The Label Complexity of Mixed-Initiative Classifier Training

Jerry Zhu / University of Wisconsin-Madison
Jina Suh, Saleema Amershi / Microsoft Research

June 22, 2016

http://aka.ms/mixedinitiative

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Microsoft Research

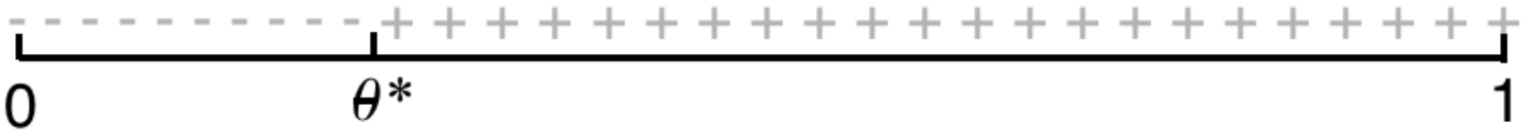
MOTIVATION

- Train a classifier
- Start with no labeled data
- Use human annotators
- Main message: Do not run active learning. We have a better procedure

THEORETICAL LABEL COMPLEXITY

Interactive classifier training

- (Example) train this 1D threshold classifier:



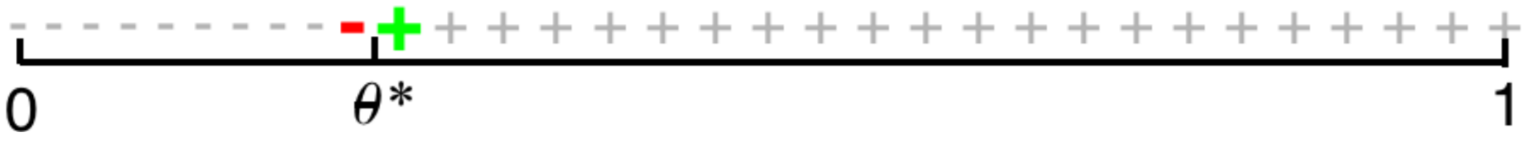
- Human oracle, starting from no labeled data
- Cost $=$ label complexity
- Shall we do active learning?

- Training is computer-initiated: computer picks x to query, no idea where θ^* is
- Active learning label complexity (realizable, Kulkarni et al. 1993)

$$AL \geq \log(1-\delta) + \max\left(VC, \log\frac{1}{\epsilon}\right)$$

But an optimal oracle can do better!

- Human knows θ^*, provides optimal teaching set:



- Teaching dimension $TD =$ smallest training set size to teach a concept
- $TD \leq AL$ always [Cakmak & Thomaz 2011; Angluin 2004; Goldman & Kearns 1995]
- $TD \ll AL$ often
- Training is human-initiated: human must pick the teaching set
-

	optimal oracle
computer-initiated	AL
human-initiated	TD

But humans are not always optimal oracles

- Naive oracle: can
  be arbitrarily bad in picking x (but always gives correct labels y)

	optimal oracle	naive oracle
computer-initiated	AL	AL
human-initiated	TD	∞

- Best of the two worlds: mixed-initiative training

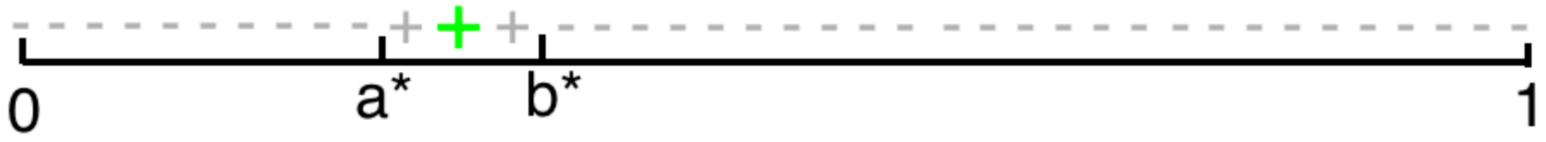
The mixed-initiative algorithm

Data $D = \emptyset$
for $i = 1$ to TD **do**
  if human no longer wants to lead **then**
    break;
  else
    human chooses (x_i, y_i)
    append (x_i, y_i) to D
  end if
end for
run active learning starting from D until completion

	optimal oracle	naive oracle
computer-initiated	AL	AL
human-initiated	TD	∞
mixed-initiated	TD	$TD + AL$

“Neither optimal nor naive” oracle

- Seed oracle: provides one point per positive region



-

	optimal	seed	naive
computer-initiated	AL	AL	AL
human-initiated	TD	∞	∞
mixed-initiated	TD	$TD + AL - AL_B$	$TD + AL$

Teacher education

- Goal: naive or seed $\rightarrow$ optimal oracle
- Show analogues: "To teach θ' you could have used D' (optimal teaching set)"
- Show expert-written explanation

EMPIRICAL
LABEL COMPLEXITY

Experiment Setup

Mechanical Turk, between-subjects

481 participants

Integer 1D threshold and interval classifier

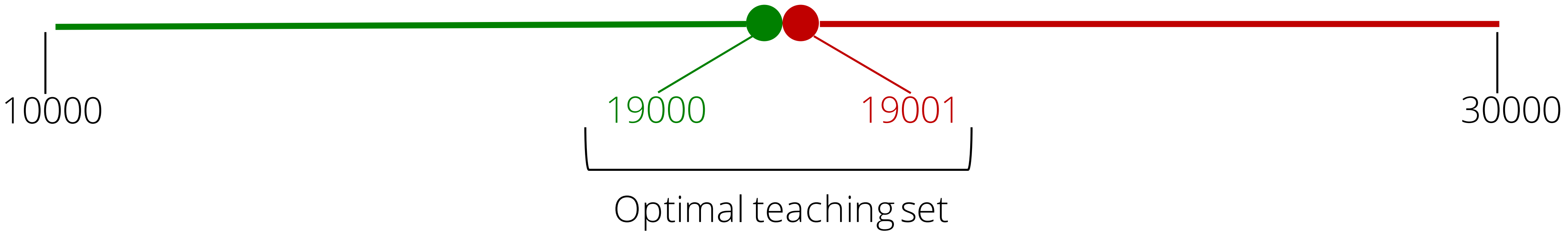
Goal: Teach a robot assistant acceptable car prices

1D Threshold Classifier Task

	Human	Mixed	Computer
No Ed.	1	2	3
Analogue	4	5	x

Conditions	5
ϴ*	19000; inclusive
Input range	[10000, 30000]
AL complexity	14

"If your price threshold was $20000 or below, you could show your robot these 2 examples: $20000 is acceptable, $20001 is unacceptable"



