

AutoML相关

为什么要有AutoML?

当前机器学习的应用需要大量的人工干预，具体体现在特征提取，模型选择，参数调节等方面，也就是在专家的经验下提高模型性能.为了使得机器学习技术更加容易应用，减少对机器学习专家的依赖,AutoML应运而生并且成为工业界和学术界的一个热门话题。

引入和贡献

没有免费的午餐理论表明，没有一种算法可以在所有问题上性能都超过其他算法，因此从特征提取，模型选择和算法选择的每一步都需要精心设计。专家资源稀缺，造成机器学习代价巨大.AutoML本着使机器学习摆脱专家的束缚的理想，使得机器学习更加平民化，加快工业和学术的发展，出于这些学术理想和工业需求，AutoML成为机器学习中一个新的子领域并且在计算机视觉，数据挖掘和自然语言处理中得到广泛关注。它已经在很多问题中取得成功。

主要贡献如下:

- 1.讨论AutoML的定义
- 2.针对当前的AutoML方法提出一个统一的框架
- 3.将AutoML按照什么需要自动化和如何自动化分类
- 4.和相关的综述相比，本文基于所提出的框架，对已存在的方法进行详细分析
- 5.从存在的问题,技术，应用理论四个方面提出未来研究者在AutoML中的研究方向。

什么是AutoML?定义

非正式定义:

max performance of learning tools,

s.t.

no human assistance

limited computational budget

正式定义:

AutoML是在没有人工帮助和有限的计算负担下建立的机器学习算法使得对于数据集E上的任务T，具有良好的性能P. AutoML的三个目标:

(A) 良好的性能:对于不同输入和学习任务,算法具有良好泛化性 (B)不需要人的帮助 (C)高效

统一框架

AutoML controller 取代人为为学习工具寻找正确的设置。它有两个关键组成部分，优化器和验证器。验证器用于衡量学习工具在给定设置下的性能；优化器用于更新学习工具的设置。

AutoML approaches

what to automate:by problem setup

1.特征工程 2.模型选择 3.优化算法选择 三个部分组成一般机器学习的核心;NAS作为特殊情况同时对这三个方面进行设置。

how to automate :by techniques

1.基础技术:网格搜索和随机搜索,强化学习和自动分化 2.经验丰富的技术:元学习和迁移学习

细节

1.特征工程:特征降维,特征生成,特征编码;搜索空间:特征增强工具在超参数设置,特征生成和选择 2.模型选择:自动选择一系列分类器并且设置超参数，使得具有良好学习性能。分类器工具如树分类器，线性分类器等;搜索空间包含这些分类器及其对应超参数 3.优化算法选择:GD,L-BFGS,SGD

AutoML 概况

完整的AutoML过程可以分为两部分:将特征选择，模型选择和优化算法整合一起的完整的pipeline;另一类是网络架构搜索，自动学习最优网络结构