Mapping Reported Term for the Adverse Event into MedDRA Using Deep Learning

Yoshihiro Nakashima
Standardization and Management Group, Data Science, Development
Astellas Pharma Inc.
September 4, 2018

Agenda

Introduction

Introduction

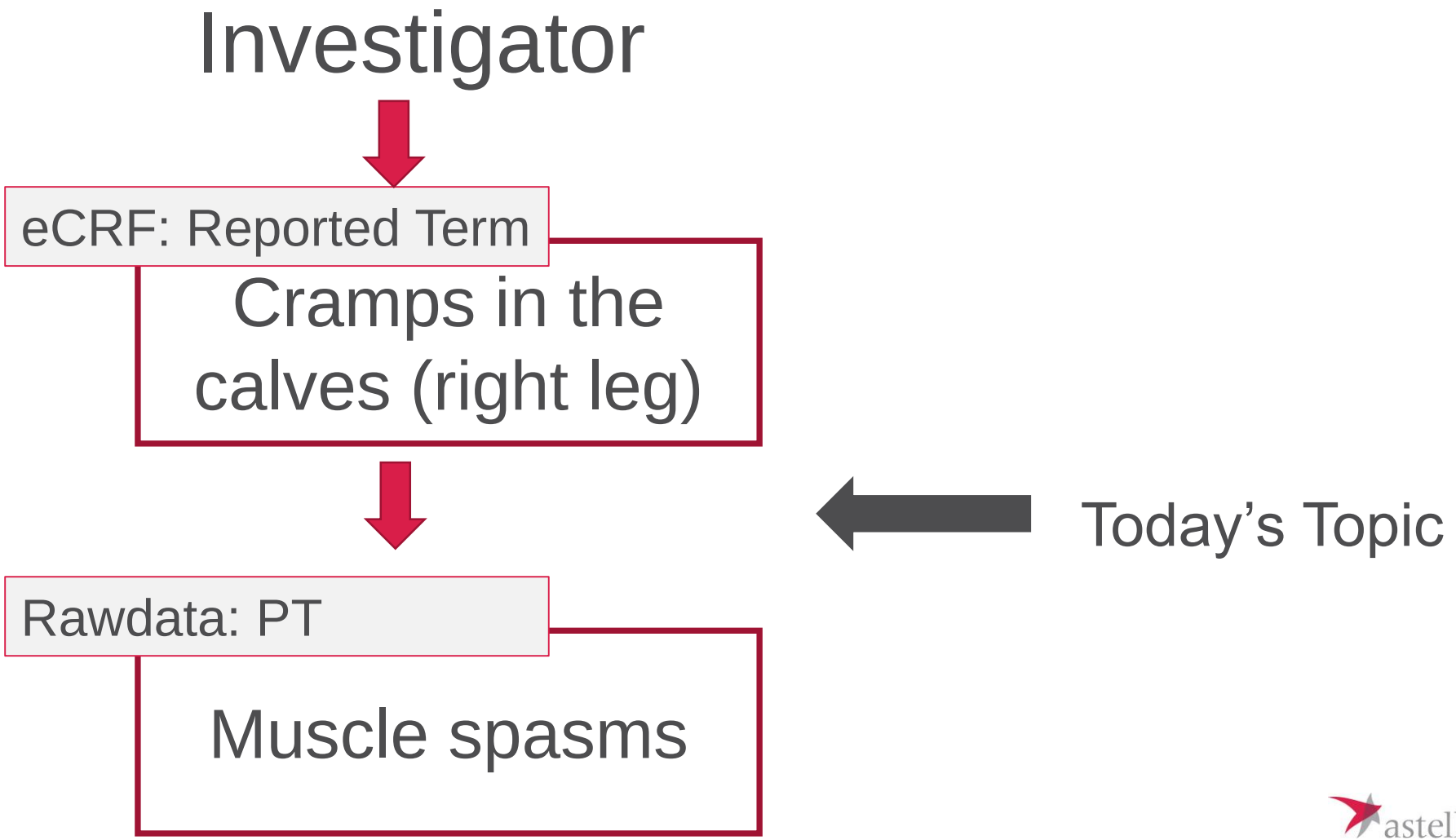
Today, it is said that artificial intelligence (AI) is the third boom, and not a day passed without hearing AI. This boom is led by deep neural network (DNN) using technique of deep learning. DNN made tasks (e.g. image recognition, natural language processing, speech recognition) more advanced than ever.

