Large-Scale Social Network Data Mining with Multi-View Information

Hao Wang

Dept. of Computer Science and Engineering
Shanghai Jiao Tong University

Supervisor: Wu-Jun Li

2013.6.19

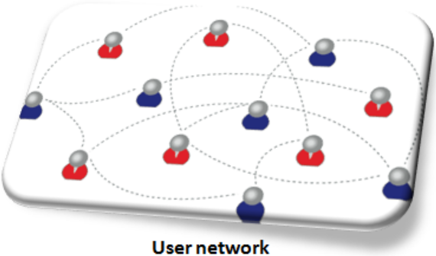
Our work

1. Hao Wang（王灏）, Binyi Chen, Wu-Jun Li. Collaborative Topic Regression with Social Regularization for Tag Recommendation. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), 2013.
2. Hao Wang（王灏）, Wu-Jun Li. Relational Collaborative Topic Regression for Recommendation Systems. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2013. (submitted)
3. Hao Wang（王灏）, Wu-Jun Li. Online Egocentric Models for Citation Networks. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), 2013.

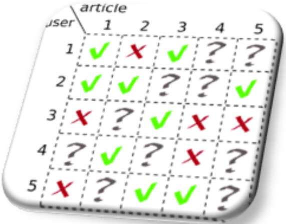
Outline

Outline

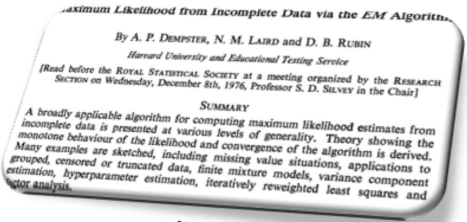
Motivation



User network



Ratings



Maximum Likelihood from Incomplete Data via the EM Algorithm

By A. P. DEMPSTER, N. M. LAIRD and D. B. RUBIN

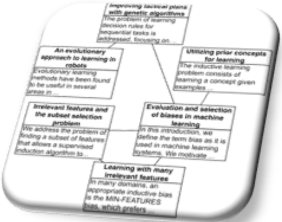
Harvard University and Educational Testing Service

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, December 8th, 1976, Professor S. D. SILVEY in the Chair]

SUMMARY

A broadly applicable algorithm for computing maximum likelihood estimates from incomplete data is presented at various levels of generality. Theory showing the monotone behaviour of the likelihood and convergence of the algorithm is derived. Many examples are sketched, including missing value situations, applications to grouped, censored or truncated data, finite mixture models, variance component estimation, hyperparameter estimation, iteratively reweighted least squares and factor analysis.

Item content



Item network

Motivation

Content:

1. Lee et al., 2010
2. Chen et al., 2010
3. Lipczak et al., 2009

Ratings:

1. Salakhutdinov et al., 2007
2. Herlocker et al., 1999

Hybrid:

1. Purushotham et al., 2012
2. Wang and Blei, 2011
3. Agarwal and Chen, 2010
4. A. Said, 2010

Contribution

1. Integrate ratings, contents and item networks
2. Significantly improve the accuracy
3. Even less training time
4. Extend to dynamic networks

Outline

Comparison of article representation

Maximum Likelihood from Incomplete Data via the *EM* Algorithm

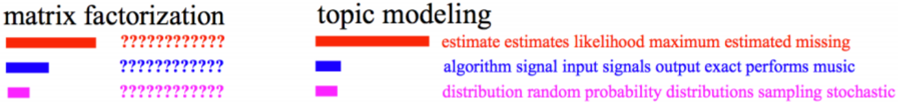
By A. P. Dempster, N. M. Laird and D. B. Rubin

Harvard University and Educational Testing Service

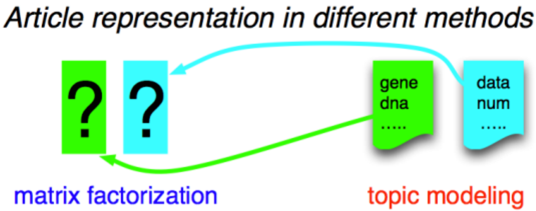
[Read before the Royal Statistical Society at a meeting organized by the Research Section on Wednesday, December 8th, 1976, Professor S. D. Silvey in the Chair]

Summary

A broadly applicable algorithm for computing maximum likelihood estimates from incomplete data is presented at various levels of generality. Theory showing the monotone behaviour of the likelihood and convergence of the algorithm is derived. Many examples are sketched, including missing value situations, applications to grouped, censored or truncated data, finite mixture models, variance component estimation, hyperparameter estimation, iteratively reweighted least squares and factor analysis.

