Knowledge Transfer Using Latent Variable Models

Ayan Acharya

UT Austin, Department of ECE

September 23, 2014

Outline

- Motivation & Background
- Active Multitask Learning Using Both Supervised and Latent Shared Topics
- Gamma Process Poisson Factorization
- Conclusion & Future Work

Motivation & Theme

Motivation:

- Labeled data is sparse in applications like document categorization and object recognition.
- Distribution of data changes across domains or over time.

Theme:

- Shared low dimensional space for transferring information across domains
- Careful adaptation of the model parameters to fit new data

Transfer Learning & Active Learning

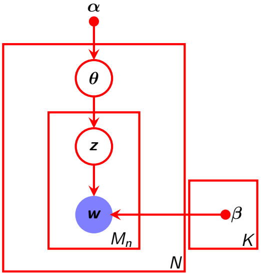
Transfer Learning:

- Concurrent knowledge transfer (or multitask learning): multiple domains learnt simultaneously
- Continual knowledge transfer (or sequential knowledge transfer): models learnt in one domain are carefully adapted to other domains

Active Learning:

- Only the most informative examples are queried from the unlabeled pool.

Topic Models: Latent Dirichlet Allocation (LDA)



- $\theta_n \sim \text{Dir}(\alpha)\ \forall n.$
- $z_{nm} \sim \text{mult}(\theta_n)\ \forall n, m.$
- $w_{nm} \sim \text{mult}(\beta_{z_{nm}})\ \forall n, m.$
- $\beta_k \sim \text{Dir}(\eta)\ \forall k.$

Figure : LDA

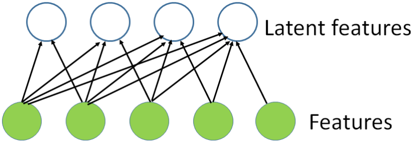


Figure : Visual Representation

Topic Models: Labeled LDA

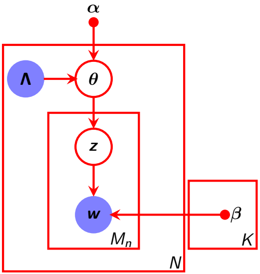


Figure : LLDA – supervision at topic level

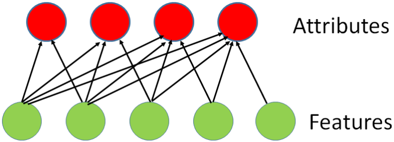


Figure : Visual Representation

Topic Models: MedLDA

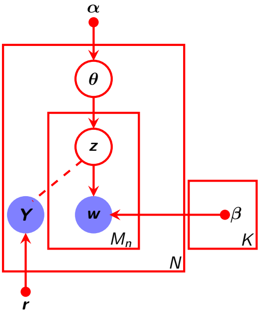


Figure : MedLDA – supervision at category/class level

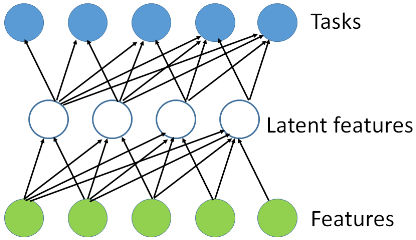


Figure : Visual Representation

Relation between Distributions

- If $x_k \sim \text{Gamma}(\alpha_k, 1/c)\ \forall k$, then $\boldsymbol{Y} = (y_k)_{k=1}^{K} \sim \text{Dir}(\alpha_1, \cdots, \alpha_K)$ where $y_k = x_k / \sum_{k=1}^{K} x_k$.
- Let $x_k \sim \text{Poisson}(\lambda_k)\ \forall k$, $X = \sum_{k=1}^{K} x_k$, $\lambda = \sum_{k=1}^{K} \lambda_k$. Suppose $(y_1, \cdots, y_K) \sim \text{mult}(X; \lambda_1/\lambda, \cdots, \lambda_K/\lambda)$. Then $P(x_1, \cdots, x_K) = P(y_1, \cdots, y_K)$.

LDA as Poisson Factorization

- $\theta_{dk} \sim \text{Gamma}(a_k, 1/c_d)\ \forall d, k.$
- $\beta_{wk} \sim \text{Gamma}(b_k, 1/c_w)\ \forall w, k.$
- $y_{dw} = \sum_{k=1}^{K} y_{dwk}, y_{dwk} \sim \text{Poisson}(\theta_{dk}\beta_{wk}).$

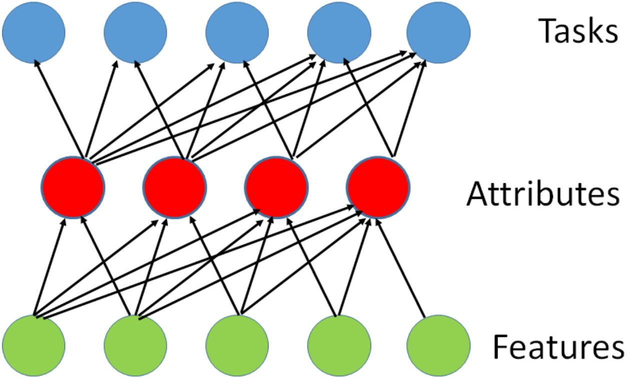
Problem Setting: Multitask Learning

- In training corpus each document/image belongs to a known class and has a set of attributes (supervised topics).
- Classes from aYahoo data: carriage, centaur, bag, building, donkey, goat, jetski, monkey, mug, statue, wolf, and zebra
- Attributes: “has head”, “has wheel”, “has torso” and 61 others
- Train models using words, supervised topics and class labels, and classify completely unlabeled test data (no supervised topic or class label)

Class: Carriage

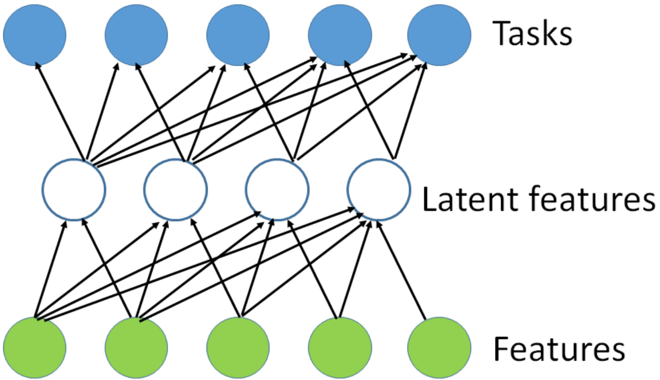
Attributes:
“has wheel?” Yes.
“has wood?” Yes.

