Solving Nonnegative Matrix Factorization via Coordinate Descent

Pan, Lanfeng[1] Lu, Xiaoling[2] Wang, Xin[3]

[123]School of Statistics
Renmin University of China

NMF

- Nonnegative Matrix Factorization is a useful exploratory data analysis method.
- Given a nonnegative matrix V, it tries to find smaller nonnegative matrics W and H to approximate V, i.e. $V_{n\times m} \approx W_{n\times r} H_{r\times m}$.
- The rank r is small, so W and H are much more smaller than V.

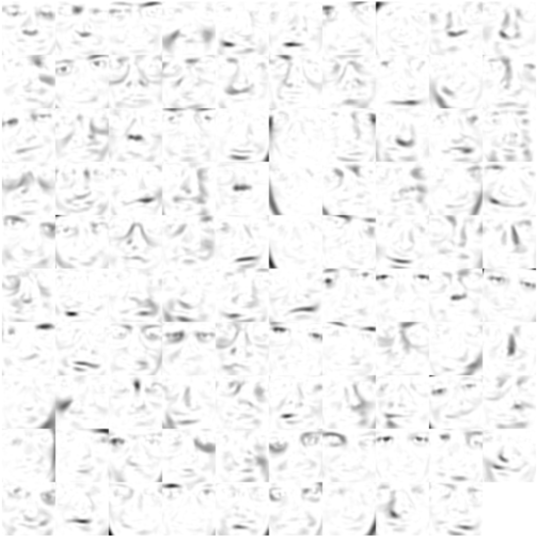
NMF and other unsupervised methods

- Principle Components Analysis, Factor Analysis and Kmeans all need to find some basis matrix W and coefficient matrix H.
- Different motivations result in different optimization methods and different explanations.
- The nonnegative constrains of NMF make W and H sparse.

Result of Sparseness

In text mining and image data, sparseness makes W and H easy to be explained.

Result of Sparseness



Multiplicative Update

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T V)_{a\mu}}{(W^T WH)_{a\mu}} \qquad W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia}}$$

Problem: Slow

Once one element of W or H becomes 0 during the iteration, it will remain 0 after that. This is a very bad property.

Alternating Least Squares

$$W^{k+1} = \arg\min_{W\geq 0} f(W, H^k), \qquad (1)$$

$$H^{k+1} = \arg\min_{H\geq 0} f(W^{k+1}, H). \qquad (2)$$

Projected Gradient Descent solved above bounded optimization problem by project updated parameters into feasible space.

$$x^{k+1} = P[x^k - \alpha_k \nabla f(x^k)].$$

This algorithm is quite fast but not fast enough.

New Algorithm

Solve bounded optimization problem via coordinate descent.

Coordinate Descent

Given an optimization problem

$$\max_{\beta} f(\beta_1, \beta_2, \ldots, \beta_p),$$

update one parameter one time by gradient descent.

$$\beta_i^{k+1} = \arg\max_{\beta_i} f(\beta_1^k, \beta_2^k, \ldots, \beta_i, \ldots, \beta_p^k)$$

Nonnegative Regression

Given nonnegative regression problem

$$\min_{\beta \geq 0} \| Y - X\beta \|^2$$

and its initial solution $\widetilde{\beta}$, we calculate current residuals. For $i = 1, 2, \ldots, p$, calculate their regression coefficient $\Delta\beta_i$ with X_i. Add $\Delta\beta_i$ on β_i, update current residuals and go to next i.

Our algorithm for NMF

- Treat $W^{k+1} = \arg\min_{W \geq 0} f(W, H^k)$ as several nonnegative regressions, we can update W. H is treated in the same way.
- In each iteration, W^k can be take as the initial value for W which could largely accelerate the algorithm.
- The computing complexity is $O(t(mnr + \frac{1}{2}mr^2 + \frac{1}{2}nr^2))$.

