Personalized Interactive Faceted Search

Jonathan Koren*, Yi Zhang*, and Xue Liu†
*University of California, Santa Cruz
†McGill University

Baskin Engineering
UC SANTA CRUZ

Outline

- Introduce Faceted Search
- Identify Problems with Current FS Tech
- Propose a Solution
- Novel Evaluation Methodology
- Experiments
- Conclusions

Faceted Search is Everywhere

Formal Definition

- Interactive Structured Search Using Key-Value Metadata
- Parallel Hierarchies of Documents
- Point and Click Structured Query Generation

Problems

- Too Many Facets and Values
- Existing approach: Ad Hoc Value Presentation
- Proposed Solution: **Personalization** and **Collaborative** faceted search for interactive system **utility** optimization

Statistical Modeling Framework

- Document Model
- User Relevance Model

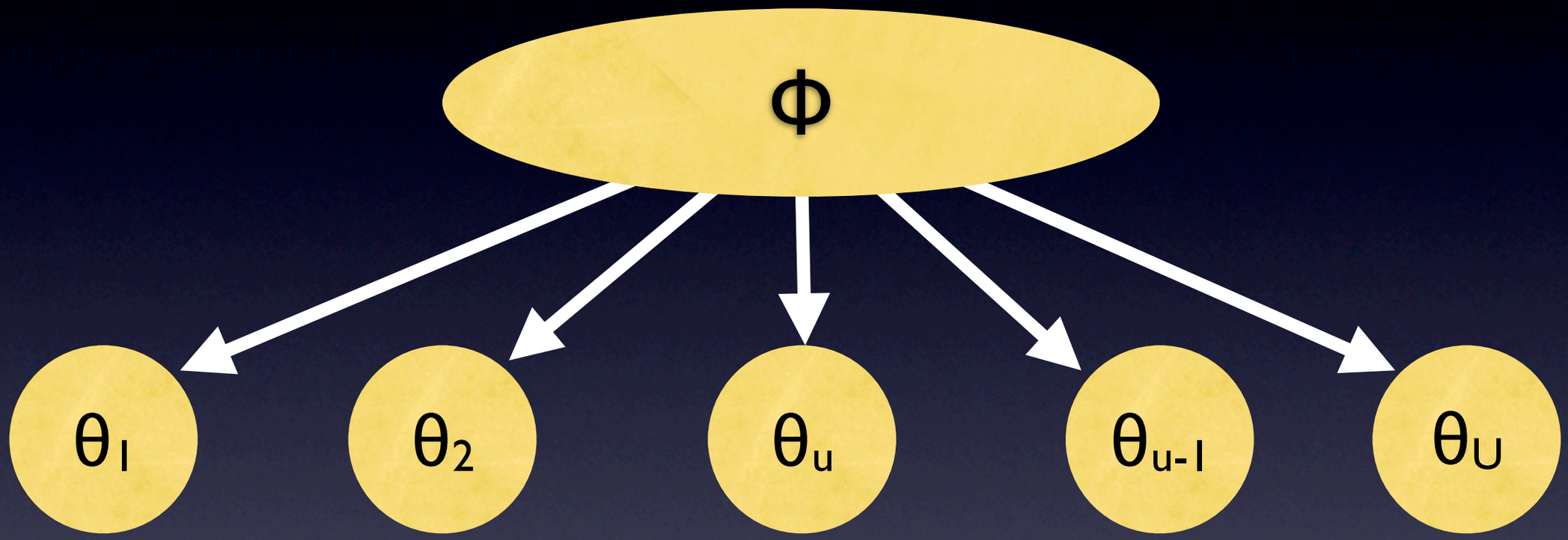
Document Model

- Docs are Unique Facet-Value Pairs
- Facets Come in Different Types
  - Facet-Type Suggests Statistical Model
- Docs Modeled as a Combination of Statistical Models

User Relevance Model

$$\theta_u = \{\mathrm{P}(rel \mid u), \mathrm{P}(x_k \mid rel, u), \mathrm{P}(x_k \mid non, u)\}$$

