

チャット悩み相談に寄せられた相談内容の深刻度判定とその成果の相談事業への応用

代表研究者 上田路子

早稲田大学 政治経済学術院 准教授

1. 研究目的

日本をはじめ、世界中の国において自殺は最も深刻な社会問題の一つである。死にたいほどつらい悩みを抱える人たちの自殺念慮の背景にある悩みに寄り添って自殺以外の解決手段を提示し、適切な支援につなげることは重要な社会的課題である。

自殺や生きづらさ全般に関する悩みを傾聴し必要な援助に結びつけるための相談事業は昔から存在し、近年は若者向けに SNS やチャットを用いた相談事業（以下 SNS 相談）が多く実施されている。しかし、従来の電話相談に限らず、SNS を用いた事業であっても相談への対応が追いついていないのが現状である。特に 2020 年春以降は新型コロナウイルスの影響により、各種相談機関に寄せられる相談件数が急増しており、支援を必要とする相談者に対応が行き届かない状況が続いている。その結果、悩みを抱えた相談者は相談に乗ってもらうまで長時間待つことを余儀なくされているのが現状であり、このような状況下では自殺を企図しながら連絡をしている相談者や、強い自殺念慮を持つ相談者など緊急性の高い相談者に即座に対応することは困難である。緊急に介入する必要がある相談者をできるだけ早く特定し、優先的に対応することは喫緊の課題であるが、そのような仕組みは技術的な理由から日本の相談事業では現在実装されていない。そこで、本研究は自然言語処理の技術を用いてチャット相談事業に寄せられる相談者の自殺リスクの緊急度や悩みの深刻度を判定する手法を開発することを主な目的とする。この研究開発を通じて、深刻な相談事例をリアルタイムで判定し、優先的に対応ができるようなシステムが実際の相談事業で実装することにつながると期待される。

2. 研究手法

本研究では、最終的には実際の SNS 相談のデータを対象に LSS を用いて個別発言の深刻度の推定を行なったが、SNS 相談のテキストデータは外部に出せるデータではないため、本報告書では開発過程で実施したツイッターデータを対象とした分析結果を紹介する。分析対象が SNS 相談のテキストデータであっても、ツイートであっても分析手法は同じであり、本稿で紹介する研究成果は SNS 相談に応用可能である。ただし、ツイッターデータと SNS 相談データの主な違いとして、ツイートにはリツイートが多く含まれること、そして不完全な文章や短い文章が多いという特徴が挙げられる。リツイートについては、投稿者の感情や精神状態を直接的に示しているとは考えにくいことから、削除をした上で分析を行なった。また、主なニュース記事も削除した上で分析をした（後述）。また、ツイッターには文字数制限がある一方、SNS 相談にはそのような制限がないことから、前者のほうが完全でない文章が多く含まれる傾向にあるが、実際の分析の際にはその点は特に問題とはならず、ツイートを対象とした場合でも SNS 相談を対象とした場合と同程度の精度を得ることができた。

2-1 データ収集方法の概要

本研究の最終的な目的は発言内容から発言者の状態や悩みの深刻度を推定することである。そのため、手法の開発段階では発言者の発言、及び検証データとしてその人の状態（主に精神的状態）の両方が必要である。そこで、検証データを入手するために、ツイッターの一般ユーザーを対象とした大規模なサーベイ調査を実施し、解析を行なった。サーベイでは回答者の属性や精神的状態とともにツイッターの公開アカウント名を尋ね、同意が得られたユーザーからはツイッター社の API を用いて過去及び将来の投稿を収集し分析対象とした。

サーベイ調査の結果とツイッター上の発言を組み合わせる研究手法はまだ確立されていないため、本研究では、まずは研究代表者らが同様の手法を用いて過去に収集したパイロット（テスト）サーベイ調査（2020 年 4 月に実施）を通じて得られたデータの品質チェックを行った。この調査では、回答者の基本属性を収集するとともに、サーベイの最後にツイッターを使っているかを尋ね、もし使っているのであれば、研究目的でのツイートの収集への同意を求めた。その上

で、同意が得られた回答者のみを対象にツイッターのアカウント名を尋ね、その結果得られた情報をもとに API を用いてそれぞれの回答者のツイートを過去（2019 年 1 月 1 日）から遡って収集した。この過程でアカウントが存在しているかどうかを判明するため、存在しないアカウント名を入れた回答者については分析対象から外した。

データの品質チェックの具体的な作業としては、研究補助者に一つ一つのアカウントの投稿内容の確認を依頼し、回答者が自己申告したアカウントが企業のものであるケース、日本語の投稿が一切ないケースなど、明らかにおかしいアカウントをリスト化していった。なお、ツイッター社は発言ごとにその言語を推定しており、データとして入手可能であるが、精度については不明であるため、人によるチェックを追加で行なった。さらに、ツイッターでの発言内容のみから投稿者の性別、年代の推定する作業を研究補助者に依頼し、サーベイで回答者が入力した性別及び年代と合致するかどうかの確認を行なった。これらの作業の結果、別人や有名企業のアカウント名、いい加減な（存在しない）アカウント名を入力する回答者がかなりの数いることが明らかになった。また、ツイッターのアカウントは存在するものの、投稿数が少ないため（主に閲覧目的でアカウントを保持しているケース）、分析対象とならないケースも見受けられた。最終的には、ツイッターのアカウント名を提供した 1,476 名の回答者のうち、分析対象に含めることができたものは 339 名であった。

