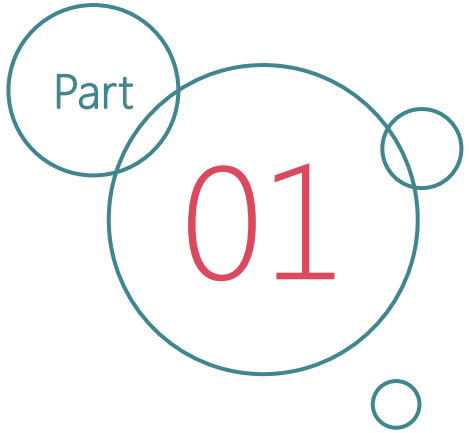
用于推荐系统的协同深度学习算法

Collaborative Deep Learning for Recommender Systems

贺语盈

OUTLINE

1. Overview for Recommender Systems
2. Probabilistic Matrix Factorization
3. Collaborative Topic Regression
4. Collaborative Deep Learning Models
5. Experiment Results
6. Possible Extensions



Overview for Recommender Systems



Recommender Systems













shopping web

music player

SNS

...Where are those recommender systems?



Recommender Systems

This paper is to develop recommendation algorithms

- CDL can extract an effective deep feature representation from content the learned representation may also be used for tasks other than recommendation.
- Unlike previous deep learning models which use simple target like classification and reconstruction, we propose to use CF as a more complex target in a probabilistic framework.
- To the best of our knowledge, CDL is the first hierarchical Bayesian model to bridge the gap between state-of-the-art deep learning models and RS.
- Extensive experiments on three real-world datasets from different domains show that CDL can significantly advance the state of the art.



Recommender Systems

User's interests are summarized using top topics he/she is interested in.

user 2832

top 5 interested topics

{ learning machine training vector learn machines kernel learned classifiers classifier }

{ retrieval documents collection document relevance topic index collections indexing latent }

{ terms relevant query queries term effectiveness expansion keywords log refinement }

{ search internet strategies searching engine engines searches google millions locate }

{ distribution random probability distributions sampling stochastic markov density estimation statistics }

Recommender Systems

Recommendations made by the algorithm:

top 25 recommendations

Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Lik

Combining Labeled and Unlabeled Data with Co-training

Conditional random fields: Probabilistic models for segmenting and labeling sequence

Experiments with a New Boosting Algorithm

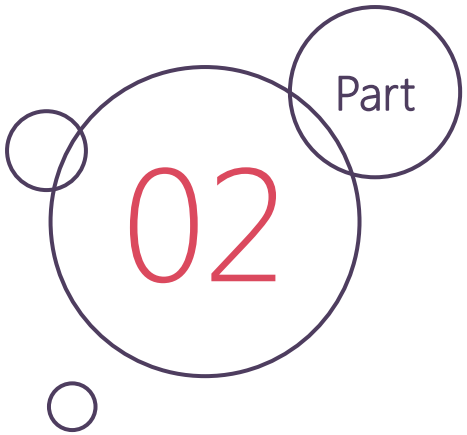
Statistical Learning Theory

Text Classification from Labeled and Unlabeled Documents using {EM}

Transductive Inference for Text Classification using Support Vector Machines

Indexing by Latent Semantic Analysis

Probabilistic Matrix Factorization



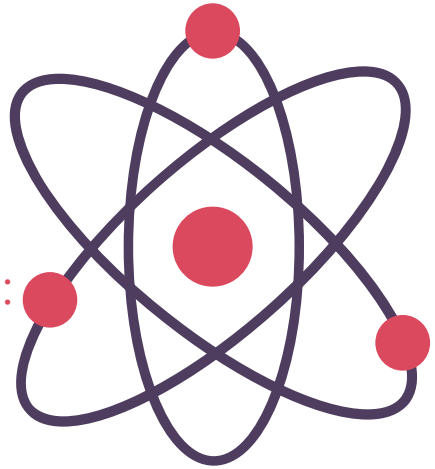
Probabilistic Matrix Factorization

Three important elements:

- users
- items
- ratings

Three steps to make recommendation:

- Collect users preferences
- Find similar users or items
- Generate recommend results.



Three categories of method for RS

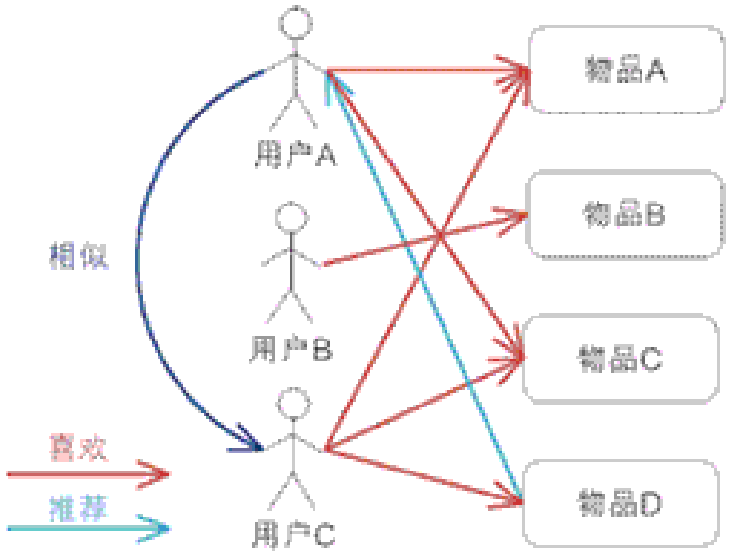
- Content-based: user profiles or product descriptions
- CF-based: past activities or preferences
- Hybrid methods: combine content-based and CF-based methods.

