

みんな **Chainer** で 何やってるの？

～コミュニティのユースケース探索～
Chainer Meetup #7, 9th Jun. 2018

Who am I?

- 是枝祐太
- 某電機会社リサーチャー
- 研究歴
 - ～ 2015: 医療+ロボット(大学)
 - ～ 2016: ロボット+応用機械学習
 - ～現在: 応用機械学習+自然言語処理



koreyou



@koreyou_

LT発表枠

発表時間:1人5分

Chainerの活用事例等について発表していただける方を募集しています。

タイトルをフィードに記載してください(まだ未定の場合は、あとでご連絡ください) **発表して頂いた方はChainerのTシャツをお配りします!** (コメント欄やDMでサイズをお送りしていただけるとピッタリのサイズをご用意できます)

本日はコミュニティのユーザが
どんなものを、どのように開発して
いるのか検証してみました

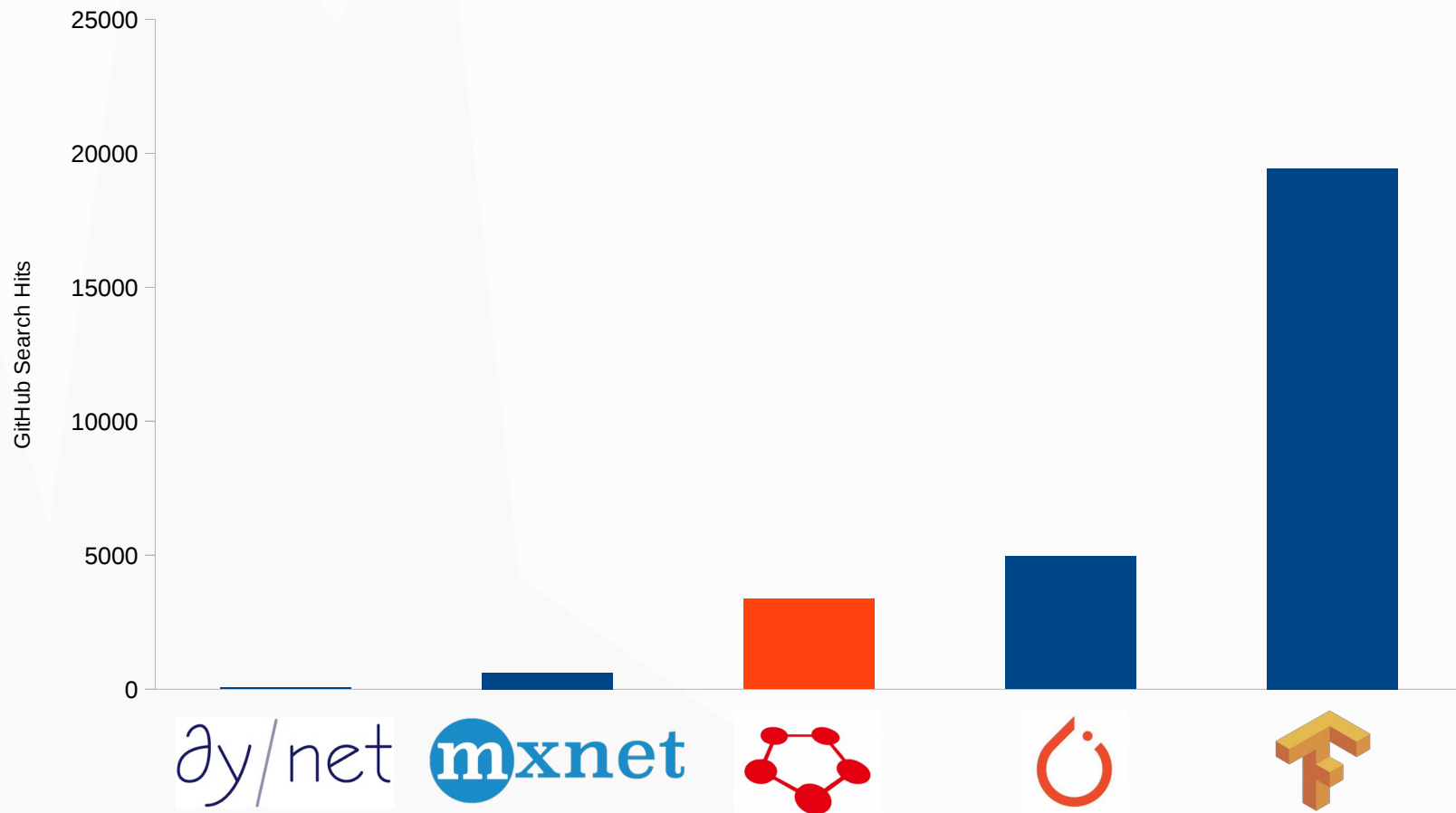
Method

- **GitHub API** を使い **chainer** を使っているレポジトリを検索
- あとは全プロジェクト **clone!**
 - 特にリミットはないらしいし、**sleep** 挟んでるから大丈夫(たぶん)
- 正規表現ベースでマッチング

こんなことがわかりました

(まずはどんなものを開発しているのか)

