

第331回NRIメディアフォーラム

SNS情報から読み取る「日本の空気感指数」 ～空気感を定量化する方法～

株式会社野村総合研究所

未来創発センター データサイエンスラボ

田村光太郎

2022年4月8日



本日のアジェンダ

1. 「空気感指数」開発の背景

- NRIが開発した「空気感指数」とは
- 指数開発の背景
- 「SNSを利用した社会の可視化」研究について

2. 「空気感指数」の求め方

- 共同研究における体制と役割
- 具体的な計算方法

3. 「空気感指数」の特徴

- 事象への反応
- 長期トレンド、周期性
- 指標別分析と将来予測
- 日別の空気感指数の変化

4. 外部変数との連動

- 日経平均株価との相関

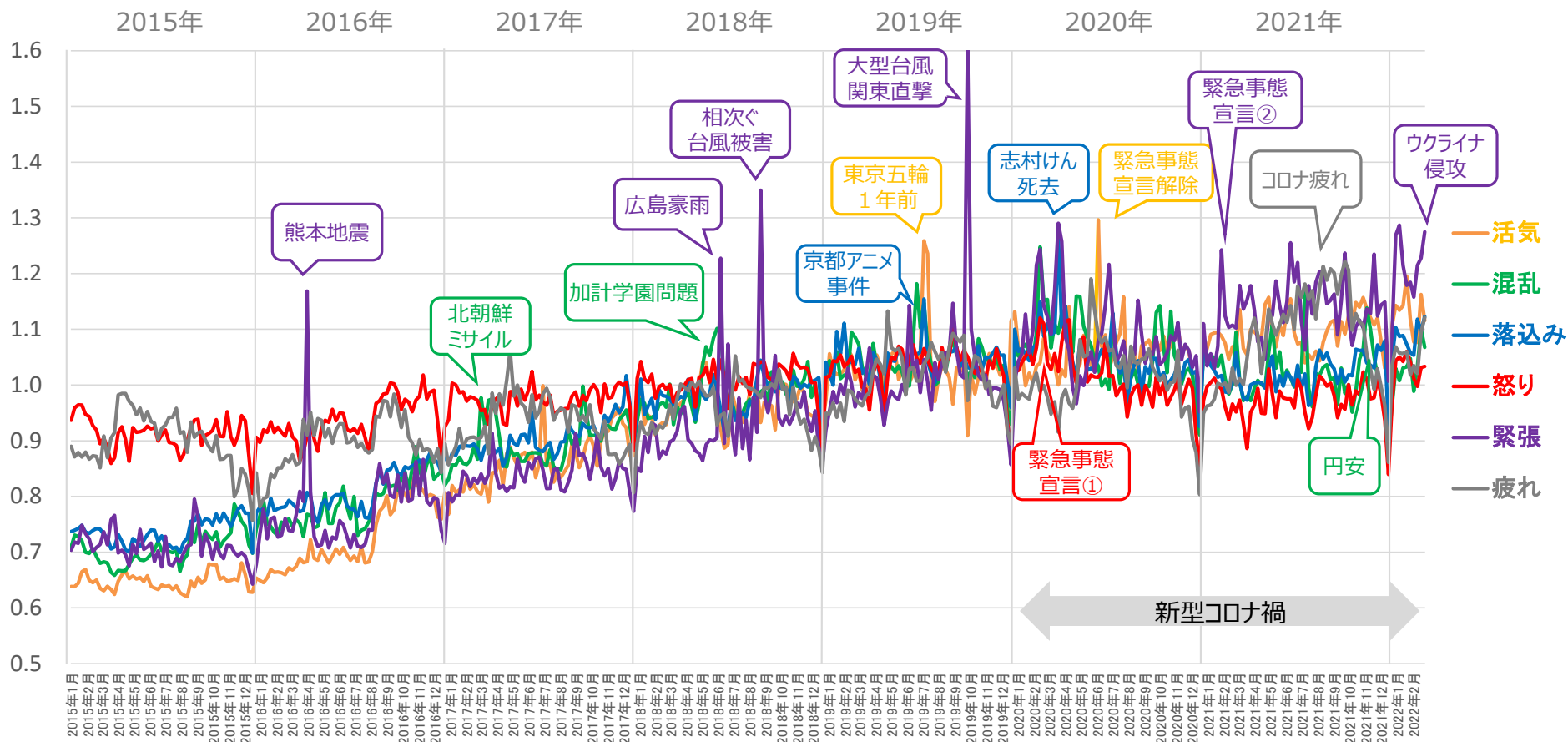
5. 今後の展望

「空気感指数」開発の背景

開発の背景：NRIが開発した「空気感指数」とは

6つの要素で「日本の空気感」を定量的に把握した指数

- 心理状態に対応する6つの指標「怒り」「混乱」「落込み」「疲れ」「緊張」「活気」で構成
- NRIではこれらの指標を継続的に計測することで、その時々々の空気感を客観的に把握するとともに、将来の政策検討、経済予測、マーケティング戦略などへも活用することを目指す