Transfer with Shared Supervised Attributes



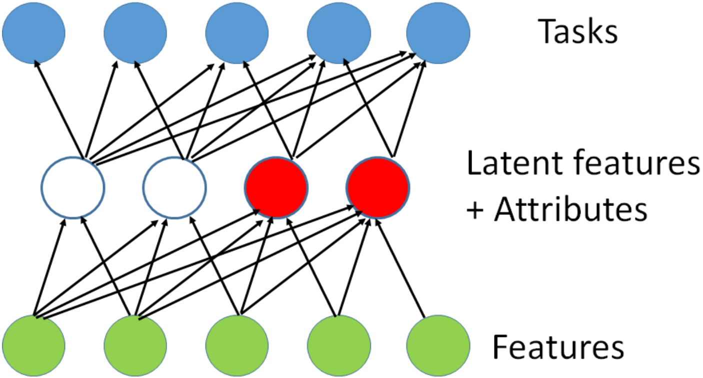
- Train to infer attributes from visual features
- Train to infer categories from attributes [Lampert et al., 2009]

Multitask Learning with Shared Latent Features



Reference: [Caruana, 1997]

Transfer with Shared Supervised and Latent Attributes



Doubly Supervised LDA

- Doubly Supervised LDA [Acharya et al., 2013]
- $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}$: priors over supervised and latent topics

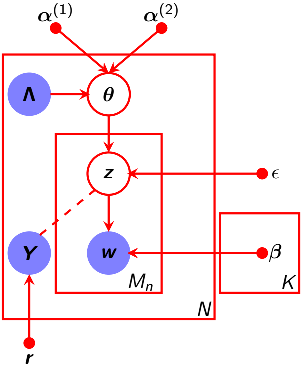


Figure : DSLDA – Supervision at both topic and category level

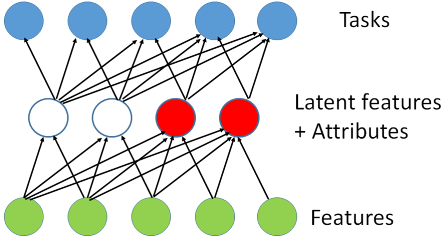


Figure : Visual Representation

Non-parametric Doubly Supervised LDA (NPDSLDA)

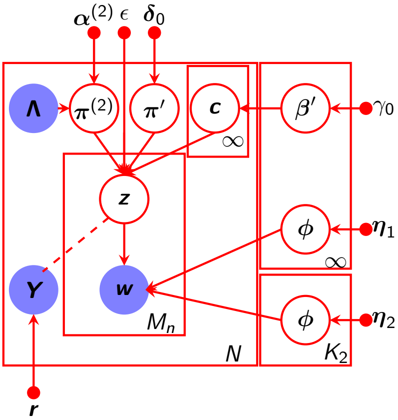


Figure : NPDSLDA

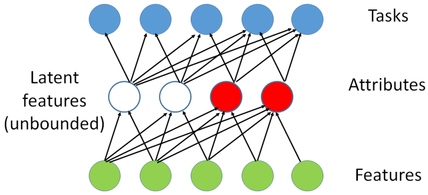


Figure : Visual Representation

Problem Setting: Active Multitask Learning

- In training corpus each document/image belongs to a known class and has a set of attributes (supervised topics).
- Classes from aYahoo data: carriage, centaur, bag, building, donkey, goat, jetski, monkey, mug, statue, wolf, and zebra
- Attributes: “has head”, “has wheel”, “has torso” and 61 others
- Train models using words, supervised topics and class labels
- An active MTL framework that can use and query over both attributes and class labels

Class: Carriage

Attributes:
“has wheel?” Yes.
“has wood?” Yes.

Active DSLDA (Act-DSLDA)

- $\boldsymbol{r}_1$: weights for multiclass SVM
- $\boldsymbol{r}_2$: weights for binary SVMs

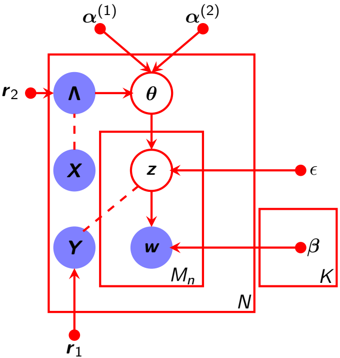


Figure : Act-DSLDA

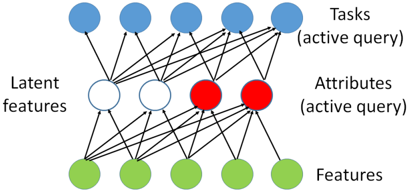


Figure : Visual Representation

Active NPDSLDA (Act-NPDSLDA)

- Non-parametric Doubly Supervised LDA [Acharya et al., 2013]

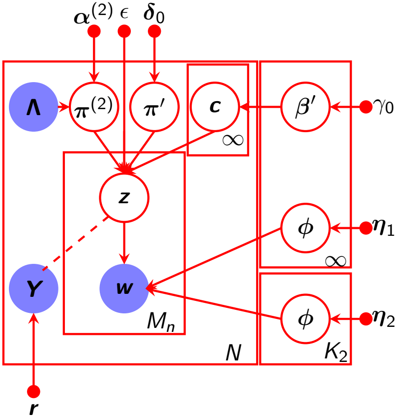


Figure : NPDSLDA

Active NPDSLDA (Act-NPDSLDA)

- Non-parametric Doubly Supervised LDA [Acharya et al., 2013]

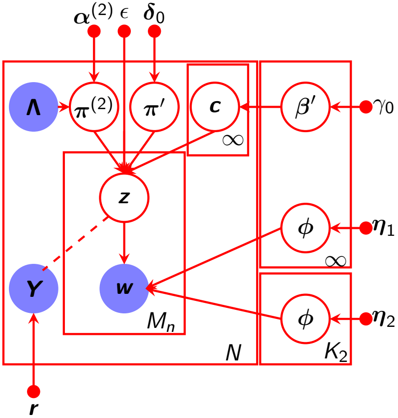


Figure : NPDSLDA

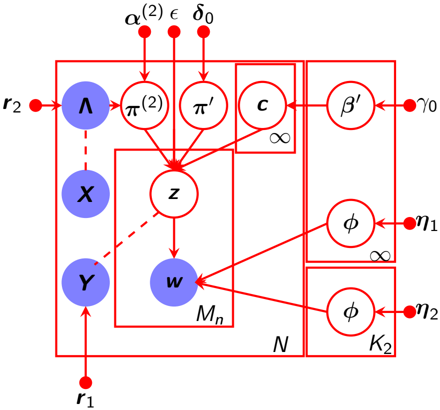


Figure : Act-NPDSLDA

Visual Representation of Act-NPDSLDA

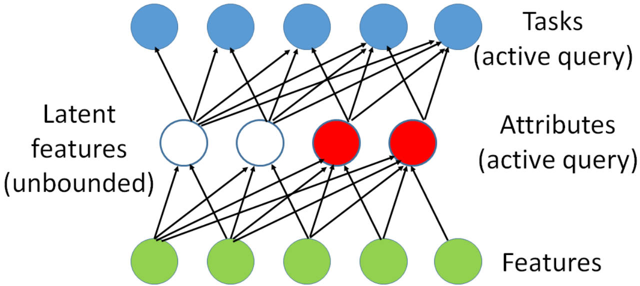


Figure : Visual Representation of Act-NPDSLDA

Inference and Learning

- Active learning measure: expected error reduction [Nigam et al., 1998]
- Batch mode: variational EM with completely factorized approximation to posterior, online SVM [Bordes et al., 2007]
- Active selection mode: incremental EM [Neal and Hinton, 1999], online SVM

Description of Dataset: ACM Conference

- **Classes:** Conference names: WWW, SIGIR, KDD, ICML, ISPD, DAC; abstracts of papers are treated as documents
- **Supervised topics:** keywords provided by the authors

Experimental Methodology

- Multitask training that evaluates benefits of sharing information among classes on the predictive accuracy of all classes
- Start with a completely labeled dataset $\mathcal{L}$ consisting of 300 documents
- In every active iteration, 50 labels (class labels or supervised topics) are queried for.

