

## タイトル

195702D 奥間 涼  
195748B 德里 光陽  
195761K 塩月 流星  
195769E 佐藤 和周  
195770J 栗原 瑠威

2022 年 1 月 18 日

## 目次

1	テーマ	2
2	目的	2
3	目標	2
4	アプローチの全体像	2
5	予定していた実験計画	2
6	データセットの構築方法	2
7	機械学習の進め方	3
8	実験	3
8.1	実験設計 . . . . .	3
8.2	実験結果 . . . . .	3
9	考察	3
10	自己評価や振り返り	3
11	時間の都合上省いた項目	3

## 1 テーマ

最近、オンラインショップを活用する人々が増えている。その一方でサクラレビューによって購入者が求めている商品が届かないケースが増えている。

購入者が求めている商品が届かないということは、購入者の金銭や時間の損失だけではなく質の悪い商品が出回ることによって社会に悪影響を与えることが考えられる。

特に有名なオンラインショッピングサイトである Amazon に着目し、その中から良くない商品、サクラレビューを中心にサクラ判定を行っていかうと考えた。

よって本グループでは『Amazon レビューのサクラ判定』というテーマを設定しそれを元に実験を行った。

## 2 目的

今回の実験を通して、サクラレビューの特徴を獲得し実際のレビューに適用できるモデルを作成する。

## 3 目標

サクラレビューの認識率 70% 以上のモデルを作成する。

## 4 アプローチの全体像

## 5 予定していた実験計画

## 6 データセットの構築方法

Amazon の商品ページの中からサクラレビューの含まれていそうな URL を収集し、サクラチェッカーで判定する。サクラが多いと判断された商品と健全なレビューと判断された商品をデータセットの作成した。

データの取得には BeautifulSoup4(以下 bs4) と selenium を用いた。

bs4 の機能を用いることで以下の用にタグからデータを取得した。

- .product-title = 商品の名前
- .averageStarRating = 商品の平均評価
- .review-title-content = レビュータイトル
- .review-rating = レビュー評価
- .review-text = レビュー本文

収集した商品のレビューをサクラレビューかどうかを各メンバーが5段階評価を行い、平均をその商品の評価としてデータを作成する。

BERT と spaCy を使用してレビュー文をベクトル化した。

## 7 機械学習の進め方

(a) どの学習器を使ったのか。

RandomForest

(b) パラメータはどのように調整したのか。

デフォルト値

## 8 実験

### 8.1 実験設計

### 8.2 実験結果

## 9 考察

## 10 自己評価や振り返り

## 11 時間の都合上省いた項目

## 参考文献

- [1] ジコログ, 初心者でも簡単にできる Selenium のインストール [Python]  
[https://self-development.info/初心者でも簡単にできる selenium のインストール【python】/](https://self-development.info/初心者でも簡単にできる_seleniumのインストール【python】/),  
参照 2021-10-26.
- [2] sakura-checker, サクラチェッカー,  
[https://sakura-checker.jp](https://sakura-checker.jp/),  
参照 2021-11-09.
- [3] Kazuyoshi Sato, Word2Vec+RandomForest で笑点の回答者ともらえる座布団の枚数を予測する,  
<https://qiita.com/WhitePepperKS/items/ef9ab61bd921a5fb7085>,  
参照 2021-12-14.
- [4] ENOKI - AI を活用した無人チャットボットのプラットフォームを提供, BERT を使用した文章ベクトル作成,  
<https://www.enoki.enoki-inc.co.jp/tech-6/>,

参照 2021-12-21.