Probabilistic Matrix Factorization

Popular solutions: collaborative filtering (CF)

User CF

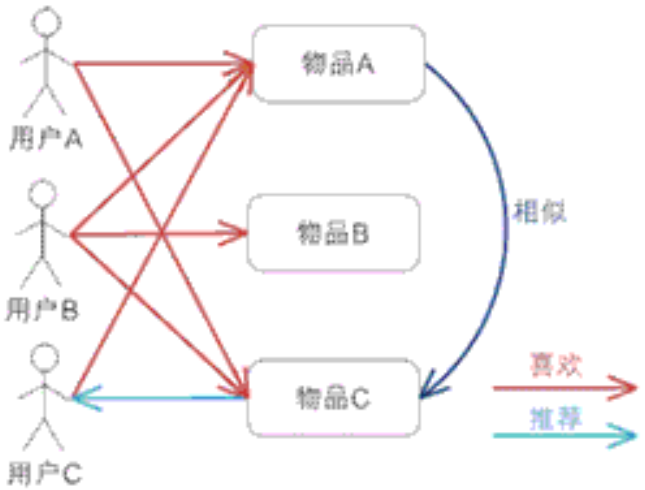
用户/物品	物品A	物品B	物品C	物品D
用户A	√		√	推荐
用户B		√		
用户C	√		√	√



Probabilistic Matrix Factorization

Item CF

用户/物品	物品A	物品B	物品C
用户A	√		√
用户B	√	√	√
用户C	√		推荐



Probabilistic Matrix Factorization

matrix factorization: one of the most popular algorithms for recommender system.

User-item matrix:

user \ item	1	2	3
1	✓	×	?
2	✓	?	×
3	?	✓	×

$$\Rightarrow \begin{bmatrix} 1 & 0 & ? \\ 1 & ? & 0 \\ ? & 1 & 0 \end{bmatrix}$$



Probabilistic Matrix Factorization

- Users and items are represented in a shared but unknown latent space,

$$\text{user } i \text{—} \; u_i \in R^K .$$
$$\text{Item } j \text{—} \; v_j \in R^K .$$

- The rating of item j by user i is achieved by the dot product,

$$r_{ij} = u_i^T v_j$$

where $r_{ij} = 1$ indicates like and 0 may be due to the fact that the user is not interested in the item, or that the user is not aware of its existence .

- In the matrix form,

$$R = U^T V.$$

Probabilistic Matrix Factorization

- For each user i, draw user latent vector $u_i \sim N(0, \lambda_u^{-1} I_K)$.
- For each item j, draw user latent vector $v_i \sim N(0, \lambda_v^{-1} I_K)$.
- For each user-item pair(i,j),draw the rating

$$r_{ij} \sim N(u_i^T v_j, c_{ij}^{-1})$$

where c_{ij} is the precision parameter.

When $c_{ij} = 1$, the maximum a posteriori estimation (MAP) of PMF corresponds to the solution in:

$$\min_{U,V} \sum_{i,j} (r_{ij} - u_i^T v_j)^2 + \lambda_u ||u_i||^2 + \lambda_v ||v_j||^2,$$

where λ_u and λ_v are regularization parameters

. Here, c_{ij} serves as a confidence parameter for rating r_{ij}. If c_{ij} is large, we trust r_{ij} more.

$$c_{ij} = \begin{cases} a, & if\ rij = 1, \\ b, & if\ rij = 0, \end{cases}$$

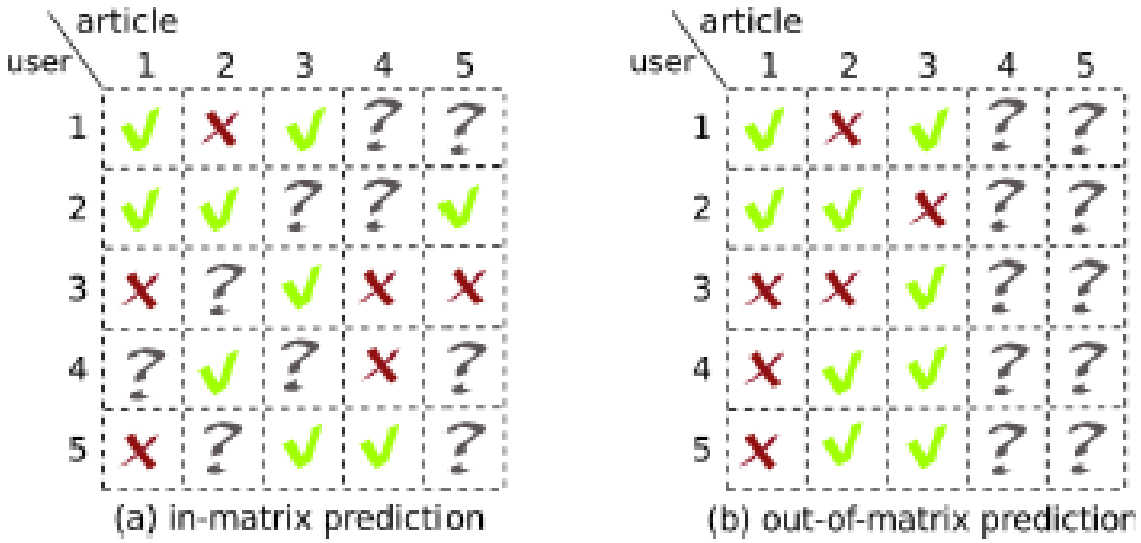
where a and b are tunning parameters saitifying $a > b > 0$.

Probabilistic Matrix Factorization

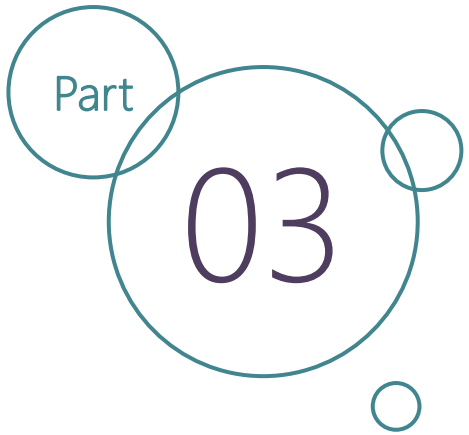
There are two main disadvantages to matrix factorization for recommendation.

First, the learnt latent space is not easy to interpret;

second, as mentioned, matrix factorization only uses information from other users—it cannot generalize to completely unrated items.



