Interactive View Recommendation

Xiaozhong Zhang
Department of Computer Science, University of Pittsburgh
xiaozhong@cs.pitt.edu

Background and Motivation

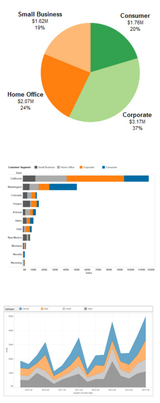
Visual data analysis tools use **utility functions** (UFs) to rank and recommend interesting views.

tableau

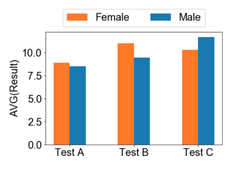
looker

Qlik

Utility Functions



A UF combines different **utility measures** (UMs) to derive a score representing **view interestingness**.



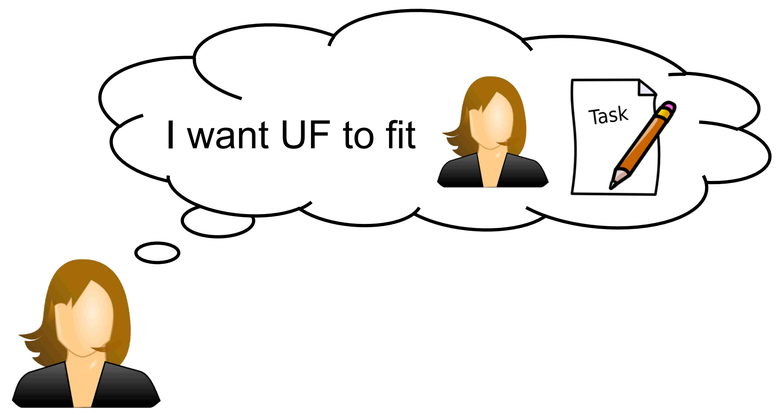
calculate

combine

0.7

$Relevance\ (R)$
$Conciseness\ (C)$
$Deviation\ (D)$
$\vdots$

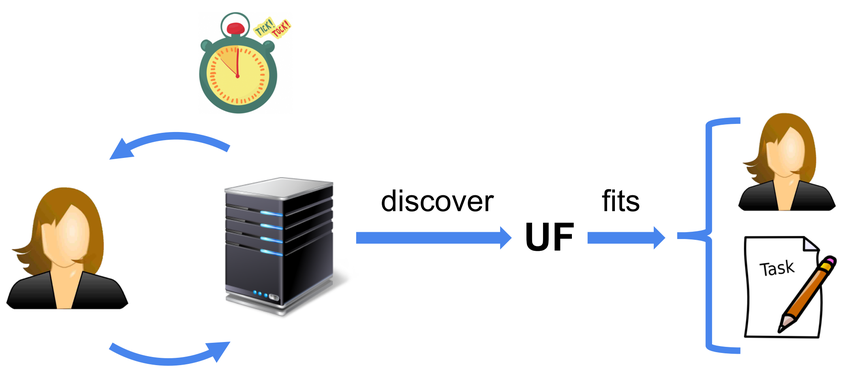
The UF in traditional view recommendation is **defined a priori**, so it **cannot adapt to** the analysis context .



ViewSeeker: An Interactive View Recommendation Tool

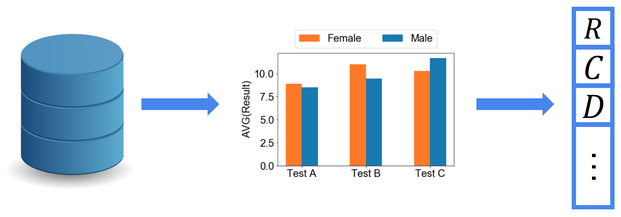
Interactive View Recommendation (IVR)

- IVR interacts with the user to **discover** the UF **most suitable** to the analysis context.
- IVR needs to have **interactive** response time.