※2017年から2021年の平均を1とするように規格化

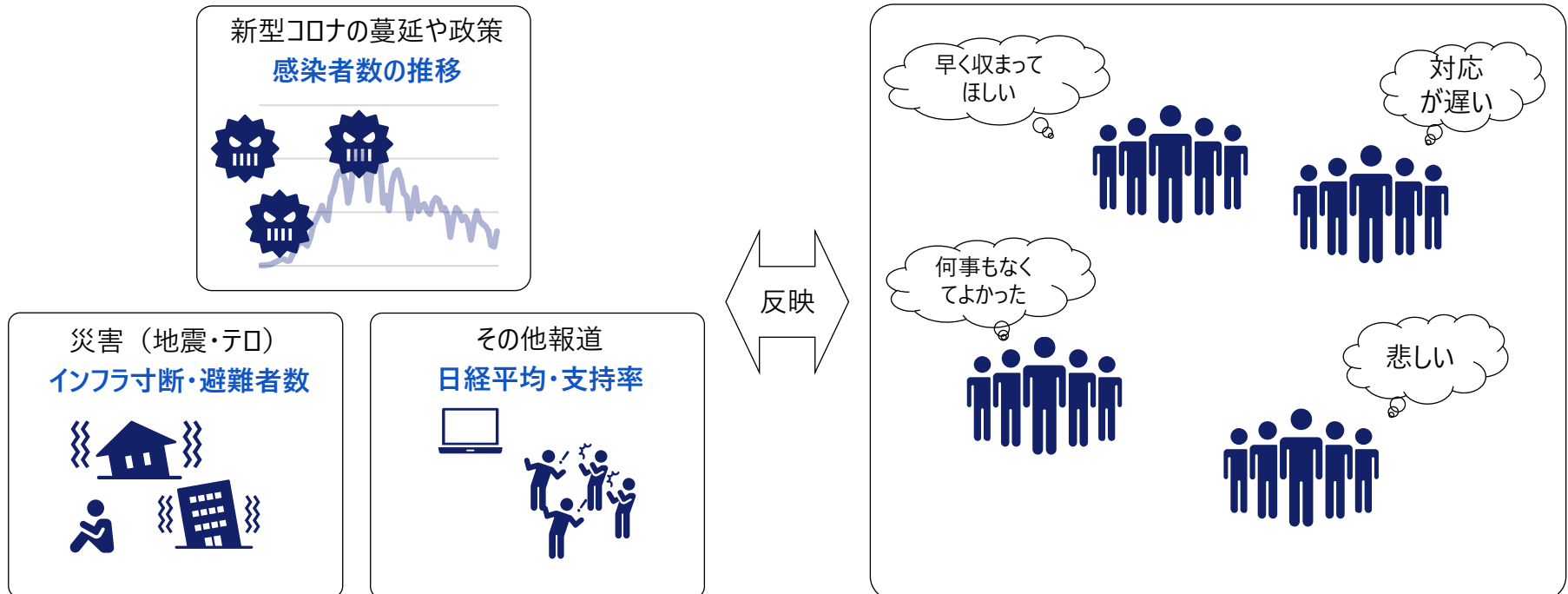
出所：NRI

Web上の情報をリアルタイムに収集すれば「空気感」を定量化できるのではないか

- 多くの社会的な事象は、その反応がWeb上に投影されることも多く、また逆に実社会に影響を及ぼすことも多い。
- 社会や経済動態の把握のために、その反応が投影されるウェブ情報を定量的かつ科学的に測定する方法を確立することが必要と考えた。

各事象ごとに個別に状態を判断
各事象に対応した指数でインパクトを測定

ウェブ上でさまざまな投稿（実社会での事象を反映）
ウェブ上の投稿をもとに事象に対する人々の反応を測定



- 東日本大震災時のTwitterのユーザ間のリツイート
ネットワーク上のデマ情報（赤）と訂正情報（青）
の広がり の可視化



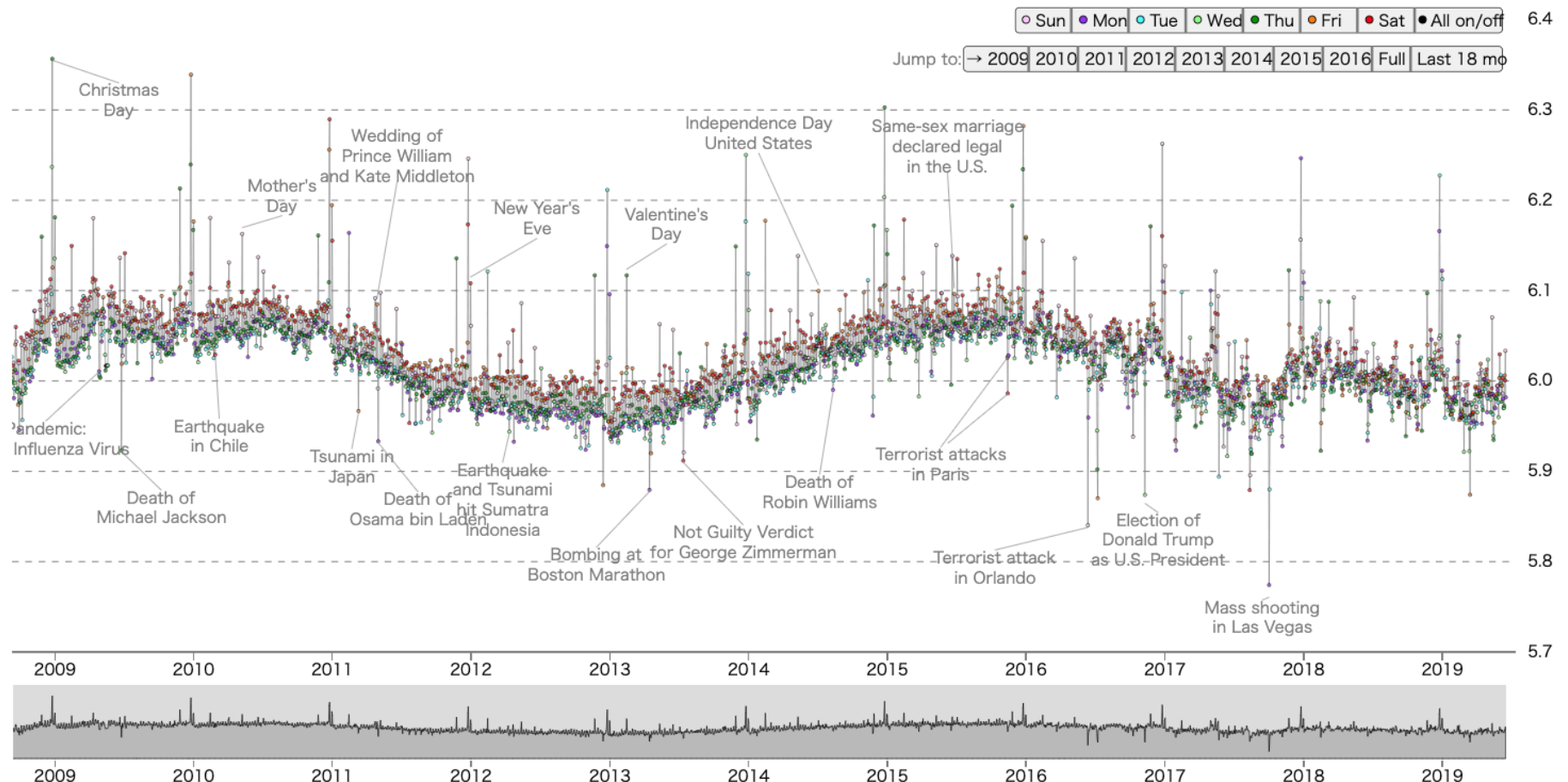
出所：Q. Wu et al., Scientific Reports, Vol. 12, 3120 より Fig.7

(参考) Twitter投稿からアメリカ社会の幸福度を測定した事例

- 海外の大学研究者グループ（Hedonometer Team）が、Twitterの投稿からアメリカ社会の幸福度を測定し、指数を配信。

Average Happiness for Twitter

出所：<https://hedonometer.org/>



Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M., "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter", *PLoS ONE* 6(12) e26752. (2011).

(参考) 経済産業省 SNS×AI 鉱工業生産予測指数

- 経済産業省「令和元年度 ビッグデータを活用した新指標開発事業」で開発された指標「SNS×AI 鉱工業生産予測指数」(野村證券) (2022年4月時点更新終了)

