Homework 7

Team 7 : 김성원, 김희수

**Train Data / Valid Data / Test Data**

* 제공받은 Data : wsj0(4195개) / wsj1(37167개)
* Training Data : wsj0 data 중 3000개 (+ wsj1 data 중 일부에 대하여 추가 학습 진행)
* Validation Data : wsj0 data 중 500개
* Test Data : wsj0 data 중 695개

**Input Normalization**

* X

**RNN Model**

* Bidirectional LSTM 3층 / 각 layer의 hidden dimension : 400
* Layer Normalization / Ba, Jimmy Lei, Jamie Ryan Kiros, and Geoffrey E. Hinton. "Layer normalization." *arXiv preprint arXiv:1607.06450* (2016).
* Loss : CTC loss
* Regularization : X
* Batch Size : 40
* Optimizer : Adam Optimizer
* Gradient Clipping – Maximum gradient norm : 5
* 초기 Learning Rate : 0.01

**Language Model**

* KenLM: 독자적인 discount와 smoothing algorithm을 이용하는n-gram lm 프레임워크이다. 원래는 word level lm을 위한 프레임워크지만 데이터셋을 character level로 잘라넣으면 char level lm을 만들어 준다. 실제 likelihood를 계산을 위해 python interface를 제공하기 때문에 tensorflow 프레임워크에 embed 하기 용이했다.
* 실험에는 5-gram CLM을 사용했고 lm을 만들기 위한 데이터셋은 제공된 transcription만을 사용했다.

**Decoding**

* Greedy Decoding : 가장 확률이 높은 sequence로 decoding하는 방식 (language model 고려 X)
* 프레임워크는 [Stanford-ctc](https://github.com/amaas/stanford-ctc/tree/master/ctc_fast/new_decoder)를 사용했고, 이 중 new\_decoder의 구현을 사용했다.
* Decoding with Character Level Language Model : Stanford ctc decoding 참조 (Maas, Andrew L., et al. "Lexicon-Free Conversational Speech Recognition with Neural Networks." *HLT-NAACL*. 2015.)
* Beam width : 10
* Level Language Model : Kenlm (5-gram CLM)
* alpha: 1.25, beta: 1.5

**Codes**

* main.py : Hyperparameter 설정 / Data, Model loading / train & evaluate
* model/model.py : CTC-RNN Model
* data\_processing.py : 각종 parsing / CER, WER 연산
* decoding.py, decoder.c, decoder.pyx, Makefile, Setup.py, wsj\_5gram.binary, build, \_\_pycache\_\_, \_\_init\_\_.py, char\_set\_reverse.txt : LM Decoding에 필요한 codes
* train.py : training
* evaluate.py : evaluation
* params/BiLSTM3\_ln\_ADAM\_400.data-00000-of-00001, BiLSTM3\_ln\_ADAM\_400.index, BiLSTM3\_ln\_ADAM\_400.meta : model

**Results**

**1) Qualitative Analysis**

Decoding w/o lm – Greedy Decoding

Decoding with lm – Decoding with CLM

\* Case 1 (두 가지 decoding 모두 잘 맞추는 경우)

|  |
| --- |
| **Decoding w/o lm** : UNUSUALLY HIGH LEVELS OF RADIATION WERE DETECTED IN MANY EUROPEAN CUNTRIES  **Decoding w/ lm** : UNUSUALLY HIGH LEVELS OF RADIATION WERE DETECTED IN MANY EUROPEAN COUNTRIES  **Transcription** : UNUSUALLY HIGH LEVELS OF RADIATION WERE DETECTED IN MANY EUROPEAN COUNTRIES |
| **Decoding w/o lm** : THE POPULATION LIVES BY HERDING GOATS AND SHEEP OR BY TRADING  **Decoding w/ lm** : THE POPULATION LIVES BY HERDING GOATS AND SHEEP OR BY TRADING  **Transcription** : THE POPULATION LIVES BY HERDING GOATS AND SHEEP OR BY TRADING |

