言い換えエラー分析のための言い換え分解ツールの試作

吉仲真人 棉原智之 荒瀬由紀*

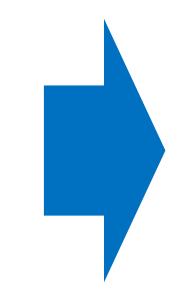
+大阪大学工学部電子情報工学科 #大阪大学データビリティフロンティア機構 *大阪大学大学院情報科学研究科

BERTでAccuracyが2ポイント改善!? それって具体的に何ができるようになったんですか?

言い換え技術の評価: AccuracyやBLEUなど

問題点

具体的にどのような言い換えの認識・生成が 可能になったのかわからない



「言い換え認識モデル・生成モデルは, どのような言い換えに成功/失敗しやすいか」

を明らかにし、モデルの詳細な分析を可能に

最小単位の言い換えへの分解

文レベルの言い換えを 最小単位の言い換え操作の連鎖に分解する

A federal **magistrate** in Fort Lauderdale ordered him held without bail.

Same polarity subs.(habitual)

A federal judge in Fort Lauderdale ordered him held without bail.

Same polarity subs.(contextual)

A federal judge in Fort Lauderdale ordered **Zuccarini** held without bail.

Diathesis alternation

Zuccarini was ordered held without bail by **a federal judge** in Fort Lauderdale.

Addition/Deletion

Zuccarini was ordered held without bail **Wednesday** by a federal judge in Fort Lauderdale.

(言い換え文対はMSRP Corpus [1]より)

言い換え文対を入力し, 最小単位の言い換え操作の連鎖を出力する

入力1:言い換え元の文

出力: 最小単位の 言い換え操作の連鎖

言い換え
分解ツール

入力2:言い換え先の文

先行研究:言い換え操作の分類

最小単位の言い換え操作の分類: EPT [2]

(例)

- Type 1. Inflectional changes
- Type 5. Same polarity subs. (habitual)
- Type 9. Opposite polarity subs. (habitual)
- Type 22. Addition/Deletion

など,全27種類

今後

- ・全ての言い換え操作の実装
- 既存のモデルが成功/失敗した文対を分解
 - → どの種類の言い換え操作が含まれる場合 に成功/失敗しやすいかを明らかにする