

明示的及び暗黙的データを活用した Twitter における 体験的ストレスの特定

飯田静空 原一夫
山形大学 理学部

s180572@st.yamagata-u.ac.jp kazuo.hara@gmail.com

概要

ストレスは社会問題となっており、Twitter 上では多くの人の愚痴や悩みが日々呟かれる。しかし、ストレスの表現方法は人によって異なるため、どれほどの人がストレスを抱えているのかを把握することは困難である。そこで本研究では、ツイートをした人自身が経験したストレス(体験的ストレス)を示すツイートを検出する分類器を構築した。具体的には、「ストレス」という言葉を明示的に含むデータと、ストレスの内容を暗黙的に含むデータの2種類を利用することでBERTを訓練した。実験の結果、我々の手法は単体データのみの学習よりも高い精度で体験的ストレスを検知でき、モデルが両データの特徴を学習できていることを分析により示した。

1 はじめに

ストレス社会という言葉があるように、現代でストレスは深刻な社会問題となっている。Twitterをはじめとするマイクロブログ上でも多くの人々の不平不満や心配事が日々散見され、最近ではCOVID-19の影響によるストレスの増加が確認されている[1]。

その一方で、Twitter上において何人のユーザーがストレスに苦しんでいるのかを我々が把握することは困難である。なぜなら、ツイート(以下「発言」と呼ぶ)におけるストレスの表現方法は人によって異なるからである。多くのユーザーは、「ストレス」という言葉を使わずに自身のストレス状態を報告する(図1 発言例1)。そのため「ストレス」という言葉が明示的に書かれた発言のみを数えたとしても、ユーザーが経験している全ストレスを考慮することはできない。さらに「ストレス」という言葉が明示的に書かれていたとしても、発言をしたユーザー自身が経験したストレス(以下「体験的ストレス」と呼ぶ)を必ずしも表すとは限らない(図1 発言例2)。

発言例1.) 上司に怒られてイライラする怒

「ストレス」という言葉を含んでいないが、上司に対しての怒りを表しているため、ストレスを感じている可能性がある。

発言例2.) 猫ちゃんが心配... 昨日からストレスのせいか体調が悪いみたいで早く帰る

「ストレス」という言葉を含んでいるが、文脈的にストレスを受けているのは猫である。

図1 発言のストレス表現が異なる例

本研究では、Twitterに現れる体験的ストレスの量を総合的に把握するために、発言中に体験的ストレスが含まれるかどうかを判断する分類器を構築する。そのために我々は、「ストレス」と「#ストレス」の言葉を含む発言から作成した2種類のデータを利用した。一つは「ストレス」という言葉を明示的に含むデータで、「今日も仕事か...とてもストレスなんだが…」のような発言である。これを**明示的データ**と呼ぶ。もう一つはストレスの内容を暗黙的に含むデータで、「隣マジでうるさいからキレそう」のような発言である。これを**暗黙的データ**と呼ぶ。我々は全ての発言から体験的ストレスを検出するために、「ストレス」の言葉を含まない暗黙的データから体験的ストレスを分類する実験を行った。しかし、暗黙的データはその性質上、データ数を増やすことが難しい。そこで我々は、収集が容易な明示的データも使い、2つのデータの特徴をBERT[2]にfine-tuningで学習させる方法を提案した。実験と分析の結果、我々の提案手法は暗黙的データのみを使用した場合よりも高い精度で体験的ストレスを分類でき、両データの特徴をモデルが学習できていることを示した。

2 関連研究

ストレスは、人や環境との相互作用によって引き起こされる苦痛の認識と定義される [3]。ストレス

には主観的な側面があるため、物理的信号である脳波[4]や音声データ[5]などの特徴でストレスを調査する研究がある。しかし、物理的な測定には莫大なコストや時間がかかる。加えて、ストレスを感じる期間は人によって異なるため[6]、この方法では生活で時々刻々と変化する精神状態の把握は難しい。

このような理由から、最近ではマイクロブログ上に現れるストレスを分析する研究が行われている。Doan ら[7]は Twitter 上の明示的データで訓練した SVM を使うことで、米国の地理ごとにストレス分析を行った。他方で、Turcan ら[8, 9]は Reddit という米国で人気のあるマイクロブログから暗黙的データを作成した。そして BERT[2]を使い、感情分類とのマルチタスク学習を行うことで体験的ストレスの分類精度を向上させた。しかし、これらの研究では明示的データか、暗黙的データかのどちらか一方のデータしか使用してはいない。それらとは対照的に、我々は両方のデータを活用することでその有効性を示す。

