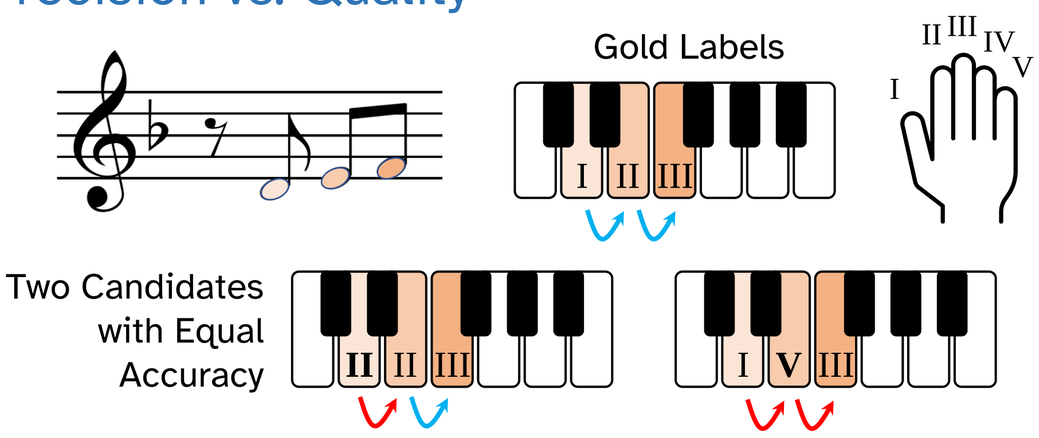
Checklist Models for Improved Output Fluency in Piano Fingering Prediction

Nikita Srivatsan & Taylor Berg-Kirkpatrick
nsrivats@cmu.edu, tberg@eng.ucsd.edu

Summary

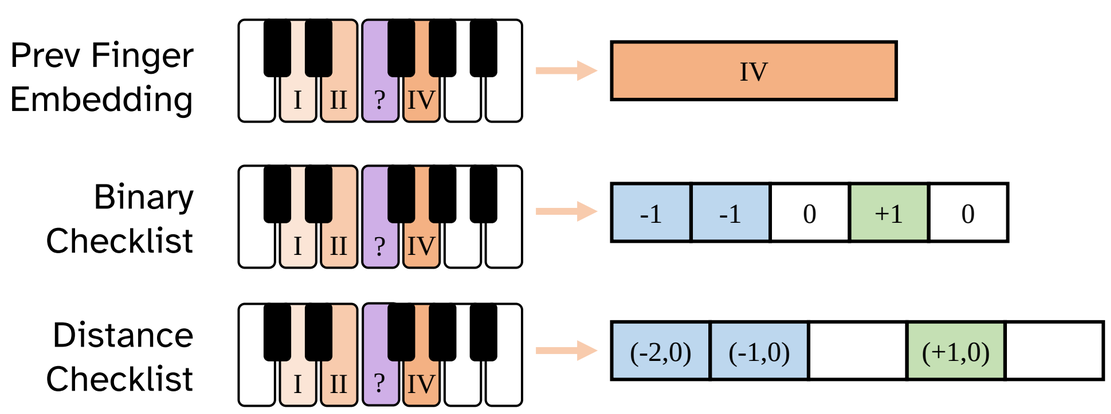
In this work we present a new approach for the task of predicting fingerings for piano music that maintains a checklist representation of recent predictions, allowing it to learn soft constraints on output structure. We note that per-note labeling precision does not adequately measure the human playability of a model's output, and therefore compare methods across several statistics which track the frequency of challenging patterns, and implement a reinforcement learning strategy to minimize these as part of our training loss.

