MoAn: User documentation

Eivind Valen

November 21, 2009

1   Introduction

MoAn is a discriminative motif finder for DNA sequences. It works by contrasting a positive and a negative set trying to establish which motifs that best discriminates between the two sets. One of its main strength the capability to handle very large negative sets (typically about 1000 sequences) which gives a representative sample of the promoter architecture in an organism.

2   Standard usage

The basic usage of MoAn is:

```
moan [OPTIONS] POSITIVE NEGATIVE
```

where POSITIVE is a FASTA sequence set and NEGATIVE likewise. To get good results it is important that the negative set is significantly larger than the positive set. The basic options are:

-D Double stranded search.

-c Search co-occurrence of two matrices

-i N Number of iterations. N is an integer. 30 million is default.

-l N Sets the verbosity. N is an integer from 0 to 4. 1 is default.

-R N,M Range of the size of the WM, from N to M. E.g.: 7,20 (default: 5,15)

-w N Optimize N WMs (default is 1). -w 2 is not the same as co-occurrence, which specifically avoids overlaps. Note also that the WMs are optimized collectively rather than as seperate chains. This will be fixed in future versions.

The default run optimizes one weight matrix, searches only a single strand and iterates 30 million steps. The iteration time is very conservative and in many cases the method will identify the correct motif long before this.

Unfortunately, as mentioned in the paper optimization of the likelihood is very hard and you should always run the program multiple times if you can. This is the main focus of our work for the next release.

2.1 Co-occurrence

Using the "-c" option enables co-occurrence of two matrices. This is in general a much harder optimization task and requires many more iterations than a single run (the 30 million iteration default is probably ok). Note, that it is therefore a good idea to always run single occurrence runs first to see if this produces any result.

3 Advanced options

3.1 Setting the prior

-P XX,YY Sets the prior probabilities on occurrences for the positive set (v^1). XX is the prior on no-occurrences (v^1_0) and YY is the prior on one occurrence (v^1_1). These are expressed as percentages so a valid argument would be: "-P 01,99" yielding 1% prior on no occurrence and 99 % on a single occurrence (this is the default). If co-occurrence is considered, the prior on the co-occurrence $(v^1_{1,2})$ is calculated as the remainder of 100% after XX and YY has been subtracted.

-N XX,YY Same priors for the negative set.

Note that you should never set the prior on single occurrence higher in the negative set than in the positive set even when considering co-occurrence. This will lead to unresolvable optimization problems.

4 Reference

- Valen, E., Winther, O., Sandelin, A. and Krogh, A.
  Discovery of regulatory elements is improved by a discriminatory approach
  PLoS Computational Biology, 5(11) 2009

5 Contact

For any questions or requests contact **eivind@binf.ku.dk**. Also please notify me if you have successfully used your program in a published work or if you are experiencing any problems with compiling or running.