Application of tropical semiring for matrix factorization

Amra Omanović, Polona Oblak, Tomaž Curk
University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, 1000 Ljubljana, Slovenia
amra.omanovic@fri.uni-lj.si, polona.oblak@fri.uni-lj.si, tomaz.curk@fri.uni-lj.si

Abstract
Matrix factorization methods employ standard linear algebra, i.e. linear models, for recommender systems. With the introduction of the tropical semiring, we can achieve non-linearity. We review algorithms that use the tropical semiring for matrix factorization and provide their strengths and limitations. We show that the tropical matrix factorization yields better results than non-negative matrix factorization for the synthetic data created by the underlying process of the tropical semiring.

Keywords: Data embedding, data mining, matrix factorization, subtropical semiring, tropical semiring.

Izvleček
Metode matrične faktorizacije uporabljajo za priporočilne sisteme standardno linearno algebro, torej linearne modele. Z zamenjavo operacij in z uvedbo tropskega polkolobarja lahko dodamo metodam komponento nelinearnosti. V članku pregledamo algoritme, ki za faktorizacijo matrike uporabljajo tropski polkolobar, in podamo njihove prednosti in omejitve. Pokažemo, da tropska matrična faktorizacija daje boljše rezultate kot nenegativna matrična faktorizacija na sintetičnih podatkih, ustvarjenih z množenjem matrik v tropskem polkolobarju.

Ključne besede: Vložitev podatkov, podatkovno rudarjenje, matrična faktorizacija, subtropski polkolobar, tropski polkolobar

1 INTRODUCTION

Data mining is one of the main challenges in computer science. There is a need to develop methods to embed data into a lower-dimensional latent space, which may help with various machine learning tasks. A data embedding model, such as matrix factorization (MF), gives us a more compact representation of the data and simultaneously finds a latent structure. MF algorithms (*e.g.*, non-negative matrix factorization (NMF) [Lee and Seung, 1999]) decompose the original matrix into a product of a base matrix and a coefficient matrix of lower dimensions. Most of machine learning methods for data embedding, *e.g.*, [Lee and Seung, 1999, Žitnik and Zupan, 2015, Zhang et al., 2007, Laurberg et al., 2008], use stan- dard linear algebra.

Recently, several authors considered substituting the standard linear algebra with other semir- ing operations, *e.g.*, [Karaev and Miettinen, 2016a, Karaev et al., 2018, Karaev and Miettinen, 2016b, Karaev and Miettinen, 2019]. In this paper we review some algorithms that use alternative nonstandard op- erations for matrix factorization and provide their strengths, limitations and potential of discovering interesting patterns. In our work, we are motivated by the question from [Karaev and Miettinen, 2016a], asking if a tropical matrix factorization can be used except for data analysis, also in other data mining and machine learning tasks, such as matrix completion. We expect that for the data that is not normally distributed and may contain a lot of extreme values using tropical semiring should give better results than MF methods that use standard operations of addition and multiplication.

Standard MF methods belong to the class of linear models that are unable to model complex relations. With the tropical semiring, we can introduce the non-linearity using the maximum operator. Another motivation for using tropical semiring is the work of Zhang *et al.* [Zhang et al., 2018]. They showed that linear regions of feedforward neural networks with rectified linear unit activation correspond to vertices of polytopes associated with tropical rational functions. Therefore, to understand specific neural networks, we need to understand relevant tropical geometry. Since the goal is not just to model the data, but also to understand the underlying mechanisms, the matrix factorization methods that use tropical semiring can give us a more straightforward interpretation than neural networks.

We split the remainder of the paper into the following sections. Sections 2 and 3 describe tropical semiring and related work, followed by results in Section 4. We conclude our paper in Section 5.

