知的処理グループ機械翻訳チーム

最終報告(2022/2/4)

惟高日向　寺田智哉
都甲尚志　柳本大輝

内容と目標

・1千万文対の対訳データと深層学習による英→日の機械翻訳。

利用する対訳データ

JParaCrawl (http://www.kecl.ntt.co.jp/icl/lirg/jparacrawl/)

・フロントエンドも開発し、Webアプリとして完成させる。

・目標：Google翻訳の精度を超えること。

役割

フロントエンド（Webページ制作）

→惟高、寺田

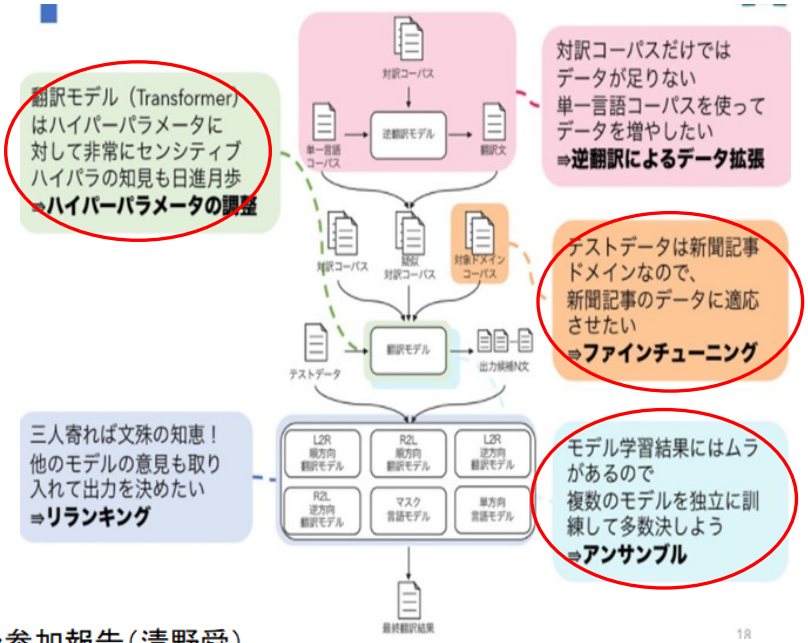
バックエンド（英日機械翻訳の精度向上）

→柳本、都甲

会議：毎週金曜日PBL授業終了後

モデル作成の際に用いた手法

- ファインチューニング
- モデルアンサンブル
- ハイパーパラメータの調整



出典：機械翻訳コンペティション参加報告（清野舜）

スケジュール表

スケジュール内容

論文翻訳・確認
担当：柳本

採用手法の検討・決定
担当：全員

Webページ作成（テスト修正）
担当：惟高・寺田

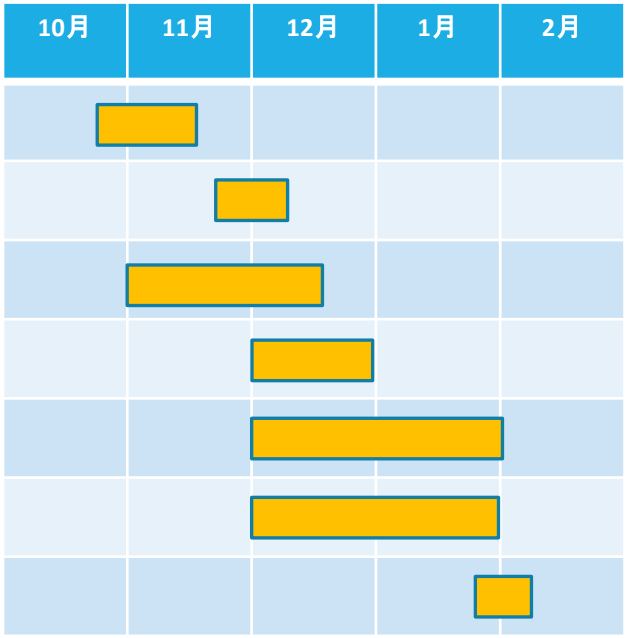
モデル構築
担当：柳本・惟高

モデル訓練
担当：柳本・惟高

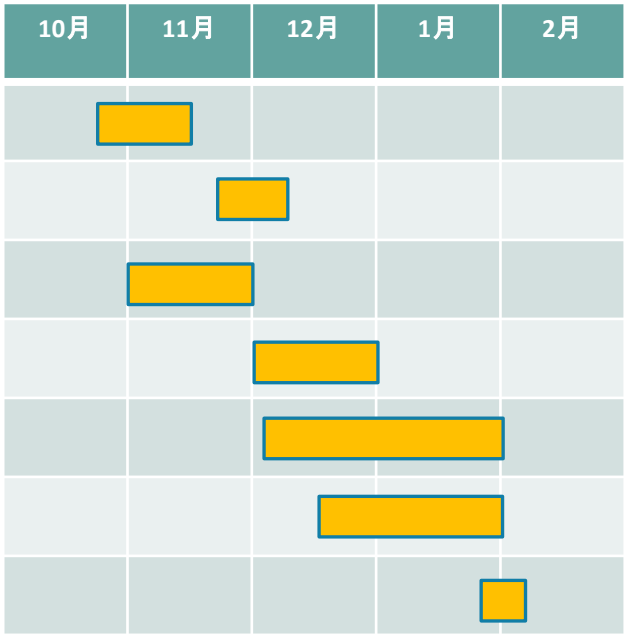
モデル改善
担当：全員

最終資料作成
担当：全員

計画



実際



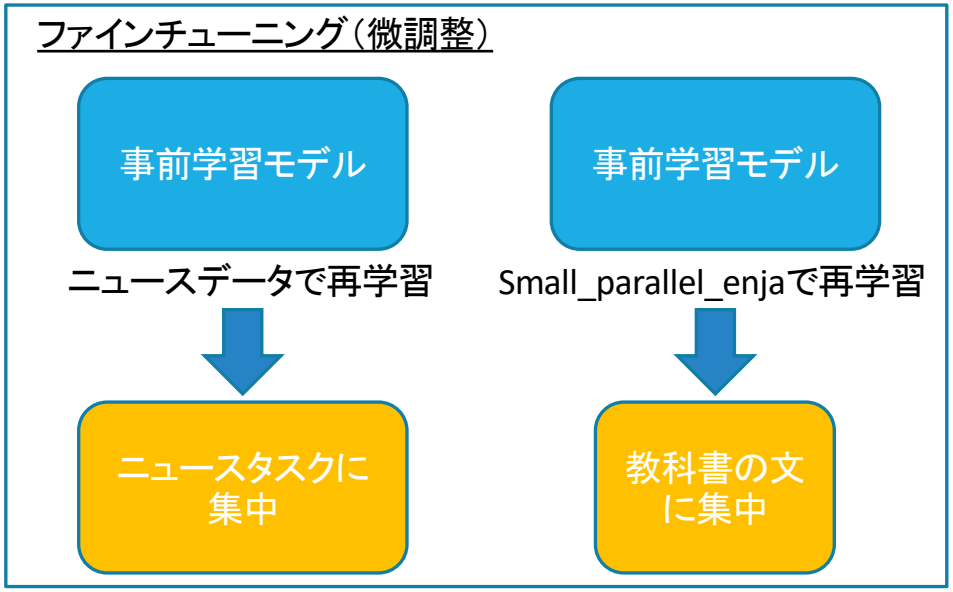
ファインチューニングについて

- 使用したデータ

1. NewsCrawl（40万5千文対のニュースデータ）

2. small_parallel_enja(5万文対)
   - 日本人の大学生によって作成された文のコーパスから文の長さを4~16語でフィルタリングされたもの - 教科書的で硬い文が多い

- 事前学習済みモデルについて
- Webからデータを収集しフィルタリングしたデータ（Jparacrawl（約1千万文対））で学習したモデル



