Multiple Relational Ranking in Tensor: Theory, Algorithms and Applications

Michael K. Ng

Centre for Mathematical Imaging and Vision and
Department of Mathematics
Hong Kong Baptist University
Email: mng@math.hkbu.edu.hk

21 July 2017

Outline

- Background
- Algorithms: PageRank and HITS
- Algorithms: MultiRank and HAR
- Numerical Examples: Information Retrieval, Community Network and Image Retrieval
- Transition Probability Tensors
- Theoretical Results
- Summary

Matrix

- Matrix can be used to describe the relationship between objects, and objects with different attributes:

	O_1	O_2	$\cdots$	O_n
O_1	$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,n}$
O_2	$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,n}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1}$	$a_{n,2}$	$\cdots$	$a_{n,n}$

	A_1	A_2	$\cdots$	A_m
O_1	$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,m}$
O_2	$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1}$	$a_{n,2}$	$\cdots$	$a_{n,m}$

  Examples: (left) a similarity matrix, an image, a Google matrix; (right) a gene expression data, multivariate data, terms and documents.

- large data (n is large); high-dimensional data (m is large)

Tensor

- Tensor can be used to describe the multiple relationships between objects. A tensor is a multidimensional array. Here a three-way array is used:

	O_1	O_2	$\cdots$	O_n
O_1	$a_{1,1,1}$	$a_{1,2,1}$	$\cdots$	$a_{1,n,1}$
O_2	$a_{2,1,1}$	$a_{2,2,1}$	$\cdots$	$a_{2,n,1}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1,1}$	$a_{n,2,1}$	$\cdots$	$a_{n,n,1}$

	O_1	O_2	$\cdots$	O_n
O_1	$a_{1,1,2}$	$a_{1,2,2}$	$\cdots$	$a_{1,n,2}$
O_2	$a_{2,1,2}$	$a_{2,2,2}$	$\cdots$	$a_{2,n,2}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1,2}$	$a_{n,2,2}$	$\cdots$	$a_{n,n,2}$

$\ldots$

	O_1	O_2	$\cdots$	O_n
O_1	$a_{1,1,p}$	$a_{1,2,p}$	$\cdots$	$a_{1,n,p}$
O_2	$a_{2,1,p}$	$a_{2,2,p}$	$\cdots$	$a_{2,n,p}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
O_n	$a_{n,1,p}$	$a_{n,2,p}$	$\cdots$	$a_{n,n,p}$

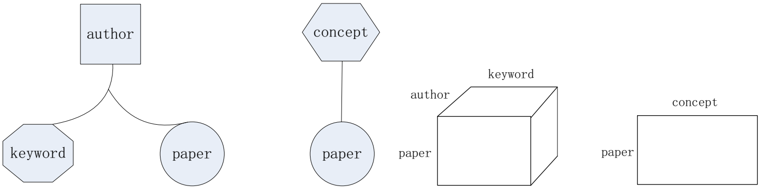
- p relationships among n objects

Tensor

- Web information retrieval is significantly more challenging than that based on web hyperlink structure
- One main difference is the multiple links based on the other features (text, images, etc)
- Example: 100,000 webpages from .GOV Web collection in 2002 TREC and 50 topic distillation topics in TREC 2003 Web track as queries
- Multiple links among webpages via different anchor texts
- 39,255 anchor terms (multiple relations), and 479,122 links with these anchor terms among the 100,000 webpages $(100000 \times 100000 \times 39255)$

Tensor

- In a social network where objects are connected via multiple relations, via sharing, comments, stories, photos, tags, keywords, topics, etc
- In a publication network where the interactions among items in three entities: author, keyword and paper



- A tensor: the interactions among items in three dimensions/entities: author, keyword and paper; A matrix: the interactions between items in two dimensions/entities: concept and paper

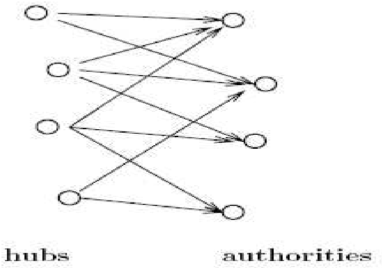
Information Retrieval

- The hyperlink structure is exploited by three of the most frequently cited Web IR methods: HITS (Hypertext Induced Topic Search), PageRank and SALSA.
- PageRank: L. Page, S. Brin, R. Motwani and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. 1998.
- HITS: J. Kleinberg. Authoritative Sources in a Hyperlinked Environment. Journal of the ACM, 46: 604-632, 1999.
- SALSA: R. Lempel and S. Moran. The Stochastic Approach for Link-structure Analysis (SALSA) and the TKC effect. The Ninth International WWW Conference, 2000.

[The survey given by A. Langville and C. Meyer, A Survey of Eigenvector Methods for Web Information Retrieval, SIAM Review, 2005.]

HITS

Each page/document on the Web is represented as a node in a very large graph. The directed arcs connecting these nodes represent the hyperlinks between the documents.



HITS

- The HITS IR method defines authorities and hubs. An authority is a document with several inlinks, while a hub has several outlinks.
- The HITS thesis is that good hubs point to good authorities and good authorities are pointed to by good hubs. HITS assigns both a hub score and authority score to a webpage.
- Webpage i has both a hub score x_i and an authority score y_i. Let E be the set of all directed edges in the Web graph and let $e_{i,j}$ represent the directed edge from node i to node j.

HITS

Given that each page has been assigned an initial hub score $x_i^{(0)}$ and authority score $y_i^{(0)}$, HITS successively refines these scores by computing

$$y_j^{(k)} = \sum_{i:e_{i,j}} x_i^{(k-1)} \quad \text{and} \quad x_i^{(k)} = \sum_{j:e_{i,j}} y_j^{(k-1)}$$

With the help of the adjacency matrix L of the directed Web graph: $L_{i,j} = 1$, if there exists an edge from node i to node j, $L_{i,j} = 0$, otherwise.

$$\mathbf{y}^{(k)} = L^T \mathbf{x}^{(k-1)} \quad \text{and} \quad \mathbf{x}^{(k)} = L\mathbf{y}^{(k)}$$

- These two equations define the iterative power method for computing the dominant eigenvector for the matrices L^TL and LL^T. Since the matrix L^TL determines the authority scores, it is called the authority matrix and LL^T is known as the hub matrix.
- L^TL and LL^T are symmetric positive semi-definite matrices. Computing the hub vector $\mathbf{x}$ and the authority vector $\mathbf{y}$ can be viewed as finding dominant right-hand eigenvectors of LL^T and L^TL, respectively.
- The structure of L allows to be a repeated root of the characteristic polynomial, in which case the associated eigenspace would be multi-dimensional. This means that the different limiting authority (and hub) vectors can be produced by different choices of the initial vector.

PageRank

- PageRank importance is determined by "votes" in the form of links from other pages on the web. The idea is that votes (links) from important sites should carry more weight than votes (links) from less important sites, and the significance of a vote (link) from any source should be tempered (or scaled) by the number of sites the source is voting for (linking to).
- The rank $r(s)$ of a given page s is

$$r(s) = \sum_{t \in B_s} \frac{r(t)}{\#(t)}$$

  where $B_s = \{$ all pages pointing to s $\}$, and $\#(t)$ is the number of out links from t.
- Compute

$$\mathbf{r} = P\mathbf{r} \quad \text{(column vector)}$$

  where P is the matrix with $p_{i,j} = 1/\#(s_j)$ if s_j links to s_i, otherwise $p_{i,j} = 0$.

PageRank

- The raw Google matrix P is nonnegative with column sums equal to one or zero.
- Zero column sums correspond to pages that that have no outlinks such pages are sometimes referred to as dangling nodes.
- Dangling nodes can be accounted for by artificially adding appropriate links to make all column sums equal one, then P is a column stochastic matrix, which in turn means that the PageRank iteration represents the evolution of a Markov chain.
- This Markov chain is a random walk on the graph defined by the link structure of the web pages in the Google database.

PageRank

An irreducible Markov chain is one in which every state is eventually reachable from every other state. That is, there exists a path from node j to node i for all i, j.

Irreducibility is a desirable property because it is precisely the feature that guarantees that a Markov chain possesses a unique (and positive) stationary distribution vector $\mathbf{r} = P\mathbf{r}$ $(\alpha = 1)$ or $\mathbf{r} = \tilde{P}\mathbf{r}$ $(0 < \alpha < 1)$, the Perron-Frobenius theorem at work:

$$\tilde{P} = \alpha P + (1 - \alpha)E$$

where E is a matrix of all ones by the number of pages. The matrix E can guarantee irreducible matrix $\tilde{P}$.