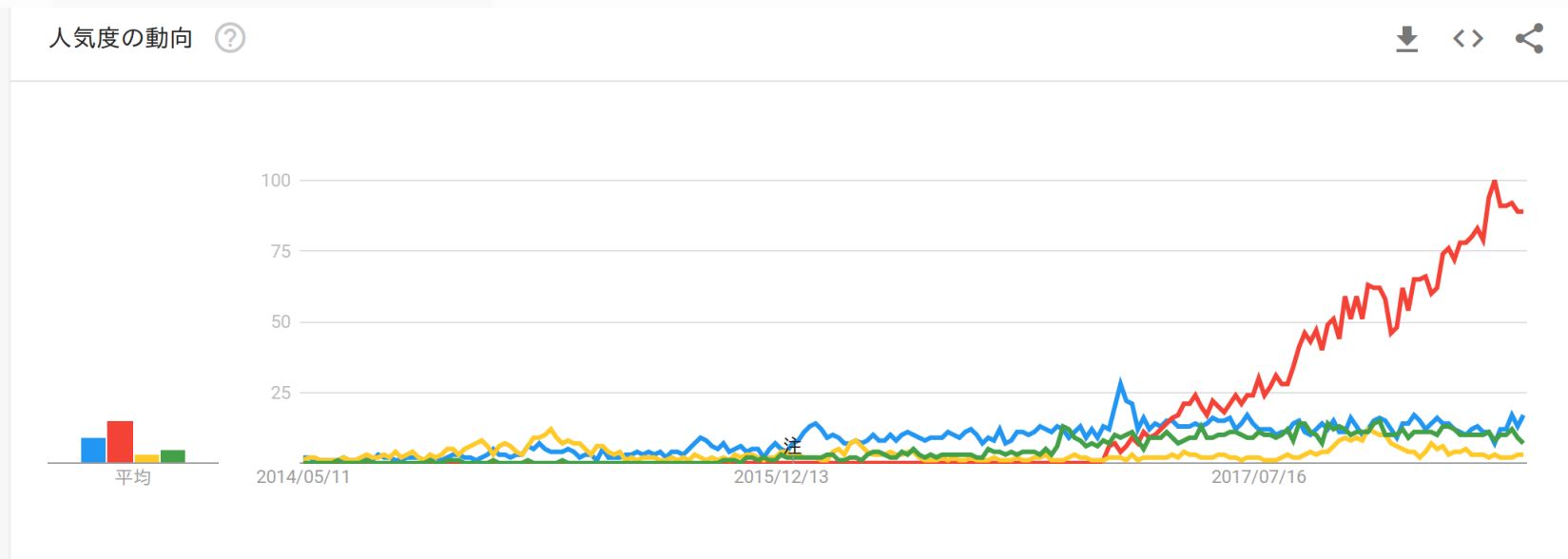
GitHub に Chainer を使ったレポジトリは 約 3000 件ある



検索ヒット数から /Chainer をインポートしていないプロジェクトは **4%** だったので無視できる

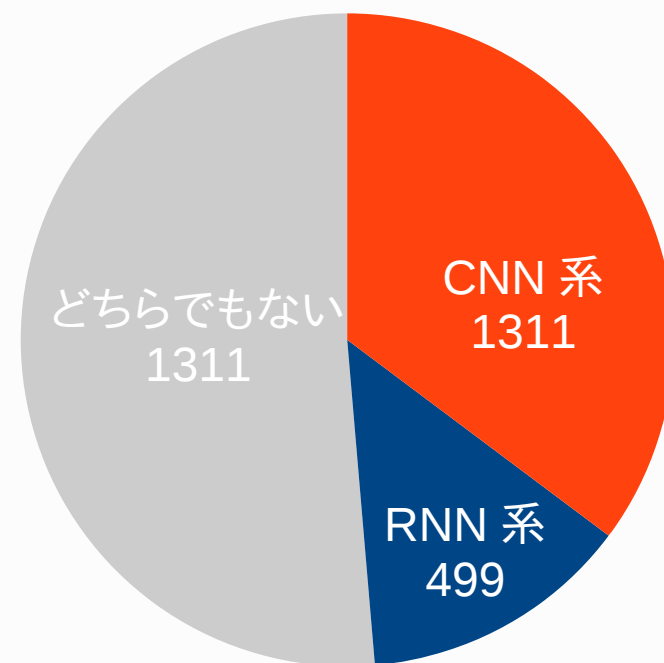
別に PyTorch に負けていない

- PyTorch の検索数が取り沙汰されるがプロジェクト数は同レベル
- **Awesome PyTorch** の論文再現を 10 件無作為抽出すると 50% は Chainer にも同じ論文の実装あり



