SPC Package ver 2.1.1 Documentation

Feb 14, 2000

The documentation includes: installation instructions, example runs, file formats and parameter definitions.

1 Installation and Running

1. Open the compressed tar file using `gzip -c -d SPC.tar.gz | tar xvf -` in the program's directory.
2. Compile the program by running `make`.
3. Prepare an ASCII data file (`xxxx.dat`) - see demo.dat.
4. Prepare an ASCII run file with all parameters (`xxxx.run`) - see demo.run.
5. Run the program by `SW xxxx.run` (log is dumped to the screen) or `SW xxxx.run >& xxxx.log` (save the output log and error messages).
6. Output files are created using the prefix defined in the param file (here `my_prefix`):
   - `my_prefix.param` (see 3.1).
   - `my_prefix.K??.edges or my_prefix.mst??.edges` depending if mstree options was used. The `??` stand for the K used in the K mutual neighbor step. (see 3.3).
   - `my_prefix.dg_01` a list of the sizes of the largest clusters in each temperature.
   - `my_prefix.dg_01.lab` a list of the cluster id. of each point in each temperature.
   - `my_prefix.mag` the magnetization and susceptability in each temperature.

2 Examples

There are two examples one with data as vectors in 2D and the other with only distances between the points.

2.1 3 concentric circles

4800 points in 2D representing 3 concentric circles taken from [1]. The example is in the directory `3Conc`. In the directory there are:

- The data file (`3Conc.dat`) - which includes the 4800 lines with 2 coordinates in each line.
- The run file (`3Conc.run`) - which includes the parameters for the algorithm.
- The output files (`orig3cout.*`) - which are out output files.

Inorder to run the example change to its directory and type: `./SW 3Conc.run`. The new output files are called `3cout.*`. A log of the run is displayed on the screen.

2.2 Clustering of images

90 by 90 matrix of distances between 90 images taken from [2]. The example is in the directory `ImageDist`. In the directory there are:

- The data file (`yoram.dat`) - which includes the 8100 matrix elements of the 90x90 distance matrix.
- The run file (`yoram.run`) - which includes the parameters for the algorithm.
- The output files (`origyout.*`) - which are out output files.

Inorder to run the example change to its directory and type: `./SW yoram.run`. The new output files are called `yout.*`. A log of the run is displayed on the screen.

3 File Formats

SPC uses a parameter file and can output several files. Following are descriptions of ths files.

3.1 Parameter file

This files contains the parameter values you want to set. Each line in this file should be in one of the following formats:

- `Par: Value` – set the parameter `Par` to the value `Value`:
- `Par|` – set the parameter `Par` without a value.
- `Par~` – unset the parameter `Par`.

In general, any word that ends with `~,| or :` is identified as a parameter name, and any other word is taken as a value if it follows :, and is ignored otherwise. For the parameter list see Section 3.1.

3.2 Data file

If the value of the parameter `Dimension` is positive than it contains the data point coordinates in a `Dimension`-dimensional space. Each row contains the coordinates of a data point, seperated by spaces. If `Dimension` is zero, if `DataIsMatrix` the file contains a distance matrix in ascii format, else each row describes the distance between two points. It has the format:

```
i1   i2   distance
```

where `i1` and `i2` are the indices of the points. Pairs the are not included in the file are cosidered to have infinite distance.

3.3 Edge file

Optional file that is used to force SPC to consider specific edges, in addition to the edges selected by the KNN and MST, or instead of them (If you choose not to use those features). Each edge is described by aline of the format:

```
i1   i2
```

4 Output files

All output file names has the format `OutFile.suffix`.

4.1 Parametes

Suffix: `.param`. Contains a list of all the parameters used by the program, and their values.

