

Weka 数据分类、回归挖掘工具

WEKA 的全名是怀卡托智能分析环境 (Waikato Environment for Knowledge Analysis) , 其源代码可从 <http://www.cs.waikato.ac.nz/ml/weka/> 得到。

2005 年 8 月, 在第 11 届 ACM SIGKDD 国际会议上, 怀卡托大学的 WEKA 小组荣获了数据挖掘和知识探索领域的最高服务奖, WEKA 系统得到了广泛的认可, 被誉为数据挖掘和机器学习历史上的里程碑, 是现今最完备的数据挖掘工具之一。

作为一个大众化的数据挖掘工作平台, WEKA 集成了大量能承担数据挖掘任务的机器学习算法, 包括对**数据进行预处理、分类、回归、聚类、关联分析以及在新的交互式界面上的可视化**等等。通过其接口, 可在其基础上实现自己的数据挖掘算法。

WEKA 中的典型算法

Bayes: 贝叶斯分类器

BayesNet: 贝叶斯信念网络

NaïveBayes: 朴素贝叶斯网络

Functions: 人工神经网络和支持向量机

MultilayerPerceptron: 多层前馈人工神经网络

SMO: 支持向量机 (采用顺序最优化学习方法)

Lazy: 基于实例的分类器

IB1: 1-最近邻分类器

IBk: k-最近邻分类器

Meta: 组合方法

AdaBoostM1: AdaBoost M1 方法

Bagging: 袋装方法

Rules: 基于规则的分类器

JRip: 直接方法—Ripper 算法

Part: 间接方法—从 J48 产生的决策树抽取规则 (不是 C4.5 规则算法)

Trees: 决策树分类器

Id3: ID3 决策树学习算法 (不支持连续属性)

J48: C4.5 决策树学习算法 (第 8 版本)

RandomForest: 基于决策树的组合方法