**User Profiles**

* What is that
* Example
* Generic
* How to lower the dimension of properties
  + Taxonomies
* Property domain

**Notion of Diversity**

* Examples of distance vs coverage based diversity
  + Using our example
* Opinions 🡪 coverage
  + We want to be as similar as can be to the source population
  + Diverse opinions exist in the source population
  + YAEL TODO: Data Ethics
* What is a good coverage
  + Coverage is ideal when each population group is proportionally represented
  + TODO: stratified sampling, like when conducting a survey, non-overlapping
* Generally Impossible
  + High dimensionality
    - Long tail – small groups, small selection
    - Over-under represented
    - Exmaples
  + Range – what does it mean to cover a range?
* What is a good compromise?
  + There is no generic solution
    - Interesting properties differ based on the application
    - Explanations and customizations

**Coverage-Based Diversity**

* Buckets
  + Intervals over score ranges
* Groups
  + A combination of union,…. Over buckets
  + Why not use all kinds of groups?
    - Hard to explain
    - We will show that we are able to cover overlapping groups
      * experiments
    - Easy to do when user specified
* Covered Group
  + What is a coverage?
    - Examples
      * One user
      * Importance
        + Proportional
      * Statsitical distribution
  + Weigths
    - Number of users
    - Equal
    - User specified \ factor
    - Expected number of users
      * In domain of a query
      * Given relevance score
    - Baseline – number of users
      * Largeset groups are most likely to be included
      * Long tail
  + Aggregation Methods
    - Sum
    - Simple monotonic
  + Max-Diversity formal definition
  + Example of max diversity
* Preprocessing
  + Bucket thresholds
    - How?
      * Naïve algorithm
    - One dimensional clustering algorithm
      * Blackbox
      * Supports incremental calucalation 🡪 our algorithm supports it as well
      * Buckets inside same property are dependent but this does not apply to buckets from different properties
    - Small number of clusers
  + Weighting algorithm
* Greedy Algorithm
  + Sub-modular
* Measures
  + Aggregated weight of selection
  + Our distribution similarity
    - Why:
      * Under-rep penalty
      * Proportional to group size
      * We will see why not to use it to choose our users
        + Not sub modular – no optimization
  + Prediction – another quality measure
    - Post mortume
    - Diversity
      * Variance
      * Dist similarity
    - We can't use it to choose users – not available to us

**Customizations and Explanations**

* No generic solution
* Motivating example and novelty
* Explanations 🡪 Customizations
* Expl.
  + Why user?
  + Why subset?
  + Why Group?
    - Explainable Buckets – artificial, our responsibility to make interpreted
      * High, med low
      * True\False
  + Does not exist in other works
  + Its is understandable to the user
* Visualizations (?)
  + A must have?
* Custimazations
  + Types
  + Because we use groups as explanations it is easily customizable
  + From custimzation into explanation and vice versa
  + Visual customizations
  + Observation: custimzations still gives submodular complexity
* Query-relevance as a customization
* Customize by distribution given by the user
  + Change definition of covered group

**Implementation**

* Datasets
  + From user data to properties with scores
  + Example – avg score for cuisine
  + Yelp / Netflix
    - Scale
    - How retrieved
    - Number of users
    - Number of properties
* System architecture
  + Explanation, visualization
* Technical Overviews

**Experiments**

* Alternative Algorithms
  + What, why
* Experiment types
  + Setup
  + Goal
  + Expectations
  + Results
  + Meaning
  + Insights