F-015

フェイクレビュー対策のためのメタ情報を利用した外れ値検出
Outlier Detection using Meta Information for Countering Fake Reviews

黒木 亮人† 成 凱†
Kurogi Akito Chang Kai

1. はじめに

近年、インターネットの発展により EC サイトなどのインターネットを介した商品の購入の需要が増加している。Gooリサーチが行った調査[1]によると、日本国内15歳以上の男女（計 2,107 名）を対象に行った「購買行動においてクチコミが与える影響」について調査したところ、全体の81.6%が商品の購入・選定の際に「レビューが気になる」と回答したとあり、多くの人がレビューを参考にすることがわかる。しかしレビューの中には「フェイクレビュー」という偽のレビューが存在する。フェイクレビューの存在によって、商品に対する正当な評価が失われる、WEB サイトの信頼性が失われるなどの影響があり、この影響が長期的に続くことにより健全なデジタル社会の実現に悪影響を及ぼす可能性がある。よって、フェイクレビューの対策を行う必要がある。

フェイクレビュー対策で問題となるのがレビューというものは非常に主観的な性質を持つため、フェイクレビューの真偽を正確に判断できるのはフェイクレビューの作成者のみ、という点である。後述するが、フェイクレビュー対策では主に機械学習が使われており、その中でも教師あり学習が最も多く使われてきた。その際のラベル付きデータセットの入手法には以下の方法がある。yelp.com が公開している、独自のフィルタリングアルゴリズムによってラベル付けされたデータセットを利用する方法、クラウドソーシングを利用して大量の架空のレビューデータセットを得る方法、人間が実際にレビューをチェックしてラベル付けする方法、ラベル付けのためのルールを決め、それに従いラベル付けする方法がある[4]。

しかし、いずれにしろ現実のレビューに真に正しくラベル付けをしたデータセットではなく、さらに人間がラベルをつける方法は時間とコストがかかる。よってこの問題を回避した上でフェイクレビュー検出を行う必要がある。

現在のフェイクレビュー検出のための研究においては、分析するデータについて主に３種類に分けることができる。レビューテキストによる検出、レビュアーの行動による検出、偽のレビュアーグループの行動による検出の３種類である。レビューテキストによる検出とは、レビューテキストを言語学的に分析する手法であり、フェイクレビューとそうでないレビューはテキストに異なる傾向を持つことを利用し n-gram などの言語的分析で分類を行う。レビュアーの行動による検出では、偽のレビュアーと通常のレビュアーが異なるレビューに関した行動をする傾向がある、ということを利用して偽のレビュアーを特定する。例えば、偽のレビュアーは早期に極端な評価をする、１日に異なる商品に対して複数のレビューをする、という傾向は典型的な偽のレビュアーの行動傾向である。

†九州産業大学　Kyushu Sangyo University

偽のレビュアーグループによる検出とは、偽のレビュアーが同時期に同じ商品をレビューし商品の平均評価を操作する、など集団で特定の行動をとることがあるので、その傾向を利用して偽のレビュアーを特定する手法である。[5]

偽のレビューは一種の詐欺とみなすことができ、そのため外れ値検出という手法が使用できる。外れ値検出とは通常のデータとは異なる点である「外れ値」を検出する技術であり、コンピュータセキュリティや医療診断など多くの場面で使用されている技術である。[6]外れ値検出では通常のデータ点と異なるデータ点とみなされたレビューをフェイクレビューであると判断する。外れ値検出には教師あり学習、教師なし学習、深層学習を使用した手法があるが、本研究では教師なし学習を使用した外れ値検出を行う。教師なし学習を使用する理由として、レビューの主観性によるラベル付きデータの入手が難しい、または入手に大きなコストがかかるという問題を回避できる。また、深層学習は優れた正確性を示すが深層学習アルゴリズムはブラックボックスとなっており、データの説明性が低いという欠点があり[4]、比較的説明性の高い教師なし学習を使用する。

2. 関連研究

外れ値検出を利用したフェイクレビュー検出の研究として David Savage ら[7]は、平均評価を操作する偽のレビュアーと大多数の通常のレビュアーの評価の違いについて二項回帰を用いて大多数の通常レビュアーの意見から乖離した偽のレビュアーを特定した。これは平均評価が最も他のサイト利用者に影響を与えると予測し、評価のみに基づくアプローチにより計算量を削減しつつ FraudEagle 手法よりも優れた精度を示した。

Wenqian Liu ら[8]は、レビューテキストだけでなく商品に関連するレビュー記録よりレビューの時間的特徴を抽出し、IsolationForest 手法で異常検知を行った。有効性の検証の際に外れ値検出の手法である ARIMA、LOF、SVM との比較を行い、IsolationForest 手法が比較した手法よりも高い正確性を示した。

これらの研究では商品評価やレビューの時間的特徴に注目することにより従来の手法よりも高い正確性を示したが、これにレビューテキスト、商品評価以外のレビューのメタデータ、偽のレビュアーグループの行動などその他の特徴を利用することでより優れた精度を示すことができることが考えられる。

3. 提案手法

本研究では商品評価とレビュアー行動に注目した外れ値検出を行う。レビュアーの信用度を示すレビュアースコアを、商品評価、レビュアー行動、レビュアーグループ行動から算出し、偽のレビュアーと通常のレビュアーを分類し、フェイクレビューを検出する。また既存の外れ値検出との比較を行い、有効性を検証する。

