User Guide for Auto-WEKA version 1.0

Lars Kotthoff, Chris Thornton, Frank Hutter
{larsko,cwthornt,hutter}@cs.ubc.ca

March 4, 2016

Contents

1 Introduction

Auto-WEKA is a tool that performs combined algorithm selection and hyperparmeter optimisation over the classification and regression algorithms implements in WEKA. More specifically, given a specific dataset, Auto-WEKA explores hyperparameter settings for many algorithms and recommends to a user which method will likely have good generalization performance, using model based optimisation techniques.

1.1 Availability

Auto-WEKA is available as a WEKA package through the WEKA package manager (WEKA version 3.7.2 and later). The source code is available at `https://github.com/larskotthoff/autoweka`, where bugs can be reported as well.

1.2 License

Auto-WEKA is open source software issued under the GNU General Public License. Note that the included SMAC optimisation method is licensed under the AGPLv3.

1.3 Requirements

Auto-WEKA does not have any additional requirements compared to WEKA. If you can run WEKA, you should be able to run Auto-WEKA.

2 Auto-WEKA Overview

Auto-WEKA is used much like any other WEKA classifier. After loading a dataset into WEKA, it can be run on it to automatically determine the best WEKA model and its parameters.

2.1 Using the GUI

There are two different ways of using Auto-WEKA through the WEKA GUI. The easiest way is to use the Auto-WEKA panel, which allows you to run Auto-WEKA directly on a loaded dataset. Figure 1 shows a screenshot of a completed run.

Alternatively, Auto-WEKA can be run through the normal "Classify" panel by selecting it from the list of classifiers (Figure 2).

When using Auto-WEKA like a normal classifier, it is important to select the Test option "Use training set". Auto-WEKA performs a statistically rigorous evaluation internally and does not require an external split into training and test sets that WEKA provides. Not selecting this option will not improve the quality of the result and cause Auto-WEKA to take much longer. Figure 3 shows the recommended setting.

Auto-WEKA has only a few options. Figure 4 shows them. Usually, you can leave them at their default values. For most users, only two options are relevant:

timeLimit The time in minutes Auto-WEKA will take to determine the best classifier and configuration. If you get bad results, try increasing this value.

memLimit The memory limit in Megabytes for running classifiers. If you have a very large dataset, you may need to increase this value.

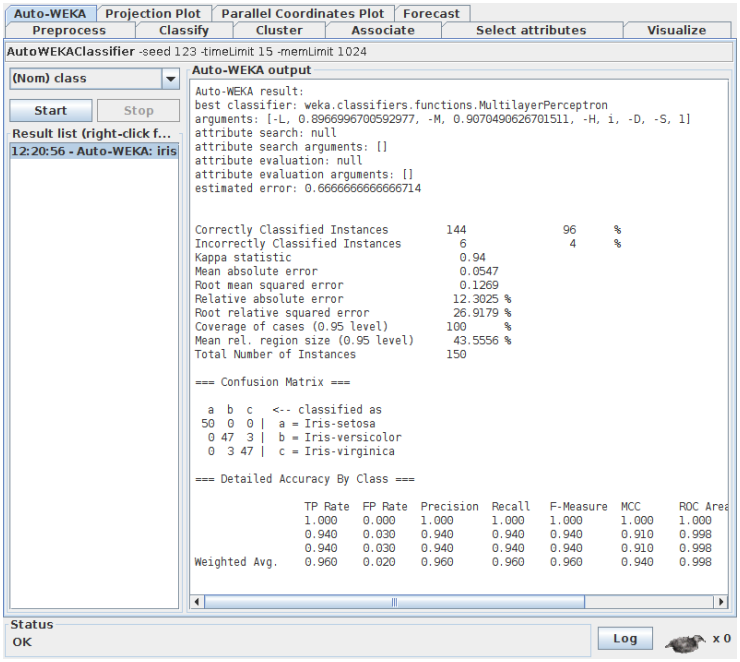

Figure 1: Auto-WEKA tab showing the output of a completed run.

