方案的第一步。为数据库内分析部署专用硬件 (如 fpga) 目前需要手动设计硬件并手动路由数据。相反, dana 会自动将高级分析查询规范映射到 fpga 加速器。加速器实现是为用户定义的函数 (udf) 生成的, 该函数表示为使用 **python** 嵌入的域特定语言 (dsl) 的 sql 查询的一部分。为了实现高效的数据库集成, dana 加速器包含一种新的硬件结构 "stri 源", 可直接与数据库的缓冲池接口。步行者提取、清理和处理由执行分析算法的多线程 fpga 引擎所消耗的训练数据元组。我们将 dana 与 postgresql 集成, 为运行不同 ml 算法的一系列真实世界和合成数据集生成硬件加速器。结果表明, dana 增强的 postgresql 平均为实际数据集提供了 8.3 倍的端到端加速, 最大速度为 28.2 倍。此外, dana 增强的 postgresql 平均比在 green 主要问题上运行的多线程 apache madlib 快 4.0 倍。dana 提供了这些好处, 同时向数据科学家隐藏了硬件设计的复杂性, 并允许他们在 **python** 的 30-60 行中表达算法。少

2018 年 9 月 18 日提交;v1 于 2018 年 1 月 8 日提交;最初宣布 2018 年 1 月。

日记本参考:[divya mahajan](#)、[joon kyung kim](#)、[jacob sacks](#)、[adel ardlan](#)、[arun kumar](#) 和 [hadi esmaeilzadeh](#)。rdbms 硬件加速高级分析。pvladb, 11 (11): 1317-1331, 2018

138. [建议: 1801.01843\[pdf,其他\]](#) Cs. 直流

土卫六上雷达导频的设计与性能表征

作者:[andre merzky](#), [matteo turilli](#), [manuel maldonado](#), [shantenu jha](#)

摘要: 许多极端规模的科学应用程序都有由大量单独的高性能任务组成的工作负载。试点抽象通过作业占位符和后期绑定分离工作负载规范、资源管理和任务执行。因此, 试点抽象的适当实现可以支持在超级计算机上集体执行大量任务。我们介绍了 radical-pot (rp) 作为一个便携式、模块化和可扩展的基于 **python** 的试验系统。我们描述了 rp 的设计、体系结构和实现。我们描述了它的性能, 并展示了它在 titan (doe 领导级别设施) 上可扩展执行由数千个 mpi 任务组成的工作负载的能力。具体而言, 我们研究 rp 的弱 (强) 缩放属性, 最高可达 131k (65k) 内核和 4096 (16384) 32 核心任务。radical-fot 可以单独使用, 也可以作为运行时系统与其他工具集成。少

2018 年 1 月 5 日提交;最初宣布 2018 年 1 月。

139. [建议: 1801.01025\[pdf,其他\]](#) cse

不同编程语言中的 bug 处理工作量如何?

作者:[张杰](#), [李峰](#), [丹浩](#), [王蒙](#), [张路](#)

摘要: 处理错误是软件开发的重要组成部分。编程语言对这一任务的影响长期以来一直是一个争论的话题。例如, 有些人认为 **python** 中的 bug 易于处理, 因为它的代码易于阅读和理解, 而另一些人认为 **python** 中没有静态类型会导致更高的 bug 处理工作。本文介绍了首次大规模研究, 以调查不同 (类别) 编程语言的生态系统是否需要不同的错误处理工作。重点是相关分析, 而不是因果分析。从 github 下载了 100 个 10 种

语言中最受欢迎的项目 (总数为 70 816, 938 sloc 和 3, 096, 009 提交), 实验结果显示了各种有趣的发现。首先, 不同的语言需要不同的错误处理工作。例如, java 和 c# 往往需要更少的时间, 但更多的线/文件修改, **python** 和 php 往往需要更少的时间和更少的线/文件修改, 而 ruby 和 javascript 往往需要更多的时间以及更多的 lin/fil 修改。其次, 弱动态语言往往比强/静态语言需要更多的时间, 而静态语言往往需要更多的绝对 lin/文件修改。玩具预测模型还提供了证据, 证明包含编程语言可以提高预测项目的 bug 处理工作的有效性。少

2018 年 1 月 3 日提交;最初宣布 2018 年 1 月。

140. [xiv:1801.00690\[pdf,其他\]](#) Cs. 艾

深度思维控制套件

作者:[yuval tassa](#), [yotam doron](#), [alastair muldal](#), [tomérez](#), [yazhe li](#), [diego de las casas](#), [david budden](#), [abbas abdolmaleki](#), [josh merel](#), [andrew lefrancq](#), [timothy lillicrap](#), [martin riedmiller](#)

摘要: 深度思维控制套件是一套具有标准化结构和可解释奖励的连续控制任务, 旨在作为强化学习代理的性能基准。这些任务是用 **python** 编写的, 由 mujoco 物理引擎提供动力, 使其易于使用和修改。我们包括几种学习算法的基准。控制套件可在 https://www.github.com/deepmind/dm_control 公开使用。所有任务的视频摘要可在 <http://youtu.be/rAai4QzcYbs>。少

2018 年 1 月 2 日提交;最初宣布 2018 年 1 月。

评论:24 页, 7 个数字, 2 个表

141. [建议: 1801.00329\[pdf,其他\]](#) Cs. Lg

zoopt: 用于无导数优化的工具箱

作者:[刘玉仁](#), [胡一琪](#), [洪谦](#), [杨宇](#), [赵谦](#)

摘要: 最近在无导数优化方面取得的进展使得可以有效地逼近复杂函数的全局最优解, 例如具有许多局部优化函数、非微函数和非连续函数的函数。本文介绍了 <https://github.com/eyounx/ZOOpt> () 工具箱, 该工具箱提供高效的无派生求解器, 易于使用。zoopt 提供了用于单线程优化的 **python** 包, 以及在 **python** 的 julia 语言帮助下提供的轻量级分布式版本。zoopt 工具箱特别关注机器学习中的优化问题, 解决高维、嘈杂和大规模的问题。在现实世界的机器学习任务中, 工具箱正朝着随时可用的工具进行维护。少

2018 年 2 月 6 日提交;v1 于 2017 年 12 月 31 日提交;最初宣布 2018 年 1 月。

142. [第 1712.09359\[pdf,其他\]](#) cs. cy

toki pona 最小和构造语言的基本概念和工具: 语言和主要问题的描述;对词汇的分析;文本合成和语法突出;wordnet 同步器

作者:[雷纳托·法布里](#)

摘要: 最小构造语言 (conlang) 对于实验很有用, 对制作工具也很舒适。toki pona (tp) 在词汇 (只有 14 个字母和 124 个引理) 和 (约) 10 个语法规则中都是最小的。这种语言是有用的, 因为它是一个使用过的, 有点建立的最小的语言, 至少有几百个流利的扬声器。本文公开了 tp 的当前概念和资源, 并提供了 **python** (和 vim) 脚本例程, 用于分析语言、文本合成、语法突出显示方案和实现初步的 tp wordnet。重点是分析基本词汇, 因为语料库分析被发现。综合以句子模板为基础, 通过跟踪使用过的单词来与上

下文相关, 并通过使用固定数量的音素 (例如诗歌) 和句子、单词和字母的数量 (例如段落) 来呈现较大的文本。语法突出显示反映了官方词典中给出的形态合成类, 在成熟的 vim 文本编辑器中描述并实现了不同的解决方案。试探性的 tp wordnet 以三种形式提供了组合和单词引理之间的关系。总之, 本文包含了潜在的新颖的概念化, 以及在分析、合成和语法方面突出 tp 语言的工具和结果。少