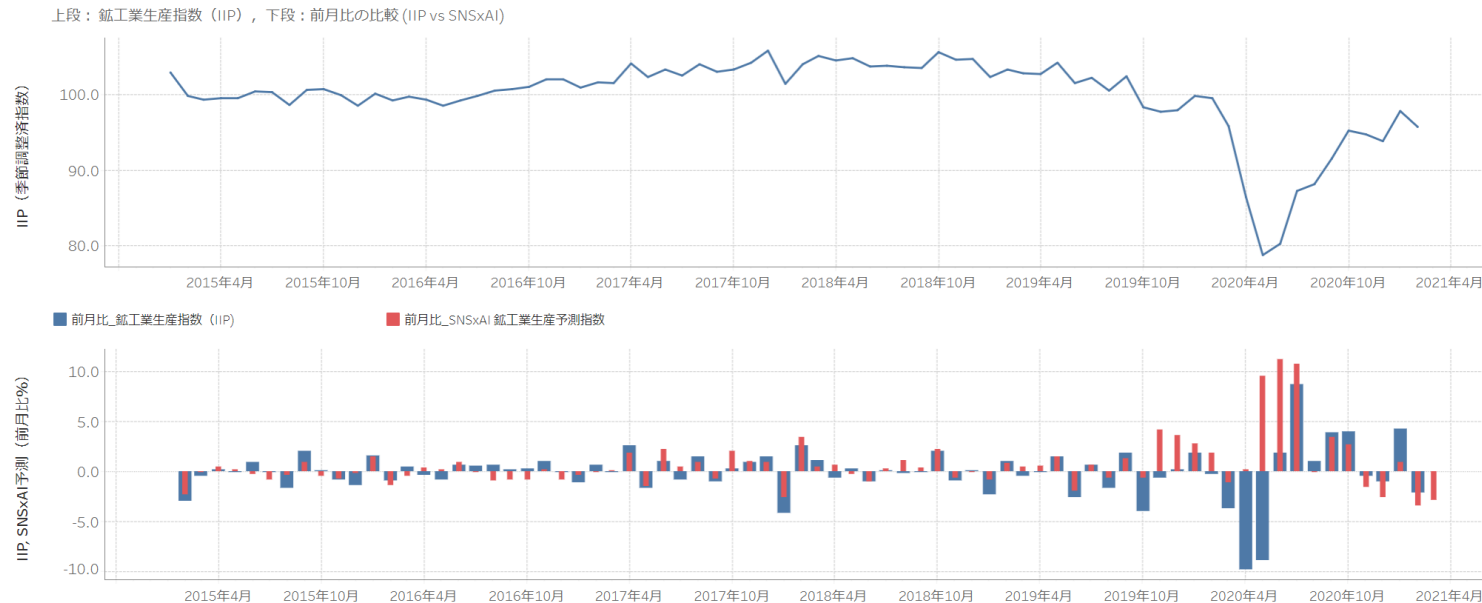
- 人工知能を活用し、TwitterやBlogから「今週は残業で疲れた」等の仕事が増えていることを示すメッセージを抽出、その件数と、製造工業予測指数の他、株価・為替といったマーケットデータを利用して、鉱工業生産指数 (IIP) を予測する指標。(下記ホームページ記載より)

- 下記の指標とともにBigData-STATSダッシュボード(β版)にて配信されている。

- METI POS小売販売額指標[ミクロ] (2022年4月時点更新終了)
- METI×NOMURA コンシューマーセンチメント・インデックス (「消費者心理指標」)

SNSxAI 鉱工業生産予測指数 (野村證券)

[利用上の注意]



出所：https://www.meti.go.jp/statistics/bigdata-statistics/bigdata_pj_2019/snsai_nomura.html よりSNS×AI鉱工業生産予測指数

「空気感指数」の求め方

NRIと筑波大学の共同研究にて指数を計算

- 筑波大学の佐野幸恵助教（システム情報系）と2021年10月から共同研究

共同研究開発における体制と役割

野村総合研究所 データサイエンスラボ

指数構成のためのアルゴリズム検討
指数利用のための外部変数との連動解析



組織ミッション

- I. 「データサイエンスに強いNRI」のブランドの構築
- II. 社内の「データサイエンス」に関するノウハウの集約・共有化
- III. データサイエンティスト人材の獲得・育成・活用

筑波大学 佐野研究室

SNSデータを利用した社会の可視化手法の検討



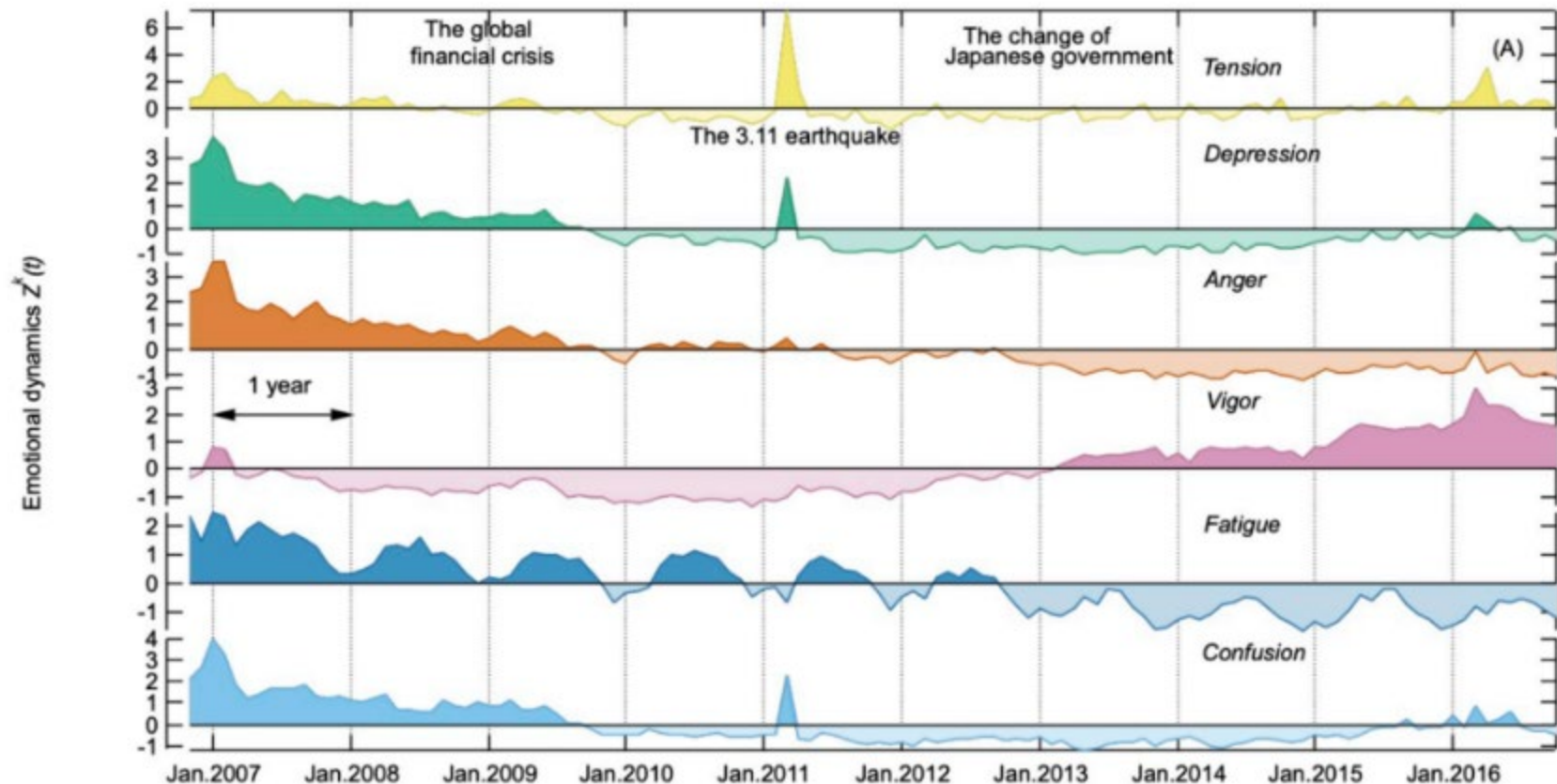
佐野幸恵助教



出所：筑波大学 佐野研究室

筑波大学の佐野助教らは、心理状態分類をブログの投稿内容に適用し、空気感を可視化

- 佐野助教らは、POMS（次ページ参照）を利用しブログの投稿内容データから空気感を可視化する技術を開発
 - ユーザ間のつながりの把握や、投稿の頻度などが課題として認識され、他のSNSプラットフォームでの研究開発を検討