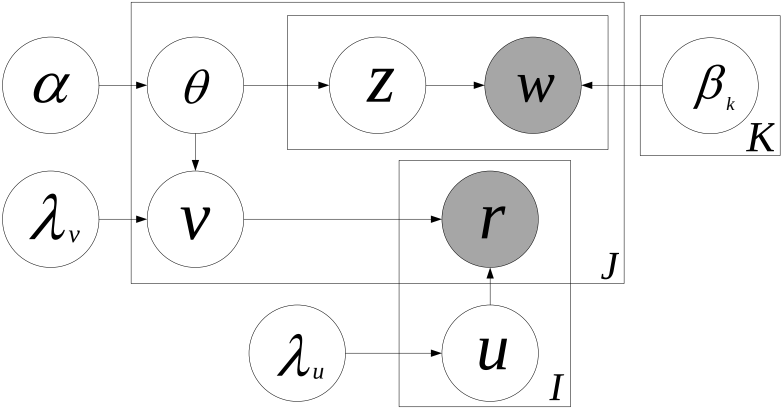


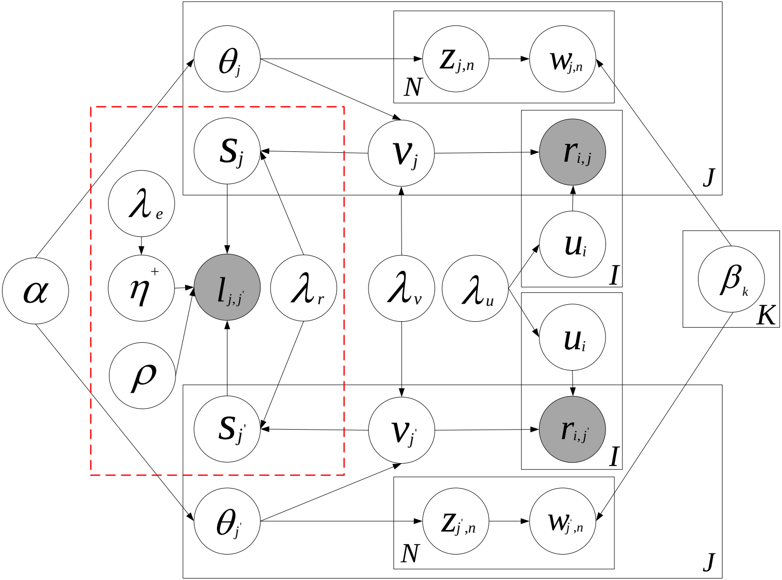
Combination of both



If we simply fix $v = \theta$, we seem to find a way to explain the unknown space using the topic space.

Graphical model of CTR




θ_j
$z_{j,n}$
$w_{j,n}$
N
s_j
v_j
$r_{i,j}$
J
λ_e
u_i
α
η^+
$l_{j,j'}$
λ_r
λ_v
λ_u
I
β_k
K
u_i
ρ
$s_{j'}$
$v_{j'}$
$r_{i,j'}$
I
$\theta_{j'}$
$z_{j',n}$
$w_{j',n}$
N
J