2 TROPICAL SEMIRING

The (max, +) semiring or tropical semiring $\mathbb{R}_{max}$, is the set $\mathbb{R} \cup \{-\infty\}$, equipped with max as addition ($\oplus$), and + as multiplication ($\otimes$). For example, $2 \oplus 3 = 3$ and $1 \otimes 1 = 2$. On the other hand, in the subtropical semiring or (max, $\times$) semiring, defined on the same set $\mathbb{R} \cup \{-\infty\}$, addition (max) is defined as in the tropical semiring, but the multiplication is the standard multiplication ($\times$). By taking the logarithm of the subtropical semiring, we obtain the tropical semiring, thus these two semirings are isomorphic.

Let $\mathbb{R}_{max}^{m \times n}$ define the set of all $m \times n$ matrices over tropical semiring. For $A \in \mathbb{R}_{max}^{m \times n}$ R*m*×*n* we denote by a_{ij} the entry in the i-th row and the j-th column of matrix A. We define the sum of matrices $A = [a_{ij}]$, $B = [b_{ij}] \in \mathbb{R}_{max}^{m \times n}$ as

$$(A \oplus B)_{ij} = a_{ij} \oplus b_{ij} = \max \{a_{ij}, b_{ij}\},$$

$i = 1, \dots, m$, $j = 1, \dots, n$, and the product of matrices

$$A = [a_{ij}] \in \mathbb{R}_{max}^{m \times p},\ B = [b_{kl}] \in \mathbb{R}_{max}^{p \times n} \quad \text{as}$$

$$(A \otimes B)_{ij} = \bigoplus_{k=1}^{p} a_{ik} \otimes b_{kj} = \max_{1 \le k \le p} \{a_{ik} + b_{kj}\},$$

$i = 1, \dots, m$, $j = 1, \dots, n$.

Matrix factorization over a tropical semiring is a decomposition of a form $A = U \otimes V$, where $A \in \mathbb{R}_{max}^{m \times n}$, $U \in \mathbb{R}_{max}^{m \times r}$, $V \in \mathbb{R}_{max}^{r \times n}$ and $r \in \mathbb{N}_0$. For small values of r such decomposition may not exist. A problem of tropical matrix factorization is thus stated as follows: given a matrix $A \in \mathbb{R}_{max}^{m \times n}$ and $r \in \mathbb{N}_0$, find $U \in \mathbb{R}_{max}^{m \times r}$ and $V \in \mathbb{R}_{max}^{r \times n}$ such that

$$A \cong U \otimes V. \tag{1}$$

Similarly, we define a subtropical matrix factorization. Note that the factorization in tropical semiring give different results and works with different methods than the factorization in subtropical semiring.

3 RELATED WORK

The most common examples of matrix factorization are the singular value decomposition (SVD) (see examples in [Golub and Reinsch, 1971]) and the non-negative matrix factorization (NMF) [Lee and Seung, 1999], where the factorization is restricted to matrices with non-negative entries. This non-negativity in resulting factor matrices U and V allows interpretation of the results. Binary matrix factorization (BMF) [Zhang et al., 2007, Zhang et al., 2010] is a variant rooted from NMF where factor matrices are binary, while probabilistic nonnegative matrix factorization (PMF) [Laurberg et al., 2008, Gaussier and Goutte, 2005] approximates data as samples from a multinomial distribution.

The Cancer algorithm [Karaev and Miettinen, 2016a] works with continuous data, performing subtropical/tropical matrix factorization (SMF) on the input matrix, and returning two factorized matrices. The algorithm's two key ideas are: iteratively updating the rank-1 factors one-by-one and approximating the reconstruction error with a polynomial of low-degree. Latitude algorithm [Karaev et al., 2018] combines NMF and SMF, where factors are interpreted as NMF or SMF features or as mixtures of both. Unfortunately, neither Can- cer not Latitude do not guarantee the convergence of the algorithms. Also, the authors [Weston et al., 2013] used subtropical semiring as part of a recommender system, which can be considered as a special kind of neural network.

De Schutter & De Moor introduced in 1997 a heuristic algorithm [De Schutter and De Moor, 1997] to compute factorization of a matrix in the tropical semiring, which we denote as Tropical Matrix Factorization (TMF). They use it to determine the minimal system order of a discrete event system (DES). In the last decades, there has been an increase in this research area, and DES is modeled as a max-plus-linear (MPL) system.