In this presentation, I will implement deep learning to map reported term for adverse event into MedDRA. The procedure consists of two steps. First, converts each word in reported term into numeric vector produced by word2vec using Wikipedia data. Second, trains DNN using words represented by numeric vectors. Through this implementation, I would like to examine the applicability of AI technology to clinical trials.

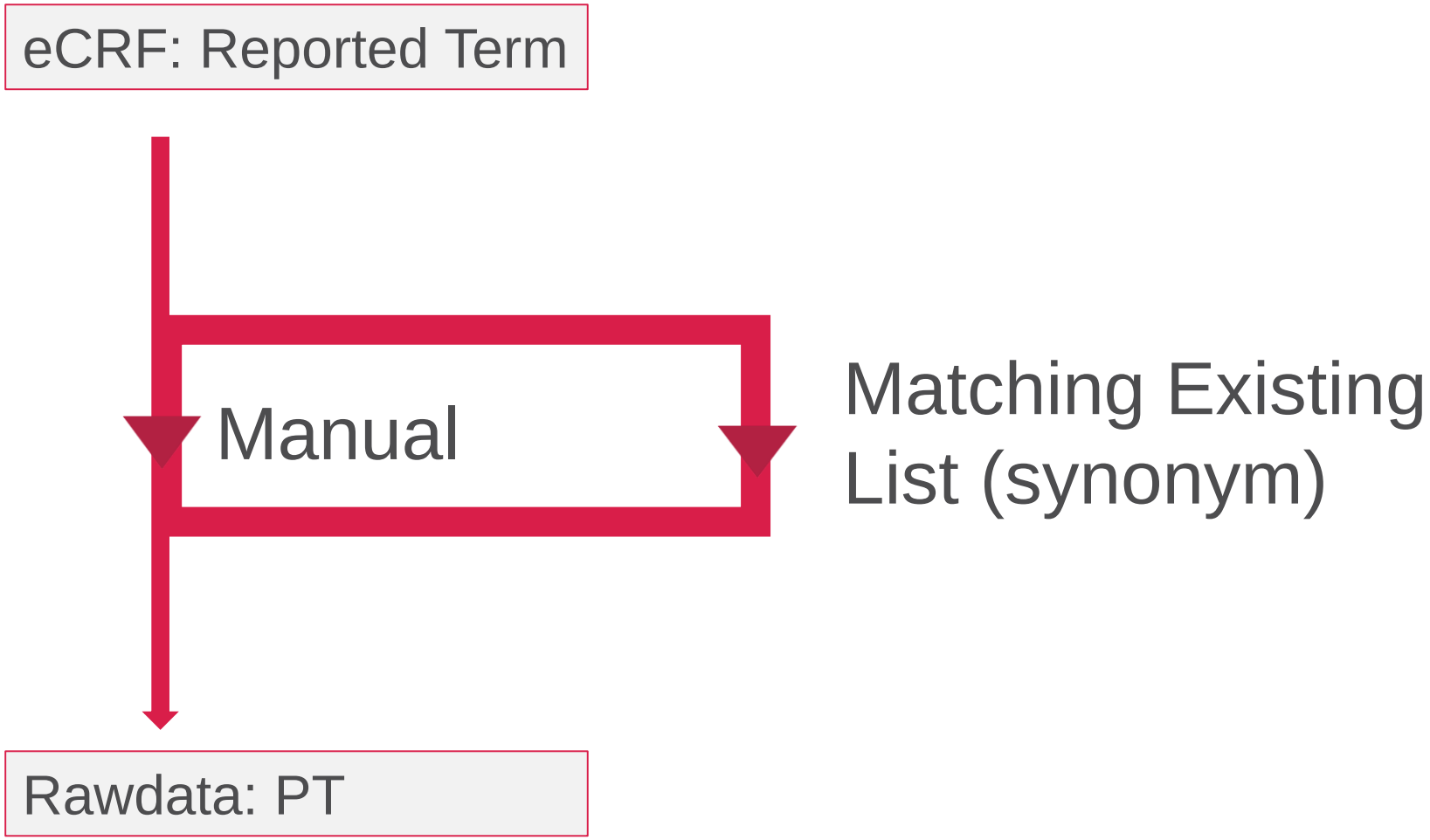