Objective function

The log-likelihood:

$$
\begin{aligned}
L = {} & \rho \sum_{(j,j')} \log \sigma(\eta^T (s_j \circ s_{j'}) + \nu) \\
& - \frac{\lambda_r}{2} \sum_j (s_j - v_j)^T (s_j - v_j) - \frac{\lambda_e}{2} {\eta^+}^T \eta^+ \\
& - \frac{\lambda_u}{2} \sum_i u_i^T u_i - \frac{\lambda_v}{2} \sum_j (v_j - \theta_j)^T (v_j - \theta_j) \\
& + \sum_j \sum_n \log(\sum_k \theta_{jk} \beta_{k,w_{jn}}) - \sum_{i,j} \frac{c_{ij}}{2} (r_{ij} - u_i^T v_j)^2.
\end{aligned}
$$

Updating rules

For U and V:

$$u_i \leftarrow (VC_iV^T + \lambda_u I_K)^{-1} VC_iR_i,$$
$$v_j \leftarrow (UC_iU^T + \lambda_v I_K + \lambda_r I_K)^{-1}(UC_jR_j + \lambda_v\theta_j + \lambda_r s_j),$$

For η^+:

$$\nabla_{\eta^+} L = \rho \sum_{l_{j,j'}=1} (1 - \sigma({\eta^+}^T \pi^+_{j,j'}))\pi^+_{j,j'} - \lambda_e \eta^+,$$

For ϕ_{jnk}:

$$\phi_{jnk} \propto \theta_{jk}\beta_{k,w_{jn}}.$$

For β_{kw}:

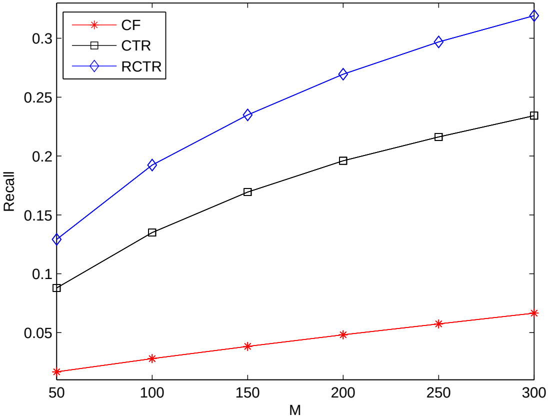
$$\beta_{kw} \propto \sum_j \sum_n \phi_{jnk} 1[w_{jn} = w].$$

Datasets

Description of datasets

	citeulike-a	*citeulike-t*
#users	5551	7947
#items	16980	25975
#tags	19107	52946
#citations	44709	32565
#user-item pairs	204987	134860
sparsity	99.78%	99.93%
#relations	549447	438722

Experimental results



The user-oriented recall of RCTR, CTR, and CF when M ranges from 50 to 300 on dataset *citeulike-t*. P is set to 1. Similar phenomenon can be observed for other values of P.

