

# ベンフォードの法則によるフェイクレビュー検出手法の検討

NECO B2 江頭叙那 (jonah)

親 ks91

2021 年 1 月 25 日

## 1 はじめに

ネットショッピングが普及し、多くの人々がオンラインで買い物をするようになった。ユーザーは商品を買う際にレビューを参考にするが、企業が金銭を見返りに偽のレビューを書かせるフェイクレビューが問題になっている。

## 2 目的

ベンフォードの法則を応用した素性を利用して、既存方式よりも少ない情報から、商品にフェイクレビューが含まれているかどうかを検出することを目指す。

## 3 ベンフォードの法則の概要

ベンフォードの法則とは、電気料金の請求書、住所の番地、株価、人口、革の長さ、物理・数学定数などの自然界に現れる数値集合の各数値における最上位桁の数値の出現確率には偏りがあるという法則である。この分布から大きく離れている場合、何らかの人手による操作などが行われた可能性がある。

## 4 仮説

偽のレビューを書くレビュアーは実際に商品を使っておらず、さらに報酬を得るために短期間で多くのレビューを書く必要があり、他のレビューを参考に書いているケースが多いのではないかと考える。その場合、偽のレビューを含んでいない商品と比べて使用される文字の頻度の分布が異なるのではないかと考えた。

## 5 既存方式

Amazon のフェイクレビューを見抜くサクラチェッカーでは、価格・製品、レビュー分布やショッピングレビューなど計 8 項目の情報を、独自のロジックや機械学習を用いて分析している。

## 6 課題

サクラチェッカーでは Amazon のレビューしか判定することができず、他 EC サイトのレビュー判定ができない。また、分析項目が多く他サイトに応用しづらい点が課題である。

## 7 提案方式

1. サクラチェッカーでサクラ度が 20% 以下の商品を安全な商品、80% 以上の商品を危険な商品とし、2 グループそれぞれ 100 商品のレビューを Selenium によって取得する。
2. 取得したレビューの文字の出現頻度をカウントし、頻度の最上位桁の割合を計算する。
3. 安全な商品と危険な商品のグループで、最上位桁の分布がどのように異なるかを分析する。
4. 分析結果をもとに商品にフェイクレビューが含まれているかどうかを判定する

## 8 結果

箇条書きとともに結果

- ちゃお
- りぼん
- なかよし

これは番号を振る箇条書きです。

1. 富士
2. 鷹
3. なすび

## 9 考察