1D Interval Classifier Task

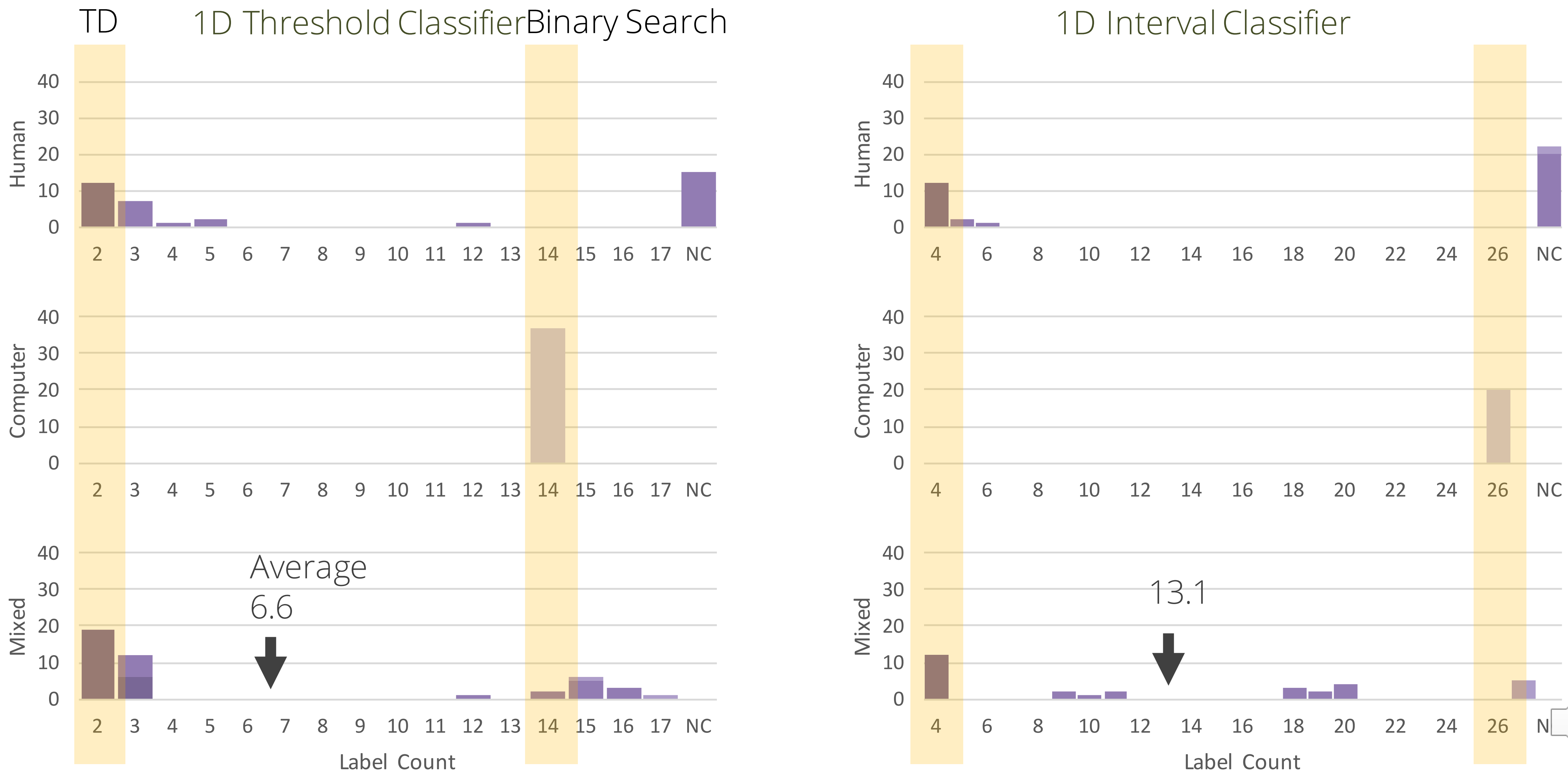
	Human	Mixed	Computer
No Ed.	1	2	3
Analogue	4	5	x
Explanation	6	7	x

Conditions	7
a*, b*	1260, 1360; inclusive
Input range	[500, 1500]
AL complexity	26

Procedure

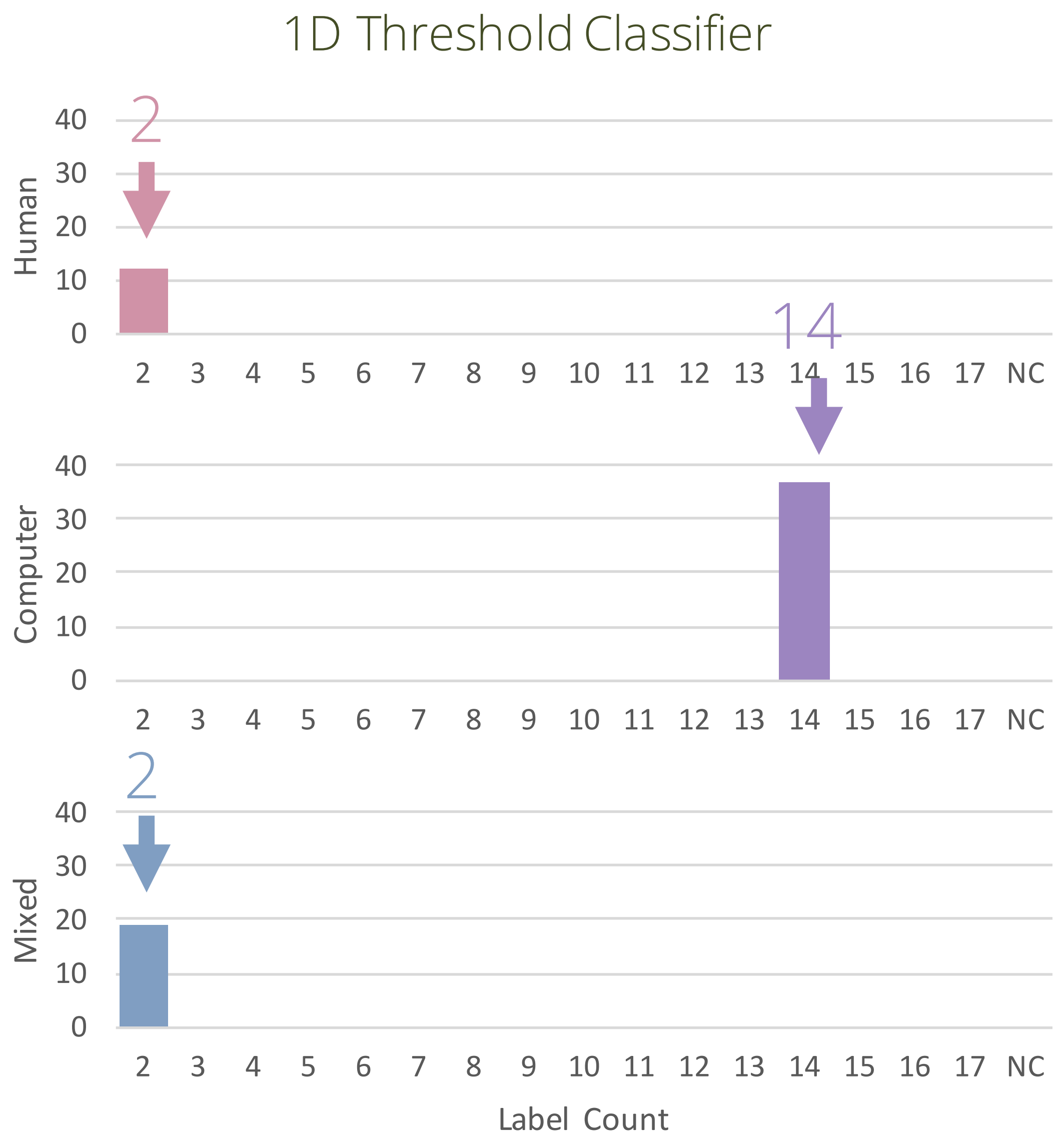
Instructions	Teacher Education	Training Task	Post-Task Survey
Cover story Target concept Instructions	Step-by-step tutorial with quizzes Hints	Human-initiated Computer-initiated Mixed-initiative	Demographics Difficulty Confidence Teaching experience Attention Numeracy Teaching strategy