Baselines: MTL Experiments

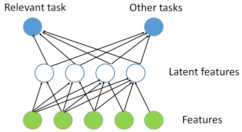


Figure : MedLDA-OVA

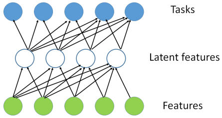


Figure : MedLDA-MTL

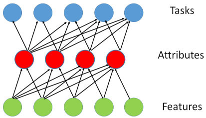


Figure : DSLDA-OSST

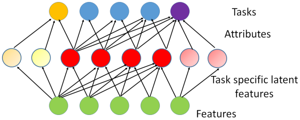
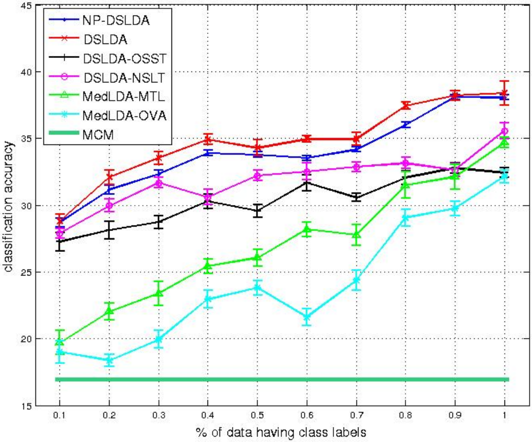


Figure : DSLDA-NSLT

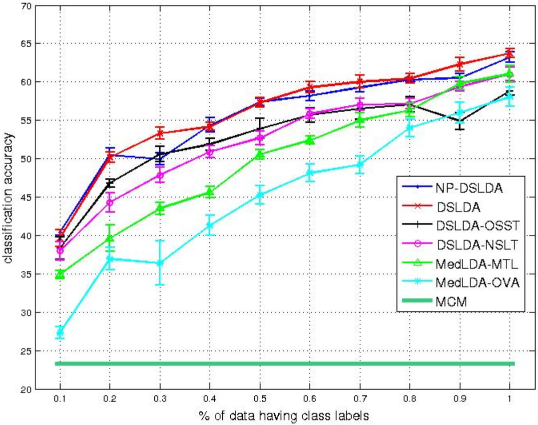
Results from aYahoo Data

- 50% training with supervised topic labels



Results from ACM Conference Text Data

- 50% training with supervised topic labels



Baselines: AMTL experiments

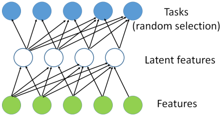


Figure : Random MedLDA-MTL (R-MedLDA-MTL)

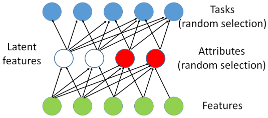


Figure : Random DSLDA (R-DSLDA)

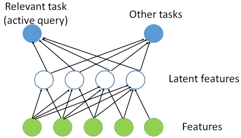


Figure : Active MedLDA-OVA (Act-MedLDA-OVA)

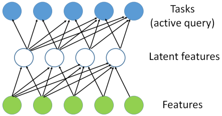
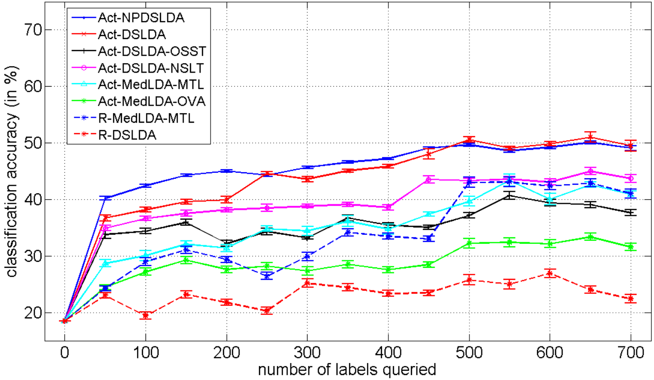
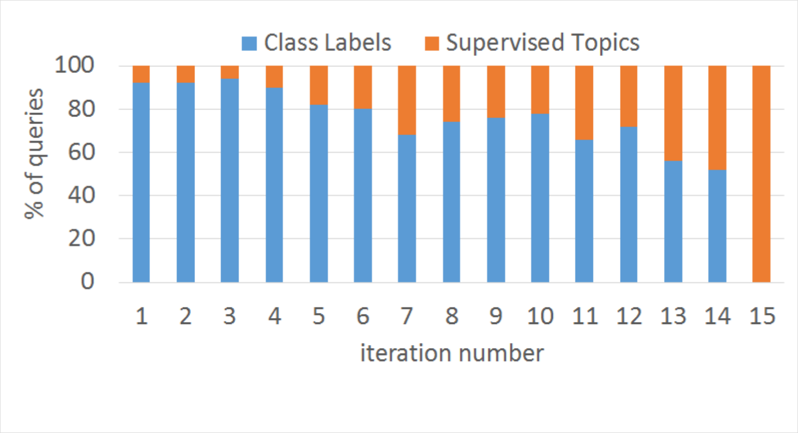


Figure : Active MedLDA-MTL (Act-MedLDA-MTL)