3 データ

3.1 体験的ストレスの定義

分類器を構築するために、我々はデータを作成する必要がある。そのために、本研究の分類基準である体験的ストレスの定義を行う。本研究では、以下の2点を満たす発言を体験的ストレスと判断した。

1. ストレスを受けた対象が発言の作者であること

「ストレス」という言葉で、発言者自身のストレスが明記されてあることとした。ストレスを受けた対象が「発言者以外の人」、「人以外の生物」、またはその発言内に「引用」された形で他者がストレスを示す発言は除かれる。

2. 発言をした時点でストレスに感じていること/ストレスの体験性を損なわないこと

発言者が発言した時点でストレスに感じていることとした。発言者が発言した時点の時制から考え、「過去」、「未来」、「習慣」を表す発言を除く。また、ストレスの体験性を損なう「仮定」、「疑問」、「推論」、「共感」、「願望」、「呼びかけ」、「応答」、「助言」、「否定」、「噂」、「主張」を表す発言も除く。

例)

今は**ストレス**で体調死んだら(体験的ストレス)

みんなは**ストレス**なのかな?(非体験的ストレス)

3.2 明示的及び暗黙的データの作成

始めに、明示的データを作成した手順を述べる。我々は Twitter API を使い、2021 年 7 月 29 日 02:21:00 から 09:42:56 の間で「ストレス」という言葉を含む発言を収集した。その結果、6535 の発言を取得することが出来た。その後、集めた発言に前処理を行った。方法としてはリツイートと重複した発言の削除、URL 部分の削除と、半角、全角スペースの切り詰め、リプライを表す発言では「@リプライ先のユーザー名」までの部分を削除するといった処理である。そして、3.1 で述べた基準のもと体験的ストレスを含む発言には 1、含まない発言には 0 のラベル付けを一人のアノテータが行った。ラベル付けには質問紙[10]を使って判断する方法などがあるが、本研究では「ストレス」という言葉の明記を、ストレス反応の根拠として信頼した。

続いて、暗黙的データを作成した手順を述べる。我々は明示的データと同じ方法で、2021 年 10 月 1 日 00:00:00 から 27 日の 23:59:59 の期間で「#ストレス」を含む発言を収集した。その結果、5861 の発言を取得することが出来た。その後集めた発言に前処理を行った。前処理の方法は明示的データの場合と共通しているが、今回は「ストレス」という言葉を含んでいないデータを作成したいため、新たに次の処理を追加した。まず、「#」部分はストレスを直接示す情報となるため、「#」以下の部分を削除した。次に、収集したデータに「ストレス」の単語を含んだ発言も存在したため、この単語の部分も削除した。これにより文脈が不自然になった発言は、不自然な表現部分を削除することで文脈を自然な形に直した。

前処理前の例) 最近、コロナで旅行いけないの本当に**ストレス!!** #ストレス#旅行

前処理後の例) 最近、コロナで旅行いけない!!

前処理をした後、明示的データの場合と同じ人物がラベル付けを行った。アノテータは 3.1 の定義をもとにラベルを付けるが、今回はデータに「ストレス」という言葉が含まれていない。そのため定義の 1 は、発言内容が作者の体験であるという基準に変更した。

作成したデータの内訳を表 1 に示す。明示的データは、短期間で比較的数の多いデータを作成できた。しかし、暗黙的データでは発言内容の重複が原因で、収集した発言の 81%が削除される結果となった。

表 1 作成したデータの内訳

データ	ラベル 1	ラベル 0	合計
明示的データ	2203(36.1%)	3897(63.9%)	6100
暗黙的データ	622(56.8%)	473(43.2%)	1095

表 2 明示的データでの分類実験の結果

Model	precision	recall	F1	accuracy
Naïve Bayes	65.79	72.82	69.12	78.03
BERT _{subword}	80.71	75.95	78.15	85.69
BERT _{char}	78.66	76.16	77.24	84.89

4 実験

4.1 明示的データによる分類実験

4.2 の提案手法で使用するモデルを選定するために、我々はまず明示的データのみを使い、体験的ストレスである発言を検知する実験を行った。モデルには、Transformer[11]に基づくニューラルネットワークモデルの BERT[2]を使用した。発言をサブワードと文字単位に分割し、BERT へ入力する場合をそれぞれ比較した。前者を BERT_{subword}、後者を BERT_{char} と名付ける。ベースラインには Naïve Bayes を選んだ。データは 8:2 の割合で訓練用とテスト用に分け、10 分割交差検証によってハイパーパラメータを調整した。ハイパーパラメータの調整は、BERT の最終層と全結合層の間の Dropout を {0.3, 0.5}、学習率を {1e-5, 2e-5, 3e-5, 4e-5, 5e-5} の中からグリッドサーチにより行った。Optimizer には AdamW[12]を使い、バッチサイズは GPU の制限上 64 とした。10 エポックが終了した時点でのモデルをテスト用のデータを使い、precision, recall, F1, accuracy の指標から評価した。BERT のモデルには、東北大学が公開している事前学習済みモデルを使用した。

