

UKRAINIAN CATHOLIC UNIVERSITY

FACULTY OF APPLIED SCIENCES

BUSINESS ANALYTICS & COMPUTER SCIENCE PROGRAMMES

Predicting Age by MRI Images Using Regression Based Non-Negative Matrix Factorization

Linear Algebra final project report

Authors:

Volodymyr CHERNETSKYI

Olesya TRETYAK

Hermann YAVORSKYI

15 May 2019



APPLIED
SCIENCES
FACULTY ●

Abstract

Non-negative matrix factorization is an excellent tool for unsupervised parts-based learning, but proves to be ineffective when parts of a whole follow a specific pattern [1]. Analyzing such local changes is particularly important when studying anatomical transformations. We implement [2] an algorithm called Regression based Non-Negative Matrix Factorization (RNMF), which incorporates a regression constraint into the NMF framework, so that it learns maximally changing parts in the basis images.

This project is focused on implementing RNMF, rather than implementing NMF or predicting model.

1 Introduction

Magnetic Resonance imaging (MRI) is a popular technique used to study the anatomy of the brain. Due to the advancement in technology involved in MRI, the need for automation in related processes has increased. One of such processes is distinguishing brain changes caused by aging between those caused by diseases.

2 Problem Setting

To accurately predict a patient's age by MRI image, we need to extract the most relevant regions of the brain image. We use a matrix decomposition method to simultaneously reduce the dimension of image, thus reducing the number of variables for predictor, and get those relevant regions.

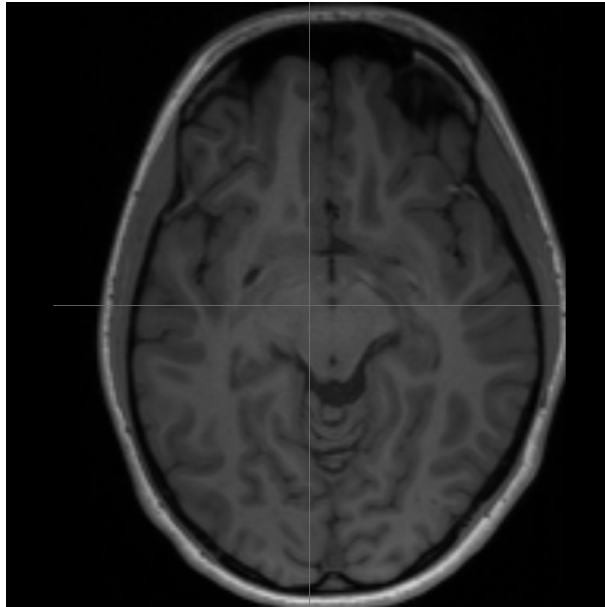


Figure 1: Example of input MRI image.

There are some well established matrix factorization algorithms in unsupervised learning: Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Non-negative matrix factorization (NMF), that represent data as a linear combination of basis images.

Our solution uses modified NMF to get the work done. NMF [3], unlike other matrix factorization algorithms, represents visual information as a sum of its parts - just like people. We combine NMF with regression methods not to only decompose an image into smaller ones, but to choose the most influential ones.

3 Solution Description

Non-negative matrix factorization (NMF) is distinguished by its use of non-negativity constraints. NMF decomposes a non-negative matrix (V) to a set of non-negative bases (W) and corresponding non-negative coefficients (H),

$$V_{n \times n_t} \approx W_{n \times m} H_{m \times n_t}$$

where $V = [v_{i,j}] = [v_1, \dots, v_{n_t}]$ is a $n \times n_t$ matrix, where n is the total number of pixels in each image, v_j is the j th input image represented as a column vector, and n_t is the number of images. We denote the basis matrix $W = [w_{i,j}] = [w_1, \dots, w_m]$ as $n \times m$ matrix, $(n + n_t)m < nn_t$. The low dimensional embedding of every column of V is the corresponding column in $H = [h_{i,j}] = [h_1, \dots, h_{n_t}]$. This factorization is achieved by minimizing the divergence between V and WH applying non-negative constraints. The divergence between V and $Y = WH$ is defined as [3] [4]

$$D(V||WH) = \sum_{i,j} (v_{ij} \log \frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij})$$