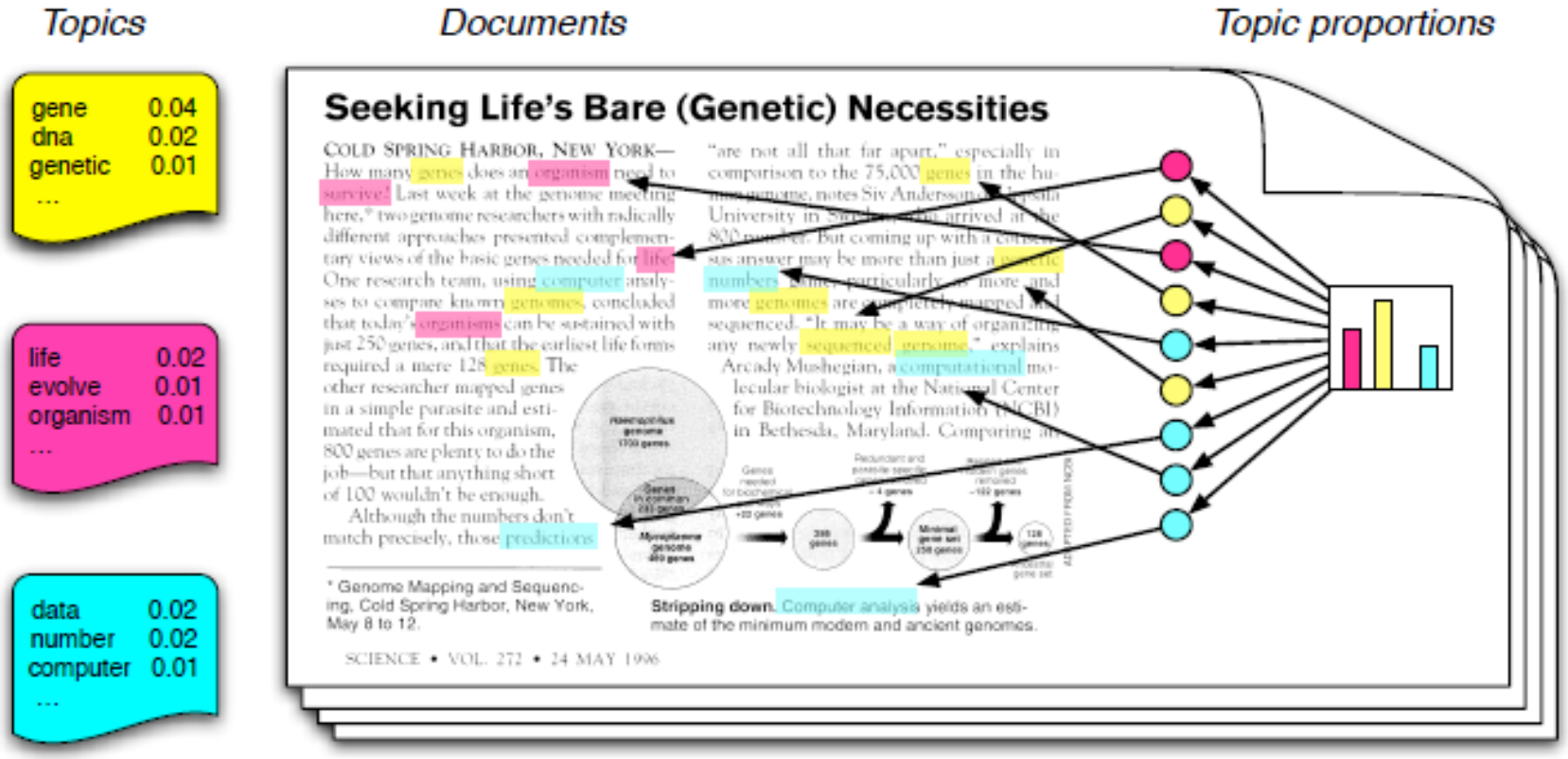
cold start



Collaborative Topic Regression

Collaborative Topic Regression

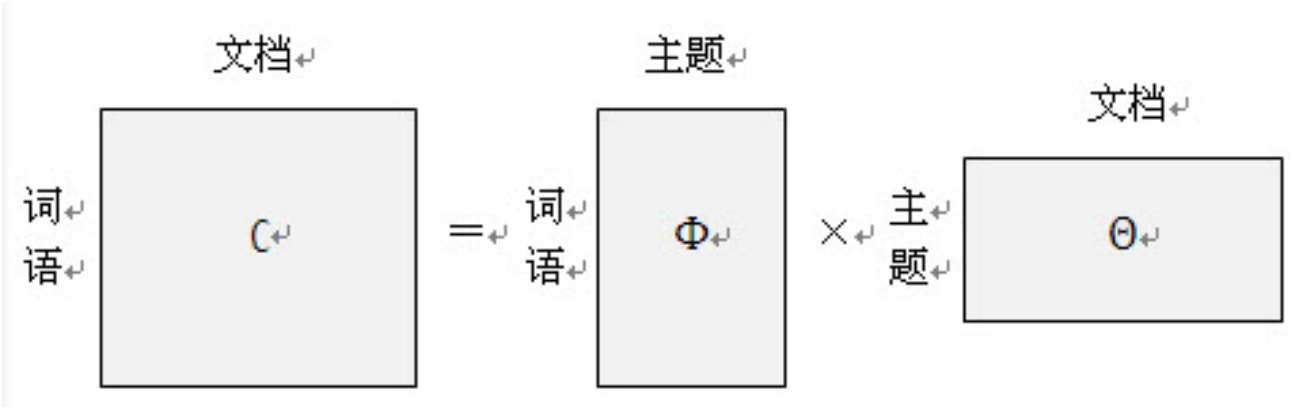
Topic modeling



Collaborative Topic Regression

Latent Dirichlet allocation (LDA) is a popular topic model. Assumes there are K topics $\beta=\beta_{1:k}$.

- For each article, topic proportions $\theta_j \sim$ Dirichlet(α).
- For each word n,
  (a)Draw topic assignment z_{jn}~Mult(θ_j).
  (b)Draw word wj_n~Mult(β_{zjn}).



Collaborative Topic Regression

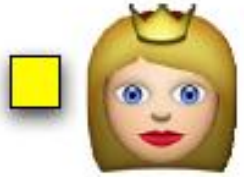
Two topics:

machine learning social network analysis

Two people:



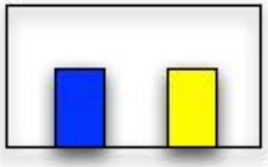
boy



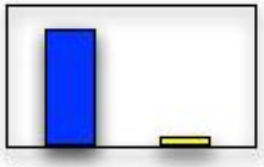
girl

Two articles:

article A



article B



Preferences:

Boy likes both A and B.