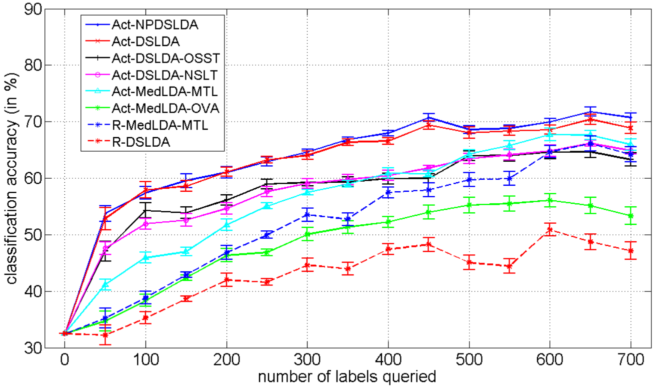
aYahoo Learning Curves



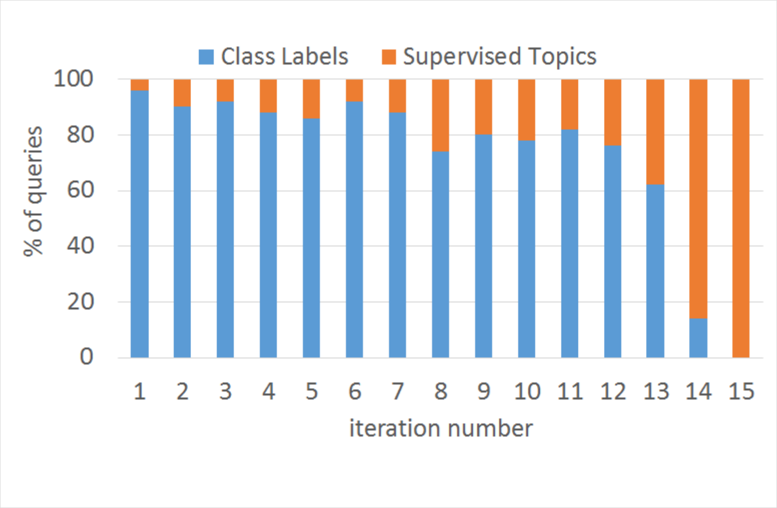
aYahoo Query Distribution



ACM Conference Learning Curves



ACM Conference Query Distribution



Negative Binomial & Chinese Restaurant Table Distribution

Negative Binomial Distribution (NB):

- $m \sim \text{NB}(r, p)$.
- $m \sim \text{Poisson}(\lambda), \lambda \sim \text{Gamma}(r, p)$.
- $m \sim \sum_{t=1}^{l} u_t,\ u_t \sim \text{Log}(p),\ l \sim \text{Poisson}(-r \log(1-p))$.

Chinese Restaurant Table Distribution (CRT):

- m : number of data points (number of customers), K : number of distinct atoms (number of tables), $s(m, l)$: Stirling number of the first kind.

$$\Pr(K = l | m, \gamma_0) = \frac{\Gamma(\gamma_0)}{\Gamma(m + \gamma_0)} |s(m, l)| \gamma_0^l,\ \ l = 0, 1, \cdots, m.$$

Gamma Process (GP)

- The Gamma Process $G \sim \mathrm{GaP}(c, G_0)$ is a completely random measure defined on the product space $\mathbb{R}_+ \times \Omega$ with concentration parameter c and a finite and continuous base measure G_0 over a complete separable metric space Ω, such that $G(A_i) \sim \mathrm{Gamma}(G_0(A_i), 1/c)$ are independent gamma random variables for disjoint partition $\{A_i\}_i$ of Ω.

Gamma Process (GP)

- The Gamma Process $G \sim \mathrm{GaP}(c, G_0)$ is a completely random measure defined on the product space $\mathbb{R}_+ \times \Omega$ with concentration parameter c and a finite and continuous base measure G_0 over a complete separable metric space Ω, such that $G(A_i) \sim \mathrm{Gamma}(G_0(A_i), 1/c)$ are independent gamma random variables for disjoint partition $\{A_i\}_i$ of Ω.
- A draw from the GP consists of countably infinite atoms, which can be expressed as:

$$G = \sum_{k=1}^{\infty} r_k \delta_{\omega_k}, (r_k, \omega_k) \overset{iid}{\sim} r^{-1} e^{-cr} dr G_0(d\omega).$$

Gamma Process (GP)

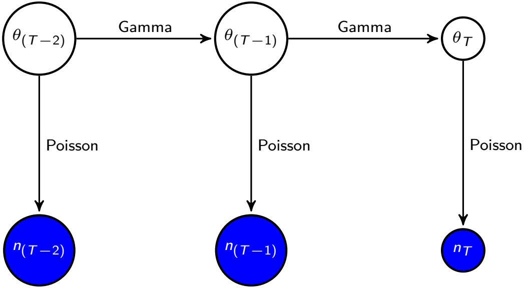
- The Gamma Process $G \sim \mathrm{GaP}(c, G_0)$ is a completely random measure defined on the product space $\mathbb{R}_+ \times \Omega$ with concentration parameter c and a finite and continuous base measure G_0 over a complete separable metric space Ω, such that $G(A_i) \sim \mathrm{Gamma}(G_0(A_i), 1/c)$ are independent gamma random variables for disjoint partition $\{A_i\}_i$ of Ω.
- A draw from the GP consists of countably infinite atoms, which can be expressed as:

$$G = \sum_{k=1}^{\infty} r_k \delta_{\omega_k}, (r_k, \omega_k) \overset{iid}{\sim} r^{-1} e^{-cr} dr G_0(d\omega).$$

- Finite approximation of GP:

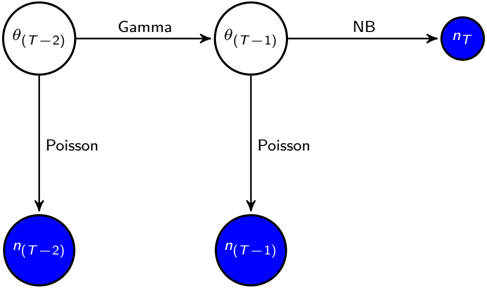
$$G = \sum_{k=1}^{K} r_k \delta_{\omega_k}, (r_k, \omega_k) \overset{iid}{\sim} r^{-1} e^{-cr} dr G_0(d\omega).$$

Gamma Poisson Autoregressive Model (GPAR)



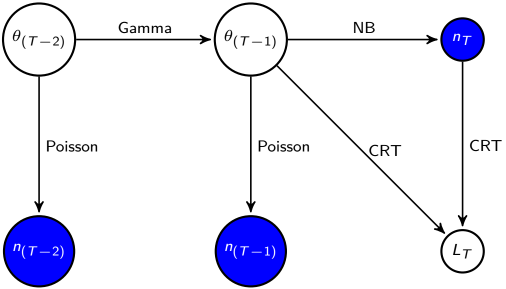
- $\theta_t \sim \text{Gamma}(\theta_{(t-1)}, 1/c), n_t \sim \text{Poisson}(\theta_t)$

Inference in GPAR



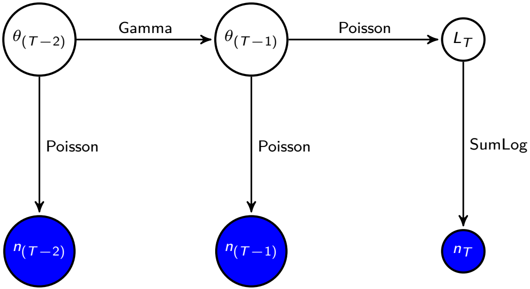
- $n_T \sim \mathrm{NB}(\theta_{(T-1)}, 1/(c+1))$.

Inference in GPAR



- $n_T \sim \text{NB}(\theta_{(T-1)}, 1/(c+1))$. Augment $L_T \sim \text{CRT}(n_T, \theta_{(T-1)})$.