出所：Y. Sano et al., Identifying long-term periodic cycles and memories of collective emotion in online social media, PLoS ONE 14(3): e0213843, (2019) より Fig.1(A)

心理学の分野で行われる心理状態を測る検査をもとに、6つの指標を設定

- Profile of mood states（POMS※）と呼ばれる心理学の分野で行われる心理状態を測る検査手法を応用
- POMSでは、各感情に関連した質問項目に回答することで被験者の6つの心理状態を把握
- 医療現場のカウンセリング、職場のストレスマネジメント、スポーツ分野のメンタルヘルス向上などに活用されている

※POMSは1971年Douglas M. McNairらが開発し、現在は、POMS 2 が開発されている。

活気	元気、活発、陽気、精力的、積極的、イキイキなどの感情
混乱	混乱、錯乱、当惑、自信がない、啞然、ゴタゴタなどの感情
落込み	悲しみ、憂うつ、孤独、怯え、ガッカリなどの感情
怒り	怒る、困る、反抗、不機嫌、イライラなどの感情
緊張	緊張、心配、不安、落ち着かない、ソワソワなどの感情
疲れ	疲れた、くたびれた、だるい、うんざり、ヘトヘトなどの感情

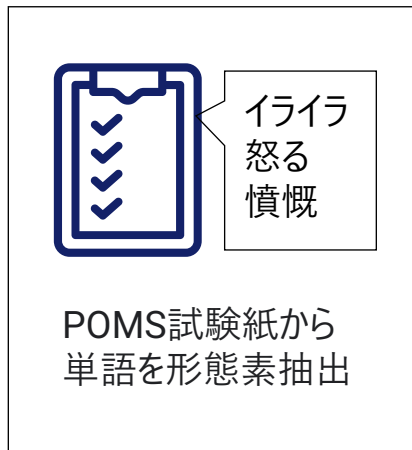
「空気感指数」の求め方：具体的な計算方法

各指標算出のための「単語辞書」を作成

- POMSの質問紙から抽出された主要な単語と類義語・表記ゆれを追加したものをもととし、指数の時系列の性質が安定する程度までの単語数を採用 ※辞書内単語の精査・追加・削除などは、随時行なっていく予定
- 当該単語を含んだ投稿をtwitterより取得

単語リストの構築（下記は「怒り」の例）

試験紙から単語の抽出



表記ゆれ単語の追加

苛々
いらいら
苛苛

類義語単語の追加

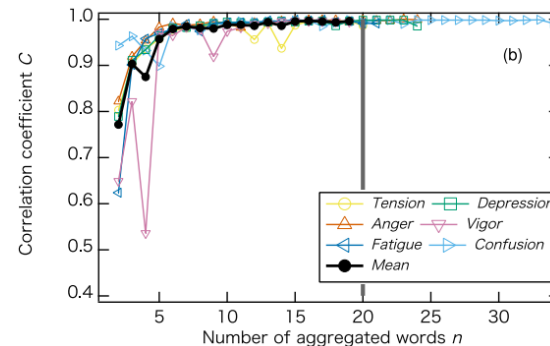
癪だ
腹立たしい
許せない

リスト内単語の統計的な精査

25単語程度/項目で時系列の統計的性質が安定する。

⇒ 1項目の単語数の基準を25単語としている。

- ・ 特定月の投稿数が5回/月以下の単語を排除
- ・ 特定日の投稿数が1000回/日以上単語を冗長化



出所：Y. Sano, Nonlinear Theory and Its Applications, IEICE 9(3), pp.380-390 (2018) より Fig.6(b)