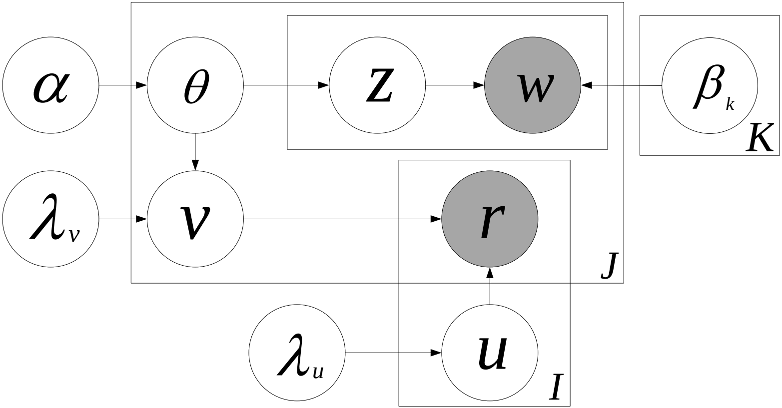
Case study

	user I (RCTR)	in user's lib?
top 3 topics	**1. activity, neural, neurons, cortex, cortical, neuronal, stimuli, spike, visual, stimulus** 2. processing, conditions, sensitivity, perception, music, sound, filters, filter, simultaneous, auditory 3. positive, correlation, hypothesis, negative, correlations, bias, intrinsic, costs, codon, aggregation	
top 10 articles	1. The variable discharge of cortical neurons	**yes**
	2. Refractoriness and neural precision	no
	3. Neural correlates of decision variables in parietal cortex	**yes**
	4. Neuronal oscillations in cortical networks	**yes**
	5. Synergy, redundancy, and independence in population codes	**yes**
	6. Entropy and information in neural spike trains	no
	7. The Bayesian brain: the role of uncertainty in neural coding and computation	**yes**
	8. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action	**yes**
	9. Psychology and neurobiology of simple decisions	**yes**
	10. Role of experience and oscillations in transforming a rate code into a temporal code	**yes**
	user I (CTR)	in user's lib?
top 3 topics	1. coding, take, necessary, place, see, regarding, reason, recognized, mediated, places 2. genetic, variation, population, populations, variants, snps, individuals, genetics, phenotypes, phenotypic 3. **activity, neural, neurons, cortex, cortical, neuronal, stimuli, spike, visual, stimulus**	
top 10 articles	1. Chromatin modifications and their function	no
	2. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution	no
	3. Lateral habenula as a source of negative reward signals in dopamine neurons	**yes**
	4. Two types of dopamine neuron distinctly convey positive and negative motivational signals	no
	5. Proportionally more deleterious genetic variation in European than in African populations	no
	6. The primate amygdala represents the positive and negative value of visual stimuli during learning	**yes**
	7. Genetic variation in an individual human exome	no
	8. Behavioural report of single neuron stimulation in somatosensory cortex	no
	9. Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus	no
	10. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli	**yes**

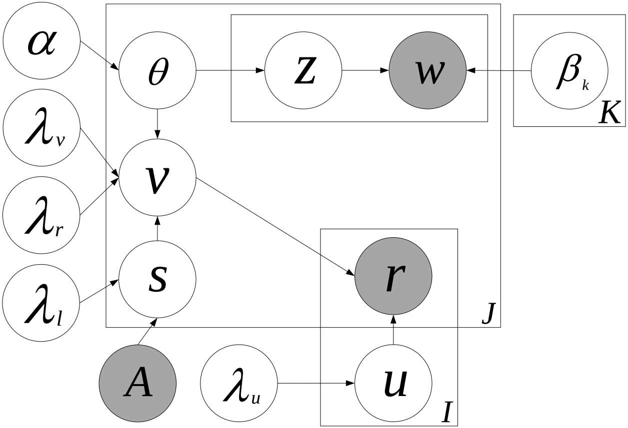
Interpretability of the learning latent structures

Outline

Graphical model of CTR



Graphical model of CTR-SR



Objective function

The log-likelihood:

$$L = -\frac{\lambda_l}{2} tr(SL_aS^T) - \frac{\lambda_r}{2} \sum_j (s_j - v_j)^T (s_j - v_j)$$

$$- \frac{\lambda_u}{2} \sum_i u_i^T u_i - \frac{\lambda_v}{2} \sum_j (v_j - \theta_j)^T (v_j - \theta_j)$$

$$+ \sum_j \sum_n \log(\sum_k \theta_{jk} \beta_{k,w_{jn}}) - \sum_{i,j} \frac{c_{ij}}{2} (r_{ij} - u_i^T v_j)^2.$$

where

$$tr(SL_aS^T) = \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} A_{jj'} ||S_{*j} - S_{*j'}||^2$$

$$= \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} [A_{jj'} \sum_{k=1}^{K} (S_{kj} - S_{kj'})^2]$$

Updating rules

For U and V:

$$u_i \leftarrow (VC_iV^T + \lambda_u I_K)^{-1} VC_iR_i,$$
$$v_j \leftarrow (UC_iU^T + \lambda_v I_K + \lambda_r I_K)^{-1}(UC_jR_j + \lambda_v \theta_j + \lambda_r s_j),$$

For S:

$$S_{k*}(t+1) \leftarrow S_{k*}(t) + \delta(t) r(t)$$
$$r(t) \leftarrow \lambda_r V_{k*} - (\lambda_l L_a + \lambda_r I_J) S_{k*}(t)$$
$$\delta(t) \leftarrow \frac{r(t)^T r(t)}{r(t)^T (\lambda_l L_a + \lambda_r I_J) r(t)}$$

