APCluster 使用手册

安装

1、环境要求：已安装 JDK1.6 及其以上，Weka3.7.X 以上

2、集成 APCluster

APCluster 提供了使用 package manager 集成到 Weka 的方式，可方便地通过 URL 在线获取 APCLuster.zip 进行集成。URL 地址为

http://datamining.xmu.edu.cn/main/~chenwq/downloads/APCluster.zip

安装过程按下图顺序：

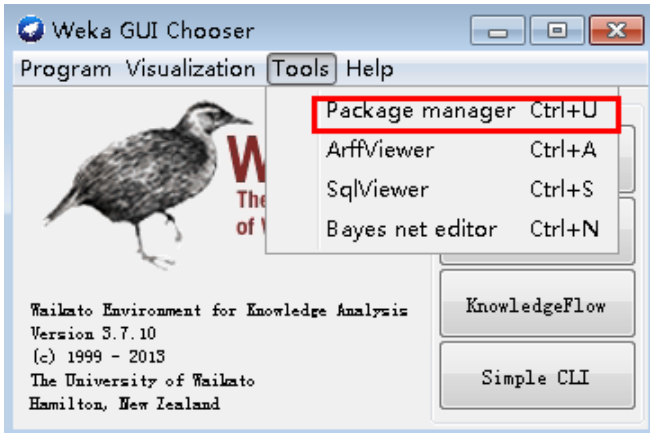


图 1 打开 Weka 的 Package manager

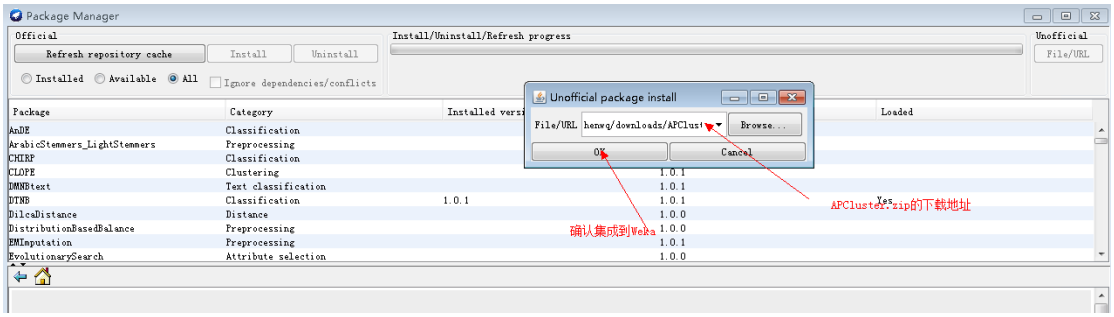


图 2 将 APCluster.zip 的 URL 复制到 File/URL 位置，OK

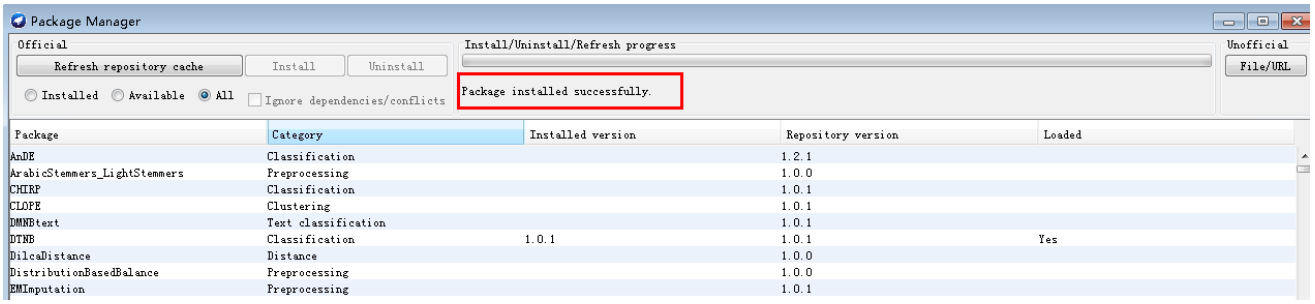


图 3 安装成功



图 4 重启 Weka

使用说明

重启 Weka 之后，APCluster 可在 Cluster 选项卡中找到，如图 5：

APCluster 的使用方法同 Weka 的其他聚类算法，下面说明该算法的参数的意义：

Parameter	Requirement	Description
convit	int, optional, default, 50	Number of iterations with no change in the clusters that stop the convergence
dampingFactor	double, optional, default, 0.9	Damping factor between 0.5 and 1
distanceFunction	Optional, default, weka.core.EuclideanDistance	The distance function to use for instances comparison
maxIterations	int, optional, default, 1000	Maximum number of iterations
numExecutionSlots	int optional, default, 1	The number of execution slots (threads) to use. Set equal to the number of available cpu/cores
Preference	Optional, default, maximum of similarity	The number of exemplars, ie of clusters, is influenced by the input preferences value. Preferences for each point - points with larger values of preferences are more likely to be chosen as exemplars.
useLastSims	Bool, optional, default, False	Whether use the similarity matrix calculated latest by APCluster