上記予備分析の結果を受けて、2022 年 5 月にツイッターアカウント名を他の項目とともに尋ねるサーベイ調査を実施した際には、冒頭にスクリーニング調査として、ツイッターの公開アカウントの有無とその使用（投稿）頻度を尋ねた。そしてある程度頻繁に使用する公開アカウントを保有していると答えた回答者について、過去（2019 年以降）から将来（2022 年 6 月まで）にかけての投稿を収集することについての同意を求めた。同意が得られた場合、アカウント名を尋ね、そこでスクリーニング調査を一旦終了した。その後、サーベイ会社からアカウント名の提供を受け、明らかにおかしいアカウント名（有名企業のもの、メールアドレスが入力してあるものなど）を排除し、その上でツイッター社の API を用いて彼らの投稿を収集した。その後、存在しないアカウントを入力した回答者及び投稿数が非常に少ないアカウントを排除した上で、有効と思われるアカウント名を入力した回答者に再度サーベイ会社経由で連絡を取り、属性や精神状態についての各種尺度への回答を依頼した。

本稿では、上記の方法で 2020 年 5 月に収集したサーベイ調査のデータ及びツイッターデータを分析対象とした。ツイッターのアカウント名及びデータ収集に同意した回答者は 2,850 名、そのうち最終的に分析対象となった回答者は 1,174 名であった（排除基準については 2.2 参照）。サーベイ調査では回答者の基本属性に加え、孤独感（UCLA3 項目版）や抑うつ状態を測定する PHQ-9（Patient Health Questionnaire-9）など回答者の状態を測る質問項目を含めた。総質問項目数は 24 である（スクリーニング調査項目を除く）。なお、サーベイ調査の実施はサーベイリサーチセンターに委託した。

2.2 ツイッターデータ

上述のように、2,850 名がツイッターのアカウント名及びデータ収集に同意したが、その後 2 問のアテンションチェック（いい加減に調査に回答をしていた場合、間違えた選択をしてしまうよう設計された質問）及び回答時間（1 分 30 秒未満の回答を無効として扱う）によって回答者を排除し、さらにアカウント名に問題があるケース、アカウントが公開されていないケース、及び性別について回答がなかった回答者を排除した結果、1,703 名がツイートの収集対象者となった。

これら 1,703 のアカウントから 2019 年 1 月 1 日から 2022 年 5 月 30 日までの投稿をツイッター社の API を用いて収集し、2,493,682 件のツイートを収集した。収集に際しては、学術目的の利用が許可された研究者向けアカウントを用いたため、収集可能な投稿数についての上限は適応されなかった。したがって、2019 年以降のすべての投稿をそれぞれのアカウントから収集することが可能であった。上記期間中にツイートの総投稿数が 28 未満（下位 10%に該当）であった回答者、リツイートしかしてない回答者を排除したところ、1,174 名が最終的な分析対象となった。さらに、当初収集した 250 万件近いツイートのうち、リツイートを排除し、Yahoo!ニュースのハッシュタグが含まれているツイートを排除し、最終的に 1,459,923 件を本研究の分析対象とした。リツイートとニュース記事を排除したのは投稿者の状態を直接的に表しているとは考えにくいためである。表 1 は収集したツイートの年別の分布を示す。

表 1：分析に含めたツイート数（年別）

2019 年	339, 284
2020 年	412, 655
2021 年	483, 951
2022 年	224, 033

注：2022 年は 5 月 30 日までに行われた投稿を収集した。

2.3 発言を通じた投稿者の精神状態の推定

本研究では、相談者の悩みの深刻度を判定するために、LSS（Latent Semantic Scaling, Watanabe 2021）の手法を用いて、それぞれの発言（ツイート）について数値化を行なった。LSS とは、semisupervised のアルゴリズムで文書を一次元の尺度に落とし込む方法であり、その過程で word embedding を用いている。supervised の方法とは異なり、教師付きのデータを必要としないのが大きな特徴である。

LSS は分析者が指定した種語を元に、文書の極性値（例：感情のポジティブあるいはネガティブ）を、種語に関連する特定の次元上で計算する。今回は、発言者（ツイッターの投稿者）の精神的状態の悪さ（psychological distress）を推定することを目的として、表 2 に掲載した種語を用いて推定を行なった。なお、推定に際しては、これらの種語そのものが分析対象の文章中に入っている必要はなく、例えば「悪い状態」カテゴリーの種語と「似ている」（つまり、word embedding の結果意味上の「距離」（semantic proximity）が近い）と判断された単語については、「悪い状態」を表すと判定されるようになっている。種語の選定に際しては、いくつか異なるパターンを試した上で最終的に使用する単語を決定した。

