

# 今週の進捗

知能数理研究室 12056 外山 洋太

# 新しいモデルの実装

## GW 以前のモデル (char2doc)

- ▶ 文字の embeddings から直接文書の embeddings を作る
- ▶ sequence の長さ  $\sim 1024 \rightarrow$  辛い

## IMDb データセット

- ▶ average sents / doc : 12.33
- ▶ average words / sent : 21.45
- ▶ average chars / word : 4.02
  - ▶ 英語平均単語長 : 約 5.1 (<http://www.wolframalpha.com/input/?i=average+english+word+length>)
  - ▶ ' ', '!', '?' の記号
  - ▶ They'll の "ll"

# 新しいモデルの実装

## char2word2sent2doc by someone, NAACL 2016

- ▶ 三輪先生から貰った論文が元
- ▶ 元論文は word2sent2doc だった
  - ▶ 単語の embeddings から文，文書の embeddings を順に Bi-directional atttentioned GRU RNN で作成
- ▶ 同じ embedding 生成の方法
  - ▶ Attention 付き
  - ▶ Bi-directional
- ▶ パラメータは主に文字の embeddings と GRU，分類のための全結合層
- ▶ char2word, word2sent, sent2doc の RNN は全て同じ実装
  - ▶ パラメータは別で持つ
  - ▶ 実験ではハイパーパラメータはほとんど一緒

# 新しいモデルの実装

## ハイパーパラメータ

- ▶ embedding サイズ（一方向）
  - ▶ 文字 : 32
  - ▶ 単語 : 32
  - ▶ 文 : 32
  - ▶ 文書 : 32
- ▶ 隠れ層の数 : 1
- ▶ 隠れ層ニューロン数 : 64
- ▶ (出力層ニューロン数 : 1)
- ▶ ドロップアウト 率 : 0.6
- ▶ L2 正則化係数 :  $1e-6$
- ▶ context vector サイズ : 32

# 新しいモデルの実装

## その他実験設定

- ▶ 学習回数 : 1024
- ▶ バッチサイズ : 1024 (今後減らします.)
- ▶ 訓練データサイズ : 25'000
- ▶ 評価データサイズ : 25'000
  - ▶ 各々 positive, negative が 12'500 ずつ

# 新しいモデルの実装

## 実験結果

実験が終わっていません

# TensorFlow と Docker と CUDA

## TensorFlow の現在

- ▶ Linux の公式サポートは Ubuntu 14.04 LTS のみ
- ▶ CPU 版は Fedora でも動く
- ▶ GPU 版は Ubuntu と Fedora でディレクトリ構造が少し違うのでそのままでは動かない
  - ⇒ 他の方法
    - ▶ ☉ 公式の Docker イメージ
    - ▶ chroot?

# TensorFlow と Docker と CUDA

## 公式の Docker イメージ

- ▶ ホスト 側 Linux の公式サポート は Ubuntu 14.04 LTS のみ
- ▶ Github のリンク (<https://github.com/tensorflow/tensorflow/tree/master/tensorflow/tools/docker>)
- ▶ とても 速い (TITAN X × 3 で 100 iterations / hour 以上)
  - ▶ CPU × 16 (hyperthreading) で半日 100 iterations くらい