To implement regression features we should make our data smoothly separated. Which means if labels of Y for corresponding labels of V are close, then corresponding labels in H should reflect this proximity. This is imposed by following constraint:

$$S_R = \frac{\sum_{i,j} f(y_i, y_j) ||h_i - h_j||}{\sum_{i,j} f(y_i, y_j)}$$

Function f is a heat kernel - weighting function with positive values

$$f(y_i, y_j) = \exp(\frac{-||y_i - y_j||}{t})$$

t depends on the range of labels. So f will decrease when the distance between data increase. This constraint only considers the labels assigned to each sample in the cross-sectional data. However, if a pair of points have similar labels, but are highly distant in the input space, the influence of such points should be minimized. We compute weights that are inversely proportional to the geodesic distance between the sample points, thus penalizing outliers.

$$m_{ij} = \frac{d_{ij}}{\|v_i - v_j\|} f(y_i, y_j)$$

where d shortest path distance matrix calculated using Floyd's algorithm [5] on the graph. Including this in the constraint above we get

$$S_{Rm} = \frac{\sum_{i,j} m_{ij} \|h_i - h_j\|}{\sum_{i,j} m_{ij}}$$

To outline most regressing features, we are localizing the bases. We are computing bases images W_j that are less noisy and restricted to the inclusion of compact regions. Also, local parts should be captured as contiguous regions. To achieve this we minimize the energy of the gradient of each of the basis W_j in the image coordinates.

$$S_G = \sum_j \int |\nabla W_j(x)|^2 dx$$

And finally, to minimize the number of basis components required to represent V and to make bases as orthogonal as possible to reduce redundancy between them we minimize [6]

$$S_O = \sum_{i,j} W^T W$$

The inclusion of the above three constraints leads to the following constrained divergence to be minimized:

$$D(V||WH) = \sum_{i,j} (v_{ij} \log \frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}) + \alpha S_{Rm} + \beta S_G + \gamma S_O$$

where $WH = Y$ and $\alpha, \beta, \gamma > 0$ are constants, which are empirically determined such that we get non-negative updates.

The following update rules can be found by minimization of the above constrained divergence:

$$h_{kl} \leftarrow \frac{-b + \sqrt{b^2 + \left(\frac{\sum_i v_{il} w_{ik} h'_{kl}}{\sum_j w_{ij} h'_{jl}}\right) \left(\frac{16\alpha m_{kl}}{\sum_{i,j} m_{ij}}\right)}}{\frac{8\alpha m_{kl}}{\sum_{i,j} m_{ij}}}, \text{ where } b = 1 - 4\alpha \frac{\sum_j h_{kj} m_{jl}}{\sum_{i,j} m_{ij}}$$

$$w_{kl} \leftarrow \frac{w_{kl} \frac{\sum_j v_{kj} h_{lj}}{\sum_i w_{il} h_{lj}}}{\sum_j h_{lj} + \beta g_{kl} + \gamma \sum_j w_{kj}}, \text{ where } g = -\nabla^2 W$$

$$w_{kl} \leftarrow \frac{w_{kl}}{\sum_j w_{jl}}$$

The proof of convergence is provided in [1].

The output is predicted using the linear regression on the lower dimensional representation (H).

4 Implementation

NMF algorithm taken from scikit-learn module decomposition.

Linear regression algorithm taken from scikit-learn module linear_model.

RNMF implemented using numpy.

```
get  $V$  from known and unknown images;
apply NMF to  $V$ ;
 $Y \leftarrow WH$ ;
calculate divergence;
while difference between old and new divergences is bigger than fixed value do
    update  $H$ ;
    update  $H$ ;
     $Y \leftarrow WH$ ;
    calculate divergence;
end
linear_regression_model fit(known part of  $H$ );
linear_regression_model predict(unknown part of  $H$ );
```

Algorithm 1: Pseudocode

5 Conclusions

Using Regression based Non-negative matrix factorization for decomposing MRI images for further predicting the age of the patient is more effective than using basic Non-negative matrix factorization for the same purposes. Moreover, RNMF is more effective than using linear regression without any preprocessing, which means having only the most significant parts of brain image is more effective than having the whole image itself.

References

- [1] Swapna Joshi., *Anatomical Parts-Based Regression Using Non-Negative Matrix Factorization*, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2863–2870.
- [2] <https://github.com/chernetskyi/anatomical-regression-using-nmf>
- [3] Lee D, Seung H. *Learning the parts of objects by nonnegative matrix factorization*, 1999
- [4] Lee D, Seung H. *Algorithms for non-negative matrix factorization. Advances in neural information processing systems*, 2001
- [5] Cormen T, Leiserson C, Rivest R, Stein C. *Introduction to algorithms*, 2001
- [6] Li S, Hou X, Zhang H, Cheng Q. *Learning spatially localized, parts-based representation*, 2001.