To implement TMF we need to know how to solve tropical linear systems. A tropical linear system is not solvable in general. For $A = [a_{ij}] \in \mathbb{R}_{max}^{m \times n}$ and $c = [c_k] \in \mathbb{R}_{max}^{m}$ we call the solutions $x \in \mathbb{R}_{max}^{n}$ of the inequality $A \otimes x \leqslant c$ the subsolutions of the linear system $A \otimes x = c$. The greatest subsolution $x = [x_1\ x_2 \dots x_n]^T$ of $A \otimes x \leqslant c$ can be computed by

$$x_i = \max_{1 \le j \le m} (c_j - a_{ji}),$$

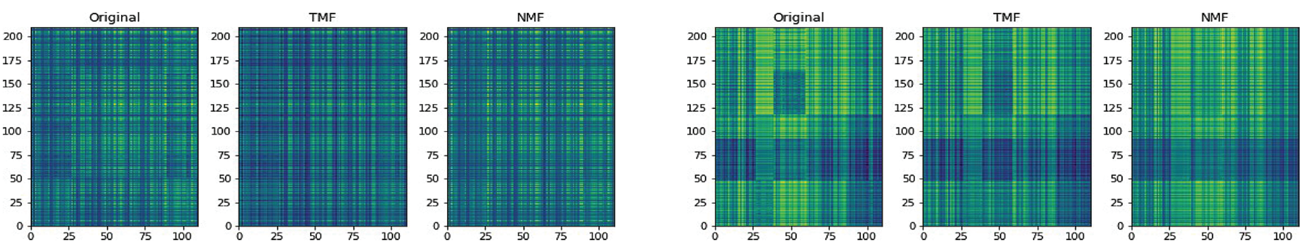
for $i = 1, 2, \ldots, n$, where symbol »−« denotes the standard subtraction in $\mathbb{R}$ [Gaubert and Plus, 1997].

TMF starts with an initial guess for the matrix U in (1), denoted by U_0, and then computes V as the greatest subsolution X of the equation $U_0 \otimes X = A$. Then authors use the iterative procedure by selecting and adapting an entry of U or V and recomputing it aLs the greatest subsolutions of $Y \otimes V = A$ and $U \otimes X = A$, respectively. The b-norm defined as $\|A_b\| = \sum_{i,j} |a_{ij}|$ is used as the objective function to get a good approximation of the input data.

In contrast to Cancer and Latitude, TMF update rules gradually reduce the approximation error and thus TMF algorithm is convergent. However, none of the existing tropical and subtropical algorithms Cancer and TMF, as defined, cannot be used for prediction tasks in data mining problems. Note that in TMF method there is no non-negativity constraint compared to the NMF and its variants. However, a weakness of TMF compared to NMF is in its computational efficiency.

4 RESULTS

We compare TMF and NMF on synthetic data created as a product of two non-negative random matrices. The objective of synthetic experiments is to show that the TMF can identify the (max, +) structure when it exists. Therefore, we construct two synthetic matrices: $D_s \in \mathbb{R}^{210 \times 110}$ as the standard product $(+, \times)$ of two random matrices of sizes 210×2 and 2×110; and $D_t \in \mathbb{R}^{210 \times 110}$ as the tropical product (max, +) of the same two matrices.



(a) Original synthetic matrix D_s and its two approxima- tions. Correlation between TMF and NMF approximated matrices and matrix D_s is equal to 0.983 and 0.990, re- spectively.

(b) Original synthetic matrix D_t and its two approxima- tions. Correlation between TMF and NMF approximated matrices and matrix D_t is equal to 0.974 and 0.966, re- spectively.

Figure 1: **Original and approximated matrices of rank 3, obtained by TMF and NMF.**

As expected, NMF reconstructs the matrix D_s better as TMF, see Figure 1a. Results on Figure 1b show that NMF cannot successfully recover the patterns when dealing with specific synthetic data. Moreover, for matrix D_t TMF returns a better approximation as NMF.