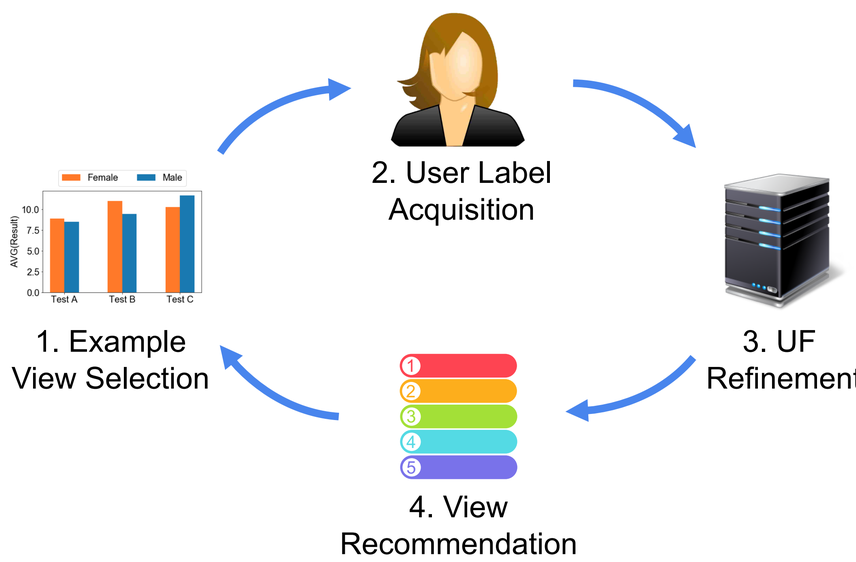


ViewSeeker
Phase 1: View Generation

It creates the views from the data and uses the **UM vector** of each view as its representation.

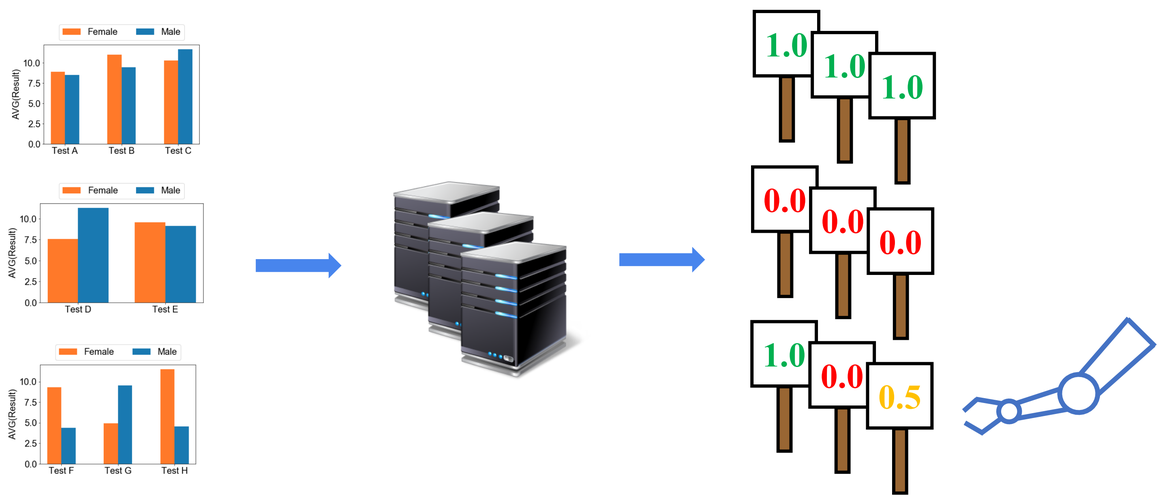


ViewSeeker
Phase 2: User Interaction



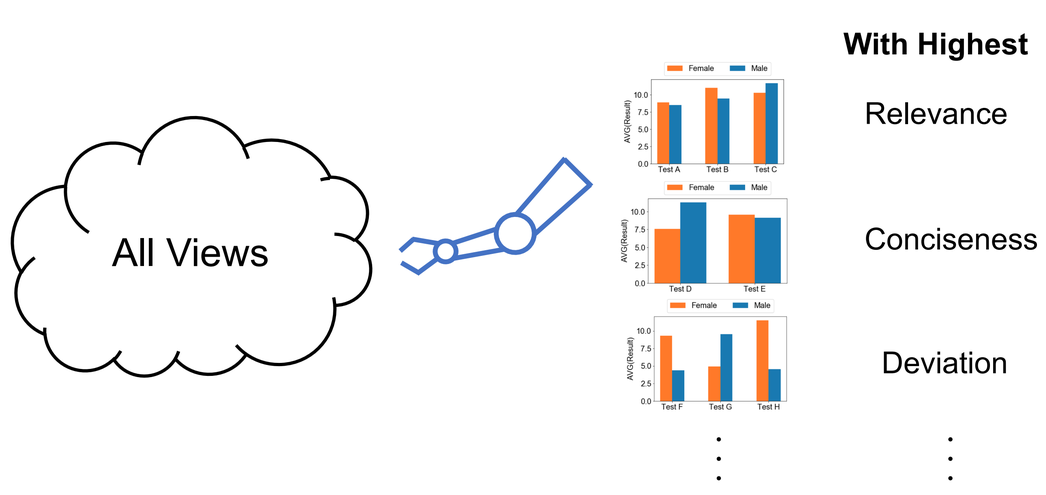
Example Selection Strategy: Query-By-Committee (QBC)

- User labeling effort is very expansive in IVR.
- **ViewSeeker** uses QBC to reduce user effort and achieve fast **model parameter convergence**.
- QBC selects the example on which the learner committee has the **largest disagreement**.



