



暨南大学
JINAN UNIVERSITY

本科生课程论文

论文题目: Combinatorial Testing Metrics for Machine Learning

学 院: 信息科学技术学院

学 系: 计算机科学系

专 业: 网络工程

课程名称: 软件测试与质量保证

学生姓名: 阮炜霖

学 号: 2020101603

指导教师: 孙玉霞

2021 年 12 月 5 日

目标：

- (1) 提交**至少 1500 字的中文写的**课程报告, 讲述论文的摘要；研究问题、解决问题的基本思想和具体方案、实验设计和结果。最好能亲自复现实验。
- (2) **制作 PPT，并录制至少 30 分钟的讲解视频**（播放 PPT+讲解论文+录屏），应当清楚地讲解论文。
- (3) 如果能复现实验或提出新想法，并**与老师深入讨论文文的，有额外加分**。

[摘要]

[问题描述]

[解决思路]

[具体方案]

[实验设计]

[实验结果]

简单翻译一下论文——

【摘要】为比较机器学习数据集的差异，本论文定义了一种新的组合覆盖率并且以函数形式返回其差异。本文举例阐明了这种方法在估计和评测机器学习模型的性能上的性能。识别和测量数据集之间的差异在机器学习领域是一个非常重要的问题，模型的准确性很大程度上取决于训练数据能否充分体现应用程序中遇到的数据。迁移学习（即是否能在其他数据集上也能有良好的预测结果）能说明模型的性能。

【1. 介绍】

在软硬件中，组件系统经常能很好地被设计和测试，但是组件对接的时候往往会因为意想不到的交互出现错误。一份基于经验数据的研究发现几乎所有的软件错误都是由有限数量的正在交互的组件造成的，并且得出一个结论：测试四到六个组件之间的交互足以抵御软件系统中所能考虑到的错误。组件之间的交互在复杂系统中称之为“影响因素”，如果测试用例不足以覆盖这些互动，它会驱动一些意想不到的行为，通常有效的互动有其限定规模，这导致在软件和硬件系统中伪穷举测试中能采用组合测试。

带测试系统的嵌入式处理机使用老套的软件方法因为大规模输入域、白盒测试需求效果以及集成或系统级别才会遇到的紧急行为等特性而遇到难题。CT 是为了大规模输入域提供伪穷举策略以测试集成系统的一个黑盒方法。进而，CT(组合测试)拥有嵌入式 ML（机器学习组件），测试一个神经网络的内状态空间和可解释机器学习的自治交通系统。

在机器学习中数据是算法表现的基础。在本文我们通过对数据集的比较来充分利用 CT 测试 ML 系统。我们考虑以下问题：

- ①两类成员的差别如何去影响分类决策？
- ②数据集之间的差异如何用来预测在一个数据集
- ③训练的模型可以适用于另外一个模型？

我们通过组合比较数据集大概有三种颗粒等级：

- ①组合显示或者隐藏的数量；
- ②那些显示或者隐藏的独特组合；
- ③那些组合的贡献。

在这次工作中，为了比较第二部分第一颗粒等级的 ML 数据集，我们定义了一个新的组合覆盖率。在第三部分，我们标亮了两个不同的应用程序规则。故障点测试和可解释性分类部分表明了规则

是怎样在数据中的可说明失误中起作用的。第二个应用程序使用这套规则去描述模型的极限数据。操控极限数据的模型的概念拥有应用程序在传迁移学习，训练的选择和测试用例，以及指挥数据收集和标签成果。我们会在第四部分谈论以后的工作。

【2. 规则】

我们将机器学习中的“特征”看作影响因素，因此每个因素在所给的数据点中都有一个特殊的值。连续赋值的因素在应用于 CT 之前必须离散化使得每个因素都有一个对应的值的集合。一个 t 路集成是 t 的一个具体赋值，或者 t 元组组数。如果有 k 个可识别特征，那么每个数据点都有 $C(k, t)$ 个集成。

组合覆盖率，也称 total t -way 覆盖，是一篇描述集合中出现的可能的 t -way 集成的部分的 CT 文章引入的规则。在集合中出现的集成会被集合所覆盖。定义一个 k 因素的体系和他们各自的等级， U 是所有可能的数据点的集合， U_t 是所有可能的 t -way 集成的集合。如果某些集成不可能，它会被束缚并从 U_t 中去除。给一个数据集 D 是 U 的子集，定义

D_t 是 D 中出现的 t -way 组合的集合。（我们承认稍微滥用标记 D 可能是一个多重集合，这不影响）