PURDUE UNIVERSITY | Department of Computer Science

CS47300: Web Information Search and Management

Collaborative Filtering: Model-Based Approaches

Prof. Chris Clifton

16 October 2020

Material adapted from course created by Dr. Luo Si, now leading Alibaba research group

Indiana Center for Database Systems ™

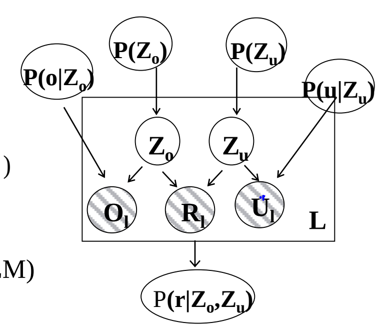
PURDUE UNIVERSITY

Department of Computer Science

Collaborative Filtering

- Flexible Mixture Model (FMM):

  Cluster users and objects separately AND allow them to belong to different classes



$$P(o_{(l)},u_{(l)},r_{(l)}) = \sum_{Z_o,Z_u} P(Z_o)P(Z_u)P(o_{(l)}\,|\,Z_o)P(u_{(l)}\,|\,Z_u)P(r_{(l)}\,|\,Z_o,Z_u)$$

- Training Procedure:
  Annealed Expectation Maximization (AEM) algorithm

  E-Step: Calculate Posterior Probabilities

$$P(z_o,z_u\,|\,o_{(l)},u_{(l)},r_{(l)}) = \frac{(P(Z_o)P(Z_u)P(o_{(l)}\,|\,Z_o)P(u_{(l)}\,|\,Z_u)P(r_{(l)}\,|\,Z_o,Z_u))^b}{\sum_{Z_o,Z_u}(P(Z_o)P(Z_u)P(o_{(l)}\,|\,Z_o)P(u_{(l)}\,|\,Z_u)P(r_{(l)}\,|\,Z_o,Z_u))^b}$$

Collaborative Filtering

$$P(Z_o);\ P(Z_u);\ P(o_{(l)} \mid Z_o);\ P(u_{(l)} \mid Z_u);\ P(r_{(l)} \mid Z_o, Z_u)$$

M-Step: Update Parameters

- Prediction Procedure:
  Fold-In process to calculate joint probabilities

$$P(o, u^t, r_{(l)}) = \sum_{Z_o, Z_u} P(Z_o)P(Z_u)P(o \mid Z_o)P(u^t \mid Z_u)P(r \mid Z_o, Z_u)$$

Fold-in process by EM algorithm

Calculate expectation for prediction

$$\hat{R}_{u^t}(o) = \sum_r r \frac{P(o, u^t, r)}{\sum_{r'} P(o, u^t, r')}$$

"**Flexible Mixture Model for Collaborative Filtering", ICML'03**

Collaborative Filtering

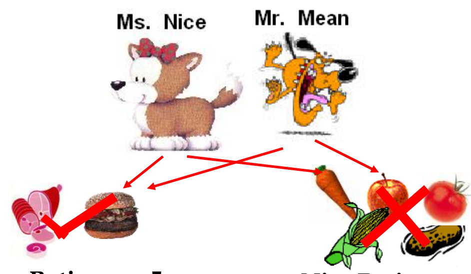
Thoughts:

- Previous algorithms address the problem that users with similar tastes may have different rating patterns implicitly (Normalize user rating)

PURDUE UNIVERSITY
Department of Computer Science

Previous Work: Thoughts

- Thoughts:



Nice Rating: 5
Mean Rating: 2

Nice Rating: 3
Mean Rating: 1

Explicitly decouple users preference values out of the rating values → **Decoupled Model (DM)**

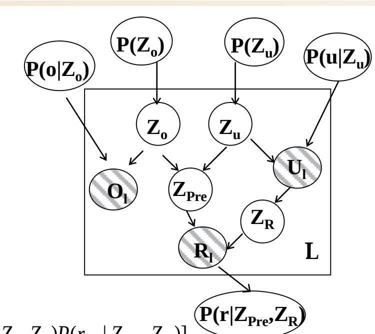
PURDUE UNIVERSITY
Department of Computer Science

Decoupled Model (DM)

- Decoupled Model (DM):
  Separate preference value

$Z_{pref} \in [1, \ldots, k]$ (1 disfavor, k favor)

from rating $r \in \{1,2,3,4,5\}$



Joint Probability:

$$P(o_{(l)}, u_{(l)}, r_{(l)}) = \sum_{Z_o, Z_u, Z_R} P(Z_o)P(Z_u)P(o_{(l)} \mid Z_o)P(u_{(l)} \mid Z_u)P(Z_R \mid u_{(l)})\left[\sum_{Z_{pre}} P(Z_{pre} \mid Z_u, Z_o)P(r_{(l)} \mid Z_{pre}, Z_R)\right]$$

"Preference-Based Graphical Model for Collaborative Filtering", UAI'03

"A study of Mixture Model for Collaborative Filtering", Journal of IR

Experimental Data

Datasets:

MovieRating and EachMovie

	MovieRating	EachMovie
Number of Users	500	2000
Number of Movies	1000	1682
Avg. # of rated items/User	87.7	129.6
Scale of ratings	1,2,3,4,5	1,2,3,4,5,6

Evaluation:

MAE: average absolute deviation of the predicted ratings to the actual ratings on items.

$$MAE = \frac{1}{L_{Test}} \sum_{l} | r_{(l)} - \hat{R}_{o_{(l)}}(u_{(l)}) |$$

Collaborative Filtering

Vary Number of Training Users

Test behaviors of algorithms with different amount of training data

- For MovieRating
  100 and 200 training users
- For EachMovie
  200 and 400 training users