2018 年 7 月 3 日提交;v1 于 2017 年 12 月 26 日提交;最初宣布 2017 年 12 月。

评论:此存储库中的 python 和 vim 脚本: <https://github.com/ttm/prv/>

143. 第 1712.08618[[pdf](#)] Cs. Db

特征提取与特征选择: 利用阿帕奇火花降低数据复杂性

作者:[Dimitrios sisiridis](#), [olivier markowitch](#)

摘要: 为了在利用机器学习的同时检测网络安全威胁和攻击, 特征提取和特征选择是输入日志预处理的首要任务。当涉及到分析来自不同来源的异构数据时, 发现这些任务非常耗时, 难以有效地管理。本文提出了一种处理来自不同网络传感器的异构数据的特征提取和特征选择的方法。该方法在 apache spark 中实现, 使用其巨蟒 api, 名为 pyspark。少

2017 年 12 月 11 日提交;最初宣布 2017 年 12 月。

日记本参考:国际网络安全及其应用杂志 (ijnsa), 第 9 卷, 第 6 期, 2017 年 11 月

144. 第 (xiv:1712. 08364)[[pdf](#),其他] Cs. Cg

具有深度学习数值的微分几何和随机动力学

作者:[line kühnelt](#),[亚历克西斯·阿诺登](#), [stefan sommer](#)

摘要: 在本文中, 我们演示了如何确定和随机动力学流形, 以及微分几何结构可以实现简洁和有效地使用现代计算框架, 将符号表达式与有效的数值计算。特别是, 我们使用了巨蟒库 theano 的符号表达和自动分化特征, 它最初是为深度学习中的高性能计算而开发的。我们展示了如何利用天野的任何阶导数的自动计算, 用微分几何和李群理论、连接、度量、曲率、左/右不变性、大地测量和并行传输等各个方面来制定。我们还将展示如何使用几行代码来制定和优化符号随机积分器和非线性统计中的概念。然后, 我们将在低维经典流形上给出明确的示例, 以进行可视化, 并演示此方法如何允许对高维问题进行简洁的实现和有效的缩放。少

2017 年 12 月 22 日提交;最初宣布 2017 年 12 月。

msc 类: 53a35;53c17;53c44;70h05;22e30 类: g.3;g.4;G.1。4

145. 第 1712.07449[[pdf](#)] Cs. Lg

基于 lstm 神经网络的新型药物类化学物质在硅素生成中的应用

作者:[peter ertl](#), [richard lewis](#), [eric martin](#) , [valery polyakov](#)

摘要: 在支持药物发现过程中, 探索新的化学空间是化学信息学最重要的任务之一。设计合理和训练良好的深层神经网络可以为新的暴力方法或各种其他机器学习技术提供一种可行的替代方法, 从而产生新型的类似药物的分子。在本文中, 我们提出了一种利用长时间短期记忆 (lstm) 神经网络生成分子的方法, 并对结果进行了分析, 包括虚拟筛选测试。利用该网络在 2 小时内产生了 100 万个类似药物的分子。这些分子新颖多样 (含有许多新的化学型), 具有良好的物理化学特性, 具有良好的合成可及性, 尽管这些品质不是特定的制约因素。虽然新颖, 但它们的结构特征和功能基团仍在 chembl 的生物活性分子所定义的空间内。使用轮廓 qsar 方法进行的虚拟筛选证

实, 这些新分子显示生物活性的潜力可与从中衍生出来的 chembl 集合相当。本研究中使用的 **python** 中的分子发生器可应要求提供。少

2018 年 1 月 8 日提交;v1 于 2017 年 12 月 20 日提交;最初宣布 2017 年 12 月。

评论:在这个版本固定了一些参考编号

146. 特别报告: 1712.07438[[pdf](#),[其他](#)] **cs. ms**

相机转换: 一种用于透视相机校正的科学 python 包

作者:[richard gerum](#), [sebastian richter](#),[亚历山大 winterl](#), [ben fabry](#), [daniel zterterbart](#)

摘要: 科学应用通常需要精确重建物体位置和与数字图像的距离。因此, 需要对图像进行透视扭曲的校正。我们提出了一个巨蟒包, 执行透视图像校正, 通过这种程序, 如果有其他信息 (如 gps), 可以从图像中自动获得相机的高度、倾斜角度和航向。提供坐标或对象大小。我们提供了企鹅群体的图像的例子, 这些图像是用固定的摄像机和直升机记录的。少

2017 年 12 月 20 日提交;最初宣布 2017 年 12 月。

评论:8 页, 5 个数字

147. 第: 1712.06933[[pdf](#), [ps](#),[其他](#)] **Cs. Hc**

vim 文本编辑器的人类学描述: 使用 10 年后的功能和调整

作者:[雷纳托·法布里](#)

摘要: vim 文本编辑器功能非常丰富, 因此非常复杂。本文是对 vim 的描述, 并对其使用和设计进行了一组注意事项。它来自于十多年来使用 vim 编写和编辑各种类型的文档 (例如 **python**、c++、javascript、chuck 程序) 的经验;LaTeX、标记、html、rdf、make 和其他标记文件;% ttm 二进制文件。在 vim 用户和开发人员社区中, 说掌握 (或开始掌握) 这个文本编辑器大约需要十年的时间是很常见的, 我发现其他有经验的用户对 vim 有不同的看法, 他们使用的是一组不同的功能。因此, 本文档揭示了我的理解, 以便面对我对其他 vim 用户的用法。另一个目标是提供一份参考文件, 使新用户能够通过阅读它及其可能产生的讨论来掌握一个健全的概览。此外, 它应该是有用的任何程度的经验的用户, 包括我, 作为命令, 命名空间和调整的汇编。根据反馈, 并成熟我的 vim 用法, 本文档可能会得到增强和扩展。少

2017 年 12 月 18 日提交;最初宣布 2017 年 12 月。

评论:脚本和其他文件在此存储库中: <https://github.com/ttm/vim>

148. 第: 1712.06919[[pdf](#)] **Cs. 红外**

维基数据中的破坏性检测的生产导向方法--2017 年 wsdm 杯上的 buffaloberry vandarverds 检测器

作者:[rafael crescenzi](#), [marcelo fernandez](#), [fedico a. garcia calabria](#), [pablo alani](#), [diego tauziet](#), [adriana baravalle](#), [andr s sebasti n d ' ambrosio](#)

摘要: 维基数据是维基媒体基金会有一个自由和开放的知识库, 它不仅为组织的其他项目提供结构化数据的中央存储, 而且还为包括搜索引擎在内的越来越多的信息系统提供中央存储。与维基百科一样, 维基数据的内容也可以由任何人创建和编辑;这是其力量的主要来源, 但也允许恶意用户破坏它, 有可能通过所有依赖它作为结构化事实来源的系统传播错误信息。我们在 2017 年 wsdm 杯上的任务是拿出一个快速可靠的预测系统, 缩小对人类修订的可疑编辑。阐述了 heindorf 等人以前的作品, 我们能够

超越所有其他参赛者,同时纳入了新的有趣的功能,统一了仅用于 **python** 的编程语言,并将功能提取器重构为更简单、更紧凑的代码库。少

2017 年 12 月 19 日提交;最初宣布 2017 年 12 月。

评论:2017 年 wsdm 杯上的 Vandalism 探测器,见 [arxiv:1712.0556](https://arxiv.org/abs/1712.0556)

类:h。3

149. 第: 1712.06658[[pdf](#),[其他](#)] Cs。Lg

多伊 [10.1109/SKIMA.2017.8294128](https://arxiv.org/abs/10.1109/SKIMA.2017.8294128)

megooost: 将估计值与不平衡数据分类的提升混合在一起

作者:[farshid rayhan](#), [sajid ahmed](#), [asif mahbub](#), [md.rafsan jani](#), [swakkhar shatabda](#), [dewan md.farid](#), [chowdhury Mofizur rahman](#)