表 2：分析に用いた種語

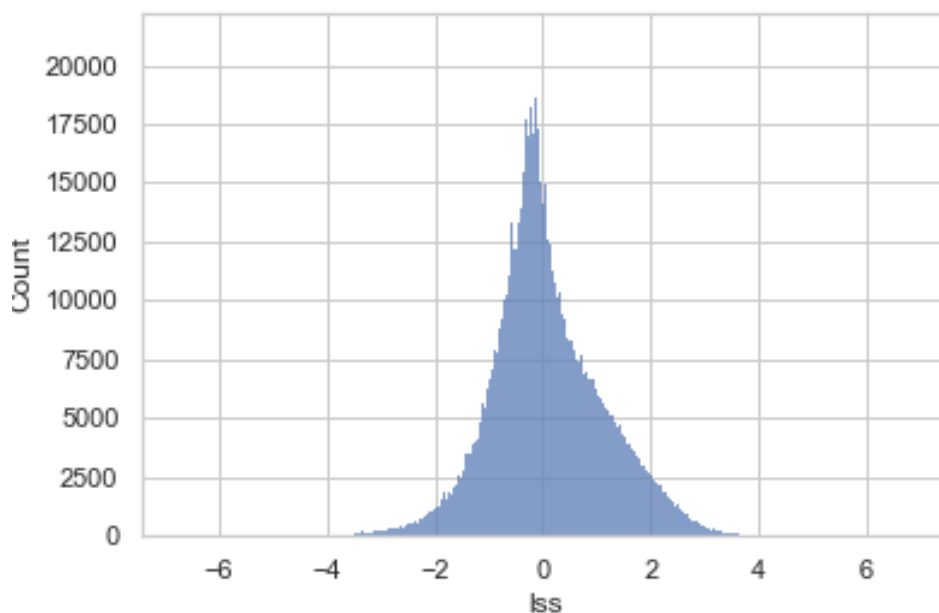
状態	単語
悪い	[悩み, 心配, 嘆く, 泣く, 悔しい, 困る, しんどい, 苦痛]
良い	[楽しみ, 絶好調, 喜ぶ, 笑う, 嬉しい, 幸せ, のんびり, 元気]

上記種語を用いて推定した極性スコアの記述統計とスコアの分布をそれぞれ表 3 と図 1 に示す。スコアは発言（ツイート）ごとに推定した。正の値は状態が悪いことを示し、負の値は状態が良いことを示している。なお、LSS スコアは初期に収集したツイートすべて（約 250 万件）について推定し、その後平均値が 0 になるように正規化を行っている。

表 3：LSS スコアの記述統計

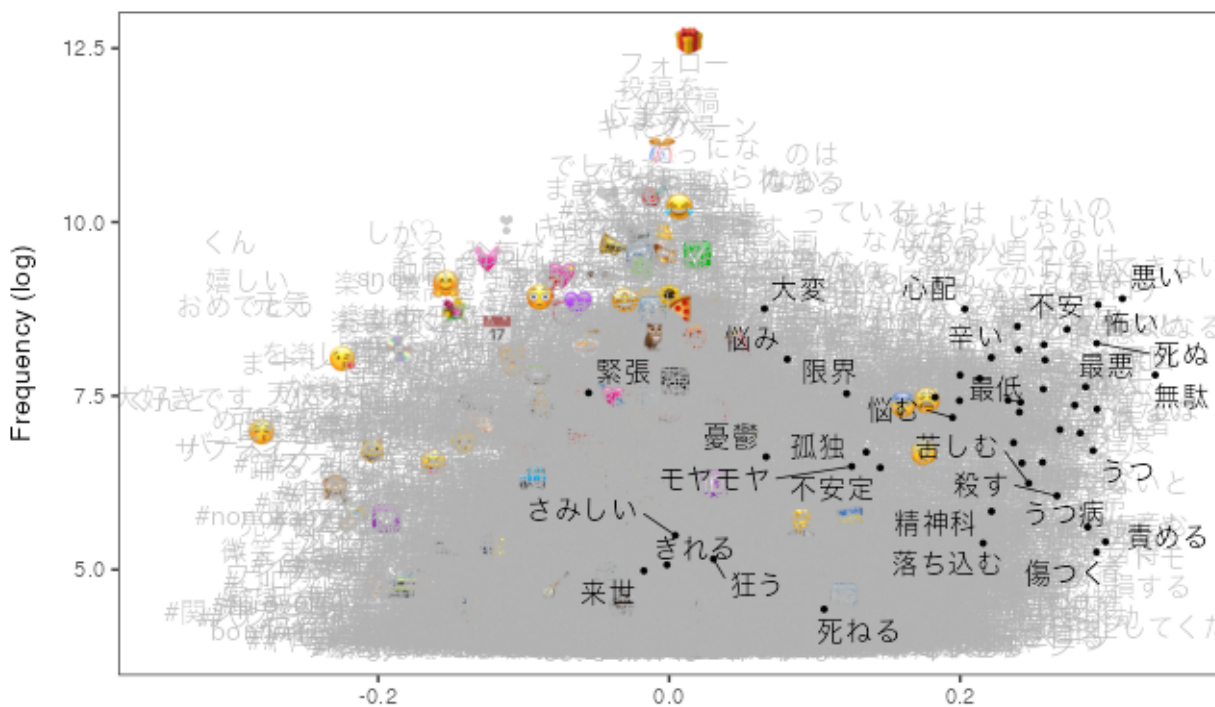
総数	1, 459, 923
平均値	0. 07577
標準偏差	1. 02828
中央値	-0. 06047
最小値	-6. 737523
最大値	6. 821659

図 1: LSS スコアの分布



これらの LSS スコアが実際に精神的状態の悪さを正しく測定しているかを目視にて確認するために、単語とそれぞれのスコアについて表示したものが図 2 である。図中に黒字で表示されている単語は、悩み相談の掲示板（本研究が分析対象とする SNS 相談とは異なるが、類似の性質を持つもの）で頻出する単語で、悩みを抱える人や、精神的状態が悪い人が頻繁に使用する傾向にある単語である。図 2 によると、これらの単語の LSS スコアは正の値を取っており、正しく判定される傾向にあることがわかる。

