

User Manual for Code for Learning Exponential Models over Rankings

Bhushan Mandhani¹ and Marina Meila²

¹ Dept. of Computer Science

² Dept. of Statistics

University of Washington

Seattle, USA 98195

This document contains usage instructions for the code that we have made available for learning the Mallows [1] and Generalized Mallows [2] models. The code employs the search method proposed in [3], and the search heuristic is computed as described in [4]. It also includes implementations of the greedy method of [5] and the local search method of [6].

The code is in Java version 1.5. The `ExptsManager` class contains the “main” method. It can be invoked using the following command.

```
java -Xms100M -Xmx1500M -classpath pathToDirectory gmm.ExptsManager  
parametersFile.txt resultsFile.txt dumpFile.txt
```

The above command specifies 100 MB as the starting size of the JVM (Java Virtual Machine) and 1500 MB as the maximum allowed size. Set these values according to the total memory available on your machine. We strongly recommend that you give the JVM as much memory as is available. However, if you are in a hurry, one way of getting the code to run fast at the expense of obtaining inferior results is to reduce the maximum allowed size of the JVM. The process will then run out of memory sooner, and revert to beam search from A* search sooner.

The `ExptsManager` class requires three command-line arguments: the file containing the input parameters, the file to which the results should be written, and the file to which a detailed dump of all results on a per iteration basis should be written. In most cases, we expect the reader to not need to bother with this detailed dump.

The parameters file needs to specify values for the following parameters. Each line of this file is of the form `parameter=value`.

- `n`: the number of items being ranked.
- `inputType`: specifies how the input Q matrix is obtained. It can take the following three values.
 1. `model` indicates Q is obtained from rankings sampled from an instance of the Mallows or the GM model.
 2. `random` indicates each Q_{ij} with $i < j$ is set to a random value in $(0, 1)$. The values Q_{ij} with $i > j$ are thus automatically determined.
 3. `filename.txt` specifies the name of the file containing the rankings from which Q is obtained. Each line of the file contains a ranking with the items separated by whitespace.

- **N**: the number of rankings to sample from the input model.
- **theta**: the value θ_1 for the input model.
- **decay**: the value **linear** indicates the θ_j parameters for the input model decrease linearly. Otherwise, they decay exponentially.
- **GMM**: **true** and **false** respectively specify whether to learn a GM model or Mallows model. Also specifies the same thing about the input model.
- **beamwidth**: the beam width to use if the search has to revert to beam search. We used a value of 1000 in our experiments.
- **iterations**: the number of times the experiment is to be repeated. The results file contains the mean values obtained over the specified number of iterations. We did 25 iterations in our experiments.

References

1. Mallows, C.L.: Non null ranking models. *Biometrika* (1957)
2. Fligner, M.A., Verducci, J.S.: Distance based ranking models. *Journal of the Royal Statistical Society* (1986)
3. Meila, M., Phadnis, K., Patterson, A., Bilmes, J.: Consensus ranking under the exponential model. In: *UAI*. (2007)
4. Mandhani, B., Meila, M.: Better search for generalized rank aggregation. In: *Submitted*. (2008)
5. Cohen, W.W., Schapire, R.E., Singer, Y.: Learning to order things. *J. Artif. Intell. Res. (JAIR)* **10** (1999) 243–270
6. Fligner, M.A., Verducci, J.S.: Multistage ranking models. *Journal of the American Statistical Association* **83** (1988)