Nonnegative Matrix Factorization

MaoQiang Xie
College of Software, NKU

Introduction

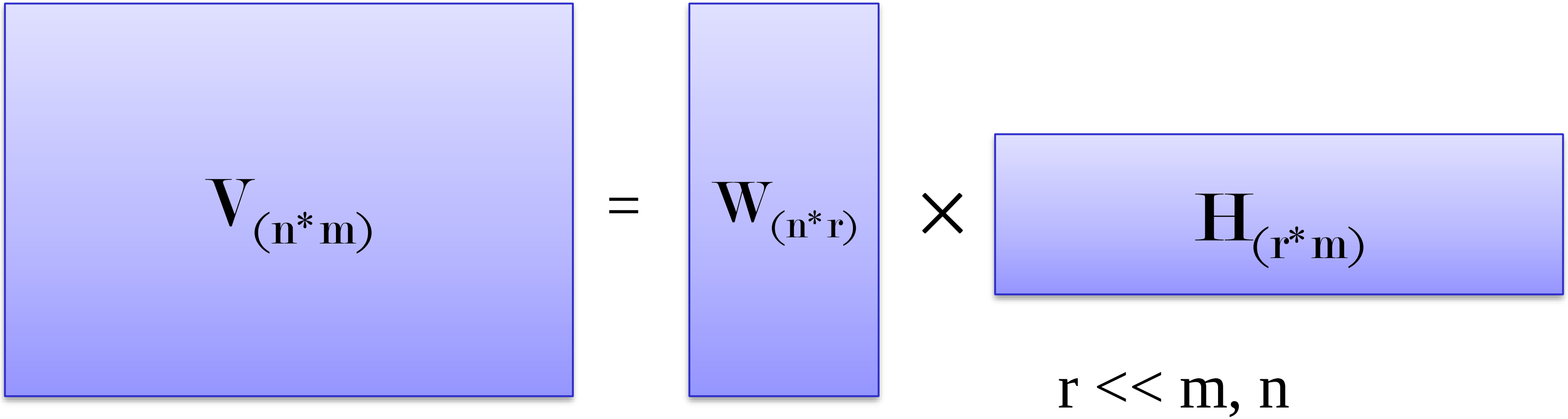
- Learning the parts of objects by nonnegative matrix factorization（Nature， 1999）
- D.D. Lee (Bell Lab. 现在MIT脑认知科学实验室负责人), H.S. Seung (from MIT)

- 功能：特征压缩、软聚类
- 应用领域：图像特征抽取、文本语义模型、基因数据分析、语音信号处理、商品推荐等

Nonnegative Matrix Factorization

- Given a non-negative matrix V, find non-negative matrix factors W and H such that:

