Appendix D

GARCON: Genetic Algorithm for Rectangular Cuts OptimizatioN

Typically HEP analysis has quite a few selection criteria (cuts) to optimise for example a significance of the "signal" over "background" events: transverse energy/momenta cuts, missing transverse energy, angular correlations, isolation and impact parameters, etc. In such cases simple scan over multi-dimensional cuts space (especially when done on top of a scan over theoretical predictions parameters space like for SUSY e.g.) leads to CPU time demand varying from days to many years... One of the alternative methods, which solves the issue is to employ a Genetic Algorithm (GA), see e.g. [819–821].

We wrote a code, GARCON [62], which automatically performs an optimisation and results stability verification effectively trying $\sim 10^{50}$ cut set parameters/values permutations for millions of input events in hours time. Examples of analyses are presented in the Physics TDR, see for example Sections 3.1, 8.4.1, 13.6, 13.7, 13.14 and recent papers [51, 316, 675, 676].

The GARCON program among many other features allows user:

- to select an optimisation function among known significance estimators, as well as to define user's own formula, which may be as simple as signal to background ratio, or a complicated one including different systematic uncertainties separately on different signal and background processes, different weights per event and so on;
- to define a precision of the optimisation;
- to restrict the optimisation using different kind of requirements, such us minimum number of signal/background events to survive after final cuts, variables/processes to be used for a particular optimisation run, number of optimisations inside one run to ensure that optimisation converges/finds not just a local maximum(s), but a global one as well (in case of a complicated phase space);
- to automatically verify results stability.

GARCON, GA-based programs in general exploit evolution-kind algorithms and uses evolution-like terms:

- Individual is a set of qualities, which are to be optimised in a particular environment or set of requirements. In HEP analysis case Individual is a set of lower and upper rectangular cut values for each of variables under study/optimization.
- Environment or set of requirements of evolutionary process in HEP analysis case is a Quality Function (QF) used for optimisation of individuals. The better QF

value the better is an Individual. Quality Function may be as simple as $S/\sqrt{B}$, where S is a number of signal events and B is a total number of background events after cuts, or almost of any degree of complexity, including systematic uncertainties on different backgrounds, etc.

- A given number of individuals constitute a Community, which is involved in evolution process.
- Each individual involved in the evolution: breeding with possibility of mutation of new individuals, death, etc. The higher is the QF of a particular individual, the more chances this individual has to participate in breeding of new individuals and the longer it lives (participates in more breeding cycles, etc.), thus improving community as a whole.
- Breeding in HEP analysis example is a producing of a new individual with qualities (set of min/max cut values) taken in a defined way from two "parent" individuals.
- Death of an individual happens, when it passes over an age limit for it's quality: the bigger it's quality, the more it lives.
- Cataclysmic Updates may happen in evolution after a long period of stagnation in evolution, at this time the whole community gets renewed and gets another chance to evolve to even better quality level. In HEP analysis case it corresponds to a chance to find another local and ultimately a global maximum in terms of quality function. Obviously, the more complicated phase space of cut variables is used the more chances exist that there are several local maximums in quality function optimisation.
- There are some other algorithms involved into GAs. For example mutation of a new individual. In this case newly "born" individual has not just qualities of its "parents", but also some variations, which in terms of HEP analysis example helps evolution to find a global maximum, with less chances to fall into a local one. There are also random creation mechanisms serving the same purpose.

There is nothing special involved in GARCON input preparation. One would need to prepare a set of arrays for each background and a signal process of cut variable values for optimisation. Similar to what is needed to have to perform a classical eye-balling cut optimisation.

In comparison to other automatised optimisation methods GARCON output is transparent to user: it just says what rectangular cut values are optimal and recommended in an analysis. Interpretation of these cut values is absolutely the same as with eye-balling cuts when one selects a set of rectangular cut values for each variable in a "classical" way by eye.

All-in-all it is a simple yet powerful ready-to-use tool with flexible and transparent optimisation and verification parameters setup. It is publicly available along with a paper on it [62] consisting of an example case study and user's manual.