MedDRA Mapping Process

astellas

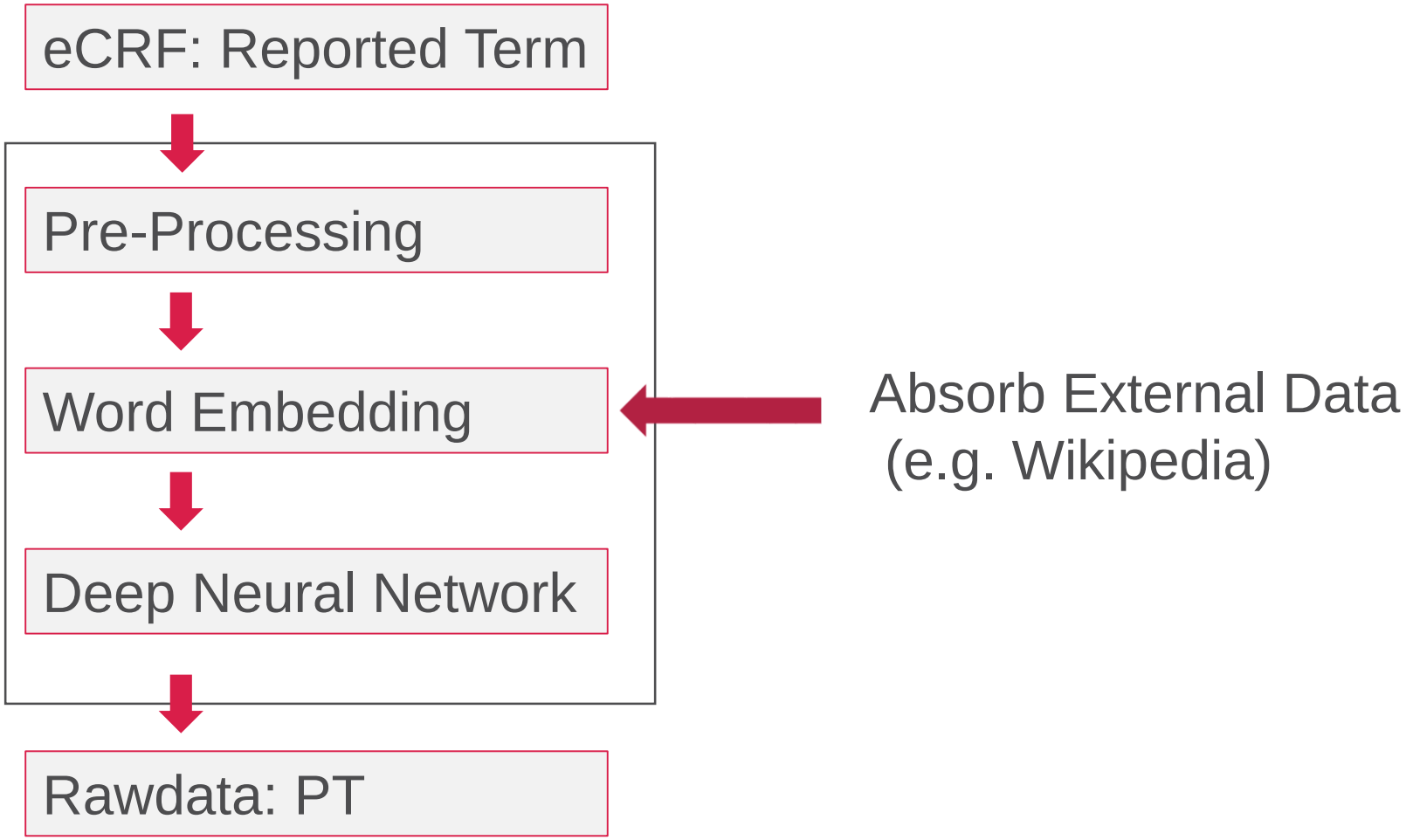
MedDRA Mapping Process: Over View



MedDRA Mapping Process: Basic



MedDRA Mapping Process: Today's Topic



Word Segmentation

Original: “Cramps in the calves (right leg)”

-> Replace text: “cramps in the calves right leg ”

-> Separate by space: “cramps”, “in” “the”, ”calves”, “right”, “leg”

Drop Stop-words

“cramps”, “in” “the”, ”calves”, “right”, “leg”