画像系ユーザが多い

- 画像：言語ユーザの比 7:3 くらい
- どちらでもないの中にはツール等も含まれている



- CNN を使っていたら画像系、RNN を使っていたら RNN
- links への正規表現マッチングで抽出

関数名を分解し、頻度を分析してみる

NLP

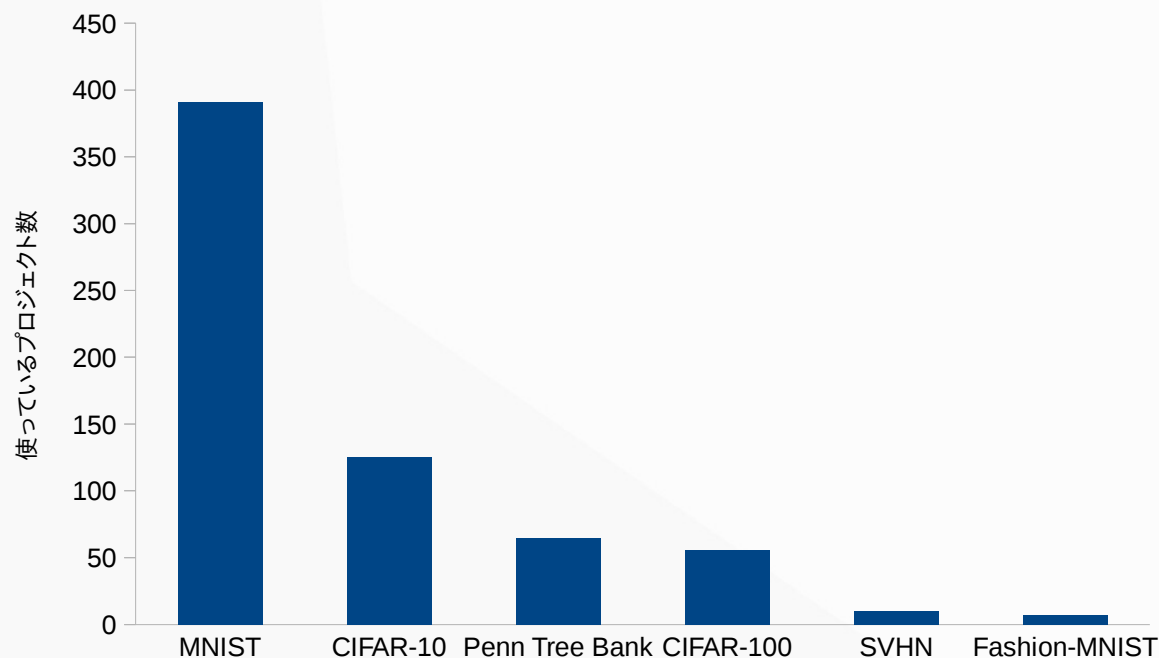
GAN

RL

General

例題をやったあと何をしたいのかわからないユーザーが多いのでは

- MNIST を活用しているユーザーが多い
- Chainer を活用したレポジトリ保有者の 77% は 1 つだけしか Chainer レポジトリがない



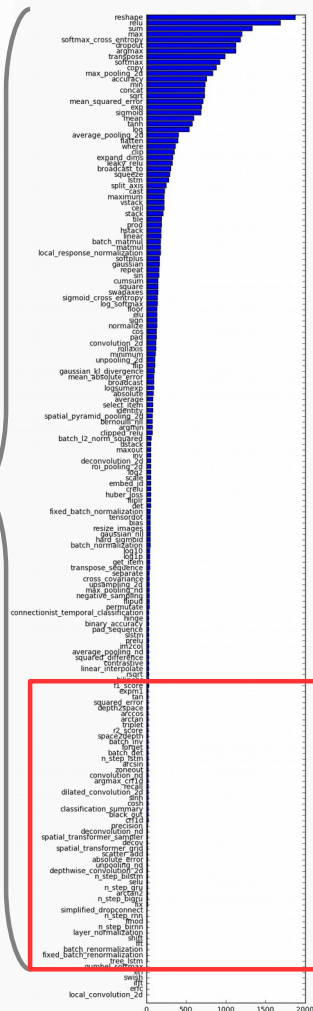
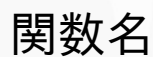
その他の雑感

- **PyTorch** とはコミュニティの色が違いそう
 - **PyTorch** は「硬い」レポジトリが多い気がする
 - **Chainer** は「たーのしー」レポジトリが多い気がする
- 「たーのしー」そうなレポジトリの例
 - “It can calculate the degrees of moe from your pictures.”
 - “ディープラーニングの力で結月ゆかりの声になるレポジトリ”
 - “eminem, snoop dogg classifier”
 - “ストリートファイター V における画像内のリュウの技を画像ポーズ認識を使い判断する”
- 日本では実業よりなレポジトリを公開しない傾向？
 - **Chainer** の最大コミュニティは日本