SALSA

- Stochastic Approach for Link Structure Analysis (SALSA). SALSA, developed by Lempel and Moran in 2000, was spawned by the combination of ideas from both HITS and PageRank.
- Like HITS, both hub and authority scores for webpages are created, and like PageRank, they are created through the use of Markov chains:

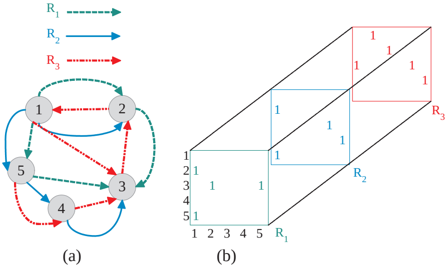
  $$\mathbf{y}^{(k)} = Normalize(L^T)\mathbf{x}^{(k-1)} \quad \text{and} \quad \mathbf{x}^{(k)} = Normalize(L)\mathbf{y}^{(k)}$$

  Both $Normalize(L^T)$ and $Normalize(L)$ are transition probability matrices (column sum is equal to 1).

Multi-relation Data: The Representation

Example: five objects and three relations (R1: green, R2: blue, R3: red) among them.

A tensor is a multidimensional array. Example: a three-way array is used, where each two dimensional slice represents an adjacency matrix for a single relation. The data can be represented as a tensor of size $5 \times 5 \times 3$ where (i, j, k) entry is nonzero if the ith object is connected to the jth object by using the kth relation.



Notation

Let $\mathbb{R}$ be the real field. We call $\mathcal{T} = (t_{i_1,i_2,j_1})$ where $t_{i_1,i_2,j_1} \in \mathbb{R}$, for $i_k = 1, \cdots, m$, $k = 1, 2$ and $j_1 = 1, \cdots, n$, a real $(2, 1)$th order $(m \times n)$-dimensional rectangular tensor.

(i_1, i_2) to be the indices for objects and j_1 to be the indices for relations.

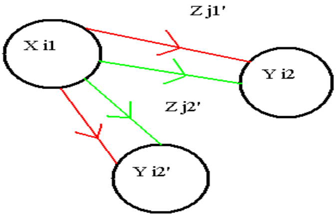
For instance, five objects $(m = 5)$ and three relations $(n = 3)$ are used in the example. When there is a link from the i_1th object to the i_2th object when the j_1th relation is used, we set $t_{i_1,i_2,j_1} = 1$, otherwise $t_{i_1,i_2,j_1} = 0$.

$\mathcal{T}$ is called non-negative if $t_{i_1,i_2,j_1} \geq 0$.

The Contribution

- Propose a framework to study the hub and authority scores of objects in multi-relational data for query search.
- Besides hub score $\mathbf{x}$ and authority score $\mathbf{y}$ for each object, we assign a relevance score $\mathbf{z}$ for each relation to indicate its importance in multi-relational data.
- These three scores have the following mutually-reinforcing idea:
  1. An object that points to many objects with high authority scores through relations of high relevance scores, receives a high hub score.
  2. An object that is pointed by many objects with high hub scores through relations of high relevance scores, receives a high authority score.
  3. A relation that is connected in between objects with high hub scores and high authority scores, receives a high relevance score.

The Idea



$$\sum_{i_2=1}^{m}\sum_{j_1=1}^{n} h_{i_1,i_2,j_1} y_{i_2} z_{j_1} = x_{i_1}, \quad 1 \le i_1 \le m$$

$$\sum_{i_1=1}^{m}\sum_{j_1=1}^{n} a_{i_1,i_2,j_1} x_{i_1} z_{j_1} = y_{i_2}, \quad 1 \le i_2 \le m$$

$$\sum_{i_1=1}^{m}\sum_{i_2=1}^{m} r_{i_1,i_2,j_1} x_{i_1} y_{i_2} = z_{j_1}, \quad 1 \le j_2 \le n$$

The HAR Model

Transition Probability Tensors:
$\mathcal{H} = (h_{i_1,i_2,j_1})$, $\mathcal{A} = (a_{i_1,i_2,j_1})$ and $\mathcal{R} = (r_{i_1,i_2,j_1})$ with respect to hubs, authorities and relations by normalizing the entry of $\mathcal{T}$ as follows:

$$\begin{aligned}
h_{i_1,i_2,j_1} &= \frac{t_{i_1,i_2,j_1}}{\sum_{i_1=1}^{m} t_{i_1,i_2,j_1}}, \quad i_1 = 1, 2, \cdots, m, \\
a_{i_1,i_2,j_1} &= \frac{t_{i_1,i_2,j_1}}{\sum_{i_2=1}^{m} t_{i_1,i_2,j_1}}, \quad i_2 = 1, 2, \cdots, m, \\
r_{i_1,i_2,j_1} &= \frac{t_{i_1,i_2,j_1}}{\sum_{j_1=1}^{n} t_{i_1,i_2,j_1}}, \quad j_1 = 1, 2, \cdots, n.
\end{aligned}$$

The Interpretation

- To compute hub and authority scores of objects and relevance scores of relations by considering a random walk in a multi-relational data/tensor, and studying the limiting probabilities having objects as hubs or as authorities, and using relations respectively.
- h_{i_1,i_2,j_1} (or a_{i_1,i_2,j_1}) can be interpreted as the probability of having the i_1th (or i_2th) object as an hub (or as an authority) by given that the i_2th (or i_1th) object as an authority (or as a hub) is currently considered and the j_1th relation is used;
- r_{i_1,i_2,j_1} can be interpreted as the probability of using the j_1th relation given that the i_2th object as an authority is visited from the i_1th object as a hub.

The Interpretation

$$\begin{aligned} h_{i_1,i_2,j_1} &= \text{Prob}[X_t = i_1 | Y_t = i_2, Z_t = j_1] \\ a_{i_1,i_2,j_1} &= \text{Prob}[Y_t = i_2 | X_t = i_1, Z_t = j_1] \\ r_{i_1,i_2,j_1} &= \text{Prob}[Z_t = j_1 | Y_t = i_2, X_t = i_1] \end{aligned}$$

X_t, Y_t and Z_t are random variables referring to consider at any particular object as a hub and as an authority, and to use at any particular relation respectively at the time t respectively. Here the time t refers to the time step in a Markov chain (higher order generalization).

The construction of a_{i_1,i_2,j_1} is related to the transpose of h_{i_1,i_2,j_1}. This is similar to the construction of SALSA algorithm to incorporate the link structure among objects for the role of hub and authority in the single relation data.

HAR - Tensor Equations

hub score: $\bar{\mathbf{x}}$
authority score: $\bar{\mathbf{y}}$
relevance score: $\bar{\mathbf{z}}$

$$\mathcal{H}\bar{\mathbf{y}}\bar{\mathbf{z}} = \bar{\mathbf{x}}, \quad \mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} = \bar{\mathbf{y}}, \quad \mathcal{R}\bar{\mathbf{x}}\bar{\mathbf{y}} = \bar{\mathbf{z}},$$

with

$$\sum_{i_1=1}^{m} \bar{x}_{i_1} = 1, \quad \sum_{i_2=1}^{m} \bar{y}_{i_2} = 1, \quad \sum_{j_1=1}^{n} \bar{z}_{j_1} = 1.$$

hub and authority scores for objects and relevance scores for relations by solving tensor equations based on mutually-reinforcing relationship among hubs, authorities and relations.

Generalization

When we consider a single relation type, we can set $\bar{\mathbf{z}}$ to be a vector $\mathbf{1}/n$ of all ones, and thus we obtain two matrix equations

$$\mathcal{H}\bar{\mathbf{y}}\mathbf{1}/n = \bar{\mathbf{x}} \quad \mathcal{A}\bar{\mathbf{x}}\mathbf{1}/n = \bar{\mathbf{y}}.$$

We remark that $\mathcal{A}$ can be viewed as the transpose of $\mathcal{H}$.
This is exactly the same as that we solve for the singular vectors to get the hub and authority scoring vectors in SALSA. As a summary, the proposed framework HAR is a generalization of SALSA to deal with multi-relational data.

