

共学習によるレビュー文書からのネガティブな意見文の抽出

三戸尚樹¹ 古宮嘉那子² 佐々木稔³

¹東京農工大学 工学部 ²東京農工大学 工学研究院 ³茨城大学 工学部

s202436r@st.go.tuat.ac.jp, kkomiya@go.tuat.ac.jp, minoru.sasaki.01@vc.ibaraki.ac.jp

概要

本研究では、評点のみが付与された一般的なレビューを用いて、文レベルのネガティブな意見の抽出を行う。具体的にはレビューを文単位に分割してもとのレビュー全体の評点を付与し、それらのデータを利用して共学習を行うことでデータ不足を補い、ネガティブなレビューを抽出する。分類器にはBERTを利用し、対象商品には工具を用いた。本研究では、文単位の共学習を利用することにより、より性能の高い抽出を行うことができることを示す。

1 はじめに

評判分析とは、テキスト情報をテキストマイニングや機械学習の技術を用いて、肯定的な評価をしている記述と否定的な評価をしている記述に分類する技術である。SNS やブログ、EC サイトの普及により、ユーザーの生の声を企業が把握することができる評判分析は多くの企業や研究者から注目を集めている。

Amazon や楽天で商品を販売している企業にとって、購入者のレビューは商品改良の重要な手掛かりであるため、レビュー本文からネガティブな意見を抽出する事が求められる。ところが、(1) 評点は悪くないレビューの中にネガティブな意見が含まれている場合があり、そのような意見抽出は従来の評点の予測による評判分析の技術では抽出できない。また、(2) ネガティブなレビューの割合はポジティブなレビューに比べて少ないことが多いため、二値分類の単純な分類器では抽出が難しい、という問題がある。

そのため本研究では、評点のみが付与された一般的なレビューを用いて、文レベルのネガティブな意見の抽出を行う。本研究では、まずレビューを文単位に分割してもとのレビュー全体の評点を付与して疑似的な訓練事例とすることで、文単位のネガティブな意見抽出を行う。また、共学習を利用してデータ不足を補い、ネガティブなレビューを抽出す

る。

共学習とは、教師-生徒のように一方が指導し一方が学習するのではなく、自律した個同士が教え・教わる関係を同時に成立させている状態を表すモデルである。一方の知識レベル向上が他方の知識レベルの向上を促進することで相乗的な学習効果が期待できる[1]。

本研究では、Amazon と楽天の工機ホールディングスの製品のレビューデータからネガティブなデータの抽出を行った。Amazon のレビューデータのみで作成したモデルと楽天のレビューデータのみで作成したモデルを共学習させる事により、一つの分類器で学習したモデルでの精度よりもネガティブなデータの抽出の精度が高くなる事を示す。

2 関連研究

Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019)[1]とは、2018年10月にGoogleのJacob Devlinらの論文で発表された事前学習モデルである。BERTはTransformer[2]のEncoderを使ったモデルである。文章の文脈を読むことが可能なモデルであり、文書分類や質問応答など、言語処理の多種多様なタスクにおいて当時の最高スコアを叩き出した。GRUとBERTを使用し、略語やスラング、スペルミス等のデータとしてノイズが多いTwitterの評判分析を行っている研究もある[3]。

3 共学習によるデータ選別を利用した評判分析

た評判分析

本研究では、評点のみが与えられたレビューの文集合から、ネガティブな意見抽出を行う。そのため、まずレビューを文単位に区切り、それぞれの文に暫定的にレビューの評点を付与する。この際、評点が1と2についてはネガティブ、評点が3、4、5についてはポジティブ（非ネガティブ）とした。

これにより文ごとにおおまかに正しいポジティブ/ネガティブの評点のついたデータが生成できるが、評点は悪くないレビューの中にネガティブな意見が含まれている場合や、その逆の場合には不正確なデータが混じった状態になる。また、ネガティブな意見は全体のレビューの中で少ない傾向があり、二値分類の単純な分類器では抽出が難しい。そこで本研究では、共学習を用いてデータ不足を解消しつつ、より正しいと思われる評点のついた文データを訓練事例に足してゆくことで、これらの問題を解決することを試みる。