$$\mathbf{V}_{(n*m)} \approx \mathbf{W}_{(n*r)}\,\mathbf{H}_{(r*m)}$$



$$\mathbf{V}_{(n*m)} = \mathbf{W}_{(n*r)} \times \mathbf{H}_{(r*m)}$$

$r << m, n$

$(m+n) * r < nm$

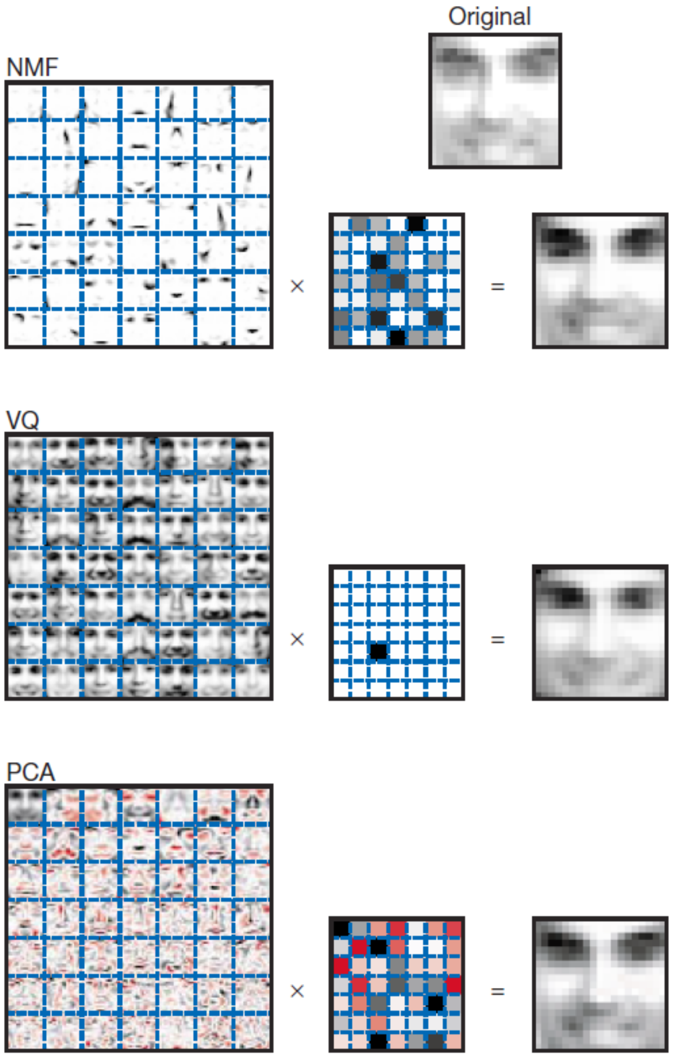


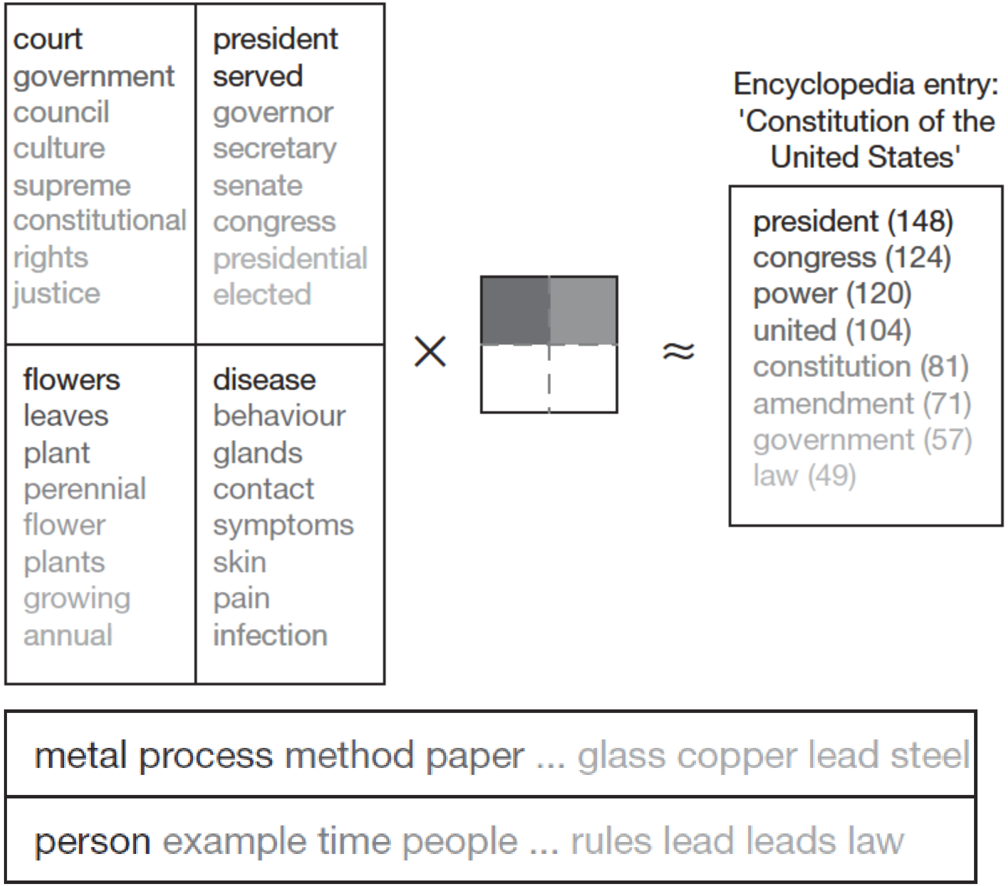
Figure 1 Non-negative matrix factorization (NMF) learns a parts-based representation of faces, whereas vector quantization (VQ) and principal components analysis (PCA) learn holistic representations. The three learning methods were applied to a database of $m = 2{,}429$ facial images, each consisting of $n = 19 \times 19$ pixels, and constituting an $n \times m$ matrix V. All three find approximate factorizations of the form $V \approx WH$, but with three different types of constraints on W and H, as described more fully in the main text and methods. As shown in the 7 × 7 montages, each method has learned a set of $r = 49$ basis images. Positive values are illustrated with black pixels and negative values with red pixels. A particular instance of a face, shown at top right, is approximately represented by a linear superposition of basis images. The coefficients of the linear superposition are shown next to each montage, in a 7 × 7 grid, and the resulting superpositions are shown on the other side of the equality sign. Unlike VQ and PCA, NMF learns to represent faces with a set of basis images resembling parts of faces.

V：n×m，n行：像素，m列：一幅图像

W：n×r，r列，每列一个basis image

H：r×m, m列，每列一个encoding

- 每个basis topic为W中该列出现次数最多的8个单词，颜色表示程度
- 一词多义被成功分在不同basis topic中
- Lead 领导；铅



court
government
council
culture
supreme
constitutional
rights
justice

president
served
governor
secretary
senate
congress
presidential
elected

flowers
leaves
plant
perennial
flower
plants
growing
annual

disease
behaviour
glands
contact
symptoms
skin
pain
infection

×

≈

Encyclopedia entry:
'Constitution of the
United States'

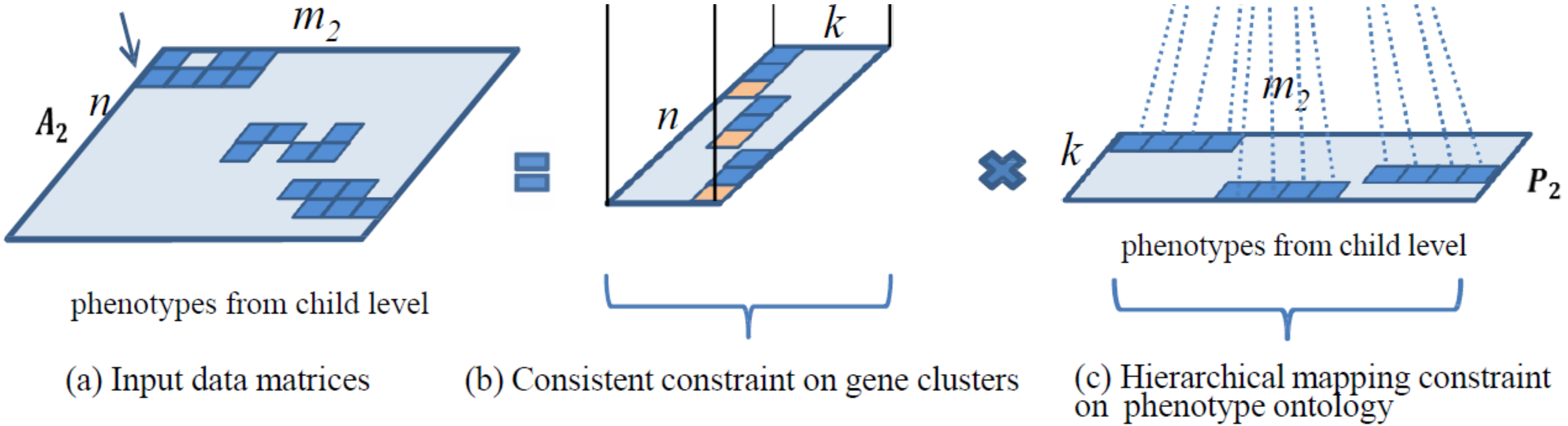
president (148)
congress (124)
power (120)
united (104)
constitution (81)
amendment (71)
government (57)
law (49)

metal process method paper ... glass copper lead steel

person example time people ... rules lead leads law

Figure 4 Non-negative matrix factorization (NMF) discovers semantic features of $m = 30,991$ articles from the Grolier encyclopedia. For each word in a vocabulary of size $n = 15,276$, the number of occurrences was counted in each article and used to form the 15,276 × 30,991 matrix V. Each column of V contained the word counts for a particular article, whereas each row of V contained the counts of a particular word in different articles. The matrix was approximately factorized into the form WH using the algorithm described in Fig. 2. Upper left, four of the $r = 200$ semantic features (columns of W). As they are very high-dimensional vectors, each semantic feature is represented by a list of the eight words with highest frequency in that feature. The darkness of the text indicates the relative frequency of each word within a feature. Right, the eight most frequent words and their counts in the encyclopedia entry on the 'Constitution of the United States'. This word count vector was approximated by a superposition that gave high weight to the upper two semantic features, and none to the lower two, as shown by the four shaded squares in the middle indicating the activities of H. The bottom of the figure exhibits the two semantic features containing 'lead' with high frequencies. Judging from the other words in the features, two different meanings of 'lead' are differentiated by NMF.

