AnchorViz

Facilitating Classifier Error Discovery through Interactive Semantic Data Exploration

Nan-Chen Chen (University of Washington)
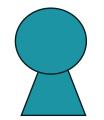
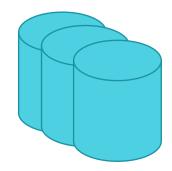Jina Suh, Johan Verwey, Gonzalo Ramos, Steven Drucker, Patrice Simard (Microsoft Research)

Interactive Machine Learning (iML)
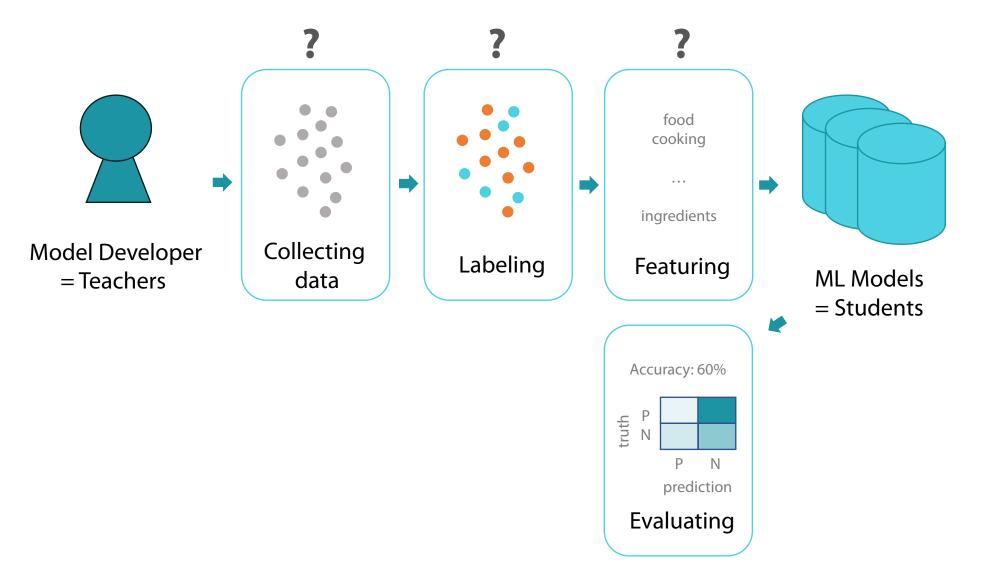
Student-Teacher Metaphor for ML

