

BÁO CÁO BÀI TẬP MÔN XỬ LÝ TIẾNG NÓI

Thành viên:

- Hoàng Sơn Tùng – 17021351
- Nguyễn Công Tuấn Anh – 17021205

Model:

- Từ **“bệnh nhân”**
 - Phân tích âm vị:

|b|~|e|~|nh|~|silent|~|nh|~|a|~
 - Bộ tham số huấn luyện: (States: 21)

[illegible]

Data:

- File Wav của 5 từ: “bệnh nhân”, “có thể”, “hai”, “người”, “tôi”
- Mỗi từ có khoảng 60-80 file, test độc lập khoảng 15 file.

Thư viện:

- Sử dụng thư viện MultinomialHMM của HMMlearn

Từ “có thể”

- Phân tích âm vị: `|c|~|o|~|silent|~|t|~|h|~|e|`
- Bộ tham số huấn luyện: (States: 21)

[illegible]

- Từ “hai”:

- Phân tích âm vị: $|h|\sim|a|\sim|i|$
- Bộ tham số huấn luyện: (States: 9)

```
if cname == "hai":
    hmm = hmmlearn.hmm.MultinomialHMM(
        n_components=9, init_params='e', params='ste', verbose=True, n_iter= 1000
    )
    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]
    ])
}
```

- Từ “người”:

- Phân tích âm vị: $|ng|\sim|uo|\sim|i|$
- Bộ tham số huấn luyện: (States: 15)

```
if cname == "người":
    hmm = hmmlearn.hmm.MultinomialHMM(
        n_components=15, init_params='e', params='ste', verbose=True, n_iter= 1000
    )
    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, 0.5, 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, 1.0]
    ])
}
```

- Từ “tôi”:

- Phân tích âm vị: $|t|\sim|o|\sim|i|$
- Bộ tham số huấn luyện: (States: 9)

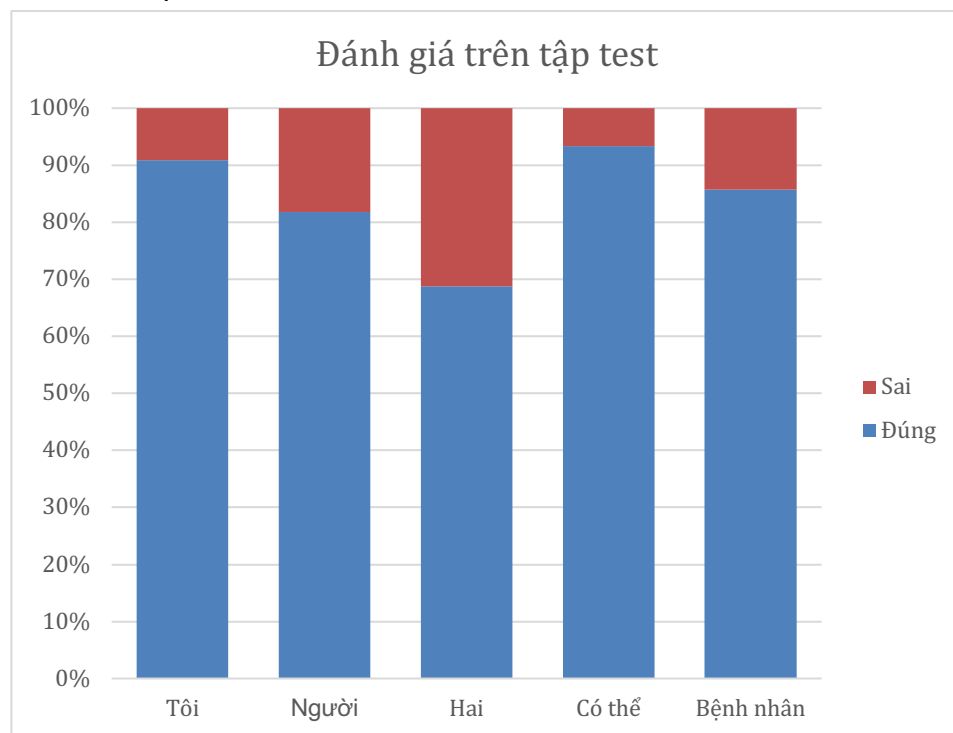
```
if cname == "tôi":
    hmm = hmmlearn.hmm.MultinomialHMM(
        n_components=9, init_params='e', params='ste', verbose=True, n_iter= 1000
    )
    hmm.startprob_ = np.array([0.6, 0.3, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0])
    hmm.transmat_ = np.array([
        [0.6, 0.3, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0],
        [0.0, 0.6, 0.3, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0],
        [0.0, 0.0, 0.6, 0.3, 0.1, 0.0, 0.0, 0.0, 0.0],
        [0.0, 0.0, 0.0, 0.6, 0.3, 0.1, 0.0, 0.0, 0.0],
        [0.0, 0.0, 0.0, 0.0, 0.6, 0.3, 0.1, 0.0, 0.0],
        [0.0, 0.0, 0.0, 0.0, 0.0, 0.6, 0.3, 0.1, 0.0],
        [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.6, 0.3, 0.1],
        [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.6, 0.4],
        [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0]
    ])
}
```