User Collaboration



- Φ is the Conjugate Prior to θ_u
- Φ Fills in Gaps in Individual User Models

Interface Evaluation

- User Studies are Expensive
- New Complementary Approach
  - Expected User Interface Utility
  - Simulated Interaction with Pseudousers

User Interface Utility

- Identify Types of Actions
- Assign Costs to Actions
- Reward for Relevant Docs Retrieved
- Calculate Utility for Entire Search Session

Expected User Interface Utility

$$E[\mathrm{U}] = \sum_{u \in \mathcal{U}} \sum_{D \in \mathcal{D}} E[\mathrm{U}(u, D)] \mathrm{P}(D \mid u) \mathrm{P}(u)$$

$$E[\mathrm{U}(u, D)] = \sum_{t=0} \sum_{a \in \mathcal{A}_t} \mathrm{R}(q_{t+1}, a, q_t) \mathrm{P}(q_{t+1} \mid a, q_t, u)$$
$$\mathrm{P}(a \mid q_t, u, D) \mathrm{P}(q_t \mid q_{t-1}, u, D)$$

Assumptions

1. Users Need to Satisfy a Need with a Set of Documents
2. Users Can Recognize Relevant Documents and Facet-Value Pairs
3. Users Continue to Perform Actions Until Their Need is Met

Pseudousers

- Stochastic Users
- First-Match Users
- Myopic Users
- Optimal Users

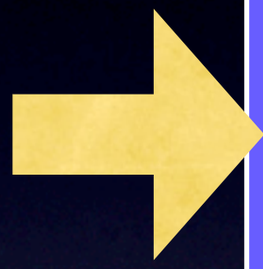
Stochastic Users

- Picks Relevant FVP at Random

A Nonrelevant	(14 matches)
B Relevant	(17 matches)
C Relevant	(11 matches)
D Nonrelevant	(12 matches)
E Nonrelevant	(12 matches)
F Relevant	(15 matches)
G Relevant	(13 matches)
H Nonelevant	(4 matches)
I Relevant	(13 matches)
J Nonrelevant	(16 matches)

First-Match Users

- Scans list for Relevant FVPs from Top to Bottom, Picking the First



A Nonrelevant	(14 matches)
B Relevant	(17 matches)
C Relevant	(11 matches)
D Nonrelevant	(12 matches)
E Nonrelevant	(12 matches)
F Relevant	(15 matches)
G Relevant	(13 matches)
H Nonelevant	(4 matches)
I Relevant	(13 matches)
J Nonrelevant	(16 matches)

Myopic Users

- Picks Relevant FVP that is Contained in the Least Number of Documents

A Nonrelevant	(14 matches)
B Relevant	(17 matches)
C Relevant	(11 matches)
D Nonrelevant	(12 matches)
E Nonrelevant	(12 matches)
F Relevant	(15 matches)
G Relevant	(13 matches)
H Nonelevant	(4 matches)
I Relevant	(13 matches)
J Nonrelevant	(16 matches)

Optimal Users

- Examines the Complete Interface
- Executes the Action that Maximizes the Utility

A Nonrelevant	(14 matches)
B Relevant	(17 matches)
C Relevant	(11 matches)
D Nonrelevant	(12 matches)
E Nonrelevant	(12 matches)
F Relevant	(15 matches)
G Relevant	(13 matches)
H Nonelevant	(4 matches)
I Relevant	(13 matches)
J Nonrelevant	(16 matches)

Evaluation Review

- Each Pseudouser Logs into the Search Interface
- Pseudouser Interacts with Interface to Retrieve a Set of Documents.
- Interface Receives a Score for the Session.
- Expected Utility = Average Score for all Sessions