Vary Amount of Given Information from the Test User

Test behaviors of algorithms with different amount of given information from test user

- For both testbeds

Vary among given 5, 10, or 20 items

Experimental Results
Improved by Combining FMM and DM

Training Users Size	Algorithms	5 Items Given	10 Items Given	20 Items Given
100	FMM	0.829	0.822	0.807
	FMM+DM	0.792	0772	0.741
200	FMM	0.800	0.787	0.768
	FMM+DM	0.770	0.750	0.728

Results on Movie Rating

Training Users Size	Algorithms	5 Items Given	10 Items Given	20 Items Given
200	FMM	1.07	1.04	1.02
	FMM+DM	1.06	1.01	0.99
400	FMM	1.05	1.03	1.01
	FMM+DM	1.04	1.00	0.97

Results on Each Movie

PURDUE UNIVERSITY

Department of Computer Science

Combine Collaborative Filtering and Content-Based Filtering

Content-Based Filtering (CBF): Recommend by analyzing the content information

Content information is very useful when few users have rated an object.

	Science Fiction?
A group of aliens visit earth……………………….. kind of friendship in which E.T learns…………	Yes ✓
Young Harry is in love and wants to marry an actress, much to the displeasure of his family….	No ✗

Unified Filtering (UF): Combining both the content-based information and the collaborative rating information for more accurate recommendation

PURDUE UNIVERSITY
Department of Computer Science

Content-Based Filtering and Unified Filtering

Content-Based Filtering (CF):

- Generative Methods (e.g. Naïve Bayes)
- Discriminative Methods (e.g. SVM, Logistic Regression)
  - **Usually more accurate**
  - **Can be used to combine features (e.g., actors for movies)**

Unified Filtering by combining CF and CBF:

- Linearly combine the scores from CF and CBF
- Personalized linear combination of the scores
- Bayesian combination with collaborative ensemble learning

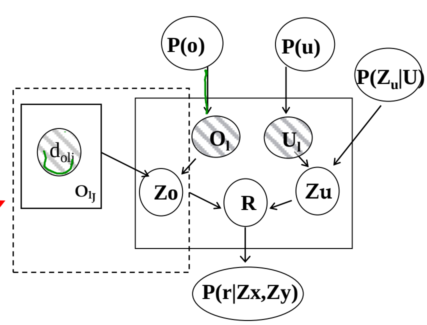
PURDUE UNIVERSITY
Department of Computer Science

Unified Filtering by flexible mixture model and exponential model

- Unified Filtering with mixture model and exponential model (UFME):

Mixture model for rating information:

$$P(o_{(l)},u_{(l)},r_{(l)}) = \sum_{Z_o,Z_u} P(Z_o|\overrightarrow{d_{ol}})P(Z_u|u)P(o_{(l)})P(u_{(l)})P(r_{(l)}|Z_o,Z_u)$$



Exponential model for content information

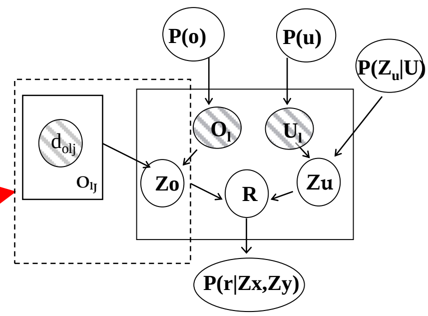
Unified Filtering by flexible mixture model and exponential model

- Unified Filtering with mixture model and exponential model (UFME):

Mixture model for rating information:

Exponential model for content information:

$$P_\theta(Z_o \mid \vec{d}_{ol}) = \frac{\exp(\sum_j \theta_{z_o j} * d_{olj})}{\sum_{Z_o'}(\exp(\sum_j \theta_{z_o j} * d_{olj}))}$$



Specific word

Unified Filtering by flexible mixture model and exponential model

- Training Procedure:

E-Step: Calculate posterior probabilities

Expectation Step of EM

M-Step: Update parameters

Second, refine the object cluster distribution with content information by maximizing

Iterative Scaling Training

"Unified Filtering by Combining Collaborative Filtering and Content-Based Filtering via Mixture Model and Exponential Model ", CIKM'04

Experiment Results

Table. MAE results for four filtering algorithms on EachMovie testbed. Four algorithms are pure content-based filtering (CBF), pure collaborative filtering (CF), unified filtering by combining mixture model and exponential model (UFME)

Training Users Size	Algorithms	0 Items Given	5 Items Given	10 Items Given	20 Items Given
50	CBF	1.43	1.21	1.24	1.19
	CF	1.21	1.14	1.13	1.12
	UFME	1.19	1.11	1.10	1.09
100	CBF	1.43	1.23	1.21	1.19
	CF	1.17	1.08	1.07	1.05
	UFME	1.17	1.08	1.06	1.05

Experiment Results

$P_\theta(Z_o \mid w)$

Table. Five most indicative words (with highest values) for 5 movie clusters, sorted by

Each column corresponds to a different movie cluster. All listed words are stemmed.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
forev	previou	mad	inhabit	custom
depress	passion	hang	dress	hang
mate	court	rape	relat	forev
broken	forget	finish	door	water
abandon	sea	arrest	younger	food

PURDUE UNIVERSITY
Department of Computer Science

Summary

What we talked about so far?

- Proposed the flexible mixture model
  – Demonstrates the power of clustering users and objects separately AND allowing them to belong to different classes
- Proposed the decoupled model
  – Demonstrates the power of extracting preference values from the surface rating values
- Proposed the unified probabilistic model for unified filtering
  – Demonstrates the power of taking advantage of content information with limited rating information