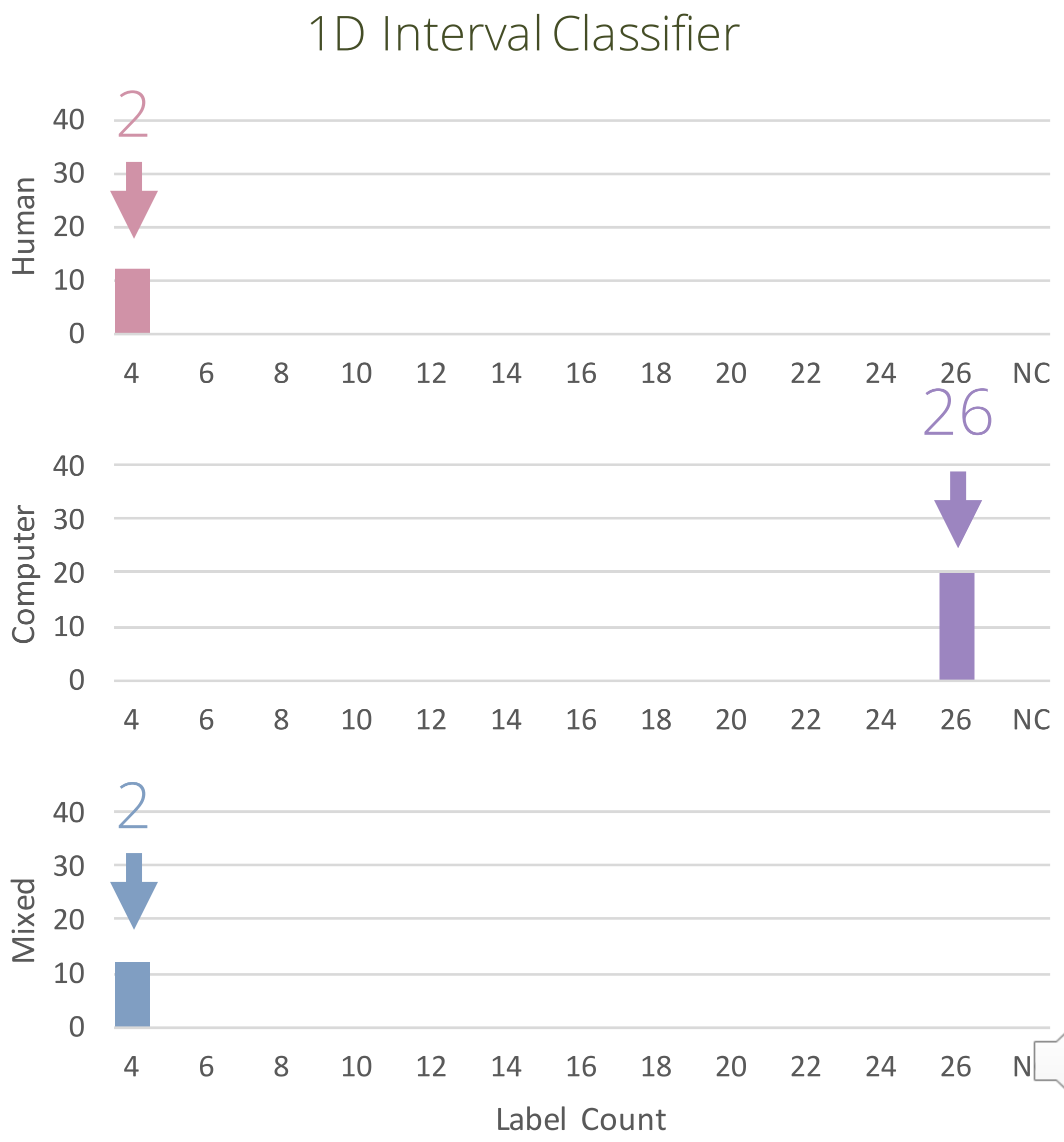
Empirical Label Complexity



Optimal Teachers

Human	TD
Computer	AL
Mixed	TD

