

MOTIVATION

Category	Count	Average
Males	117,061	8.1
Females	22,165	7.9
Aged under 18	6,419	8.5
Females under 18	4,776	8.6
Males 18-29	1,576	8.2
Females 18-29	87,795	8.2
Males 30-34	30,346	7.9
Females 30-34	3,987	7.7
Aged 35+	6,095	7.6
Males Aged 35+	4,957	7.7
Females Aged 35+	1,272	7.9
Top 100 users	45	8.2
US users	22,848	7.5
Non-US users	96,401	8.0

- Examining reviews vs. Trusting overall aggregate rating
- IMDB ratings demographic breakdown not meaningful enough

PROBLEM

We propose a novel and powerful third option that explains ratings by leveraging user and item (movie) attribute information

- MRI: Meaningful Rating Interpretation**
- Example:

Black Swan (2010) - Drama, Thriller - 17 December 2010 (USA)

Young female reviewers love this movie, average rating: 9.3

Reviewers from New York love this movie, average rating: 8.7

Young male student reviewers hate this movie, average rating: 6.1

Woody Allen

Female reviewers love romantic movies by him, average rating: 8.9

Middle-aged male reviewers like movies by him, average rating: 8.4

Young male reviewers dislike drama movies by him, average rating: 7.1

SUB-PROBLEMS

SM : Similarity Mining

- Identify groups of reviewers who consistently share **similar** item ratings

Titanic (1997)

Teen-aged female reviewers have rated this movie uniformly. Their average rating: 9.2

DM : Diversity Mining

- Identify groups of reviewers who consistently **disagree** on item ratings

Schindler's List (1993)

Teen-aged female reviewers and male middle-aged reviewers have rated this movie inconsistently, their average rating: 7.5

Middle-aged male reviewers love this movie, their average rating: 9.1

Teen-aged female reviewers hate this movie, their average rating: 6.2

DATA MODEL & ALGORITHMS

- Collaborative rating site: <Set of Items, Set of Users, Ratings>
- Rating tuple: <item attributes, user attributes, rating>

Movie ID	Title	Genre	Director	User ID	Name	Gender	Location	Rating
1	Titanic	Drama	James Cameron	1	Amy	Female	New York	8.5
2	Schindler's List	Drama	Steven Spielberg	3	John	Male	New York	7.0

- Groups:** Set of ratings describable by a set of attribute values
- Notion of group based on data cube
- OLAP literature for mining multidimensional data

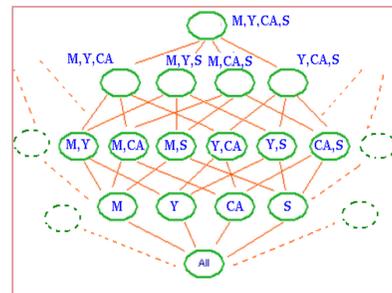
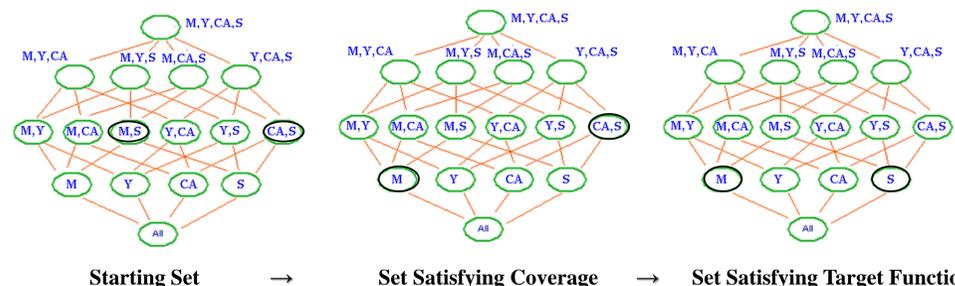


Figure: Partial Rating Lattice for a Movie
(M:Male, Y:Young, CA:California, S:Student)

Description Error	Measures how well a cuboid average rating approximates the numerical score of each individual rating belonging to it
Difference Balance	Measures whether the positive and negative ratings are "mingled together" (high balance) or "separated apart" (low balance)
Coverage	Measures the percentage of ratings (or, positive and negative ratings in case of DM) covered by the returned cuboids

Task: Quickly identify "good" groups (or cuboids) in the lattice that help users understand ratings effectively

- Both Similarity Mining and Diversity Mining problems are **NP-Hard**
- Exact Algorithm:** Brute-force enumerating all possible combinations of cuboids in lattice to return the exact (i.e., optimal) set as rating descriptions
- Random Restart Hill Climbing Algorithm:**
 - Often fails to satisfy Coverage constraint; Large number of restarts required
 - Need an algorithm that optimizes both Coverage and Target Function constraints simultaneously
- Randomized Hill Exploration Algorithm (RHE):** Two phase algorithm in which the hill around a randomly selected set of initial cuboids is explored to satisfy Coverage and then that result set is adopted as the starting point to optimize the Target Function



VISUALIZING RATING INTERPRETATIONS

- Perform Similarity and Diversity mining on ratings
- Constrain resulting groups to always include location
- Visualize mined results over a map
- Exploration over geographical location and time help in understanding rating patterns and evolution

RATING VISUALIZATION IN ACTION



CONCLUSIONS

- Novel problem of meaningful rating interpretation in collaborative rating sites
 - Similarity Mining
 - Diversity Mining
- Meaningful interpretations and visualization of ratings by reviewers of interest
- Exploration of rating patterns over location and time

REFERENCES

Data Cubes

- Mahashweta et. al, MRI: Meaningful Interpretations of Collaborative Ratings, VLDB 2011

Data Cubes

- Gray et. al, A relational aggregation operator generalizing group-by, cross-tab, and sub-totals, ICDE 1996
- Sathe et. al, Intelligent rollups in multidimensional olap data, VLDB 2001
- Lakshmanan et. al, Quotient cube: how to summarize the semantics of a data cube, VLDB 2002
- Ramakrishnan et. al, Exploratory mining in cube space, ICDM 2006
- Wu et. al, Promotion analysis in multi-dimensional space, VLDB 2009

Recommendation Explanation

- Herlocker et. al, Explaining collaborative filtering recommendations, CSCW 2000
- Bilgic et. al, Explaining recommendations: Satisfaction vs. promotion, IUI 2005