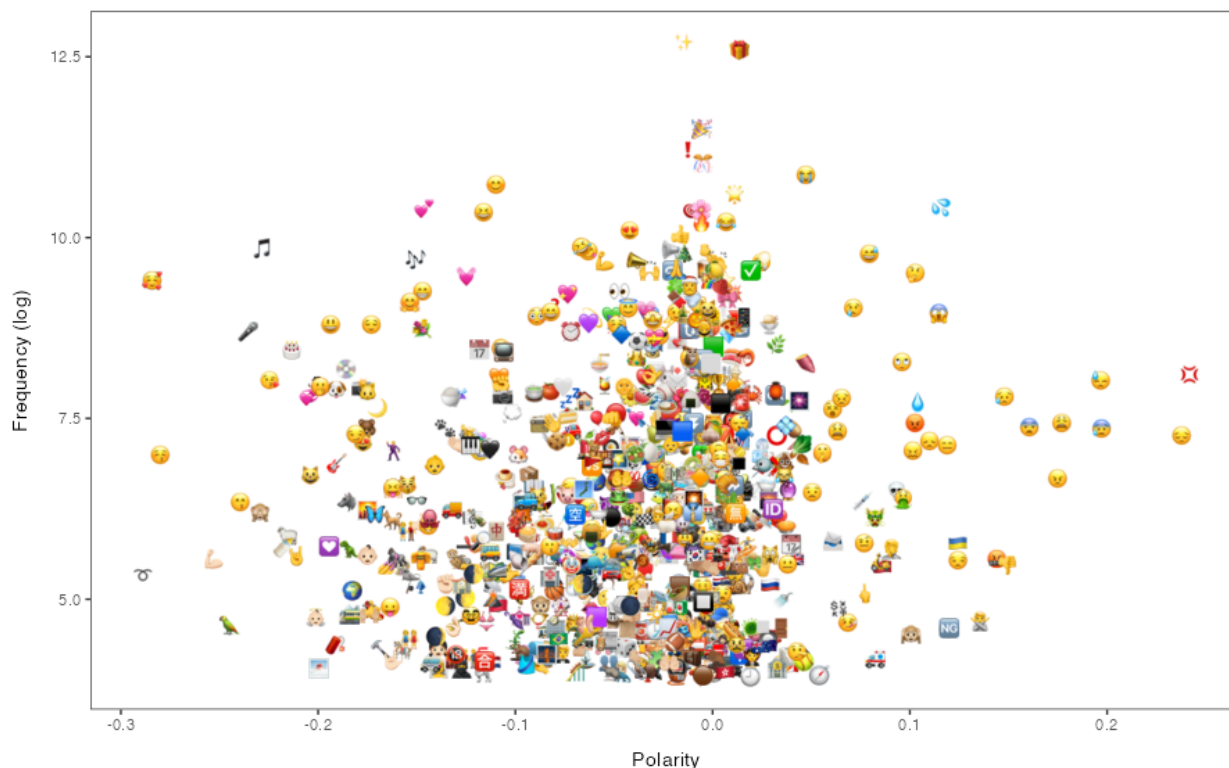
図 2: 単語別 LSS スコア



注：横軸の LSS スコアを示し、正の値が悪い状態を表している。縦軸は頻度。

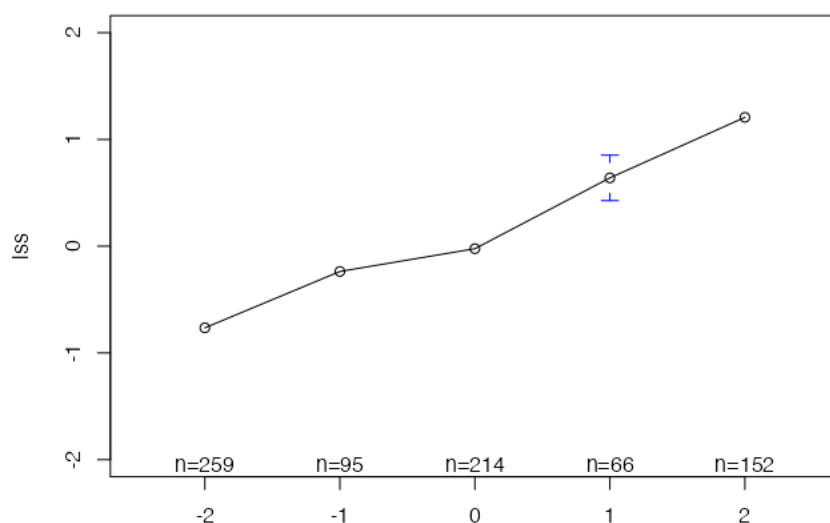
同様の確認をツイート内の絵文字のみで行ったものが図3である。嬉しい気持ちや楽しい気分を示す絵文字は負の値を取り、悲しみや怒りなどを示す絵文字は正の値を取っていることから、スコアは投稿者の状態や気分を反映する傾向にあることがうかがえる。

図3：単語別 LSS スコア（絵文字のみ）



以上の目視による簡易な検証に加え、人によるコーディング（判定）を用いた検証も行った。具体的には、3名の研究補助者にそれぞれ800のツイートを読んでもらい、発言の内容のみを判断基準として、投稿者の状態が「悪い」、「どちらでもない」、「良い」かについて判定を依頼した。「悪い」と判断した場合には1、「どちらでもない」と判断した場合には0、「良い」との判断を下した場合には-1のスコアを割り当てた。3名は独立に作業をし、他のコーダーによる判定結果は一切見していない。最終的には3名のコーディングの一致度に基づき、各発言について2, 1, 0, -1, -2のスコアを付与した。例えば、3名が全員「悪い」と判断した場合は、2のスコアをつけ、2名が「悪い」と判断したが、他の1名がそうでないと判断した場合には1のスコアをつけた。つまり、最終スコアはコーダー間の一致度が高いほど大きい絶対値を取り、その符号は状態が悪いか良いかを示している。人によるコーディングとLSSのスコアの関係を見たものが図4である。横軸は人の手による総合スコア、縦軸はそれぞれのスコアの値についてLSSの平均値を示している。図4によると、人が「悪い」と判定したツイート（つまり、横軸のスコアが正の値を取るもの）については、LSSの平均値が正の値を取る傾向にあること、そして反対にコーダーが「良い」状態であると判断したものについてはLSSの値が負の値を取る傾向にあることがわかる。したがって、LSSスコアと人による判断は方向性が一致していることが明らかになった。