構築された辞書

項目	単語数
緊張	21
落込み	25
怒り	25
活気	20
疲れ	22
混乱	35

※混乱は低頻度の単語が多いため、単語数を多くしています。

インプットデータとしては、Twitterの投稿データを毎日大量に収集・処理

■ 本共同研究では、データのリアルタイム性やユーザ情報の取得性の点でTwitterの投稿データを利用。

- 設定した「単語辞書」の単語とその類語が含まれることを条件として、Twitter APIにより取得したユーザの投稿データから指数を算出。



個々のツイートと、取得単語の例 ※実際の投稿ではありません

@XYZZ

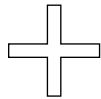
なにあの地震。余震がまだ続くかと思うと不安・

@ADFV

まじか...入学式も中止か。悲しい。

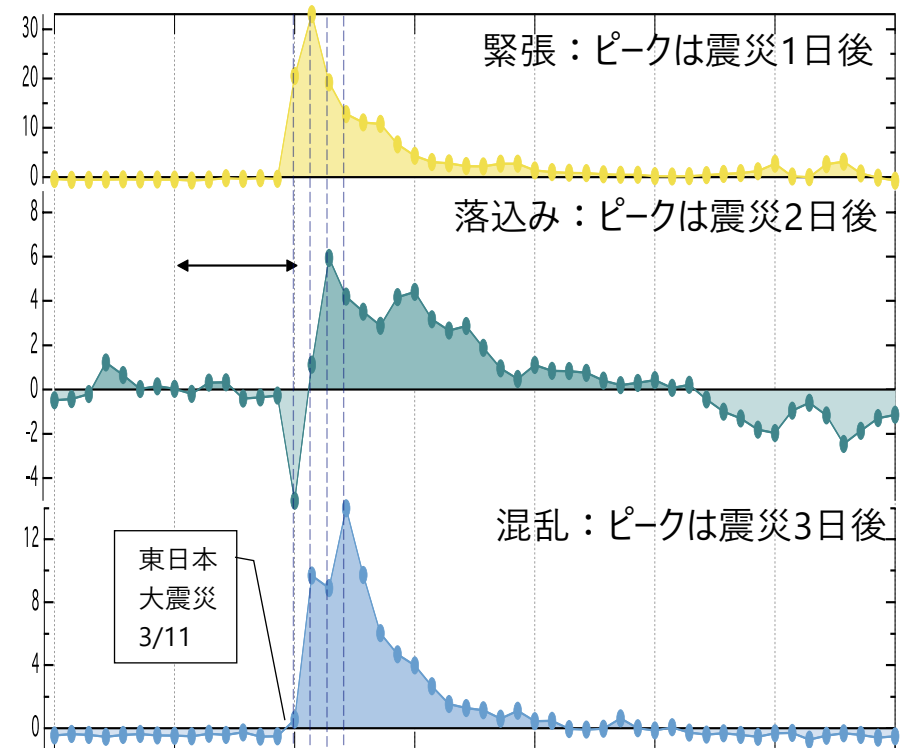
@123rR

歩いても朝までに家だどりつけないし、駅どこも混乱してる。



設定した単語辞書 + 類語

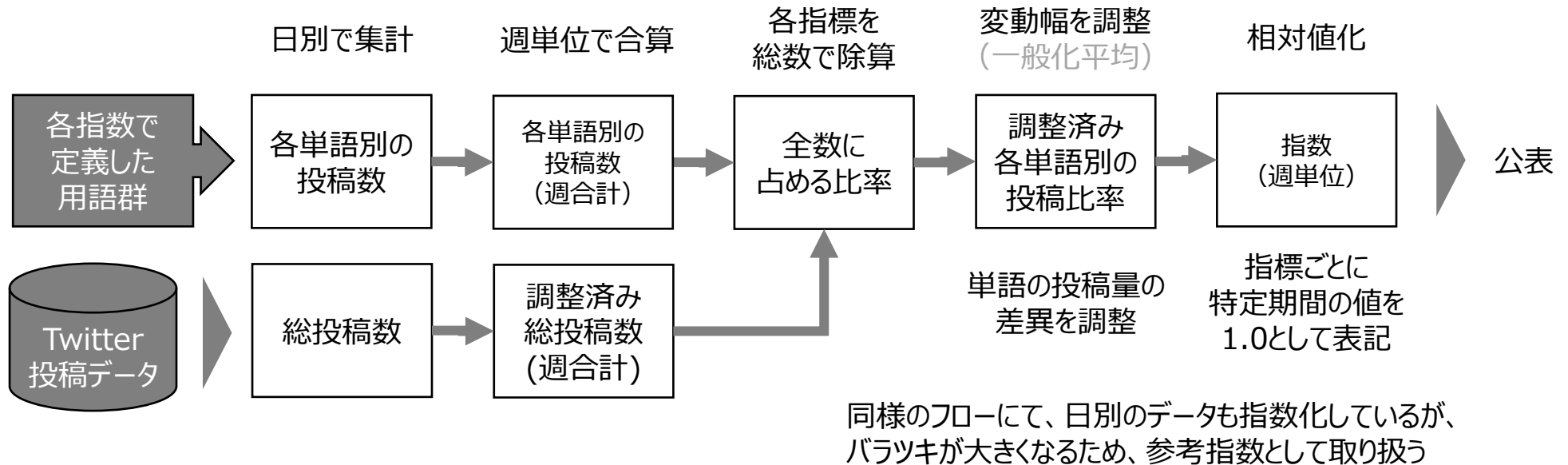
心理状態を表す6つの指数を構成（例として3つを表示）



出所：Y. Sano et al., Identifying long-term periodic cycles and memories of collective emotion in online social media, PLoS ONE 14(3): e0213843, (2019) より Fig.3(A)

各指標ごとに日次の投稿数時系列を取得し、全数での規格化や単語間格差を調整後算出

- 単語辞書をもとにTwitterから投稿を取得し、日次の単語投稿数時系列を取得。
※日本語の投稿であり、返信・リツイート、URL含やメディアのツイートを取得除外するなどのフィルタールールあり
- 全体トレンドとみなせる投稿数時系列の比率に変換したうえで、単語別の投稿件数の格差等の影響を除き、各指数を構成。



一部の単語の影響が過大にならないように調整

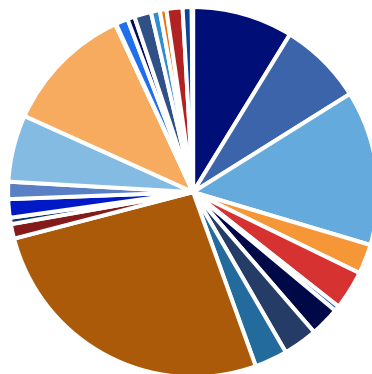
- 辞書内単語の投稿数をベースとして、特定の単語の寄与が大きくなるように、統計的に変換している。
 - 特定の単語へのスパム的な投稿や指数算出に不適切と考えられる投稿等は、精査したうえで取得を除外。
- 今後は、辞書ベースの判定ロジックから、自然言語処理技術によって、構築したモデルと併用する。

投稿件数の割合
(2015年10月の1日当たりの平均)

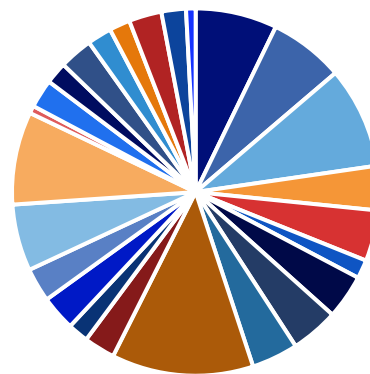
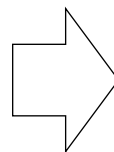
調整後の指標への寄与割合
(2015年10月の1日当たりの平均)

「緊張」と判別される単語群
(全25単語)

- ・怒る
- ・イライラ
- ・憤慨
- etc.