The MultiRank Model

Transition Probability Tensors:
$\mathcal{A} = (a_{i_1,i_2,j_1})$ and $\mathcal{R} = (r_{i_1,i_2,j_1})$ with respect to nodes and relations by normalizing the entry of $\mathcal{T}$ as follows:

$$a_{i_1,i_2,j_1} = \frac{t_{i_1,i_2,j_1}}{\sum_{i_2=1}^{m} t_{i_1,i_2,j_1}}, \quad i_2 = 1, 2, \cdots, m,$$

$$r_{i_1,i_2,j_1} = \frac{t_{i_1,i_2,j_1}}{\sum_{j_1=1}^{n} t_{i_1,i_2,j_1}}, \quad j_1 = 1, 2, \cdots, n.$$

$$\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} = \bar{\mathbf{x}}, \quad \mathcal{R}\bar{\mathbf{x}}^2 = \bar{\mathbf{z}},$$

with

$$\sum_{i}^{m} \bar{x}_i = 1, \quad \sum_{j}^{n} \bar{z}_j = 1.$$

HAR - Query Search

To deal with query processing, we need to compute hub and authority scores of objects and relevance scores of relations with respect to a query input (like topic-sensitive PageRank):

$$
\begin{aligned}
(1-\alpha)\mathcal{H}\bar{\mathbf{y}}\bar{\mathbf{z}} + \alpha\mathbf{o} &= \bar{\mathbf{x}}, \\
(1-\beta)\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} + \beta\mathbf{o} &= \bar{\mathbf{y}}, \\
(1-\gamma)\mathcal{R}\bar{\mathbf{x}}\bar{\mathbf{y}} + \gamma\mathbf{r} &= \bar{\mathbf{z}},
\end{aligned}
$$

where $\mathbf{o}$ and $\mathbf{r}$ are two assigned probability distributions that are constructed from a query input, and $0 \leq \alpha, \beta, \gamma < 1$, are three parameters to control the input.

$$F(\mathbf{x},\mathbf{y},\mathbf{z}) = \begin{pmatrix} (1-\alpha)\mathcal{H}\bar{\mathbf{y}}\bar{\mathbf{z}} + \alpha\mathbf{o} \\ (1-\beta)\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} + \beta\mathbf{o} \\ (1-\gamma)\mathcal{R}\bar{\mathbf{x}}\bar{\mathbf{y}} + \gamma\mathbf{r} \end{pmatrix}$$

$$\Omega_m = \{\mathbf{u} = (u_1, u_2, \cdots, u_m) \in R^m | u_i \geq 0, 1 \leq i \leq m, \sum_{i=1}^{m} u_i = 1\}$$

and

$$\Omega_n = \{\mathbf{w} = (w_1, w_2, \cdots, w_n) \in R^n | w_j \geq 0, 1 \leq j \leq n, \sum_{j=1}^{n} w_j = 1\}$$

HAR - Theory

Theorem

Suppose $\mathcal{H}$, $\mathcal{A}$ and $\mathcal{R}$ are constructed from $\mathcal{T}$, $0 \leq \alpha$, β, $\gamma < 1$, and $\mathbf{o} \in \Omega_m$ and $\mathbf{r} \in \Omega_n$ are given. If $\mathcal{T}$ is irreducible, then there exist $\bar{\mathbf{x}} > 0$, $\bar{\mathbf{y}} > 0$ and $\bar{\mathbf{z}} > 0$ such that $(1-\alpha)\mathcal{H}\bar{\mathbf{y}}\bar{\mathbf{z}} + \alpha\mathbf{o} = \bar{\mathbf{x}}$, $(1-\beta)\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} + \beta\mathbf{o} = \bar{\mathbf{y}}$, and $(1-\gamma)\mathcal{R}\bar{\mathbf{x}}\bar{\mathbf{y}} + \gamma\mathbf{r} = \bar{\mathbf{z}}$, with $\bar{\mathbf{x}}, \bar{\mathbf{y}} \in \Omega_m$ and $\bar{\mathbf{z}} \in \Omega_n$.

Using Brouwer's Fixed Point Theorem.

Theorem

Suppose $\mathcal{T}$ is irreducible, and $\mathcal{H}$, $\mathcal{A}$ and $\mathcal{R}$ are constructed from $\mathcal{T}$, $0 \leq \alpha, \beta, \gamma < 1$ and $\mathbf{o} \in \Omega_m$ and $\mathbf{r} \in \Omega_n$ are given. If 1 is not the eigenvalue of the Jacobian matrix of the mapping F, then the solution vectors $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$ and $\bar{\mathbf{z}}$ are unique.

Using Kellogg's Unique Fixed Point Results.

HAR - Theory ($\alpha, \beta, \gamma > 1/2$)

Theorem

Suppose $\mathcal{H}$, $\mathcal{A}$ and $\mathcal{R}$ are constructed from $\mathcal{T}$, and $\mathbf{o} \in \Omega_m$ and $\mathbf{r} \in \Omega_n$ are given. If $1/2 < \alpha, \beta, \gamma < 1$ then the solution vectors $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$ and $\bar{\mathbf{z}}$ are unique.

Check the Jacobian matrix

$$J = \begin{pmatrix} 0 & (1-\alpha)J_{12} & (1-\alpha)J_{13} \\ (1-\beta)J_{21} & 0 & (1-\beta)J_{23} \\ (1-\gamma)J_{31} & (1-\gamma)J_{32} & 0 \end{pmatrix}$$

J_{st} are transition probability matrices (their column sum are equal to 1). The largest magnitude of the eigenvalue of J is less than 1.

The HAR Algorithm

Input: Three tensors $\mathcal{H}$, $\mathcal{A}$ and $\mathcal{R}$, two initial probability distributions $\mathbf{y}_0$ and $\mathbf{z}_0$ with $(\sum_{i=1}^{m}[\mathbf{y}_0]_i = 1$ and $\sum_{j=1}^{n}[\mathbf{z}_0]_j = 1)$, the assigned probability distributions of objects and/or relations $\mathbf{o}$ and $\mathbf{r}$ $(\sum_{i=1}^{m}[\mathbf{o}]_i = 1$ and $\sum_{j=1}^{n}[\mathbf{r}]_j = 1)$, three weighting parameters $0 \le \alpha, \beta, \gamma < 1$, and the tolerance ϵ
Output: Three limiting probability distributions $\bar{\mathbf{x}}$ (authority scores), $\bar{\mathbf{y}}$ (hub scores) and $\bar{\mathbf{z}}$ (relevance values)
Procedure:

1: Set $t = 1$;
2: Compute $\mathbf{x}_t = (1-\alpha)\mathcal{H}\mathbf{y}_{t-1}\mathbf{z}_{t-1} + \alpha\mathbf{o}$;
3: Compute $\mathbf{y}_t = (1-\beta)\mathcal{A}\mathbf{x}_{t-1}\mathbf{z}_{t-1} + \beta\mathbf{o}$;
4: Compute $\mathbf{z}_t = (1-\gamma)\mathcal{R}\mathbf{x}_{t-1}\mathbf{y}_{t-1} + \gamma\mathbf{r}$;
5: If $||\mathbf{x}_t - \mathbf{x}_{t-1}|| + ||\mathbf{y}_t - \mathbf{y}_{t-1}|| + ||\mathbf{z}_t - \mathbf{z}_{t-1}|| < \epsilon$, then stop, otherwise set $t = t + 1$ and goto Step 2.

Convergence $(\alpha, \beta, \gamma > 1/2)$

Theorem

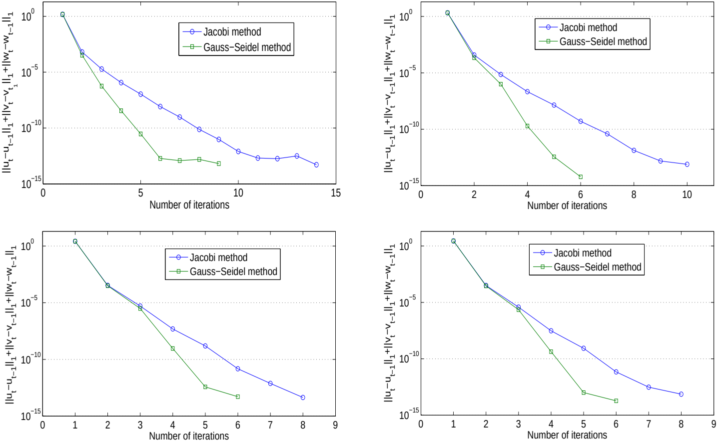
Suppose $\mathcal{H}$, $\mathcal{A}$ and $\mathcal{R}$ constructed, and $\mathbf{o} \in \Omega_m$ and $\mathbf{r} \in \Omega_n$ are given. If $1/2 < \alpha, \beta, \gamma < 1$ then the HAR algorithm converges to the unique vectors $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$ and $\bar{\mathbf{z}}$.

Gauss Seidel iterations:

Compute $\mathbf{x}_t = (1 - \alpha)\mathcal{H}\mathbf{y}_{t-1}\mathbf{z}_{t-1} + \alpha\mathbf{o}$;
Compute $\mathbf{y}_t = (1 - \beta)\mathcal{A}\mathbf{x}_t\mathbf{z}_{t-1} + \beta\mathbf{o}$;
Compute $\mathbf{z}_t = (1 - \gamma)\mathcal{R}\mathbf{x}_t\mathbf{y}_t + \gamma\mathbf{r}$;

Gauss Seidel is faster than Jacobi

Example in Information Retreival:



Experiment 1

- 100,000 webpages from .GOV Web collection in 2002 TREC and 50 topic distillation topics in TREC 2003 Web track as queries
- Links among webpages via different anchor texts
- 39,255 anchor terms (multiple relations), and 479,122 links with these anchor terms among the 100,000 webpages
- If the i_2th webpage links to the i_1th webpage via the j_1th anchor term, we set the entry t_{i_1,i_2,j_1} of $\mathcal{T}$ to be one. The size of $\mathcal{T}$ is $100,000 \times 100,000 \times 39,255$
- Sparse tensors: the percentage of nonzero entries is 1.22×10^{-7}%

Evaluation

- P@k: Given a particular query, we compute the precision at position k as follows: P@k $=$ $\#$ relevant documents in top k results / k
- NDCG@k the normalized version of DCG@k that discounts the contribution of low-ranking relevant documents
- MAP: Given a query, the average precision is calculated by averaging the precision scores at each position in the search results where a relevant document is found. MAP is then the mean of the average precision scores of all queries.
- R-prec is the precision score after R documents are retrieved, i.e., R-prec=P@R, where R is the total number of relevant documents for such query.

