Bayesian Multi-relational Clustering

Hanhuai Shan and Arindam Banerjee

Department of Computer Science and Engineering
University of Minnesota, Twin Cities

Introduction

Multi-relational data –capture relations among multiple entities of interest.

Multi-relational clustering –discover latent clusters in different entities *and their relationships*, for better recommendation, serving, etc., e.g., to obtain a better prediction of customer behavior, it is important to discover a group of customers who *tend to buy* similar products *sold by* a certain group of sellers. **Bayesian multi-relational clustering** –allow each object to belong to multiple clusters with varying degrees, applicable to various data types

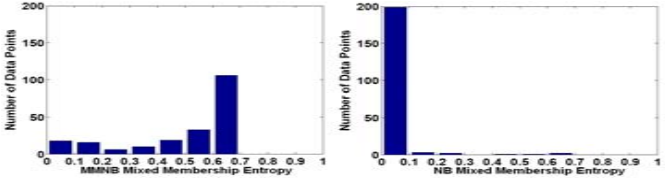
Mixed membership naïve Bayes models

–One-way Bayesian clustering

- A naïve Bayes model with Dirichlet priors to allow mixed membership

	Micro-precision		Perplexity	
	MMNB	NB	MMNB	NB
Ecoli	**0.8334±0.1127**	0.7656±0.1009	0.0134±0.0018	**0.0115±0.0001**
Glass	**0.6389±0.0744**	0.5976±0.0748	**0.0783±0.0232**	0.2383±0.1415
Ionosphere	0.6514±0.0500	**0.6886±0.0767**	**1.5035±0.1878**	1.7093±0.1863
Statlog-seg	**0.5874±0.0544**	0.5619±0.0675	**1.3032±0.0517**	2.2422±0.2536
Segmentation	**0.6476±0.0930**	0.6333±0.1100	**1.4215±0.1770**	2.6365±1.3470
Sonar	**0.6051±0.0550**	0.6050±0.0438	**0.3043±0.0147**	0.3161±0.0146
Vowel	**0.3990±0.0258**	0.3222±0.0450)	0.7709±0.0220	**0.6416±0.0209**
Wdbc	**0.9161±0.0253**	0.9089±0.0351	**0.7974±0.0784**	0.8090±0.0874
Wine	**0.9294±0.0668**	0.6882±0.0834	**4.5722±0.6454**	5.0251±0.4230

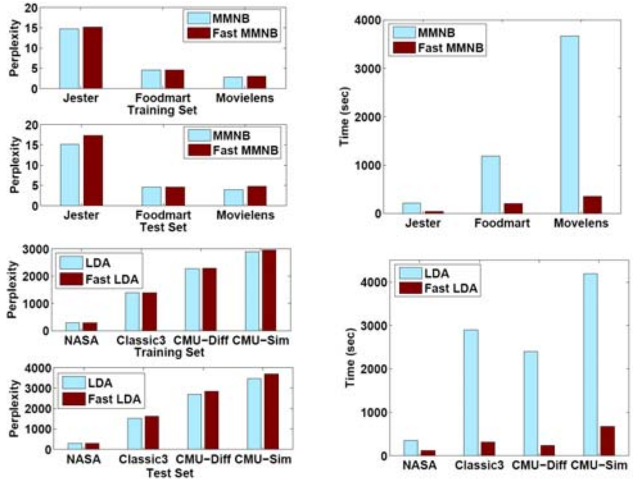
MMNB has better precision and perplexity than NB on UCI dataset



Mixed membership entropy histogram from MMNB and NB

MMNB's mixed membership vs NB's sole membership on Sonar

Fast mixed membership naïve Bayes & Fast latent Dirichlet allocation



By using a fast variational inference algorithm, Fast MMNB and Fast LDA are about *10 times faster* than standard MMNB and LDA, with a similar perplexity.