本研究では Amazon のレビューデータ[5]から作成した BERT モデルと楽天のレビューデータ[6]から作成した BERT モデルの 2 つのモデルを使って共学習を行った。

本研究では、Amazon のレビューデータからのみ訓練させた BERT モデルと楽天のレビューデータのみで訓練された BERT モデルの 2 つのモデルで共学習を行った。共学習は以下の手順によって行う。

(1) Amazon データと楽天データを訓練データとテストデータに分ける

(2) モデルを二つ作成する。この際、一つ目のモデルは Amazon の訓練データのみ、ふたつ目のモデルは楽天の訓練データのみで学習を行う。また、ハイパーパラメータの調整には、一つ目のモデルには楽天の訓練データを、ふたつ目のモデルには Amazon の訓練データを利用する。

(3) モデルにテストデータを適用する。この際、一つ目のモデルには楽天のテストデータを、ふたつ目のモデルには Amazon のテストデータを適用する。

(4) テストデータのうち、高い確信度でネガティブであると判定されたデータを抽出する

(5) (4)で抽出したネガティブなデータをネガティブな訓練データとしてもとの訓練データに追加する。また、この際に、ポジティブな訓練データをランダムに選択して同時に訓練データに追加する。この際、ポジティブなデータは推論結果ではなく、レビューの評点を利用して追加した

(6) これらのデータを訓練データとして、次のモデルを学習する

(7) (3)-(6)を繰り返す

(8) ふたつのモデルの最終的な訓練データを合計して、新しいモデルを学習し、これを最終的なモデルとする

4 データ

本研究は、ネガティブな意見を文単位で抽出することを目的とする。実験に際して、文ごとのアノテーションつきデータを含む対象商品のレビューと文ごとのアノテーションのない別商品のレビューデータの二つを使用した。

まず、対象商品のデータは工機ホールディングスの商品と工機ホールディングスの競合企業の商品の Amazon と楽天のレビューデータである。工機ホールディングスの Amazon データ 515 件、楽天データ 234 件、競合企業の Amazon データ 770 件、楽天データ 820 件から成る。その中でも極性とレビュー本文のセットが 1,963 件、極性とレビュー本文、工機ホールディングスが抽出したいと考えているネガティブな文がアノテーションされているデータが 376 件である。極性ごとのデータ件数を表 1 に示す。アノテーションは工機ホールディングスの社員によって行われた。

表 1 対象商品データの極性の分布

	Amazon	楽天
極性 1	70	17
極性 2	22	5
極性 3	61	92
極性 4	205	143
極性 5	927	797
合計	1,285	1,054

表 2 レビューデータの極性の分布

	Amazon	楽天
極性 1	1,000	1,538
極性 2	1,000	1,176
極性 3	1,000	3,866
極性 4	1,000	11,967
極性 5	1,000	28,070
合計	5,000	46,617

次に、文ごとのアノテーションのないレビューデータは、ランダムに抽出した Amazon データ 5,000 件と工機ホールディングスが販売している製品と同

じジャンルの工具の楽天データ 46,617 件である。Amazon データのジャンルについては、ベビー用品など多岐にわたる。極性ごとのデータ件数を、表 2 に示す。

5 実験

BERT の訓練には、最適化関数は SGD、損失関数はクロスエントロピー誤差を用いた。

Amazon のレビューデータは、訓練データを 1,165 件、テストデータを 3,001 件に分けた。楽天データは、訓練データ 2,000 件、テストデータを 1,730 件に分けた。そのうえで、レビューデータを文レベルに分割し、レビューの評点を文に付与した。ベースラインの文レベルの評点の分布およびポジティブとネガティブのデータ件数を表 3 に示す。なお、a はアノテーションでネガティブとラベル付けしたデータである。また、列を示す 1 から 5 は五分割交差検定のテストデータを示す。レビューごとに評点を付与した後、データ件数が均等になるように五分割交差検定のデータセットを分けたため、アノテーションでネガティブとラベル付けしたデータには偏りが生じている。なお、訓練データのネガティブの割合は、5.52%から 8.86%になる。