	P@10	P@20	NDCG@10	NDCG@20	MAP	R-prec
HITS	0.0000	0.0000	0.0000	0.0000	0.0041	0.0000
SALSA	0.0160	0.0140	0.0157	0.0203	0.0114	0.0084
TOPHITS (500-rank)	0.0020	0.0010	0.0044	0.0028	0.0008	0.0002
TOPHITS (1000-rank)	0.0040	0.0020	0.0088	0.0057	0.0016	0.0010
TOPHITS (1500-rank)	0.0040	0.0030	0.0063	0.0049	0.0011	0.0018
BM25+ DepInOut	0.0280	0.0180	0.0419	0.0479	0.0370	0.0370
HAR (rel. query)	0.0560	0.0410	0.0659	0.0747	0.0330	0.0552
HAR (rel. and obj. query)	**0.1100**	**0.0800**	**0.1545**	**0.1765**	**0.1035**	**0.1051**

The results of all comparison algorithms on TREC data set.

ranks	HAR
1	**www.tempe.gov/library/youth/teach.htm**
2	**www.get.wa.gov/reading.htm**
3	www.tempe.gov/library/content.htm
4	www.ed.gov/pubs/Paraprofessionals/norfolk.html
5	www.loc.gov/rr/international/int-gateway.html
6	**graham.sannet.gov/public-library/searching-the-net/childrensites.shtml**
7	**www.cde.ca.gov/ci/literature/**
8	www.cde.ca.gov/news/releases2001/rel36.asp
9	www.hud.gov/lea/leastand.html/
10	libwww.library.phila.gov/databases/keywordsearch.taf

ranks	TOPHITS
1	www.students.gov/link_search/listlinks.cfm?cfid=1139239&cftoken=44790442&Topic=0101
2	www.students.gov/link_search/listlinks.cfm?cfid=1139177&cftoken=3776877&Topic=0404
3	www.crh.noaa.gov/dlh/firewx.htm
4	www.students.gov/link_search/listlinks.cfm?cfid=1139251&cftoken=50170076&Topic=0101
5	www.bls.gov/oes/2000/oestec2000.htm
6	www.students.gov/link_search/listlinks.cfm?cfid=1139205&cftoken=97602529&Topic=0203
7	www.students.gov/link_search/listlinks.cfm?cfid=1130081&cftoken=49786299&Topic=0101
8	www.students.gov/link_search/listlinks.cfm?cfid=1139251&cftoken=50170076&Topic=0103
9	www.crh.noaa.gov/sgf/hydro/reports/hydrorpt.shtml
10	graham.sannet.gov/directories/privacy.shtml

ranks	HAR
1	**www.npwrc.usgs.gov/**
2	pastel.npsc.nbs.gov/resource/2001/impplan/appenE.htm
3	**water.usgs.gov/eap/env_guide/fish_wildlife.html**
4	**fedlaw.gsa.gov/legal2a.htm**
5	**envirotext.eh.doe.gov/data/uscode/16/**
6	envirotext.eh.doe.gov/data/uscode/16/ch5A.html
7	envirotext.eh.doe.gov/data/uscode/16/ch6.html
8	www.nwr.noaa.gov/1salmon/salmesa/pubs.htm
9	www.dfg.ca.gov/hcpb/conplan/mitbank/banking_report.shtml
10	**www.hawaii.gov/dlnr/IdxConservation.htm**

ranks	TOPHITS
1	**www.npwrc.usgs.gov/**
2	www.students.gov/link_search/listlinks.cfm?cfid=1133639&cftoken=90444830&Topic=0101
3	www.students.gov/link_search/listlinks.cfm?cfid=1133456&cftoken=20839758&Topic=0101
4	www.fedstats.gov/qf/states/29000.html
5	www.students.gov/link_search/listlinks.cfm?cfid=1133457&cftoken=22178553&Topic=0404
6	www.students.gov/link_search/listlinks.cfm?cfid=1139247&cftoken=47516940&Topic=0101
7	www.crh.noaa.gov/sgf/products/mtruno.shtml
8	www.students.gov/link_search/listlinks.cfm?cfid=1139201&cftoken=34746433&Topic=0101
9	www.students.gov/link_search/listlinks.cfm?cfid=1130149&cftoken=13070299&Topic=0101
10	www.students.gov/link_search/listlinks.cfm?cfid=1139266&cftoken=37252845&Topic=0203

ranks	HAR
1	usembassy.state.gov/seoul/wwwh42x4.html
2	**usembassy.state.gov/seoul/wwwhe404.html**
3	**www.state.gov/r/pa/bgn/2792pf.htm**
4	**www.state.gov/r/pa/bgn/2792.htm**
5	**www.voa.gov/korean/**
6	www.st.nmfs.gov/st3/multilateral.html
7	www.usinfo.state.gov/regional/ea/easec/nkoreapg.htm
8	usembassy.state.gov/seoul/wwwh42yt.html
9	usembassy.state.gov/seoul/wwwh0108.html
10	citrus.sbaonline.sba.gov/nd/ndoutreach.html

ranks	TOPHITS
1	www.students.gov/link_search/listlinks.cfm?cfid=1139239&cftoken=44790442&Topic=0101
2	www.students.gov/link_search/listlinks.cfm?cfid=1139177&cftoken=3776877&Topic=0404
3	www.crh.noaa.gov/dlh/firewx.htm
4	www.students.gov/link_search/listlinks.cfm?cfid=1139251&cftoken=50170076&Topic=0101
5	www.bls.gov/oes/2000/oestec2000.htm
6	www.students.gov/link_search/listlinks.cfm?cfid=1139205&cftoken=97602529&Topic=0203
7	www.students.gov/link_search/listlinks.cfm?cfid=1130081&cftoken=49786299&Topic=0101
8	www.students.gov/link_search/listlinks.cfm?cfid=1139251&cftoken=50170076&Topic=0103
9	www.crh.noaa.gov/sgf/hydro/reports/hydrorpt.shtm
10	graham.sannet.gov/directories/privacy.shtml

Experiment 2

1. five conferences (SIGKDD, WWW, SIGIR, SIGMOD, CIKM)
2. Publication information includes title, authors, reference list, and classification categories associated with publication
3. 6848 publications and 617 different categories
4. 100 category concepts as query inputs to retrieve the relevant publications
5. Tensor: $6848 \times 6848 \times 617$, If the i_2th publication cites the i_1th publication and the i_1th publication has the j_1th category concept, then we set the entry t_{i_1,i_2,j_1} of $\mathcal{T}$ to be one, otherwise we set the entry t_{i_1,i_2,j_1} to be zero.
6. 24901 nonzero entries, the percentage of the nonzero entries is $8.61 \times 10^{-5}\%$

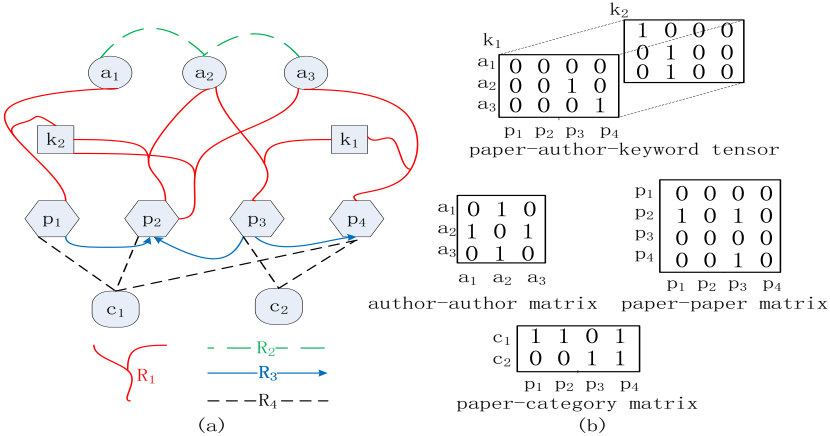
	P@10	P@20	NDCG@10	NDCG@20	MAP	R-prec
HITS	0.2260	0.1815	0.3789	0.3792	0.2522	0.2751
SALSA	0.4100	0.3105	0.5606	0.5352	0.3462	0.3929
TOPHITS (50-rank)	0.1360	0.1145	0.1684	0.1557	0.0566	0.0617
TOPHITS (100-rank)	0.1640	0.1340	0.2012	0.1857	0.0646	0.0732
TOPHITS (150-rank)	0.1920	0.1410	0.2315	0.1998	0.0732	0.0765
BM25+ DepInOut	0.0170	0.0145	0.0147	0.0138	0.0162	0.0109
HAR (rel. query)	**0.5880**	**0.4155**	**0.7472**	**0.6760**	**0.4731**	**0.4683**

The results of all comparison algorithms on DBLP data set.

