

**VIETNAM NATIONAL UNIVERSITY  
UNIVERSITY OF ENGINEERING AND TECHNOLOGY**



**SPEECH PROCESSING**

**HMM MODEL TRAINING  
REPORT**

Lecturer: Dr. Tran Quoc Long

Members:

Dao Tuan Anh	16020190
Pham Khanh Ly	17021291
Nguyen Thi Anh Trang	17021346

**HANOI, 05/2020**

---

# HMM MODEL

List word to recognition:

- "tôi"
- "dịch"
- "theo"
- "người"
- "bệnh\_nhân"

For each word, we build a model for it based on the Multinomial HMM model and the number of its phonemes. In details:

## 1. "tôi":

+ Phonemes:        /t/     /o/     /i/

+ Metrics: Accuracy

```
hmm = hmmlearn.hmm.MultinomialHMM(
    n_components=9, init_params='e', params='ste', verbose=True
)
hmm.startprob_ = np.array([0.6, 0.2, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0])
hmm.transmat_ = np.array([
    [0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.4, 0.6],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0]
])
```

## 2. "dịch":

+ Phonemes:        /d/     /i/     /c/

+ Metrics: Accuracy

```
hmm = hmmlearn.hmm.MultinomialHMM(
    n_components=9, init_params='e', params='ste', verbose=True
)
hmm.startprob_ = np.array([0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0])
hmm.transmat_ = np.array([
    [0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.5],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0]
])
```

## 3. "theo":

+ Phonemes:        /t<sup>h</sup>/   /e/     /o/

## + Metrics: Accuracy

```
hmm = hmmlearn.hmm.MultinomialHMM(
    n_components=9, init_params='e', params='ste', verbose=True
)
hmm.startprob_ = np.array([0.6, 0.2, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0])
hmm.transmat_ = np.array([
    [0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.4, 0.6],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0]
])
```

#### 4. "người":

+ Phonemes:        /ŋ/        /iə/        /j/

## + Metrics: Accuracy

```
hmm = hmmlearn.hmm.MultinomialHMM(
    n_components=15, init_params='e', params='ste', verbose=True
)
hmm.startprob_ = np.array([0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0])
hmm.transmat_ = np.array([
    [0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.5],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0]
])
```

### 5. "bệnh nhân":

+ Phonemes: /b/ /ê/ /p/ /silent/ /ŋ/ /ə/ /n/

## + Metrics: Accuracy

```
hmm = hmmlearn.hmm.MultinomialHMM(
    n_components=21, init_params='e', params='ste', verbose=True
)
hmm.startprob_ = np.array([0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0])
hmm.transmat_ = np.array([
    [0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
    [0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
```

```

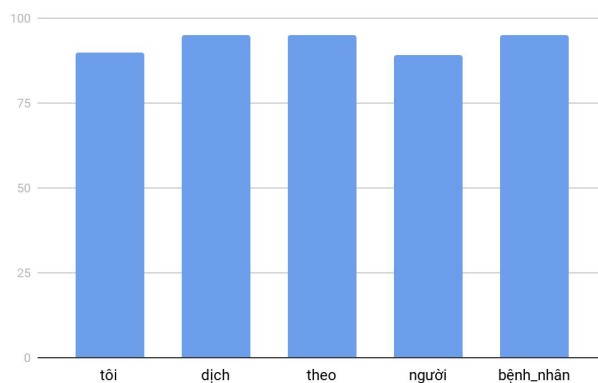
[0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3, 0.2,
0.0, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 0.3,
0.2, 0.0, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5,
0.3, 0.2, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.5, 0.3, 0.2, 0.0, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.5, 0.3, 0.2, 0.0],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.5, 0.3, 0.2],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.5, 0.5],
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 1.0],
] )

```

## RESULTS

With the test dataset:

Accuracy of HMM models on test dataset



With the recorded dataset:

Accuracy of HMM model on recorded dataset

