A Model For Mixed Linear-Tropical Matrix Factorization

James Hook, Sanjar Karaev, Pauli Miettinen

University of Birmingham: 18th June 2018

Low-Rank Approximate Factorization

Given a matrix $A \in \mathbb{R}^{n \times m}$, an approximate factorization of rank k is a pair $B \in \mathbb{R}^{n \times k}$ and $C \in \mathbb{R}^{k \times m}$, such that

$$A \approx BC.$$

Such approximate factorizations are used throughout applied mathematics in...

- Compression
- Visualization/interpretation
- Matrix completion/prediction

Huge number of variations

- Constrains on factor matrices e.g. orthogonal, triangular, non-negative...
- Measure of closeness e.g. Frobenius norm, KL divergence...
- What about the matrix-matrix product itself?

Tropical Semirings

Tropical algebra concerns any semiring whose 'addition' operation is max or min.

E.g. the min-plus semiring $\mathbb{R}_{\min +} = [\mathbb{R} \cup \{\infty\}, \oplus, \otimes]$, where

$$a \oplus b = \min\{a, b\}, \quad a \otimes b = a + b, \quad \forall\ a, b \in \mathbb{R}_{\min +}.$$

Min-plus matrix multiplication is defined in analogy to the classical case. For $A \in \mathbb{R}^{n\times m}_{\min +}$ and $B \in \mathbb{R}^{m\times d}_{\min +}$ we have $A \otimes B \in \mathbb{R}^{n\times d}_{\min +}$, with

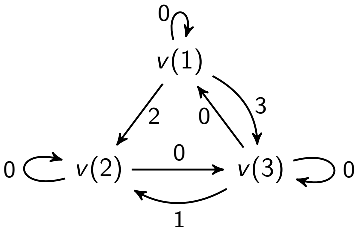
$$(A \otimes B)_{ij} = \bigoplus_{k=1}^{m} a_{ik} \otimes b_{kj} = \min_{k=1}^{m}(a_{ik} + b_{kj}).$$

For example

$$\begin{bmatrix} 0 & 2 & 3 \\ \infty & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 2 & 3 \\ \infty & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Paths through graphs viewpoint

$$\begin{bmatrix} 0 & 2 & 3 \\ \infty & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 2 & 3 \\ \infty & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$



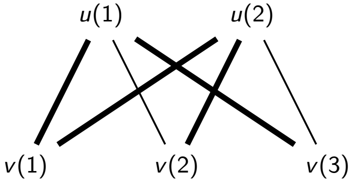
- For $A \in \mathbb{R}^{n\times n}_{\min +}$, precedence graph $\Gamma(A)$.

Proposition

$\left(A^{\otimes \ell}\right)_{ij}$ = *the weight of the minimally weighted path of length* ℓ*, through* $\Gamma(A)$*, from* $v(i)$ *to* $v(j)$*.*

Paths through graphs viewpoint

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ \cdot & 0 & 1 \\ \cdot & \cdot & 0 \end{bmatrix}$$



- For $A \in \mathbb{R}^{n \times d}_{\min +}$, precedence bipartite graph $\mathcal{B}(A)$.

Proposition

$\left(A \otimes A^T\right)_{ij}$ = *the weight of the minimally weighted path (of length 2) through* $\mathcal{B}(A)$ *from* $v(i)$ *to* $v(j)$.

Min-Plus Low-Rank Matrix Approximation

Min-plus low-rank matrix approximation

For $M \in \mathbb{R}_{\min +}^{n \times m}$ and $0 < k \leq \min\{n, m\}$, we seek

$$\min_{A \in \mathbb{R}_{\min +}^{n \times k},\ B \in \mathbb{R}_{\min +}^{k \times m}} \|M - A \otimes B\|_F^2.$$

Network interpretation

Given a network with shortest path distances M build a new network with k 'transport hub' vertices whose shortest path distances approximate M.