	P@10	P@20	NDCG@10	NDCG@20	MAP	R-prec
HAR (rel. query)	0.2000	0.1312	0.1488	0.1647	0.1312	0.2075
HAR (rel. and (obj. query)	**0.6000**	**0.7500**	**0.6995**	**0.7697**	**0.5422**	**0.6226**

The results of HAR with two settings when we query "clustering" concept and "document" related papers. In this case, we judge a paper to be relevant if it has "clustering" concept and a "document" related title for evaluation.

Community Discovery



(a) An example of an academic publication network where R_1 indicates paper-author-keyword interaction, R_2 indicates author-author collaboration, R_3 indicates paper-paper citation, R_4 indicates paper-category interaction. (b) One tensor and three matrices are used to represent the interactions among authors, papers, keywords and categories.

For SIAM journal data, we consider the papers published in SIMAX, SISC and SINUM in the recent 15 years

Multi-dimensional networks:
the paper-author-keyword (x_1-x_2-x_3) tensor $\mathcal{P}^{(1)}$ $[3736 \times 1807 \times 7630; nz = 31257\ 6.07 \times 10^{-5}$%];
the paper-paper (x_1-x_1) citation tensor $\mathcal{P}^{(2)}$ $[3736 \times 3736;$ $nz = 4523$ 0.032%] ;
the author-author (x_2-x_2) collaboration tensor $\mathcal{P}^{(3)}$ $[1807 \times 1807;$ $nz = 4410$ 0.14%];
the paper-category (x_1-x_4) concept tensor $\mathcal{P}^{(4)}$ $[3736 \times 1202;$ $nz = 11897$ 0.26%]

$\mathcal{P}^{(1)} \rightarrow \mathcal{P}^{(1,1)}, \mathcal{P}^{(1,2)}, \mathcal{P}^{(1,3)}$
$\mathcal{P}^{(2)} \rightarrow \mathcal{P}^{(2,1)}$
$\mathcal{P}^{(3)} \rightarrow \mathcal{P}^{(3,1)}$
$\mathcal{P}^{(4)} \rightarrow \mathcal{P}^{(4,1)}, \mathcal{P}^{(4,2)}$

Multi-dimensional networks tensor equations:

$$
\begin{aligned}
\text{(paper) } \mathbf{x}_1 &= (1-\alpha)\left[\frac{1}{3}\mathcal{P}^{(1,1)}\mathbf{x}_2\mathbf{x}_3 + \frac{1}{3}\mathcal{P}^{(2,1)}\mathbf{x}_1 + \frac{1}{3}\mathcal{P}^{(4,1)}\mathbf{x}_4\right] + \\
&\quad \alpha\mathbf{z}_1 \\
\text{(author) } \mathbf{x}_2 &= (1-\alpha)\left[\frac{1}{2}\mathcal{P}^{(1,2)}\mathbf{x}_1\mathbf{x}_3 + \frac{1}{2}\mathcal{P}^{(3,1)}\mathbf{x}_2\right] + \alpha\mathbf{z}_2 \\
\text{(keyword) } \mathbf{x}_3 &= (1-\alpha)\mathcal{P}^{(1,3)}\mathbf{x}_1\mathbf{x}_2 + \alpha\mathbf{z}_3 \\
\text{(category) } \mathbf{x}_4 &= (1-\alpha)\mathcal{P}^{(4,2)}\mathbf{x}_1 + \alpha\mathbf{z}_4
\end{aligned}
$$

z_1, z_2, z_3, z_4 are seeds for papers, authors, keywords and categories respectively.

Numerical Example

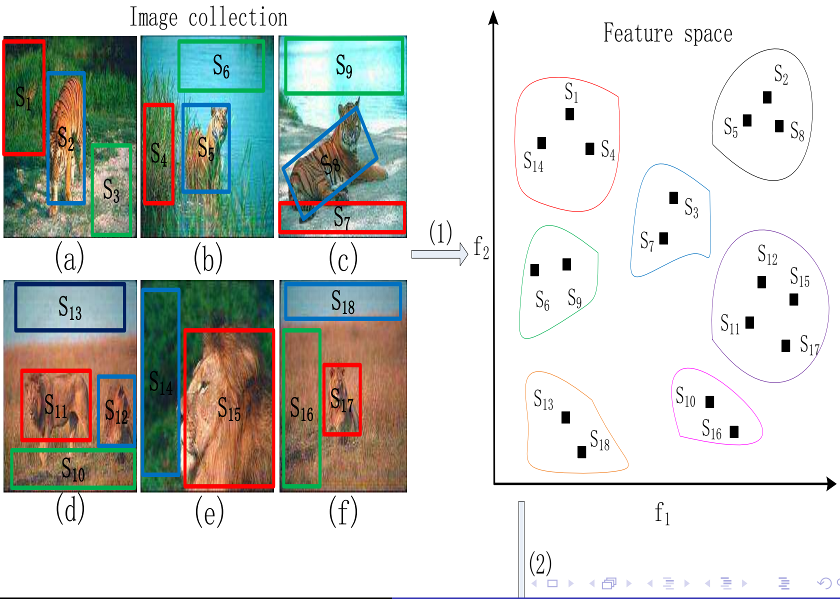
Community	MetaFac	Clauset	LWP	MultiComm
SIMAX	0.3544	0.0612	0.0022	**0.6396**
SINUM	0.4591	0.4321	0.0001	**0.5113**
SISC	0.3868	0.3047	0.0056	**0.4321**
average F-measure	0.4001	0.2660	0.0026	**0.5277**
KDD	0.5741	0.7662	0.0020	**0.7722**
SIGIR	0.6497	0.6664	0.0015	**0.7350**
average F-measure	0.6119	0.7163	0.0018	**0.7536**

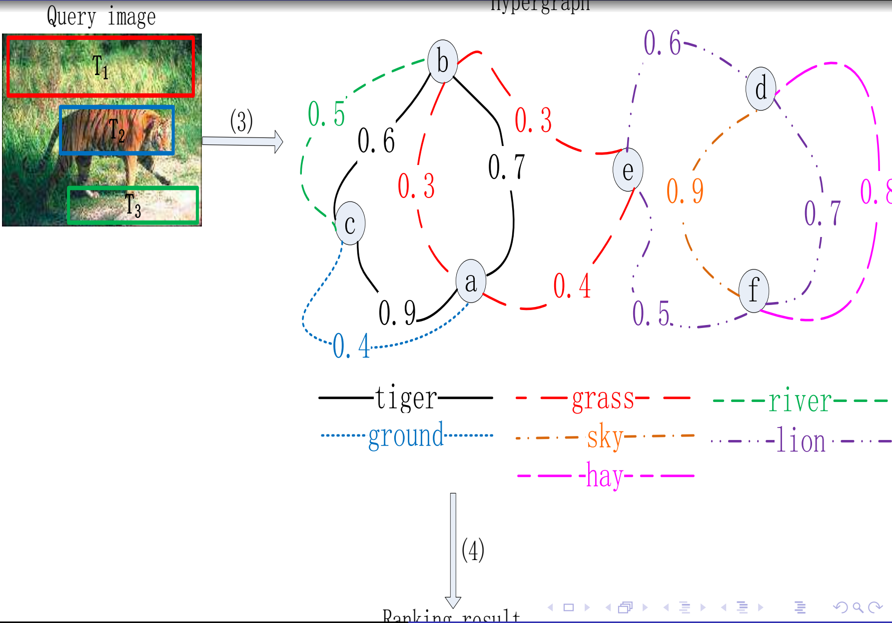
Community(AMS code)	MetaFac	Clauset	LWP	MultiComm
65F10	0.2123	0.1143	0.0090	**0.4602**
65F15	0.1641	0.2197	0.0222	**0.4282**
65N30	0.2133	0.0701	0.0059	**0.3672**
65N15	0.1341	0.0802	0.0152	**0.3244**
65N55	0.1743	0.2555	0.0001	**0.3321**
average F-measure	0.1796	0.1480	0.0105	**0.3824**

Image Retrieval

- An image can be represented by several visual concepts, and a tensor is built based on visual concepts as its entry contain the affinity between two images in the corresponding visual concept.
- A ranking scheme can be used to compute the association scores of images and the relevance scores of visual concepts by employing input query vectors to handle image retrieval: MultiRank and HAR algorithms.