Model Developer
= Teachers

ML Models
= Students

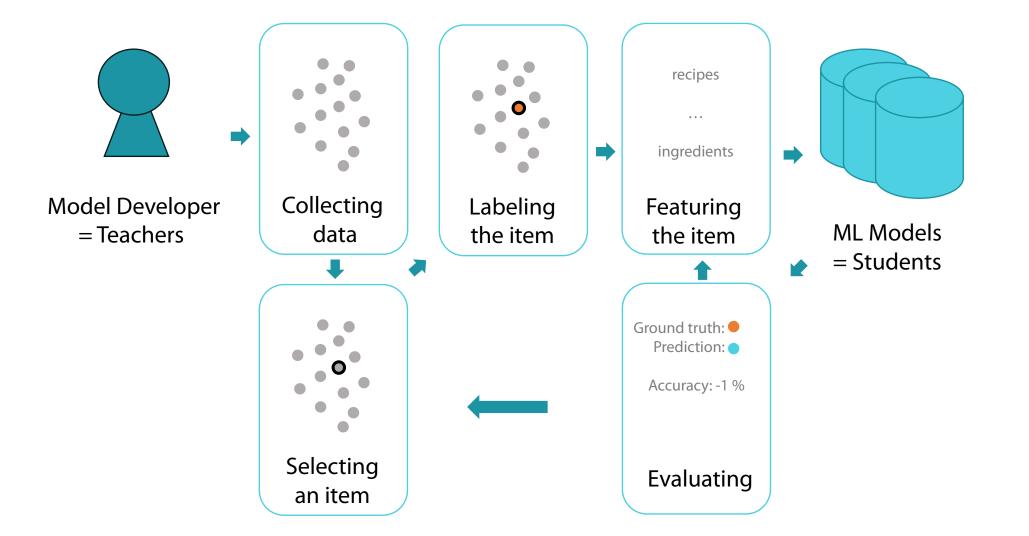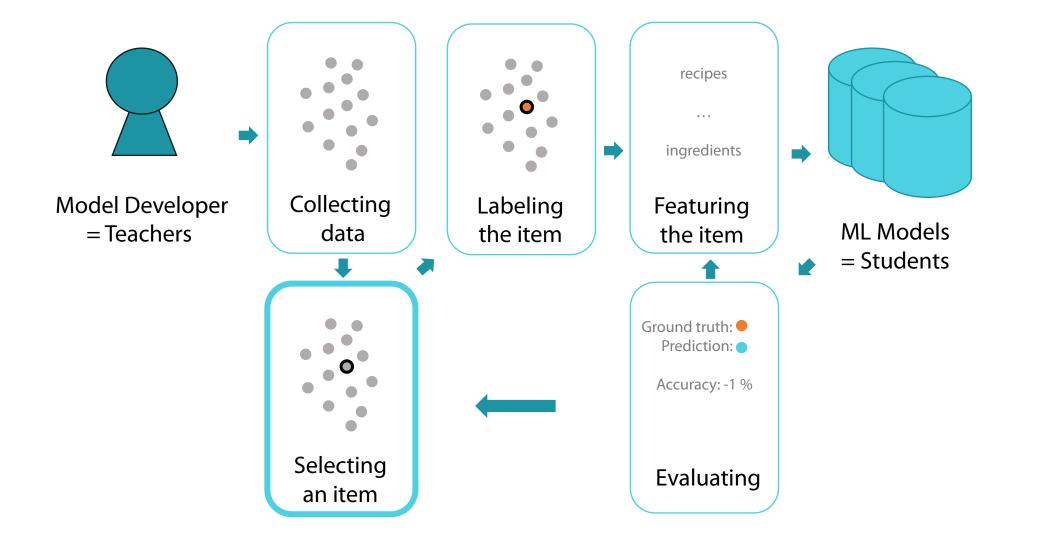# Traditional ML Paradigm

iML Paradigm

Model Developer
= Teachers

Collecting
data

Labeling
the item

recipes

…

ingredients

Featuring
the item

ML Models
= Students

Selecting
an item

Ground truth: ●
Prediction: ●

Accuracy: -1 %

Evaluating
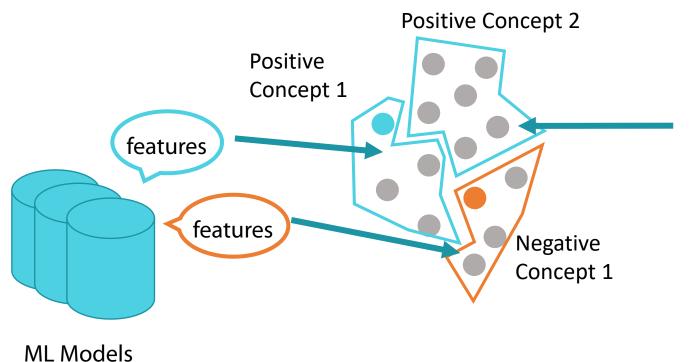
Key Challenge: Selecting Items for iML