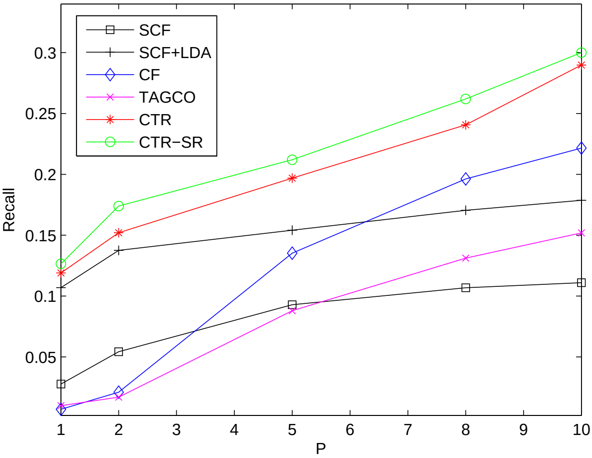
For ϕ_{jnk}:

$$\phi_{jnk} \propto \theta_{jk} \beta_{k,w_{jn}}.$$

For β_{kw}:

$$\beta_{kw} \propto \sum_j \sum_n \phi_{jnk} 1[w_{jn} = w].$$

Experimental results



Recall@50 for all methods in *citeulike-a*.

Case study

Article I	Title: How much can behavioral targeting help online advertising?			
	Top topic 1: web, search, engine, pages, keyword, click, hypertext, html, searchers, crawler			
	Top topic 2: mobile, phones, attitudes, advertising, consumer, marketing, commerce, sms, m-learning			
	True tags: behavioral_targeting, advertising, ads, computational_advertising, recommend, user-behavior, user_profile			
Top 10 recommended tags	CTR	True tag?	CTR-SR	True tag?
	1. random-walks	no	1. behavioral_targeting	**yes**
	2. page-rank	no	2. ads	**yes**
	3. computational_advertising	**yes**	3. computational_advertising	**yes**
	4. citizen-science	no	4. random-walks	no
	5. natural_history	no	5. page-rank	no
	6. search_engine	no	6. developing	no
	7. engine	no	7. recommend	**yes**
	8. searchengine	no	8. advertising	**yes**
	9. what	no	9. what	no
	10. re-ranking	no	10. need	no
Article II	Title: Lowcost multitouch sensing through frustrated total internal reflection			
	Top topic 1: molecular, molecules, surface, chemical, formation, forces, reaction, shapes, sensing, kinetics			
	Top topic 2: design, interface, principles, interfaces, interactive, devices, usability, application			
	True tags: tech, screen, gestures, touch, interface, multitouch, multi-touch, table, visualization, computer_vision			
Top 10 recommended tags	CTR	True tag?	CTR-SR	True tag?
	1. guide	no	1. touch	**yes**
	2. gamma	no	2. field	no
	3. optical	no	3. gestures	**yes**
	4. nanoparticles	no	4. table	**yes**
	5. nano	no	5. multi-touch	**yes**
	6. dna-nanotecnology	no	6. screen	**yes**
	7. tirf	no	7. multitouch	**yes**
	8. sms	no	8. dna-nanotecnology	no
	9. touch	**yes**	9. nano	no
	10. field	no	10. superlist	no

Example articles with recommmended tags

Outline

From DEM to OEM

Three parts of variables:

1. Model parameters $\boldsymbol{\beta}$
2. Link features
3. Topic features

Dynamic Egocentric Models (DEM, adapting only Model parameters):

$$L(\boldsymbol{\beta}) = \prod_{e=1}^{m} \frac{\exp(\boldsymbol{\beta}^T \mathbf{s}_{i_e}(t_e))}{\sum_{i=1}^{n} Y_i(t_e) \exp(\boldsymbol{\beta}^T \mathbf{s}_i(t_e))}$$