-> “cramps”, ”calves”, “right”, “leg”

(#words = 4)

Vectorize the word using Word2Vec to "calculate" words (i.e. Plus, Minus, Distance)

Dictionary	Vocabulary	Embeddable Words in Reported Term / All Words
Wikipedia*1	2.2 million	2300 / 2600
Google News*2	3 million	2000 / 2600

*1: https://dumps.wikimedia.org/enwiki/latest/
*2: https://code.google.com/archive/p/word2vec/

Structure of Neural Network

Simple Neural Network (NN)

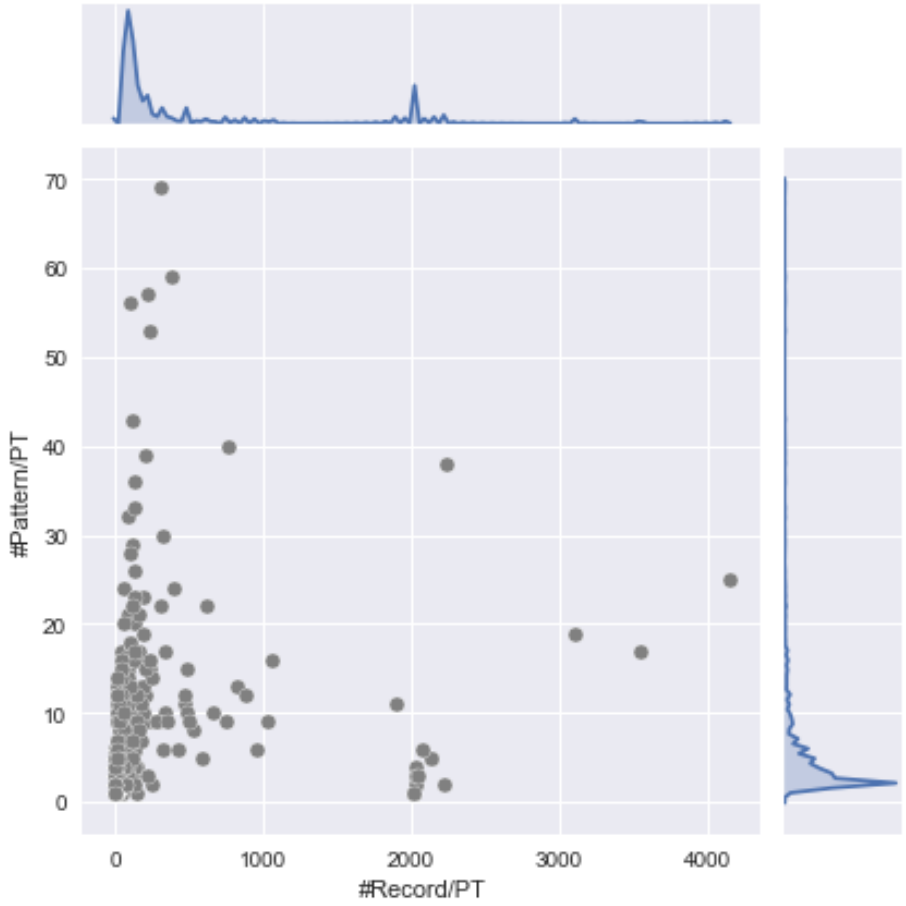
- No structure
- Each layers are fully connected

Bidirectional Recurrent Neural Network (BRNN)

- Treat sequence of the data
- Network is trained by Bidirectional information (Past and Future information)

