Probabilistic Piano Fingering Prediction with Hidden Markov Models

ANASTASIA ISKALIYEVA
Dept Information and Computer Sciences
a.iskaliyeva@gmail.com

ABSTRACT

Determining optimal fingerings for piano music is a complex, cognitively demanding task that is crucial for performance. This project explores the application of a generative probabilistic model, the Hidden Markov Model (HMM), for predicting piano fingerings. Building upon previous work that established baselines using discriminative models like Random Forest (86% accuracy), RNN (51% accuracy), and LSTM (73% accuracy), this study implements an HMM to model the sequential nature of fingering choices.

The model's states correspond to the five fingers, and observations are the MIDI pitches of the notes. We investigate two approaches: one where the HMM parameters are learned directly from the data and another where the model is initialized with priors for transition, emission, and initial state probabilities calculated from the annotated dataset.

The results show that initializing the model with data-derived priors significantly improves performance from 21.4% to 36.5% accuracy. While the HMM did not outperform the feature-engineered Random Forest baseline, the analysis provides a thorough investigation into a generative approach for this sequence labeling task and highlights key challenges, such as the constraints of the Markov assumption and the importance of feature representation.

1. INTRODUCTION

The process of selecting fingerings for a piano piece is fundamental to a musician's learning process. A well-chosen fingering pattern can facilitate fluid execution, technical accuracy, and musical expression, while a poor one can create unnecessary difficulty. While many musical scores come with fingering annotations, a vast number do not, leaving the task to the performer. This manual annotation is often a time-consuming and challenging process.

My prior research, "Predicting Piano Fingerings with Machine Learning", established the viability of using machine learning to automate this task. By comparing Random Forest, Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) models, the study found that a Random Forest model, explicitly engineered with features representing the context of surrounding notes, achieved the highest accuracy of 86%. This highlighted the importance of local note context in making fingering decisions.

This project extends that research by exploring a different modeling paradigm: the Hidden Markov Model (HMM). As a generative model, an HMM is naturally suited for sequence labeling tasks like piano fingering. It models the problem as a sequence of hidden states (the fingers) that produce a sequence of observable outputs (the notes). This approach allows us to probabilistically model the underlying structure of fingering patterns, including the likelihood of transitioning from one finger to another and the probability of a certain finger being used for a specific note.

This report details the implementation of a Categorical HMM using the `hmmlearn` library. We analyze the model's performance both with and without the use of pre-calculated priors and compare its accuracy against the previously established baselines.

2. DATASET AND PREPROCESSING

The dataset used for this project is the same one established in the previous work. It was created by parsing 43 publicly available, fingering-annotated piano scores in MusicXML format from Musescore.com.

- **Total Notes:** ~39,000
- **Annotated Notes:** ~16,000 (for both hands)
- **Features Extracted:** Note pitch (as a MIDI number), chord information, and rest information.

For the HMM implementation, the following preprocessing steps were performed:

1. **Hand Selection:** The analysis was focused exclusively on the right-hand part, as is common in fingering prediction research.
2. **Phrase Segmentation:** Rests break the continuity of fingering sequences. To handle this, the musical data was segmented into "phrases," where each phrase is a continuous sequence of notes between rests or at the beginning/end of a piece. Sequences of length 1 were discarded, as they provide no transitional information.
3. **Data Splitting:** The dataset of phrases was split into a training set (80%) and a testing set (20%). The training data consists of MIDI pitch sequences and their corresponding lengths, while the testing data also includes the ground-truth finger sequences for accuracy evaluation.

3. METHODOLOGY: THE HIDDEN MARKOV MODEL

A Hidden Markov Model was chosen for its inherent ability to handle temporal sequences. The model was defined with the following components:

- **Hidden States (N):** The five fingers of the right hand, represented as states {1, 2, 3, 4, 5}.
- **Observations (M):** The discrete MIDI pitch values of the notes being played [1: 108].
- **HMM Parameters (λ):**
    - **Initial State Probabilities (π):** A vector representing the probability of starting a musical phrase with a given finger. This was estimated by calculating the distribution of the first finger used after a rest or at the beginning of a piece.
    - **Transition Probabilities (A):** A 5x5 matrix where the probability of transitioning from finger *i* to finger *j*. This was estimated by counting the occurrences of consecutive fingerings in the annotated data.