Geometrical interpretation

Given m points $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_m \in \mathbb{R}_{\max}^n$ find a k-dimensional min-plus linear space C to minimize

$$\sum_{i=1}^{m} \text{dist}(\boldsymbol{m}_i - C)^2.$$

Min-Plus Low-Rank Matrix Approximation

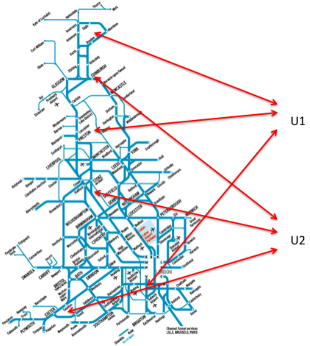


Figure: Original image taken from Network Rail

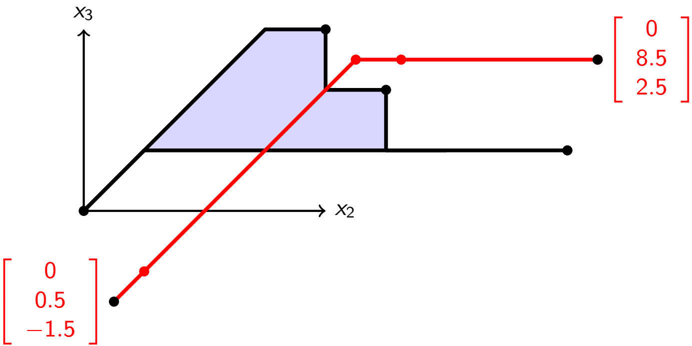
J. Hook.
Min-plus algebraic low rank matrix approximation: a new method for revealing structure in networks.
arXiv:1708.06552.

J. Hook.
Linear regression over the max-plus semiring: algorithms and applications.
arXiv:1712.03499.

Column space geometry viewpoint

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 5 & 8 \\ 0 & 3 & 2 & 1 \end{bmatrix} \approx \begin{bmatrix} 0 & 0 \\ 0.5 & 8.5 \\ -1.5 & 2.5 \end{bmatrix} \otimes \begin{bmatrix} 0.5 & 4 & 4.5 & \infty \\ 0 & 0 & -0.25 & -0.67 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & -0.25 & -0.67 \\ 1 & 4.5 & 5 & 7.83 \\ -1 & 2.5 & 2.25 & 1.83 \end{bmatrix}$$



Max-Times Semiring

The max-times semiring $\mathbb{R}_{\max\times} = [\mathbb{R}_+, \boxplus, \boxtimes]$, where

$$a \boxplus b = \max\{a, b\}, \quad a \boxtimes b = a \times b, \quad \forall\ a, b \in \mathbb{R}_{\max\times}.$$

Max-times matrix multiplication is defined in analogy to the classical case. For $A \in \mathbb{R}_{\max\times}^{n\times m}$ and $B \in \mathbb{R}_{\max\times}^{m\times d}$ we have $A \boxtimes B \in \mathbb{R}_{\max\times}^{n\times d}$, with

$$(A \boxtimes B)_{ij} = \bigboxplus_{k=1}^{m} a_{ik} \boxtimes b_{kj} = \max_{k=1}^{m}(a_{ik} b_{kj}).$$

For example

$$\begin{bmatrix} 0 & 100 & 100 \\ 0 & 1 & 1 \\ 1 & 10 & 1 \end{bmatrix} \boxtimes \begin{bmatrix} 0 & 100 & 100 \\ 0 & 1 & 1 \\ 1 & 10 & 1 \end{bmatrix} = \begin{bmatrix} 100 & 1000 & 100 \\ 1 & 10 & 1 \\ 1 & 100 & 100 \end{bmatrix}.$$

Max-Times Low-Rank Approximation

Max-Times Low-Rank Approximation

Given an input matrix $A \in \mathbb{R}_{\max\times}$ and an integer $k > 0$, find $B \in \mathbb{R}_{\max\times} R_{+}^{n \times k}$, $C \in \mathbb{R}_{\max\times} R_{+}^{k \times m}$, such that

$$\|A - B \boxtimes C\|_F$$

is minimized.