Model Developer
= Teachers

Collecting
data

Labeling
the item

recipes

…

ingredients

Featuring
the item

ML Models
= Students

Selecting
an item

Ground truth: ●
Prediction: ●

Accuracy: -1 %

Evaluating

Islands of concepts for the target class

Positive Concept 2

Positive Concept 1

Feature blindness error: When there is no feature representing a concept

features

features

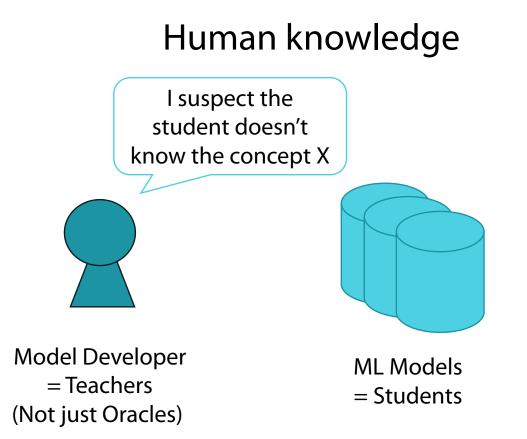Negative Concept 1

ML Models = Students

Two Typical Ways to Select Items

Active learning methods

- Uncertainty Sampling

- Uniform Sampling

- Stratified Sampling

Random or ambiguous items

Human knowledge

I suspect the student doesn't know the concept X
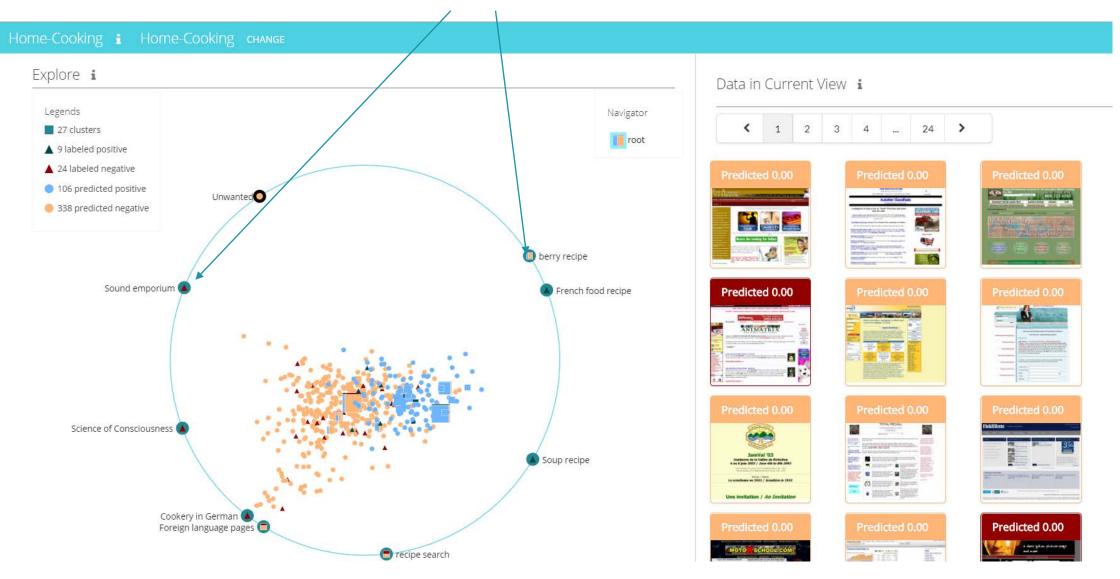
Model Developer
= Teachers
(Not just Oracles)
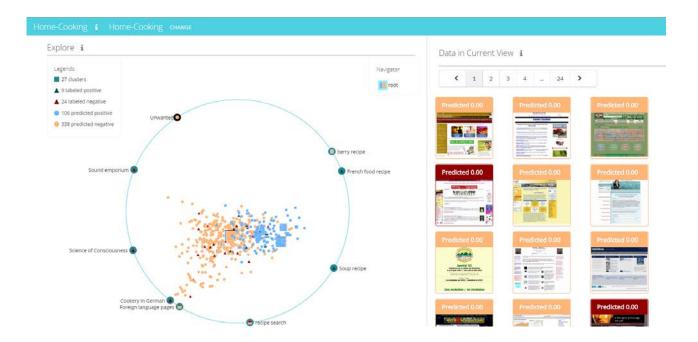
ML Models
= Students

Can't enumerate all concepts!