摘要: 由于大多数现实生活数据集不平衡, 班级不平衡问题一直是机器学习和数据挖掘领域的一个具有挑战性的研究问题。几种现有的机器学习算法试图通过正确识别多数类样本而忽略少数类来最大限度地提高精度分类。然而, 少数群体实例的概念通常比多数类代表更高的兴趣。近年来, 文献中采用了几种成本敏感方法、集成模型和采样技术对不平衡数据集进行分类。本文提出了一种新的不平衡数据集提升算法 meboost。meboost 将两个不同的弱学习者与提升混合在一起, 以提高不平衡数据集的性能。megooost 是现有技术的替代方案, 如 sometoost、rusboost、adaboost 等。在 12 个基准不平衡数据集上评估了 meboost 的性能, 这些数据集具有最先进的集成方法, 如 smotboost、ruboost、easy 集成、eusboost、dataost。实验结果表明, 与其他方法相比, meboost 是一种有效且有希望的处理不平衡数据集的算法。该代码的巨蟒版本可在此处找到: <https://github.com/farshidrayhanui/>少

2018 年 1 月 13 日提交;v1 于 2017 年 12 月 18 日提交;最初宣布 2017 年 12 月。

评论:斯基马-2017

150. 第: 1712.05878[[pdf](#),[其他](#)] Cs。直流

基于 **mpi** 的 **python** 框架, 用于 **keras** 的分布式培训

作者:[dustin anderson](#), [jean-roch vlimant](#), [maria spiropulu](#)

摘要: 我们提出了一个轻量级的 **python** 框架, 用于在多个 **gpu** 或 **cpu** 上对神经网络进行分布式训练。该框架是建立在流行的 **keras** 机器学习库。消息传递接口 (**mpi**) 协议用于协调训练过程, 该系统非常适合在超级计算站点提交作业。我们详细介绍了软件的功能, 描述了它的使用情况, 并演示了它在从高能物理研究中提取的基准问题上的不同尺寸系统上的性能。少

2017 年 12 月 15 日提交;最初宣布 2017 年 12 月。

评论:4 页, 4 个数字, 1 个表, ds @ hep, sc17

151. 第 1712.05690[[pdf](#),[其他](#)] Cs。CI

sockeye: 神经机器翻译工具包

作者:[felix hieber](#), [tobias domhan](#), [michael denkowski](#), [david vilar](#), [artem sokolov](#), [ann 克利夫 ton](#), [matt poster](#)

摘要: 我们描述了 sockeye (版本 1.12), 这是一个用于神经机器翻译 (nmt) 的开源序列到序列工具包。sockeye 是一个生产就绪的培训和应用模型框架, 也是研究人员的实验平台。该工具包以 **python** 编写, 构建在 **mxnet** 上, 为三种最突出的编码器解码器架构提供了可扩展的培训和推理: 注意力递归神经网络、自注意变压器和完全卷积网

络。sockeye 还支持广泛的优化器、规范化和正则化技术, 以及当前 nmt 文献中的推理改进。用户可以轻松地运行标准培训配方, 探索不同的模型设置, 并加入新的想法。在本文中, 我们重点介绍了 sockeye 的功能, 并将其与 2017 年机器翻译会议 (wmt) 的两种语言弧线上的其他 nmt 工具包进行了比较: 英语-德语和拉脱维亚-英语。我们报告了所有三个架构的竞争性 BLEU 分数, 包括 sockeye 变压器实现的整体最佳得分。为了便于进一步比较, 我们发布了实验中使用的所有系统输出和培训脚本。

sockeye 工具包是在 apache2.0 许可证下发布的自由软件。少

2018 年 6 月 1 日提交;v1 于 2017 年 12 月 15 日提交;最初宣布 2017 年 12 月。

152. 第 1712.05181[[pdf](#),[其他](#)] Cs. CI

rasa: 开源语言理解和对话管理

作者:[tom bocklisch](#), [joey faulkner](#), [nick palowski](#), [alan nichol](#)

摘要: 我们介绍了一对工具, rasa nlu 和 rasa core, 这是用于构建会话软件的开源巨蟒库。其目的是使非专业软件开发人员能够使用基于机器学习的对话管理和语言理解。在设计理念方面, 我们的目标是易于使用, 并从最小 (或不) 初始训练数据进行引导。这两个软件包都有广泛的文档记录, 并附带了一套全面的测试。该代码可

<https://github.com/RasaHQ/>

2017 年 12 月 15 日提交;v1 于 2017 年 12 月 14 日提交;最初宣布 2017 年 12 月。

评论:在 NIPS 关于对话 ai、代码的研讨会上发表

153. 第 1712.04786[[pdf](#),[其他](#)] Cs. 哦

自动计算器: 一个简单的, 基于 pyth 的, 用于数值计算的自动化框架

作者:[prabhu ramachandran](#)

摘要: 我们提出了一个易于使用的、基于 python 的框架, 该框架允许研究人员自动进行计算模拟。特别是, 该框架便于组装几个长时间运行的计算, 并从这些计算产生的数据生成各种图形。该框架使复制使用单个命令为发布创建的每个图形成为可能。它还允许一个人在计算机网络中分发计算。该框架已被用于在数值计算中撰写研究论文。本文讨论了框架的设计, 以及使用该框架的好处。所提出的想法是一般性的, 应该可以帮助研究人员组织他们的计算, 以获得更好的重现性。少

2018 年 2 月 4 日提交;v1 于 2017 年 12 月 11 日提交;最初宣布 2017 年 12 月。

msc 类: 68u20

154. 第 1712.04546[[pdf](#), [ps](#),[其他](#)] Cs. Ce

用整数混沌博弈表示对 dna 序列进行编码

作者:[尹昌川](#)

摘要: dna 序列是编码遗传信息的基础。遗传信息不仅可以通过符号序列来理解, 也可以通过序列内隐藏的信号来理解。符号序列需要转化为数值序列, 以便通过信号处理技术来揭示隐藏信号。所有当前的转换方法都将 dna 序列编码为相同长度的数值。这些表示在基因组信号压缩、加密和隐写的应用中存在局限性。我们提出了一个整数混沌博弈表示 (icgr) 的 dna 序列和一个无损编码方法 dna 序列的 icgr。在 icgr 方法中, dna 序列由核苷酸的迭代函数及其在序列中的位置表示。然后, dna 序列可以唯一地编码和恢复使用三个整数从 icgr。一个整数是序列长度, 另外两个整数表示序列中核苷酸的累积分布。整数编码方案可以将每个核苷酸的 dna 序列压缩 2 位。dna 序列的

整数表示为序列压缩、加密和隐写提供了一个前瞻性的工具。本研究中的 **python** 项目可在 <https://github.com/cyinbox/iCGR> 免费提供给公众

2017 年 12 月 19 日提交;v1 于 2017 年 12 月 12 日提交;最初宣布 2017 年 12 月。

msc 类: 92d20

155. 第 1712.0882[[pdf](#),其他] Cs. Sd

理论: 利用深部递归神经网络从音频中不受监督地学习表现

作者:[michael freitag](#), [shahin Amiriparian](#), [sergey pugachevskiy](#), [nicolas cummins](#), [björn schuller](#)

摘要: audeep 是一个 **python** 工具包, 用于从声学数据中学习深度无监督表示。它是基于一个循环序列的自动编码器方法, 可以学习时间序列数据的表示, 考虑到他们的时间动态。除了为用户和开发人员提供 **python** api 外, 我们还提供了广泛的命令行界面, 这两者都有全面的文档记录, 并可在 <https://github.com/auDeep/auDeep> 公开获得。实验结果表明, 奥源特征与最先进的音频分类具有竞争力。少