Nonnegative Matrix Factorization



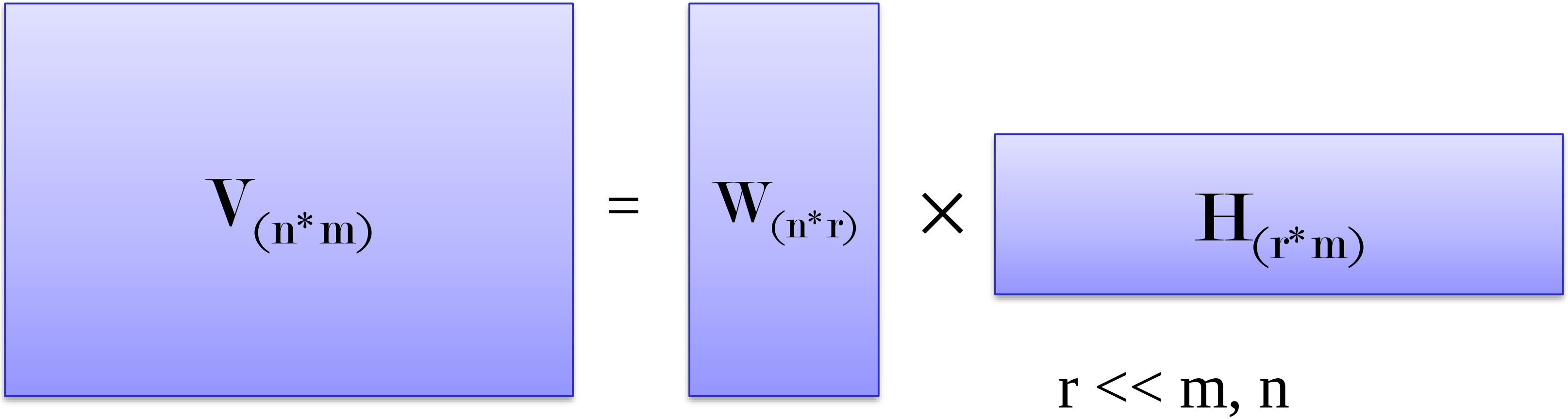
(a) Input data matrices (b) Consistent constraint on gene clusters (c) Hierarchical mapping constraint on phenotype ontology

- V：表型-基因矩阵，用户-商品矩阵
- 分解后
  - W：表型-簇，用户-簇
  - H：压缩特征-基因，压缩特征-商品

Nonnegative Matrix Factorization

- Given a non-negative matrix V, find non-negative matrix factors W and H such that:

$$\mathbf{V}_{(n*m)} \approx \mathbf{W}_{(n*r)}\,\mathbf{H}_{(r*m)}$$



$r << m, n$

$(m+n) * r < nm$

Cost Function

- 分解前后的误差：$V = WH + E$

- 损失函数：$min_{W,H}\|V - WH\|$

使用欧式距离衡量误差 $$min_{W,H}\sum_{i,j}(V_{ij} - (WH)_{ij})^2$$

使用KL散度衡量误差 $$min_{W,H}\sum_{i,j}(V_{ij}log\frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij})$$

Cost Function

- 分解前后的误差：$V = WH + E$
- 损失函数： $min_{W,H}\|V - WH\|$
- 假设噪声E服从高斯分布：

$$p(E|W,H) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} exp(-\frac{E_{ij}^2}{2\sigma_{ij}^2})$$

- 构造似然函数：

$$\{W,H\} = argmax_{W,H}p(V|W,H) = argmin_{W,H}\{-logp(V|W,H)\}$$

Optimization

- 假设噪声E服从高斯分布：

$$p(E|W,H) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} exp(-\frac{E_{ij}^2}{2\sigma_{ij}^2})$$

- 构造似然函数：

$$\{W,H\} = argmax_{W,H} p(V|W,H) = argmin_{W,H}\{-logp(V|W,H)\}$$

$$L(W,H) = \frac{1}{2\sigma_{ij}^2}\sum_{i,j}(V_{ij} - (WH)_{ij})^2 + \sum_{i,j} log(\sqrt{2\pi}\sigma_{ij})$$

Optimization

$$L(W,H) = \frac{1}{2\sigma_{ij}^2}\sum_{i,j}(V_{ij} - (WH)_{ij})^2 + \sum_{i,j} log(\sqrt{2\pi}\sigma_{ij})$$

$$\frac{\partial L(W,H)}{\partial W_{ik}} = c[\sum_j H_{kj}(V_{ij} - (WH)_{ij})]$$

$$= c[\sum_j V_{ij}H_{kj} - \sum_j (WH)_{ij}H_{kj}]$$

$$= c[(VH^T)_{ik} - (WHH^T)_{ik}]$$

$$\frac{\partial L(W,H)}{\partial H_{kj}} = c[\sum_i W_{ik}V_{ij} - \sum_i (WH)_{ij}W_{ik}]$$

$$= c[(W^TV)_{kj} - (W^TWH)_{kj}]$$

Optimization

从负梯度方向进行更新

$$W_{ik}^{(n+1)} \leftarrow W_{ik}^{(n)} + \lambda_1 * [(VH^T)_{ik} - (WHH^T)_{ik}]$$

$$H_{kj}^{(n+1)} \leftarrow H_{kj}^{(n)} + \lambda_2 * [(W^TV)_{kj} - (W^TWH)_{kj}]$$

设学习步长：

$$\lambda_1 = \frac{W_{ik}^{(n)}}{(WHH^T)_{ik}}, \lambda_2 = \frac{H_{kj}^{(n)}}{(W^TWH)_{kj}}$$

将学习步长代入得：

$$W_{ik} \leftarrow W_{ik} \frac{(VH^T)_{ik}}{(WHH^T)_{ik}} \qquad H_{kj} \leftarrow H_{kj} \frac{(W^TV)_{kj}}{(W^TWH)_{kj}}$$

Algorithm

```
Algorithm NMF_Euclid
INPUT：  V
OUTPUT： W, H
Randomize W, H;
Repeat
    Update  $W_{ik} \leftarrow W_{ik} \frac{(VH^T)_{ik}}{(WHH^T)_{ik}}$
    Update  $H_{kj} \leftarrow H_{kj} \frac{(W^TV)_{kj}}{(W^TWH)_{kj}}$
Until converge or maximal iterations
```

Cost Function with KL Divergence

$$min_{W,H} \sum_{i,j} (V_{ij} log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij})$$

Kullback–Leibler Divergence

$$D_{\mathrm{KL}}(P \| Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)}$$

Information Entropy

$$\mathrm{H}(X) = \sum_i \mathrm{P}(x_i)\, \mathrm{I}(x_i) = -\sum_i \mathrm{P}(x_i) \log_b \mathrm{P}(x_i),$$