AnchorViz

anchor: the representation of a concept

Outlines

1. Motivating example & AnchorViz design

2. Evaluation & Analysis

3. Discussion & Future work

Outlines
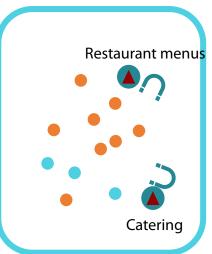
Motivating Example:
anchor & nearby item discrepancy
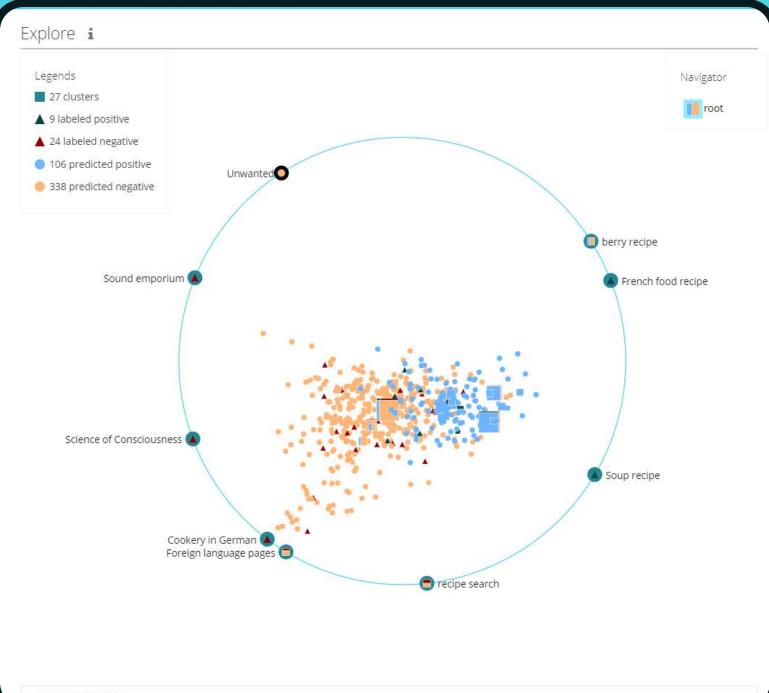
Design Objectives

1. Let users define concepts of the target class and unrelated classes

2. Spread the dataset based on concepts

3. Show how user-defined concepts impact the positions of items

4. Provide information about the model's current prediction along the user's labels

5. Optimize for efficient reviewing process

Explore ℹ

Legends

- 🟩 27 clusters
- ▲ 9 labeled positive
- 🔺 24 labeled negative
- 🔵 106 predicted positive
- 🟠 338 predicted negative

Navigator

🟧 root

Unwanted ◉

🟧 berry recipe

🔺 Sound emporium

🟢 French food recipe

🔺 Science of Consciousness

🟢 Soup recipe

Cookery in German
Foreign language pages 🔺 🟧

🟧 recipe search

Data in Current View ℹ

‹ **1** 2 3 4 … 24 ›

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Explore ℹ

Items for exploration

Legends
- ■ 27 clusters
- ▲ 9 labeled positive
- ▲ 24 labeled negative
- ● 106 predicted positive
- ● 338 predicted negative

Navigator
- ▮ root

- Unwanted
- berry recipe
- French food recipe
- Sound emporium
- Soup recipe
- Science of Consciousness
- Cookery in German Foreign language pages
- recipe search

Data in Current View ℹ

< | 1 | 2 | 3 | 4 | ... | 24 | >

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Predicted 0.00

Legends

- 🟦 27 clusters
- 🔺 (dark green) 9 labeled positive
- 🔺 (dark red) 24 labeled negative
- 🔵 106 predicted positive
- 🟠 338 predicted negative

...ms for exploration

Navigator

🔲 root

berry recipe

French food recipe

Science of Consciousness

Cookery in German
Foreign language pages

recipe search

Pre-grouping using
hierarchical clustering

The position is the sum of all anchors' effects

$$\vec{V}_k = (r \cdot cos\theta, r \cdot sin\theta)$$

$$Position(i) = (x_i, y_i) = \sum_k \frac{value_k(i)}{\sum_j value_j(i)} \cdot \vec{V}_k$$

Visualization Layout

Explore i

Legends
- 27 clusters
- 9 labeled positive
- 24 labeled negative
- 105 predicted positive
- 338 predicted negative

Navigator

root

1000:3825

Unwanted

berry recipe

French food recipe

Sound emporium

Science of Consciousness

Soup recipe

Cookery in German
Foreign language pages

recipe search

Data in Current View i

ADD THIS ITEM AS AN ANCHOR CL

| Predicted 0.59 | Positive | Unsure | Negative |

BUILD A FREE WEBSITE
OF YOUR OWN ON
Angelfire

BJ's Restaurants Healthy
Discover our delicious Enlightened Entrees - all under 575 calorie
www.bjsrestaurants.com

Check the Current Florida Lottery (For those that might not want to deliver mail next week)   Cape
Delivery Confirmation Check   Postal Rate Chart   Stamps On-line   Links   +Bugman's Posta

SW Florida Postal Wo

Check out these favorite Recipes of SW
Click on to one of these Creative Alternatives the next time you a