Girl prefers A.

Collaborative Topic Regression

- For each user i, draw user latent vector $u_i \sim \mathcal{N}(0, \lambda_u^{-1} I_k)$.

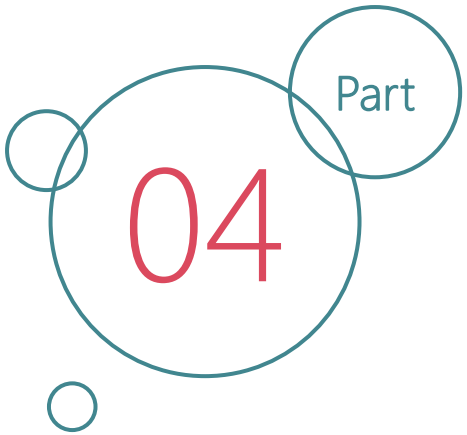
- For each item j,
  (a)Draw topic proportions $\theta_j \sim$ Dirichlet(α).
  (b)Draw item latent offset $\varepsilon_j \sim \mathcal{N}(0, \lambda_v^{-1} I_k)$ and set the item latent vector as $v_j = \theta_j + \varepsilon_j$.
  (c) For each word n,
    i. Draw topic assignment $z_{jn} \sim$ Mult(θ_j).
    ii. Draw word $w_{jn} \sim$ Mult(β_{zjn}).

- For each user-item pair(i,j),draw the rating

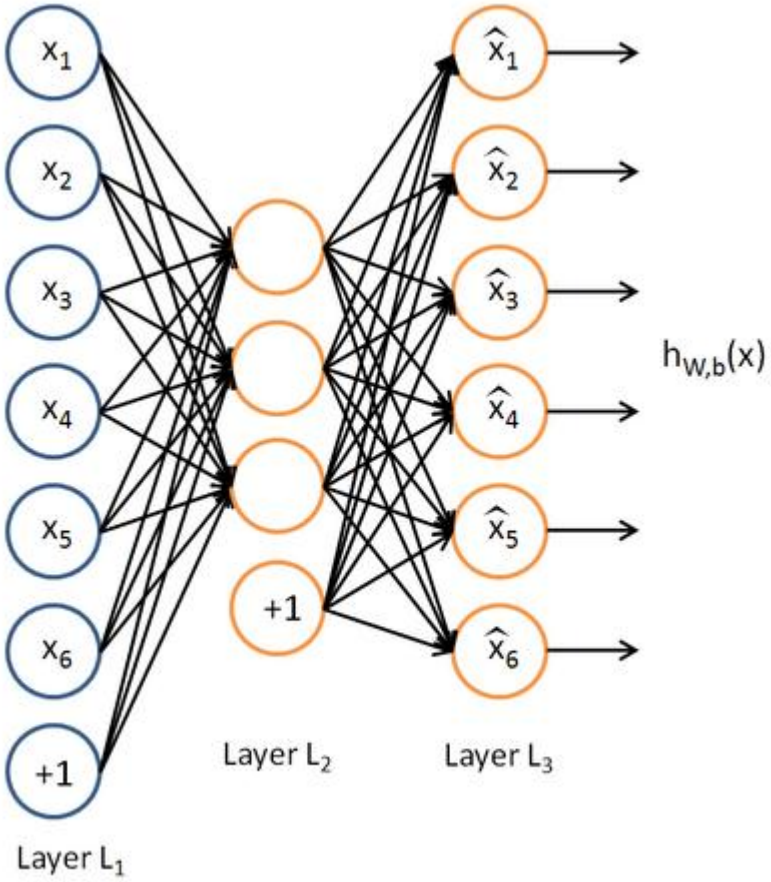
$$r_{ij} \sim \mathcal{N}(u_i^T v_j, c_{ij}^{-1})$$

We call the model Collaborative Topic Regression (CTR). CTR combines traditional collaborative filtering(CF) with top modeling.

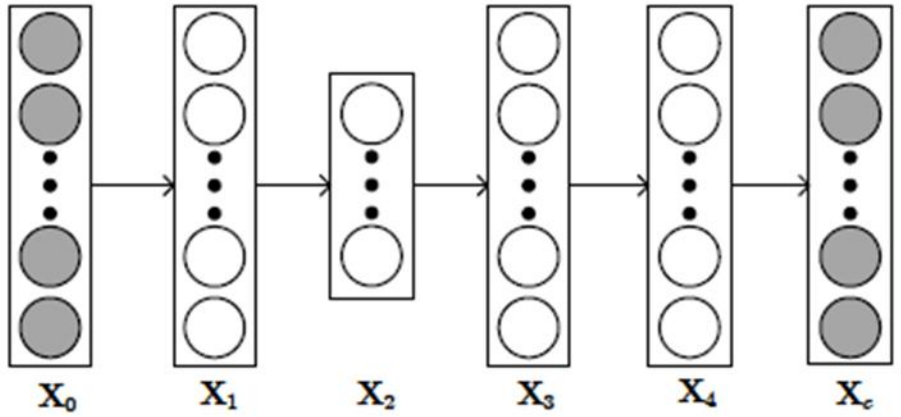
Collaborative Deep Learning



Collaborative Deep Learning



Stacked Denoising Autoencoders



SDAE solves the following optimization problem:

$$\min_{\{\mathbf{W}_l\},\{\mathbf{b}_l\}} \|\mathbf{X}_c - \mathbf{X}_L\|_F^2 + \lambda \sum_l \|\mathbf{W}_l\|_F^2,$$

