

doi:10.1631/FITEE.1400398

题目：一种概念漂移情况下数据流分类的整体方法

目的：数据流（data stream）管理和处理是计算机科学领域的热点问题。本文提及的“数据流”指连续且快速生成的数据包。数据流的专有特性有数据量极大、生成率高、处理时间有限和数据概念漂移（concept drift）等。这些特性将数据流区别于其他标准数据形式。数据流的一个重要问题即为输入数据分类。本文提出一种新型的整体分类器（ensemble classifier）。

创新点：在数据流分类器的基础上，提出一种包含概念漂移检测、基分类器移除和动态加权机制的方法。

方法：（1）针对不同数据输入条件，对基分类器使用两种加权函数；（2）利用 Kappa 系数确定概念漂移，提升算法精度；（3）基于基分类器的质量，移除不同数目的基分类器；（4）在决策阶段对基分类器应用加权机制，提升算法对漂移的适应性，提高分类器效率。

结论：在标准数据集上测试，本文方法较现有整体分类器和单分类器可获得更高的精度；在某些情况下可节省运行时间和内存用量。

关键词：数据流；分类；整体分类器；概念漂移