



Uncover the factors to help measure how young children learn

Yu, Shenglan
danyus@Stanford.edu

Introduction

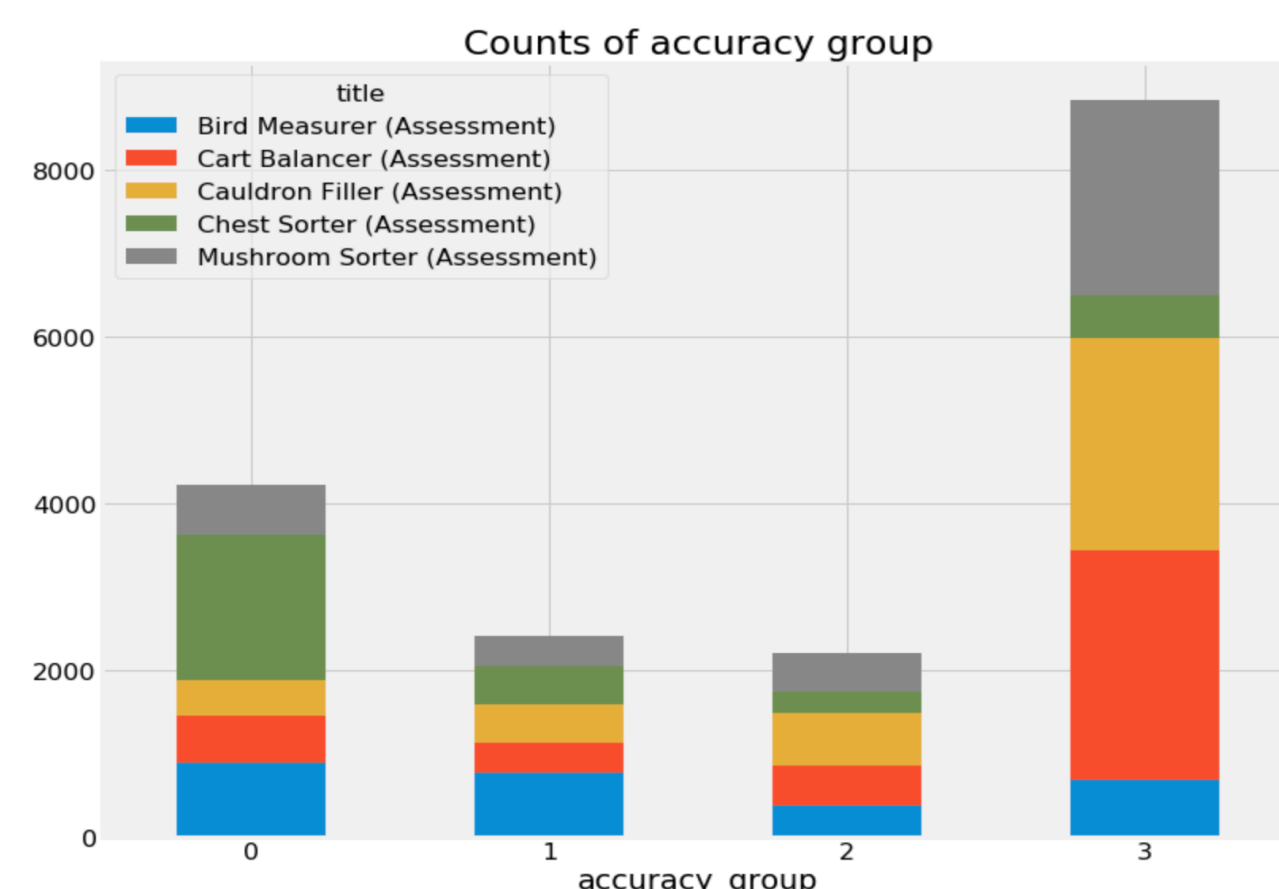
Use Apps to teach young children early STEM concepts like length, width, capacity and weight.

- Through interactive approaches like short video and games in App.
- Child take assessments to get evaluation.
- Goal: use the gameplay data to forecast how many attempts a child will take to pass a given assessment.



Data

- Data comes from Kaggle competition(1).
- The game data collected from the App: PBS Kids Measure Up! Activities(2)
- Each row represents one game event in one game session on one device.
- E.g. row describe the event of dragging a dinosaur using finger. (start_x, start_y, end_x, end_y ...)