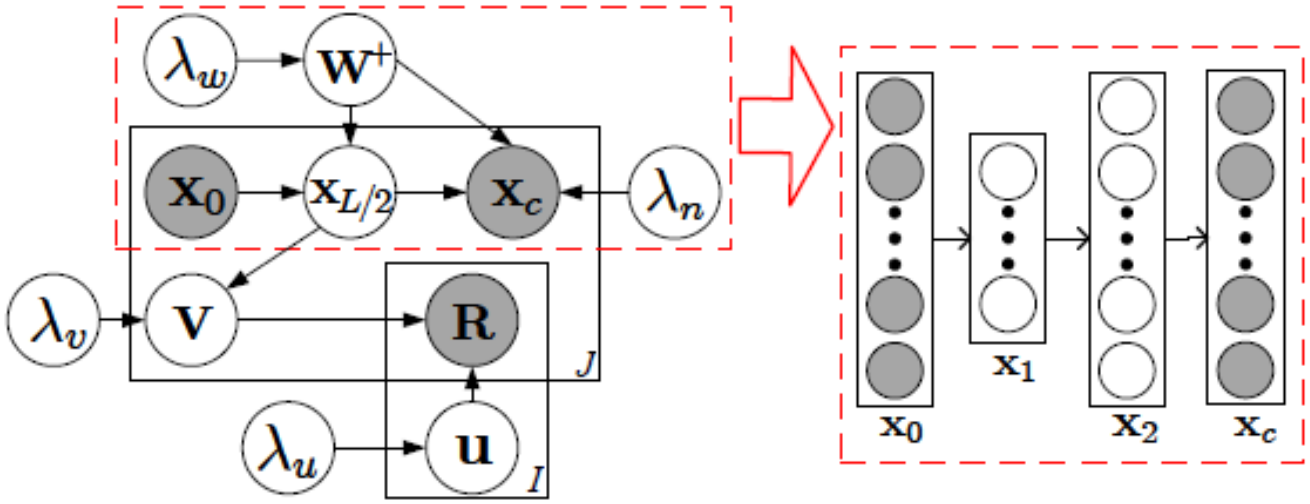
Collaborative Deep Learning

Collaborative Deep Learning(CDL): Using the Bayesian SDAE as a component, the generative process of CDL is denoted as follows:

1. For each layer l of the SDAE network,
   (a) For each column n of the weight matrix $\mathbf{W}_l$, draw
   $$\mathbf{W}_{l,*n} \sim \mathcal{N}(\mathbf{0}, \lambda_w^{-1}\mathbf{I}_{K_l}).$$
   (b) Draw the bias vector $\mathbf{b}_l \sim \mathcal{N}(\mathbf{0}, \lambda_w^{-1}\mathbf{I}_{K_l})$.
   (c) For each row j of $\mathbf{X}_l$, draw
   $$\mathbf{X}_{l,j*} \sim \mathcal{N}(\sigma(\mathbf{X}_{l-1,j*}\mathbf{W}_l + \mathbf{b}_l), \lambda_s^{-1}\mathbf{I}_{K_l}).$$
2. For each item j,
   (a) Draw a clean input $\mathbf{X}_{c,j*} \sim \mathcal{N}(\mathbf{X}_{L,j*}, \lambda_n^{-1}\mathbf{I}_J)$.
   (b) Draw a latent item offset vector $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \lambda_v^{-1}\mathbf{I}_K)$ and then set the latent item vector to be:
   $$\mathbf{v}_j = \boldsymbol{\epsilon}_j + \mathbf{X}_{\frac{L}{2},j*}^T.$$
3. Draw a latent user vector for each user i:
   $$\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \lambda_u^{-1}\mathbf{I}_K).$$
4. Draw a rating $\mathbf{R}_{ij}$ for each user-item pair (i, j):
   $$\mathbf{R}_{ij} \sim \mathcal{N}(\mathbf{u}_i^T\mathbf{v}_j, \mathbf{C}_{ij}^{-1}).$$

Collaborative Deep Learning

The graphical model



Collaborative Deep Learning

Maximum A Posteriori Estimates

$$\mathscr{L} = -\frac{\lambda_u}{2}\sum_i \|\mathbf{u}_i\|_2^2 - \frac{\lambda_w}{2}\sum_l (\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2) - \frac{\lambda_v}{2}\sum_j \|\mathbf{v}_j - \mathbf{X}_{\frac{L}{2},j*}^T\|_2^2 - \frac{\lambda_n}{2}\sum_j \|\mathbf{X}_{L,j*} - \mathbf{X}_{c,j*}\|_2^2 - \frac{\lambda_s}{2}\sum_l\sum_j \|\sigma(\mathbf{X}_{l-1,j*}\mathbf{W}_l + \mathbf{b}_l) - \mathbf{X}_{l,j*}\|_2^2 - \sum_{i,j}\frac{\mathbf{C}_{ij}}{2}(\mathbf{R}_{ij} - \mathbf{u}_i^T\mathbf{v}_j)^2.$$

$$\mathscr{L} = -\frac{\lambda_u}{2}\sum_i \|\mathbf{u}_i\|_2^2 - \frac{\lambda_w}{2}\sum_l (\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2) - \frac{\lambda_v}{2}\sum_j \|\mathbf{v}_j - f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T\|_2^2 - \frac{\lambda_n}{2}\sum_j \|f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+) - \mathbf{X}_{c,j*}\|_2^2 - \sum_{i,j}\frac{\mathbf{C}_{ij}}{2}(\mathbf{R}_{ij} - \mathbf{u}_i^T\mathbf{v}_j)^2,$$



Collaborative Deep Learning

Update rules:

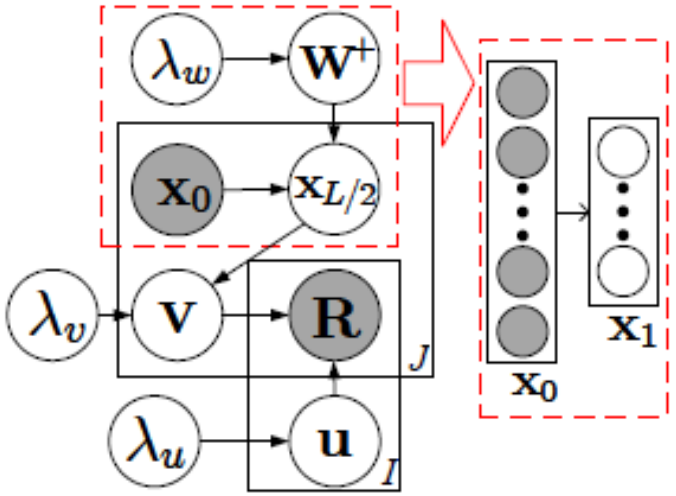
$$\mathbf{u}_i \leftarrow (\mathbf{V}\mathbf{C}_i\mathbf{V}^T + \lambda_u \mathbf{I}_K)^{-1}\mathbf{V}\mathbf{C}_i\mathbf{R}_i$$
$$\mathbf{v}_j \leftarrow (\mathbf{U}\mathbf{C}_i\mathbf{U}^T + \lambda_v \mathbf{I}_K)^{-1}(\mathbf{U}\mathbf{C}_j\mathbf{R}_j + \lambda_v f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T),$$

