In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA: Penn State.

Two methods for search and optimising cognitive model parameters

David Peebles (d.peebles@hud.ac.uk)
Department of Behavioural and Social Sciences, University of Huddersfield
Queensgate, Huddersfield, HD1 3DH, UK

Keywords: Parameter optimisation, differential evolution, high throughput computing, HTCondor, ACT-R.

Introduction

Searching for the best set of parameter values is a key component of cognitive modelling and one in which a great deal of uncertainty lies. Parameter search can be a slow, laborious process when done by hand, particularly when a model has several interacting parameters, and can be challenging when models are non-differentiable, non-continuous, non-linear, stochastic, or have many local optima.

There a several methods for searching parameter spaces for such models. Here I present two: *differential evolution* (DE) and using a *High Throughput Computing* (HTC) environment managed by *HTCondor*. The two methods are similar in that they both explore parameter spaces by generating populations of models, but there the similarity ends. Below I describe both, explain the circumstances where choosing one may be preferable over the other, and provide an example of each using a simple ACT-R model for the reader to investigate.

Differential evolution

Differential evolution is an evolutionary strategy for real numbers that has been used and refined extensively for multidimensional numerical optimisation since it was devised in the mid 1990s (Storn & Price, 1995, 1997). The main attractions of DE are its simplicity, wide applicability, relatively few control parameters (three: NP, the population size, F, a scale factor applied to the mutation process, and Cr, a constant that regulates the crossover process), and the accuracy, convergence rate and robustness of its performance. To use DE for optimising cognitive model parameters, the model is conceptualised as an objective function of the parameters being optimised that produces a single fitness value (e.g., R^2) to be maximised.

The DE algorithm

In common with many evolutionary algorithms, DE applies repeated cycles of *mutation*, *recombination*, and *selection* on an initial, randomly generated population of vectors to create a single vector that produces the best solution to a problem. The DE process is started by creating an NP sized population of real-valued vectors of D dimensions, one dimension for each of the model parameters to be optimised. The vectors are initialised with uniformly distributed random numbers within maximum and minimum bounds set for each dimension.

To create the next population of vectors, each vector **i** in the current population is selected in sequence, designated as the *target* vector, and subjected to a competitive process. The competition involves the three mutation, recombination, and selection steps described below.

Mutation Mutation randomises the search process, but unlike many other evolutionary strategies that mutate vectors by adding Gaussian noise, DE does so by computing the weighted difference between two vectors in the current population. This ensures that differences in the scale and sensitivity of different vector parameters are taken into account and that the search space is explored equally on all dimensions.

A mutated *donor* vector is created by randomly selecting three unique vectors, **j**, **k** and **l**, which are not equal to **i**, from the population and adding the difference between **j** and **k** (scaled by the F parameter) to **l**.

Recombination Once the donor vector has been created it is crossed with the target vector to create the *trial* vector. This recombination allows successful solutions from the previous generation to be incorporated into the trial vector.

Crossover is achieved by a series of Bernoulli trials which determine for each of $D-1$ dimensions which parent will donate its value. The process is moderated by the crossover rate parameter Cr (where $0 \leq Cr \leq 1.$). For each dimension, a uniformly distributed random number, x between 0 and 1 is generated and compared to Cr. If $x \leq Cr$, the donor vector's parameter is passed on to the trial vector, otherwise the parameter comes from the target vector. To ensure that the trial vector does not emerge identical to the target vector, one dimension is selected at random to inherit its value from the donor vector.

Selection The model is then run with the parameter values from the trial vector and if the resulting fitness value is better than or equal to that of the target vector, the trial vector replaces it in the next generation, or else the target vector is retained in the next generation. This process of mutation, recombination, and selection is carried out for each vector in the current population until the next population is created and the evolutionary process continues for a user-defined number of cycles. The vector with the highest fitness is recorded for each population and the winning vector in the final population is considered the best solution to the problem.

Setting DE parameters

The performance of the DE algorithm is quite sensitive to its three control parameters and numerous attempts have been made to determine the optimal values for various problems (e.g., Pedersen, 2010; Neri & Tirronen, 2010; Gämperle, Müller, & Koumoutsakos, 2002). For example in the mutation process the F constant scales all of the vector parameters

equally and determines the size of the distance between the target and trial vectors. Effective values for F are generally regarded to fall between 0.4 and 1.0 with a good initial value being 0.5 (Das & Suganthan, 2011; Storn & Price, 1995).