Characteristics of Reported Term and MedDRA

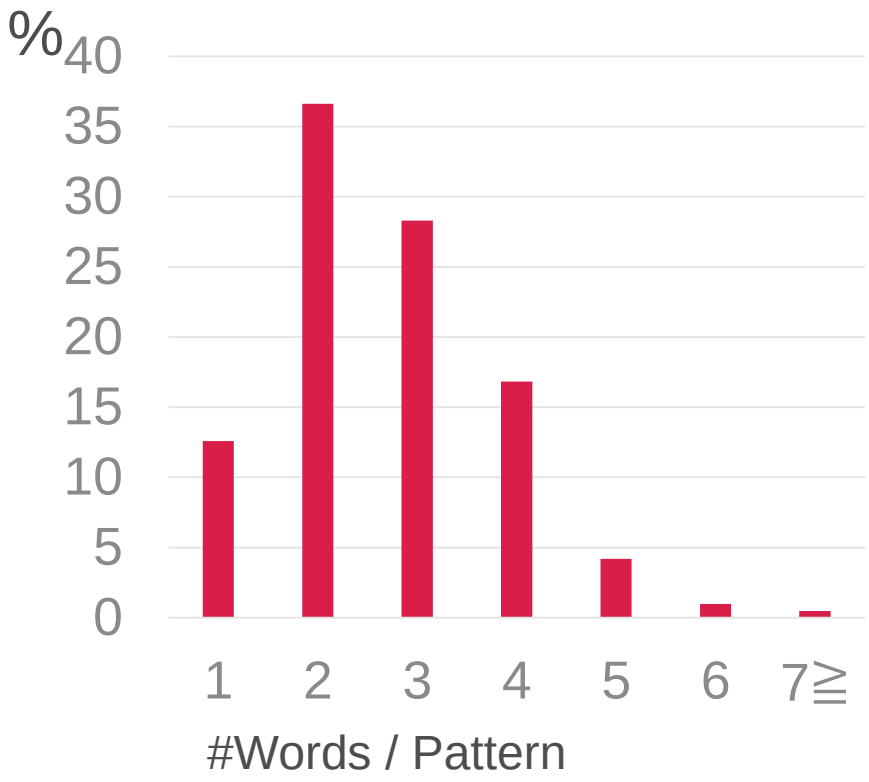
Distribution of Pattern and Record



Pattern: Unique Pre-processed Reported Term
Record: CRF Record

astellas

Distribution of #Words per Pattern



Building Deep Neural Network

astellas

Split Data into Training Data, Validation Data, and Test Data

Split Data

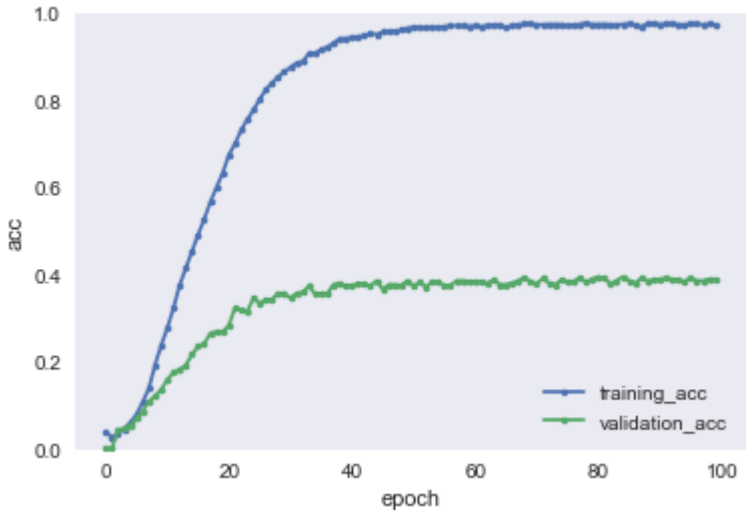
- Training Data
- Validation Data
- Test Data (at least 1 reported term per each PT)

Build Neural Network

- Construct neural network using training data and validation data
- Evaluate neural network by test data

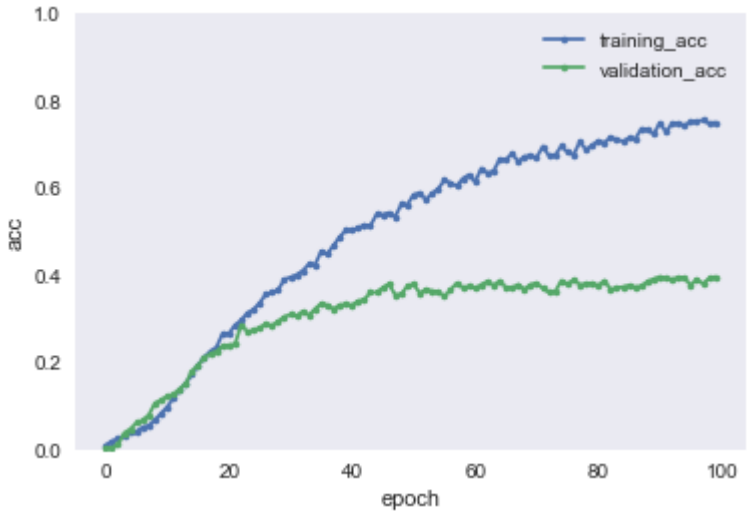
Accuracy of each Structure by Original Data