Inference in GPAR

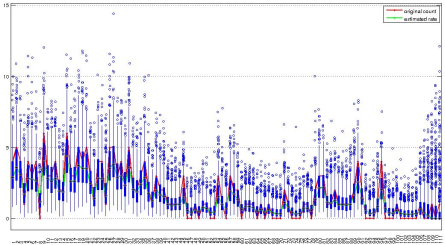


- $n_T \sim \sum_{t=1}^{L_T} \mathrm{Log}(1/(c+1)), L_T \sim \mathrm{Poisson}(\theta_{(T-1)} \log((c+1)/c)).$

Gibbs Sampling in GPAR

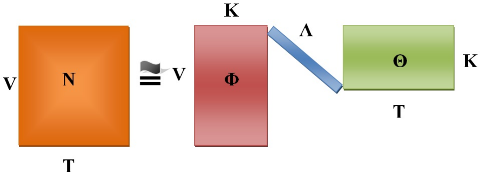
- Backward Sampling: for $t = T$ to 1, $L_t \sim \text{CRT}(n_t, \theta_{(t-1)})$.
- Forward Sampling: for $t = 1$ to T,
  $\theta_t \sim \text{Gamma}(\theta_{(t-1)} + n'_t, p_t)$, $p_t = 1/(1 + c - \log(p_{(t-1)}))$,
  $n'_t = n_t + L(t+1)$.

Result from GPAR



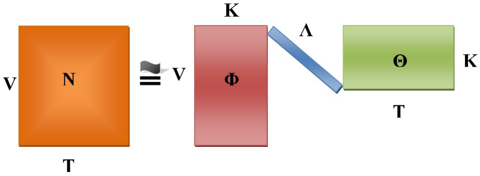
DataSet	Measure	GP-DPFA	SGCP	KS	LGCP10	LGCP25	LGCP100
SDS1	MSE	4.23	**4.20**	6.65	5.96	6.12	5.44
	PMSE	3.09	**2.98**	5.43	6.92	4.53	4.28
SDS2	MSE	**27.12**	38.38	73.71	70.34	53.27	43.51
	PMSE	**10.14**	12.01	13.49	14.73	12.91	12.52
SDS3	MSE	**10.94**	11.41	30.56	90.76	22.14	10.79
	PMSE	**5.81**	7.19	25.17	28. 72	23.49	20.08

Gamma Process Dynamic Poisson Factor Analysis (GP-DPFA)



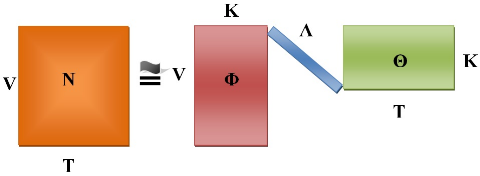
- $\mathbf{N} \sim \text{Poisson}(\mathbf{\Phi\Lambda\Theta})$ $\mathbf{N} \in \mathbb{Z}_{+}^{V\times T}$.

Gamma Process Dynamic Poisson Factor Analysis (GP-DPFA)



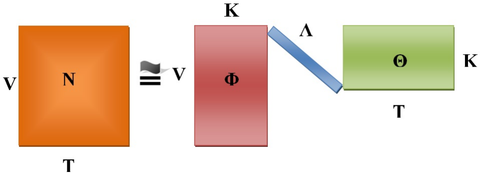
- $\mathbf{N} \sim \text{Poisson}(\mathbf{\Phi\Lambda\Theta})$ $\mathbf{N} \in \mathbb{Z}_+^{V \times T}$.
- $n_{vt} = \sum_{k=1}^{K} n_{vtk}, \; n_{vtk} \sim \text{Poisson}(r_k \phi_{vk} \theta_{tk})$.

Gamma Process Dynamic Poisson Factor Analysis (GP-DPFA)



- $\mathbf{N} \sim \text{Poisson}(\mathbf{\Phi\Lambda\Theta})$ $\mathbf{N} \in \mathbb{Z}_+^{V\times T}$.
- $n_{vt} = \sum_{k=1}^{K} n_{vtk},\ n_{vtk} \sim \text{Poisson}(r_k \phi_{vk} \theta_{tk})$.
- $\phi_k \sim \text{Dir}(\eta_1, \cdots, \eta_V)$, $\theta_{tk} \sim \text{Gamma}(\theta_{(t-1)k}, 1/c_t)$.

Gamma Process Dynamic Poisson Factor Analysis (GP-DPFA)

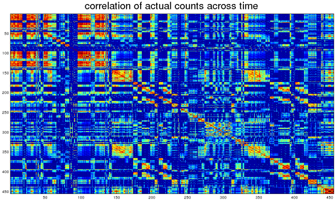


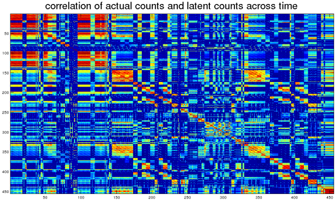
- $\mathbf{N} \sim \text{Poisson}(\mathbf{\Phi\Lambda\Theta})$ $\mathbf{N} \in \mathbb{Z}_+^{V\times T}$.
- $n_{vt} = \sum_{k=1}^{K} n_{vtk},\ n_{vtk} \sim \text{Poisson}(r_k \phi_{vk} \theta_{tk})$.
- $\phi_k \sim \text{Dir}(\eta_1, \cdots, \eta_V)$, $\theta_{tk} \sim \text{Gamma}(\theta_{(t-1)k}, 1/c_t)$.
- $\lambda_k \sim \text{Gamma}(r_0/K, 1/c)$.

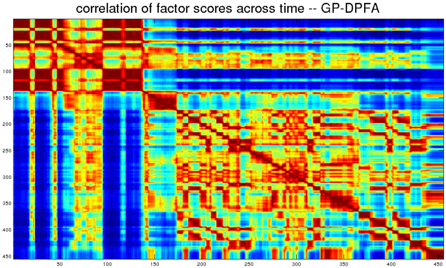
Inference in GP-DPFA

- Introduce $l_{tk} \sim \mathrm{CRT}(n_{tk} + l_{(t+1)k}, \theta_{(t-1)k})$, sample l_{tk} backwards.
- Sample $\theta_{tk}|- \sim \mathrm{Gamma}(\theta_{(t-1)k} + n_{tk} + l_{(t+1)k}, (1 - p_{tk})/c_t)$ in the forward direction.
- $\phi_k|- \sim \mathrm{Dir}(\eta_1 + n_{1.k}, \cdots, \eta_V + n_{V.k})$.
- $\lambda_k|- \sim \mathrm{Gamma}(n_{..k} + r_0/K, 1/(c + \sum_t \theta_{tk}))$.
- $(n_{vt1}, \cdots, n_{vtK})|- \sim \mathrm{mult}(n_{vt}, \frac{\lambda_1 \phi_{v1} \theta_{t1}}{\sum_k \lambda_k \phi_{vk} \theta_{tk}}, \cdots, \frac{\lambda_K \phi_{vK} \theta_{tK}}{\sum_k \lambda_k \phi_{vk} \theta_{tk}})$.