Main Dishes
- **Coquina Chowder** Out of IONA -Ft Myers
- **Mediterranean Calzones** Out of IONA -Ft Myers
- **The SECRET Teriyaki Sauce Recipe** (From Japanese Restaura
- **Mexican Surprise** Out of IONA -Ft Myers

Beverages
- **ORANGE WAKE-UP DRINK** Out of IONA -Ft Myers

Details view

Explore ℹ

Legends

■ 20 clusters
▲ 39 labeled positive
▲ 2 labeled negative
● 394 predicted positive
● 56 predicted negative

Use Navigator to go back →

Navigator

■ root
■ 7759

Unwanted

berry recipe

Sound emporium ▲

▲ French food recipe

Science of Consciousness ▲

▲ Soup recipe

Cookery in German
Foreign language pages

recipe search

Data in Current View ℹ

ADD THIS WHOLE LIST AS AN ANC

< 1 2 3 4 ... 25 >

Predicted 0.00

Predicted 0.01

Predicted 0.01

Predicted 0.01

Predicted 0.04

Predicted 0.06

Predicted 0.08

Predicted 0.11

Predicted 0.12

Predicted 0.13

Predicted 0.13

Predicted 0.14

Navigate to a cluster

Explore ℹ

Legends
- 🟩 27 clusters
- 🔺 9 labeled positive
- 🔺 24 labeled negative
- 🔵 105 predicted positive
- 🟠 338 predicted negative

Navigator

🟧 root

Unwanted

P 364884

berry recipe

French food recipe

Sound emporium

P 92657

Science of Consciousness

Soup recipe

Cookery in German
Foreign language pages

recipe search

Anchor Details ℹ

Remove   CLOSE

Name   P 364884

Example items in the anchor:

✕ Predicted 0.98

RECIPES

Inn RECIPES

OUTSTANDING RECIPES FOR
...TODAY

AARDVARK &
ASSOCIATES

Breakfast Re

A L'Orange Blossom Croissants
From: Austin's Wildflower Inn

Almond French Toast
From: Coconut Grove Bed & Breakfast

Almond French Toast
From: Swan House South

Almond Skillet Coffee Cake
From: Hoyt House Bed & Breakfast

Apples & Spice Bake Toast
From: Village Green Inn

Baked Eggs with Brie Cheese and Black F
From: Fleur De Lys Bed & Breakfast

breakfast
lunch
afternoon tea
hors d'oeuvres
dinner
dessert
beverage
holiday

comments

Add an anchor

Explore ℹ

Legends

■ 27 clusters
▲ 9 labeled positive
▲ 24 labeled negative
● 105 predicted positive
● 338 pred

Navigator

Change the anchor name →

Move the anchor around →

Unwanted

P 364884

berry recipe

French food recipe

Sound emporium ▲

Science of Consciousness ▲

Soup recipe

Cookery in German
Foreign language pages

recipe search

Anchor Details ℹ

Remove CLOSE

Name P 364884

Example items in the anchor:

✕
Predicted 0.98

RECIPES

Inn
RECIPES

OUTSTANDING RECIPES FOR
...TODAY

AARDVARK &
ASSOCIATES

Breakfast R

A L'Orange Blossom Croissants
From: Austin's Wildflower Inn

Almond French Toast
From: Coconut Grove Bed & Breakfast

Almond French Toast
From: Swan House South

Almond Skillet Coffee Cake
From: Hoyt House Bed & Breakfast

breakfast
lunch
afternoon tea
hors d'oeurvres
dinner
dessert
beverage
holiday

Apples & Spice Bake Toast
From: Village Green Inn

Baked Eggs with Brie Cheese and Black F
From: Fleur De Lys Bed & Breakfast

comments

Edit an anchor

Outlines

Instructions

(1) find items where the classifier is making a mistake

(2) find a set of items that are diverse from each other

(3) try to understand the dataset and classifier performance in the process

100% accuracy on the training set but ~75% on the test set

Build the model using an iML tool

16 Participants
7 Females & 9 Males
Distributing nearly equally across
4 levels of ML experiences

A binary classifier
for cooking webpages
(12 existing features)