Optimization

- 假设噪声E服从泊松分布：

$$p(V_{ij}|W,H) = \frac{(WH)_{ij}^{X_{ij}}}{X_{ij}!} exp(-(WH)_{ij})$$

- 构造似然函数：

$$L(W,H) = \sum_{ij} [V_{ij} log(WH)_{ij} - (WH)_{ij} - log(X_{ij}!)]$$

Optimization

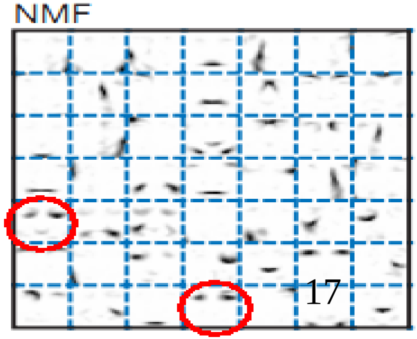
- 更新规则：

$$W_{ik} \leftarrow W_{ik} \frac{\sum_j H_{kj} X_{ij} / (WH)_{ij}}{\sum_j H_{kj}}$$

$$H_{kj} \leftarrow H_{kj} \frac{\sum_i W_{ik} X_{ij} / (WH)_{ij}}{\sum_i W_{ik}}$$

Limitation of NMF

- 不适合所有的问题：比如采集到非固定点拍摄的图像，或者高清晰物体就不适合用NMF做。
- 对于这种复杂问题part学习的处理，就需要一个多层隐变量的结构模型(类似DL)，而不像NMF中只用一层表示隐变量。
- 尽管非负这个约束可以进行part-based representation的学习，它们在编码的相关性方面也是有不足的，
- NMF只约束了W和H的非负性（这是唯一先验，只要求满足这个），而没有考虑V、W、H内部元素间的相关性。



Application 1

疾病相关、基因相关成组分析

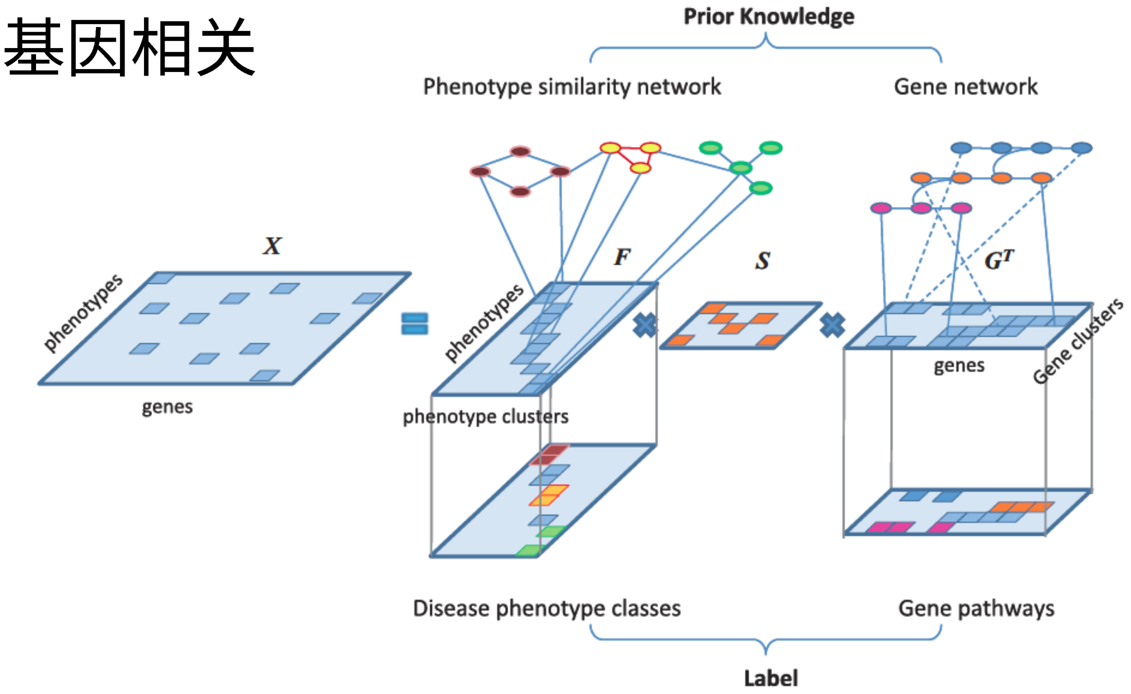
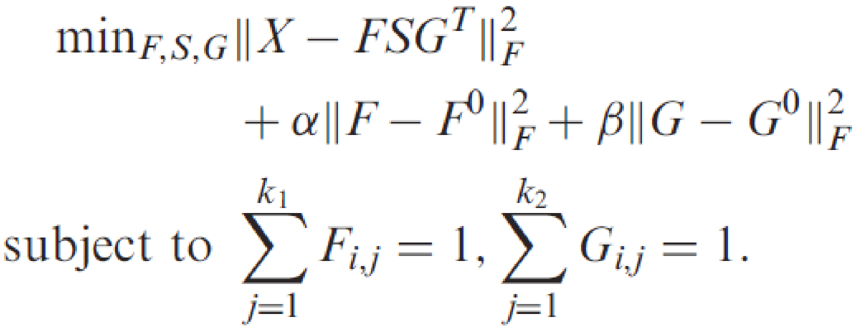


Figure 1. NMTF of disease phenotype–gene associations. The phenotype–gene association matrix X is factorized into products of three matrices, phenotype cluster membership F, gene cluster membership G and phenotype cluster–gene cluster association S for supervised co-clustering of phenotypes and genes. Label information for the disease classes and the pathways are available for a small number of phenotypes and genes. Prior knowledge is also introduced from phenotype similarity network and gene network. For better visualization, different colors are used to distinguish the phenotypes and the genes in different clusters.