一部単語の指数への
寄与を調整



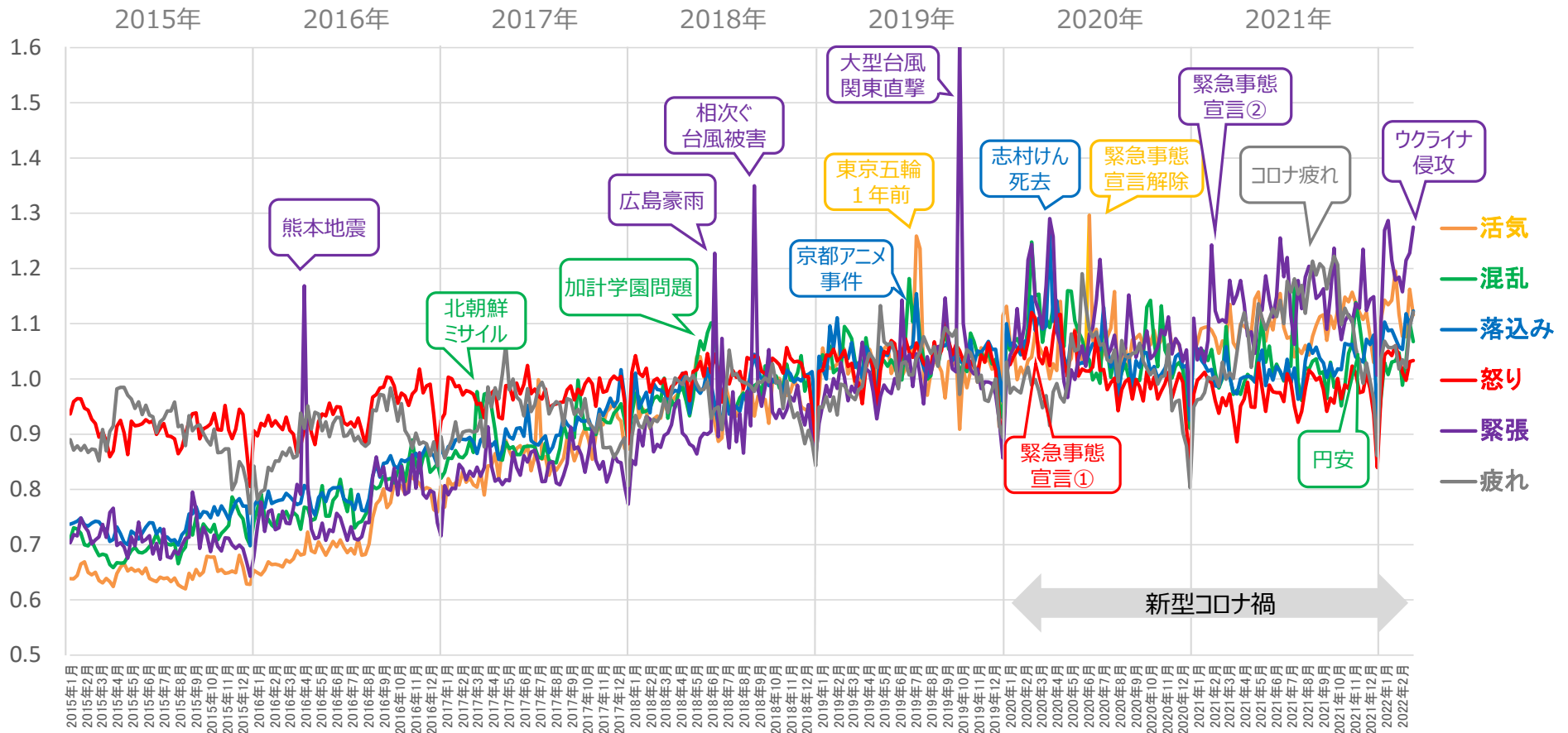
出所：NRI

「空気感指数」の特徴

「空気感指数」の特徴：事象への反応

事象によって6つの「空気感指数」（活気・混乱・落込み・怒り・緊張・疲れ）は大きく変動

- 「活気」指標はイベントや規制の緩和、「混乱」は政治的事件や経済的ニュース、「落込み」は事件や芸能で変動
- 「緊張」は大規模な災害や国家的事象、「怒り」「疲れ」は長期の政策や世相に対する反応して変動



※2017年から2021年の平均を1とするように規格化

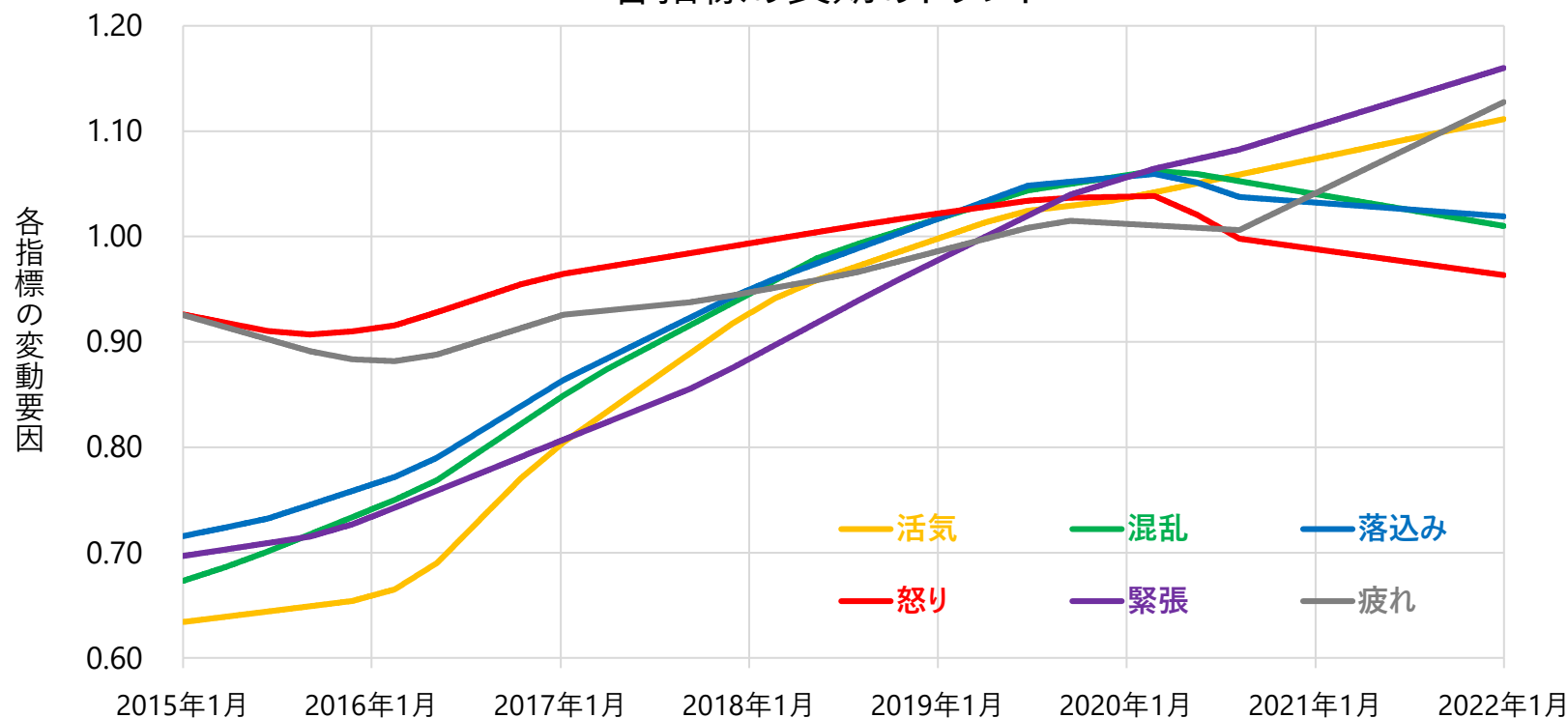
出所：NRI

「空気感指数」の特徴：長期トレンド

長期トレンドをみると、「活気」や「緊張」は右肩上がり。「混乱」「怒り」はダウントレンド

- 各指標の変化要因のうち、長期トレンド、季節性（月間変動）、週内変動などを抽出
- 長期トレンドをみると、「活気」「緊張」指標は、日本経済の好景気などを反映して、右肩上がりで推移
- コロナ禍における2020年以降では、「混乱」「落込み」「怒り」がダウントレンドに反転、「疲れ」指標は大幅に上昇