Dataset

4,000 webpages
400 labeled
(50% positive)
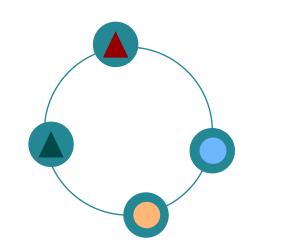
Study Procedure
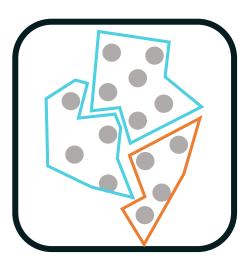
Introduction
Videoconferencing
Starting audio & screen recording
ML concepts review

Training session
A binary classifier
for travel webpages
to introduce AnchorViz
and practice

Real study session
A binary classifier
for cooking webpages

Analysis

Anchor Effectiveness

Error Analysis
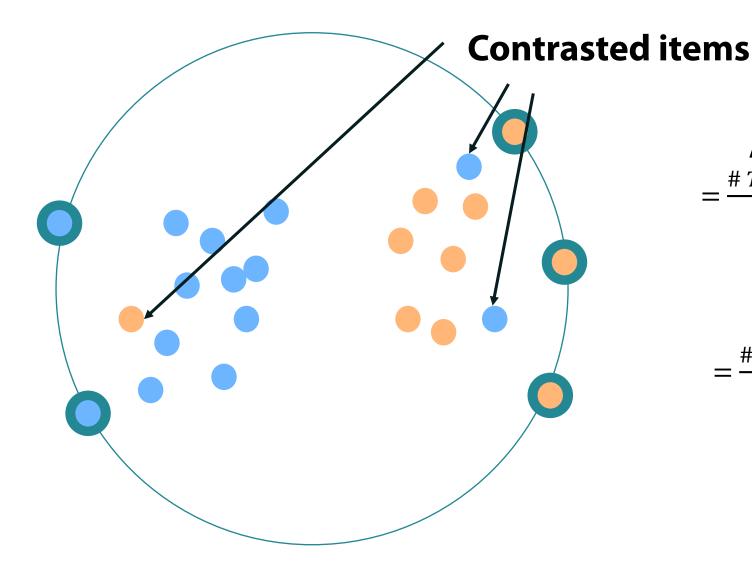
Qualitative coding
on user behaviors

Anchor Effectiveness

Contrasted items

Anchor Error Precision (AEP)

$$= \frac{\textit{\# True Errors in the Contrasted Items}}{\textit{\# Items in the Contrasted Items}}$$

Anchor Error Recall (AER)

$$= \frac{\textit{\# True Errors in the Contrasted Items}}{\textit{\# True Errors}}$$