Image Retrieval





Query image
T_1
T_2
T_3
(3)
Hypergraph
b
c
a
e
d
f
0.5
0.6
0.3
0.3
0.7
0.9
0.4
0.4
0.6
0.9
0.7
0.5
tiger
grass
river
ground
sky
lion
hay
(4)

Image Retrieval

ground sky lion
hay

(4)

Ranking result

(a) score=0. 3908 | (b) score=0. 2238 | (c) score=0. 1588 | (e) score=0. 1362 | (d) score=0. 0452 | (f) score=0. 0452

Image Retrieval (MSRC data)

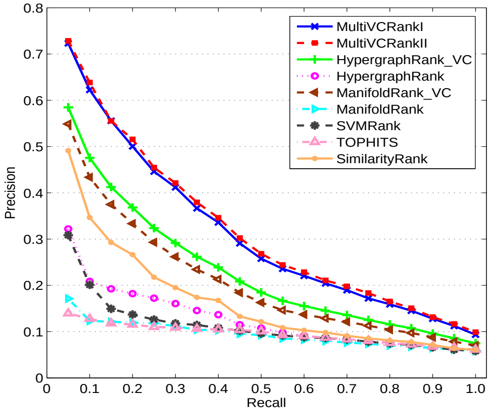


Image Retrieval (Corel data)

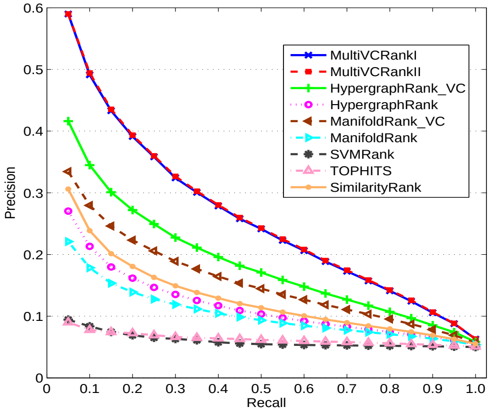
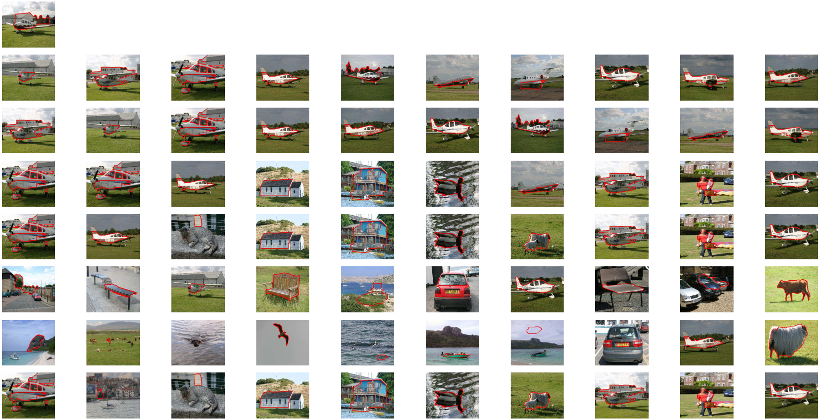
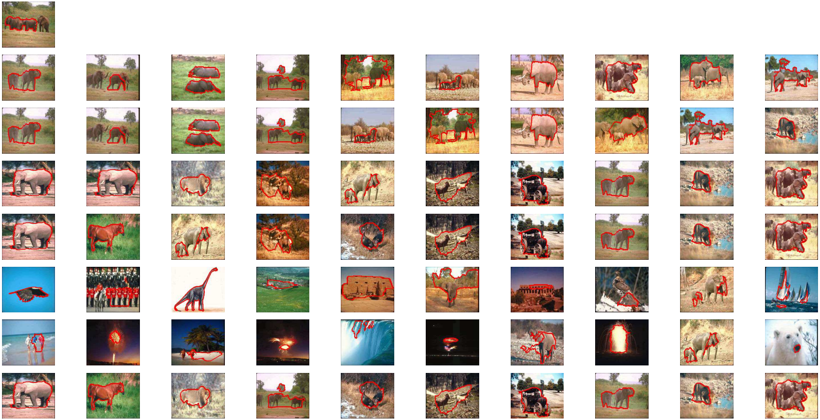


Image Retrieval



The top ten results of different algorithms on MSRC data for a query image. The red lines mark the salient regions. The first row is the query image, and the second, third, fourth, fifth, sixth, seventh and eighth rows show the results of MultiVCRankI, MultiVCRankII, HypergraphRank_VC, ManifoldRank_VC, RankSVM, TOPHITS, and SimlarityRank, respectively.



The top ten results of different algorithms on Corel data for a query image. The red lines mark the salient regions produced by the algorithm in [?]. The first row is the query image, and the second, third, fourth, fifth, sixth, seventh and eighth rows show the results of MultiVCRankI, MultiVCRankII, HypergraphRank_VC, ManifoldRank_VC, RankSVM, TOPHITS, and SimlarityRank, respectively.

The performance comparison in terms of Precisoin@k and NDCG@k on MSRC and Corel datasets.

	MSRC				Corel			
	P@5	P@10	NDCG@5	NDCG@10	P@5	P@10	NDCG@5	NDCG@10
MultiVCRankI	0.4867	0.4506	0.5271	0.4715	0.5417	0.4860	0.5699	0.5214
MultiVCRankII	0.5000	0.4683	0.5327	0.4708	0.5433	0.4887	0.5717	0.5237
HypergraphRank	0.3521	0.3196	0.4013	0.3512	0.4021	0.3521	0.4286	0.4073
ManifoldRank	0.3267	0.2983	0.3892	0.3393	0.3505	0.3076	0.3759	0.3837
SVMRank	0.1433	0.1183	0.1485	0.1280	0.1015	0.0840	0.1101	0.1021
TOPHITS	0.1301	0.1083	0.1521	0.1122	0.1086	0.0819	0.1027	0.0964
SimilarityRank	0.2633	0.2367	0.3005	0.2696	0.2479	0.2046	0.2565	0.2161

Image Retrieval with Feedback Training (MSRC data)

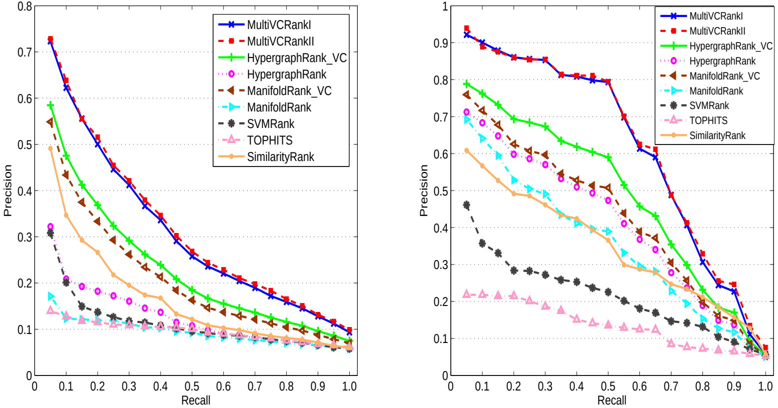
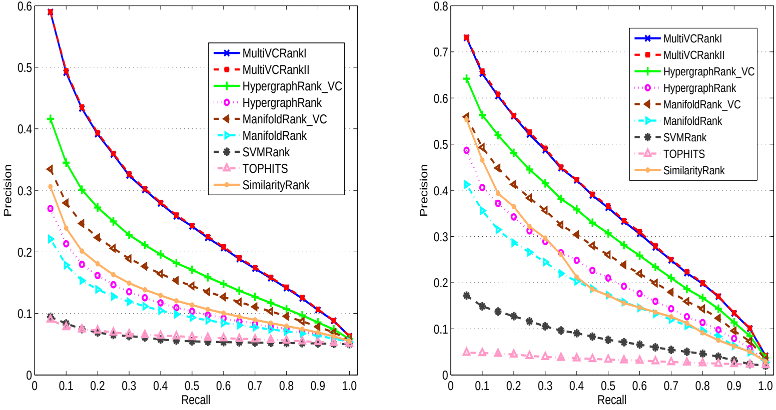


Image Retrieval with Feedback Training (Corel data)



Transition Probability Matrix

Suppose there is a fixed probability $p_{i,j}$ independent of time such that

$$\begin{aligned} & \text{Prob}(X_{t+1} = i | X_t = j, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) \\ = & \text{Prob}(X_{t+1} = i | X_t = j) = p_{i,j} \end{aligned}$$

where $i, j, i_0, i_1, \ldots, i_{t-1} \in \langle n \rangle$. Then this $\{X_t\}$ is called a Markov chain process. The probability $p_{i,j}$ represents the probability that the process will make a transition to the state i given that currently the process is in the state j. Clearly one has

$$p_{i,j} \geq 0, \quad \sum_{i=1}^{n} p_{i,j} = 1, \quad j = 1, \ldots, n.$$

The matrix $P = (p_{i,j})$ is called the one-step transition probability matrix of the process.