$$\begin{aligned}\nabla_{\mathbf{W}_l}\mathscr{L} &= -\lambda_w \mathbf{W}_l \\ &- \lambda_v \sum_j \nabla_{\mathbf{W}_l} f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T (f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T - \mathbf{v}_j) \\ &- \lambda_n \sum_j \nabla_{\mathbf{W}_l} f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+)(f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+) - \mathbf{X}_{c,j*})\end{aligned}$$

$$\begin{aligned}\nabla_{\mathbf{b}_l}\mathscr{L} &= -\lambda_w \mathbf{b}_l \\ &- \lambda_v \sum_j \nabla_{\mathbf{b}_l} f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T (f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T - \mathbf{v}_j) \\ &- \lambda_n \sum_j \nabla_{\mathbf{b}_l} f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+)(f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+) - \mathbf{X}_{c,j*}).\end{aligned}$$

Collaborative Deep Learning

Extreme cases

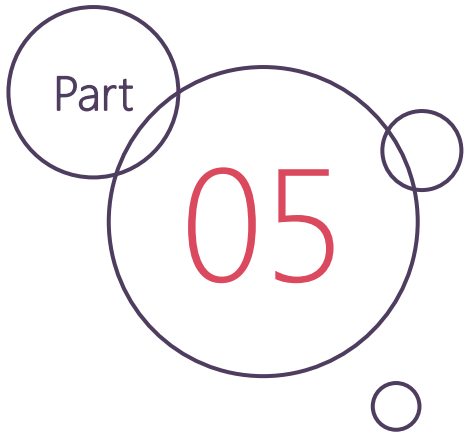


1 $\lambda n/\lambda v \to +\infty$

it will degenerate to a two-step model in which the latent representation learned using SDAE is put directly into the CTR.

2 $\lambda n/\lambda v \to 0$

the decoder of the SDAE essentially vanishes. On the right of Figure is the graphical model of the degenerated CDL when $\lambda n/\lambda v$ goes to zero.



Experiment Results

Experiment Results

Datasets from

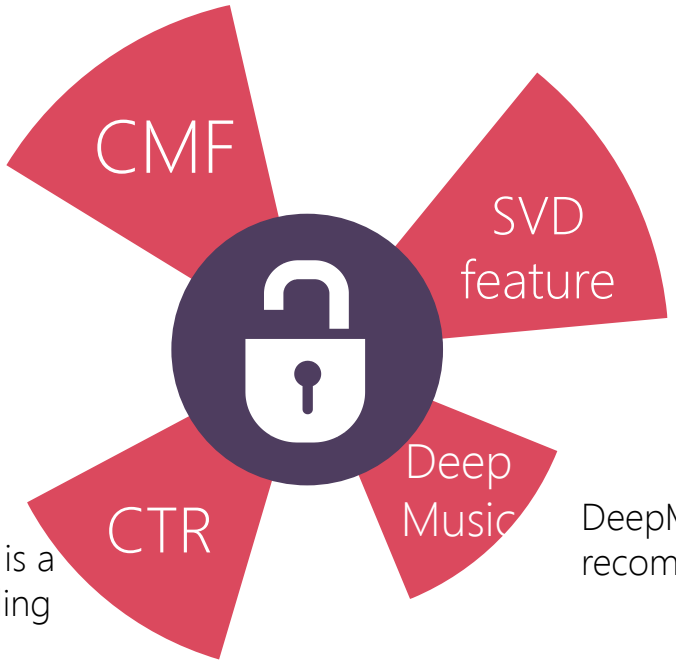
- CiteULike
  (a)citeulike-a: 5551 users and 16980 items
  (b)citeulike-b: 7947 users and 25975 items
- Netfix: 40761 users, 9928 movies and 15348808 ratings

Evaluation: five-fold cross-validation with recall,

$$\text{recall@M} = \frac{\text{number of articles the user likes in top M}}{\text{total number of article the user likes}}$$

Experiment Results

Collective Matrix Factorization is a model incorporating different sources of information by simultaneously factorizing multiple matrices.



SVDFeature is a model for feature-based collaborative filtering.

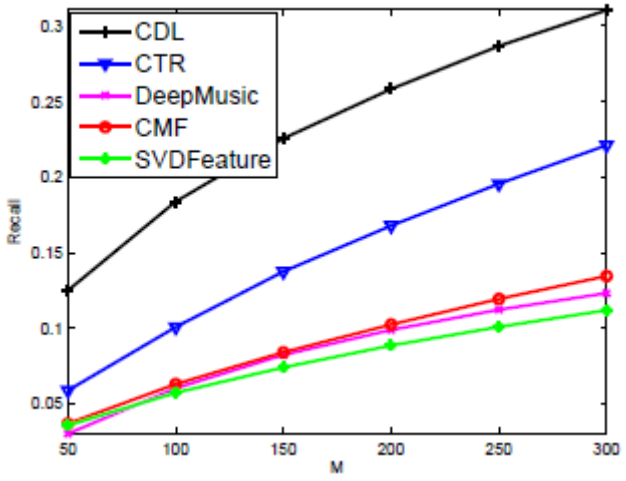
Collaborative Topic Regression is a model performing topic modeling and collaborative filtering simultaneously

