P06　Sentence BERTを用いたスパムレビュー検出
－楽天市場レビューへの適用－

窪崎巧真，福澤和久（愛知工業大学）

1. 研究背景

ECサイトに投稿されるレビューには商品の評判を不正に操作しようとするものが存在する。
それらをスパムレビューと呼ぶ。

- ほとんどの消費者は一つの商品当たり数十件程度のレビューしか見ないため、スパムレビューの存在は有害
- 肯定的なレビューを書くことに対するインセンティブの提供の蔓延
- スパムレビューの検出は時間の掛かる作業であり、スパムレビューが投稿されてから削除までに時間が掛かってしまう
- 日本語のスパムレビューを対象とする研究は少ない

2. 提案手法

レビュー文を Sentence BERT[1] で768次元でベクトル化し、コサイン距離が0.08以下のレビューをスパムとする。
得られたベクトルとラベルから Embedding Projector[2] で分析。

データセット：楽天市場レビューの2019年のデータセットからUSBメモリ、キャットフード、インクカートリッジを抽出した

$$\text{コサイン距離} = 1 - \cos(\theta)$$

$$\text{コサイン類似度} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=0}^{n} A_i B_i}{\sqrt{\sum_{i=0}^{n} A_i^2}\sqrt{\sum_{i=0}^{n} B_i^2}}$$

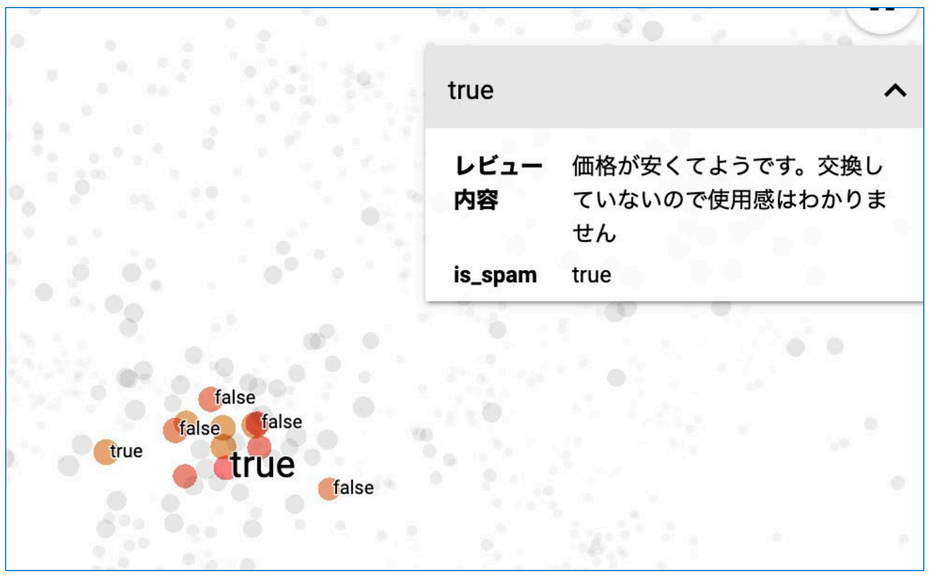
3. 結果・考察

■スパムレビュー検出の例

index	コサイン距離	レビュー文
463	0.0000	互換インクですが、なんら遜色なく使えますので大満足です。
1031	**0.0702**	**互換インクですが、何の不都合もなく使えます。 コストパフォーマンスも最高です。**
1033	**0.0702**	**互換インクですが、何の不都合もなく使えます。 コストパフォーマンスも最高です。**
24630	**0.0777**	**互換インクですが、全く支障がなく大満足です。**
32751	0.0810	互換性インクですが、全く問題なく使用できています。これで十分ですね！

■シンプルなスパム検出と提案手法による検出のスパム率の比較

ジャンル	合計レビュー数	シンプルなスパム		シンプルなスパム+加工されたスパム（提案手法）	
		スパム数	スパム率	スパム数	スパム率
USBメモリ	4807	139	2.89%	**535**	**11.13%**
キャットフード	20699	2396	11.58%	**4585**	**22.15%**
インクカートリッジ	43920	4143	9.43%	**15015**	**34.19%**



Embedding Projector による分析

4. 結論

- スパムレビュー投稿者(スパマー)は全く同じレビューを複数回に渡って投稿することが多い。これはレビュー数の多い商品の方が消費者が安心して買うからだと考えられる。
- しかし、全く同じレビューではシステム的に簡単に検出できる。そのためスパマーは少しだけ加工したスパムレビューを投稿する。多くは半角スペースを挿入したり、漢字を平仮名にしたり単語の言い換え表現にする。またはそれらを組み合わせたものを投稿する。提案手法はこれらを検出するものである。

参考文献
[1] Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, Nils Reimers and Iryna Gurevych, 2019, 1908.10084, arXiv, cs.CL
[2] Embedding Projector: Interactive Visualization and Interpretation of Embeddings, Daniel Smilkov and Nikhil Thorat and Charles Nicholson and Emily Reif and Fernanda B. Vi é gas and Martin Wattenberg, 2016, 1611.05469, arXiv, stat.ML

謝辞

本研究では、国立情報学研究所IDRデータセット提供サービスにより楽天グループ株式会社様からご提供頂いた「楽天データセット」(https://rit.rakuten.com/data_release/)を利用しました。心より感謝いたします。