

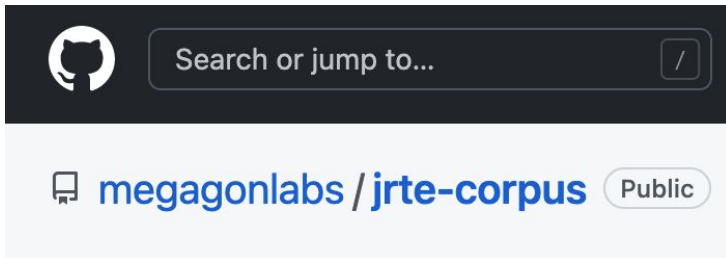


Megagon Labs

Japanese Realistic Textual Entailment Corpus (JRTE) の紹介

@JED2022
2022/03/18

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



Japanese Realistic Textual Entailment Corpus



Overview

This corpus contains examples labeled whether the premise entails the hypothesis or not as follows.

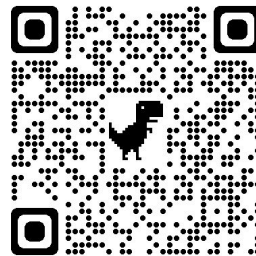
```
Hypothesis: 部屋から海が見える。(You can see the ocean from your room.)
Premise   : 部屋はオーシャンビューで景色がよかったです。(The room had an ocean view and a nice view.)
Label    : Entailment
```

All examples utilize texts in Japanese hotel reviews posted on [Jalan](#), which is a travel information web site. This corpus also contains sentences with sentiment polarity labels and labels whether the text is hotel reputation or not as follows.

```
Text       : 朝食が美味しいです。(The breakfast is delicious.)
Sentiment  : Positive
Hotel reputation: True
```

⚠ Because some of the data have been removed for various reasons, this corpus does not exactly correspond to one used in the reference papers.

GitHubからダウンロードし
てすぐに使えます



<https://github.com/megagonlabs/jrte-corpus>

JRTE Corpusとは

- 旅行情報サイト『じゃらんnet』上で公開されている宿泊施設へのクチコミを用いて作ったコーパス
- 3種類のラベル
 1. 宿の特徴の有無ラベル (2値, 5,553件)
 2. 感情極性ラベル (3値, 5,553件)
 3. 含意関係ラベル (2値, 54,996件)

文	特徴 (RHR)	感情 (PN)
部屋も清潔でとても過ごしやすかったです。	1	1
全く美味しくない。	1	-1
また利用したいですありがとうございました。	0	1
結婚記念日旅行に利用させていただきました。	0	0

仮説H	前提P	含意関係 (RTE)
部屋から海が見える。	お部屋からの眺めは、最高でした。	0
部屋から海が見える。	部屋はオーシャンビューで景色がよかったです。	1

一部には非含意根拠ラベル有り

仮説H	とても綺麗なカプセルホテル！
前提P	館内とても綺麗でした。
含意関係	0 (無し)
非含意根拠	[["とても", 0], ["綺麗な", 0], [<u>"カプセルホテル"</u> , 3], ["!", 0], ["<1>", 0], ["<unknown>", 0]]

自動分類のサンプル

感情極性分類

```
$ poetry run transformers-cli serve --task sentiment-analysis --model ./model-pn --port 8900
$ curl -X POST -H "Content-Type: application/json" "http://localhost:8900/forward" -d
'{"inputs":["ご飯が美味しいです。", "3人で行きました。", "部屋は狭かったです。"]}'
{"output":[{"label":"pos","score":0.8015708923339844},{label":"neu","score":0.47732535004615784
},{label":"neg","score":0.42699119448661804}]}
```

宿特徴有無判定

```
$ poetry run transformers-cli serve --task sentiment-analysis --model ./model-rhr --port 8901
$ curl -X POST -H "Content-Type: application/json" "http://localhost:8901/forward" -d
'{"inputs":["ご飯が美味しいです。", "3人で行きました。"]}'
{"output":[{"label":"yes","score":0.9653761386871338},{label":"no","score":0.8748807907104492}]
}
```

含意関係認識

```
$ poetry run transformers-cli serve --task sentiment-analysis --model ./model-rte --port 8902
$ curl -X POST -H "Content-Type: application/json" "http://localhost:8902/forward" -d
'{"inputs":[["風呂がきれいです。", "食事が美味しいです"], ["暑いです。", "とても暑かった"]]
}'
{"output":[{"label":"NE","score":0.9982748627662659},{label":"E","score":0.9790723323822021}]}
```