\* Case 2 (Greedy Decoding은 틀리는 경우)

|  |
| --- |
| **Decoding w/o lm** : SWO NARROW GAUGE RAILROADS FROM CHINA ENTER THE CITY FROM THE NORTHEAST ANS NORTHWOST  **Decoding w/ lm** : TWO NARROW GAUGE RAILROADS FROM CHINA ENTER THE CITY FROM THE NORTHEAST AND NORTHWEST  **Transcription** : TWO NARROW GAUGE RAILROADS FROM CHINA ENTER THE CITY FROM THE NORTHEAST AND NORTHWEST |
| **Decoding w/o lm** : SOME MAPS USE BANDS OF COLOR TO INDICATE DIFFERENTANTORVALS AF VALU  **Decoding w/ lm** : SOME MAPS USE BANDS OF COLOR TO INDICATE DIFFERENT INTERVALS OF VALUE  **Transcription** : SOME MAPS USE BANDS OF COLOR TO INDICATE DIFFERENT INTERVALS OF VALUE |
| **Decoding w/o lm** : A TANKER IS A SHIPP DESIND TO CARRYE LARGE VOLUMES OF OIL OR OTHER LIQUIDD CARGO  **Decoding w/ lm** : A TANKER IS A SHIP DESIGNED TO CARRY LARGE VOLUMES OF OIL OR OTHER LIQUID CARGO  **Transcription** : A TANKER IS A SHIP DESIGNED TO CARRY LARGE VOLUMES OF OIL OR OTHER LIQUID CARGO |
| **Decoding w/o lm** : PRIVAETFREE SCHOOLS WORE FORMED BOTTH IN PORNEIGHBORHOODS AND IN MIDDLE CLASS COMMUNITIES  **Decoding w/ lm** : PRIVATE FREE SCHOOLS WERE FORMED BOTH IN POOR NEIGHBORHOODS AND IN MIDDLE CLASS COMMUNITIES  **Transcription** : PRIVATE FREE SCHOOLS WERE FORMED BOTH IN POOR NEIGHBORHOODS AND IN MIDDLE CLASS COMMUNITIES |
| **Decoding w/o lm** : THESE CHANGES AROUSED ORTHODOX OPPOSITION AN SOMETNOMS GOVERNLMENT INTRVENTION  **Decoding w/ lm** : THESE CHANGES AROUSED ORTHODOX OPPOSITION AND SOMETIMES GOVERNMENT INTERVENTION  **Transcription** : THESE CHANGES AROUSED ORTHODOX OPPOSITION AND SOMETIMES GOVERNMENT INTERVENTION |

\* Case 3 (둘 다 틀리는 경우)

|  |
| --- |
| **Decoding w/o lm** : BISMARCK SERVS HAS AT TRADE AND SHIPPING POINT FOR AN AREAOF LARGE MECHANIZED FARMS  **Decoding w/ lm**  : BISMARCK SERVES HAS A TRADE AND SHIPPING POINT FOR AN AREA OF LARGE MECHANIZED FARMS  **Transcription** : BISMARCK SERVES AS A TRADE AND SHIPPING POINT FOR AN AREA OF LARGE MECHANIZED FARMS |
| **Decoding w/o lm** : HE CLIPSES OF THE SUNND MOON HAVE LONG MADE A DEEP IMPRESSION ON HUMANKIND  **Decoding w/ lm** : HE CLIPSES OF THE SUN AND MOON HAVE LONG MADE A DEEP IMPRESSION ON HUMANKIND  **Transcription** : ECLIPSES OF THE SUN AND MOON HAVE LONG MADE A DEEP IMPRESSION ON HUMANKIND |