Cost Function



$$\min_{F,S,G} \|X - FSG^T\|_F^2 + \alpha\|F - F^0\|_F^2 + \beta\|G - G^0\|_F^2$$

$$\text{subject to } \sum_{j=1}^{k_1} F_{i,j} = 1, \sum_{j=1}^{k_2} G_{i,j} = 1.$$

- 矩阵分解误差
- 疾病聚类同先验吻合
- 基因聚类同先验吻合

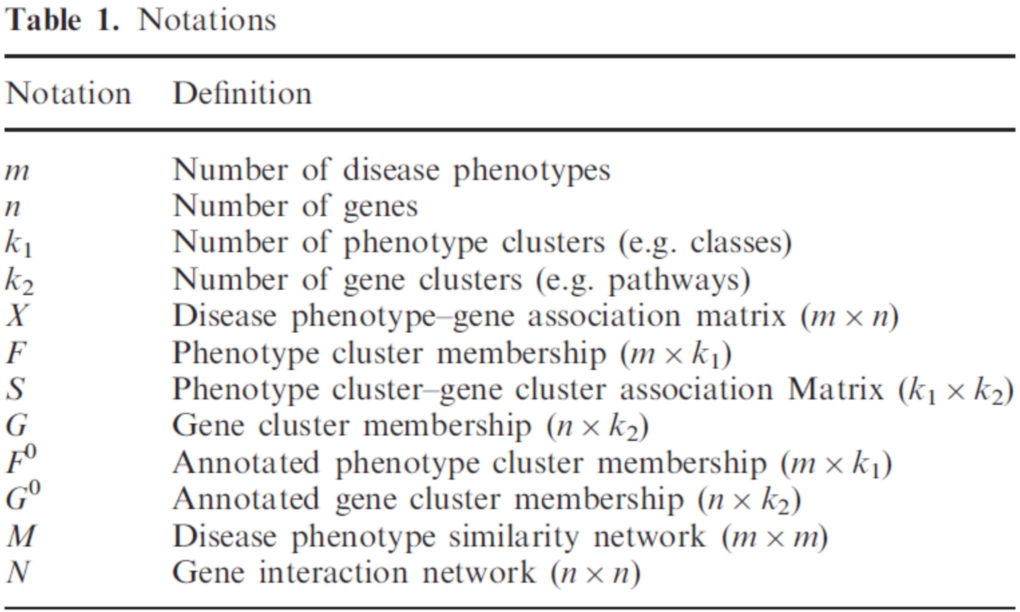
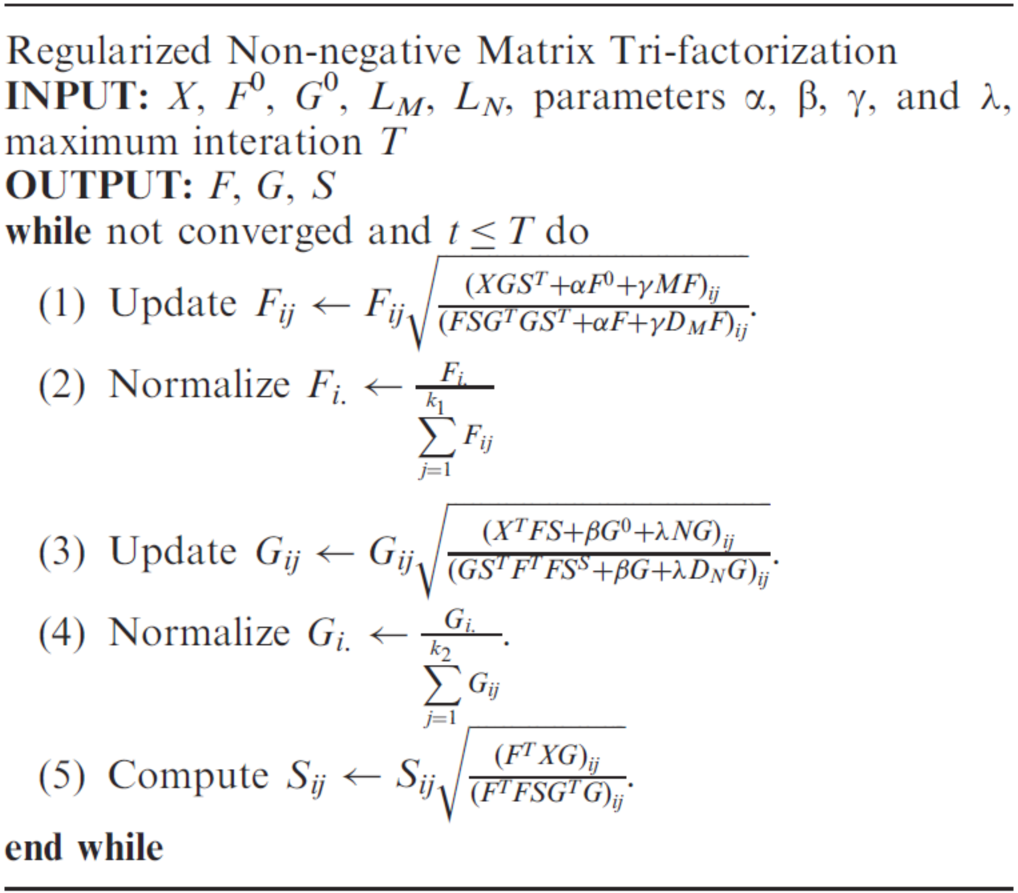


Table 1. Notations

Notation	Definition
m	Number of disease phenotypes
n	Number of genes
k_1	Number of phenotype clusters (e.g. classes)
k_2	Number of gene clusters (e.g. pathways)
X	Disease phenotype–gene association matrix ($m \times n$)
F	Phenotype cluster membership ($m \times k_1$)
S	Phenotype cluster–gene cluster association Matrix ($k_1 \times k_2$)
G	Gene cluster membership ($n \times k_2$)
F^0	Annotated phenotype cluster membership ($m \times k_1$)
G^0	Annotated gene cluster membership ($n \times k_2$)
M	Disease phenotype similarity network ($m \times m$)
N	Gene interaction network ($n \times n$)

Algorithm



Regularized Non-negative Matrix Tri-factorization

INPUT: X, F^0, G^0, L_M, L_N, parameters α, β, γ, and λ, maximum interation T

OUTPUT: F, G, S

```
while not converged and $t \leq T$ do
    (1) Update $F_{ij} \leftarrow F_{ij}\sqrt{\frac{(XGS^T+\alpha F^0+\gamma MF)_{ij}}{(FSG^TGS^T+\alpha F+\gamma D_M F)_{ij}}}$.
    (2) Normalize $F_{i.} \leftarrow \frac{F_{i.}}{\sum_{j=1}^{k_1} F_{ij}}$
    (3) Update $G_{ij} \leftarrow G_{ij}\sqrt{\frac{(X^TFS+\beta G^0+\lambda NG)_{ij}}{(GS^TF^TFS^S+\beta G+\lambda D_N G)_{ij}}}$.
    (4) Normalize $G_{i.} \leftarrow \frac{G_{i.}}{\sum_{j=1}^{k_2} G_{ij}}$.
    (5) Compute $S_{ij} \leftarrow S_{ij}\sqrt{\frac{(F^TXG)_{ij}}{(F^TFSG^TG)_{ij}}}$.
end while
```