どのように開発しているのか

Chainer 組み込みの関数のほとんどは めったに使われない

- どんどん新しい関数がたされているので使わなければもったいない

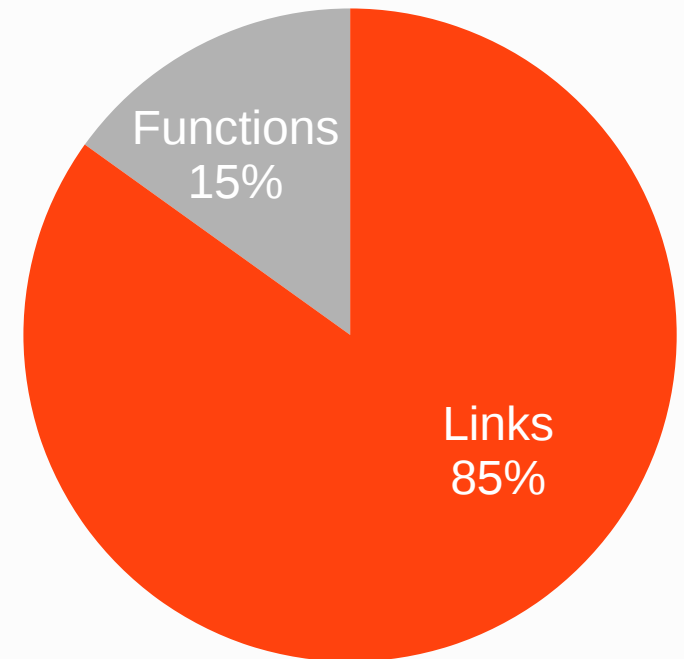


最近実装されたものや、便利関数など宝の山

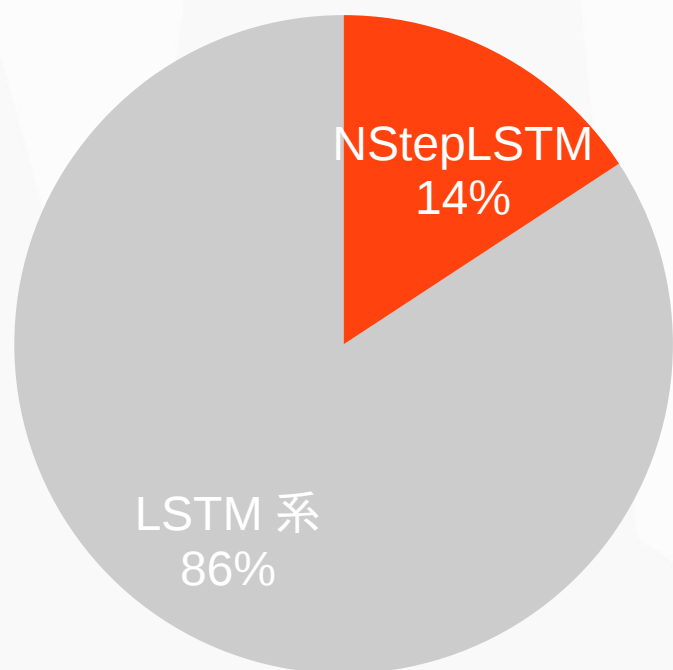
links を選ぶのが ベストプラクティス (そう)

- **links/functions** 両方の実装がある関数において、約 **85%** のプロジェクトは **links** 側を採用

links	functions
Convolution2D	convolution_2d
NStepGRU	n_step_gru
LayerNormalization	layer_normalization