Online Egocentric Models (OEM, adapting all three parts):

$$\begin{aligned} &minimize \quad -\log L(\boldsymbol{\beta}, \boldsymbol{\omega}) + \lambda \sum_{k=1}^{n} \|\boldsymbol{\omega}_k - \boldsymbol{\theta}_k\|_2^2 \\ &subject\ to: \quad \boldsymbol{\omega}_k \succeq \mathbf{0},\ \mathbf{1}^T \boldsymbol{\omega}_k = 1, \end{aligned}$$

Updating rules

Using coordinate descent:

1. Online $\boldsymbol{\beta}$ Step:

$$L_w(\boldsymbol{\beta}) = \prod_{e=x+q-W_t}^{x+q-1} \frac{\exp(\boldsymbol{\beta}^T \mathbf{s}_{i_e}(t_e))}{\sum_{i=1}^{n} Y_i(t_e) \exp(\boldsymbol{\beta}^T \mathbf{s}_i(t_e))}.$$

2. Online Topic Step:

$$\begin{aligned} \frac{\partial f}{\partial \boldsymbol{\omega}_k} = & -\sum_{i=1}^{p} \mathbf{a}_i + \sum_{i=1}^{p} \frac{\mathbf{a}_i \alpha_i \exp(\mathbf{a}_i^T \boldsymbol{\omega}_k)}{A_i + \alpha_i \exp(\mathbf{a}_i^T \boldsymbol{\omega}_k)} \\ & + \sum_{u=p+1}^{q} \frac{\mathbf{b}_u \gamma_u \exp(\mathbf{b}_u^T \boldsymbol{\omega}_k)}{B_u + \gamma_u \exp(\mathbf{b}_u^T \boldsymbol{\omega}_k)} \\ & + 2\lambda(\boldsymbol{\omega}_k - \boldsymbol{\theta}_k) \end{aligned}$$

Datasets

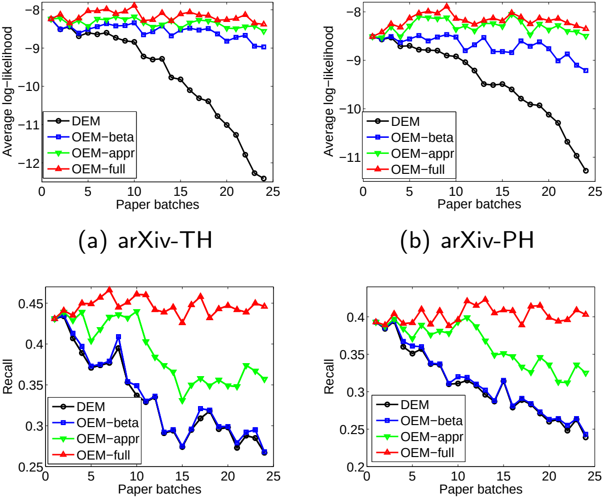
Information of data sets

Data Set	#Papers	#Citations	#Unique Times
arXiv-TH	14226	100025	10500
arXiv-PH	16526	125311	1591

Data set partition for building, training and testing.

Data Sets	Building	Training	Testing
arXiv-TH	62239	1465	36328
arXiv-PH	82343	1739	41229

Experimental results



(a) arXiv-TH (b) arXiv-PH

(c) arXiv-TH(K=250) (d) arXiv-PH(K=250)

Outline

Conclusion

Data Mining with Multi-View Information:

1. **Relational CTR**:
   Social information as prior
2. **CTR with Social regularization**:
   Social information as observation
3. **Online Egocentric Models**:
   Dynamic social information

Publication

1. Hao Wang（王灏）, Binyi Chen, Wu-Jun Li. Collaborative Topic Regression with Social Regularization for Tag Recommendation. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), 2013.
2. Hao Wang（王灏）, Wu-Jun Li. Online Egocentric Models for Citation Networks. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), 2013.
3. Hao Wang（王灏）, Wu-Jun Li. Relational Collaborative Topic Regression for Recommendation Systems. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2013. (submitted)

Q & A