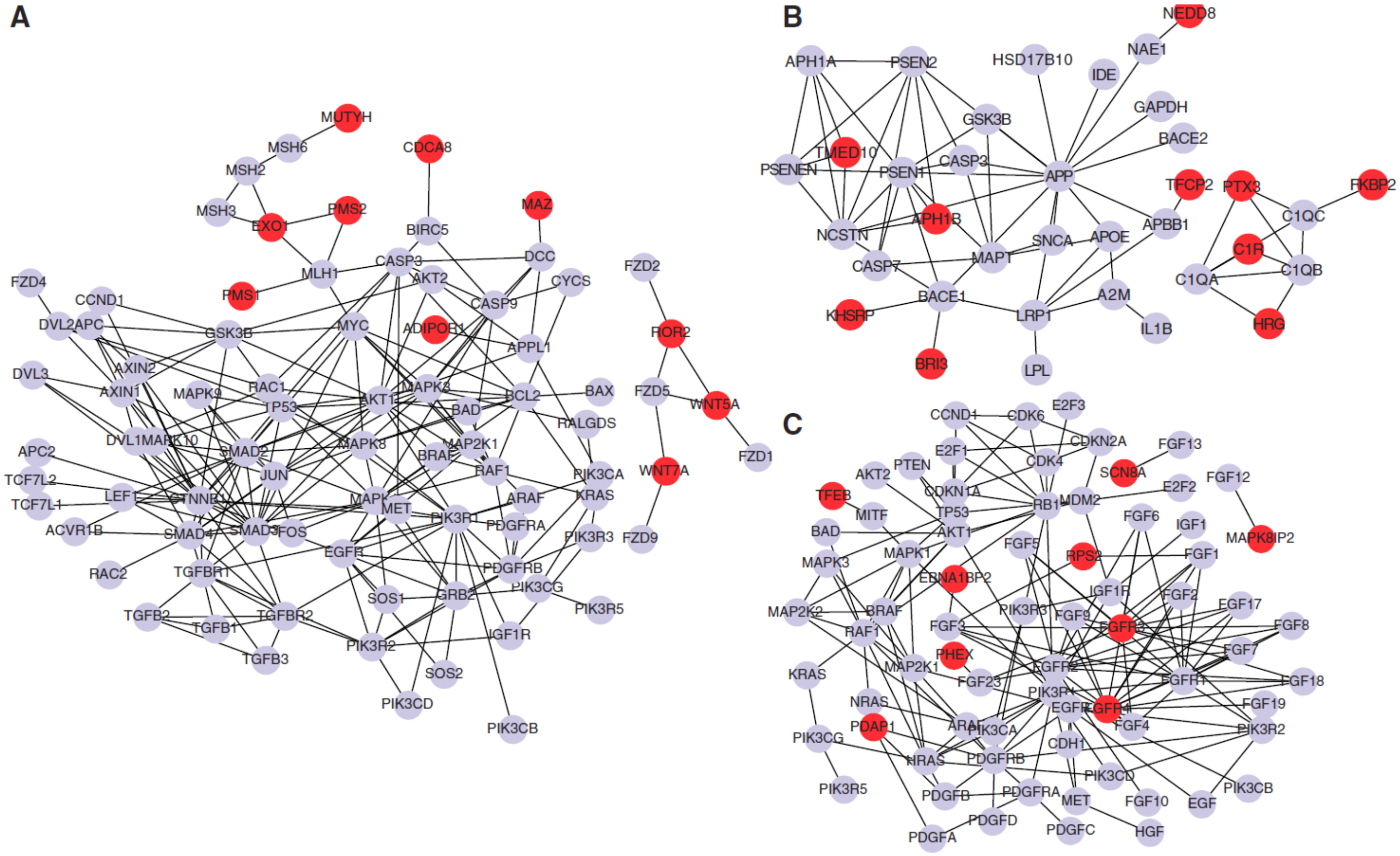
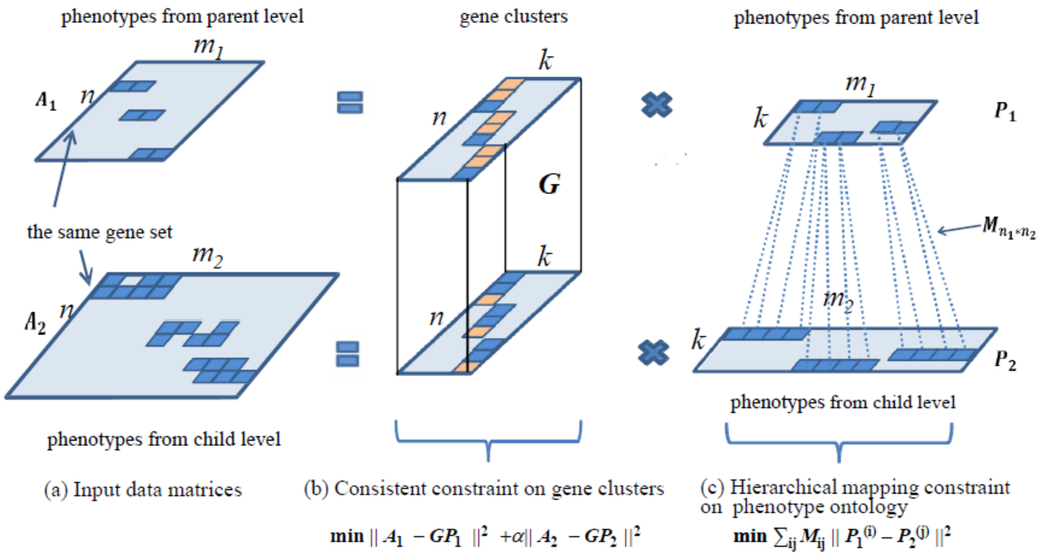


Figure 5. PPI subnetworks of the extended disease pathways. In each pathway, gray nodes are known member genes in the disease pathways and red nodes are newly predicted member genes. Edges represent PPI between two genes. Note that if a known or a newly predicted member gene is not interacting with any other member genes in the pathway, the gene is not included. (**A**) Colorectal cancer pathway. The predicted colorectal cancer genes EXO1 and ADIPOR1 are interacting with many other genes in the colorectal cancer pathway. (**B**) Alzheimer pathway. Over-expression of C1R is known for involving alzheimer disease. (**C**) Melanoma pathway. Mutation and copy number changes in new member gene FGFR3 were recently discovered in melanoma.