図4：3名のコーダーによる投稿者の状態の判定結果とLSSスコアの関係

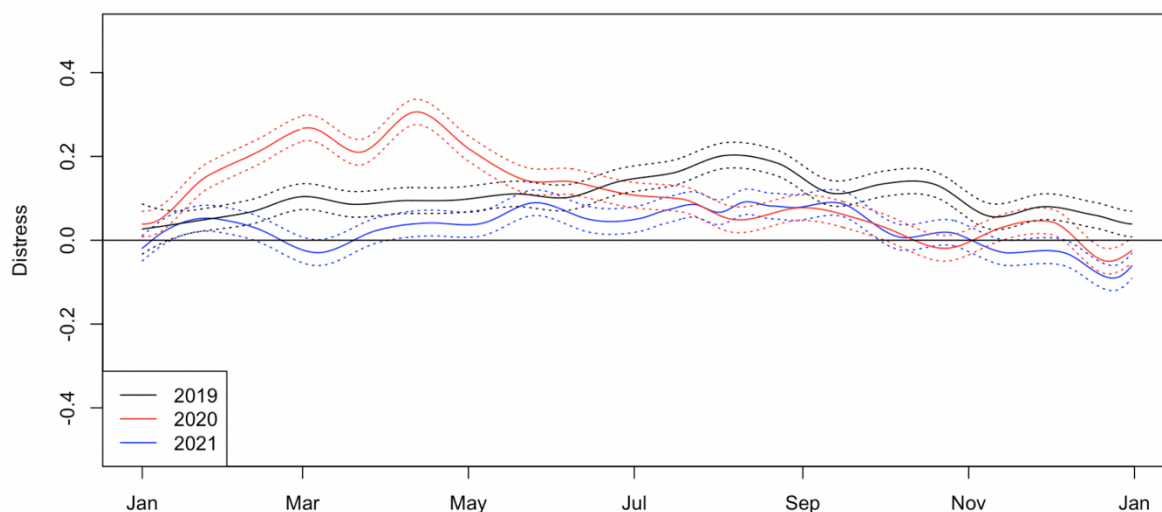


さらに、LSSは連続値を取り、人による総合スコアは離散値を取ることに留意した上で、両者の相関係数を計算すると、0.71となった。以上から、LSSによる推定は人による判断と相当程度の一致度を持つと結論づけることができる。

3 分析結果

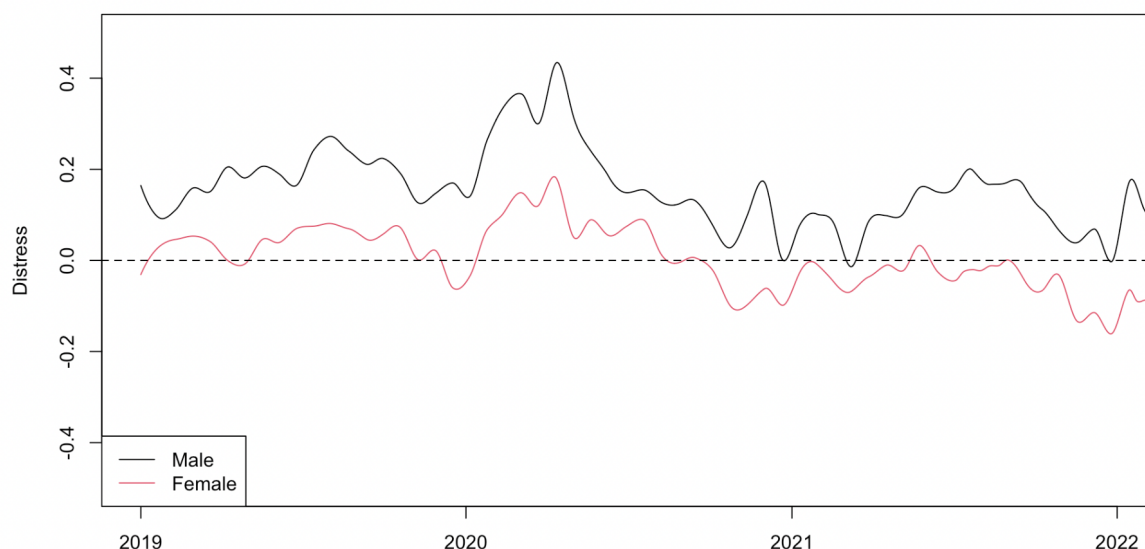
上記作業によりLSSスコアの妥当性が確認できたことから、本節では主に得られた結果をいくつか紹介する。図5は、LSSの日別平均値を用いたデータを年別にプロットしたものである。点線は95%信頼区間を示す。2022年については本稿作成時点で5月までのデータしか収集できていないため、図の作成時には除いた。図5において赤線で示された2020年のスコアを見ると、新型コロナウイルス感染症の拡大が広がった2020年春頃から、人々の状態が前年の2019年（黒線）と比較して大幅に悪くなっている（正の値は精神的状態が比較的悪いことを示す）。最初の緊急事態宣言が宣言されたのは2020年4月初頭であるが、それより前の2月くらいから状態が悪くなっていることが確認できる（日本で最初にコロナに感染した患者が確認されたのは2020年1月16日、同年の2月にはダイヤモンド・プリンセス号で感染者が発生したことが大きく報道された）。また、同年の3月の頭にも大きな山があり、日本全国で一斉休校が要請されたことが影響している可能性がある。その後、7月くらいからLSSのスコアは低下しており、2020年夏頃には人々の状態が回復したことが見て取れる。また、2021年（青線）には2019年とほぼ同じ程度に戻っており、同年の後半においては2019年よりも全体的に状態が良いという結果が得られた。

