

综述. 特征选择方法

MF1733062 万晨 weanl_jc@163.com

2018 年 4 月 8 日

1. 介绍

FS 在 ML 中能解决那些问题? 已经解决得怎么样了? 还有那些问题? feature construction = FS + FE 特征选择 (Feature, Variable and Attribution Selection), 是机器学习 feature construction 的重要组成部分。在筛选原始数据, 构造有用的特征集合方面, 特征选择不同于特征提取 (Feature Extraction): 后者会通过线性或非线性的方式从原始数据中构造出全新的特征, 具有特征学习和表示学习的能力 [”Representation Learning: A Review and New Perspectives”]; 特征选择通过设计一些简单高效或者精致巧妙的方法, 实现从原始特征集合中选出最优的特征子集, 能够保持特征对应的原始物理意义。所谓最优特征子集, 理论上定义为没有信息丢失的最小特征子集, 以 Markov blanket 的形式给出 [D. Koller, Toward optimal feature selection][C.Aliferis, Local causal and markov blanket induction for causal discovery and feature selection]; 实际中理论的 ground-truth 很难找, 所以经验上一般我们用预测器性能 (如分类器的精度) 来评估特征子集选择的结果。

2. 过滤式与包裹式