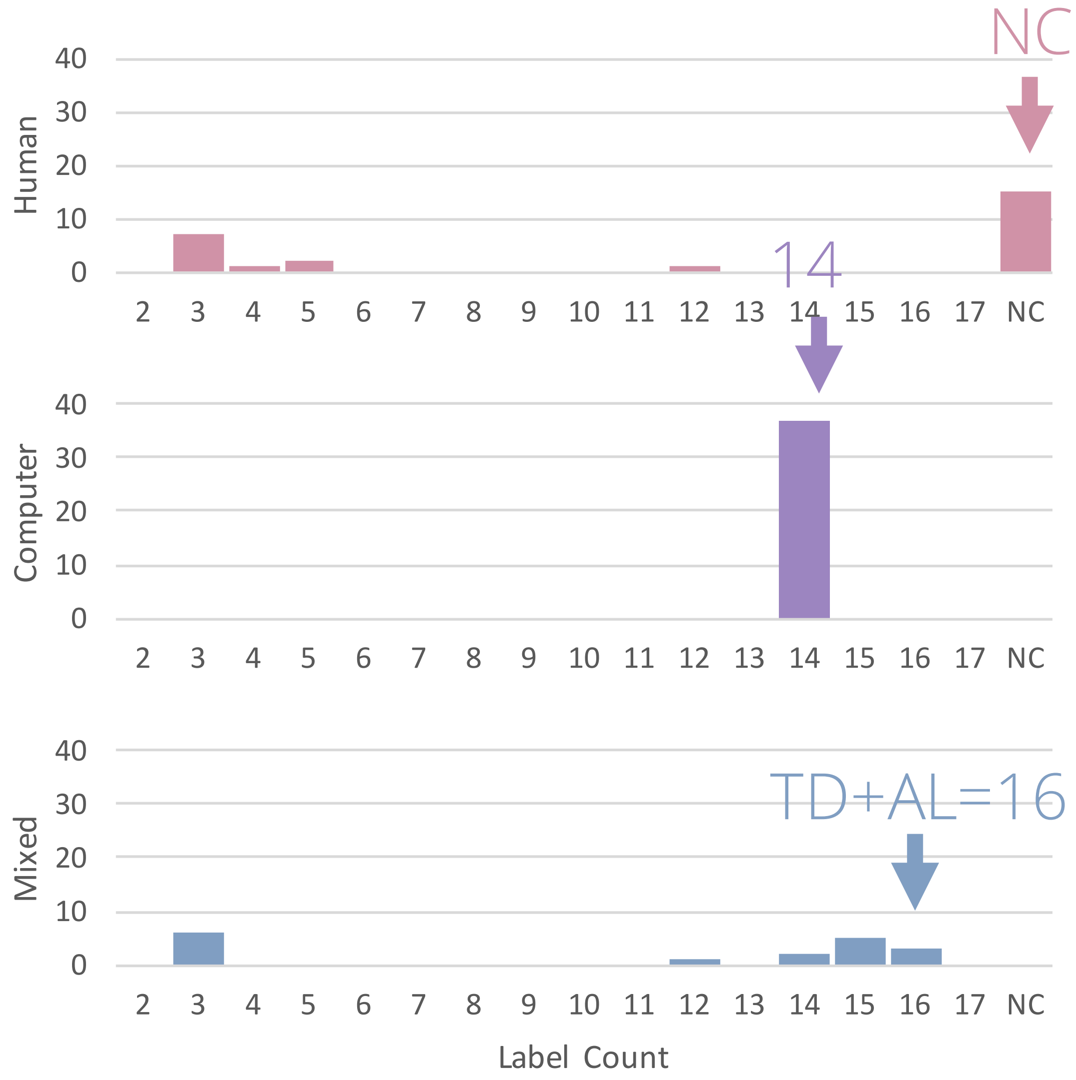




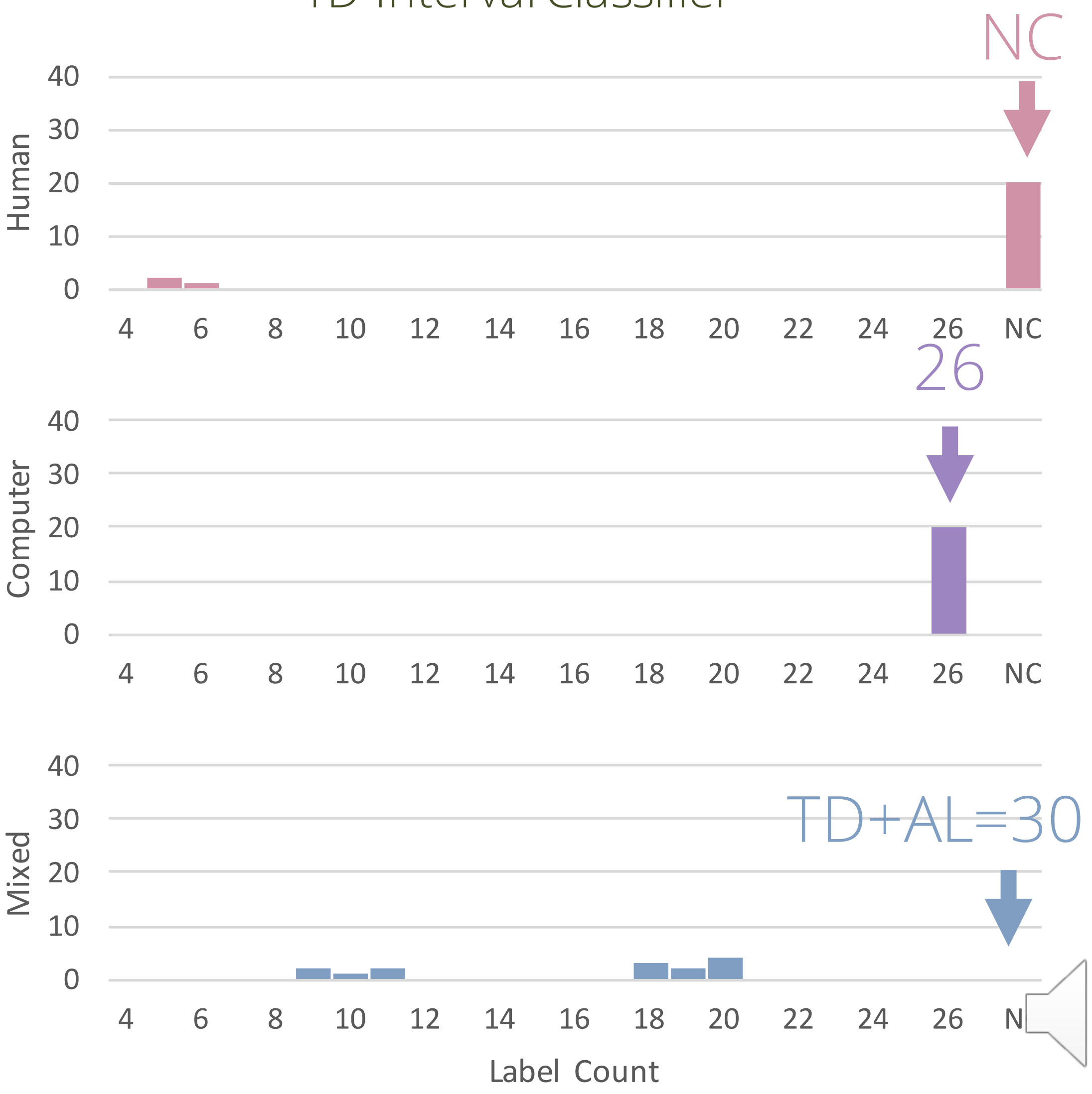
Seed Teachers

Human	∞
Computer	AL
Mixed	TD + (AL – AL_B)