図 5 : LSS の推移 (2019-2021 年、全回答者)



次に、性別のスコアの傾向に目を向けると（図 6）、分析対象となった回答者については男性（Male、N=634）のほうが女性（Female、N=540）に比べて、分析期間を通じて LSS のスコアが高く、状態が悪いことが分かる。前述の 2020 年 3 月の一斉休校や、同年 4 月から 5 月にかけての緊急事態宣言の影響がここでも見て取れるが、それらのイベントが与えた影響の大きさに著しい男女差はないようである。なお、今回実施したサーベイ調査では回答者のサンプリングの際に代表的サンプルの確保を目指していないため、男女間に年齢などの属性に大きな違いがあることには注意が必要である。そのため、図 6 の結果だけをもって、男性のほうが一概に精神的状態が悪いと結論づけることはできないことに留意すべきである。

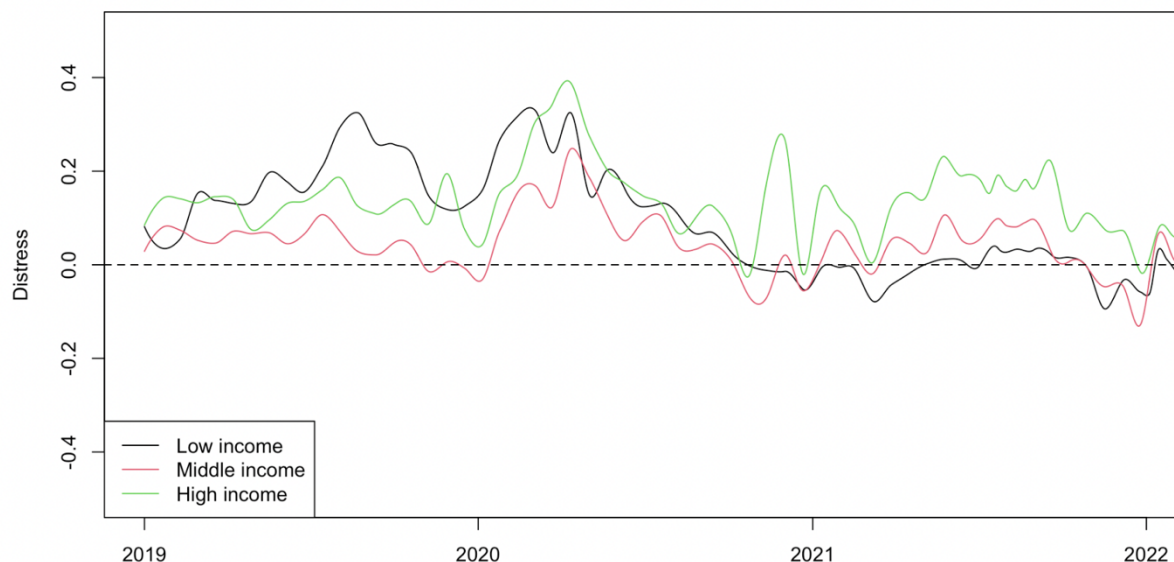
図 6 : 性別 LSS スコア



本研究ではサーベイ調査内の質問により、様々な回答者の基本属性を把握しているが、そのうちの一つである世帯年収別にスコアの傾向を見たものが図 7 である。世帯年収が 400 万円未満の回答者を「Low income」（N=422）、400 万円以上 800 万円未満を「Middle income」（N=490）、800 万円以上を「High income」（N=262）として分類した。コロナ前の 2019 年のデータを見ると、Low income の回答者の精神的状態が一番悪く、その次に悪い傾向にあるのが High income の回答者であった。コロナの感染拡大が始まった 2020 年春頃からは、どの年収グループにおいても状態が悪くなっており、Low income の回答者が特に強い影響を受けたわけでもないことが明らかになった。意外なことに、2021

年に入ってからLow incomeの回答者の状態のほうが、より年収が多い回答者に比較して良かったという結果も得られている。ただし、正確な結論を導くためには、回答者の属性（特に職業など）を統制した結果を見る必要があり、以下の図はあくまで全体的な傾向を示していると解釈すべきである。

図7：世帯年収別 LSS スコア



今回行ったサーベイ調査では、国際的にもよく使用されている、回答者の抑うつ状態を把握することを目的とする PHQ-9 という 9つの項目からなる質問を尋ねた。そのうちの一つの質問は、過去2週間に「死んだ方がましだ、あるいは自分を何らかの方法で傷つけようと思ったことがある」頻度を尋ねており、その回答をもって回答者の自殺念慮の有無を測った。上記質問に「半分以上」（つまり、2週間のうち7日くらい上記のような考えを持っていた回答者）あるいは「ほとんど毎日」と回答した人を「自殺念慮あり」と定義し、それ以外の人との比較を示したものが図8である。

