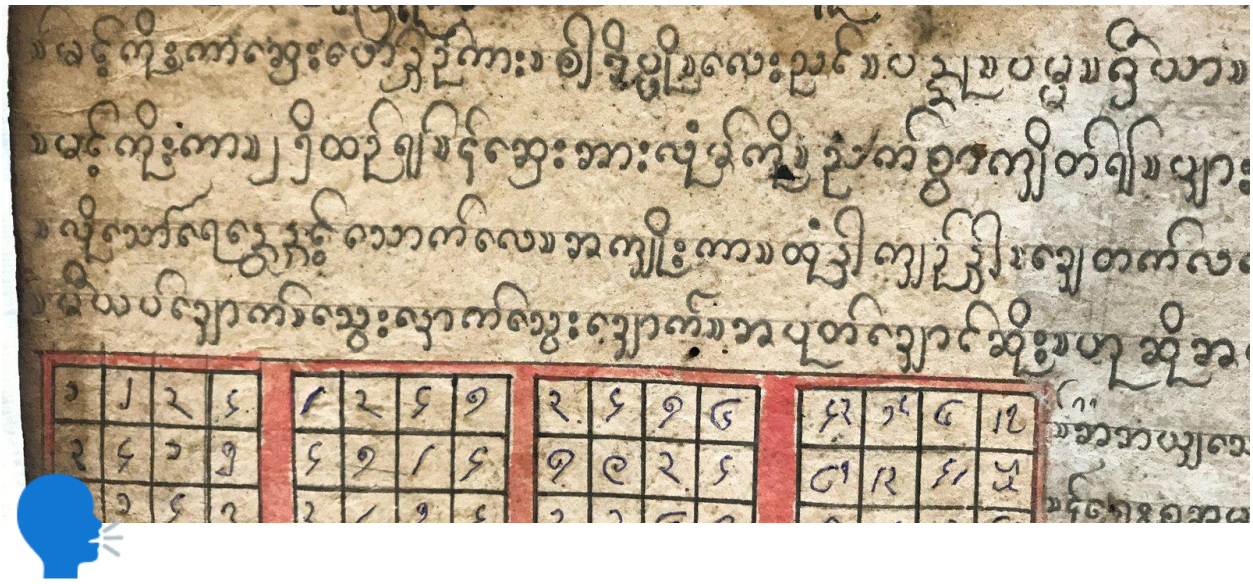


Tutorial for Burmese G2P Converter

CRFSuite Installation

> CRFsuite - A fast implementation of Conditional Random Fields (CRFs)
> Naoaki Okazaki's website
> https://www.chokkan.org/software/crfsuite/

1. Install https://github.com/chokkan/liblbfgs

```
$./autogen.sh
$./configure
$ sudo make
$ sudo make install
```

2. Install https://github.com/chokkan/crfsuite.git
    a. `$./autogen.sh`
        ▼ If you got this error, install automake and libtoolize: `$sudo apt-get install automake`, `$sudo apt-get install libtool`

```
ibtoolize: error: One of these is required:
libtoolize: gm4 gnum4 m4
libtoolize: error: Please install GNU M4, or 'export M4
./autogen.sh: 20: ./autogen.sh: aclocal: not found
aclocal failed!
```

    b. Run as usual

```
$ ./configure
$ sudo make
$ sudo make install
$ ./configure --with- liblbfgs=/usr/local/lib/liblbfgs-1.1
```

        ▼ After installation, when you run `crfsuite –help`, if you got the following error, you need to update `ld.so.conf` file: `$sudo vi /etc/ld.so.conf` and add this line `LIBDIR=/usr/local/lib/`

```
crfsuite: error while loading shared libraries: libcrfs
```

Training

1. Change the current directory into either `ver1` or `ver2`.
2. Create preprocessed training and testing data by running

```
$ ../scripts/preprocess.sh
```

3. Start training a model. Training time depends on your data size, no. of features and learning options. For the case of provided model `g2p.model`, it takes approximately 12 hours.

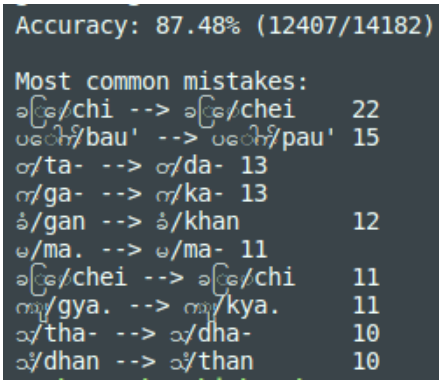
```
$ crfsuite learn -m ./g2p.model -e2 ./train ./test
```

Evaluating

Have the current directory into either `ver1` or `ver2`. There're three ways to evaluate your model.

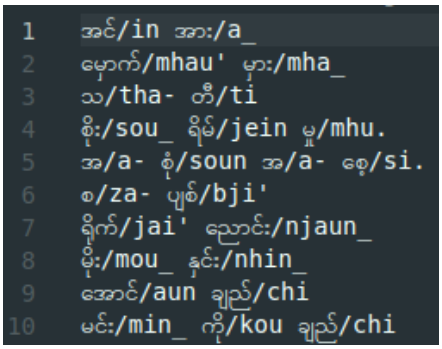
1. Using testing data, observe accuracy and most common mistakes.

```
$ ../scripts/evaluate.sh -e g2p.mo
```



2. Observe the predictions for testing data.

```
$ ../scripts/evaluate.sh -p g2p.mo
```



3. Interactively input a custom sentence and observe the prediction.

```
# require the model's name to be 'g2p.model'
$ ../scripts/predict.py
```

```
Enter a Burmese sentence: သဘာဝပတ်ဝန်းကျင်၏ ဆန့်ကျင်ဘက်အရာသည် လုပ်ယူဖန်တီးထားသော ပတ်ဝန်းကျင်ဖြစ်သည်
Phonemes of the sentence: dha- ba wa. bjo_ wun_ kyin gyin za- kyin be' a- ja dhi lou' ju hpan ti_ hta_ dho_ ba- wun_ dha- din_ ga_
----------------------------------------
```