NStepLSTM 系を使っている人が少ない

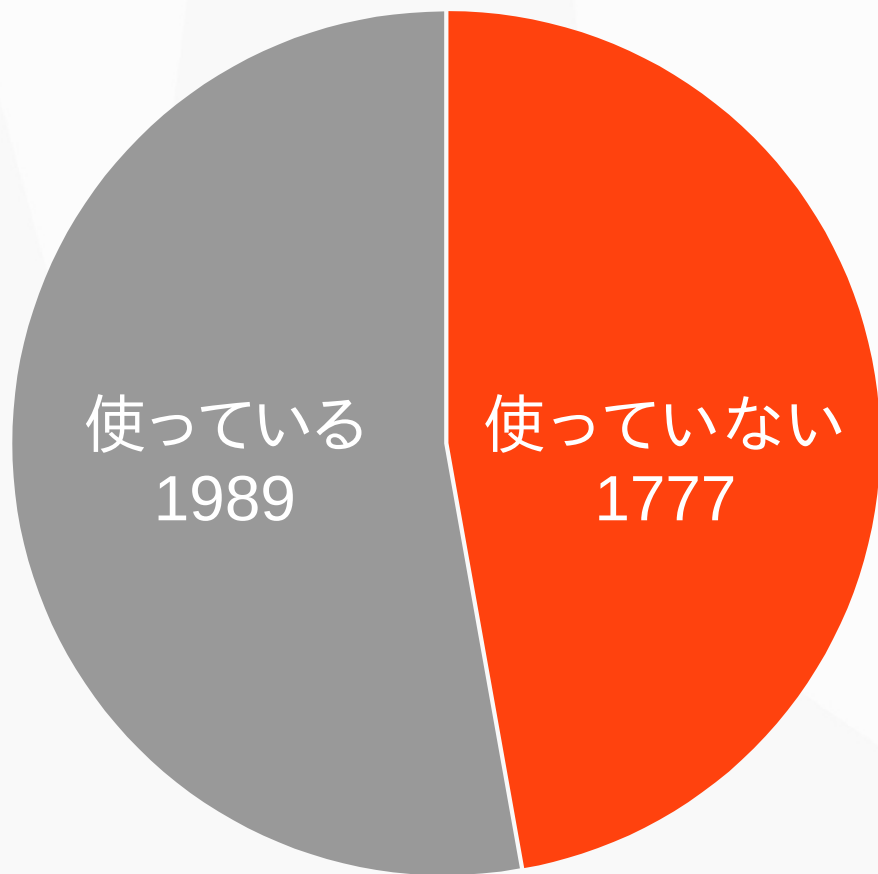


- 単なる文分類などであれば
NStepLSTM 系を使ったほうが良いはず
- どちらを使えばよいのか迷うので指針があると親切

links	LSTM 系	NStepLSTM 系
可変長への対応	△	○
Deep 化	△	○
両方向化	△	○
実装の単純さ	○	×