各指標の長期のトレンド



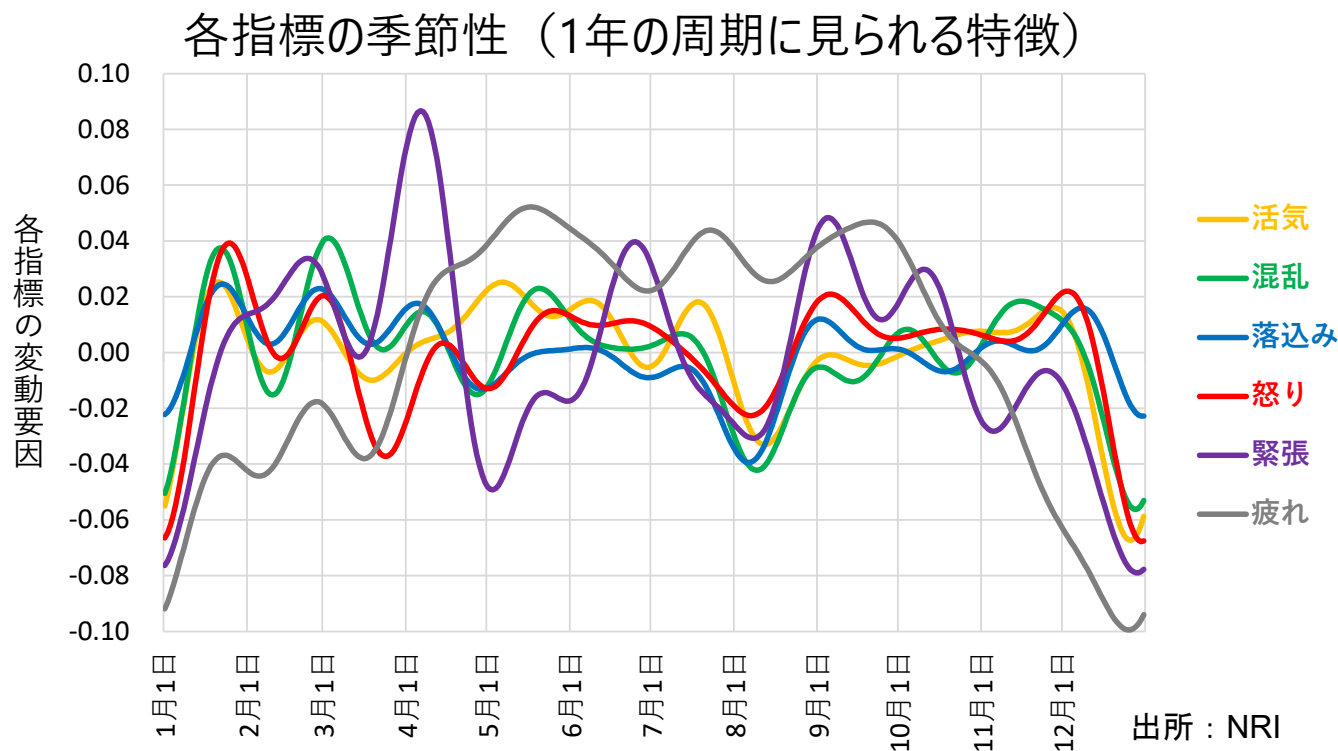
出所：NRI

注）要因の分解は「Prophet」（Facebook社が提供している時系列データを分析、将来予測するためのアルゴリズム）にて実施（以下同様）
データの変動要因を長期トレンド、周期性、時事（イベント）効果、外れ値などに分解できる。2015年～2021年のデータをもとに実施。

「空気感指数」の特徴：季節性（月間変動）

過去のデータから季節性をみると、「緊張」は4月に高まり、「疲れ」は5～10月で高い。

- 活気…… 年間を通じて安定している傾向。5月のゴールデンウィークが年間で最も高い。
- 混乱…… 全体的に月別の変動が大きい。3月初旬に高まる傾向。8月は夏休みの影響もあり低い水準。
- 落込み…… 年間を通じて、特に大きな季節変動はない。季節によらず特定の社会現象で上下する傾向。
- 怒り…… 年末年始、春休み、夏休みなどで谷になり、休み明けに高まる傾向。それ以外は一定している。
- 緊張…… 年度開始の4月、下期開始の10月に高まる傾向。夏休み明けの9月初旬も高い。GWは低い。
- 疲れ…… 年間の動きが大きい指標。夏休みなどで人々の活動が活発になる5～10月が高位安定。

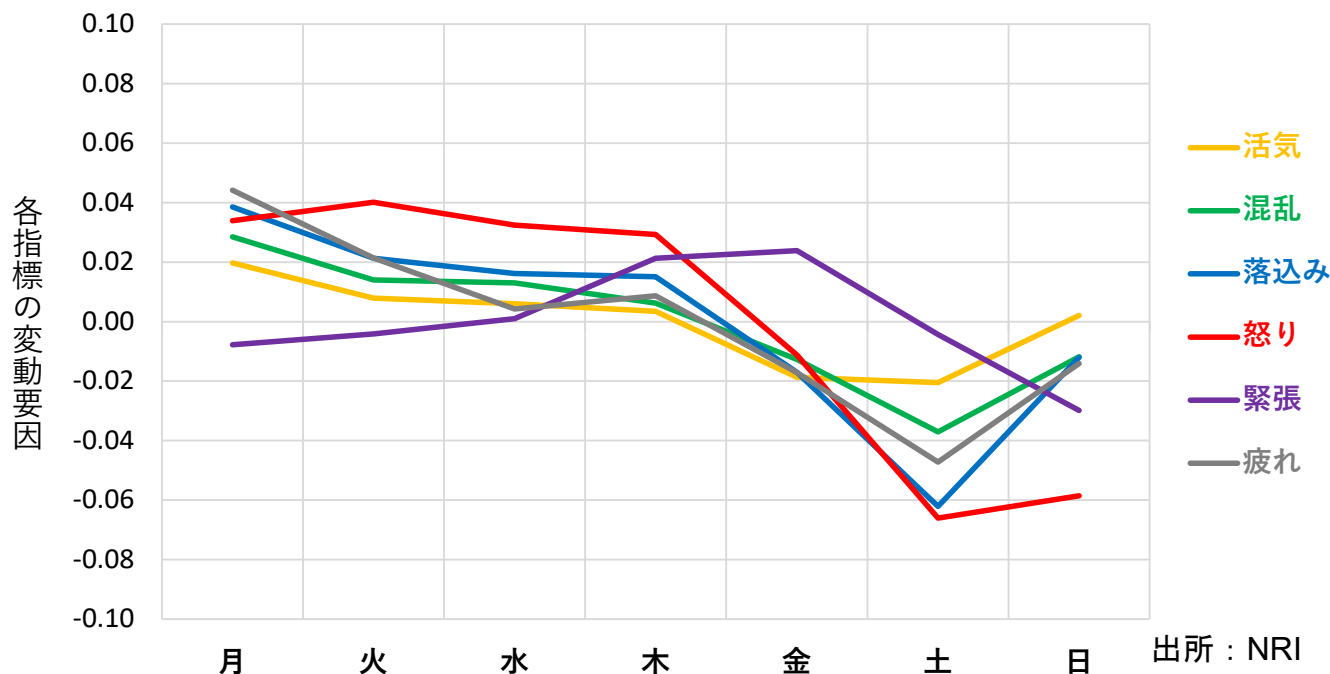


「空気感指数」の特徴：週内変動

曜日別の特徴では、「疲れ」は月曜がピーク、「怒り」は土日に低下

- 活気…… 平日と週末の差が少ない指標。平日の場合は週末に向かって減少傾向。
- 混乱…… 「活気」指標と同様に、平日の場合は週末に向かって減少傾向。土曜日には大きく減少する。
- 落込み…… 平日と週末の差が大きく、特に土曜日は低い水準。月～木曜日で高くなる傾向。
- 怒り…… 月～木曜日で高く、金曜日から土日にかけて低い。週末を除く曜日で「怒り」のきっかけが多い。
- 緊張…… 1週間の変動は小さい指標。平日は他の指標とは異なり、週末に向けて指標が高まる。
- 疲れ…… 月曜日がピークとなる指標で、そこから週末にかけて減少していく。日曜日にまた高まる傾向。