- [じゃらんnetに投稿された宿クチコミを用いた感情極性分析・含意関係認識の一例](#) - Megagon Labs

作ったモチベーション

宿の提案のために,

1. 宿の特徴に関して,
2. 好評な意見を抽出し,
3. 関連する意見を含意関係でまとめたかった

から

データの活用例

いろいろな方が利用してくださっています

- blog記事
- Kaggle community competition

紹介 @dcm_ozaki (NTTドコモ R&D)

投稿日 2020年12月14日

Organization

BERTを用いて日本語レビューコーパスの含意関係認識やってみた

Python, 自然言語処理, NLP, MachineLearning, bert

はじめに

こんにちは。

NTTドコモサービスイノベーション部1年目社員の尾崎です。

この記事は、NTTドコモ R&D Advent Calendar 2020 14日目の記事です。

通常業務では、位置情報を活用したレコメンド基盤の開発を担当しています。

先日、自然言語処理コーパスについて知らべていたら、LREC 2020に通っている含意関係認識(後記述)の日本語コーパスが公開されていたため

早速BERTにて含意関係認識タスクを行ってみました。

https://qiita.com/dcm_ozaki/items/1c770a7bb6994c2b073f



日本語文章の感情分析APIをつくる

📅 2021.12.27に公開 ⌚ 5 min read

こんばんは、寒いですね
さすがに故郷は真っ白に染まっているようです
久しぶりに帰りたいかな

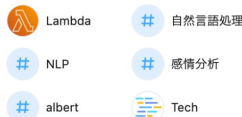
さて、年末年始と呼ばれる期間に突入しますが皆さんいかがお過ごしですか？
「年末年始、時間あるからなにかやってみようかな」
そう考えてる方にちょうどいい(と思ってもらえそうな)ものを用意しました！

日本語文章の感情分析APIをつくる

というわけで本題です。
今回は日本語の文章の感情を分析するモデルを作成→デプロイしてAPIとして利用できるようにし
てみよう、という話です。
ちなみに「APIまではちょっと…」という方も、モデル作成して推論結果を見るところまででも楽しめると思うので是非に。

まあ感情分析と言ってもよくあるネガティブ・ポジティブ・ニュートラルの分析です。
COTOHA APIの感情分析みたいなものを自分でつくってみましょう、という話ですね。
余談ですがCOTOHA API大好きです。こんなお手軽にいろいろなタスクが実行できるのは素晴らしいと思います。

Topics



ken



Follow

睡眠爆殺芸人☆です

👉 サポートする

目次

- 日本語文章の感情分析APIをつくる
- モデルをつくる
 - ALBERT
 - 学習データ
 - その他
- Lambdaにデプロイする

https://zenn.dev/ken_11/articles/8f37dea8f93221

感情極性の予測コンペ

<https://www.kaggle.com/c/rkcup-1/overview/acknowledgement>

+ Create

- Home
- Competitions**
- Datasets
- Code
- Discussions
- Courses
- More

Leaderboard

Raw Data

Search leaderboard

Public **Private**

The private leaderboard is calculated with approximately 75% of the test data. This competition has completed. This leaderboard reflects the final standings.

#	△	Team	Members	Score	Entries	Last
1	↔	後は折る時間		0.99493	7	8mo
2	↔	kanna hashimoto		0.99336	16	8mo
3	↑	こんなデータあったんだ		0.99258	1	8mo
4	↓	Bowser in the Sky		0.99246	31	8mo
5	—	GPU使わない縛りやめた(BERT)		0.99172	5	8mo
6	↓	BERTにHello worldした		0.99144	6	8mo
7	—	【運営】BERT Baseline		0.99084	6	8mo
8	↓	uhahahahaha		0.99079	12	8mo

Search

Sign In

Community Prediction Competition

RKcup #1

Jalan Sentiment Analysis Challenge

15 teams · 8 months ago

Overview Data Code Discussion Leaderboard Rules

Overview

Description

RKcup #1

Evaluation

ハンズオン：2021/6/29 19:00 ~

Timeline

コンペ：2021/6/29 19:00 ~ 2021/7/6 19:00

Acknowledgement

この期間中は各人が1日20回まで自由に予測結果をサブミットすることができます

データ：じゃらん口コミ

目的：自然言語処理の基礎から応用までをハンズオンを通して身につけて、コンペを楽しみましょう

Launch

8 months ago

Close

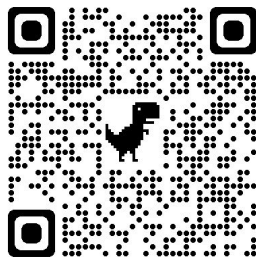
8 months ago

15 Teams 15 Competitors 130 Entries

Points This competition did not award ranking points
Tiers This competition did not count towards tiers

This is a Community competition provided for anyone to use.
Host your own Community competition »

ぜひ、みなさまご利用ください



<https://github.com/megagonlabs/jrte-corpus>