表 3 ベースラインの文レベルの評点の分布

	1	2	3	4	5
a	128	19	0	0	0
評点 1	44	50	74	21	19
評点 2	11	16	37	5	18
ネガ	183	85	111	26	37
評点 3	21	63	50	99	182
評点 4	165	183	142	181	134
評点 5	804	842	870	867	823
ポジ	990	1088	1062	1147	1139
合計	1,173	1,173	1,173	11,73	1,176

また、提案手法において、データを追加する前の文レベルの評点の分布とポジティブとネガティブのデータ件数を表 4 に示す。このデータを 1 回目のモデルの訓練事例とした。なお、訓練データのネガティブの割合は、Amazon データは 15.98%、楽天データは 7.34%になる。

表 4 共学習の 1 回目のモデルの訓練データに使用した文レベルの評点の分布

	Amazon	楽天
評点 1	284	221
評点 2	283	158
ネガ	567	379
評点 3	948	535
評点 4	1,011	1336
評点 5	1,023	2,912
ポジ	2982	4783
合計	3,549	5,162

共学習を行う上でポジティブなデータの足し方として、以下の 2 つの手法で実験を行った。

- (1) 事前確率の割合でデータを足す手法
- (2) 同数のデータを足す手法

(1)事前確率の割合でデータを足す手法では、例えば Amazon のみで訓練を行った BERT モデルがネガティブであると推論した件数が 100 件だった場合、事前に訓練させた楽天の BERT モデルと同じ割合 (1:9) を保つために、900 件のポジティブのデータを追加した。(2)同数のデータを足す手法では、例えば Amazon のみで訓練を行った BERT モデルがネガティブであると推論した件数が 100 件だった場合、100 件のポジティブのデータを追加した。モデルの作成は三回繰り返した。つまり、データの追加は二度行った。

推論時の確信度は、softmax 関数のネガティブの値が 0.8 以上のデータを抽出した。

提案手法の訓練には、評点のみ付与された Amazon および楽天のレビューデータを用い、評価する時のみ、対象商品のデータを用いた。これに対し、本研究のベースラインには、対象商品のデータで五分割交差検定を行ったものを用いる。まず、文に分割し、文にはレビューの評点を付与した。さらに、376 件のアノテーションつきデータについては、レビューの評点にかかわらずネガティブなデータとして使用した。

また、モデルにはベースラインを含めて、どれも BERT を fine-tuning して利用した。ベースラインおよび共学習を行った全てのモデルのハイパーパラメータは、0.0001 となった。

6 結果

ベースラインの結果を表5に、事前確率の割合でデータを足した手法の結果を表6に、また、同数のデータを足した手法の結果を表7に示す。表のEPOCHは30エポック回した中のF値が最高だったときのエポック数である。またその時点での精度、再現率、F値を表に示した。表5-7から、ふたつの提案手法の精度、F値がベースラインを超える事を確認できた。また、同数のデータを足した場合には再現率もベースラインを上回った。

表5 ベースラインの結果

	EPOCH	精度	再現率	F 値
1	15	42.08%	42.54%	0.42
2	21	47.06%	55.56%	0.50
3	14	45.05%	57.47%	0.50
4	3	38.46%	30.30%	0.33
5	30	0.00%	0.00%	0.00
平均		34.53%	37.17%	0.35

表6 事前確率の割合で足した場合の結果

	EPOCH	精度	再現率	F 値
1	5	64.18%	23.50%	0.34
2	5	43.55%	31.76%	0.37
3	15	49.33%	33.33%	0.40
4	7	58.33%	26.92%	0.37
5	15	39.39%	35.14%	0.37
平均		50.96%	30.13%	0.37