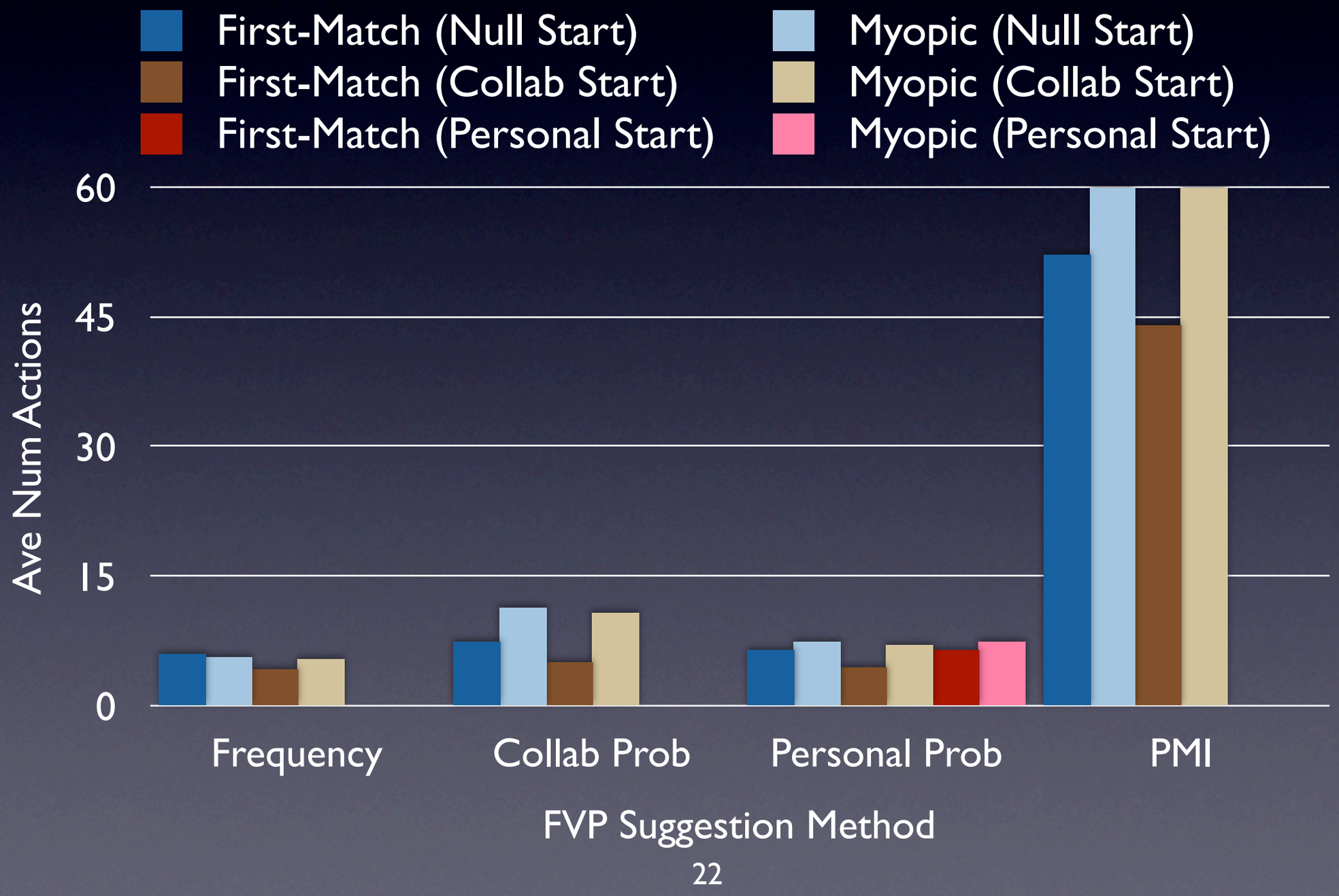
Personalization Experiments

- Facet-Value Pair Suggestion
  - Most Frequent
  - Most Probable (Collaborative)
  - Most Probable (Personalized)
  - Mutual Information
- Start Page Personalization
  - Empty Page
  - Collaborative Page
  - Personalized page