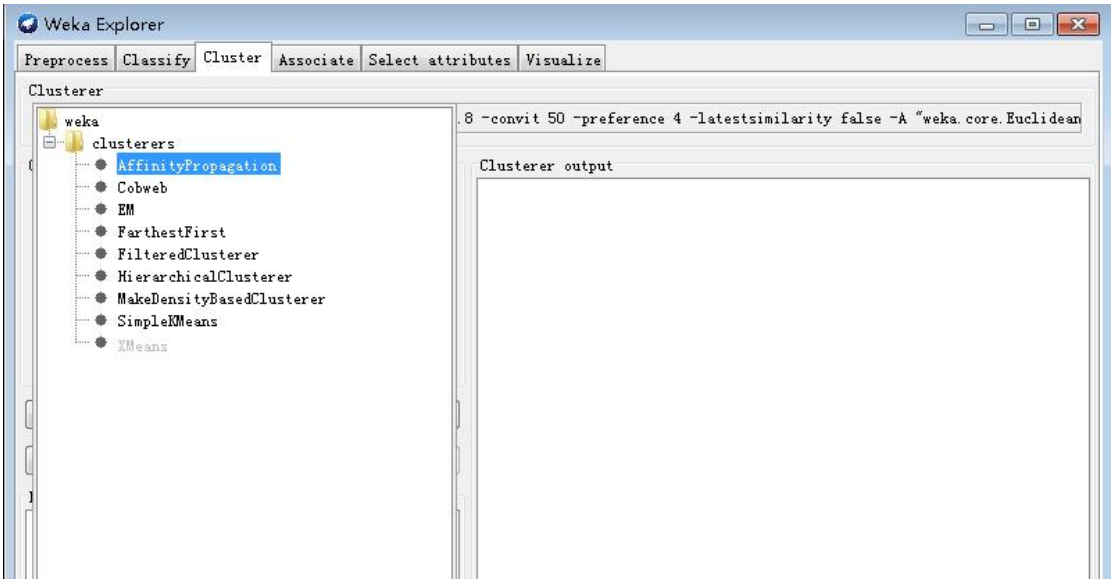


图 5 APCluster 的选项卡

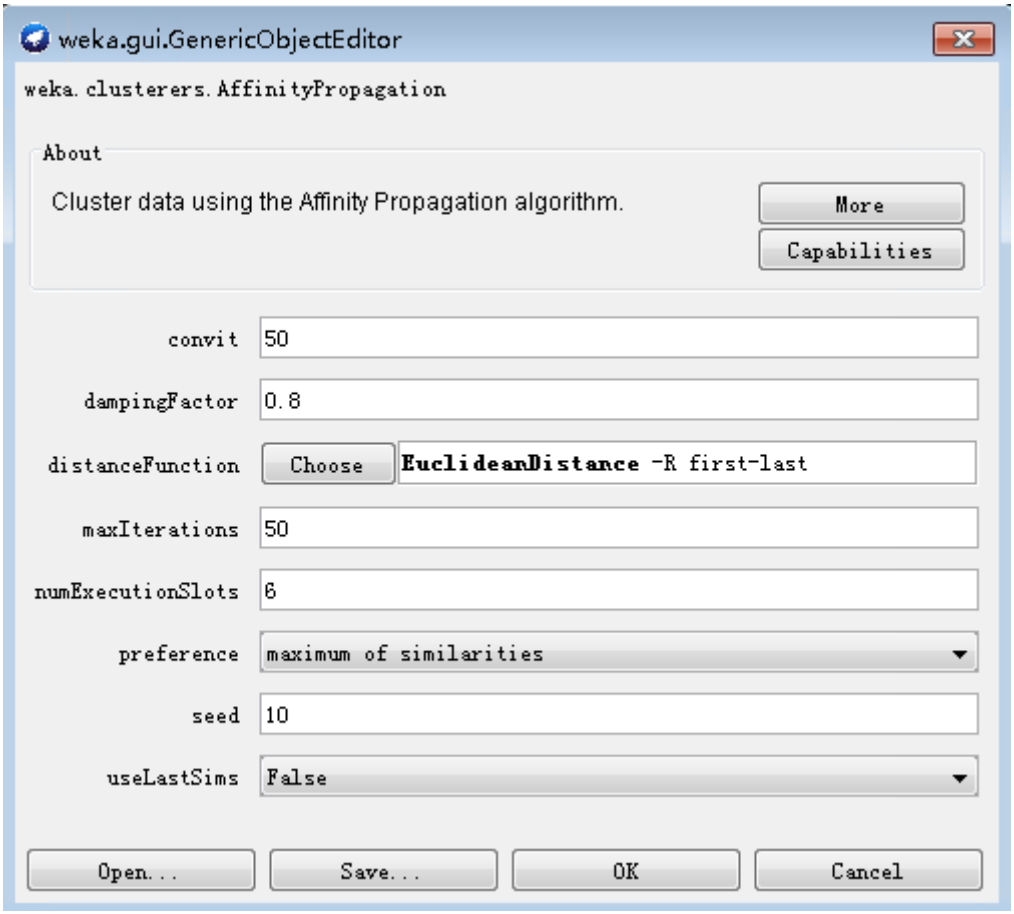


图 6 APCluster 参数

例子

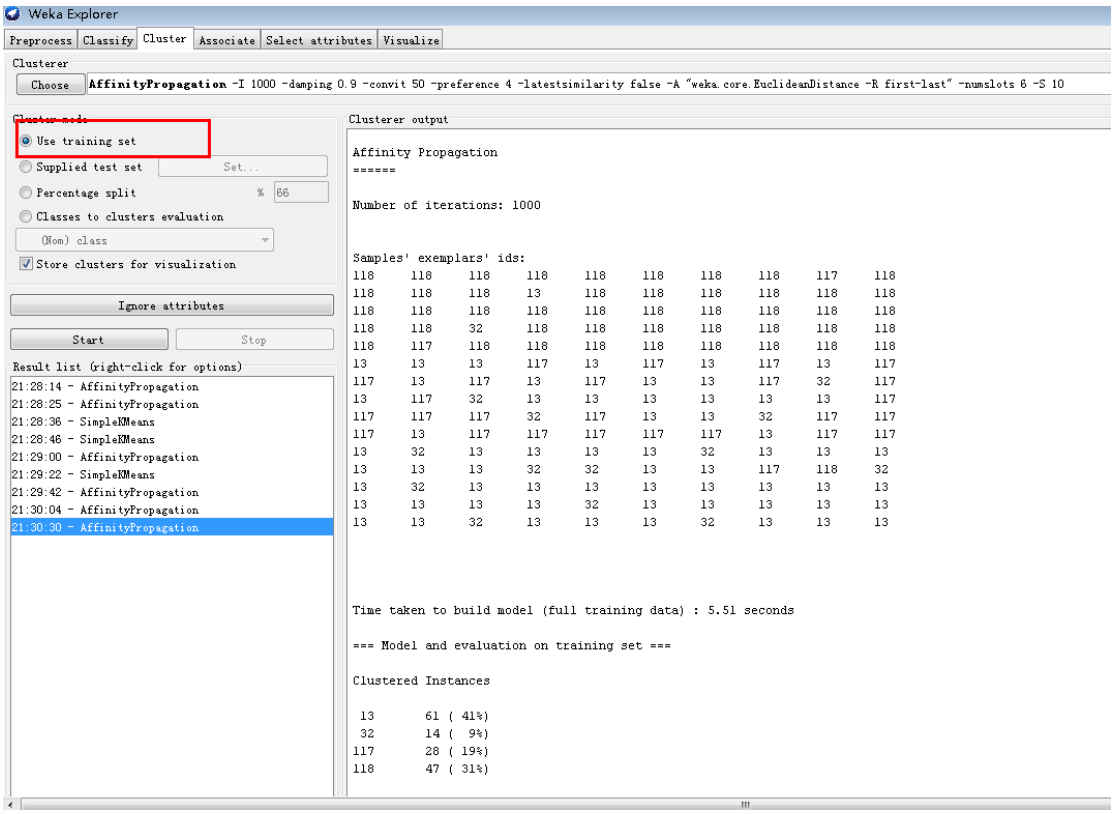


图 7 APCluster 例子

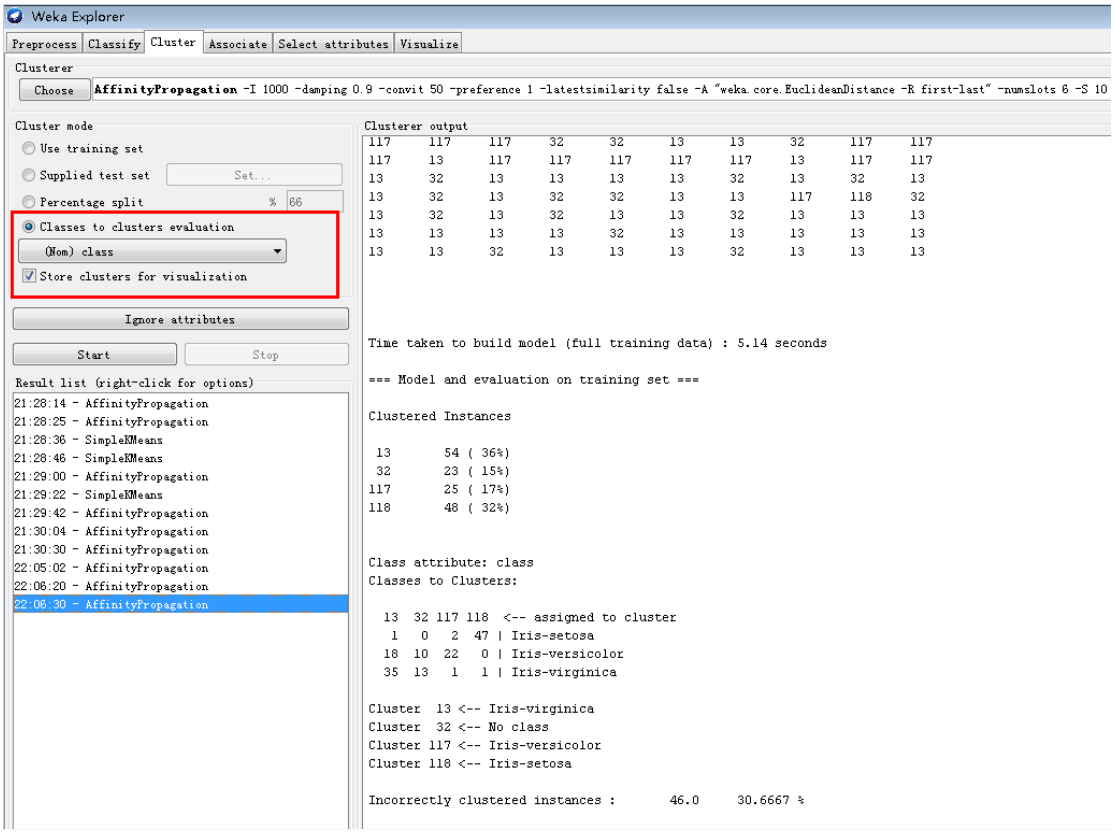


图 8 APCluster 例子 2

图 7 采用 Using Training set（上图的方框位置）：Samples' exemplars' ids 表示每个样本所属的聚类中心 id。例如第一行表示 id 为 0-7，9 的样本的聚类中心是 id 为 118 的样本，id 为 8 的样本的聚类中心是 id 为 117 的样本。

图 8 采用 Classes to clusters evaluation 的运行结果，相比图 7，多了一些类别信息。

并行化实现技术

并行化技术主要采用 JDK1.6 及其以上支持的线程池相关技术，分别是 ExecutorService，ThreadPoolExecutor 以及 Future 设计模式，具体实现细节参考 Java API 文档。

集成新开发分类器到 Weka 的方法

分类器编写规范参考 Weka 的文档：WekaManual.pdf

陈文强
lrwenqiang@gmail.com
2013.12.18