図8：自殺念慮の有無別 LSS スコア

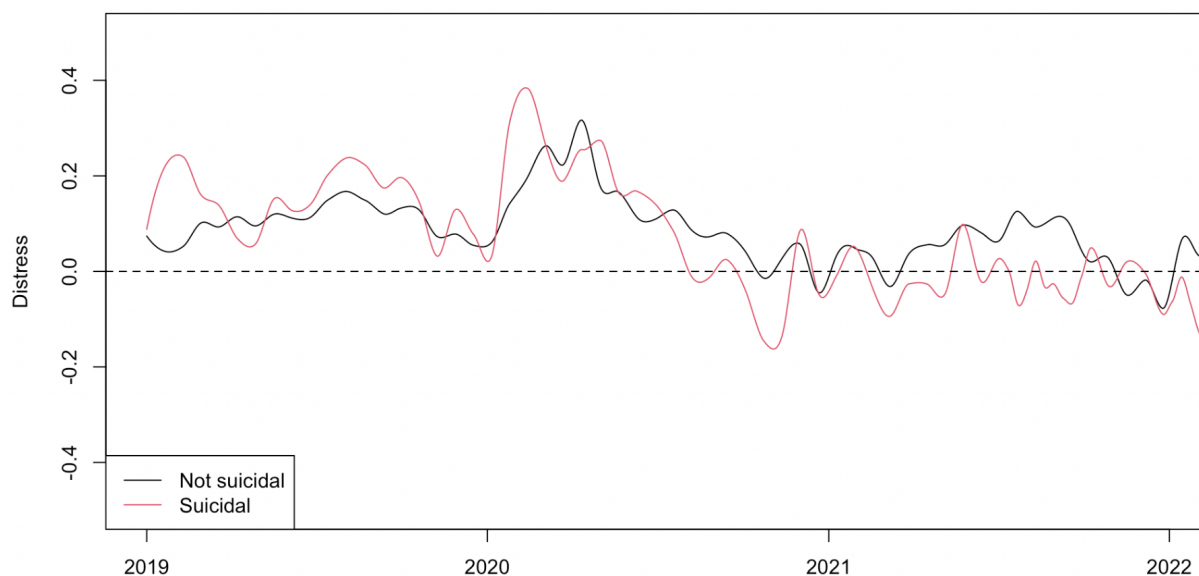


図8によると、コロナ前の2019年において、自殺念慮を持つ人（Suicidal、N=144）はそうでない人（Not suicidal、N=1,030）よりも、ツイッター上の発言から推定された精神的状態が全体的に悪

く、想定通りの結果となった。これは、LSS スコアが発言者の精神的状態をある程度正確に測れているということを示唆している。コロナ拡大の初期である 2020 年の 2 月から 3 月にかけては、自殺念慮のある人の状態は自殺をほとんど考えない人と比べて悪い傾向にあったが、同年 4 月の緊急事態宣言のころからはあまり差がなくなっている。この結果は、緊急事態宣言下ではコロナ以前に自殺念慮のあった人のほうが、そうでない人に比べて精神的状態が良かったという先行研究の結果と整合的である (Sueki and Ueda, 2021)。2020 年後半から 2021 年にかけては、自殺念慮のある人のほうがそうでない人に比べて、発言から推定される精神的状態が良いという結果も得られている。この点については、長引くコロナ禍が人々に与える複雑な影響の存在が推察され、今後詳細な分析が必要であろう。

最後に、困った時に頼れる家族以外の知り合いの存在の有無別に精神的状態の違いを検討した。サーベイでは、回答者に家族以外の知り合いで以下のような人がいるか尋ね（「悩みを聞いてくれる人」「困ったときに相談に乗ってくれたり助けてくれる人」「落ち込んでいる時に元気付けてくれる人」）いずれについても「いない」と答えた回答者を「Has no friend」(N=278)と定義した。困った時に頼れる人のいない人と、頼れる知り合いがいる人たち (Has friends, N=896) の LSS スコアを図示したのが図 9 である。