DeepMusic is a model for music recommendation

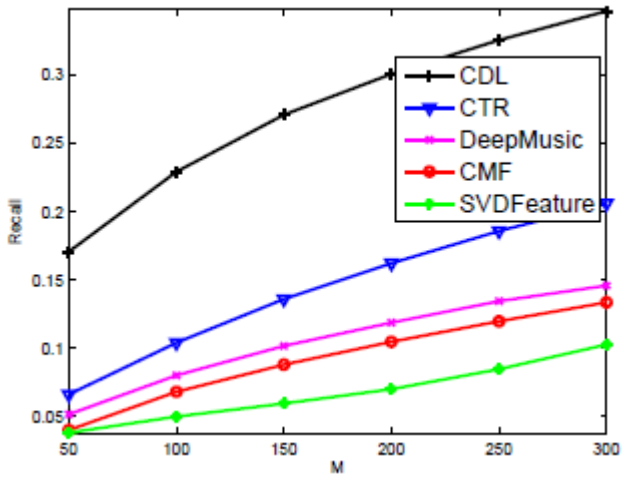
Experiment Results

recall@M in the sparse setting

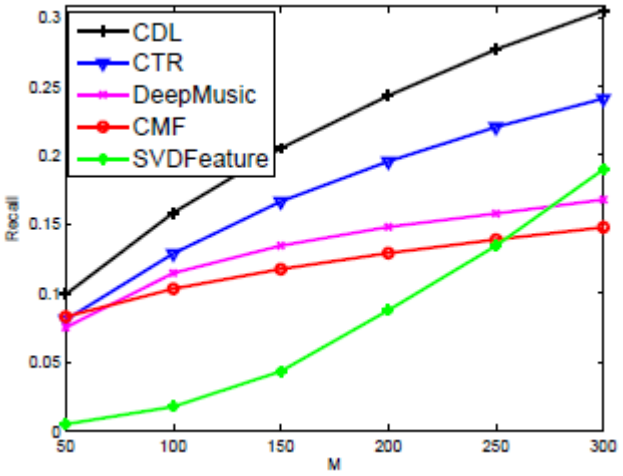
citeulike-a



citeulike-t



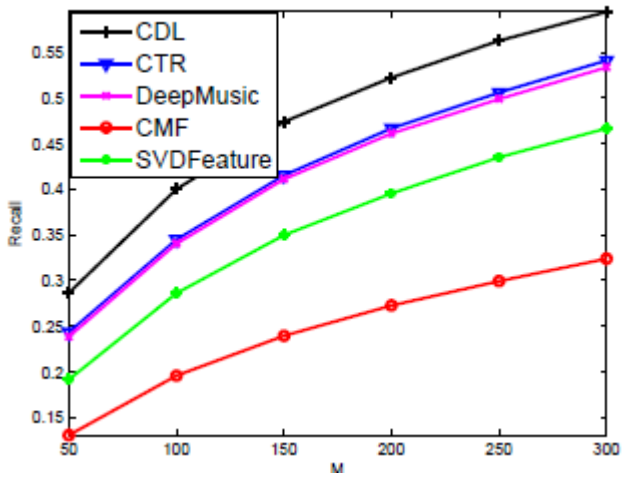
Netflix



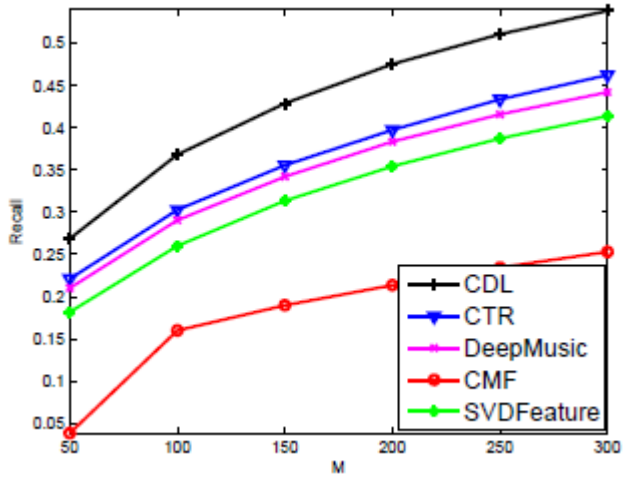
Experiment Results

recall@M in the dense setting

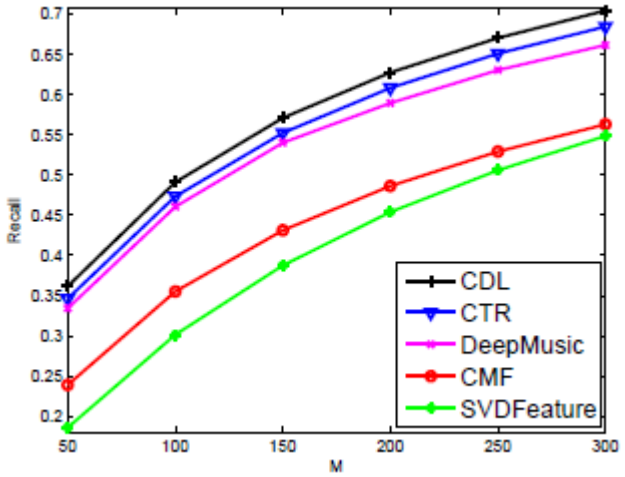
citeulike-a



citeulike-t



Netflix





Experiment Results

Table 1: mAP for three datasets

	citeulike-a	*citeulike-t*	*Netflix*
CDL	**0.0514**	**0.0453**	**0.0312**
CTR	0.0236	0.0175	0.0223
DeepMusic	0.0159	0.0118	0.0167
CMF	0.0164	0.0104	0.0158
SVDFeature	0.0152	0.0103	0.0187

Table 2: Recall@300 in the sparse setting (%)

#layers	1	2	3
citeulike-a	27.89	**31.06**	30.70
citeulike-t	32.58	34.67	**35.48**
Netflix	29.20	30.50	**31.01**

Table 3: Recall@300 in the dense setting (%)

#layers	1	2	3
citeulike-a	58.35	**59.43**	59.31
citeulike-t	52.68	53.81	**54.48**
Netflix	69.26	70.40	**70.42**