BRNN



Accuracy for Test: 43%

Simple NN



Accuracy for Test: 42%

astellas

Improve Training

Data Acquisition

- Medical history data
- MedDRA dictionary (LLT is linked to PT)

Data Augmentation

Image data

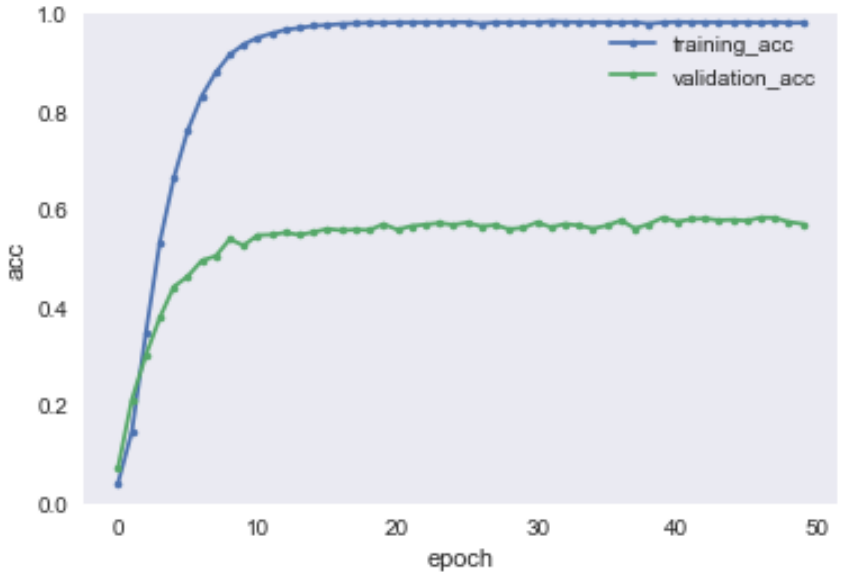
- Rotation, Shift, etc.