2017 年 12 月 22 日提交;v1 于 2017 年 12 月 12 日提交;最初宣布 2017 年 12 月。

156. 第 1712.01662[[pdf](#)] Cs. 简历

多伊 [10.1371/journal.pone.0199239](https://doi.org/10.1371/journal.pone.0199239)

在考虑色觉不足的情况下优化颜色, 以便能够准确地解释科学数据

作者:[jamie r. nuñez](#), [christopher r.anderton](#), [ryan s. renslow](#)

摘要: 颜色视力缺乏 (cvd) 影响 4% 以上的人群, 并导致不同的视觉感知颜色。虽然这已经知道了几十年, 在视觉光谱中具有多种颜色的颜色经常被用来表示数据, 从而导致有这种缺陷的人有可能被误解或难以解释。在创建此处介绍的模块之前, 没有使用现代颜色外观模型对 cvd 进行数学优化的颜色映射。虽然有一些尝试, 使美观或主观上可以容忍的颜色为那些与 cvd, 我们的目标是使优化的颜色为最准确的感知科学数据的尽可能多的观众。我们开发了一个 **python** 模块 cmaputil, 用于创建 cvd 优化的颜色映射, 它导入颜色, 并将其修改为 cvd 安全的色彩空间中的感知均匀, 同时线性化并最大限度地提高亮度范围。该模块提供给科学界, 使其他人能够轻松地创建自己的 cvd 初优化颜色。在这里, 我们展示了一个使用此模块创建的 cvd 优化色图示例, 该图经过优化, 可供没有 cvd 的用户以及红绿色色盲患者查看。这种颜色图 cividis 使这两组的可视数据解释几乎相同, 在色调和亮度上都是感知均匀的, 并且亮度呈线性增加。少

2018 年 8 月 1 日提交;v1 于 2017 年 11 月 29 日提交;最初宣布 2017 年 12 月。

日记本参考:[j. r. nuñez](#)、[c. r. anderton](#) 和 [r. s. renslow](#), "考虑到色觉缺陷, 优化颜色, 以便能够准确解释科学数据," plos one, 2018 年。13 (7): e00199239 页

157. 第: 1712.00825[[pdf](#),其他] Cs. Cg

从人脸法线和面积重建凸多面体的算法

作者:[giuseppe ellaroli](#)

摘要: 由于明科夫斯基的研究, 凸多面体的研究有一个众所周知的结果, 那就是凸多面体是由其面的方向和区域唯一地确定的 (直至翻译)。该定理保证了与给定面法线和面积相关的多面体的存在, 但没有提供明确找到它的建设性方法。本文根据 lasserre 计算中凸多面体体积的方法, 提出了一种从其面法线和面积重建三维凸多面体的算法。Rn.该算法的 **python** 实现可在 <https://github.com/gsellaroli/polyhedrec>。少

2017 年 12 月 3 日提交;最初宣布 2017 年 12 月。

评论:12 页, 3 个数字。该算法的 python 实现可在

<https://github.com/gsellaroli/polyhedrec>

158. 第: 1711. 10713[pdf,其他] cs.PL

多伊 [10.1016/j.scico.2017.11.005](#)

字符串 api 的批判性分析: pharo 的案例

作者:[damien pollet](#) , [stphane ducasse](#)

摘要: 除 c 之外, 大多数编程语言都为字符串提供了本机抽象, 但字符串 api 在大小、表现力和各种语言的主观便利性方面差别很大。在 pharo 中, 虽然乍一看字符串类的 api 似乎很丰富, 但实际上往往会感到繁琐;为了提高其可用性, 我们面临着评估其设计的挑战。然而, 我们几乎没有发现任何关于设计力及其如何构造设计空间的准则, 也没有对预期的字符串操作及其不同的变化进行全面分析。在本文中, 我们首先分析 pharo 4 字符串库, 然后将其与 haskell、java、python、ruby 和 rust 对应的库进行对比。我们收集描述字符串 api 的标准, 并反映功能和设计紧张关系。这种分析应有助于语言设计师理解字符串的设计空间, 并将作为今后重新设计 pharo 字符串库的基础。少

2017 年 11 月 29 日提交;最初宣布 2017 年 11 月。

日记本参考:计算机编程科学, elsevier, 2017, \& \ #x3008; 10.1016/j.scico.2017.11.005\& \ #x3009

159. 第: 1711. 10464[pdf] Cs。简历

openmv: python 动力、可扩展的机器视觉相机

作者:[ibrahim abdelkader](#) , [yasser el-sonbaty](#) , [mohamed el-habro](#) 行

摘要: 半导体制造工艺的进步和大规模集成使要求苛刻的应用程序远离集中处理, 更接近网络边缘 (即边缘计算)。使用低功耗嵌入式智能相机执行复杂的网络内图像处理已成为可能, 从而实现了众多新的协作图像处理应用。本文介绍了一种新型的低功耗智能相机 opensmv, 它自然地适用于无线传感器网络和机器视觉应用。此平台的独特之处在于运行嵌入式 python3 解释器, 允许在 python 中编写外围设备和机器视觉库脚本。此外, 它的硬件可以通过通过具有新功能 (如热成像和网络模块) 增强平台的模块进行扩展。少

2017 年 11 月 1 日提交;最初宣布 2017 年 11 月。

日记本参考:2017 年计算机图形学、可视化、计算机视觉和图像处理国际会议

160. 第: 1711. 0886[pdf,其他] cs. ms

多伊 [10.1063/1.5019320](#)

将微分方程与集 tcode、jdtcdde 和 jtcsde 高效、轻松地集成

作者:[gerrit ansmann](#)

摘要: 我们提出了一个 python 模块的家族, 用于普通微分方程、延迟方程或随机微分方程的数值积分。其主要特点是, 用户象征性地输入导数, 并对其进行实时编译, 使用户能够有效地集成来自更高级解释语言的微分方程。所提出的模块特别适用于大型微分方程组, 如用于描述复杂网络上的动力学。通过选定的输入方法, 所提出的模块还可以完全自动化估计普通方程和时滞微分方程的正则和横向李雅普诺夫指数的过程。我

们从概念上讨论了模块的设计, 分析了它们的性能, 并通过应用来演示它们的能力, 以及时解决问题。少

2018 年 3 月 22 日提交;v1 于 2017 年 11 月 25 日提交;最初宣布 2017 年 11 月。

日记本参考:混沌 28, 043116 (2018)

161. 第: 1711.07731[[pdf](#), [ps](#),其他] q-fin。公关

多伊 [10.1016/j.apenergy.2018.06.003](#)

统一采购价格和可削减区块订单的电力市场清算新方法

作者:[iacopo savelli](#), [bertrand cornélusse](#), [antonio giannitrapani](#), [simone paoletti](#), [antonio vicino](#)

