PRIMPing Boolean Matrix Factorization by Proximal Alternating Linearized Minimization

Sibylle Hess, Nico Piatkowski, and Katharina Morik

TU Dortmund, Computer Science, LS 8, 44221 Dortmund, Germany

Abstract. We propose a novel Boolean matrix factorization algorithm, based on recent results from optimization theory. We demonstrate the superior robustness of the new approach in the presence of several kinds of noise and the interpretability on synthetic and real-world data.

Keywords: Boolean Matrix Factorization, Minimum Description Length, Proximal Minimization, Nonconvex-Nonsmooth Minimization

Given the task to explore binary data, Boolean Matrix Factorization (BMF) is a method of choice. BMF yields a simultaneous clustering of rows and columns of the data matrix into binary, thus interpretable, cluster representatives. Furthermore, state-of-the art methods automatically estimate the number of prevalent clusters through the application of the minimum description length principle [2].

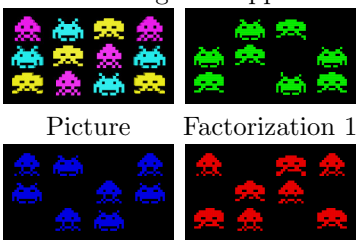

Fig. 1: Reconstructions of the image top left by three outer products returned by PRIMP. Best viewed in color.

Unfortunately, existing algorithms are greedy and rely on heuristics to solve the **NP**-hard problem of BMF. We propose with the procedure PRIMP to apply the optimization scheme PALM to a real-valued relaxation of the objective. PALM enables the minimization of the generally nonconvex description length and the nonsmooth penalization of non-binary values under convergence guarantees. A rounding procedure rounds the result to binary values and decides over the number of returned clusters. For more information, we refer to [1].

Acknowledgments SFB 876, projects A1 and C1.

References

1. Sibylle Hess, Katharina Morik, and Nico Piatkowski. 2017. The PRIMPING routine–Tiling through proximal alternating linearized minimization. Data Min. Knowl. Discov. 31, 4 (July 2017), 1090-1131.
2. Pauli Miettinen and Jilles Vreeken. 2014. MDL4BMF: Minimum Description Length for Boolean Matrix Factorization. ACM Trans. Knowl. Discov. Data 8, 4, Article 18 (October 2014).

In: M. Leyer (Ed.): Proceedings of the LWDA 2017 Workshops: KDML, FGWM, IR, and FGDB.
Rostock, Germany, 11.-13. September 2017, published at http://ceur-ws.org