1D Threshold Classifier



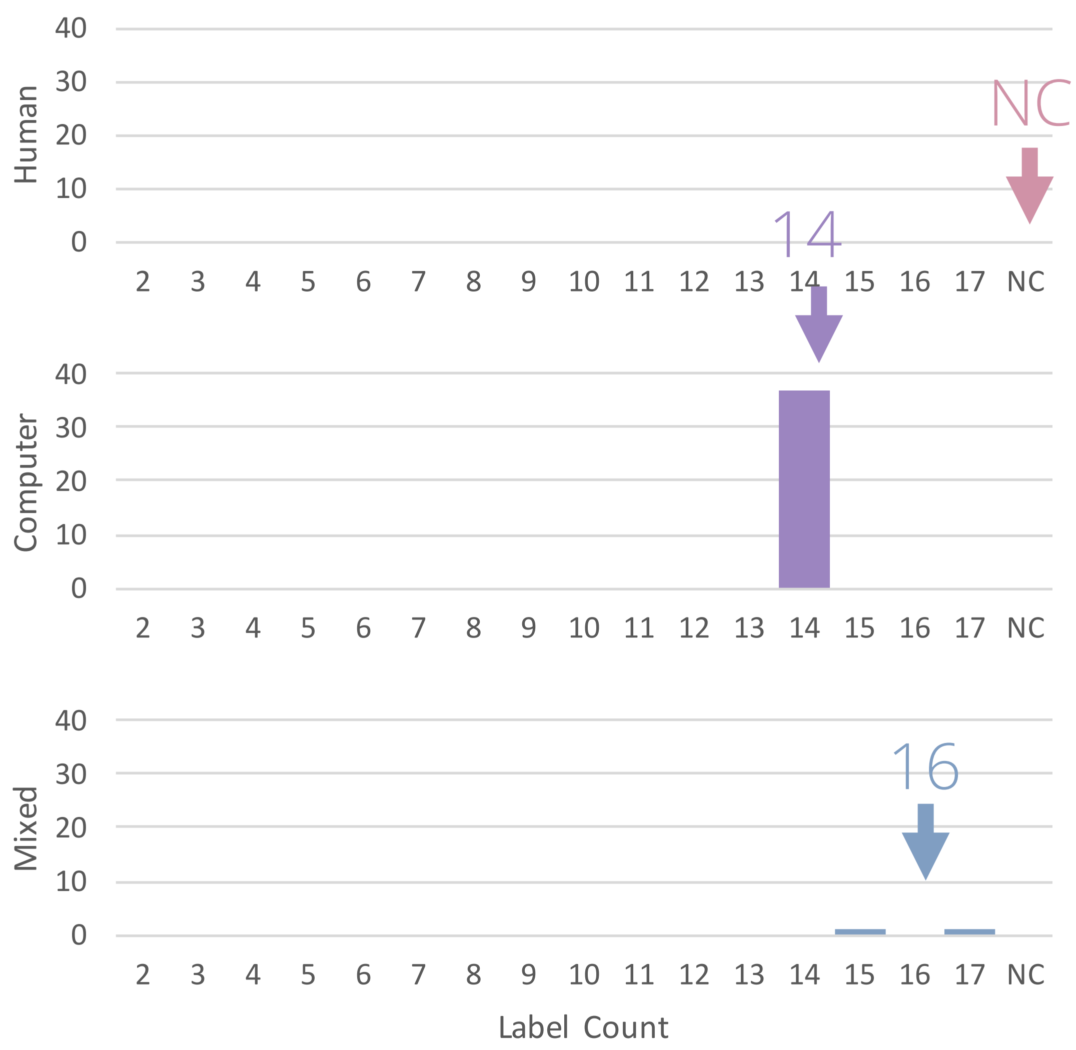
1D Interval Classifier



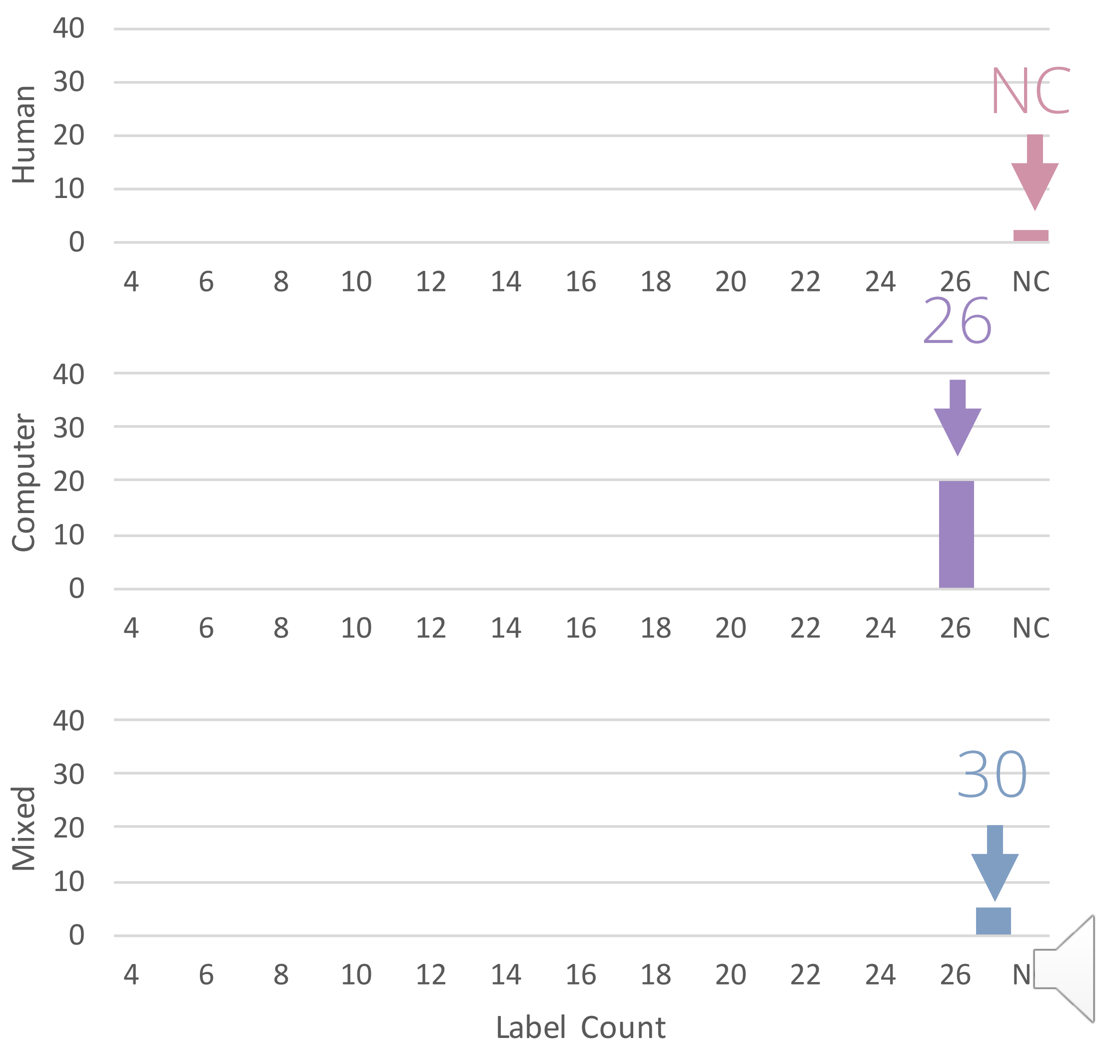
Naïve Teachers

Human	∞
Computer	AL
Mixed	TD + AL

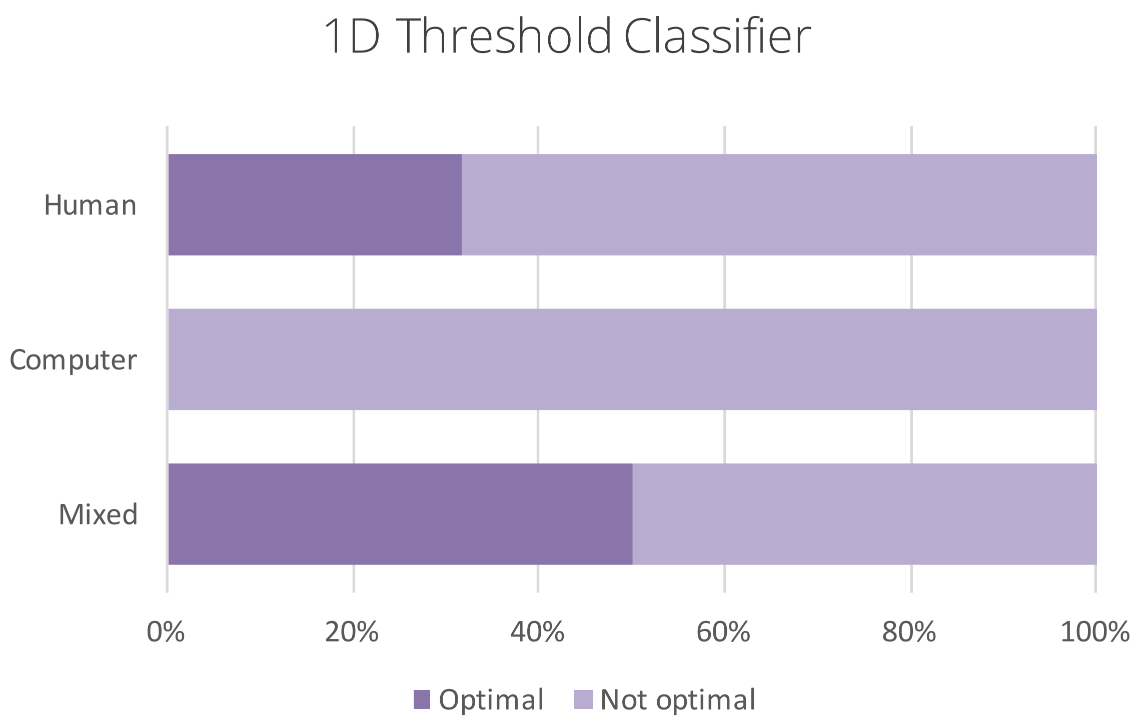
1D Threshold Classifier

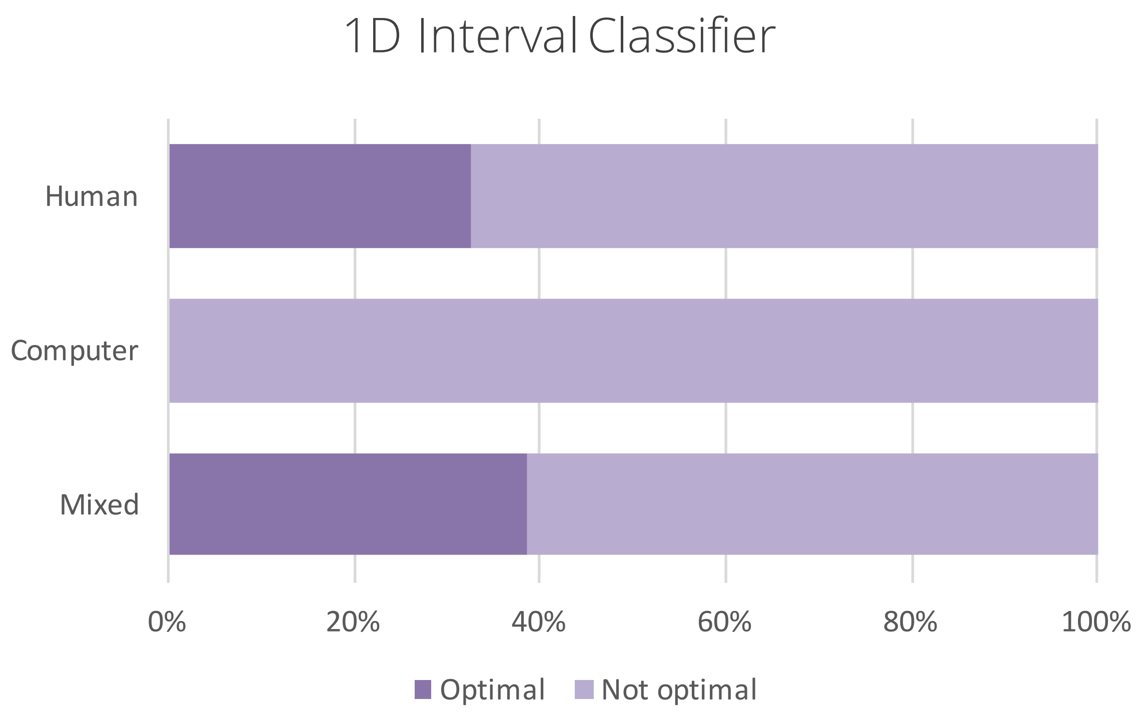


1D Interval Classifier



Effects of Training Paradigms on Optimal Teachers





Benefits of Mixed-Initiative Training

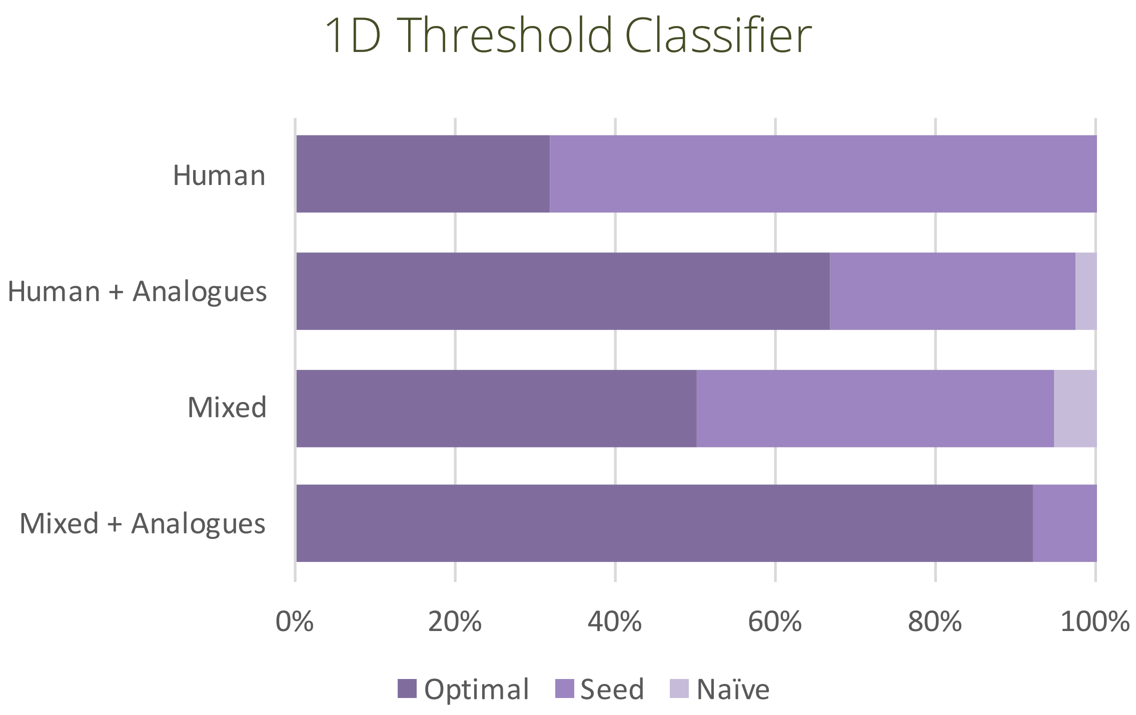
Enables optimal teaching

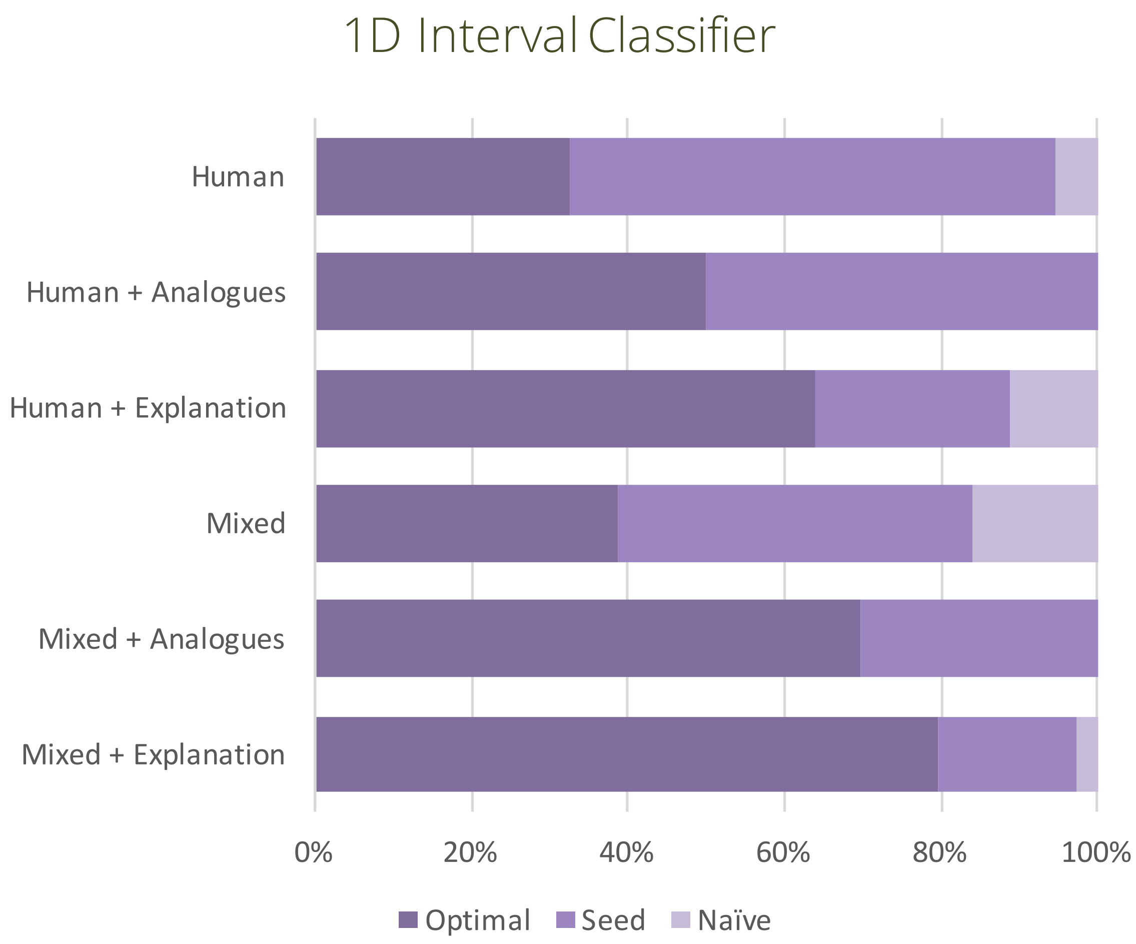
Prevents over-teaching

Eliminates not-completed (NC) participants

