

はんなり自然言語処理ハンズオンに向けたMTG

2022年2月3日(木)

太田博三

アジェンダ案

10分ぐらいで実演(講義 + Colab)

1. これまでのハンズオンの方向性の共有
2. 想定する参加者・レベル感などの共有
3. 従来の伝統的な自然言語処理とディープラーニング・ベースのとの違いの認識の共有
4. 目次案(アイデアの捕捉)
5. 講義(理論的な説明) + Colabで60ー90分で完結の仮案
6. 日程等
7. その他

共有資料

1. たたき台 : GoogleSpredsheel

https://docs.google.com/spreadsheets/d/1eH4RBq_7NkwBTGZAF8rHkEIW-rtCeE-ly6xVaAhx-VM/edit#gid=0

2. Github https://github.com/otanet/hannari_NLP_hands_on_2022

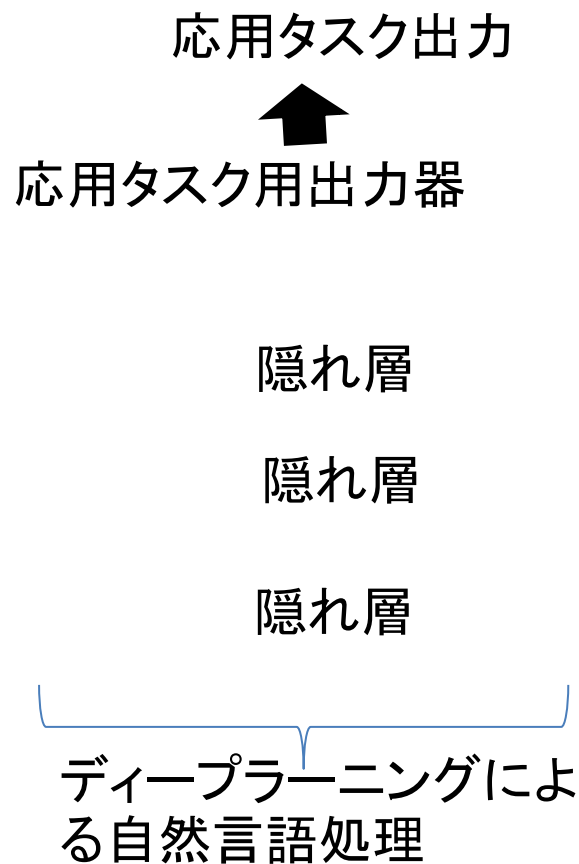
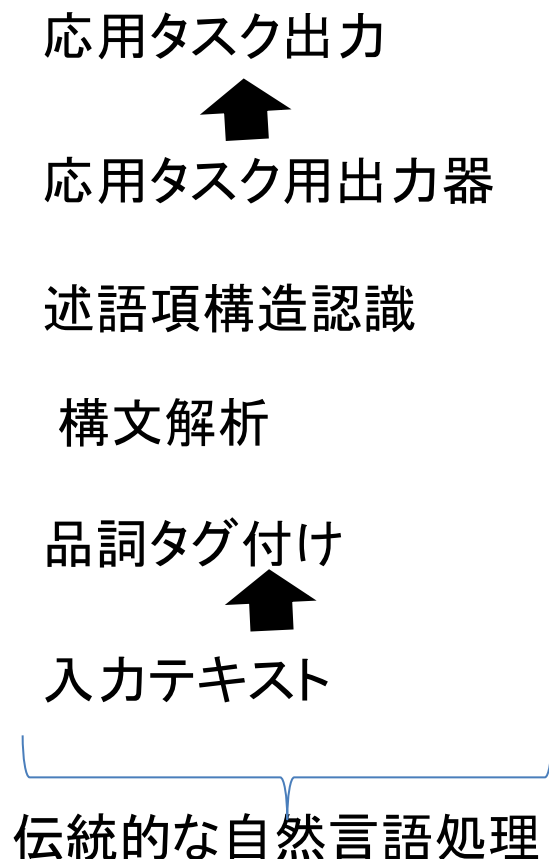
1. これまでのハンズオンの方向性の 共有

- 過去のハンズオンの事例で、注意事項など。

2. 想定する参加者・レベル感などの共有

- 私見では、レベルは高めなのかなという感じです。
- 自然言語処理はあまり詳しくなくても、画像認識とか、隣接分野は詳しい方が少なくない!?

3. 従来の伝統的な自然言語処理とディープラーニング・ベースのとの違いの認識の共有



3. 従来の伝統的な自然言語処理とディープラーニング・ベースのとの違いの認識の共有

伝統的な自然言語処理

形態素解析

意味解析

構文解析

...

・メリット: 1つ1つがわかりやすい
デメリット: 逐次段階であり、Codeで動かすより、理論の理解が中心になる。

ディープラーニングによる自然言語処理

ある特定のタスクを想定して行う。

Embedding?

BERT?

言語モデル(事前学習モデル)

各タスクを実行(ex. 要約や文生成、感情分析)

出力

・メリット: 文生成などの目的に直線でアプローチできる。
デメリット: Codeを動かしても、応用はできない。

目次案(アイデアの捕捉)

伝統的な自然言語処理

- 形態素解析
- 意味解析
- 構文解析

→形態素解析のMeCabや総合的にテキストマイニング(立命館大の名前を忘れて...)が実装例に近い。

※ 係り受け解析は機械翻訳の精度に影響するなど。

- ディープラーニングによる自然言語処理
- (導入編:超入門)
- **TransformerでBERTを扱う-超入門**
- 分散表現: Word2Vecなど
- BERT導入: Masked model
- (基礎編: 主要なタスク)
- BERTによる文章の穴埋め
- BERTによる文章分類
- BERTによる固有表現抽出
- BERTによる文章校正
- T5による自動要約
- GPT-2による文生成
- (応用編)
- (論文の紹介)言語モデルの関係性の検討(アナロジーを題材に)
- (論文の紹介)含意関係認識・オントロジーなど

講義(理論的な説明)＋Colabで60－90分で完 結の仮案

- 別紙参照してください。

日程等

- いつから始めるか？
- 時間は何時間か？
- どのくらい続けるか？

その他

- 特にご協力をお願いしたいこと
- Colanの使い方
- .ipnbをColabに置き換えて実行させるなど