表7 同数のデータを足した場合の結果

	EPOCH	精度	再現率	F 値
1	20	49.64%	37.70%	0.43
2	30	48.68%	43.53%	0.46
3	17	39.68%	45.05%	0.42
4	12	43.75%	26.92%	0.33
5	4	39.58%	51.35%	0.45
平均		44.27%	40.91%	0.42

7 考察

提案手法では、文レベルでアノテーションされたテストデータの情報は一切用いていないのに対し、ベースラインではこれらの用例を五分割交差検定に

用いている。それにもかかわらず、提案手法のF値はベースラインを上回った。このことから、評点のみが付与されているレビューデータからのネガティブな意見抽出のタスクに共学習を用いた提案手法が有効であると分かる。また、事前確率の割合でポジティブなデータを追加した場合より、ネガティブなデータと同数のポジティブなデータを追加した場合の方がF値が高かった。これは、偏りのある事前確率の割合でポジティブデータを足したため、ネガティブなデータの割合が非常に低くなった結果、再現率が低くなったためであると考えられる。一般に、精度と再現率はトレードオフの関係にあるが、今回の実験ではこの偏りの影響をうまく回避できたため、同数のデータを足した場合のF値が他手法を上回ったと考えられる。

共学習を終えた時の事前学習の割合でデータを足した場合と同数のデータを足した場合のデータのネガティブとポジティブの分布を表8に示す。同数のデータを足した場合にはネガティブなデータの割合が高まっているのが見て取れる。レビューの評点を文に付与した「おおよそ正しいデータ」をもとに、信頼性の高いネガティブデータを抽出し、なおかつ同数のポジティブデータを足すことで共学習において正解率を高めることができたと考えられる。

表8 共学習を終えた時のネガ/ポジの分布

	事前学習の割合でデータを足した場合	同数のデータを足した場合
ネガ	1,104	1,715
ポジ	10,779	8,534
合計	11,883	10,249
ネガの割合	9.29%	16.73%

8 おわりに

本研究では、Amazonと楽天のレビューデータから工機ホールディングスが抽出したいネガティブなデータの抽出を行った。その結果、共学習を利用したBERTモデルは、ベースラインを超える精度でデータ抽出を行える事が明らかになった。今後は、精度の向上とポジティブのデータ抽出方法について検討する必要がある。

謝辞

工機ホールディングスの西河智雅様、吉崎陽祐様には、本研究のベースライン、評価データで使用するデータを準備していただきました。また、楽天データにつきましては、国立情報学研究所の楽天市場データを利用させていただきました。この場を借りて御礼申し上げます。本研究は JSPS 科研費 17H00917, 17KK0002, 18K11421 の助成を受けたものです。

参考文献

1. Anders Sogaard, *Semi-Supervised Learning and Domain Adaptation in Natural Language Processing (Synthesis Lectures on Human Language Technologies)*, Morgan & Claypool Publishers, 2013
2. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Jacob

Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, *Computation and Language* 2018

3. Transformers for Language Understanding, Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, In *Proceedings of NAACL-HLT 2019*, pp. 4171-4186, (2019)

4. GRUBERT A GRU-Based Method to Fuse BERT Hidden Layers for Twitter Sentiment Analysis, Leo Horne, Matthias Matti, Pouya Pourjafar, Zuowen Wang, *Association for Computational Linguistics* 2020, pp. 130-138, (2020)

5. Amazon データセット
<https://s3.console.aws.amazon.com/s3/buckets/amazon-reviews-ml?region=us-west-2&prefix=json/&showversions=false>

6. 楽天グループ株式会社 (2021): 楽天市場データ. 国立情報学研究所情報学研究データリポジトリ. (データセット).
<https://doi.org/10.32130/idr.2.1>