- **Emission Probabilities (B):** A 5x108 matrix where B_{jk} is the probability of playing note *k* (observation) with finger *j* (state). This was estimated from the frequency of each finger being used for each specific MIDI pitch.

Two models were trained and evaluated:

1. **HMM without Priors:** A `CategoricalHMM` trained directly on the note sequences, allowing the Baum-Welch algorithm to estimate all parameters (A, B, π) from scratch.
2. **HMM with Priors:** A `CategoricalHMM` initialized with the pre-calculated A, B, and π matrices estimated directly from the training data.

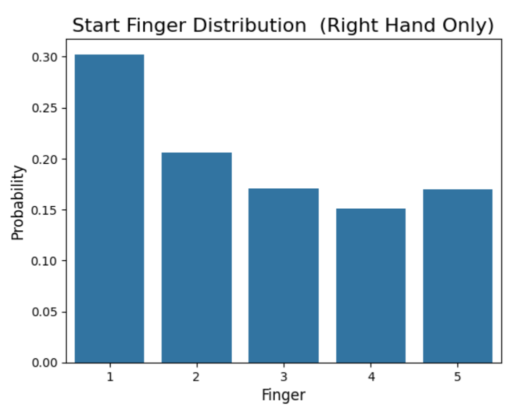
Inference to predict the most likely sequence of fingerings for a new piece of music is done using the **Viterbi algorithm**.

4. RESULTS AND DISCUSSION

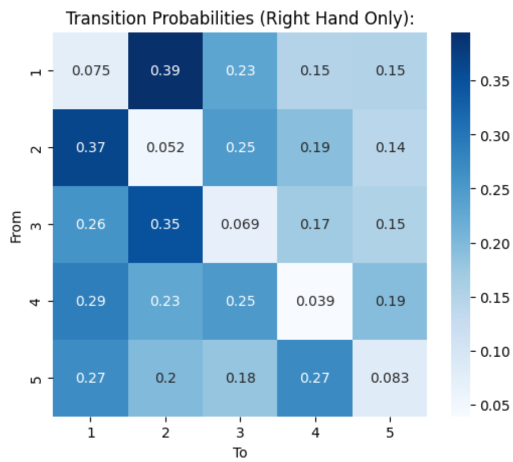
4.1. Estimated Priors

Before training the full model, the initial, transition, and emission probabilities were calculated from the dataset to serve as priors. These distributions provide valuable insight into common fingering practices.

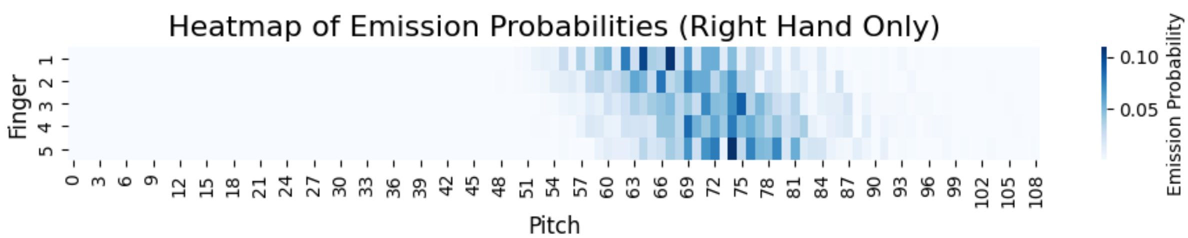
Initial State Distribution: The probability distribution for the first finger of a phrase shows a clear preference for the thumb (1) and pinky (5), which are common starting fingers for scales and arpeggios.



Transition Probabilities: The transition matrix heatmap shows the probability of moving from one finger to the next. High probabilities are seen along the near-diagonal, representing stepwise motion (e.g., 1→2, 2→3, 2→1). It is very uncommon to reuse the same finger for consecutive notes, as indicated by the low probabilities on the main diagonal.



Emission Probabilities: The emission probability heatmap visualizes which fingers are most likely to play which pitches. It reveals an ergonomic pattern where lower MIDI pitches (to the left) are frequently played by the thumb (1) and higher pitches (to the right) are more commonly played by the outer fingers (4 and 5).