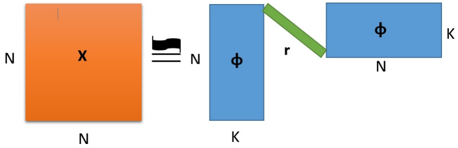
Results from GP-DPFA – Piano





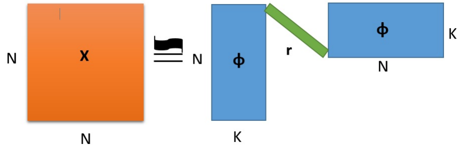


Gamma Process Poisson Factorization for Network Modeling (N-GPPF)



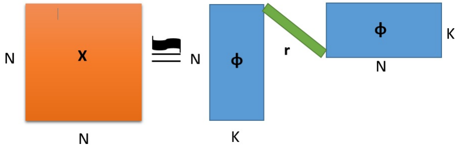
- $\phi_k \sim \prod_{n=1}^{N} \text{Gamma}(a_0, 1/c_n)$, $c_n \sim \text{Gamma}(c_0, 1/d_0)$.

Gamma Process Poisson Factorization for Network Modeling (N-GPPF)



- $\phi_k \sim \prod_{n=1}^{N} \text{Gamma}(a_0, 1/c_n)$, $c_n \sim \text{Gamma}(c_0, 1/d_0)$.
- $r_k \sim \text{Gamma}(\gamma_0/K, 1/c)$, $c \sim \text{Gamma}(g_0, 1/h_0)$, $\gamma_0 \sim \text{Gamma}(e_0, 1/f_0)$.

Gamma Process Poisson Factorization for Network Modeling (N-GPPF)



- $\phi_k \sim \prod_{n=1}^{N} \mathrm{Gamma}(a_0, 1/c_n)$, $c_n \sim \mathrm{Gamma}(c_0, 1/d_0)$.
- $r_k \sim \mathrm{Gamma}(\gamma_0/K, 1/c)$, $c \sim \mathrm{Gamma}(g_0, 1/h_0)$, $\gamma_0 \sim \mathrm{Gamma}(e_0, 1/f_0)$.
- $x_{nm} = \sum_{k=1}^{K} x_{nmk}, x_{nmk} \sim \mathrm{Pois}(r_k \phi_{nk} \phi_{mk})$.
- $b_{nm} = I_{\{x_{nm} \geq 1\}}$.

Inference in N-GPPF

- $(x_{nmk})_{k=1}^{K} |- \sim \text{mult}\left(x_{nm}\frac{r_k\phi_{nk}\phi_{mk}}{\sum_{k=1}^{K} r_k\phi_{nk}\phi_{mk}}\right)_{k=1}^{K}.$

Inference in N-GPPF

- $(x_{nmk})_{k=1}^{K} | - \sim \text{mult}\left(x_{nm} \frac{r_k \phi_{nk} \phi_{mk}}{\sum_{k=1}^{K} r_k \phi_{nk} \phi_{mk}}\right)_{k=1}^{K}.$
- $(\phi_{nk} | -) \sim \text{Gamma}\left(a_0 + x_{n.k}, \frac{1}{c_n + r_k \phi_k^{-n}}\right).$

Inference in N-GPPF

- $(x_{nmk})_{k=1}^{K}|- \sim \text{mult}\left(x_{nm}\frac{r_k\phi_{nk}\phi_{mk}}{\sum_{k=1}^{K} r_k\phi_{nk}\phi_{mk}}\right)_{k=1}^{K}$.
- $(\phi_{nk}|-) \sim \text{Gamma}\left(a_0 + x_{n.k}, \frac{1}{c_n + r_k\phi_k^{-n}}\right)$.
- $(r_k|-) \sim \text{Gamma}\left(\frac{\gamma_0}{K} + x_{..k}, \frac{1}{c + \sum_n \phi_{nk}\phi_k^{-n}}\right)$.

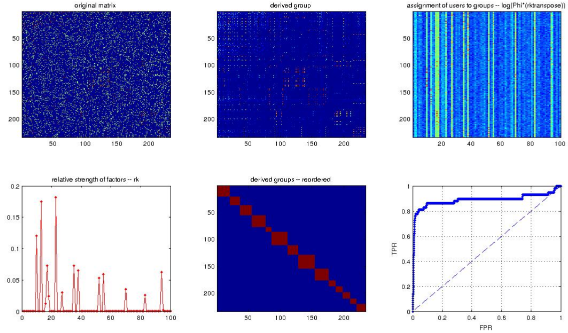
Inference in N-GPPF

- $(x_{nmk})_{k=1}^{K}|- \sim \text{mult}\left(x_{nm}\frac{r_k\phi_{nk}\phi_{mk}}{\sum_{k=1}^{K} r_k\phi_{nk}\phi_{mk}}\right)_{k=1}^{K}.$
- $(\phi_{nk}|-) \sim \text{Gamma}\left(a_0 + x_{n.k}, \frac{1}{c_n + r_k\phi_k^{-n}}\right).$
- $(r_k|-) \sim \text{Gamma}\left(\frac{\gamma_0}{K} + x_{..k}, \frac{1}{c + \sum_n \phi_{nk}\phi_k^{-n}}\right).$
- $(c_n|-) \sim \text{Gamma}\left(c_0 + Ka_0, \frac{1}{d_0 + \phi_{n.}}\right).$

Computational Complexity

Method	N-GPPF	IRM	ILA	MMSB
Complexity	$O(SK + NK)$	$O(N^2K)$	$O(N^2K)$	$O(N^2K)$

Results from N-GPPF – NIPS Authorship Network

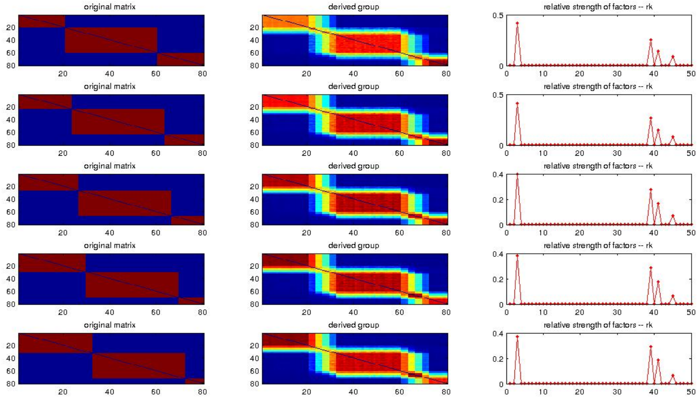


Measure	N-GPPF	IRM	MMSB	MedLFRM
AUC	0.9234	0.8924	0.8713	0.9647
Running time (mins)	4	13*60	17	22*60

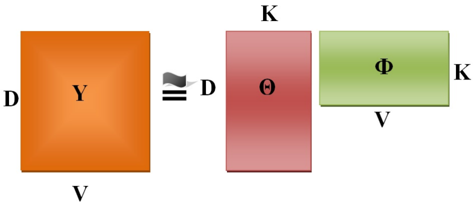
Dynamic Network Modeling using GPPF (D-GPPF)

- $r_{tk} \sim \text{Gamma}(r_{(t-1)k}, 1/c)$, $c \sim \text{Gamma}(g_0, 1/h_0)$, $\gamma_0 \sim \text{Gamma}(e_0, 1/f_0)$.
- $x_{tnm} = \sum_{k=1}^{K} x_{tnmk}, x_{tnmk} \sim \text{Pois}(r_{tk}\phi_{nk}\phi_{mk})$.

