

知識の整理のための根拠付き自然文間含意関係コーパスの構築

林部祐太 (株式会社リクルート Megagon Labs, Tokyo, Japan)

動機

じゃらんnetのレビューから、
「**どのような宿の特徴がカスタマーに好評か**」
という知識を得て整理したい！



BERTのPre-Training

- じゃらんnetに投稿された口コミ2,000万文以上
- SentencePiece (size=32k)でトークナイズ
- 12-layer, 768-hidden, 12-heads
- 100万steps, バッチサイズ512
- TPuv3-8で約63時間 (約55,000円)

語彙の例

語彙の例
、のがではもに！をまたですや
2たでした（）1のでからしとてもて？
3ただありがとうございます...なのでには
ですが良かったです部屋はでもなど
でしたがな今回は」ホテル少しか朝食



コーパスの特徴

1. 自然文のみで構築

アノテータによる作文
→コストが高い、事例が偏る
→自然文だけを使う！

- ・**頻度フィルタ**：頻度2以上の文を利用
- ・**文長フィルタ**：5～40文字の文を利用
- ・**宿の特徴フィルタ**：関係有の文を利用
5,370文で学習、F=94.0
- ・**感情極性フィルタ**：肯定的な文を利用
5,370文で学習、F=90.2

21,868種類の文をアノテーション候補

文	特徴	感情	対象
立地バツグンです。	有	肯定	
全く美味しくない。	有	否定	×

2. 複雑な構造の文は除外

応用と分析のしやすいコーパスとする
ため、原子的な事例のみを含めたい

× 立地は抜群で、部屋も綺麗です。

述語項構造フィルタ

- ・JUMAN++とKNP++の構文・格解析の結果に基づき、複数の述語項構造を含む文を除外
- ・含まれる述語項構造と名詞構造が全く同じとなる文が複数含まれないように

以上により、21,868 種の文を候補に

また利用したいですありがとうございました。	無	肯定	×
結婚記念日旅行に利用させていただきました。	無	中立	×

3. 根拠付きのアノテーション

レビュー：お部屋からもスカイツリーが見えました。

- ☐ お部屋から
 - ☐ 花火
 - ☐ が見れる！
 - ☐ <正しい>
 - ☐ <意味不明、回答困難>
- 上記の宿のレビュー文を踏まえて考えたとき、
仮説文：お部屋から花火が見れる！
は正しいですか？

- ・正しい場合は<正しい>のみを選ぶ
- ・正しいといえない場合は、その箇所を選ぶ (複数選択可)
- ・意味不明な文がある場合や、特定の箇所での指摘が困難な場合は<意味不明、回答困難>を選ぶ

クラウドワーカー3人でアノテーション



コーパスの構築手順

- ・類似文検索で”そこそこ”類似している文を前提文として利用
- ・2段階で作成

BASE

- ・21,868文から頻度順に2,017 文を選んだ
- ・同一の述語は最大5文
- ・各文に類似文を5文選択
- ・2017文x5文から8,065組選択

APPEND

- ・3,843文選択し、類似文を10文選択
- ・仮説・前提入れ替え
- ・(3,843文x15文x2)事例をBASEで学習したモデルで分類し、各スコア帯から計9,000文抽出

類似文検索デモ (Jalan)



実家のような雰囲気でした。
0.30 秒 [CLS], "実家", "どのような雰囲気", "でした", ":", "[SEP]" [4, 5667, 19874, 23,
"bert": 0.27742505073547363, "faiss": 0.020982980728149414, "make_response": 0.00042510
>>

RTE	文	件数
0	実家のような居心地の良さでした。	
1	実家のような安心感です。	
2	実家のような、アットホームな感じでした。	
3	部屋は実家のような雰囲気で落ち着きます。	
4	実家に帰ってきたような感覚でした。	
5	自宅のような雰囲気でも過ごしやすいかったです。	
6	おばあちゃん家のような雰囲気です。	
7	実家のような温かさ。	

	BASE		APPEND	
	学習	評価	学習	評価
含意	2,826	278	2,401	269
非含意 (うち回答困難)	4,432 (833)	529 (93)	5,699 (692)	631 (81)
合計	7,258	807	8,100	900

性能評価実験

- ・BERTモデルをfine-tuning
- ・バッチサイズ32, 最大入力トークン数は25, エポック数3
- ・APPENDに対する性能は、BASEに対する性能を下回る傾向→難しい問題が多い

	BASE			APPEND		
	P	R	F	P	R	F
BASE	88.5	96.8	92.4	72.3	79.6	75.8
APPEND	80.4	91.4	85.5	80.0	83.3	81.6
ALL	87.1	97.1	91.8	82.1	83.3	82.7

今後の課題

- ・より対象を限定して質の向上・知識の整理への適用



判断が難しい事例

語句に曖昧性がある事例

H:ご飯が美味しい P:米が美味しい。
“ご飯”=食事?米? “刺身”, “休む”等

項の省略がある事例

H:のんびり入る事が出来ました。 P:ゆったりと入浴出来ました。

評価を直接的に記述していない事例

H:金目の煮付けは、最高でした
P:あの味が忘れられません。

反例の想定に個人差がある事例

H:目の前がビーチです。 P:目の前が海です。
H:温泉、良かったです！ P:ゆっくり露天風呂に入れました。

評価対象・評価値が曖昧・不明な事例

H:全てにおいて文句なしでした。 P:接客サービスも素晴らしかったです。
H:スタッフの方の対応。 P:お手頃価格。

強調表現が仮説文にある事例

H:朝食は軽食でした。
P:しかも軽食付き。

参考文献

[Devlin+, 2019] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
[Arora+, 2017] A Simple but Tough-to-beat baseline for sentence embeddings
faiss: <https://github.com/facebookresearch/faiss>
[Camburu+ 2018]e-SNLI: Natural Language Inference with Natural Language Explanations
[平+, 2010]仮説オントロジの構築方法の検討