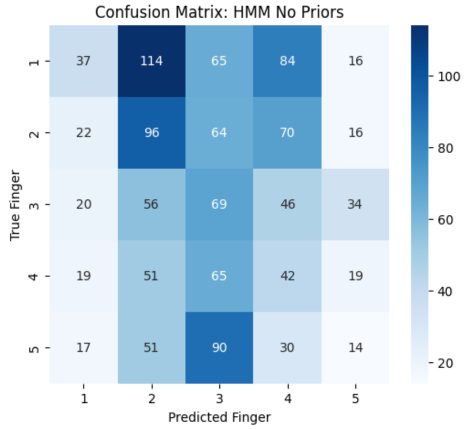


4.2. Model Performance

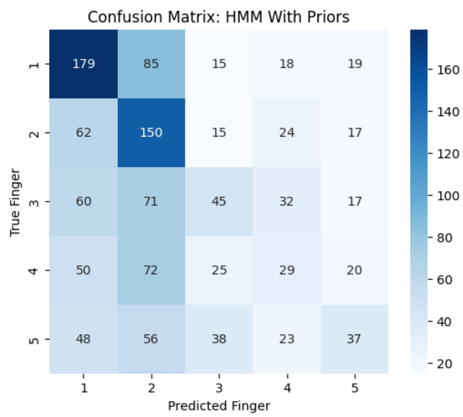
The models were evaluated on the test set using fingering accuracy, which is the percentage of correctly predicted fingers on notes that had a ground-truth annotation.

Model	**Test Set Accuracy**
HMM (without priors)	*21.38%*
HMM (initialized with priors)	***36.45%***
Baseline: Random Forest	*86%*
Baseline: LSTM	*73%*
Baseline: RNN	*51%*

HMM without Priors: The model trained from scratch achieved an accuracy of **21.38%**. The confusion matrix below shows that the model heavily favors predicting fingers 2 and 3, struggling to differentiate finger usage correctly.



HMM with Priors: Initializing the model with the data-derived priors led to a notable improvement, with accuracy reaching **36.45%**. The confusion matrix shows a much more reasonable distribution of predictions, with higher values along the diagonal, indicating more correct predictions for each finger.



4.3. Discussion

The results clearly demonstrate that initializing the HMM with well-estimated priors is crucial for this task. However, even with this improvement, the HMM's accuracy of 36.5% falls significantly short of the 86% achieved by the Random Forest baseline.

Several factors likely contribute to this performance gap:

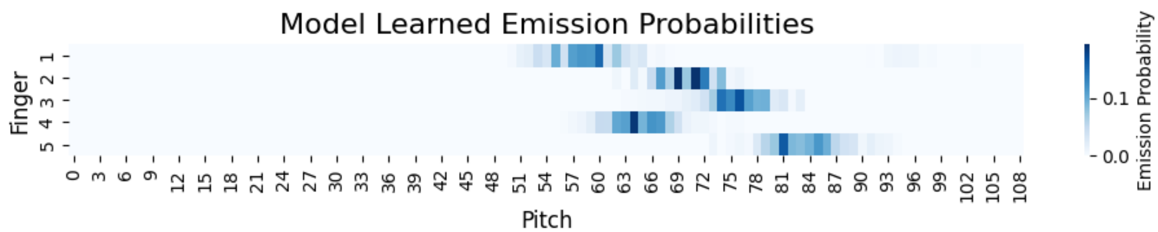
1. **The Markov Assumption:** The HMM assumes that the choice of the current finger only depends on the previous finger. This is a significant oversimplification. Pianists often make decisions based on the context of several preceding and upcoming notes to optimize hand position.
2. **Feature Representation:** The HMM uses only a single feature for its observations—the MIDI pitch. In contrast, the high-performing Random Forest model was engineered with features representing a window of four previous notes and two upcoming notes (both pitch and finger), providing it with much richer contextual information for each prediction.
3. **Generative vs. Discriminative:** The HMM is a generative model that learns the joint probability distribution P(fingers, notes). The Random Forest is a discriminative model that

directly learns the conditional probability P(finger | notes). For a classification task like this, a discriminative approach is often more direct and effective, especially when provided with strong, contextual features.