Simulations

We randomly generate matrices of different size and apply our algorithms with different ranks.

rank	20		40		60		80	
columns	CD	PG	CD	PG	CD	PG	CD	PG
200	0.8	1.7	1.3	3.8	1.8	7.1	2.7	10.9
2000	15.5	30.9	23.6	63.3	33.5	103	46.1	153
20000	717	2852	1057	4819	1474	6687	1966	8427

Example

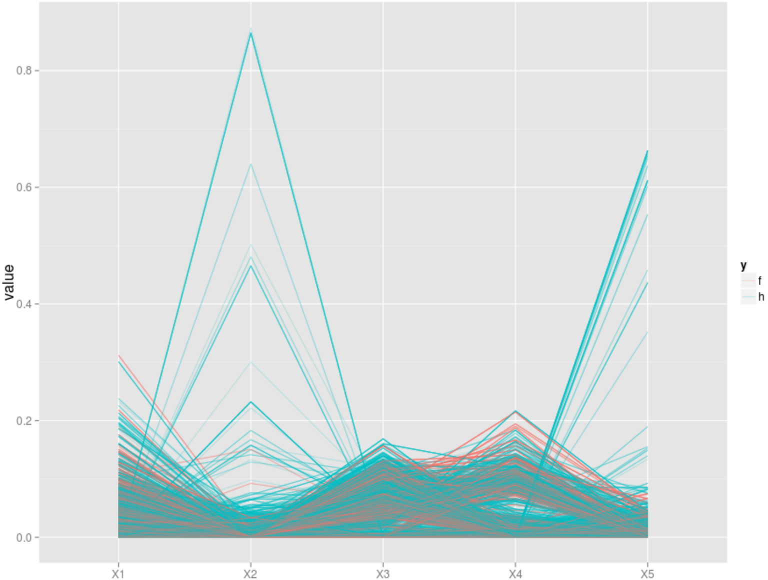
For a real word frequency dataset of the size of 1706×4170, we tried different ranks.

rank	20	40	60	80
CD	11.8	19	26.8	34.8
PG	32.4	62.4	100.7	166

Example

A real weibo dataset grasped from weibo.com of the size of 32147×10711. The weibos are of two different topics, "Han, Han"and "Fang, Zhouzi". The following graph visualized the W matrix.

Example



Example

“一部分, 不, 不只, 中, 九, 也是, 人民日, 令人, 作家, 創, 助人, 動, 司, 善良, 大, 已, 常, 心, 意, 感, 拾, 文, 文化, 日常生活, 味, 最近, 核心, 機, 歷, 灣, 熱, 物, 生活, 發, 眼, 知名, 程, 經, 老, 與, 英, 融入, 行, 表, 親, 計, 訪, 認, 讓, 身, 車, 這, 都, 都是, 鏡, 闆, 陸, 震撼, 韓寒”

”171,5,http, 一, 上, 不少, 为, 举行, 也, 人, 作者, 六六, 其, 博, 厘米, 发出, 吗, 图片, 在, 墙, 宴会, 宴请, 家里, 席, 引发, 当时, 微, 手里, 拿着, 据, 日前, 昨日, 有, 此举, 测, 测量, 热, 特地, 真的, 站着, 等, 网友, 蜗居, 袜子, 认为, 议, 请来, 贴, 赤脚, 身高, 还, 还有, 透露, 那么, 重要, 量身, 随后, 高的”

Summary

- We have implemented the new algorithm in C++ and also wrapped the code in a R package bignmf. The source code has been push on Github. See https://github.com/panlanfeng/bignmf

References

[1] D Lee and H Seung. Algorithms for non-negative matrix factorization. In in Neural Information Processing Systems, 2000, pages 556–562.

[2] Chih jen Lin. Projected gradient methods for non-negative matrix factorization. Technical report, Neural Computation, 2007.

[3] Noah Simon, Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 39(5):1–13, 2011.