Application 2

一致性矩阵分解

不同粒度描述表征不影响聚类结果



(a) Input data matrices

(b) Consistent constraint on gene clusters

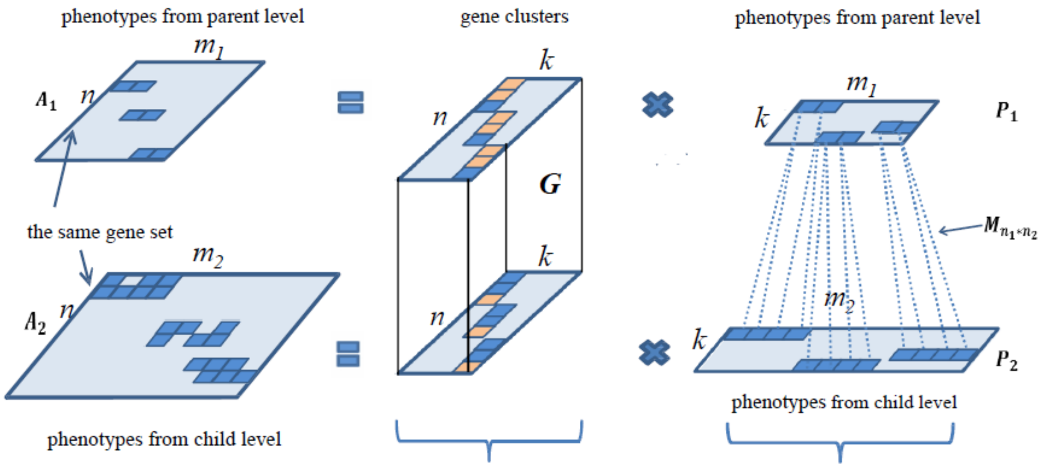
$$\min \| A_1 - GP_1 \|^2 + \alpha \| A_2 - GP_2 \|^2$$

(c) Hierarchical mapping constraint on phenotype ontology

$$\min \sum_{ij} M_{ij} \| P_1^{(i)} - P_2^{(j)} \|^2$$

Figure 1: Illustration of the CMNMF framework. (a) The gene-phenotype associations are divided into two matrices according to the level of phenotype ontologies, and the two matrices share the same gene set. (b) consistent constraint on factorized gene clusters. (c) Hierarchical mapping constraint on the phenotype ontologies at parent and child levels.

Cost Function



(a) Input data matrices

(b) Consistent constraint on gene clusters

$$\min \| A_1 - GP_1 \|^2 + \alpha \| A_2 - GP_2 \|^2$$

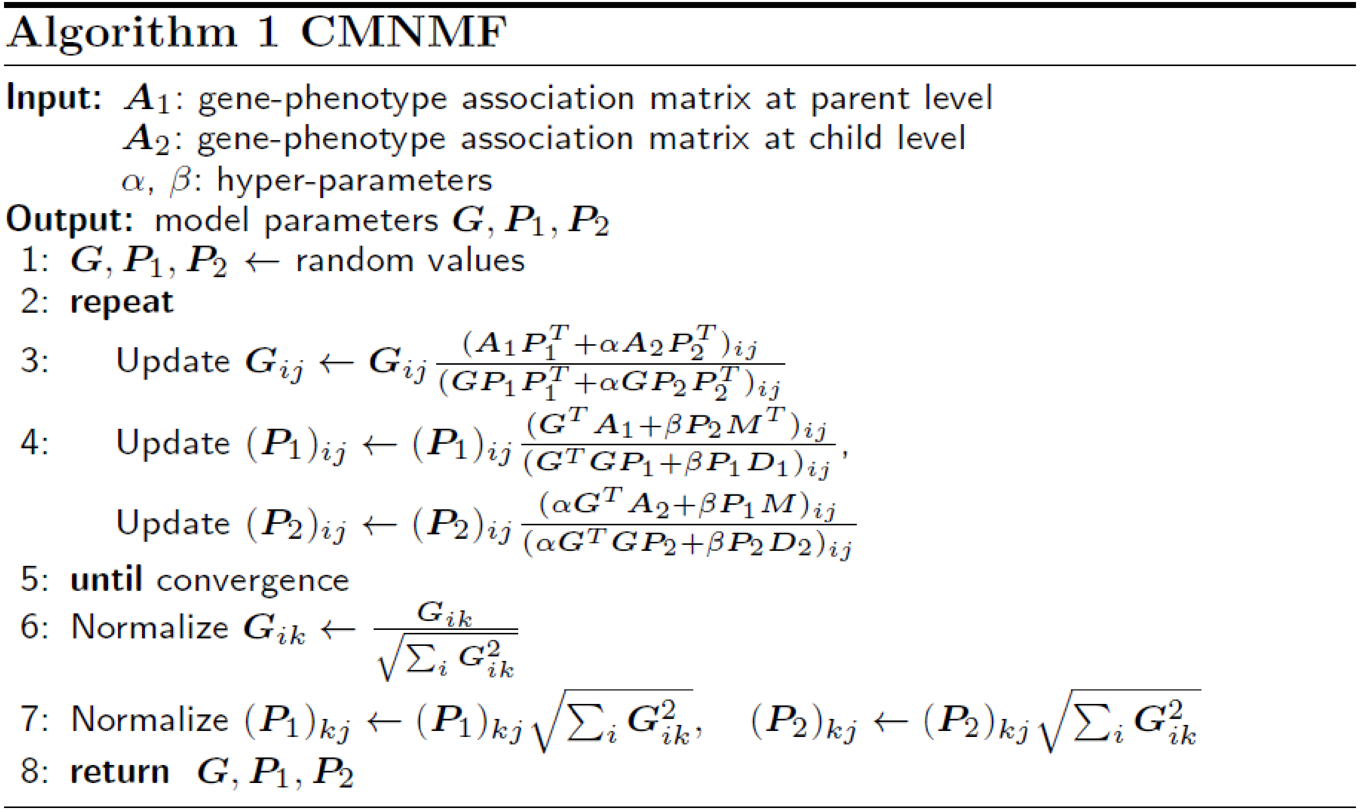
(c) Hierarchical mapping constraint on phenotype ontology

$$\min \sum_{ij} M_{ij} \| P_1^{(i)} - P_2^{(j)} \|^2$$

$$L = \|A_1 - GP_1\|_F^2 + \alpha \|A_2 - GP_2\|_F^2 + \beta \sum_{ij} M_{ij} \|P_1^{(i)} - P_2^{(j)}\|^2$$