5 CONCLUSION

Standard matrix factorization methods perform learning tasks over matrices equipped with addition and multiplication. The constructed models are linear and thus unable to model complex, non-linear relations. This can be addressed by introducing the tropical semiring with (max, +) operations.

To the best of our knowledge, we are the first to implement and apply TMF in data analysis. We showed that TMF gives better results than NMF, when the data is created by an underlying process of (max, +) semiring.

In our future work, we plan to adapt TMF to be able to predict missing values and test methods on real data. Because the resulting structure can be simpler to interpret than with standard linear algebra, we believe that future research will show that semirings are useful in many scenarios.

REFERENCES

[1] De Schutter, B. and De Moor, B. (1997). Matrix factorization and minimal state space realization in the max-plus algebra. In *Proceedings of the 1997 American Control Conference (Cat. No. 97CH36041)*, volume 5, pages 3136–3140. IEEE.

[2] Gaubert, S. and Plus, M. (1997). Methods and applications of (max,+) linear algebra. In *Annual symposium on theoretical aspects of computer science*, pages 261–282. Springer.

[3] Gaussier, E. and Goutte, C. (2005). Relation between plsa and nmf and implications. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 601–602. ACM.

[4] Golub, G. H. and Reinsch, C. (1971). Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer.

[5] Karaev, S., Hook, J., and Miettinen, P. (2018). Latitude: A model for mixed linear-tropical matrix factorization. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 360–368. SIAM.

[6] Karaev, S. and Miettinen, P. (2016a). Cancer: Another algorithm for subtropical matrix factorization. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 576–592. Springer.

[7] Karaev, S. and Miettinen, P. (2016b). Capricorn: An algorithm for subtropical matrix factorization. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 702–710. SIAM.

[8] Karaev, S. and Miettinen, P. (2019). Algorithms for approximate subtropical matrix factorization. *Data Mining and Knowledge Discovery*, 33(2):526–576.

[9] Laurberg, H., Christensen, M. G., Plumbley, M. D., Hansen, L. K., and Jensen, S. H. (2008). Theorems on positive data: On the uniqueness of nmf. *Computational intelligence and neuroscience*, 2008.

[10] Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788.

[11] Weston, J., Weiss, R. J., and Yee, H. (2013). Nonlinear latent factorization by embedding multiple user interests. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 65–68.

[12] Zhang, L., Naitzat, G., and Lim, L.-H. (2018). Tropical geometry of deep neural networks. *arXiv preprint arXiv:1805.07091*.

[13] Zhang, Z., Li, T., Ding, C., and Zhang, X. (2007). Binary matrix factorization with applications. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 391–400. IEEE.

[14] Zhang, Z.-Y., Li, T., Ding, C., Ren, X.-W., and Zhang, X.-S. (2010). Binary matrix factorization for analyzing gene expression data. *Data Mining and Knowledge Discovery*, 20(1):28.

[15] Žitnik, M. and Zupan, B. (2015). Data fusion by matrix factorization. *IEEE transac- tions on pattern analysis and machine intelligence*, 37(1):41–53.

▪

Amra Omanović received a master's degree from the Faculty of Computer and Information Science, University of Ljubljana, in 2018. She is a junior researcher at the Faculty of Computer and Information Science, University of Ljubljana. Her research is focused on the application of linear algebra over semirings in data embedding and fusion methods.

▪

Polona Oblak received the doctoral degree from the Faculty of Mathematics and Physics, University of Ljubljana, in 2008. She is an associate professor at the Faculty of Computer and Information Science, University of Ljubljana. Her research interests include inverse eigenvalue problems for graphs, linear algebra over semirings and combinatorial matrix theory

▪

Tomaž Curk received a doctoral degree from the Faculty of Computer and Information Science, University of Ljubljana, in 2007. He is an assistant professor and serves as a vice-dean for research at the Faculty of Computer and Information Science, University of Ljubljana. His research is focused on the application of machine learning and data integration methods in bioinformatics.