Cross Domain Collaborative Filtering for Recommendation Systems:

Background:

Mining of Massive Data Sets
Distance measures, clustering, recommendation systems, matrix factorization
Tensor Decompositions and Applications
Basic tensor notation, operations, and factorizations
Graph Mining: Laws, tools, and case studies
Ch 14: SVD random walks and Tensors
Ch 15: Tensors

Mining of Massive Data Sets: Coursera (Jan. 31, 2015)
https://class.coursera.org/mmds-002
Introduction to Recommender Systems: Coursera (video lectures)
https://class.coursera.org/recsys-001/lecture

Toolkits:

Comparative Recommender System Evaluation: Benchmarking Recommendation Frameworks (2014)

Compares common recommendation algortithms as implemented in three frameworks.

Evaluation dimensions: dataset, data splitting, evaluation strategies, and metrics

RiVal: recommender evaluation toolkit (open source):

http://rival.recommenders.net/

MyMediaLite: Recommender System Library (open source):

http://www.mymedialite.net/

Lenskit: open-source tools for recommender systems

http://lenskit.org/

Apache mahout: scalable machine learning library (free?):

http://mahout.apache.org/

Data

Recommendation and Ratings Public Data Sets For Machine Learning

https://gist.github.com/entaroadun/1653794

Recsys Challenge Data

Data available from past Recsys: ACM Conference Series on Recommender Systems

http://recsys.acm.org/

Netflix Challenge Dataset

Number of reviews: Over 100 million
Number of users: 480,000
Timespan Oct 1998 – Dec 2005

Amazon Review Data

Compiled by Jure Leskovec
Stanford Network Analysis Project (SNAP) Datasets
http://snap.stanford.edu/data/web-Amazon.html

Dataset statistics	
Number of reviews	34,686,770
Number of users	6,643,669
Number of products	2,441,053
Users with > 50 reviews	56,772
Median no. of words per review	82
Timespan	Jun 1995 - Mar 2013

Data format

```
product/productId: B00006HAXW
product/title: Rock Rhythm & Doo Wop: Greatest Early Rock
product/price: unknown
review/userId: A1RSDE90N6RSZF
review/profileName: Joseph M. Kotow
review/helpfulness: 9/9
review/score: 5.0
review/time: 1042502400
review/summary: Pittsburgh - Home of the OLDIES
review/text: I have all of the doo wop DVD's and this one is as good or better than the
1st ones. Remember once these performers are gone, we'll never get to see them again.
Rhino did an excellent job and if you like or love doo wop and Rock n Roll you'll LOVE
this DVD !!
```

Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text

- Leskovec, McAuley (2013)
- Researchers utilize the Amazon Review data set
- Uses review texts to understand rating behavior, model hidden rating dimensions

From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise through Online Reviews (Leskovec, McAuley 2013)

Each of our datasets were obtained from public sources on the web using a crawler, and are made available for others to use.

Data Exploration

Explore Amazon Data

Create a document with the following format:

UserID #Reviews #Domains DomainEntropy

Let,

$\mathbf{P}(\, x_i \,)$ = The probability a user $\boldsymbol{x}$ made a recommendation in domain $\boldsymbol{i}$.

Then the domain entropy is

$$H(X) = \sum_i P(x_i)\, I(x_i) = -\sum_i P(x_i) \log_b P(x_i)$$

Evaluation Metrics

Evaluating Recommender Systems (2009)

- Tech report form Microsoft Research.
- Properties of recommendation systems,
- Metrics and methods for online and offline evaluation.
- Suggests protocols for experimentation.
- Interesting idea: partitioning users in an online system to evaluate features

Rank and relevance in Novelty and Diversity Metrics for Recommender Systems (2011)

- Develops a general theory for evaluation metrics of novelty and diversity
- Diversity: How different items are with respect one another
- Novelty: How different items are with respect to what has previously been seen
- Discovery: an item is seen by (or is familiar to) a user
- Choice: an item is used, picked, selected, consumed, bought by a user
  Relevance: an item is liked, useful, enjoyed, etc. by a user

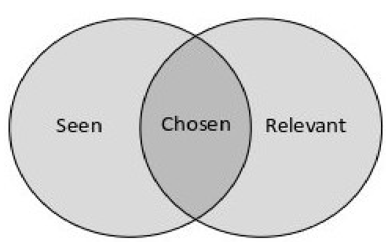

Figure 1. Discovery, choice and relevance models.

Cross Domain

Cross-domain recommendation systems: a survey of the state of the art

- Analysis and taxonomy of the cross-domain recommendation task
- “Hybrid approaches have barely been investigated”
- Advantage of cross domain recommendation may not be increased accuracy but added novelty and more diverse recommendations

Relations between domains		Recommendation models: *Adaptive*	Recommendation models: *Collective*
Content-based	*Attributes*	Azak 2010 [3]	
	Social tags	Kaminskas & Ricci 2011 [8]	Abel et al. 2011 [1] Szomszor et al. 2008 [18]
	Semantic properties	Fernández-Tobías et al. 2011 [6]	Loizou 2009 [12]
	Correlations		Shi et al. 2011 [17]
Collaborative filtering-based	*Ratings*	Azak 2010 [3] Berkovsky et al. 2008 [4] Winoto & Tang 2008 [20]	Loizou 2009 [12]
	Rating patterns	Li et al. 2009a [10]	Li et al. 2009b [11]
	Latent factors	Pan et al. 2010 [15]	Pan et al. 2011 [14]
	Correlations		Cremonesi et al. 2011 [5] Zhang et al. 2010 [21]

Personalized recommendation via cross-Domain triadic factorizations

Uses Tensor decomposition to generate recommendations (Amazon Review data)