Removes blind search complexity

Effects of Teacher Education



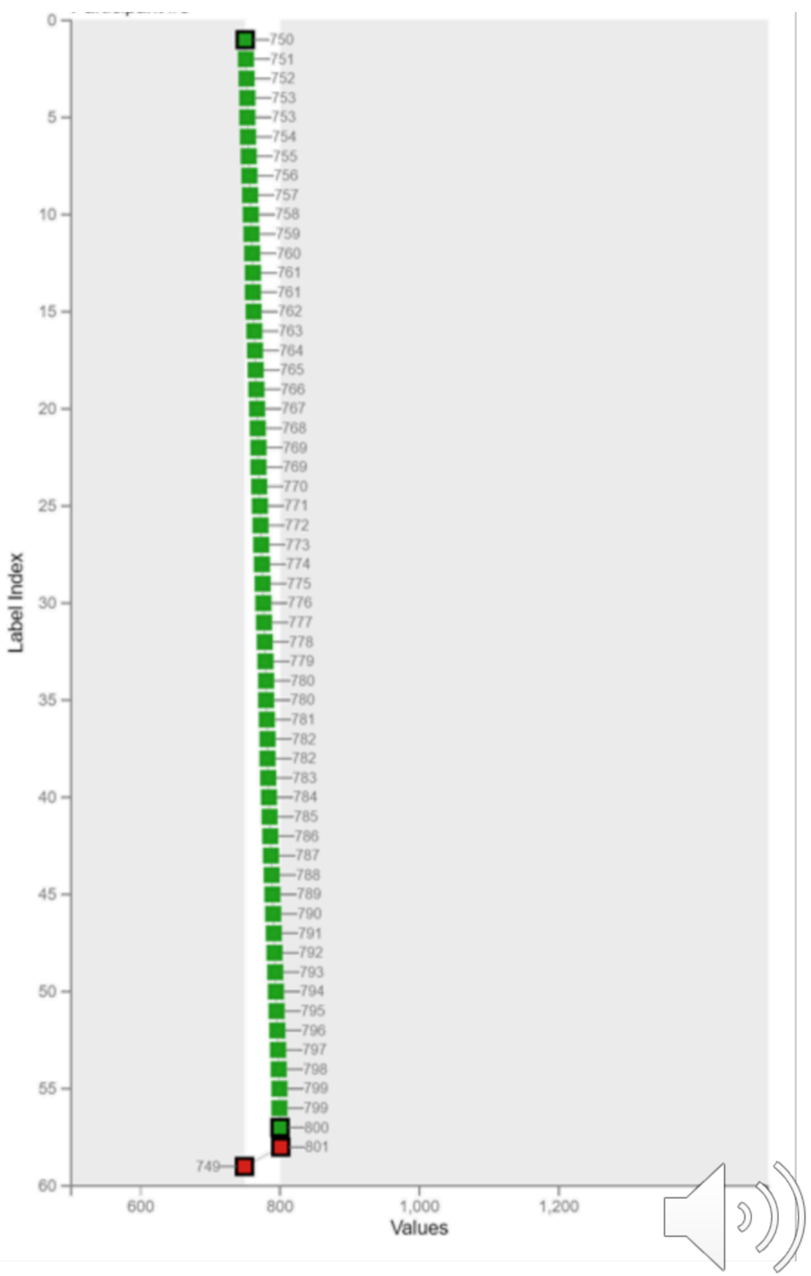


BRIDGING THEORY AND HUMANS

Humans alone are inefficient

Humans can provide more than the necessary TD training items. (29% in threshold, 8.1% in interval)

“I taught robot all acceptable price ranges.”

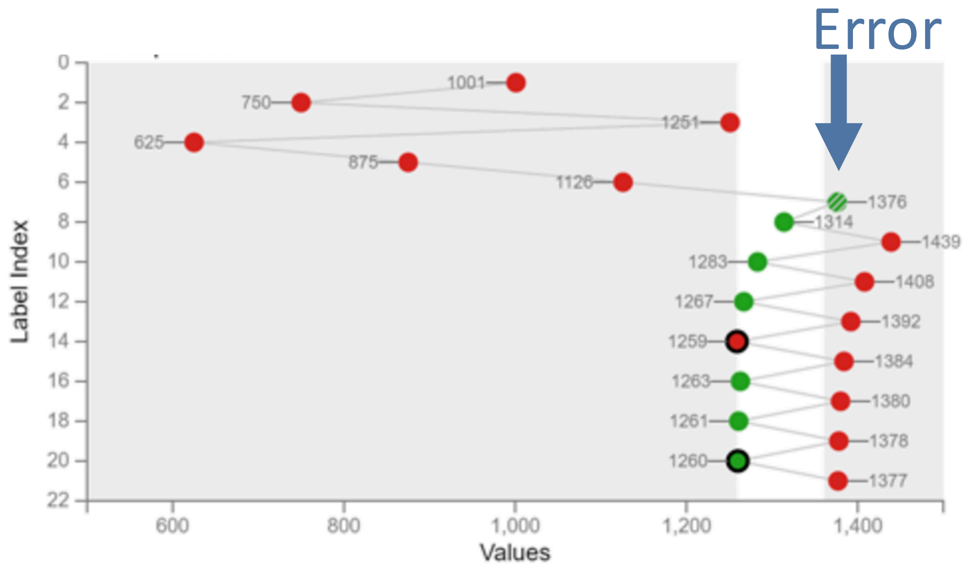


→ Support for mixed-initiative training

Humans are noisy

Human teachers provided wrong labels 3.5% of the time.

Nearly half (19/39) of the participants in computer-initiated, interval

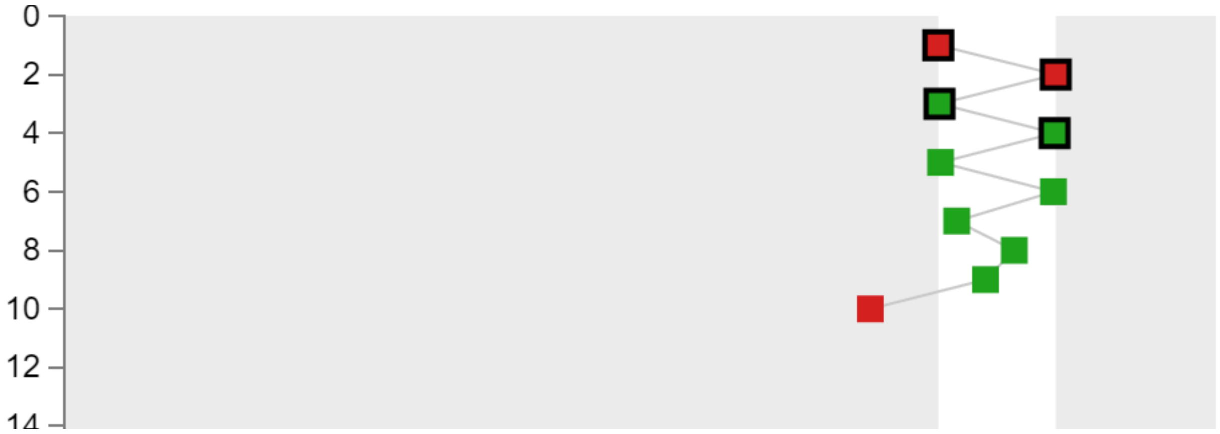


→ Allow for attentive labeling or correcting mislabels

Humans have incorrect mental models

Several participants did not understand the robot

“My teaching strategy was to provide the lowest and highest acceptable prices, then provide some acceptable prices in between the range.”



→ Educate humans how to interact with ML algorithms

Labeling effort = label complexity?

Manually selecting an example requires more cognitive effort than providing a label for a given example.

Computer-initiated: 17.3 labels/min
Human-initiated: 2.8 labels/min

→ Help humans explore data or generate examples efficiently

Future research

Teacher education strategy

Interaction techniques or translation layer

Efficient exploration

Other mixed approach

SUMMARY

- Formal justification of mixed-initiative classifier training
- Label complexity analysis and empirical verification
- Benefits of a mixed-initiative training and teacher education
- Limitations and design implications
- Future research directions
- Main message: Mixed-initiative training is a better procedure

Thank you!

Jerry Zhu / University of Wisconsin-Madison
Jina Suh, Saleema Amershi / Microsoft Research

June 22, 2016

http://aka.ms/mixedinitiative