約半数のプロジェクトは Trainer を使っていない

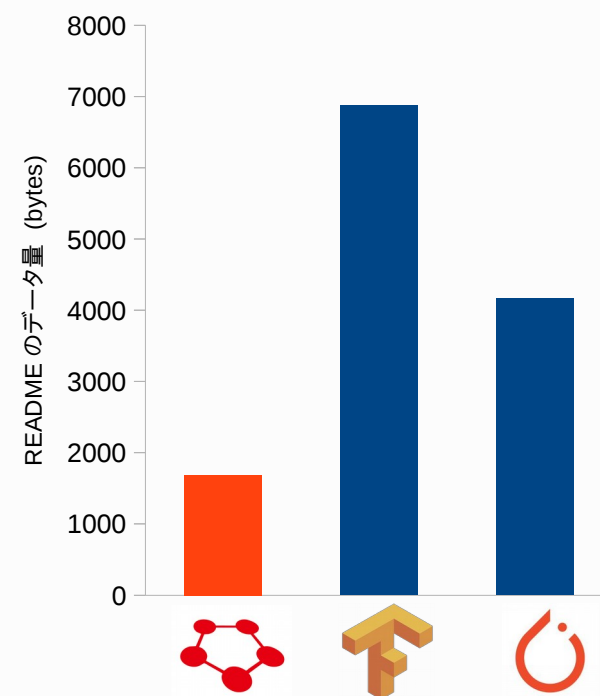
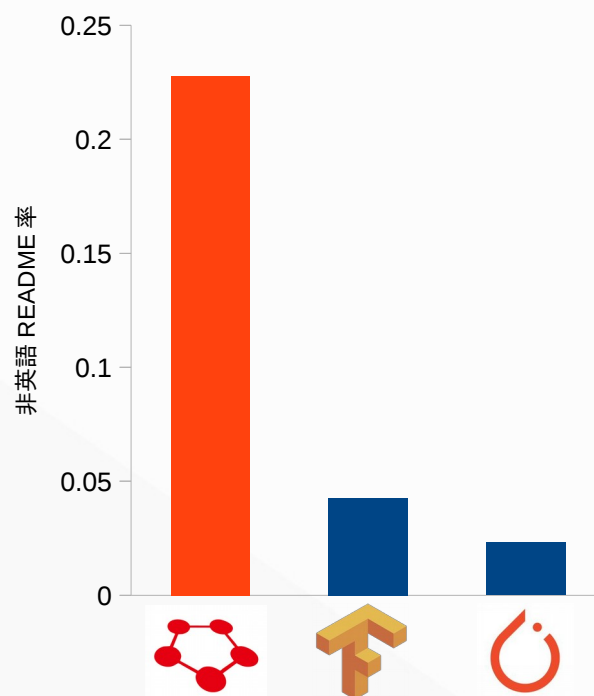
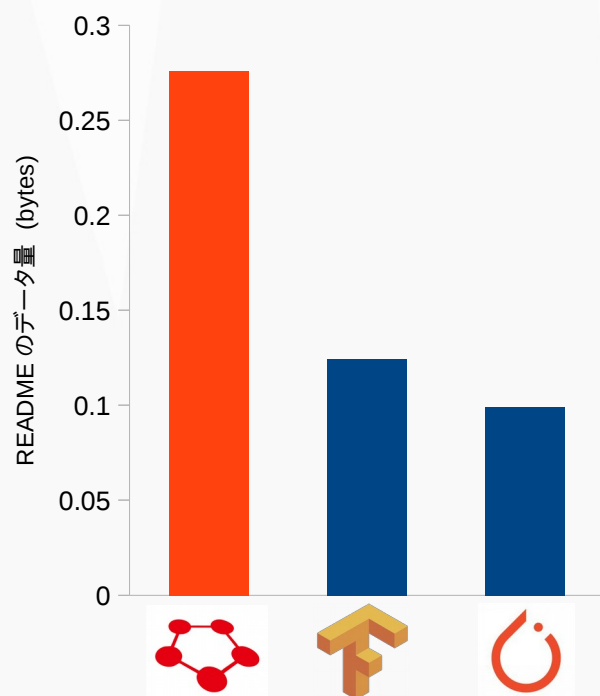
- **Trainer** 使うとフレームワークの恩恵を受けやすい!