Text data

- Replace with Synonyms
- Back translating
- Permutate the words

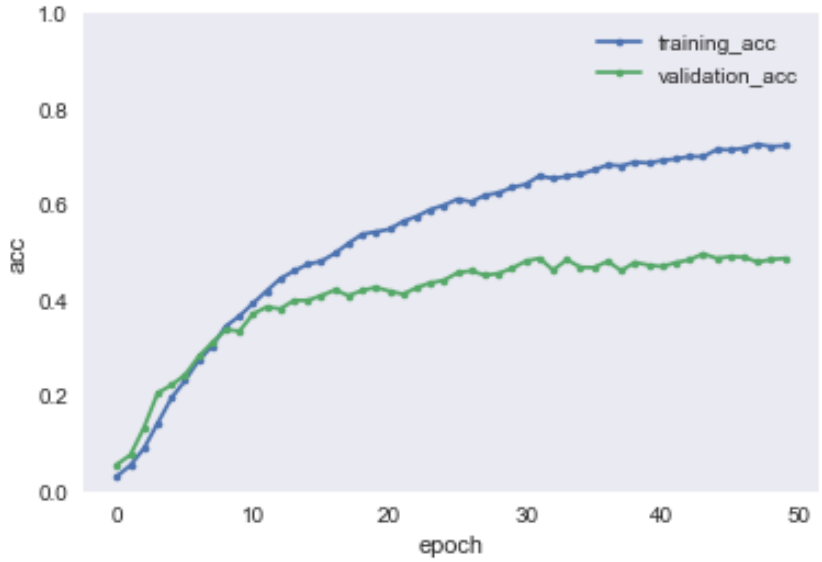
Accuracy of BRNN by each Training Data

Permutated Training Data



Accuracy for Test: 59%

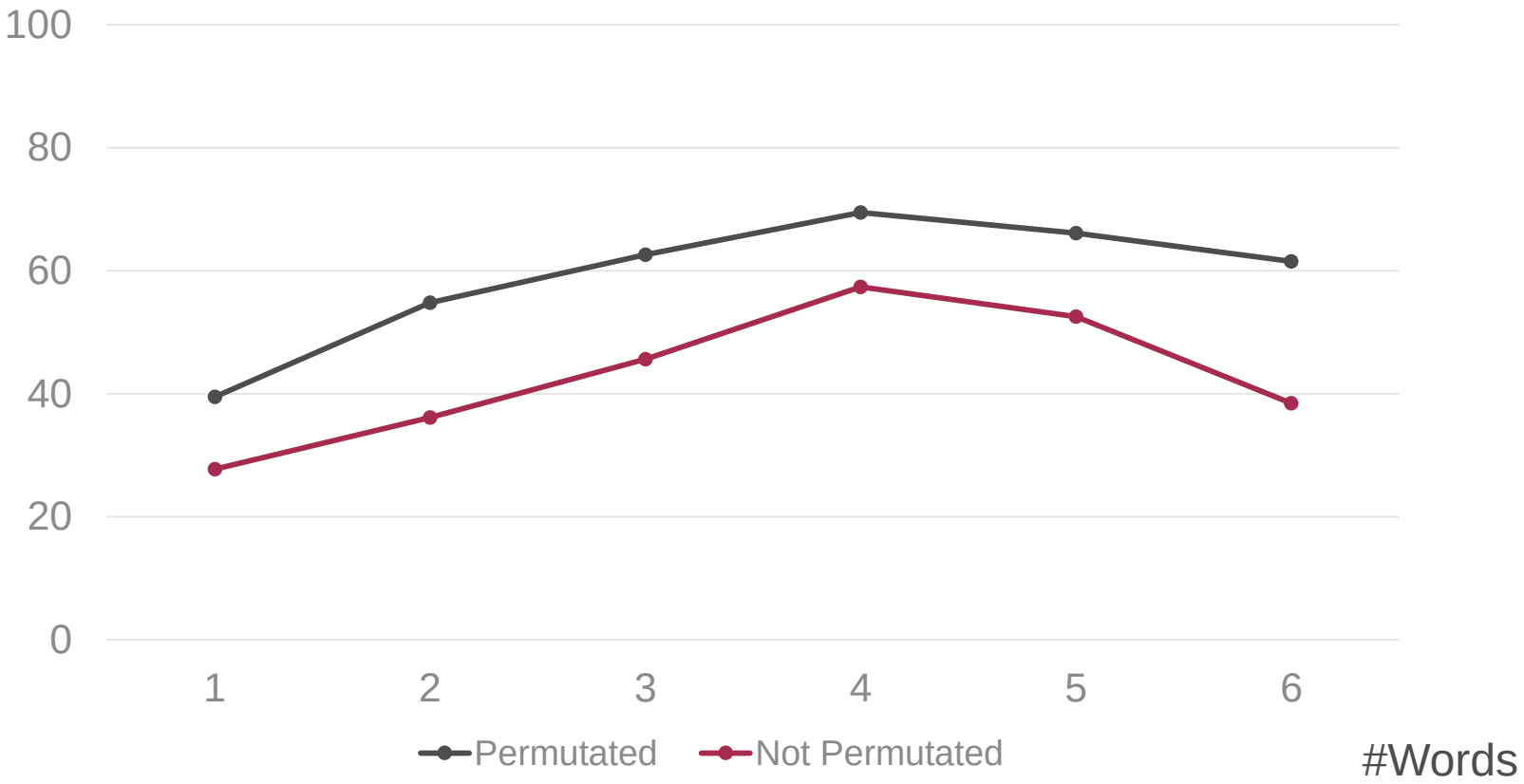
Original Training Data



Accuracy for Test: 43%

astellas

Accuracy of BRNN by Test Data



Improved accuracy using permutated training data
But is the accuracy the best measurement to your purpose?

astellas

Summary

- **Implemented MedDRA mapping process using DNN.**
- **BRNN is better than simple NN.**
- **Word permutation improved test accuracy, but still over-fitting.**
- **Consider process and performance measurement and based on your purpose before implicate to real world.**

References

- 斎藤康毅, ゼロから作るDeep Learning 2 - 自然言語処理編, オライリージャパン, 2018
- Keras document, https://keras.io/
- Mike Schuster and Kuldip K. Paliwal, Bidirectional Recurrent Neural Networks, IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 45, NO. 11, NOVEMBER 1997
- Rico Sennrich, Barry Haddow, and Alexandra Birch, Improving neural machine translation models with monolingual data, arXiv preprint arXiv:1511.06709, 2015.
- Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space, arXiv: 1301.3781, 2013