Precision vs. Quality



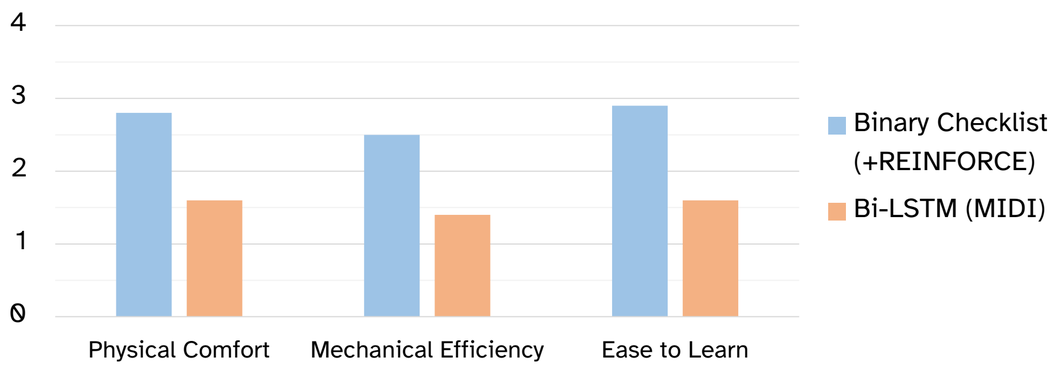
Raw accuracy often fails to reflect nuanced differences in the playability of model outputs arising from interactions between nearby predictions

Checklists



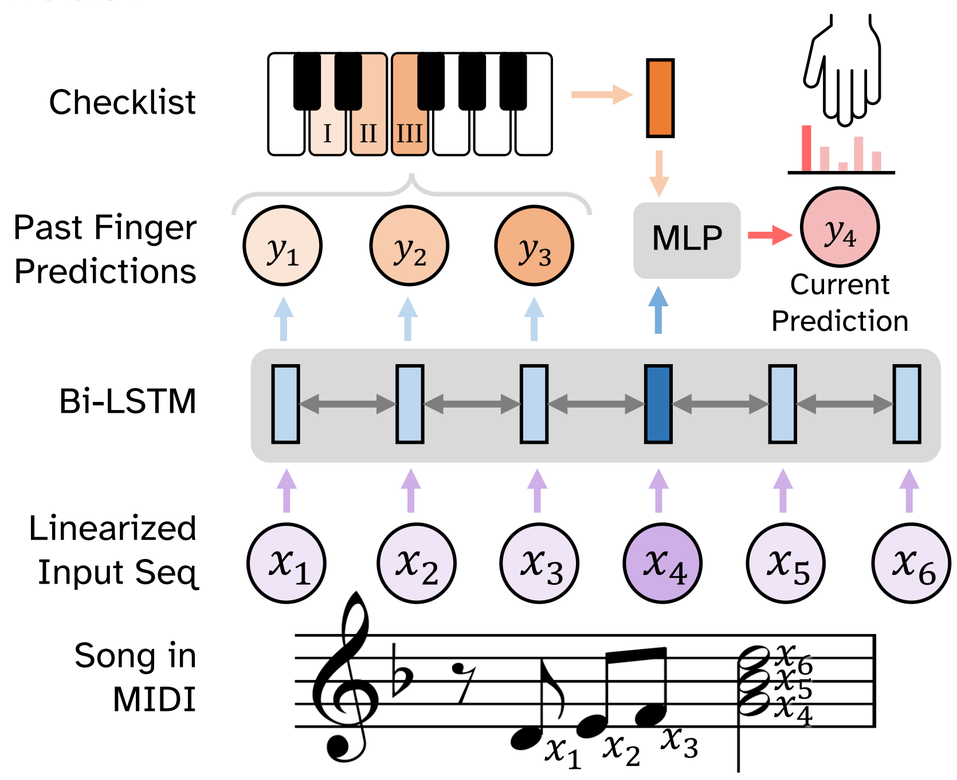
We compare various encodings of fingering history, including an embedding of just the previous finger (i.e. an MEMM), and checklists that track the current location of fingers relative to the next note