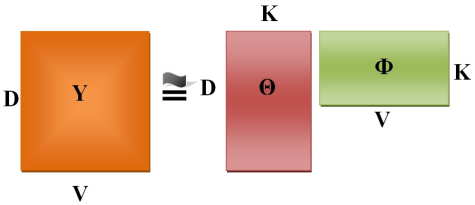
Result from D-GPPF



GPPF for Modeling Count Data (C-GPPF)

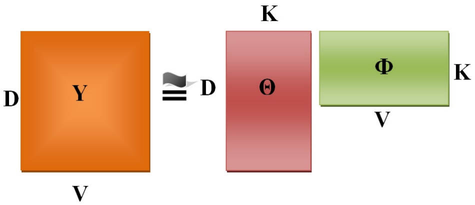


GPPF for Modeling Count Data (C-GPPF)



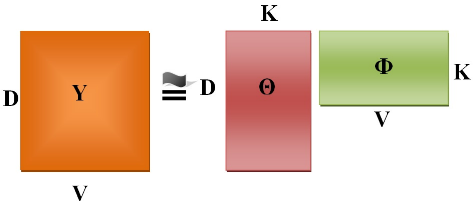
- $\theta_k \sim \prod_{d=1}^{D} \mathrm{Gamma}(a_0, 1/c_d)$, $c_d \sim \mathrm{Gamma}(c_0^{(1)}, 1/d_0^{(1)})$.

GPPF for Modeling Count Data (C-GPPF)



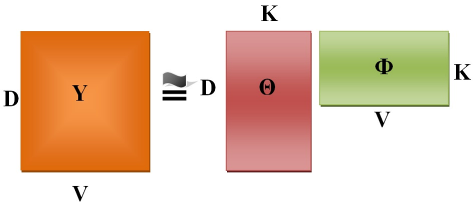
- $\theta_k \sim \prod_{d=1}^{D} \text{Gamma}(a_0, 1/c_d),\ c_d \sim \text{Gamma}(c_0^{(1)}, 1/d_0^{(1)})$.
- $\phi_k \sim \prod_{w=1}^{V} \text{Gamma}(b_0, 1/c_w),\ c_w \sim \text{Gamma}(c_0^{(2)}, 1/d_0^{(2)})$.

GPPF for Modeling Count Data (C-GPPF)



- $\theta_k \sim \prod_{d=1}^{D} \text{Gamma}(a_0, 1/c_d)$, $c_d \sim \text{Gamma}(c_0^{(1)}, 1/d_0^{(1)})$.
- $\phi_k \sim \prod_{w=1}^{V} \text{Gamma}(b_0, 1/c_w)$, $c_w \sim \text{Gamma}(c_0^{(2)}, 1/d_0^{(2)})$.
- $r_k \sim \text{Gamma}(\gamma_0/K, 1/c)$, $c \sim \text{Gamma}(g_0, 1/h_0)$, $\gamma_0 \sim \text{Gamma}(e_0, 1/f_0)$.

GPPF for Modeling Count Data (C-GPPF)

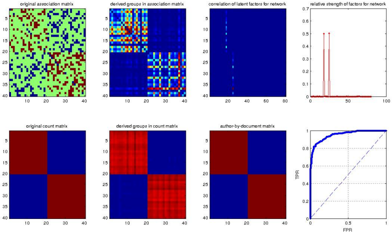


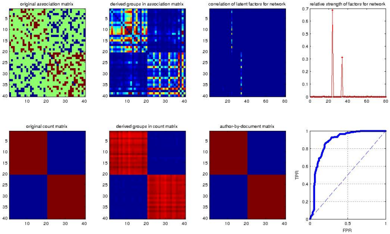
- $\theta_k \sim \prod_{d=1}^{D} \text{Gamma}(a_0, 1/c_d)$, $c_d \sim \text{Gamma}(c_0^{(1)}, 1/d_0^{(1)})$.
- $\phi_k \sim \prod_{w=1}^{V} \text{Gamma}(b_0, 1/c_w)$, $c_w \sim \text{Gamma}(c_0^{(2)}, 1/d_0^{(2)})$.
- $r_k \sim \text{Gamma}(\gamma_0/K, 1/c)$, $c \sim \text{Gamma}(g_0, 1/h_0)$, $\gamma_0 \sim \text{Gamma}(e_0, 1/f_0)$.
- $y_{dw} = \sum_{k=1}^{K} y_{dwk}$, $y_{dwk} \sim \text{Pois}(r_k \theta_{dk} \phi_{wk})$.

Computational Complexity of C-GPPF

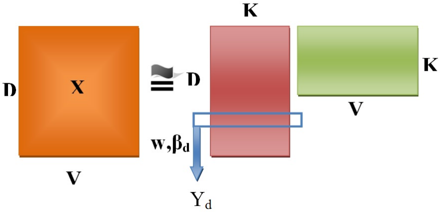
Method	C-GPPF	VEM-PF	PMF
Complexity	$O(SK + DK + VK)$	$O(K^2D + KV + KS)$	$O(DVK)$

GPPF for Joint Network and Topic Modeling (J-GPPF)



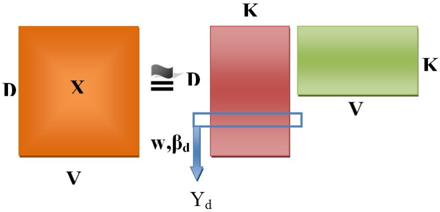


Supervised GPPF (S-GPPF)



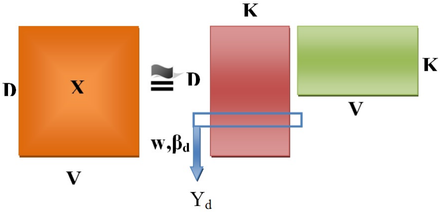
- $\theta_{dk} \sim \text{Gamma}(r_k, \exp(\beta_{dk}))$, $\beta_{dk} \sim \mathcal{N}(0, \alpha_{dk}^{-1})$, $\alpha_{dk} \sim \text{Gamma}(t_0, 1/u_0)$.

Supervised GPPF (S-GPPF)



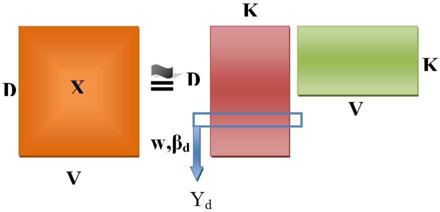
- $\theta_{dk} \sim \text{Gamma}(r_k, \exp(\beta_{dk}))$, $\beta_{dk} \sim \mathcal{N}(0, \alpha_{dk}^{-1})$, $\alpha_{dk} \sim \text{Gamma}(t_0, 1/u_0)$.
- $\phi_{wk} \sim \text{Gamma}(a_0, 1/c_w)$, $a_0 \sim \text{Gamma}(b_0, 1/c_0)$, $c_w \sim \text{Gamma}(d_0, 1/e_0)$.