摘要: 欧洲市场清算问题的特点是一组异构的订单和规则, 迫使启发式和迭代求解方法的实施。特别是, 可削减的整批订单和统一的收购价格 (upp) 造成了严重的困难。块是跨越多个小时的订单, 可以完全接受, 也可以完全拒绝。upp 规定, 所有消费者在所有地区都要支付共同的价格, 即 upp, 而生产者则获得地区价格, 而各地区的价格可能因地区而异。在 upp 和块状订单存在的情况下, 市场清算问题是欧洲背景下的一个重大未结问题。upp 方案导致一个涉及原始变量和双变量的非线性优化问题, 而块顺序则将多时间约束和二元变量引入该问题。因此, 在两个块和 upp 存在的情况下的市场清算问题可以看作是一个非线性整数规划问题, 涉及具有互补和多时间约束的原始变量和双变量。本文的目的是提出一种非迭代和无启发式的方法, 在可裁剪块订单和 upp 都存在的情况下解决市场清算问题。该解决方案是精确的, 没有接近当前市场数据的分辨率水平。通过采用等效 upp 公式, 该方法产生了一个混合整数线性规划, 它是从一个非线性整数双层规划问题开始建立起来的。利用实际市场数据报告了数值结果, 表明了该方法的有效性。该模型已在 **python** 中实现, 代码可在公共存储库中免费使用。少

2018 年 6 月 6 日提交;v1 于 2017 年 11 月 21 日提交;最初宣布 2017 年 11 月。

评论:15 页, 7 个数字

162. 建议: 1711.05412[[pdf](#), [ps](#),其他] 反渗透委员会

ikbt: 用行为树求解闭式逆运动学

作者:[张电木](#), [blake hanaford](#)

摘要: 串行机器人手臂具有复杂的运动学方程, 必须解决, 以编写有效的手臂规划和控制软件 (反向运动学问题)。现有的逆运动学软件包往往依赖于有显著缺陷的数值方法。在这里, 我们报告了一个新的符号逆运动学求解器, 它克服了数值方法的局限性, 以及以前符号软件包的缺点。我们集成了行为树, 这是一个以前用于控制智能机器人行为、组织方程求解过程的执行规划框架, 以及每种求解技术的模块化体系结构。该系统成功地解决了问题, 生成了一个 latex 报告, 并为 19 个 4-6 自由度的示例机器人中的 18 个生成了 **python** 代码模板。该系统易于扩展、维护和多平台, 几乎没有依赖关系。完整的软件包可与 github 上的修改 bsd 许可证。少

2017 年 12 月 7 日提交;v1 于 2017 年 11 月 15 日提交;最初宣布 2017 年 11 月。

评论:14 页, 6 个数字

163. 第: 1711.04913[[pdf](#)] Cs。Lg

pylememgs: 生物信息学应用的大边距多实例分类和排名

作者:am 阵地 asif, wajid arshad abbasi, farzeen munir, asa ben-hur, fayyaz ul amir afsar minhas

摘要: 动机: 计算生物学中基于机器学习的方法发展的一个主要挑战是, 由于实验注释蛋白质和 dna 序列的特性所需的时间和资源, 数据可能无法准确地标记。标准的监督学习算法假定训练数据的精确实例级标记。多实例学习是处理此类标记歧义的范例。然而, 在多实例学习中广泛使用的大边缘分类方法本质上是启发式的, 具有较高的计算要求。本文提出了用于多实例分类和排序的随机子梯度优化大边距算法, 并将其提供给一个称为 pylemmings 的软件套件。结果: 我们已经测试了 pylegmgs 的一些生物信息学问题以及基准数据集。pylemmings 已经成功地识别了功能重要的蛋白质片段: 钙调素结合蛋白中的结合位点、prion 形成区域和淀粉样蛋白核。pylegmgs 在所有这些任务中实现了最先进的性能, 展示了多实例学习的价值。此外, 与启发式解决方案相比, 我们的方法在运行时间方面提高了 100 多倍, 与基准数据集相比, 该解决方案的精度提高了 100 倍。可用性和实现: 巨蟒包可在

<http://faculty.pieas.edu.pk/fayyaz/software.html#pylemmings> 下载。少

2017 年 11 月 13 日提交;最初宣布 2017 年 11 月。

164. 第: 1711.03954[[pdf](#),[其他](#)] Cs. 简历

涡流网络: 一种用于海洋涡流像素-智慧分类的深层神经网络

作者:redouane lguensat, miao sun , ronan fablet, evan mason, pierre tandeo, ge chen

摘要: 这项工作介绍了 eddynet, 这是一种基于深度学习的架构, 用于从哥白尼海洋和环境监测处 (cmems) 提供的海面高度 (ssh) 地图中自动检测和分类涡流。eddynet 是一种类似于 u-net 的网络, 由卷积编码器解码器和像素分类层组成。输出是具有相同大小的输入的地图, 其中像素具有以下标签 \{ '0': 非涡流, '1': 反气旋涡流, '2': 旋风涡流。我们研究了 selu 激活函数的使用, 而不是经典的 relu + bn, 我们使用重叠的损失函数, 而不是交叉熵损失。keras **python** 代码、培训数据集和 eddynet 权重文件都是开源的, 可在 <https://github.com/redouanelg/EddyNet> 上免费获得。少

2017 年 11 月 10 日提交;最初宣布 2017 年 11 月。

165. 第: 1711. 03016[[pdf](#),[其他](#)] cs.PL

dlvm: 用于深度学习系统的现代编译器基础结构

作者:richard wei , lane schwartz , vikram adve

摘要: 深度学习软件需要可靠性和性能。然而, 许多现有的深度学习框架都是软件库, 在 **python** 中充当不安全的 dsl 和计算图解释器。我们提出了 dlvm, 一个编译器基础结构的设计和实现与线性代数中间表示, 算法区分与联合代码生成, 域特定的优化和代码生成器目标 gpu 通过 llvm。dlvm 是受 llvm 启发的现代编译器基础架构, 比现有的深度学习编译器框架更模块化、更通用, 并支持具有高表现力的张量 dsl。借助我们在 swift 中嵌入的典型阶段 dsl, 我们认为 dlvm 系统支持一种模块化、安全和高性能的深度学习框架。少

2018 年 2 月 2 日提交;v1 于 2017 年 11 月 8 日提交;最初宣布 2017 年 11 月。

166. 第: 1711.02712[[pdf](#),[其他](#)] cs. ms

切线: 使用 **python** 中的源代码转换实现自动区分

作者:bart van merriënboer, 亚历山大 b. wilschko, dan moldovan

摘要: 自动微分 (ad) 是机器学习编程系统必不可少的基础。切线是使用 **python** 中的源代码转换 (sct) 执行 ad 的新库。它将编写在 **python** 和 **numpy** 的句法子集中的数字函数作为输入, 并生成计算导数的新 **python** 函数。这种自动区分的方法不同于机器学习中流行的现有软件包, 如 **tensorflow** 和 **autograd**。优点是 **tangent** 在 **python** 中生成渐变代码, 用户可以读取, 易于理解和调试, 并且没有运行时开销。切线还引入了抽象, 以便轻松地将逻辑注入生成的梯度代码中, 从而进一步提高可用性。少

2017 年 11 月 7 日提交;最初宣布 2017 年 11 月。

167. 第: 1711.01229[[pdf](#),[其他](#)] Cs。直流

面向 hep 中的实时数据查询系统

作者:[jim pivarski](#), [david lange](#), [thanat jat 管](#) [hattharachat](#)

摘要: 探索性数据分析工具必须快速响应用户的问题, 以便对一个问题 (例如可视化直方图或适合) 的答案可以影响下一个问题。在工业中使用的某些基于 **sql** 的查询系统中, 即使是非常大的 (pb) 数据集也可以汇总在人工时间刻度 (秒) 上, 采用的技术包括列数据表示、缓存、索引和代码生成/jit 编译。本文介绍了实现这样一个高能物理系统 (hep) 的进展, 重点介绍了优化数据访问和计算 "查询大小" 有效载荷 (如单个直方图或直方图组) 的中间问题。比大型重建或数据浏览工作。这些技术包括将 **root t** 分支直接提取到 **numpy** 数组中, 以及编译 **python** 分析函数 (而不是 **sql**) 以非常快速地执行。我们还将讨论缓存和主动将作业传递给在缓存中预装必要输入数据的工作节点的问题。较大解决方案的所有这些部分都可作为独立的 **github** 存储库使用, 并可用于当前的分析。少