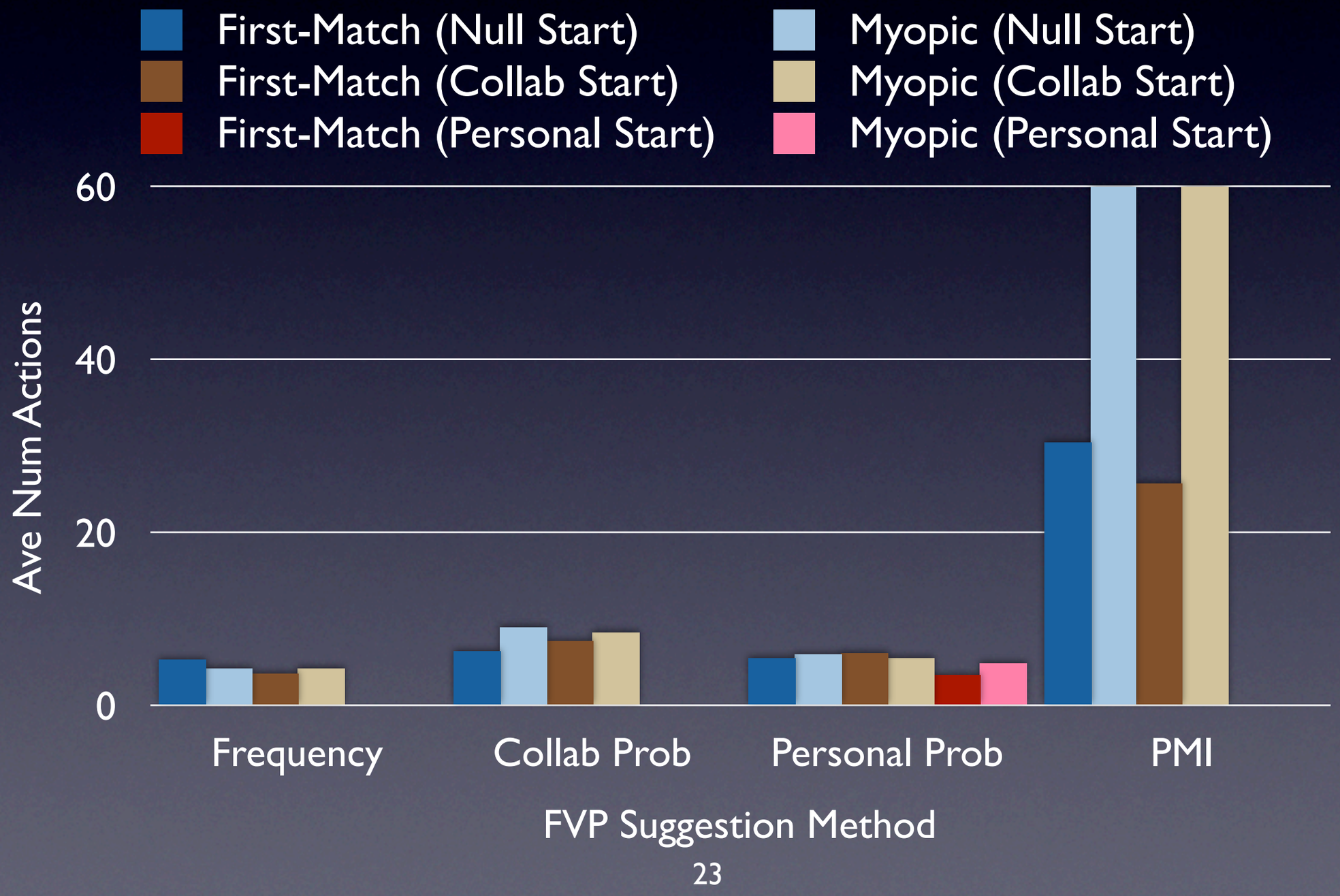
Document Corpora

- 8000 Documents from IMDB
  - 19 Facets and 367k Facet-Value Pairs
- 5000 Users Each from Netflix and MovieLens
  - 633k Ratings for Netflix
  - 742k Ratings for Movielens

Results (Netflix)



Results (MovieLens)



Conclusions

- Many Facets and Values are a Problem
  - Personalized Interfaces Can Help
- Proposed Statistical Modeling Framework for Faceted-Search
- Proposed Inexpensive Repeatable Evaluation Technique for Faceted-Search Interfaces
- Personalized Start Pages are Helpful

fin

Example: Two Myopic Users Search for “The ‘Burbs”

User: 302

certificate=PG
soundmix=Dolby
genre=Comedy

productiondesigner=SpencerJamesH

User: 1329

certificate=PG
soundmix=Dolby
genre=Comedy
country=USA
language=English
colorinfo=Color
year=1989
productiondesigner=SpencerJamesH