テストデータにおける各モデルの最良な結果を表 2 に示す。まず、BERT のモデルは Naïve Bayes よりも全ての評価指標でより良い結果であることが分かる。BERT は文脈を考慮できるモデルとされており、独立な仮定に基づくモデルよりも複雑な分類が可能であると考えられる。また BERT_{subword} は BERT_{char} と比べ、recall 以外の全ての指標でより良い結果であった。この実験では、BERT へサブワードに分割したテキストを入力とする方が適していたと考えられる。

4.2 暗黙的データによる分類実験

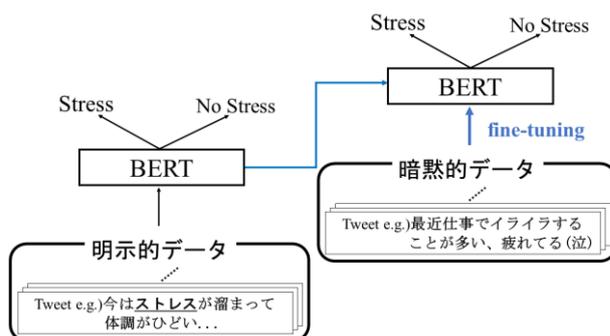


図 2 提案手法

4.1 では、明示的に「ストレス」と書かれてある発言を分類できるモデルを構築した。しかし、我々の目標は Twitter 上の全発言に出現する体験的ストレスを検知することである。そのため、次に暗黙的データを使った分類実験を行う。

暗黙的データは 3.2 で述べた発言内容の重複という理由から、データ数を増やすことが難しい。そこで我々は、明示的及び暗黙的データの両方を使う手法を提案する。図 2 に我々の提案手法を示す。まず BERT を明示的データにより訓練することで、訓練済みモデルを出力する。次にこのモデルを暗黙的データで fine-tuning を行い、BERT を再訓練する。ここで出力されたモデルを BERT_{proposed} と名付ける。

ベースラインに暗黙的データのみを使って学習した BERT を BERT_{baseline} として使用した。BERT_{proposed} は、4.1 で最良の結果であった BERT_{subword} を暗黙的データにより fine-tuning することで構築した。データは 5:5 の比率で訓練用とテスト用に分け、5 分割交差検証でハイパーパラメータを調整した。ハイパーパラメータ、Optimizer、及びバッチサイズは明示的データの場合と同じ設定である。最大を 10 エポックとして、最も高い accuracy のエポック時点の各モデルを評価した。

表 3 は各モデルの最良の結果を示している。BERT_{proposed} は BERT_{baseline} よりも全ての指標でより良い結果であることが分かる。次に BERT_{proposed} が学習によりどのような特徴をとらえたのかを分析する。

4.3 分析

我々は LIME[13]と J-LIWC[14]により、BERT_{proposed} が学習後に着目した部分の変化に関する分析を行った。LIME とは、あるモデルへの入力ベクトルの空間に摂動を与え、出力に対しての局所的な線形近似を行うことで、入力語の影響度を測定できるモデルである。J-LIWC は心理学者による信頼性や妥当性検

表 4 学習前後で BERT が判断根拠とする項目 (上から LIME の出力語を多く含む順に並んでいる)

順位	BERT _{proposed} (ラベル 1)	BERT _{subword} (ラベル 1)	BERT _{proposed} (ラベル 0)	BERT _{subword} (ラベル 0)
1	機能語	機能語	機能語	機能語
2	相対性	感情	格助詞	格助詞
3	時間	ネガティブ感情	相対性	感情
4	動詞	生物学的プロセス	認知プロセス	助動詞
5	認知プロセス	健康	感情	動因
6	格助詞	不安	動因	相対性
7	感情	相対性	助動詞	認知プロセス
8	助動詞	格助詞	社会的(相互作用)	動詞
9	ネガティブ感情	認知プロセス	動詞	ポジティブ感情
10	動因	動詞	ポジティブ感情	社会的(相互作用)

表 3 暗黙的データでの分類実験の結果

Model	precision	recall	F1	accuracy
BERT _{baseline}	85.51	86.51	86.26	83.98
BERT _{proposed}	88.76	89.47	89.11	87.12