Transition Probability Matrix

Suppose $P = (p_{i,j})$ is a transition probability matrix, then there exists a non-negative vector $\bar{\mathbf{x}}$ such that $P\bar{\mathbf{x}} = \bar{\mathbf{x}}$. In particular, if P is irreducible, then $\bar{\mathbf{x}}$ must be positive and unique. When P is primitive, for

$$\lim_{t \to \infty} \mathbf{x}_t = \lim_{t \to \infty} P\mathbf{x}_{t-1}$$

with any initial distribution vector $\mathbf{x}_0$, we have

$$\bar{\mathbf{x}} = P\bar{\mathbf{x}}.$$

Transition Probability Tensor

The $(m-1)^{\mathrm{th}}$ order Markov chain model is used to fit the observed data through the calculation of the $(m-1)^{th}$ order transition probabilities:

$$0 \leq p_{i_1,i_2,\cdots,i_m} = \mathrm{Prob}(X_{t+1} = i_1 | X_t = i_2, \ldots, X_{t-m+2} = i_m) \leq 1$$

where $i_1, i_2, \ldots, i_m \in \langle n \rangle$, and

$$\sum_{i_1=1}^{n} p_{i_1,i_2,\cdots,i_m} = 1.$$

The probability $p_{i_1,i_2,\cdots,i_m}$ represents the probability that the process will make a transition to the state i_1 given that currently the process is in the state i_2 and previously the process are in the states $i_3, \cdots, i_m$.

Transition Probability Tensor

$$\begin{aligned}
\bar{\mathbf{x}}_{i_1} &= \lim_{t\to\infty} \text{Prob}(X_t = i_1) \\
&= \sum_{i_2,\cdots,i_m=1}^{n} p_{i_1,i_2,\cdots,i_m} \times \\
&\quad \times \lim_{t\to\infty} \text{Prob}(X_{t-1} = i_2, X_{t-2} = i_3, \cdots, X_{t-m+1} = i_m) \\
&= \sum_{i_2,\cdots,i_m=1}^{n} p_{i_1,i_2,\cdots,i_m} \prod_{j=2}^{m} \lim_{t\to\infty} \text{Prob}(X_t = i_j) \\
&= \sum_{i_2,\cdots,i_m=1}^{n} p_{i_1,i_2,\cdots,i_m} \bar{x}_{i_2} \cdots \bar{x}_{i_m} = (\mathcal{P}\bar{\mathbf{x}}^{m-1})_{i_1},
\end{aligned}$$

To determine a limiting probability distribution vector of a $(m-1)^{th}$ order Markov chain by solving a tensor equation:

$$\bar{\mathbf{x}} = \mathcal{P}\bar{\mathbf{x}}^{m-1}$$

Transition Probability Tensor

Definition: An m^{th} order n-dimensional tensor $\mathcal{A}$ is called reducible if there exists a nonempty proper index subset $I \subset \{1, 2, \cdots, n\}$ such that

$$a_{i_1, i_2, \cdots, i_m} = 0, \quad \forall i_1 \in I, \quad \forall i_2, \cdots, i_m \notin I.$$

If $\mathcal{A}$ is not reducible, then we call $\mathcal{A}$ irreducible.

Theorem: If $\mathcal{P}$ is a transition probability tensor of order m and dimension n, then there exists a nonzero non-negative vector $\bar{\mathbf{x}}$ such that $\mathcal{P}\bar{\mathbf{x}}^{m-1} = \bar{\mathbf{x}}$ and $\sum_{i=1}^{n} \bar{x}_i = 1$. In particular, if $\mathcal{P}$ is irreducible, then $\bar{\mathbf{x}}$ must be positive.

Transition Probability Tensor

(Order 3) Let S be a proper subset of $\langle n \rangle$ and S' be its complementary set in $\langle n \rangle$, i.e., $S' = \langle n \rangle \backslash S$. For $\mathcal{P} = (p_{i_1,i_2,i_3})$, let

$$\gamma(\mathcal{P}) := \min_{S \subset \langle n \rangle} \left\{ \min_{i_3 \in \langle n \rangle} \left(\min_{i_2 \in S} \sum_{i \in S'} p_{i,i_2,i_3} + \min_{i_2 \in S'} \sum_{i \in S} p_{i,i_2,i_3} \right) + \min_{i_2 \in \langle n \rangle} \left(\min_{i_3 \in S} \sum_{i \in S'} p_{i,i_2,i_3} + \min_{i_3 \in S'} \sum_{i \in S} p_{i,i_2,i_3} \right) \right\}.$$

Theorem: Suppose $\mathcal{P}$ is a transition probability tensor of order 3 and dimension n. If $\gamma(\mathcal{P}) > 1$, then the nonzero non-negative vector $\bar{\mathbf{x}}$ is unique.

Transition Probability Tensor

(Order 3) A simplified condition:

$$|p_{i,i_2,i_3} - p_{i,j_2,j_3}| < \frac{1}{n}, \quad \forall i, i_2, i_3, j_2, j_3 \in \langle n \rangle.$$

A general case:

$$\delta_m(\mathcal{P}) := \min_S \left\{ \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in S'} p_{i,i_2,\cdots,i_m} + \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in S} p_{i,i_2,\cdots,i_m} \right\}.$$

Theorem: Suppose $\mathcal{P}$ is a transition probability tensor of order m and dimension n. If $\delta_m(\mathcal{P}) > \frac{m-2}{m-1}$, then the nonzero non-negative vector $\bar{\mathbf{x}}$ is unique.

Transition Probability Tensor

Theorem: Suppose $\mathcal{P}$ is a transition probability tensor of order 3 and dimension n. Then $\{\mathbf{x}_t\}$ generated by the iterative method, satisfies

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 \leq (2 - \gamma(\mathcal{P}))\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1, \quad \forall t = 0, 1, \cdots .$$

If $\gamma(\mathcal{P}) > 1$, then $\{\mathbf{x}_t\}$ converges linearly to the unique solution $\bar{\mathbf{x}}$, for any initial distribution vector $\mathbf{x}_0$.

Theorem: Suppose $\mathcal{P}$ is a transition probability tensor of order m and dimension n. Then $\{\mathbf{x}_t\}$ generated by the iterative method, satisfies

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 \leq (m-1)(1 - \delta_m(\mathcal{P}))\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1, \quad \forall t = 0, 1, \cdots ,$$

If $\delta_m(\mathcal{P}) > \frac{m-2}{m-1}$, then $\{\mathbf{x}_t\}$ converges linearly to the unique solution $\bar{\mathbf{x}}$, for any initial distribution vector $\mathbf{x}_0$.

Perturbation Results

The Perron vector $\mathbf{x}$ associated to the largest Z-eigenvalue 1 of $\mathcal{P}$, satisfies

$$\mathcal{P}\mathbf{x}^{m-1} = \mathbf{x}$$

where the entries x_i of $\mathbf{x}$ are non-negative and $\sum_{i=1}^{n} x_i = 1$.

When $\mathcal{P}$ is perturbed to an another transition probability tensor $\tilde{\mathcal{P}}$ by $\Delta\mathcal{P}$, the 1-norm error between $\mathbf{x}$ and $\tilde{\mathbf{x}}$?

Extension of matrix perturbation bound

Perturbation Results

Theorem: Let $\mathcal{P}$ and its perturbed tensor $\tilde{\mathcal{P}} = \mathcal{P} + \Delta\mathcal{P}$ be m^{th}-order n-dimensional transition probability tensors. If $\delta_m(\mathcal{P}) > \frac{m-2}{m-1}$, then the Perron vector $\mathbf{x}$ of $\mathcal{P}$ is unique, and for any Perron vector $\tilde{\mathbf{x}}$ of $\tilde{\mathcal{P}}$,

$$||\tilde{\mathbf{x}} - \mathbf{x}||_1 \le \frac{||\Delta\mathcal{P}||_1}{(m-1)\delta_m(\mathcal{P}) + 2 - m}.$$

Theorem: Let $\mathcal{P}$ and its perturbed tensor $\tilde{\mathcal{P}} = \mathcal{P} + \Delta\mathcal{P}$ be 3^{th}-order n-dimensional transition probability tensors. If $\gamma(\mathcal{P}) > 1$, then the Perron vector $\mathbf{x}$ of $\mathcal{P}$ is unique, and for any Perron vector $\tilde{\mathbf{x}}$ of $\tilde{\mathcal{P}}$

$$||\tilde{\mathbf{x}} - \mathbf{x}||_1 \le \frac{||\Delta\mathcal{P}||_1}{\gamma(\mathcal{P}) - 1}.$$

Perturbation Results

The bound is sharp. Let $\mathcal{P}_\varepsilon$ be a 3^{th} order 2-dimension tensor given by

$$\mathcal{P}_\varepsilon = \begin{pmatrix} \frac{1}{3}+\varepsilon & \frac{1}{3}+\varepsilon & \frac{2}{3}-\varepsilon & \frac{2}{3} \\ \frac{2}{3}-\varepsilon & \frac{2}{3}-\varepsilon & \frac{1}{3}+\varepsilon & \frac{1}{3} \end{pmatrix}$$

where $\varepsilon > 0$ and

$$\tilde{\mathcal{P}} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

By a simple computation, $\gamma(\tilde{\mathcal{P}}) = \frac{4}{3}$, $\delta_3(\tilde{\mathcal{P}}) = \frac{2}{3}$ and $||\Delta\mathcal{P}||_1 = \|\mathcal{P}_\varepsilon - \tilde{\mathcal{P}}\|_1 = 2\varepsilon$. We have

$$||\Delta \mathbf{x}||_1 \leq \frac{3}{2}\varepsilon.$$

Perturbation Results

A simple calculation reveals that the equality holds when $\varepsilon = \frac{2}{3}$. Indeed, the eigenvector solutions are given by

$$\mathbf{x} = \left(\frac{2 - \sqrt{2(2-3\varepsilon)}}{3\varepsilon}, 1 - \frac{2 - \sqrt{2(2-3\varepsilon)}}{3\varepsilon} \right)^T \text{ and } \tilde{\mathbf{x}} = (1/2,\ 1/2)^T,$$

respectively. When $\varepsilon = 2/3$, we have $\mathbf{x} = (1, 0)^T$, $\|\mathbf{x} - \tilde{\mathbf{x}}\|_1 = \frac{3}{2}\varepsilon = 1$.