Human Eval



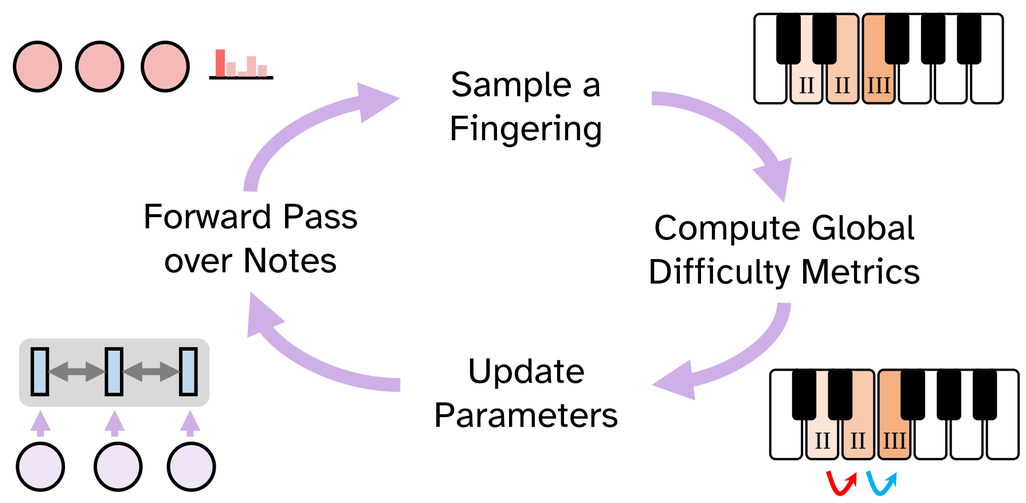
Results from human expert evaluation, showing qualitative improvements in output playability

Model



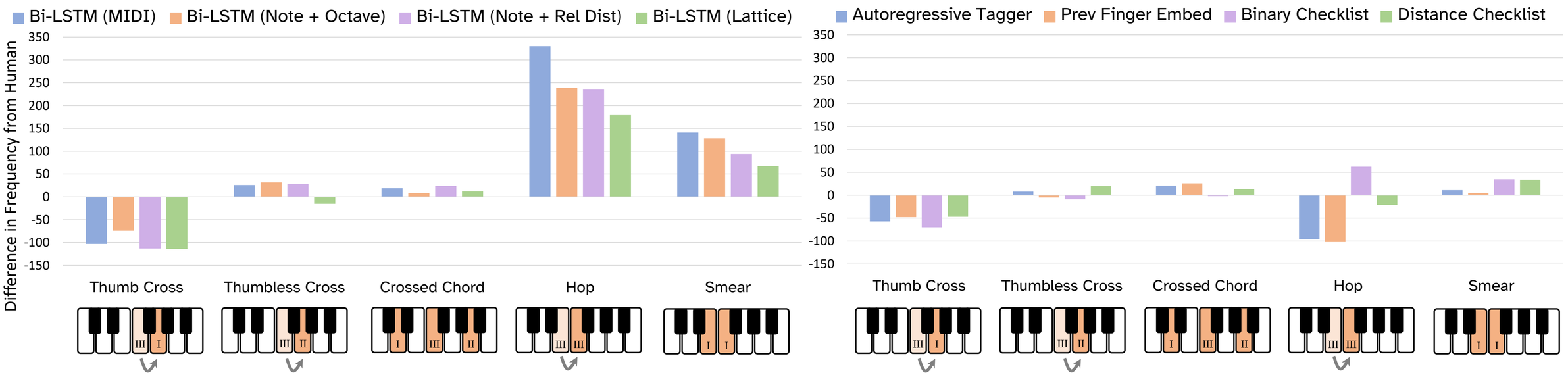
Overview of our checklist approach. Inputs are linearized and fed into a Bi-LSTM which then passes contextualized embeddings to an MLP, which predicts the next finger conditioned on a history of recent finger placements on the keyboard

REINFORCE



To encourage fluent output, we train using REINFORCE over non-differentiable difficulty metrics in addition to cross-entropy

Results



Quantitative results tracking the frequency of various fingering patterns that can significantly affect performability but may not be reflected in raw labeling accuracy. Left is an ablation of input representation strategies. Right shows our checklist models with REINFORCE over certain metrics.