S. Karaev and P. Miettinen.
Capricorn: An Algorithm for Subtropical Matrix Factorization.
SIAM International Conference on Data Mining 2016.

S. Karaev and P. Miettinen.
Cancer: Another Algorithm for Subtropical Matrix Factorization.
ECML PKDD 2016.

Factorization Models

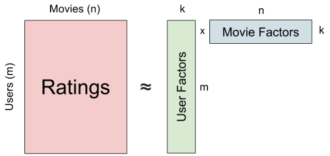


Figure: Image taken from blog2.sigopt.com

1. **SVD:** Sum of parts of different signs. Optimal with ‘classical’ product.
2. **NMF:** Sum of non-negative parts. Interpretable factors ‘parts of a whole’.
3. **Max-times:** Maximum of non-negative parts. Interpretable factors ‘winner takes all’
4. **Mixed Tropical-Linear Model:** Some entries determined by NMF some entries determined by Max-times.

The Mixed Tropical-Linear Model

Given an input matrix $A \in \mathbb{R}_+^{n \times m}$, we seek factor matrices $B \in \mathbb{R}_+^{n \times k}$ and $C \in \mathbb{R}_+^{k \times m}$ and parameters $\alpha \in \mathbb{R}^{n \times m}$, such that

$$A_{ij} \approx \alpha_{ij}(B \boxtimes C) + (1 - \alpha_{ij})(BC)_{ij}.$$

- $\alpha_{ij} \approx 1 \Leftrightarrow A_{ij}$ determined by tropical product
- $\alpha_{ij} \approx 0 \Leftrightarrow A_{ij}$ determined by linear product

We enforce

$$\alpha_{ij} = \sigma(\theta_i + \phi_j),$$

where $\theta \in \mathbb{R}^n$ and $\phi \in \mathbb{R}^m$ are vectors to be determined and σ is the logistic sigmoid

$$\sigma(x) = \frac{1}{1 + \exp(-x)}.$$

The Mixed Tropical-Linear Model

For $B \in \mathbb{R}_+^{n \times k}$, $C \in \mathbb{R}_+^{k \times m}$, $\theta \in \mathbb{R}^n$ and $\phi \in \mathbb{R}^m$ define the mixed tropical-linear product

$$(B \boxtimes_{\theta,\phi} C)_{ij} = \alpha_{ij}(B \boxtimes C) + (1 - \alpha_{ij})(BC)_{ij},$$

where $\alpha_{ij} = \sigma(\theta_i + \phi_j)$.

Mixed Tropical-Linear Low-Rank Approximation

Given an input matrix $A \in \mathbb{R}_+^{n \times m}$ and an integer $k > 0$, find $B \in \mathbb{R}_+^{n \times k}$, $C \in \mathbb{R}_+^{k \times m}$, $\theta \in \mathbb{R}^n$ and $\phi \in \mathbb{R}^m$ such that

$$\|A - B \boxtimes_{\theta,\phi} C\|_F$$

is minimized.

Our Algorithm

Algorithm 1 Latitude

Input: $A \in \mathbb{R}_+^{n\times m}, k \in \mathbb{N}, N \in \mathbb{N}$
Output: $B^* \in \mathbb{R}_+^{n\times k}, C^* \in \mathbb{R}_+^{k\times m}, \theta^* \in \mathbb{R}^{n\times 1}, \phi^* \in \mathbb{R}^{1\times m}$
Parameters: M ▷ The maximum possible value of parameter vectors. In practice 5 is a good choice