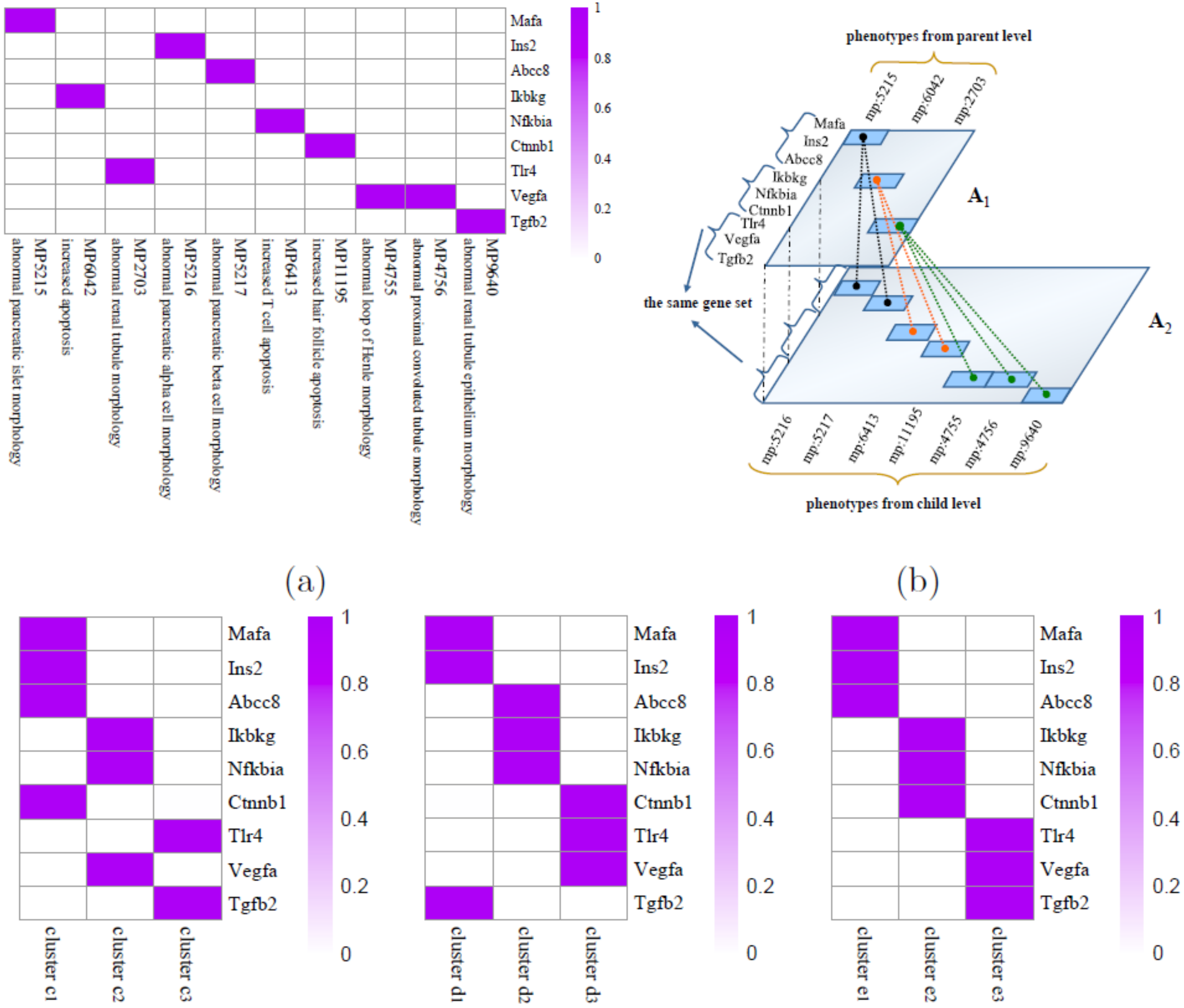
$$\text{s.t.} \quad G \geq 0, \; P_1 \geq 0, \; P_2 \geq 0$$

Algorithm



Algorithm 1 CMNMF

```
Input: $A_1$: gene-phenotype association matrix at parent level
       $A_2$: gene-phenotype association matrix at child level
       $\alpha$, $\beta$: hyper-parameters
Output: model parameters $G, P_1, P_2$
$G, P_1, P_2 \leftarrow$ random values
repeat
    Update $G_{ij} \leftarrow G_{ij} \frac{(A_1 P_1^T + \alpha A_2 P_2^T)_{ij}}{(G P_1 P_1^T + \alpha G P_2 P_2^T)_{ij}}$
    Update $(P_1)_{ij} \leftarrow (P_1)_{ij} \frac{(G^T A_1 + \beta P_2 M^T)_{ij}}{(G^T G P_1 + \beta P_1 D_1)_{ij}}$,
    Update $(P_2)_{ij} \leftarrow (P_2)_{ij} \frac{(\alpha G^T A_2 + \beta P_1 M)_{ij}}{(\alpha G^T G P_2 + \beta P_2 D_2)_{ij}}$
until convergence
Normalize $G_{ik} \leftarrow \frac{G_{ik}}{\sqrt{\sum_i G_{ik}^2}}$
Normalize $(P_1)_{kj} \leftarrow (P_1)_{kj} \sqrt{\sum_i G_{ik}^2}$,   $(P_2)_{kj} \leftarrow (P_2)_{kj} \sqrt{\sum_i G_{ik}^2}$
return $G, P_1, P_2$
```


Mafa
Ins2
Abcc8
Ikbkg
Nfkbia
Ctnnb1
Tlr4
Vegfa
Tgfb2
abnormal pancreatic islet morphology
MP5215
increased apoptosis
MP6042
abnormal renal tubule morphology
MP2703
abnormal pancreatic alpha cell morphology
MP5216
abnormal pancreatic beta cell morphology
MP5217
increased T cell apoptosis
MP6413
increased hair follicle apoptosis
MP11195
abnormal loop of Henle morphology
MP4755
abnormal proximal convoluted tubule morphology
MP4756
abnormal renal tubule epithelium morphology
MP9640
1
0.8
0.6
0.4
0.2
0
phenotypes from parent level
mp:5215
mp:6042
mp:2703
A1
A2
the same gene set
mp:5216
mp:5217
mp:6413
mp:11195
mp:4755
mp:4756
mp:9640
phenotypes from child level
cluster c1
cluster c2
cluster c3
cluster d1
cluster d2
cluster d3
cluster e1
cluster e2
cluster e3

(a)

(b)

(c)

(d)

(e)

Comparison

	VQ	PCA	NMF
Representation	holistic	holistic	Parts-based
Basis Image	Whole face	eigenfaces	Localized features
在W和H上的约束	each column of H is constrained to be a unary vector, every face is approximated by a single basis image.	W行正交，H列正交. each face is approximated by a linear combination of all the eigenfaces	Not allow negative entries in W and H. allow multiple basis images to represent a face, but only additive combinations