Evaluation Metric

Metric – Quadratic weighted kappa

$$\kappa = 1 - \frac{\sum_{i,j} w_{ij} O_{ij}}{\sum_{i,j} w_{ij} E_{ij}}$$

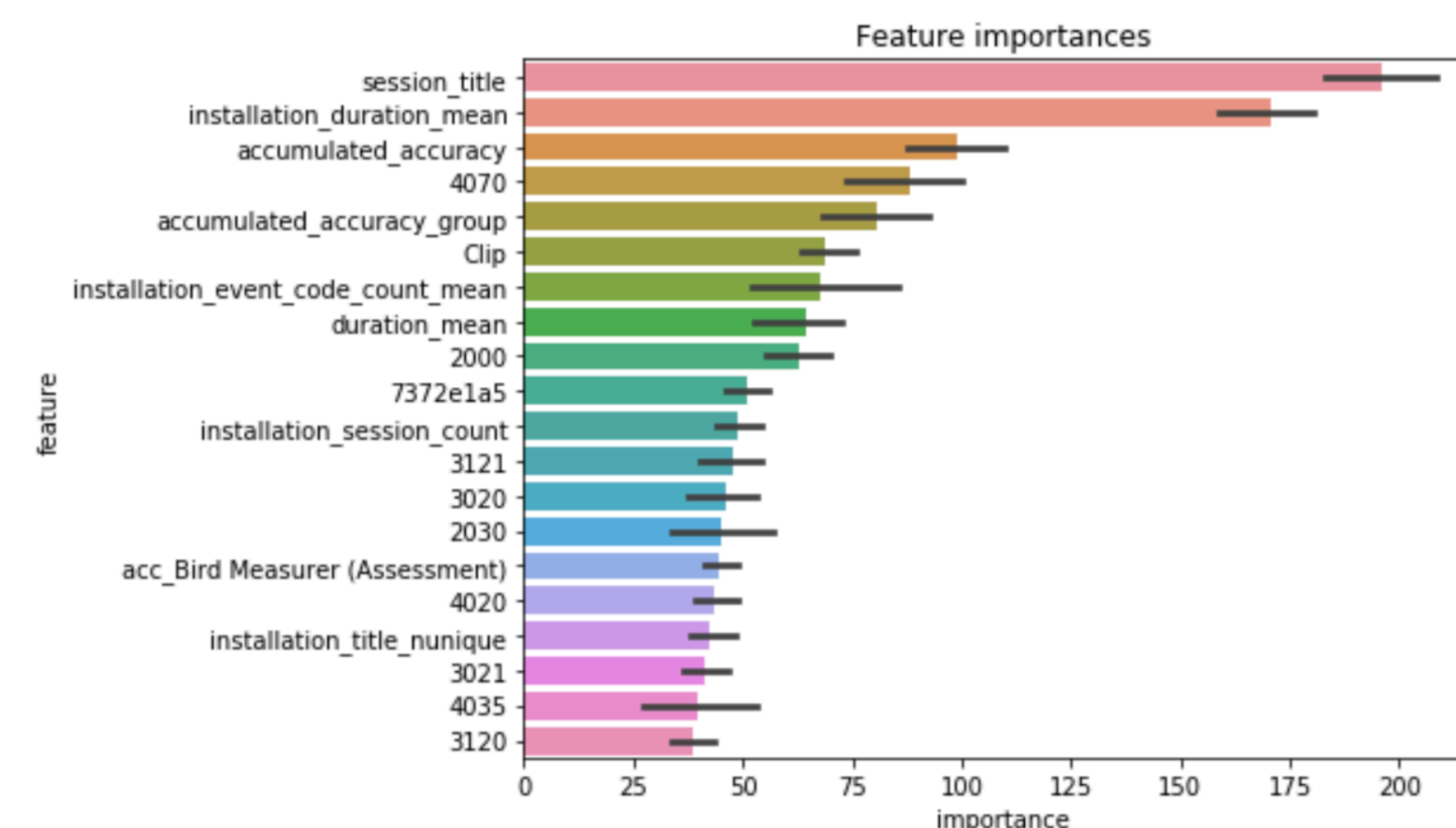
$$w_{ij} = \frac{(i - j)^2}{(N - 1)^2}$$

Supervised Multi-class classification:

- 3: the assessment solved in first attempt
- 2: the assessment solved in second attempt
- 1: the assessment solved in three or more times
- 0: never solved

Feature

- Aggregate rows under same assessment game session to make samples.
- Around 900 engineered features. Assessment feature, each event type feature, duration, history performance ...



Model

Boost regression + rounding:

- Gradient boost regression to get numeric values.
- Use optimization to find the best discrete interval.
- Discretize into classes.

Baseline: take the median of history assessment performance as prediction

Other learners: SVM, random forest...

Results

Model	Train kappa	Test kappa
Baseline	0.4	0.39
Boost classification	0.65	0.46
Boost regression + rounding	0.71	0.536

- Trained 250 epochs.
- Training: 17000 samples
- Test: 1000 samples
- SVM, random forest performs poorly

Discussion

- Assessment on geography concept has the highest performance while length concept has the worst performance.
- Playing time and history performance has strong impact.
- Events representing distraction(things the player think should do something but don't) strong impact.
- Still a lot of game play data not used well, thus high bias. (data not well structured)

Future

More feature engineering: use more game play data like the dinosaur color, cloud size in the game.

Reference

- (1) Kaggle 2019 data science bowl.
<https://www.kaggle.com/c/data-science-bowl-2019/data>
- (2) PBS Kids Measure Up!
<https://pbskids.org/apps/pbs-kids-measure-up.html>

- Video presentation link:
https://youtu.be/z3YIf_LCJOc