証がされている言語, 心理的項目を含んだ日本語の辞書である。4 で学習後の BERT_{subword} と BERT_{proposed} で, ラベル 1 と 0 の判断根拠となる LIME の出力スコアの高い 5 語を各データから抽出し, J-LIWC の各項目ごとに含まれる語を数え, 多い順に項目を算出した。データには, 4.1 と 4.2 で分割したテスト用をそれぞれ使用した。表 4 はその結果を示している。

まず, ラベル 1 を判断する際のトップ 10 に入った項目の比較として, BERT_{subword} では「生物学的プロセス」, 「健康」, 「不安」, BERT_{proposed} では「時間」, 「助動詞」, 「動因」の項目に違いが生じた。BERT_{subword} でのこれらの項目順位が高い理由は, いずれも「ストレス」という単語を項目内の語として含んでいることが挙げられる。明示的データではどの発言にも「ストレス」という言葉が出現するため, 我々はこの単語が判断根拠にはなりづらいと予想していた。しかし, 実際にこれらの項目の詳細な内訳を見ると「ストレス」の出現数が項目中の半分以上を占めていた。これは, 同じ単語でも BERT が文脈に応じて判断根拠とするか否かを使い分けているためであると我々は考えている。BERT_{proposed} で「時間」の項目の詳細な内訳を見ると, 「今日」「た」「最近」が使用頻度の高い語であった。これは BERT_{subword} においても同様であった。しかし, BERT_{proposed} では「今日」という単語の割合が BERT_{subword} よりも 6% ほど上昇しており, 学習前後での細かな違いが確認された。このように変化した理由の詳細な分析は今

後の課題であるが, 「時間」の項目語にラベル 1 の判断根拠としてより重みがついたことが分かる。

BERT_{proposed} では「助動詞」の項目がラベル 1 で 8 番目, ラベル 0 で 7 番目に高かった。一方で BERT_{subword} ではラベル 0 の判断根拠にしかなくなっておらず, 暗黙的データでは助動詞をラベル 1 の判断根拠にする割合が強まった。ラベル 1 の「動因」では, 共通して「仕事」という言葉が一番多かった。これは, 仕事に関連した語がストレスを示す傾向があるという先行研究と一致する[15]。また, 「悪い」「怖い」といったネガティブな動因が占めることも共通していた。しかし, BERT_{proposed} で二番目に多い単語が「旦那」であり, 妻が夫の愚痴を言う発言の特徴を新たに学習したことが分かった。ラベル 0 でトップ 10 の項目は全て一致しており, 順位が若干変動する結果となった。ラベル 0 の基準は両データで同じであったため, これは妥当な結果である。以上のように, モデルが両データの共通, 非共通な特徴を学習したことが確認された。

5 おわりに

本研究では, 明示的に「ストレス」と書かれてある発言と暗黙的にストレスの内容を含む発言の両データを使い, 発言者がストレスに感じている発言を検出する分類器を構築した。実験と分析の結果, 我々の手法は単体データのみの場合よりも高い精度で発言者のストレスを分類可能であり, 両データの特徴を学習することが有効であることを示した。

今後の研究として, 「ストレス」, 「#ストレス」の言葉を含まず, ストレスの自覚なしに発せられた発言に対しても, この手法で構築した分類器が適用可能かどうかを調査していきたい。