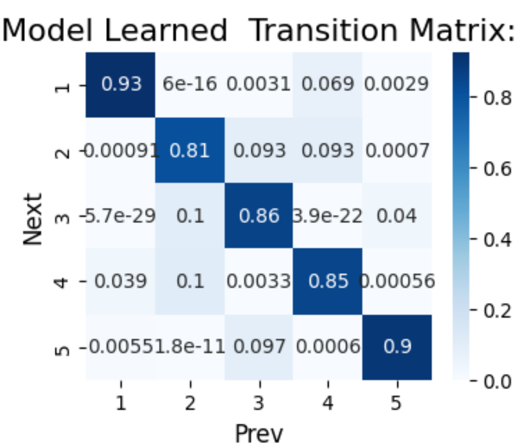
5. QUALITATIVE ANALYSIS

The HMM learned three key probability distributions—start finger, transition, and emission—that collectively represent the "knowledge" of piano fingering it acquired from the training data. From a pianist's standpoint, these distributions capture several fundamental and ergonomically sound principles of piano playing.

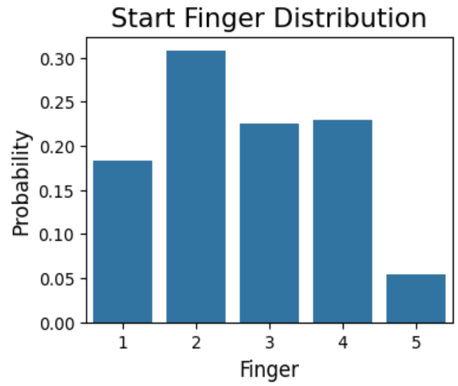
Emission Probabilities: The model has learned a rigid relationship between fingers and pitches, confining each finger to a very narrow range of notes. This fails to account for essential piano techniques like hand repositioning, crossovers, and thumb-under movements, where any finger might be required to play notes outside of a basic, static five-finger position.



Transition Matrix: The model incorrectly learned that moving from the third finger to the thumb (a core "thumb-under" technique for scales) has a near-zero probability of 6e-16. Furthermore, the probabilities on the diagonal are exceptionally high (e.g., 0.93 for finger 1 to 1, 0.9 for finger 5 to 5), suggesting the model heavily and unrealistically favors repeating notes with the same finger over fluidly moving between fingers.



Start Finger Distribution: The model shows a strong bias for starting a piece with the second finger, assigning it the highest probability of over 0.30. This undervalues the thumb (finger 1), which is one of the most common starting fingers but is given a much lower probability of less than 0.20.



In summary, the model appears to have learned a rudimentary set of rules for a static hand position, failing to capture the dynamic movements and essential techniques that are fundamental to proficient piano playing.

6. CONCLUSION AND FUTURE WORK

This project successfully implemented and evaluated a Hidden Markov Model for the task of automatic piano fingering prediction. The key finding is that while the HMM provides a sound probabilistic framework for this sequential task, its performance is highly dependent on proper initialization with data-derived priors. Even so, the model's simplifying assumptions and limited feature set prevent it from reaching the accuracy of discriminative models like Random Forest that leverage richer contextual features.

This work serves as a valuable analysis of a generative approach and lays the groundwork for future improvements. Potential directions for future work include:

- **Richer Observation Features:** Expanding the HMM's observation vector to include features like the interval to the next note, note duration, or whether a note is part of a chord.
- **Higher-Order HMMs:** Using a second-order HMM, where the current state depends on the two previous states, could capture more of the necessary musical context.
- **Hybrid Approaches:** Combining the probabilistic sequencing of an HMM with the predictive power of a neural network (e.g., using an RNN to output emission probabilities) could potentially yield superior results.

7. REFERENCES

1. Iskaliyeva, A. (Previous Work). "Predicting Piano Fingerings with Machine Learning: A Comparison of Random Forest, RNN, and LSTM."
2. Yonebayashi, Y., & Kameoka, H. (2007). "Automatic Decision of Piano Fingering Based on Hidden Markov Models".
3. Nakamura, E., & Yoshii, K. (2018). "Statistical Learning and Estimation of Piano Fingering." *International Society for Music Information Retrieval Conference (ISMIR).*
4. Code for this work: Google Colab Notebook