Discriminative latent Dirichlet allocation

- Introducing labels to latent Dirichlet allocation for classification
- Allow the number of classes to be larger than the number of topics

	Nasa	Classic3	Cmu-diff	Cmu-sim	Cmu-same
Fast DLDA (c)	0.9237 ±0.0163	0.6756 ±0.0234	0.9800 ±0.0102	0.8653 ±0.0182	0.7900 ±0.0315
Fast DLDA ($c+15$)	0.9232 ±0.0144	0.6858 ±0.0216	0.9747 ±0.0121	0.8713 ±0.0264	0.8458 ±0.0214
Fast DLDA ($c+30$)	0.9301 ±0.0128	0.6838 ±0.0234	0.9817 ±0.0099	0.8707 ±0.0228	***0.8468 ±0.0190***
Fast DLDA ($c+50$)	0.9237 ±0.0138	0.6854 ±0.0211	***0.9823 ±0.0083***	0.8700 ±0.0230	0.8150 ±0.0184
Fast DLDA ($c+100$)	0.9261 ±0.0102	***0.6866 ±0.0245***	0.9760 ±0.0108	***0.8718 ±0.0182***	0.8347 ±0.0187
vMF	0.9216 ±0.0113	0.6509 ±0.0246	0.9530 ±0.0071	0.7447 ±0.0214	0.7600 ±0.0347
NB	***0.9334 ±0.0094***	0.6766 ±0.0230	0.9813 ±0.0069	0.8613 ±0.0216	0.8410 ±0.0262
LR	0.9209 ±0.0157	0.6396 ±0.0252	0.9553 ±0.0157	0.6750 ±0.1330	0.4823 ±0.1283
SVM	0.9192 ±0.0146	0.6854 ±0.0278	0.9563 ±0.0105	0.8357 ±0.0156	0.8120 ±0.203

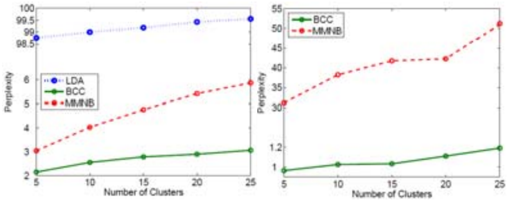
#topics increases from c to c+100. A higher accuracy is usually observed with a larger #topics.

DLDA is better than or comparable to other classification algorithms, including SVM.

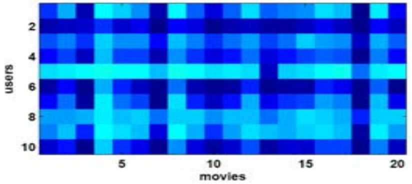
Bayesian Co-clustering

–Two-way Bayesian clustering

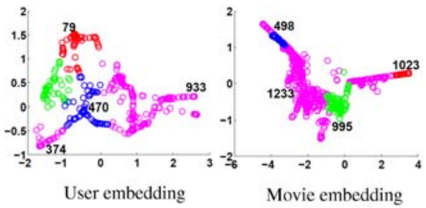
- Given a data matrix denoting the relationship between row and column entities, obtaining row and column clusters simultaneously
- A Bayesian model allowing mixed membership on both sides



Perplexity on binary Jester(left) & original Movielens(right) with varying number of user clusters. BCC has the lowest perplexity

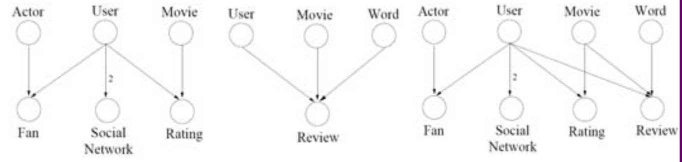


Extracted relationship between user and movie clusters on movielens. A darker block indicates a better preference from the corresponding user cluster to movie cluster



Mixed memberships for rows and columns are considered as inter-dependent low-dimensional representation of rows and columns (co-embedding). Projecting them to a two-dimensional space gives a visualization for both users and movies in movielens

Future work—Bayesian multi-relational clustering



Book Model—One entity (User) is connected with multiple other entities (Actor, Movie, Other Users) through corresponding relations.

Tensor Model—Multiple entities (User, Movie, Word) are connected through one relation.

General case—A combination of Book Model and Tensor Model. Arbitrary forms of relationships

UNIVERSITY OF MINNESOTA