Anchor Effectiveness Results
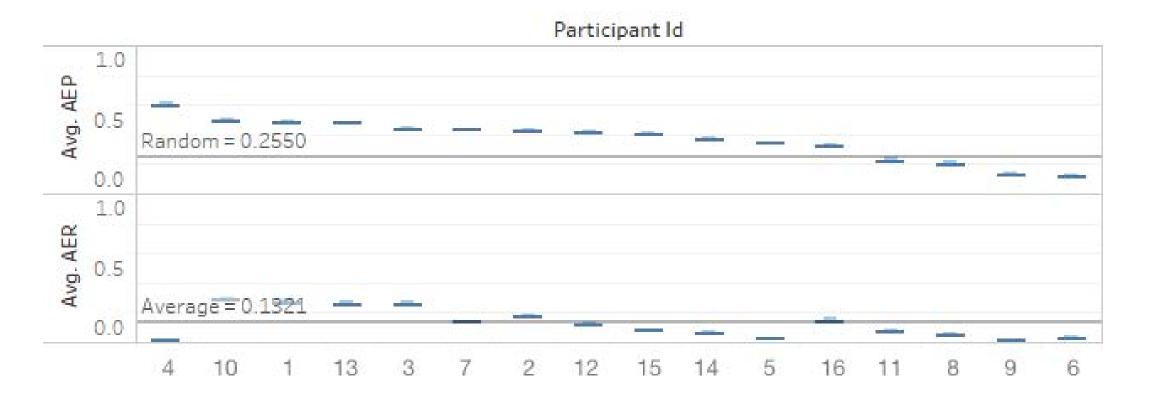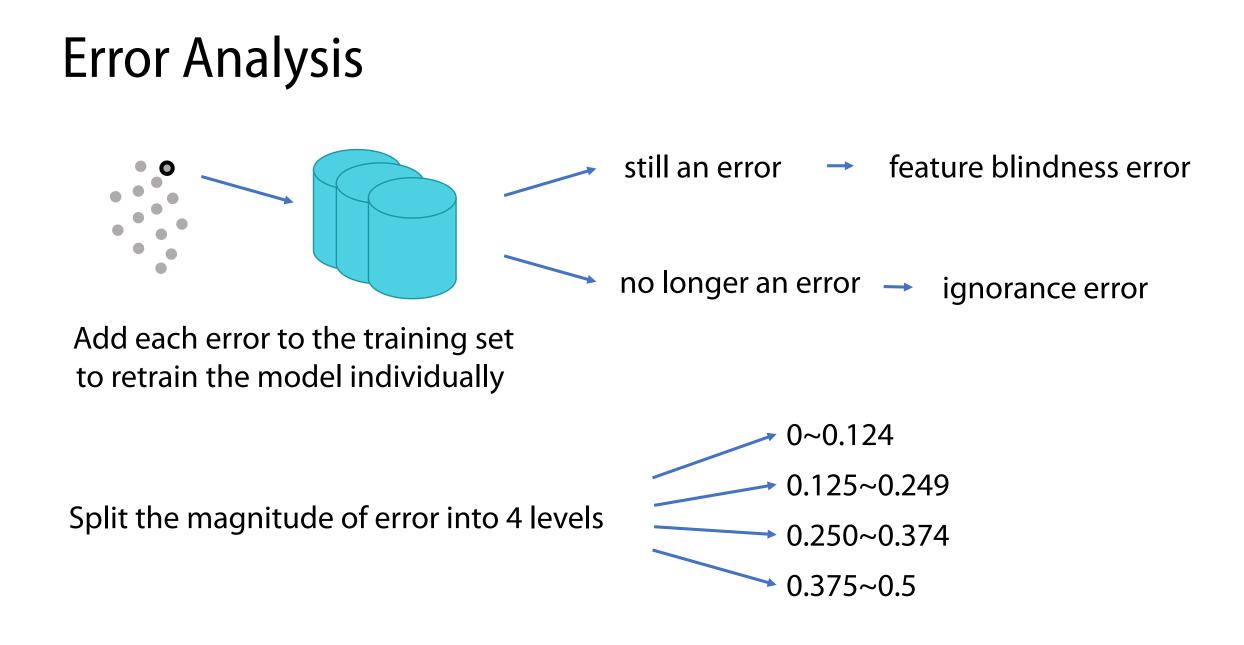
Figure 5. Average anchor effectiveness metrics (AEP/AER) for each participant. Most participants have better AEP than random.

Error Analysis

still an error $\rightarrow$ feature blindness error

no longer an error $\rightarrow$ ignorance error

Add each error to the training set
to retrain the model individually

Split the magnitude of error into 4 levels

0~0.124

0.125~0.249

0.250~0.374

0.375~0.5

Error Analysis Results
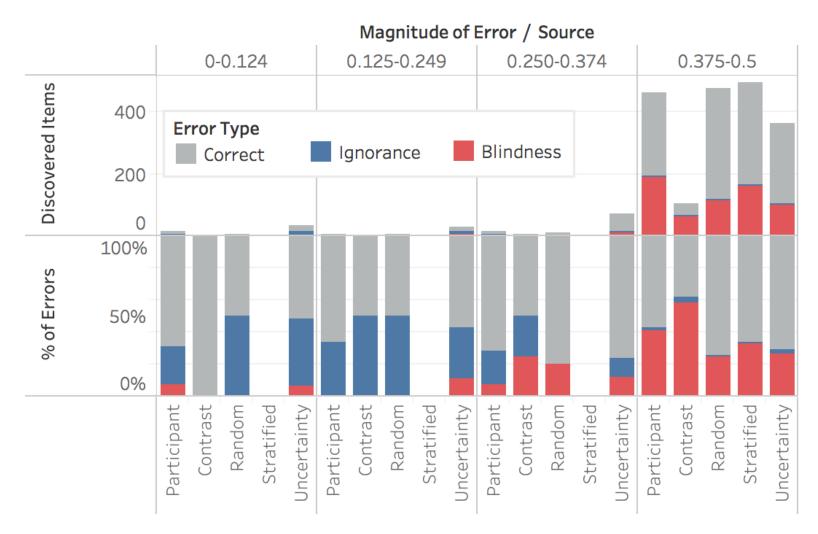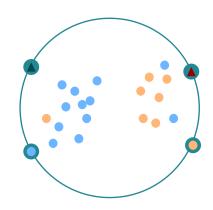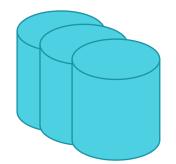
Figure 3. Distribution of discovered items (top) and the magnitude of errors (bottom) across participants and algorithmic samplers. The items are skewed towards high magnitude errors, but items discovered by participants and contrasted items higher chance of getting high magnitude items than algorithmic samplers.