Ví dụ về bộ tham số sau khi huấn luyện của từ *tôi* (các kết quả khác có thể hiện trong notebook “notebook.ipynb”)

```
[8.37978009e-01 3.53232314e-02 1.26698760e-01 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00]
[0.00000000e+00 8.98223275e-01 8.87899264e-02 1.29867985e-02
 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00]
[0.00000000e+00 0.00000000e+00 6.70342840e-01 2.89264385e-02
 3.00730721e-01 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 8.79005765e-01
 1.77100826e-02 1.03284152e-01 0.00000000e+00 0.00000000e+00
 0.00000000e+00]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 7.68172352e-01 9.65195551e-02 1.35308092e-01 0.00000000e+00
 0.00000000e+00]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 8.70778993e-01 1.07312854e-01 2.19081529e-02
 0.00000000e+00]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 8.74661838e-01 1.24784337e-01
 5.53824355e-04]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 8.93191053e-01
 1.06808947e-01]
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 1.00000000e+00]
```

Kết quả:

- Trên tập test:

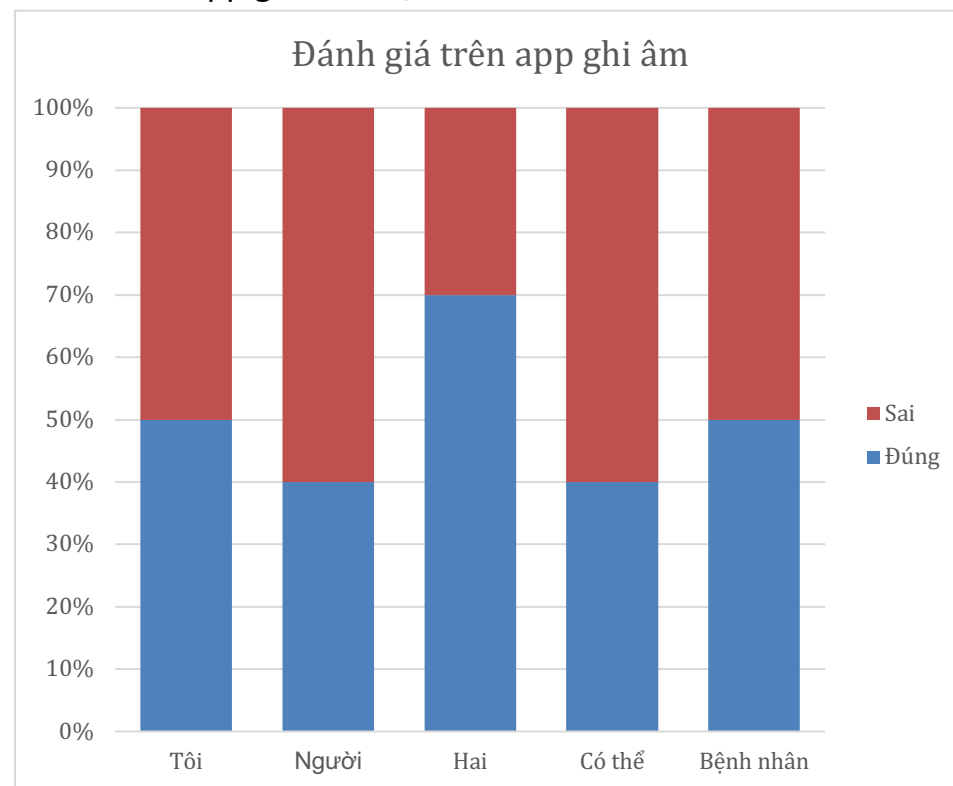


- Nhận xét: Số mẫu thử dao động 12-15 mẫu, độ chính xác tương đối cao (>80-90%) tuy nhiên có dấu hiệu overfit do thử với app thực tế cho kết quả ngược lại

Code:

- <https://github.com/megatunger/voice-class>
- Demo video app: <https://youtu.be/BR1mSaxDi5M>

- Trên app ghi âm thực tế:



Nhận xét: Số mẫu thử mỗi từ là 10 mẫu, môi trường có tạp âm. Độ chính xác thu được (~50-60%) kém hơn so với tập test.