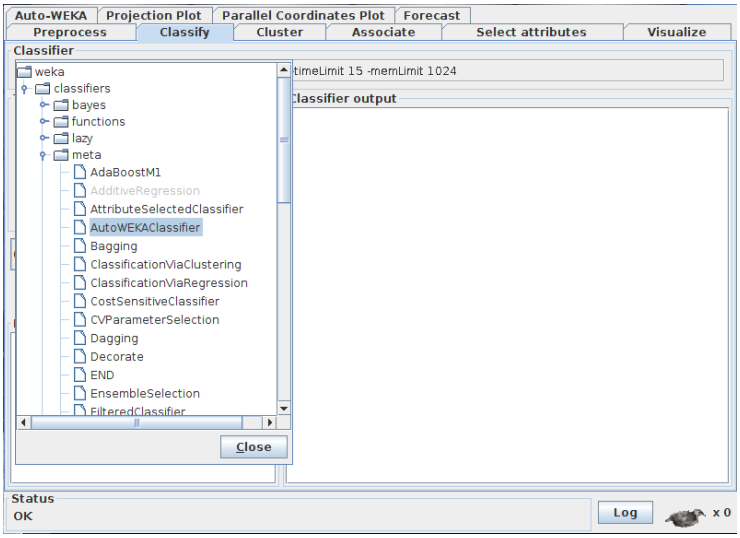

Figure 2: Location of the Auto-WEKA classifier in the list of classifiers.

While Auto-WEKA is running, it will provide the number of evaluated configurations and estimated error of the best configuration found so far in the status

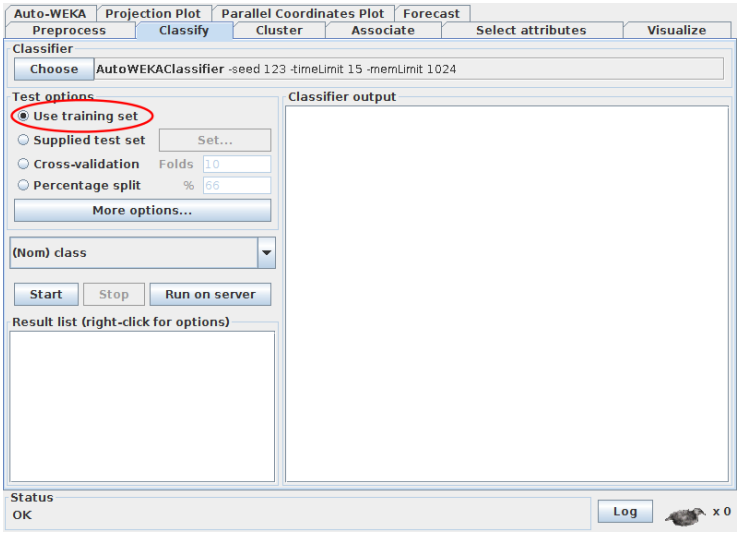

Figure 3: Recommended evaluation setting when using Auto-WEKA like a normal classifier.

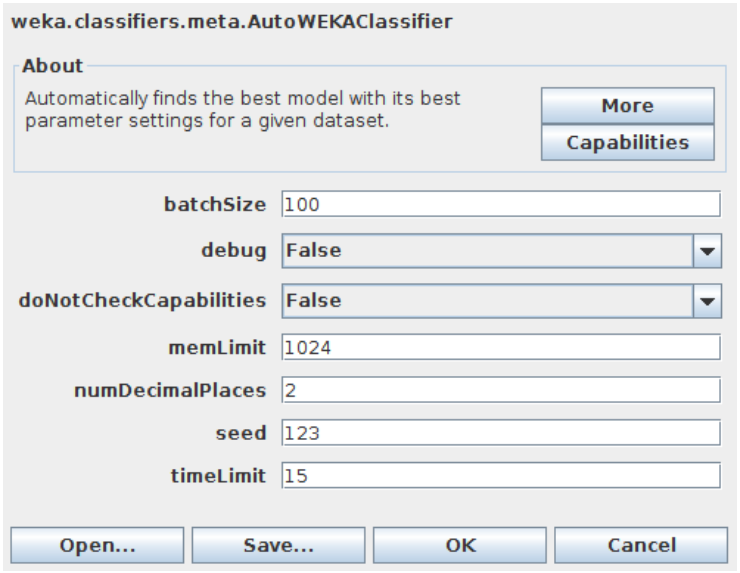

Figure 4: Auto-WEKA options.

bar.

Note that the time limit is *approximate* and Auto-WEKA may not take *exactly* as long as requested.

2.2 Running Experiments Using the CLI

Auto-WEKA can be run from the CLI like any other WEKA classifier, for example:

```
java -cp autoweka.jar weka.classifiers.meta.AutoWEKAClassifier \
    -t iris.arff -timeLimit 15 -no-cv
```