どのように公開しているのか

コミュニティ活性化のためにも ドキュメンテーションをがんばりたい

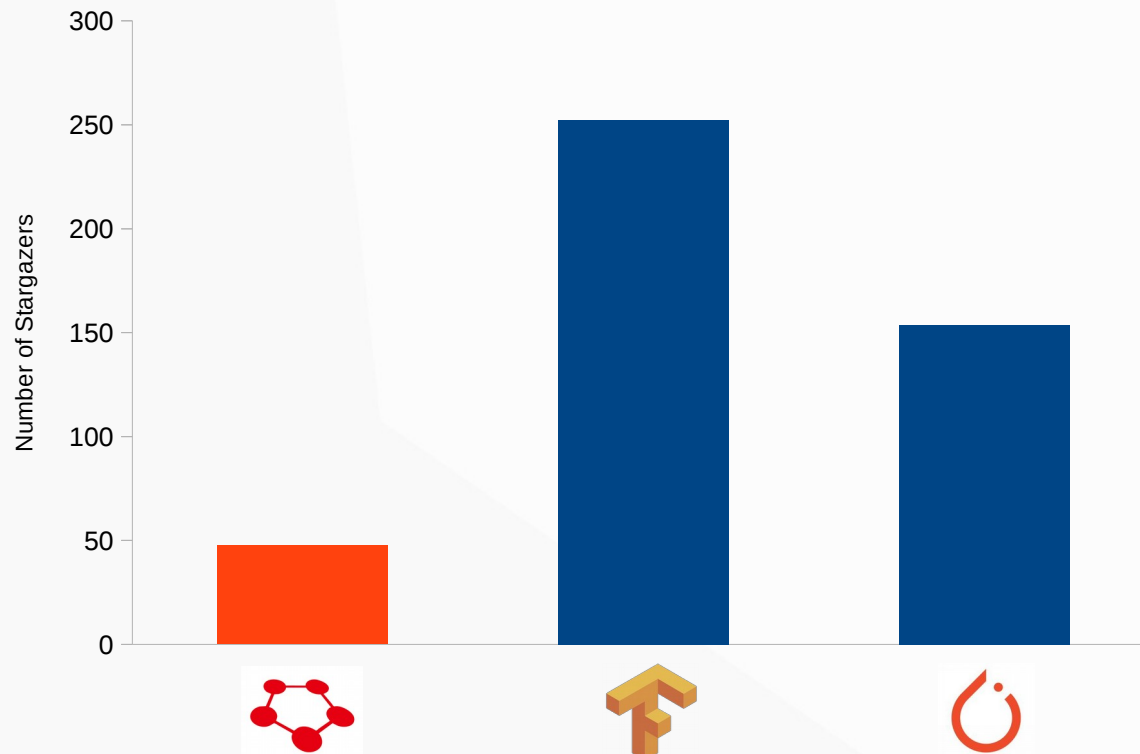
- README の整備具合は TensorFlow/PyTorch に大きく差をあけられている



※TensorFlow と PyTorch は 100 ~ 200 サンプル

みんなもっと星をつけよう

- 日本人は星をつけたがらないのか？(自分もほとんどつけたことない)



※TensorFlow と PyTorch は 100 ～ 200 サンプル

まとめ

- **Chainer** ユーザは画像、言語、強化学習などなど多様
- 新規ユーザ獲得、定着のために(コミュニティ)レポジトリのドキュメンテーション拡充が急務
- 現在の利用状況から見た知見
 - **Chainer** の新機能を活用できていないのがもったいない
 - 特に理由がなければ **functions > links**
 - **NStepLSTM** はもっと使われても良い
 - **Trainers** をもっと使おう

Thank you for your attention!