Qualitative Coding on User Behaviors

Reasons for creating anchors

Exploration strategies
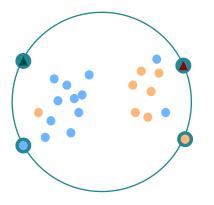
Insights about the classifier

Qualitative Coding Results

Reasons for creating anchors

- Define positive & negative concepts (recipe vs. science)

- Capture potential issues (foreign language pages)

- Validate hypotheses (suspecting and verifying the issues of index pages)
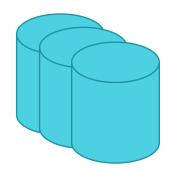
Qualitative Coding Results

Exploration strategies

- Leverage visual encoding (discrepancy in colors and positions)

- Play with the placement of anchors (changing topology to find patterns)

- Refine anchors to make them more effective (change an "cucumber" anchor to "vegetables")
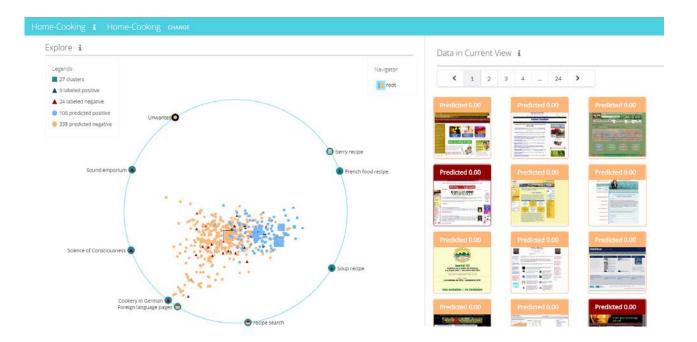
Qualitative Coding Results

Insights about
the classifier

- Recall problem, not precision problem

- Clusters of errors

- Multiple meanings of a word can be a problem

Outlines

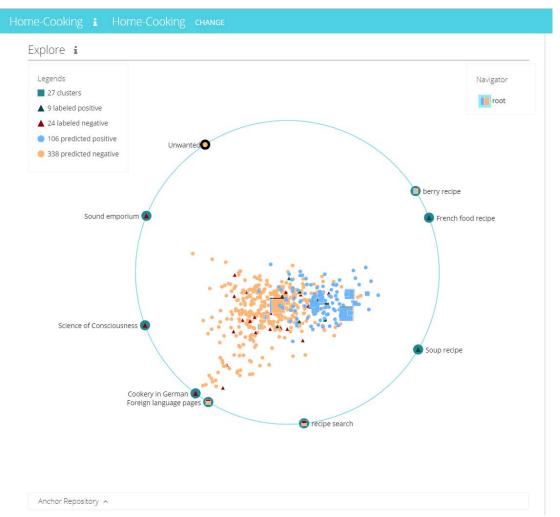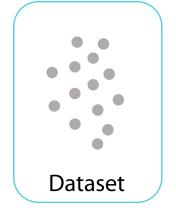# AnchorViz is not merely an error discovery tool

- Concept externalization & storage

- Concept evolution
  - Similar to structured labeling (Kulesza et al. 2014)

- Featuring

Future Work

More ways to
define & manipulate
anchors

Integration with
a complete iML loop

More evaluations with
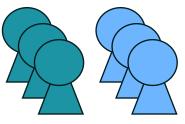various setups

Acknowledgements

Machine Teaching Group

Jina Suh

Johan Verwey

Gonzalo Ramos

Steven Drucker

Patrice Simard

Microsoft Research

All participants & reviewers

And you!

Nan-Chen Chen
nanchen@uw.edu

HUMAN CENTERED DESIGN & ENGINEERING
UNIVERSITY of WASHINGTON