4. 実験

外れ値検出に教師なし外れ値検出の手法の一つである IsolationForest によって外れ値検出を行う。IsolationForest は Fei Tony Liu ら[9]が提案したモデルであり RandomForest と同じくアンサンブル学習の一つである。 アンサンブル学習とは複数の機械学習モデルを組み合わせることでより強力なモデルを構築する手法である。入力したデータ点の集合を決定木で分割していき、特徴量をランダムに選び特徴量の最大値から最初値の区間で分割点をランダムに選ぶ。分割を繰り返していくと葉にはデータ点が 1 つだけになるように分割していき、ルートノードから葉までの距離が異常スコアを表す。早い段階で外れ値は分割される確率が高いためルートノードから葉までの距離(Average Path Length)が小さいほど異常スコアが高くなる[9]。

あるデータ点 x の異常度 s(x,n)を算出する。まず木の内部での h(x)の正規化した値 c(n)は(2)式で算出する。h(x)をパス長、H(i)を調和数とする[10]。

$$H(i) = \log(i) + 0.57721 \tag{1}$$

$$c(n) = H(n-1) - \left(\frac{2(n-1)}{n}\right) \tag{2}$$

異常度 s(x,n)は(3)式で算出する。E(h(x))を全ての木の内部での h(x)の平均値とする。

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{3}$$

異常かどうかの判定は異常度 s が 1 に近いとき異常なデータ点、0.5 より小さいとき異常な値データ点ではないと判定される。また、全ての s(x,n)=0.5 の場合異常なデータセットはデータセットにないと判定される。[10]

IsolationForest の利点として高速な外れ値検出が可能、メモリ占有率が比較的少ない、小さなデータセットでも外れ値検出が可能、不均衡データにも使いやすいというメリットがある。また、外れ値検出の問題点である Swamping と Masking という問題点を解決することができる。Swamping とは通常のデータが異常なデータと近いときに False negative、つまり異常なデータを正常なデータとして処理してしまうという現象であり、Masking とは異常なデータが多い場合に、それらが密な集合となり検出することができない現象を指す。これらの問題を IsolationForest では元データからサブサンプリングを行い、部分モデルを作成することで解決を行うことができる[10][11]。

使用レビューデータセットとしてデータ収集・提供サイト Datafiniti が提供する、Amazon や Best Buy などの Web サイトから収集した 50 種類のエレクトロニクス製品に対するオンラインレビューのデータセット 7000 件の内、Micorsoft ブランドを持つある商品についてのレビュー495 件を対象とした。

商品の評価点数 review.rating(1～5)について外れ値検出を行う。IsolationForest 使用時に設定したパラメータとして、アンサンブル内の木の数を 100、それぞれの木を作るためにランダムサンプリングされるデータ数を 256、データセットに含まれる異常値の割合を 1%として設定した。

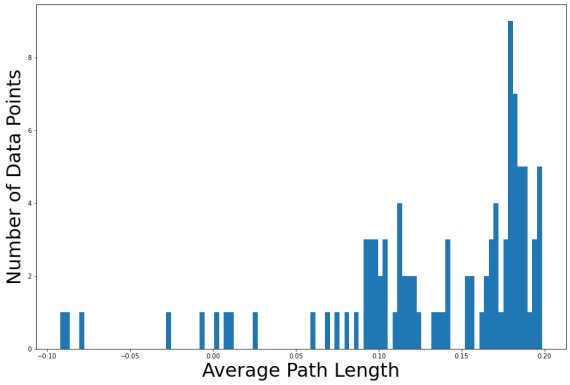

図 1 Microsoft 製品に対しての異常度グラフ

図 1 のようにヒストグラムによりレビューの異常値分布を示した。Average Path Length が小さいほど異常度が高いレビューとなっている。

5. まとめと今後の課題

本研究ではレビューテキスト、レビューメタデータ、レビュアー行動、レビュアーグループ行動などを特徴量として扱い外れ値検出によって高精度で説明性の高いフェイクレビュー検出を行う。実験では Datafiniti が提供するエレクトロニクス製品に対するレビューのうち Microsoft ブランドを持つある商品に対して IsolationForest による外れ値検出を行い、ヒストグラムとして図に示した。

今後の予定として今回行った実験を進めてレビューの分類を行う。また、今回対象とできなかった特徴量についても実験を行い、有効性を検証する。

参考文献

[1] NTT レゾナンス.”「購買行動におけるクチコミの影響」に関する調査”. NTT コム リサーチ.2012/04/27. https://research.nttcoms.com/database/data/001436/(参照 2021/01/08)

[2] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating Fake News: A Survey on Identification and Mitigation Techniques. ACM Trans. Intell. Syst. Technol. 10, 3, Article 21 (May 2019), 42pages DOI:https://doi.org/10.1145/3305260

[3] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. S. (2013, July). What yelp fake review filter might be doing?. In Icwsm (pp. 409-418)

[4] R. Mohawesh et al, "Fake Reviews Detection：A Survey." in IEEE Access.vol.9,pp.65771-65802,2021.doi：10.1109 /ACCESS.2021.3075573

[5] Liu, Bing. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge university press, 2020.

[6] 曽我部東馬 著 曽我部完 監修 , 2021 , Python による異常検知 , 東京都：オーム社

[7] Savage, David, et al. "Detection of opinion spam based on anomalous rating deviation." Expert Systems with Applications 42.22 (2015): 8650-8657.

[8] Liu, Wenqian, et al. "A Method for the Detection of Fake Reviews Based on Temporal Features of Reviews and Comments." IEEE Engineering Management Review 47.4 (2019): 67-79.

[9] “sklearn.ensemble.IsolationForest”.scikit-learn. https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html,(2021-6-14)

[10] Liu Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolationforest." 2008 eighth ieee international conference on data mining. IEEE,2008.

[11] ヤン・ジャクリン.” Isolation Forest と異常検知（ネットアクセスログを用いて）”.GRIBlog.2020-07-06. https://griblog.hatenablog.com/entry/2020/07/05/171621,(2021-6-14)