eccCL: an optimized GPU implementation of ensemble classifier chains

Mona Riemenschneider, Alexander Herbst, Ari Rasch, Sergei Gorlatch and Dominik Heider

Department of Bioinformatics, Straubing Center of Science, 94315 Straubing, Germany and Institute of Computer Science, University of Muenster, 48149 Muenster, Germany

m.riemenschneider@wz-straubing.de

Multi-label classification (MLC) has recently gained great attention in diverse fields of research, for example in text mining and biomedical applications [HSCH13, RSN$^+$16] as protein function prediction or drug resistance testing. In this context, the MLC concept of classifier chaining has been shown to improve prediction performance in comparison to single-label classification, especially when applied as ensemble classifier chains [RPHF11]. However, these techniques lack computational efficiency when applied on high-dimensional data sets. By adapting algorithms for the use of graphical processing units (GPUs), computational efficiency can be greatly improved due to parallelization of computations [OSGH14]. Here, we provide a parallelized and optimized GPU implementation of classifier chains and ensemble classifier chains that is able to process several millions of instances in adequate computing time. Additionally to the OpenCL implementation, we provide an R-Package (eccCL) with an easy to use R-interface for parallelized GPU usage. The software is available at http://www.heiderlab.de.

References

[HSCH13] D Heider, R Senge, W Cheng, and E Hüllermeier. Multilabel classification for exploiting cross-resistance information in HIV-1 drug resistance prediction. *Bioinformatics*, 29, 2013.

[OSGH14] M. Olejnik, M. Steuwer, S. Gorlatch, and D. Heider. gCUP: rapid GPU-based HIV-1 co-receptor usage prediction for next-generation sequencing. *Bioinformatics*, 30(22):3272–3273, 2014.

[RPHF11] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. *Machine Learning*, 85(3):333–359, jun 2011.

[RSN$^+$16] M Riemenschneider, R Senge, U Neumann, E Hüllermeier, and D Heider. Exploiting HIV-1 protease and reverse transcriptase cross-resistance information for improved drug resistance prediction by means of multi-label classification. *BioData Mining*, 9(1):1–6, 2016.