\* Case 4 (둘 다 꽤 틀리는 경우)

|  |
| --- |
| **Decoding w/o lm** : BUT JUST FOR WANTING TESHEAR THE PROSE IN COMS A BECAME AP PROD DARELET ROVOLUTION ARY AN THE ISAO SOME PEOPLE HE RECALS  **Decoding w/ lm**  : BUT JUST FOR WANTING THE SHEAR THE PROSE IN COLUMES THAT BECAME A PROD DERELLAT REVOLUTION ARE IN THE IS OF SOME PEOPLE HE RECALLS  **Transcription** : BUT JUST FOR WANTING TO HEAR THE PROS AND CONS I BECAME A PRO DERELICT REVOLUTIONARY IN THE EYES OF SOME PEOPLE HE RECALLS |
| **Decoding w/o lm** : THE BROADCASTING PROPERTIES OF LONG BEEN REGARDAT IS CHENCORP'SBESTE FENCEAGAINSTT HA HOSTILE TAKEOVER  **Decoding w/ lm** : THE BROADCASTING PROPERTIES OF LONG BEEN REGARDED AS THAN CORP'S BEST TO FENCE AGAINST A HOSTILE TAKEOVER  **Transcription** : THE BROADCASTING PROPERTIES HAVE LONG BEEN REGARDED AS GENCORP'S BEST DEFENSE AGAINST A HOSTILE TAKEOVER |
| **Decoding w/o lm** : ON COTATA A SIXTY TWO YEAROLD AGRICULTURAL WORKER FROM CEETHAUT WHONATES THOAA IS OUT SIDE FRIDIES AD WONE A N. LOOCKING FOR IS EMPLOYER  **Decoding w/ lm**  : ONE POTATOR A SIXTY TWO YEAR OLD AGRICULTURAL WORKER FROM SEE THOUT WHEN IT IS TO A HIS OUTSIDE FRITIES AND ONE A. M. LOOKING FOR ITS EMPLOYER  **Transcription** : JUAN CALDERA A SIXTY TWO YEAR OLD AGRICULTURAL WORKER FROM CIUDAD JUAREZ CHIHUAHUA IS OUTSIDE FREDDY'S AT ONE A. M. LOOKING FOR HIS EMPLOYER |

Case 1을 보면 일부 speech data의 경우에는 language model을 고려하지 않고 CTC training만 써서 transcription을 잘 맞추는 것을 확인할 수 있다. Case 2 / Case 3을 보면 greedy decoding하는 것 보다 decoding시 language model을 고려할 경우 transcription에 가깝게 잘 맞추는 경향을 볼 수 있다. Case 4를 보면 고유 명사나 긴 phrase들(PROS AND CONS / PRO DERELICT REVOLUTIONARY / GENCROP’S / JUAN CALDERA / CIUDAD JUAREZ CHIHUAHUA / FREDDY)의 경우 모두 많이 틀리는 것을 확인 할 수가 있다. 5-gram CLM보다 더 긴 context를 담을 수 있는 language model (7-gram CLM / RNNLM)을 사용하면 더 좋은 성능을 낼 수 있을 것이다.

**2) Quantitative Analysis**

|  |  |
| --- | --- |
| **CTC loss** | 38.4510 |
| **CER w/o LM** | 12.45% |
| **WER w/o LM** | 44.43% |
| **CER w/ LM** | 6.28% |
| **WER w/ LM** | 18.16% |

CER은 language model을 포함한 decoding으로부터 6% 상승 효과를 얻었고, WER의 경우 26%나 상승효과가 있었다.

**Evaluation method**

Kenlm을 사용하기 위해서는,

pip install https://github.com/kpu/kenlm/archive/master.zip

를 통해 kenlm python interface를 설치하면 된다.

그리고 수정된 Stanford-ctc decoder를 사용하기 위해서는 첨부된 setup.py를 아래와 같이 수행하면 된다.

python setup.py install

그러면 “decoder” package가 설치된다.

main.py와 test\_all.list, test\_all.trans, wsj0, wsj1의 위치를 같게 하고 python main.py 로 실행시키면 최종적으로 결과를 확인할 수 있고, decoded.txt파일에 각 feat 파일에 대한 greedy decoding / lm decoding / transcription 순서대로 결과를 확인할 수 있다.