4.2 Magnetization and Susceptibility

$n, T, T/ < J >$,Magnetization, chi1, chi2 Magnetization= By N_{μ}^{max} - see our paper chi1 = variance of the magnetization (we use this) chi2 = variance of the average size of SW clusters

4.3 Labels

4.4 Averages

4.5 Timming

4.6 Correlations

5 Environment Parameters

To make the program simpler and clearer we use environment parameters, in which we keep data and flags to be use by different functions, without the need to pass them as arguments. This also allows an easy way to add more flags and options to the program wuthout littering the code too much. The parameters used in the programm so far are:

AverageInteraction
: the average interaction between neighbours.

ClustersReported
: how many cluster sizes are reported.

ClusterMinSizeReported
: minimal size of cluster to be reported. all clusters above or equal this size will be reported, even if their number exceeds the `ClustersReported` value.

CharDist
: characteristic distance between neighbours (used for calculating the interaction).

DataFile
: the file containing the coordinate of the point, or the distances matrix.

DataIsInteraction
: The values givenshould be cosidered as the interaction, and not as distances.

Dimensions
: dimension of the varctors describing the points. if 0, the data is expected to be a distances matrix.

DirectedGrowth
Report cluster obtained from the correlations by directed growth.

DataIsMatrix
the data file is organized as a distance matrix and not as a list.

EdgeFile
file containing edges to be added to the list of nearest neighbours.

FourPointCorr
Sample and report four point correlations.

ForceNN
Use this (false) number of nearest neighbours for calculating the interaction.

ForceRandomSeed
Random seed to start the program with, if we want this run to be identicle to a previous one. If not set, the seed is taken as the clock value.

ForceChD
Use this (false) characteristic distance for calculating the interaction.

InfMetric
Use infinity metric to calculate distances.

KNearestNeighbours
maximal number of nearest neighbours (used in the knn algorithm).

Lambda
λ value to be used if `UseZ` is set.

MinTemp
The lowest temperature to use..

MaxTemp
The maximal temp. to use.

MSTree
add the edges of the minimal spanning tree.

NumberOfPoints
number of points.

NumberOfEdges
the total number of edges.

NearestNeighbrs
average number of nearest neighbours (used for calculating the interaction).

OutFile
prefix for the output files.

PottsSpins
The number of different spin values.

PrevTempFile
prefix for former output files, used to obtain initial states for the threshold and directed growth calculations.

RandomSeed
The random seed with which the program has started.

RandomInitialConfig
By default, the simulations start with all spins set to 0. Whenb this option is set, the initial configuration will be random.

SusceptColors
number of susciptibilities from the susceptibility vector to report.

SaveAverages
sample and report data on Swensen-Wang averages.

SaveSuscept
sample and report magnetization and susceptibility data.

SWCycles
number Swensen-Wang sweeps.

SWFraction
the fraction SW sweeps for which averages are calculated. The first (1-SWfract)*cyc sweeps are discarded.

StopRunAtBreak
Stop at the first break of the system into clusters.

StopRunClusterSize
Stop when the largest cluster has this size.

Threshold
Report cluster obtained from the correlations by thresholding.

ThresholdTheta
The threshold on correlations between neighbours, above which they are assumed to belong to the same cluster.

ThresholdMin

ThresholdMax

ThresholdStep
: If those variables are set, the programs report the results obtained for different thresholds, taken from min to max values with the step indicated.

TempStep
: the steps in temperature from min to max values in which the simulation should be ran.

TempStepMul
: The temperature update is $T(t+1) = Tstep * T(t)$.

Timing
: report the time required by the programm for each step.

UseZ
: Scale the interaction according to a given λ.

WriteLabels
: Write the label of the cluster each spin belongs to.

WriteCorFile
: Write the correlations between each pair of neighbours.

WriteFPSum
: Write four point correlations sum.

References

[1] M. Blatt, S. Wiseman and E. Domany, *Neural Computation* **9** 1805 (1997).

[2] E. Domany, M. Blatt, Y. Gdalyahu and D. Weinshall, *Comp. Phys. Commm* **5** 121 (1999).