モデルアンサンブルの評価（ニュース）

・1つのモデルのみ

モデル	モデルA	モデルB	モデルC	モデルD	モデルE
sacrebleu	22.45	**22.52**	22.31	22.46	22.22

・モデルアンサンブル

モデル	A+B	A+B+C	A+B+C+D	A+B+C+D+E	A+B+D+E	A+B+D
sacrebleu	22.87	23.07	23.31	**23.52**	23.36	23.12

→評価の高いモデルのみでアンサンブルすることが必ずしも効果的であるわけではない
　また、モデルの数を増やす≠性能の向上

実験結果(ニュース)

- ニュースタスクデータで評価

モデル	SacreBLEU
ベースライン	18.31
＋ファインチューニング	22.52
＋モデルアンサンブル	23.52
Google翻訳	**24.19**

モデルアンサンブルの評価（教科書）

・1つのモデルのみ

モデル	モデルA	モデルB	モデルC	モデルD	モデルE
BLEU	39.92	39.53	39.97	**40.07**	39.71

・モデルアンサンブル

モデル	A+B	A+B+C	A+B+C+D	A+B+C+D+E	A+C+D	C+D
BLEU	**40.13**	39.73	39.97	39.70	39.92	39.93

→評価の高いモデルのみでアンサンブルすることが必ずしも効果的であるわけではない
　また、モデルの数を増やす≠性能の向上

実験結果(教科書)

- small_parallel_enjaで評価

モデル	SacreBLEU
ベースライン	20.44
＋ ファインチューニング	40.07
＋ モデルアンサンブル	**40.13**
Google翻訳	31.53

活動の成果

Webページ

まとめ

・英→日の機械翻訳器を作成し、翻訳機能を利用できるWebアプリとして開発した。

・ファインチューニング、モデルアンサンブル等の手法を用いGoogle翻訳を超えるよう尽力した。