2017 年 11 月 8 日提交;v1 于 2017 年 11 月 3 日提交;最初宣布 2017 年 11 月。

评论:6 页, 2 个数字, 2017 年 **acat** 诉讼程序

168. 第: 1711.00837[[pdf](#),[其他](#)] Cs。Lg

基于 k-均值和 **smote** 的不平衡学习问题的过山

作者:[felixlast](#), [ge 科洛](#) [osdouzas](#), [fernando bacao](#)

文摘: 从类不平衡的数据中学习仍然是监督学习中常见和具有挑战性的问题, 因为标准分类算法是为处理平衡类分布而设计的。虽然有不同策略来解决这个问题, 但生成人工数据以实现均衡类分布的方法比对分类算法的修改更通用。这些称为 "超采样器" 的技术修改训练数据, 允许将任何分类器与类不平衡的数据集一起使用。许多算法已被提出了这项任务, 但大多数是复杂的, 往往会产生不必要的噪音。本文提出了一种简单有效的基于 **k** 均值聚类 and **smote** 过采样的过采样方法, 避免了噪音的产生, 有效地克服了类之间和类内部的不平衡。对 71 个数据集进行的大量实验的经验结果表明, 该方法对训练数据进行了过度采样, 提高了分类结果。此外, **k** 均值 **smote** 的性能始终优于其他流行的过采样方法。实现以巨蟒编程语言提供。少

2017 年 12 月 12 日提交;v1 于 2017 年 11 月 2 日提交;最初宣布 2017 年 11 月。

评论:19 页, 8 个数字

169. 第 1711.00137[[pdf](#), [ps](#),[其他](#)] Cs。艾

石榴: 巨蟒快速灵活的概率建模

作者:[jacob schreiber](#)

摘要: 我们提出了石榴, 一个开源机器学习包, 用于 **python** 中的概率建模。概率建模包含了广泛的方法, 这些方法使用概率分布显式描述不确定性。在石榴中实现的三个广

泛使用的概率模型是一般混合模型、隐藏马尔可夫模型和贝叶斯网络。石榴的一个主要重点是从其定义中抽象出训练模型的复杂性。这使得用户能够专注于为其应用程序指定正确的模型，而不是受到他们对基础算法的理解的限制。这一重点的一个方面涉及从数据集中收集足够的补充统计数据，作为培训模式的战略。这种方法可以轻松地实现许多有用的学习策略，例如内核外学习、微型学习和半监督学习，而无需用户考虑如何对数据进行分区或修改算法来处理这些任务自己。石榴是用 cython 编写的，目的是加快计算速度，并释放全局解释器锁，以允许内置多线程并行性，使其与类似算法的其他实现具有竞争力--或性能优于--。本文概述了石榴的设计选择，以及这些选择如何使复杂的功能能够通过简单的代码得到支持。少

2018 年 2 月 27 日提交;v1 于 2017 年 10 月 31 日提交;最初宣布 2017 年 11 月。

170. 第 1711.00046[[pdf](#),其他] Cs. Ds

按比例替换或检索文档中的关键字

作者:[维卡什·辛格](#)

摘要: 本文介绍了在给定文本中替换关键字或查找关键字的 flashtext 算法。flashtext 可以在文档的一次传递中搜索或替换关键字。此算法的时间复杂度并不取决于要搜索或替换的术语数。对于大小为 n (字符) 的文档和 m 关键字字典，时间复杂度为 $o(n)$ 。此算法比 regex 快得多，因为正则表达式时间复杂度为 $o(mn)$ 。它也不同于阿霍科拉西克算法，因为它不匹配子字符串。flashtext 设计为只匹配完整的单词 (两边都有边界字符的单词)。对于 {apple} 的输入字典，此算法将无法将其与 "我喜欢菠萝" 相匹配。该算法还设计为先进行最长的匹配。对于字符串 "我喜欢机器学习" 上的输入字典 {机器、学习、机器学习}，它将只考虑最长的匹配，即机器学习。我们已经将这种算法的巨蟒实现作为开源算法在 github 上提供，在许可的 mit 许可证下发布。少

2017 年 11 月 9 日提交;v1 于 2017 年 10 月 31 日提交;最初宣布 2017 年 11 月。

171. 第 1710.10296[[pdf](#)] cs. ne

一种深递归神经网络语言模型在 xilinx fpga 上的实现

作者:[郝玉峰](#),[史蒂文·奎格利](#)

摘要: 近年来, fpga 越来越多地应用于语音识别、机器学习、云计算等问题, 如微软使用的必应搜索引擎。这是因为与通用处理器相比, fpga 具有出色的并行计算能力以及较低的功耗。然而, 这些应用主要集中在大规模的 fpga 集群上, 这些集群对于执行大量矩阵或卷积操作具有极高的处理能力, 但不适合便携式或移动应用。本文介绍了单 fpga 平台的研究, 探讨了 fpga 在这些领域的应用。在本项目中, 我们设计了一个深度递归神经网络 (drnn) 语言模型 (lm), 并在 pynq 板上实现了一个带有 axxi 流接口的硬件加速器, 该主板配备了 xilinx zynq soc xc7z020 1clg400c。pynq 不仅具有丰富的可编程逻辑资源, 而且具有灵活的嵌入式操作系统, 适用于自然语言处理领域。我们使用 python 和 theano 设计 dmnn 语言模型, 在 cpu 平台上训练模型, 并在 pynq 板上部署模型, 以便使用 jupyter 笔记本电脑验证模型。同时, 设计了一种基于 pynq 的硬件库--叠加硬件加速器, 并验证了其对 pynq 板的加速效应。最后, 我们发现 dmnn 语言模型可以顺利地部署在嵌入式系统上, 而带有 axxi stream 接口的叠加加速器的执行速度为 20 个 gops 处理吞吐量, 与 30 日和第 31 号。少

2017 年 11 月 16 日提交;v1 于 2017 年 10 月 26 日提交;最初宣布 2017 年 10 月。

172. 第 1710.09578[[pdf](#), [ps](#),其他] cs. ms

python 中的快速线性转换

作者:[christoph wagner](#), [sebastian semper](#)

摘要: 本文为 **python** 编程语言引入了一个新的自由库, 它提供了结构化线性转换的集合, 这些转换不是显式二维数组, 而是以更有效的方式通过利用结构知识。这允许快速和内存愉快的向前和向后转换, 同时也推进一个干净但仍然灵活的接口, 这些高效的算法, 从而使代码更具可读性, 可扫描性和适应性。我们首先概述了这个库的目标, 然后是如何实现这些目标的, 最后我们展示了与 **python** 可用的最先进包的当前状态相比的性能。此库是在免费许可证下发布和分发的。少

2017 年 10 月 26 日提交;最初宣布 2017 年 10 月。

173. 第 1710.08195[[pdf](#),[其他](#)] **lo c**

反应系统工具集的细化计算

作者:[iulia dragomir](#), [viorel preoteasa](#), [stavros tripakis](#)

摘要: 反应系统工具集的细化微积分, 这是一个用于反应系统的组合建模和推理的环境, 建立在 [isabelle](#)、[simulink](#) 和 **python** 的基础上。

2018 年 2 月 23 日提交;v1 于 2017 年 10 月 23 日提交;最初宣布 2017 年 10 月。

174. 第: 1710.06915[[pdf](#),[其他](#)] **cs.PL**

多伊 [10.250 80/shinma-7f4c6e7-00b](#)