Supervised GPPF (S-GPPF)



- $\theta_{dk} \sim \text{Gamma}(r_k, \exp(\beta_{dk}))$, $\beta_{dk} \sim \mathcal{N}(0, \alpha_{dk}^{-1})$, $\alpha_{dk} \sim \text{Gamma}(t_0, 1/u_0)$.
- $\phi_{wk} \sim \text{Gamma}(a_0, 1/c_w)$, $a_0 \sim \text{Gamma}(b_0, 1/c_0)$, $c_w \sim \text{Gamma}(d_0, 1/e_0)$.
- $x_{dw} \sim \text{Poisson}(\langle \boldsymbol{\theta}_d, \boldsymbol{\phi}_w \rangle)$.

Supervised GPPF (S-GPPF)



- $\theta_{dk} \sim \text{Gamma}(r_k, \exp(\beta_{dk}))$, $\beta_{dk} \sim \mathcal{N}(0, \alpha_{dk}^{-1})$, $\alpha_{dk} \sim \text{Gamma}(t_0, 1/u_0)$.
- $\phi_{wk} \sim \text{Gamma}(a_0, 1/c_w)$, $a_0 \sim \text{Gamma}(b_0, 1/c_0)$, $c_w \sim \text{Gamma}(d_0, 1/e_0)$.
- $x_{dw} \sim \text{Poisson}(\langle \boldsymbol{\theta}_d, \boldsymbol{\phi}_w \rangle)$.
- $y_d \sim \text{logit}(\langle \mathbf{w}, \boldsymbol{\beta}_d \rangle)$, $w_k \sim \mathcal{N}(0, \alpha_k^{-1})$, $\alpha_k \sim \text{Gamma}(v_0, 1/z_0)$.

Result for S-GPPF

ACM Conference Data:

Method	S-GPPF	Med-LDA	S-LDA	LDA+SVM
Accuracy	68.21%	65.13%	61.93%	57.47%

Future Works

- Dynamic Poisson Factorization (Dynamic Topic Model).
- Model for streaming count data.
- Multitask learning using supervised GPPF.
- Multi-view learning using GPPF.

Publications 2014-11

1. Coletta, Luiz Fernando, Hruschka, Eduardo R., Acharya, Ayan, and Ghosh, Joydeep, A Differential Evolution Algorithm to Optimize the Combination of Classifier and Cluster Ensembles, International Journal of Bio-Inspired Computation, 2014.
2. Acharya, Ayan, Mooney, Raymond J., and Ghosh, Joydeep, Active Multitask Learning Using Both Latent and Supervised Shared Topics, Proceedings of the 2014 SIAM International Conference on Data Mining, pp.190-198, 2014.
3. Acharya, Ayan, Hruschka, Eduardo R., Ghosh, Joydeep, Sarwar, Badrul, and Ruvini, Jean-David, Probabilistic Combination of Classifier and Cluster Ensembles for Non-transductive Learning, SDM, 2013 [.pdf].
4. Gunasekar, Suriya, Acharya, Ayan, Gaur, Neeraj, and Ghosh, Joydeep, Noisy Matrix Completion Using Alternating Minimization, ECML PKDD, Part II, LNAI 8189, pp.194-209, 2013 [.pdf].
5. Acharya, Ayan, Rawal, Aditya, Mooney, Raymond J., and Hruschka, Eduardo R., Using Both Supervised and Latent Shared Topics for Multitask Learning, ECML PKDD, Part II, LNAI 8189, pp.369-384, 2013 [.pdf].
6. Ghosh, Joydeep and Acharya, Ayan, Cluster Ensembles: Theory and Applications, in Data Clustering: Algorithms and Applications, 2013 [.pdf].
7. Acharya, Ayan, Mooney, Raymond J., Ghosh, Joydeep, Active Multitask Learning Using Doubly Supervised Latent Dirichlet Allocation, NIPS Topic Model Workshop, 2013 [.pdf].
8. Ghosh, Joydeep and Acharya, Ayan, A Survey of Consensus Clustering, Appearing in Handbook of Cluster Analysis, 2013 [.pdf].

Publications 2014-11

1. Coletta, Luiz Fernando, Hruschka, Eduardo R., Acharya, Ayan, and Ghosh, Joydeep, Towards the Use of Metaheuristics for Optimizing the Combination of Classifier and Cluster Ensembles, Appearing in 11th Brazilian Congress (CBIC) on Computational Intelligence, 2013, [.pdf].
2. Acharya, Ayan, Hruschka, Eduardo R., Ghosh, Joydeep, and Acharyya, Sreangsu, An Optimization Framework for Combining Ensembles of Classifiers and Clusterers with Applications to Non-transductive Semi-Supervised Learning and Transfer Learning, Appearing in ACM Transactions on Knowledge Discovery from Data, 2013 [.pdf].
3. Acharya, Ayan, Hruschka, Eduardo R., Ghosh, Joydeep, and Acharyya, Sreangsu, Transfer Learning with Cluster Ensembles, Journal of Machine Learning Research - Proceedings Track, 27 , pp.123-132, 2012 [.pdf].
4. Acharya, Ayan, Lee, Jangwon, and Chen, An, Real Time Car Detection and Tracking in Mobile Devices, IEEE International Conference on Connected Vehicles and Expo, 2012 [.pdf].
5. Ghosh, Joydeep and Acharya, Ayan, Cluster ensembles, Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery, 1 (4) , pp.305-315, 2011 [.pdf].
6. Acharya, Ayan, Hruschka, Eduardo R., Ghosh, Joydeep, and Acharyya, Sreangsu, C^3E: A Framework for Combining Ensembles of Classifiers and Clusterers, MCS, pp.269-278, 2011 [.pdf].
7. Acharya, Ayan, Hruschka, Eduardo R., and Ghosh, Joydeep, A Privacy-Aware Bayesian Approach for Combining Classifier and Cluster Ensembles, SocialCom/PASSAT, pp.1169-1172, 2011 [.pdf].

Questions?

References

Acharya, A., Rawal, A., Mooney, R. J., and Hruschka, E. R. (2013).
Using both supervised and latent shared topics for multitask learning.
In *ECML PKDD, Part II, LNAI 8189*, pages 369–384.

Bordes, A., Bottou, L., Gallinari, P., and Weston, J. (2007).
Solving multiclass support vector machines with larank.
In *Proc. of ICML*, pages 89–96.

Caruana, R. (1997).
Multitask learning.
Machine Learning, 28:41–75.

Lampert, C. H., Nickisch, H., and Harmeling, S. (2009).
Learning to detect unseen object classes by betweenclass attribute transfer.
In *Proc. of CVPR*, pages 951–958.

Neal, R. M. and Hinton, G. E. (1999).
A view of the EM algorithm that justifies incremental, sparse, and other variants.

Nigam, K., McCallum, A., Thrun, S., and Mitchell, T. (1998).
Learning to classify text from labeled and unlabeled documents.
In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 792–799. AAAI Press.