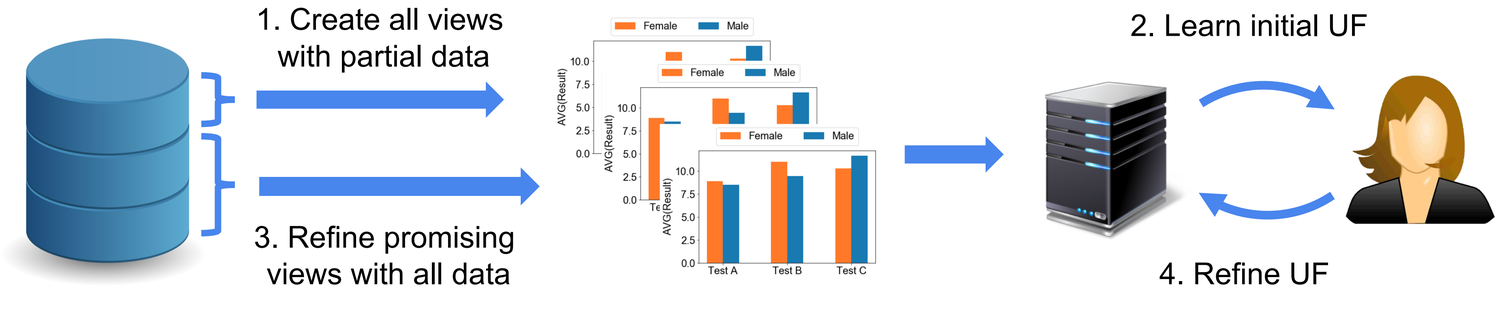
Optimization: Initial Example Selection

- In Phase 2, **ViewSeeker** first selects views with the highest score for each UM.
- These views are more likely to provide useful information about the UF in the **high view interestingness range**, in which the user has more interest.



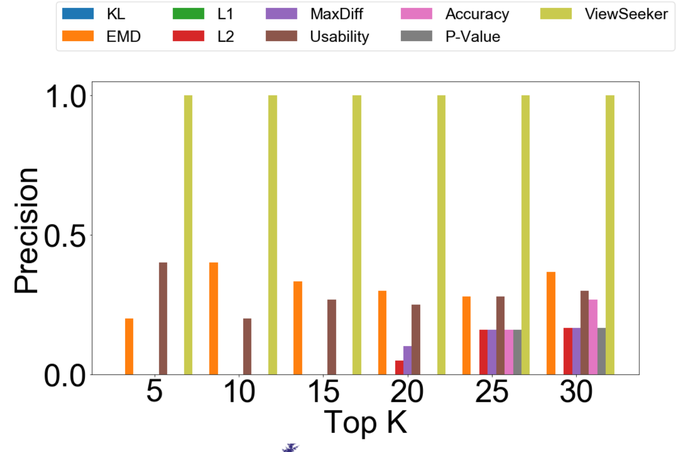
Optimization: 2-Stage View Generation

- To achieve the interactive time limit tl in Phase 1, **ViewSeeker** uses α percent of the data to generate all the views, with $\alpha = \frac{tl \cdot s}{V}$, where s is the view generation speed and V is the total view count.
- To achieve tl in Phase 2, in each interaction iteration, **ViewSeeker** selects n views with the highest interestingness based on the UF learned from user feedback for view refinement using all data, with $n = tl \cdot s$.



Experimental Results

ViewSeeker outperforms the best baseline with defined-a-priori UF by **3X** in recommendation precision, and achieves 100% precision with only **9 - 16** labeled views.



Advanced Data Management Technologies Laboratory, University of Pittsburgh

This work was partially supported by NIH award U01HL137159 and reflects the authors opinions

ACM SIGMOD/PODS International Conference on Management of Data
June 14-19th, 2020
PORTLAND
ACM SIGMOD +PODS 2020
Portland, Oregon USA