Hyperbolic Matrix Factorization

Support:

For comments, suggestions and help using the program, contact Prof. Aleksandar Poleksic at poleksic.cs.uni.edu.

Data sets:

The four data sets (*Nr*, *Gpcr*, *Ion*, and *Enz*) used in our benchmark [1] are available in the "data" directory. The protein-drug association data is stored in flat files. Each file consists of three columns (protein, drug, and association (0/1), respectively). The protein and drug pairwise similarity matrices are provided in the "data" directory.

Code:

The source code, implemented in MATLAB, is available in the "code" directory. The hyperbolic variants of the three methods analyzed in this study (NRLMF, DNILMF, and NGN) can be found in the respective directories (*h_nrlmf*, *h_dnilmf*, and *h_ngn*). General hyperbolic routines (hyperbolic inner product, hyperbolic distance, exponential map, projection, hyperbolic gradient descent, etc.) are stored in the subdirectory *hyperbolic_routines*. To generate benchmarking results, execute *RunBenchmark.m* routine. In an example setting, to benchmark the hyperbolic variant of the logistic matrix factorization algorithm (as implemented, for instance, in [2], [3], or [4]) on the *Nr* dataset in CVS1 (OFFTARGET) setting using 3 rounds of 5-fold CV, run the following command:

```
nohup matlab -nodisplay -nosplash -nodesktop -r "RunBenchmark('HNRLMF','Nr','OFFTARGET',3,3);exit"
```

Note that, for efficiency, our MATLAB implementation takes advantage of the MATLAB *parfor* loop (different iterations are run in parallel on multiple workers). Thus, in absence of any code modification, our program requires *MATLAB Parallel Computing Toolbox* (https://www.mathworks.com/products/parallel-computing.html).

References:

[1] Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, *24*(13), i232-i240.

[2] Johnson, C. C. (2014). Logistic matrix factorization for implicit feedback data. Advances in Neural Information Processing Systems, 27(78), 1-9.

[3] Liu, Y., Wu, M., Miao, C., Zhao, P., & Li, X. L. (2016). Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLoS computational biology, 12(2), e1004760.

[4] Lim, H., Gray, P., Xie, L., & Poleksic, A. (2016). Improved genome-scale multi-target virtual screening via a novel collaborative filtering approach to cold-start problem. Scientific reports, 6(1), 1-11.