参考文献

1. Sachin Thukral, Suyash Sangwan, Arnab Chatterjee, and Lipika Dey. Identifying pandemic-related stress factors from social-media posts--effects on students and young-adults. arXiv preprint arXiv:2012.00333. 2020
2. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proc. of NAACL, pp.4171-4186. 2019.
3. Sheldon Cohen, Gregory E. Miller, and Bruce Rabin. Psychological Stress and Antibody Response to Immunization: A Critical Review of the Human Literature. *Psychosomatic Medicine*, Vol. 63 - Issue 1 - pp.7-18. 2001.
4. Fares Al-Shargie, Masashi Kiguchi, Nasreen Badruddin, Sarat C. Dass, and Ahmad Fadzil Mohammad Hani. Mental stress assessment using simultaneous measurement of eeg and fnirs. *Biomedical Optics Express*, Vol. 7 - Issue 10 - pp.3882-3898. 2016.
5. Xin Zuo, Tian Li, and Pascale Fung. A multilingual natural stress emotion database. In Proc. of LREC, pp.1174-1178. 2012.
6. American Psychological Association. What's the difference between stress and anxiety? (引用日: 2022 年 1 月 8 日.) <https://www.apa.org/topics/stress/anxiety-difference>.
7. Son Doan, Amanda Ritchart, Nicholas S Perry, Juan D Chaparro, and Mike Conway. How Do You #relax When You're #stressed? A Content Analysis and Infodemiology Study of Stress-Related Tweets. *JMIR Public Health Surveill*, Vol. 3 - Issue 2 - e35. 2017.
8. Elsbeth Turcan, and Kathleen McKeown. Dreddit: A Reddit Dataset for Stress Analysis in Social Media. In Proc. of EMNLP, pp.97-107. 2019.
9. Elsbeth Turcan, Smaranda Muresan, and Kathleen McKeown. Emotion-Infused Models for Explainable Psychological Stress Detection. In Proc. of NAACL, pp.2895-2909, 2021.
10. Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C. Eichstaedt, and Lyle H. Ungar. Understanding and measuring psychological stress using social media. CoRR, abs/1811.07430, 2018.
11. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pp.6000-6010. 2017.
12. Ilya Loshchilov, and Frank Hutter. Decoupled Weight Decay Regularization. arXiv preprint arXiv:1711.05101. 2017
13. Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why should I trust you?": Explaining the predictions of any classifier. In Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135-1144. 2016.
14. Tasuku Igarashi, Shimpei Okuda, and Kazutoshi Sasahara. Development of the Japanese Version of the Linguistic Inquiry and Word Count Dictionary 2015 (J-LIWC2015). PsyArXiv Preprints: 10.31234/osf.io/5hq7d. 2021
15. Huijie Lin, Jia Heinie, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, and Ling Feng. User-Level Psychological Stress Detection from Social Media Using Deep Neural Network. In Proc. of ACM MM, pp.507-516. 2014.

A 付録

A.1 予備調査

我々は、「ストレス」と「#ストレス」という言葉を含んだ各発言の一日あたり収集可能な発言量を調査をした。まず、Twitter API を使用して2021年10月1日00:00:00から27日23:59:59までの期間で「ストレス」と「#ストレス」という言葉を含んだ各発言を収集した。その後、日ごとに発言量の集計を行った。図3と図4はその結果の棒グラフを表している。

縦軸の目盛りが100倍異なることから分かる通り、「ストレス」と「#ストレス」の発言では、一日あたりの発言量が大きく異なる。「ストレス」という言葉を含んだ発言は一日に平均して37,738の発言を収集できたが、「#ストレス」の場合は一日に平均して217の発言しか収集することができなかった。このことから、「ストレス」の言葉を含んだ発言の方が、短期間で大規模な量のデータを収集出来ることが分かる。

A.2 ハイパーパラメータ

4.1と4.2の実験におけるハイパーパラメータのチューニング結果を表5と表6に示す。

表5 4.1におけるハイパーパラメータ
チューニングの結果

パラメータ	BERT _{subword}	BERT _{char}
Dropout	0.5	0.5
学習率	3e-5	5e-5

表6 4.2におけるハイパーパラメータ
チューニングの結果

パラメータ	BERT _{baseline}	BERT _{proposed}
Dropout	0.5	0.5
学習率	5e-5	3e-5

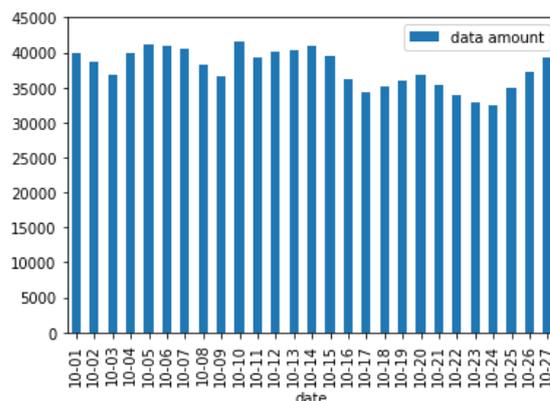


図3 「ストレス」の言葉を含む発言における一日あたりの発言量

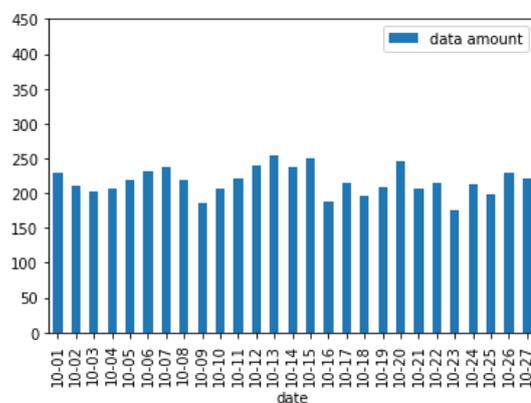


図4 「#ストレス」の言葉を含む発言における一日あたりの発言量