図 9: 困った時に頼れる知り合いの存在の有無別 LSS スコア

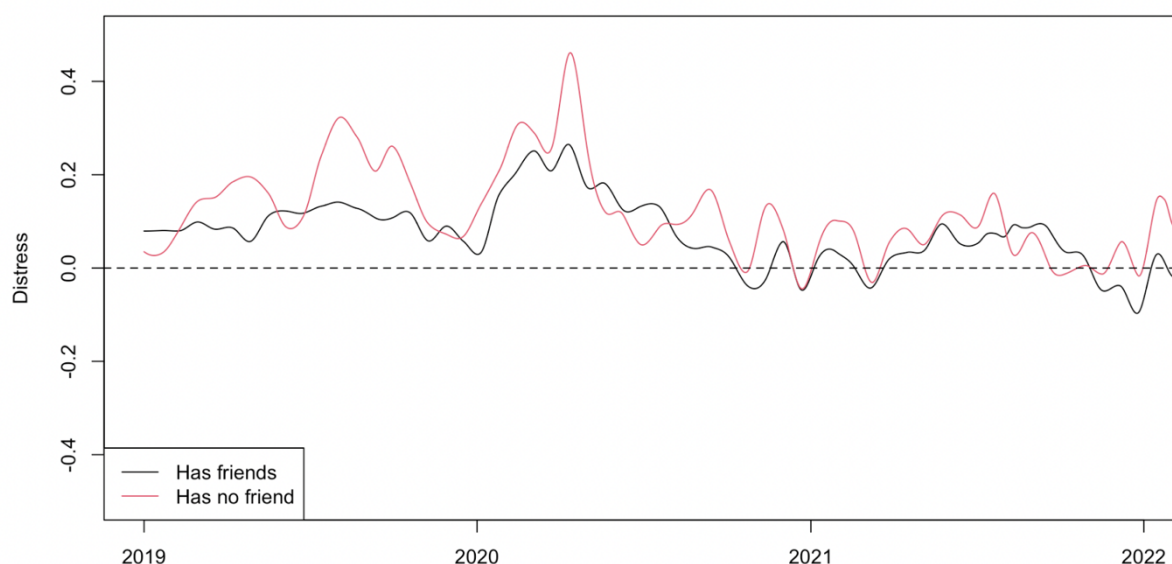


図 9 によると、困った時に頼れる知り合いのいない人たちは、コロナ前の 2019 年およびコロナ拡大期の 2020 年前半にかけて、頼れる人を持つ人に比べて状態が大幅に悪かったものの、その差は 2020 年後半から縮小していることが分かる。自殺念慮のある人の動向 (図 8) と同様に、コロナの長期化に伴う複雑な影響が示唆される。例えば、コロナ下で会食の機会や集団でのイベントなど対面コミュニケーションを必要とする場面が減り、コロナ以前には対人関係に困難を抱えていた人たちが、生きづらさを感じるようになった可能性なども考えられる。今後は彼らのツイッター上での発言の内容を分析する作業を通じて、これらの点を検証することが必要であろう。

4. まとめ

本研究では、SNS 相談において相談者の精神的状態の悪さや自殺念慮の程度を早期に把握し、悩みの深刻度や緊急度が高い相談者に即座に対応することを最終的な目的として、精神的状態の悪さを発言から推定するための方法を開発した。その結果、人によるコーディング (判定) と LSS による判定はかなりの程度一致していることが明らかになった。さらに、自殺念慮の高い人や困った時に頼れる知り合いのいない人 (すなわち、おそらく精神的状態が悪いと想定される人) の場合、発言から推定される精神的状態も悪いことが明らかになり、LSS による推定の有効性が相当程度示されたといえる。ツイッター上での個々の発言という、非常に短い文書でも推定が可能であることが明らかになったことから、この手法を用いることにより、相談者の発言が大量になくとも、ある程度数があれ

ば、彼らの悩みの深刻度を測定することが可能であると期待される。そのため、SNS 相談事業においては、冒頭の会話でチャットボットを用いて、相談者からより多くの発言を引き出すことなどを通じて、深刻度や緊急度が高い相談者の特定が可能になると考えられる。

なお、本稿ではツイッターのデータを対象に、ユーザーの精神的状態の悪さのみを測定した検証結果を紹介したが、研究期間中には実際の SNS 相談のデータを用いて相談者の精神的状態の悪さ

(Psychological distress) 及び自殺念慮 (Suicidal ideation) の二つの次元について推定も行ない、同様の成果を挙げている。SNS 相談データを対象にした分析の際には、SNS 相談に含まれる発言内容には非常にネガティブな単語が多く、ポジティブなものはほとんど含まれないことから、LSS の種語設定の際に、「悪い」状態の種語のみを用いた。SNS 相談のデータは外部に公表することができないため、本稿はツイッターデータを用いた検証結果を紹介したが、実際の相談現場での応用を見据えた技術開発ができたと考えている。

本稿の分析結果の解釈に当たって、いくつか留意点も存在する。本研究では回答者からツイッターアカウント名についての情報を得る際に二段階のステップ（ツイッターアカウント名をまず尋ね、アカウントが存在し、かつ投稿数が一定数あった場合にのみ本調査に進んでもらう）を踏むことで、存在しないアカウントあるいは投稿数が非常に少ないアカウントについては分析対象から省くことができたが、自己申告されたアカウントが実際にその回答者のものであるかは確認ができていないことには注意が必要である。有名企業アカウント名やフォロワー数が非常に多いケース（一般の個人のものとは考えにくい）は排除しているが、例えば知り合いのアカウント名を入れた場合には排除することが不可能である。将来的には、アカウント名を入力してもらう際に本人確認ができる方法を取る予定である。また、今回の調査では有効なアカウントを入手することを優先したため、例えば回答者の年齢グループによって性別の分布に大きなばらつきがあるなど、回答者が代表的なサンプルではないことにも注意が必要である。そのため、本稿の結果をもって一般的な傾向が把握できるわけではない。本研究の主な目的はあくまで発言から発言者の精神的状態を推定するための技術的開発であることに留意の上、本稿の結果は解釈するべきである。

【参考文献】

Sueki, Hajime and Ueda, Michiko. (2021) Short-Term Effect of the COVID-19 Pandemic on Suicidal Ideation: A Prospective Cohort Study. *Crisis*. DOI: 10.1027/0227-5910/a000797

Watanabe, Kohei. (2021) Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages, *Communication Methods and Measures*, 15:2, 81-102, DOI: 10.1080/19312458.2020.1832976

(注書き) 本研究においては、早稲田大学招聘研究員の渡辺耕平氏の多大なる協力を得た。

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
Female suicides during the COVID-19 pandemic in Japan	International Association for Suicide Prevention 10th Asia Pacific Conference	2022 年 5 月
Suicide mortality in Japan before and during the COVID pandemic: an update on recent trends	International Covid-19 Suicide Prevention Research Collaboration	2022 年 4 月