The crossover rate parameter Cr affects the search process by regulating the probability that noisy random values enter the trial vector (raising Cr increases the likelihood that dimension values will come from the donor vector). Although views differ, Cr values between 0.3 and 0.9 are generally considered reasonable for the majority of functions.

Recommendations for the optimum population size, NP are generally specified as a function of the number of vector parameters, D, and also vary but typically range between 3D and 10D (e.g., Storn & Price, 1995; Gämperle et al., 2002).

Research into DE is very active and a number of variants and adaptations have been developed (Neri & Tirronen, 2010). The standard version described here is still widely used and performs well on many problems.

High throughput computing and HTCondor

While DE is useful for optimising models with relatively few parameters or short run times on a single computer, if models are large, complex, or are simulating the behaviour of many participants, then the computational requirements may be such that this option becomes impractical. In these circumstances, an alternative is to search the parameter space by running a population of models over a computer network and one relatively accessible and increasingly popular way to do this is by using *HTCondor* (https://research.cs.wisc.edu/htcondor).

HTCondor is an open source, cross-platform software system for managing and scheduling computationally intensive tasks across computer networks developed over many years at the University of Wisconsin-Madison (Litzkow, Livny, & Mutka, 1988). It can be employed on dedicated server clusters or to schedule tasks over idle desktop computers on a network and it is widely used in universities and research institutions worldwide, including CERN, Fermilab, and NASA.

Using HTCondor for exploring parameter spaces for cognitive models can be achieved by submitting multiple versions of the model, each with a different set of randomly generated parameter values, analysing the returned outputs, and then iterating. The process for doing so is relatively straightforward. All that is required is the creation of a *submit description file* which specifies details about the job such as the executable to be run and upon which platform, the model files to be loaded by the executable, the command to start the program running, and the number of times to run the program. As each program may also use the standard streams, files must be defined that will substitute for stdin, stdout and stderr.

For example, the extract below is from a submit description file for a job to run 100 instances of an ACT-R model defined in the file *paired.lisp*. It specifies that only 64-bit Windows machines in the network should be used, that the executable is ACT-R for 64-bit Windows, and that the arguments to the executable are to load the model file, run it for 20 participants, and then quit. In addition, output, error, and log files are defined that will be created for each instance of the model and specifications made that both the executable and the model file should be transferred to each machine. Finally, the last command sets the job to run 100 instances of the model.

```
requirements = (OpSys == "WINNT61" && Arch == "INTEL") ||
               (OpSys == "WINDOWS" && Arch == "INTEL") ||
               (OpSys == "WINDOWS" && Arch == "X86_64"))

executable = actr-s-64.exe
arguments = "-l 'paired.lisp' -e '(collect-data 20)' -e '(quit)'"

transfer_executable = ALWAYS
transfer_input_files = paired.lisp

output = out.stdout.$(Cluster).$(Process)
error = out.err.$(Cluster).$(Process)
log = out.clog.$(Cluster).$(Process)

queue 100
```

When all of the model instances have been run, their outputs will be available in numbered output files which can then be collected together and analysed.

Example code

To enable further investigation of these methods, code to optimise an ACT-R model of paired associate learning taken from Unit 4 of the ACT-R tutorials (available from the ACT-R website), together with full instructions is available on GitHub. The repository for differential evolution can be found at https://github.com/peebz/actr-paired-de while that for running the model on HTCondor is available at https://github.com/peebz/actr-paired-htc.

References

Das, S., & Suganthan, P. N. (2011). Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, *15*(1), 4–31.

Gämperle, R., Müller, S. D., & Koumoutsakos, P. (2002). A parameter study for differential evolution. *Advances in intelligent systems, fuzzy systems, evolutionary computation*, *10*, 293–298.

Litzkow, M. J., Livny, M., & Mutka, M. W. (1988, June). Condor—A hunter of idle workstations. In *Proceedings of the 8th international conference on distributed computing systems* (pp. 104–111).

Neri, F., & Tirronen, V. (2010). Recent advances in differential evolution: A survey and experimental analysis. *Artificial Intelligence Review*, *33*(1-2), 61–106.

Pedersen, M. E. H. (2010). *Good parameters for differential evolution* (Tech. Rep. No. HL1002). www.hvass-labs.org: Hvass Laboratories.

Storn, R., & Price, K. (1995). *Differential evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces* (Tech. Rep. No. TR-95-012). Berkeley, CA: ICSI Berkeley.

Storn, R., & Price, K. (1997). Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341–359.