Perturbation Results

- (P. Schweitzer)

$$||\Delta \mathbf{x}||_1 \le ||Z||_1 ||\Delta P||_1,$$

where $Z := (I - P + \mathbf{x}\mathbf{e}^T)^{-1}$ and $\mathbf{e}$ is a vector of all ones.

- (C. Meyer)

$$||\Delta \mathbf{x}||_1 \le ||(I-P)^{\#}||_1 ||\Delta P||_1,$$

where $(I-P)^{\#}$ is the group inverse of $(I-P)$.

- (E. Seneta)

$$||\Delta \mathbf{x}||_1 \le \frac{1}{1-\eta(P)} ||\Delta P||_1,$$

where $\eta(P) := \sup_{||\mathbf{v}||_1 = 1, \mathbf{v}^T\mathbf{e}=0} ||P\mathbf{v}||_1$.

$$||\Delta \mathbf{x}||_1 \le \eta((I-P)^{\#}) ||\Delta P||_1 = \eta(Z^{\#}) ||\Delta P||_1.$$

Perturbation Results

Theorem: Let P and its perturbed matrix $\tilde{P} = P + \Delta P$ be transition probability matrices. If $\delta_2(P) > 0$, then the Perron vector $\mathbf{x}$ with $||\mathbf{x}||_1 = 1$ of P is unique, and for any Perron vector $\tilde{\mathbf{x}}$ with $||\tilde{\mathbf{x}}||_1 = 1$ of $\tilde{P}$ we have

$$||\tilde{\mathbf{x}} - \mathbf{x}||_1 \leq \max_{S \subset \langle n \rangle} \frac{2\alpha_2(\Delta P, S)}{\delta_2(P, S)} \leq \frac{||\Delta P||_1}{\delta_2(P)}.$$

Although we cannot show our bound is always better than the other bounds, the following example shows that our bound is better:

$$P = \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}, \quad \Delta P = \begin{pmatrix} -0.0156 & 0.0119 \\ 0.0156 & -0.0119 \end{pmatrix}.$$

The bounds: 0.0312, 0.0234, 0.0468, and 0.0229 (our)

Multilinear PageRank

Theorem: Let $\mathcal{P}$ be an order-m stochastic tensor, v be a stochastic vector. Then the multilinear PageRank equation

$$x = \alpha \mathcal{P} x^{m-1} + (1-\alpha) v$$

has a unique solution if there exists a vector $\sigma = (\sigma_1, \cdots, \sigma_n)$ such that

$$\alpha \left(\max_{s \in \langle n \rangle} \sum_{k=2}^{m} \max_{\{i_2, \cdots, i_m\} \setminus \{i_k\}} \sum_{i \in \langle n \rangle, i_k = s} |p_{i, i_2, \cdots, i_k, \cdots, i_m} - \sigma_i| \right) < 1.$$

Multilinear PageRank

Since for any $s \in \langle n \rangle$

$$\begin{aligned} & \sum_{k=2}^{m} \max_{\{i_2,\cdots,i_m\}\setminus\{i_k\}} \sum_{i\in\langle n\rangle, i_k=s} |p_{i,i_2,\cdots,i_k,\cdots,i_m} - \sigma_i| \\ \leq \quad & (m-1)\max_{\{i_2,\cdots,i_m\}} \sum_{i\in\langle n\rangle} |p_{i,i_2,\cdots,i_m} - \sigma_i|, \end{aligned}$$

the multilinear PageRank equation has a unique solution provided there exists a vector $\sigma = (\sigma_1, \cdots, \sigma_n)$ such that

$$\alpha < \frac{1}{(m-1)\max_{\{i_2,\cdots,i_m\}} \sum_{i\in\langle n\rangle} |p_{i,i_2,\cdots,i_m} - \sigma_i|}.$$

Taking $\sigma_i = 0$, then the condition reduces to $\alpha < \frac{1}{m-1}$.

Multilinear PageRank

The multilinear PageRank equation has a unique solution provided that

$$\alpha < \frac{2}{(m-1)\sum_{i\in\langle n\rangle}\left(\max_{\{i_2,\cdots,i_m\}} p_{i,i_2,\cdots,i_k,\cdots,i_m} - \min_{\{i_2,\cdots,i_m\}} p_{i,i_2,\cdots,i_k,\cdots,i_m}\right)},$$

or

$$\alpha < \frac{1}{(m-1)\left(1-\sum_{i\in\langle n\rangle}\min_{i_2,\cdots,i_m} p_{i,i_2,\cdots,i_m}\right)},$$

or

$$\alpha < \frac{1}{(m-1)\left(\sum_{i\in\langle n\rangle}\max_{i_2,\cdots,i_m} p_{i,i_2,\cdots,i_m}-1\right)}.$$

A real square matrix with non-negative elements all of whose row-sums and column-sums are equal to 1 is said to be doubly stochastic.

One of the most important properties for a doubly stochastic matrix is the Birkhoff-von Neumann theorem:

Theorem: Any n-by-n doubly stochastic matrix is in the convex hull of c permutation matrices for $c \leq (n-1)^2 + 1$.

Multi-Stochastic Tensor

Definition: An m^{th}-order n-dimensional nonnegative tensor $\mathcal{A} = (a_{i_1,i_2,\cdots,i_m})$ is called *multi-stochastic* if for all $i_j = 1, \cdots, n$, $j \neq k$ we have

$$\sum_{i_k=1}^{n} a_{i_1,\cdots,i_k,\cdots,i_m} = 1, \quad k = 1, ..., m.$$

We denote the set of m^{th}-order n-dimensional multi-stochastic tensors by $\mathbf{\Omega}^{(m,n)}$.

Definition: An m^{th}-order n-dimensional nonnegative tensor $\mathcal{P}$ is said to be a *permutation tensor* if $\mathcal{P}$ has exactly n^{m-1} entries of unity such that $\mathcal{P}_{i_1,\cdots,i_k,\cdots,i_m} = 1$ is the only non-zero entry in the $(i_1, \cdots, i_{k-1}, :, i_{k+1}, \cdots, i_m)$ tube of the k^{th}-class tubes of $\mathcal{P}$ for $1 \leq k \leq m$. We denote the set of m^{th}-order n-dimensional permutation tensors by $\mathbf{\Psi}^{(m,n)}$.

Multi-Stochastic Tensor

Theorem: Let $\mathcal{A}$ be a 3^{rd}-order n-dimensional triply stochastic tensor with its 1-unfolding matrix $A_{(1)} = [A_1 \mid A_2 \mid \cdots \mid A_n]$. Then $\mathcal{A}$ is a convex combination of finitely many permutation tensors if and only if A_i can be written as follows:

$$A_i = \sum_{k=1}^{c} \alpha_i^{(k)} P_i^{(k)}, \quad 1 \leq i \leq n \tag{1}$$

where $P_i^{(k)}$ is a permutation matrix,

$$\sum_{k=1}^{c} \alpha_i^{(k)} = 1, \quad \alpha_i^{(k)} = \alpha_j^{(k)} \geq 0, \quad 1 \leq i \neq j \leq n,$$

and

$$P_i^{(k)} \circ P_j^{(k)} = 0, \quad 1 \leq k \leq c, \quad 1 \leq i \neq j \leq n.$$

In general, a triply stochastic tensor cannot be expressed as a convex combination of finitely many permutation tensors. We guess that the set of extreme points in $\mathbf{\Omega}^{(3,n)}$ contains more than permutation tensors. It is known that the set of all permutation matrices is the same as the set of extreme points in $\mathbf{\Omega}^{(2,n)}$ (the set of doubly stochastic matrices).

As an example, the 1-unfolding matrix of $\mathcal{S}$ given by

$$S_{(1)} = \left[\begin{array}{ccc|ccc|ccc} 0.5 & 0 & 0.5 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 1 & 0 & 0 & 0 & 0.5 & 0.5 \end{array}\right]$$

is an extreme point (is not a permutation tensor).

Concluding Remarks

- Develop models and algorithms for multiple relational data ranking in tensors
- Relevant theoretical results are presented
- Numerical examples are given to show the usefulness of the models
- More and more applications involving tensor/multi-relational data
- Theory and Algorithms are required to be investigated