function LATITUDE(A, k, N)
  initialize B and C
  $D \leftarrow BC - A$
  $f_i \leftarrow \sum_{j=1}^{m} D_{ij}, g_j \leftarrow \sum_{i=1}^{n} D_{ij}$
  $s_i \leftarrow$ index of the i-th smallest element of f
  $t_j \leftarrow$ index of the j-th smallest element of g
  $\theta_i \leftarrow \frac{i-n}{n-1} M$
  $\phi_j \leftarrow \frac{j-m}{m-1} M$
  $B^* \leftarrow B, C^* \leftarrow C$ ▷ Initialize best factors.
  $\theta^* \leftarrow \theta, \phi^* \leftarrow \phi$ ▷ Initialize best parameters.
  $bestError \leftarrow \left\| A - B \boxtimes_{\theta,\phi} C \right\|_F$
  for $iter \leftarrow 1$ **to** N **do**
    for $j \leftarrow 1$ **to** m **do**
      $[C^j, \phi_j] \leftarrow$ MixReg$(A^j, B, C^j, \theta, \phi_j, M)$
    for $i \leftarrow 1$ **to** n **do**
      $[B_i, \theta_i] \leftarrow$ MixReg$(A_i^T, C^T, B_i^T, \phi, \theta_i, M)$
    if $\left\| A - B \boxtimes_{\theta,\phi} C \right\|_F < bestError$ **then**
      $B^* \leftarrow B, C^* \leftarrow C$
      $\theta^* \leftarrow \theta, \phi^* \leftarrow \phi$
      $bestError \leftarrow \left\| A - B \boxtimes_{\theta,\phi} C \right\|_F$
  return $B^*, C^*, \theta^*, \phi^*$

Algorithm 2 MixReg

Input: $a \in \mathbb{R}_+^{n\times 1}, B \in \mathbb{R}_+^{n\times k}, c \in \mathbb{R}_+^{k\times 1}, \theta \in \mathbb{R}^{n\times 1}, t \in \mathbb{R}, M > 0$
Output: $c \in \mathbb{R}_+^{k\times 1}, t \in \mathbb{R}$

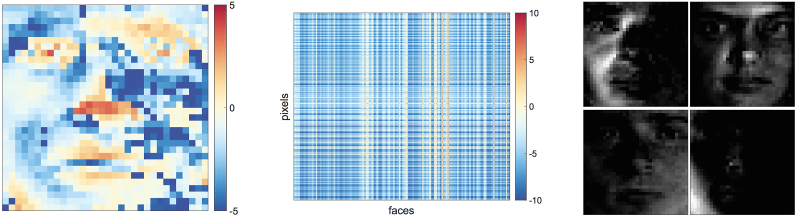
function MIXREG(a, B, c, θ, t, M)
  $X_i \leftarrow B_i \square c^T$
  $\alpha \leftarrow \sigma(\theta + t)$
  $T_{ij} \leftarrow \begin{cases} 1 & j = \arg\max_{1\le s\le k} X_{is} \\ 1 - \alpha_i & \text{otherwise} \end{cases}$
  $Y \leftarrow B \square T$
  $c \leftarrow \arg\min_{p \in \mathbb{R}_+^{k\times 1}} \|a - Bp\|_F$
  $t \leftarrow \arg\min_{s \in [-M, M]} \|a - B \boxtimes_{\theta,s} c\|_F$
  return c, t

Examples

Table: Reconstruction error for real-world datasets.

	Climate	NPAS	Face	4NEWS	HPI
$k =$	10	10	40	20	15
Latitude	**0.023**	**0.207**	0.157	0.536	0.016
SVD	0.025	0.209	**0.140**	**0.533**	**0.015**
NMF	0.080	0.223	0.302	0.541	0.124
Cancer	0.066	0.237	0.205	0.554	0.026

Examples



(a) Vector θ (b) Matrix α (c) Columns of B

Figure 3: (a) Vector θ for the Face data as an image. (b) Matrix α for the Face data. (c) Four columns of B for the Face data.

Conclusion

- 'Classical' low-rank approximate factorizations used throughout applied maths.
- Tropical low-rank approximate factorizations including min-plus and max-times provide a completely different model but with analogous algebraic structure.
- We introduced a novel model that interpolates between NNMF and max-times.
- Able to outperform SVD on some real life data sets. What is the structure being detected?

S. Karaev, J. Hook and P. Miettinen.
Latitude: A Model for Mixed Linear-Tropical Matrix Factorization.
SIAM International Conference on Data Mining 2018.