Experiment Results

Table 4: Interpretability of the latent structures learned

	user I (CDL)	in user's lib?
top 3 topics	1. search, image, query, images, queries, tagging, index, tags, searching, tag 2. social, online, internet, communities, sharing, networking, facebook, friends, ties, participation 3. collaborative, optimization, filtering, recommendation, contextual, planning, items, preferences	
top 10 articles	**1. The structure of collaborative tagging Systems** **2. Usage patterns of collaborative tagging systems** 3. Folksonomy as a complex network **4. HT06, tagging paper, taxonomy, Flickr, academic article, to read** **5. Why do tagging systems work** 6. Information retrieval in folksonomies: search and ranking **7. tagging, communities, vocabulary, evolution** **8. The complex dynamics of collaborative tagging** 9. Improved annotation of the blogosphere via autotagging and hierarchical clustering **10. Collaborative tagging as a tripartite network**	**yes** **yes** no **yes** **yes** no **yes** **yes** no **yes**
	user I (CTR)	in user's lib?
top 3 topics	1. social, online, internet, communities, sharing, networking, facebook, friends, ties, participation 2. search, image, query, images, queries, tagging, index, tags, searching, tag 3. feedback, event, transformation, wikipedia, indicators, vitamin, log, indirect, taxonomy	
top 10 articles	**1. HT06, tagging paper, taxonomy, Flickr, academic article, to read** 2. Structure and evolution of online social networks 3. Group formation in large social networks: membership, growth, and evolution 4. Measurement and analysis of online social networks 5. A face(book) in the crowd: social searching vs. social browsing 6. The strength of weak ties 7. Flickr tag recommendation based on collective knowledge 8. The computer-mediated communication network 9. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis 10. Increasing participation in online communities: A framework for human-computer interaction	**yes** no no no no no no no no no



Experiment Results

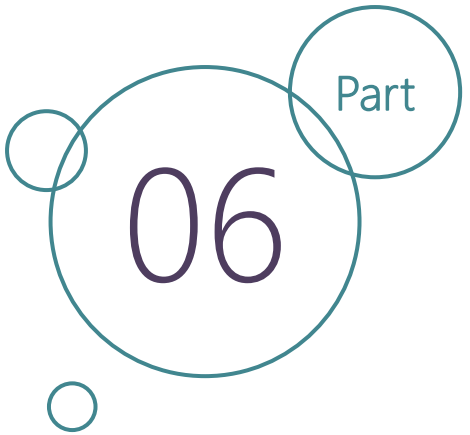
	user II (CDL)	in user's lib?
top 3 topics	1. flow, cloud, codes, matter, boundary, lattice, particles, galaxies, fluid, galaxy 2. mobile, membrane, wireless, sensor, mobility, lipid, traffic, infrastructure, monitoring, ad 3. hybrid, orientation, stress, fluctuations, load, temperature, centrality, mechanical, two-dimensional, heat	
top 10 articles	**1. Modeling the flow of dense suspensions of deformable particles in three dimensions**	**yes**
	2. Simplified particulate model for coarse-grained hemodynamics simulations	**yes**
	3. Lattice Boltzmann simulations of blood flow: non-newtonian rheology and clotting processes	**yes**
	4. A genome-wide association study for celiac disease identifies risk variants	**yes**
	5. Efficient and accurate simulations of deformable particles	**yes**
	6. A multiscale model of thrombus development	**yes**
	7. Multiphase hemodynamic simulation of pulsatile flow in a coronary artery	**yes**
	8. Lattice Boltzmann modeling of thrombosis in giant aneurysms	**yes**
	9. A lattice Boltzmann simulation of clotting in stented aneursysms	**yes**
	10. Predicting dynamics and rheology of blood flow	**yes**
	user II (CTR)	in user's lib?
top 3 topics	1. flow, cloud, codes, matter, boundary, lattice, particles, galaxies, fluid, galaxy 2. transition, equations, dynamical, discrete, equation, dimensions, chaos, transitions, living, trust 3. mobile, membrane, wireless, sensor, mobility, lipid, traffic, infrastructure, monitoring, ad	
top 10 articles	**1. Multiphase hemodynamic simulation of pulsatile flow in a coronary artery**	**yes**
	2. The metallicity evolution of star-forming galaxies from redshift 0 to 3	no
	3. Formation versus destruction: the evolution of the star cluster population in galaxy mergers	no
	4. Clearing the gas from globular clusters	no
	5. Macroscopic effects of the spectral structure in turbulent flows	no
	6. The WiggleZ dark energy survey	no
	7. Lattice-Boltzmann simulation of blood flow in digitized vessel networks	no
	8. Global properties of 'ordinary' early-type galaxies	no
	9. Proteus : a direct forcing method in the simulations of particulate flows	**yes**
	10. Analysis of mechanisms for platelet near-wall excess under arterial blood flow conditions	**yes**

Experiment Results

Table 5: Example user with recommended movies

User III	Movies in the training set: **Moonstruck, True Romance, Johnny English, American Beauty, The Princess Bride, Top Gun, Double Platinum, Rising Sun, Dead Poets Society, Waiting for Guffman**		
# training samples	2	4	10
Top 10 recommended movies by CTR	Swordfish	**Pulp Fiction**	**Best in Snow**
	A Fish Called Wanda	A Clockwork Orange	**Chocolat**
	Terminator 2	Being John Malkovich	**Good Will Hunting**
	A Clockwork Orange	**Raising Arizona**	**Monty Python and the Holy Grail**
	Sling Blade	Sling Blade	Being John Malkovich
	Bridget Jones's Diary	Swordfish	**Raising Arizona**
	Raising Arizona	A Fish Called Wanda	The Graduate
	A Streetcar Named Desire	Saving Grace	Swordfish
	The Untouchables	The Graduate	Tootsie
	The Full Monty	Monster's Ball	Saving Private Ryan
# training samples	2	4	10
Top 10 recommended movies by CDL	Snatch	**Pulp Fiction**	**Good Will Hunting**
	The Big Lebowski	Snatch	**Best in Show**
	Pulp Fiction	**The Usual Suspect**	**The Big Lebowski**
	Kill Bill	Kill Bill	**A Few Good Men**
	Raising Arizona	Momento	**Monty Python and the Holy Grail**
	The Big Chill	**The Big Lebowski**	**Pulp Fiction**
	Tootsie	**One Flew Over the Cuckoo's Nest**	The Matrix
	Sense and Sensibility	As Good as It Gets	**Chocolat**
	Sling Blade	**Goodfellas**	**The Usual Suspect**
	Swinger	The Matrix	**CaddyShack**

Possible Extensions





Possible Extensions

- The bag-of-words representation may be replaced by more powerful alternatives.
- The Bayesian nature of CDL also provides potential performance boost if other side information is incorporated
- CDL actually provides a framework that can also admit deep learning models other than SDAE. One promising choice is the convolutional neural network model

THANKS