匹配: 一种模式匹配库

作者:[manuel krebber](#), [henrik barthels](#), [paolo bientinesi](#)

摘要: 模式匹配是符号计算的有力工具, 基于明确定义的术语重写系统理论。应用程序域包括代数表达式、抽象语法树以及 [xml](#) 和 [json](#) 数据。遗憾的是, [python maics](#)、[配合比](#)、[模式](#)、[pypatt](#) 都没有像 [mathematica](#) 那样通用和灵活的模式匹配的轻量级实现。因此, 我们创建了开源模块匹配, 它在 **python** 中提供了类似的模式匹配功能, 它使用一种新的算法, 通过利用模式之间的相似性, 更有效地查找大型模式集的匹配。少

2017 年 10 月 16 日提交;最初宣布 2017 年 10 月。

评论:[arxiv](#) 管理说明: 与 [arxiv:1710.00077](#) 的实质性文本重叠

175. 第 1710.06590[[pdf](#)] **Cs. DI**

medoc: 将 medline 加载到本地 mysql 数据库中的 python 包装器

作者:[emeric dynomant](#), [mathilde gorieu](#), [helene perrin](#), [marion denorme](#), [fabien picon](#), [amaud desfeux](#)

摘要: 自 [medline](#) 数据库发布以来, 该实体索引的文档数量每年都在增加。国家卫生研究院 ([nih](#)) 开发了一些工具来质疑这一科学出版物。但是, 就大数据、文本挖掘和数据科学方面的进步而言, 构建包含 [medline](#) 上所有可用元数据的本地关系数据库的选项对于优化利用这些资源将非常有用。[medoc](#) ([medline DOWnloading contrivance](#)) 是一个 **python** 程序, 旨在下载 [ftp](#) 上的数据, 并将所有提取的信息加载到本地 [mysql](#) 数据库中。[medoc](#) 花了 4 天 17 个小时才将此服务器上可用的 2600 万个文档加载到标准计算机上。此索引关系数据库允许用户生成复杂而快速的查询。因此, 可以搜索所有字段所需的信息, 这是很难通过 [pubmed](#) 图形界面完成的任务。[medoc](#) 在 <https://github.com/MrMimic/MEDOC> 免费和公开提供。少

2017 年 10 月 18 日提交;最初宣布 2017 年 10 月。

评论:4 页, 1 个图

176. 第 1710.04196[[pdf](#),[其他](#)] Cs. Sd

pyroom 声学: 用于音频室模拟和阵列处理算法的 **python** 包

作者:[robin scheibler](#), [eric bezzam](#), [ivan Dokmanić](#)

摘要: 我们提出了一个针对音频阵列处理算法的快速开发和测试的软件包--热室声学。包的内容可分为三个主要组件: 直观的 **python** 面向对象接口, 以快速构建不同的模拟场景, 包括 2d 和 3d 房间中的多个声源和麦克风; 一种快速 c 实现的一般多面体室图像源模型, 有效地生成室内脉冲响应, 模拟源和接收机之间的传播; 最后, 参考了波束形成、测向和自适应滤波等常用算法的实现。它们共同构成了一个包, 有可能通过显著降低性能评估步骤中的实现开销来加快新算法的上市时间。少

2017 年 10 月 11 日提交;最初宣布 2017 年 10 月。

评论:5 页, 5 个数字, 描述一个软件包

177. 第 1710.03561[[pdf](#),[其他](#)] Cs. pf

ciw: 开源离散事件仿真库

作者:[geraint i. palmer](#), [vincent a.knight](#), [paul r. harper](#), [asyl i. hawa](#)

文摘: 本文介绍了 ciw, 这是一个用于执行 **python** 中开发的离散事件模拟的开源库。从最佳实践和计算研究的重现性方面说明了该库的优势。分析了 ciw 的性能, 并与几种可供选择的离散事件仿真框架进行了比较。

2017 年 9 月 27 日提交;最初宣布 2017 年 10 月。

178. 第: 1710.0013 条[[pdf](#),[其他](#)] Cs. Db

一种实用的 **python api, 用于查询 aflowlib**

作者:[conred w. rosenbrock](#)

摘要: 大型数据库 (如 [aflowlib.org](#)) 为通过机器学习发现材料趋势提供了宝贵的数据源。尽管有 rest api 和查询语言, 但与 aflux 语言相关的学习曲线对新用户起到了一定的作用。此外, 数据使用非标准序列化格式存储。在这里, 我们提供了一个高级 api, 允许使用标准的巨蟒运算符和语言功能立即访问 flowlib 数据。它提供了一种将 flowlib 数据与其他巨蟒材料包 (如 ase 和 quippy) 集成的简单方法, 并将自动反序列化到 numpy 数组和巨蟒对象中。此软件包可通过 "pip 安装流程" 获得。少

2017 年 9 月 28 日提交;最初宣布 2017 年 10 月。

评论:7 页, 3 个代码列表

179. 第: 1710.00379[[pdf](#),[其他](#)] Cs. Lg

剧本: **python 中基于池的主动学习**

作者:[杨耀元](#), [李少川](#), [钟玉安](#), [吴东恩](#), [陈思安](#), [林宣天](#)

摘要: libact 是一个 **python** 包, 旨在使一般用户更轻松地进行主动学习。该软件包不仅实现了几种流行的主动学习策略, 而且还采用了主动自学元算法, 帮助用户在飞行中自动选择最佳策略。此外, 该包还提供了一个统一的接口, 用于实现更多的策略、模型和特定于应用程序的标签。该包在 github 上是开源的, 可以很容易地从 **python** 包索引存储库中安装。少

2017 年 10 月 1 日提交;最初宣布 2017 年 10 月。

180. 第 1710.00077[[pdf](#),[其他](#)] cs.PL

****python** 中的高效模式匹配**

作者:manuel krebber, henrik barthels, paolo bientinesi

摘要: 模式匹配是符号计算的强大工具。应用程序包括术语重写系统, 以及符号表达式、抽象语法树、xml 和 json 数据的操作。它还允许以重写规则的形式直观地描述算法。我们提出了开源 **python** 模块匹配, 它提供的功能和表现力类似于 mathematica 中的模式匹配。特别是, 它包括句法模式匹配, 以及交换函数和/或关联函数、序列变量和约束匹配的匹配。match 使用新的和改进的算法, 通过利用模式之间的相似性, 有效地查找大型模式集的匹配项。对 matchpy 的性能进行了一些实际问题的研究。少

2017 年 9 月 29 日提交;最初宣布 2017 年 10 月。

181. 第: 1710.00027[[pdf](#),[其他](#)] Cs. Db

面向数据集成的系统构建议程

作者:anhai doan, adel ardalan, jeffrey r. ballard, sanjib das, yash govind, pradap konda, han li, erik paulson, paul Suganthan g. c., haojun zhang

摘要: 本文认为, 数据管理界应该投入更多的精力来构建数据集成 (di) 系统, 以便真正推进该领域。为了实现这个目标, 我们做出了三个贡献。首先, 我们借鉴最近的工业经验, 讨论当前直接投资系统的局限性。其次, 我们提出了一个议程, 以建立一种新的 di 系统, 以解决这些限制。这些系统逐步指导用户完成 di 工作流。它们提供了解决步骤 "痛点" 的工具, 这些工具是在 **python** 数据科学和大数据生态系统 (pydata) 的基础上构建的。我们讨论如何在 pydata 中培育此类工具的生态系统, 然后使用它来构建用于协作的 di 系统。最后, 我们讨论了威斯康星州正在进行的工作, 这表明这些直接投资系统非常有希望, 建立它们带来了许多有趣的研究挑战。少

2017 年 9 月 29 日提交;最初宣布 2017 年 10 月。

182. 建议: 1709.09 480[[pdf](#),[其他](#)] Cs. 艾

多伊 10.1109/SSCI.2017.8280935

工业控制问题的基准环境驱动

作者:daniel hein, stefan depeweg, michel Runkler, steffen udluft, 亚历山大 hentschel, thomas a.runkler, Runkler sterzing

摘要: 在强化学习的研究领域, 经常开发出新颖而有希望的方法, 并将其引入 rl 社区。然而, 尽管许多研究人员热衷于将他们的方法应用于现实世界的问题, 但在实际的工业环境中实施这样的方法往往是一个令人沮丧和繁琐的过程。一般来说, 学术研究小组获得实际工业数据和应用的机会有限。因此, 通常会使用人工软件基准来开发、评估和比较新方法。一方面, 这些基准旨在提供可解释的 rl 培训方案, 并详细了解现有方法的学习过程。另一方面, 它们通常与工业实际应用没有太大相似之处。为此, 我们利用我们的行业经验设计了一个基准, 以弥补免费提供、记录和激励的人为基准与实际工业问题的特性之间的差距。由此产生的工业基准 (ib) 已通过在其 github 上发布其 java 和 **python** 代码 (包括 openai gym 包装器) 向 rl 社区公开。在本文中, 我们对 ib 的动力学进行了详细的激励和描述, 并确定了能够捕获实际行业控制问题中常见情况的原型实验设置。少

2018 年 2 月 6 日提交;v1 于 2017 年 9 月 27 日提交;最初宣布 2017 年 9 月。

期刊参考: 2017 年 iee 计算智能研讨会系列 (ssci)

183. 第: 1709.09069[[pdf](#),[其他](#)] Cs. Lg

开放健身房的 mdp 环境

作者:[安德烈亚斯·基尔希](#)

摘要: openai 健身房为研究人员和爱好者提供了简单易用的环境来加强学习。即使是最简单的环境也有一定程度的复杂性, 可以模糊 rl 方法的内部工作方式, 使调试变得困难。本白皮书介绍了一个 **python** 框架, 通过在 **python** 中特定于域的语言。然后, 它显示使用此 mdp 框架创建的结果和可视化效果。少

2017 年 9 月 26 日提交;最初宣布 2017 年 9 月。

评论:6 页, 2 个数字

184. 第: [1709.08842](#)[pdf,其他] Cs. Lg

利用脉冲学习音乐的预测模型

作者:[jonas langhabel](#)

摘要:用于迭代扩展和剔除要素集, 有效地探索对于常见特征选择方法来说太大的特征空间。我为 pulse 设计了一个通用的 **python** 框架, 提出了任务优化的特征生成操作和各种音乐理论驱动的功能, 并在标准的单声道 f.. 更多

2017 年 9 月 26 日提交;最初宣布 2017 年 9 月。

评论:硕士论文

185. 建议: [1709.08363](#)[pdf, ps,其他] 反渗透委员会

节点基元: 面向社交机器人的开放式最终用户编程平台

作者:[enrique coronado](#), [fulvio mastrogiovanni](#), [gentiane venture](#)

抽象: 随着机器人的预期采用能够在真实场景中无缝、直观地与人进行交互, 因此需要为非技术熟练的用户提供易于理解的机器人行为范式。本文提出了一个交互设计机器人编程平台, 实现多学科社会机器人的研究和应用。此平台称为节点基元 (nep), 由两个主要部分组成。一方面, 开发了基于 zeromq 和 **python** 的分布式软件框架, 以提供进程间通信和机器人行为规范机制。另一方面, 开发了基于 web 的最终用户编程 (eup) 界面, 以便能够轻松、直观地编程和执行机器人行为。为了评价 nep, 我们讨论了利用手臂手势控制机器人行为的人机交互应用的开发。并对所提出的 eup 接口进行了可用性测试。少

2017 年 9 月 25 日提交;最初宣布 2017 年 9 月。

msc 类: 68t40

186. 第 [xiv:1709.07541](#)[pdf,其他] Cs. Sd

民歌研究中的声调声音基本频率估计方法

作者:[c. jarne](#)

摘要: 本文提出了一种快速实现基频估计的方法。该算法是基于频域方法的。它主要是为音调的声音而开发的, 并用于金丝雀鸟歌曲分析。该方法已实现, 但不限于此类数据。它可以很容易地适应其他建议。**python** 库被用来开发一个具有简单算法的代码来获得基本频率。本地大学存储库中提供了一个简单的开放源代码。少

2017 年 9 月 21 日提交;最初宣布 2017 年 9 月。

评论:声调声音基频估计方法的预打印版本

187. 第 [xiv:1709.06743](#)[pdf,其他] Cs. Ce

pandapower-用于方便电力系统建模、分析和优化的开源 python 工具

作者:leon thurner,亚历山大·谢德勒,florian schäfer ,jam-hendrik menke,julian dollichon,friederike meier,steffen meinecke,martin braun

摘要: pandapower 是一种基于 **python** 的 **bsd** 许可电力系统分析工具,旨在实现平衡电力系统静态和准静态分析的自动化和优化。它根据 iec 60909 提供潮流、最优潮流、状态估计、拓扑图搜索和短路计算。pandapower 包括以前基于 pypower 的牛顿-拉斐逊功率流求解器,该求解器已通过及时编译得到了加速。求解器的其他增强功能包括对恒流负载建模的功能、具有多个参考节点的网格和连接检查。pandapower 网络模型基于电气元件,如线路、两线和三绕组变压器或理想开关。所有元件都可以用铭牌参数定义,并在内部用等效电路模型进行处理,这些模型已经根据行业标准软件工具进行了验证。用于定义网络的表格数据结构基于 **python** 库熊猫,它允许轻松地处理输入和输出参数。**python** 中的实现使 pandapower 易于使用,并允许使用第三方库进行舒适的扩展。pandapower 已成功地应用于多个网格研究以及教育目的。一项全面、公开的案例研究表明,熊猫在自动时间序列计算中可能得到应用。少

2018 年 4 月 18 日提交;v1 于 2017 年 9 月 20 日提交;最初宣布 2017 年 9 月。

188. 第 07:170 9.06 169[[pdf](#),[其他](#)] Cs。Ds

改进拼接对齐,以识别正交组和多个 cds 对齐

作者:jean-david acilar, safa jammali, esie kuitche, aida ouangoua

文摘: 拼接对齐问题 (sap), 包括在一个没有拼接的基因组序列上寻找拼接 ma 序列的最佳半全局对齐,在很大程度上被考虑为基因组中基因结构的预测和注释。在这里,我们重新访问它的目的是从同源基因中识别一组 cds 正畸群,并计算多个 cds 对齐。我们引入了一个新的约束版本的拼接对齐问题,以及一个算法,利用输入 ma 和基因序列的外显子结构的完整信息,以计算高覆盖准确的对齐。我们展示了如何将 cds 和基因家族的基因序列之间的成对拼接对齐直接用于将基因家族的 cds 集合集中到一组正畸群中。我们还引入了一个被称为 "多拼接对齐问题" (msap) 的拼接对齐问题的扩展,该问题包括在同一基因家族的多个基因上同时对齐多个 ma 序列。我们为这个问题开发了一个启发式算法解决方案。我们展示了如何利用多个拼接对齐将同源 cds 聚集到正交和密切的分体组中,以及如何构造多个 cds 对齐。python 中方法的实现可在 demande 上 SFA@USherbrooke.ca。关键词: 分裂对齐, cds 正畸组, 多 cds 对齐, 基因结构, 基因家族。少

2017 年 9 月 18 日提交;最初宣布 2017 年 9 月。

评论22 页, 7 个数字