各指標の曜日別の特徴

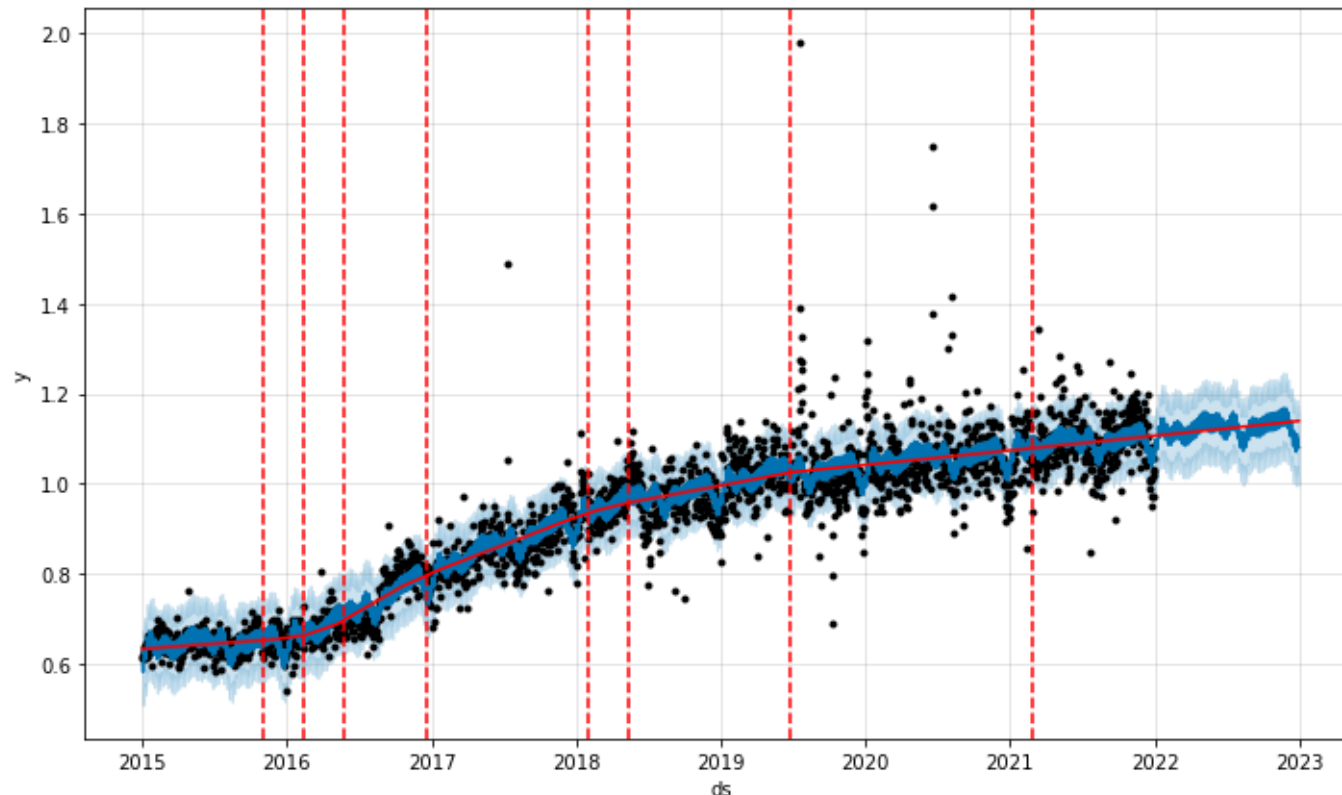


「空気感指数」の特徴：指標別分析と将来予測（「活気」指標）

「活気」指標は右肩上がりで高まっているが、2021年になり上向き傾向が高まる

- 「活気」指標の傾向は、2016・2017年に大きく上昇し、2019年の中頃から上昇トレンドが減速
- 好調な景気を反映し、ゆるやかな右肩上がりの後、2021年3月ごろにトレンドが変わり、上昇傾向が強まる
- 2020年は新型コロナ禍で活気が停滞するも、2021年には慣れもあり、長期的には活気が上向きに変化

「活気」指標の時系列変化と将来予測



出所：NRI

注) 過去のデータの特徴分析や将来予測は「Prophet」にて実施。

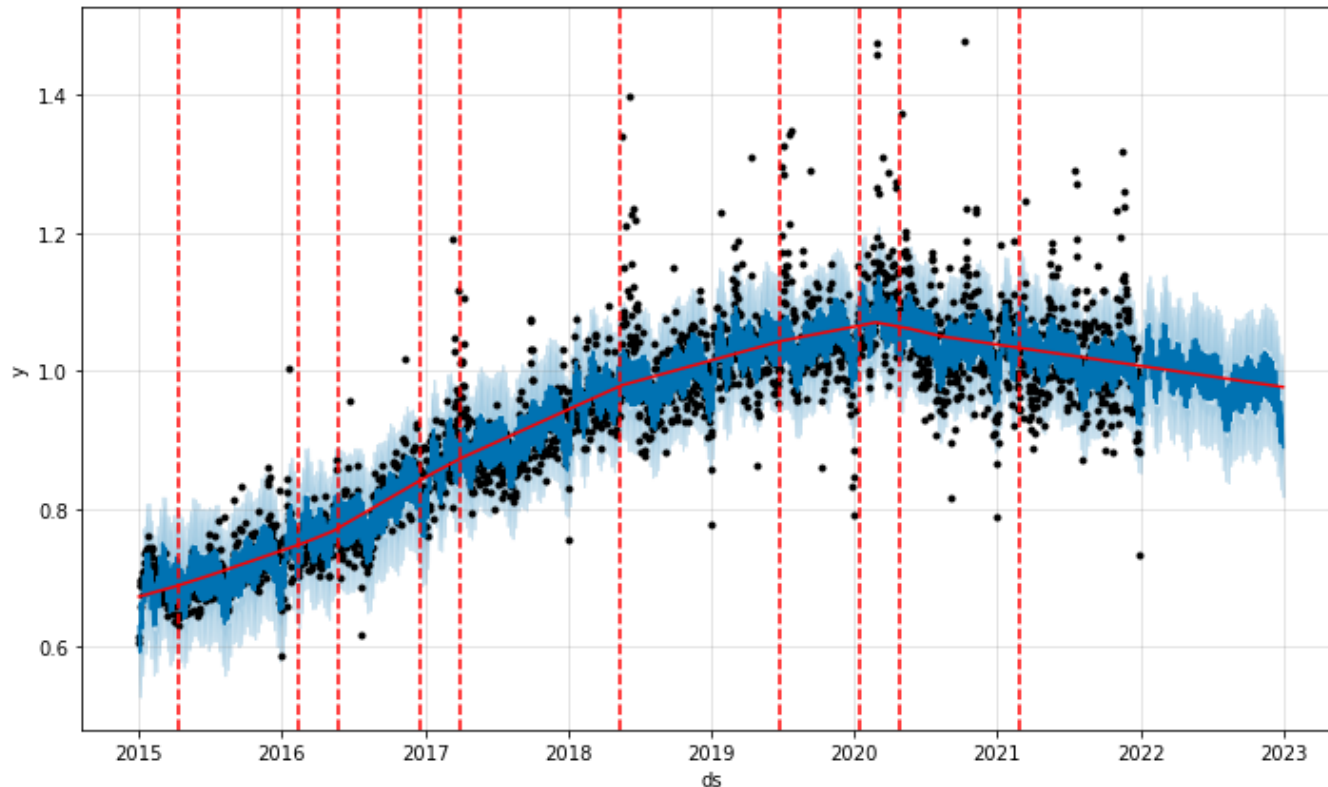
黒い点が実際の値、青色の線が数理解析モデルによる予測幅（2022年は予測値）、赤点線は変曲点（トレンドが変わったタイミング）

「空気感指数」の特徴：指標別分析と将来予測（「混乱」指標）

「混乱」指標は、2019年までは右肩上がりも、2020年初頭からダウントレンド

- 「混乱」指標は、2015年～2019年までは緩やかに右肩上がりの傾向
- 新型コロナ感染症が流行し始めた2020年の初頭を境に上昇トレンドが下降トレンドに
- 2021年になり下降トレンドはやや落ち着きをみせており、今後の動向が注目される

「混乱」指標の時系列変化と将来予測



出所：NRI

注) 過去のデータの特徴分析や将来予測は「Prophet」にて実施。

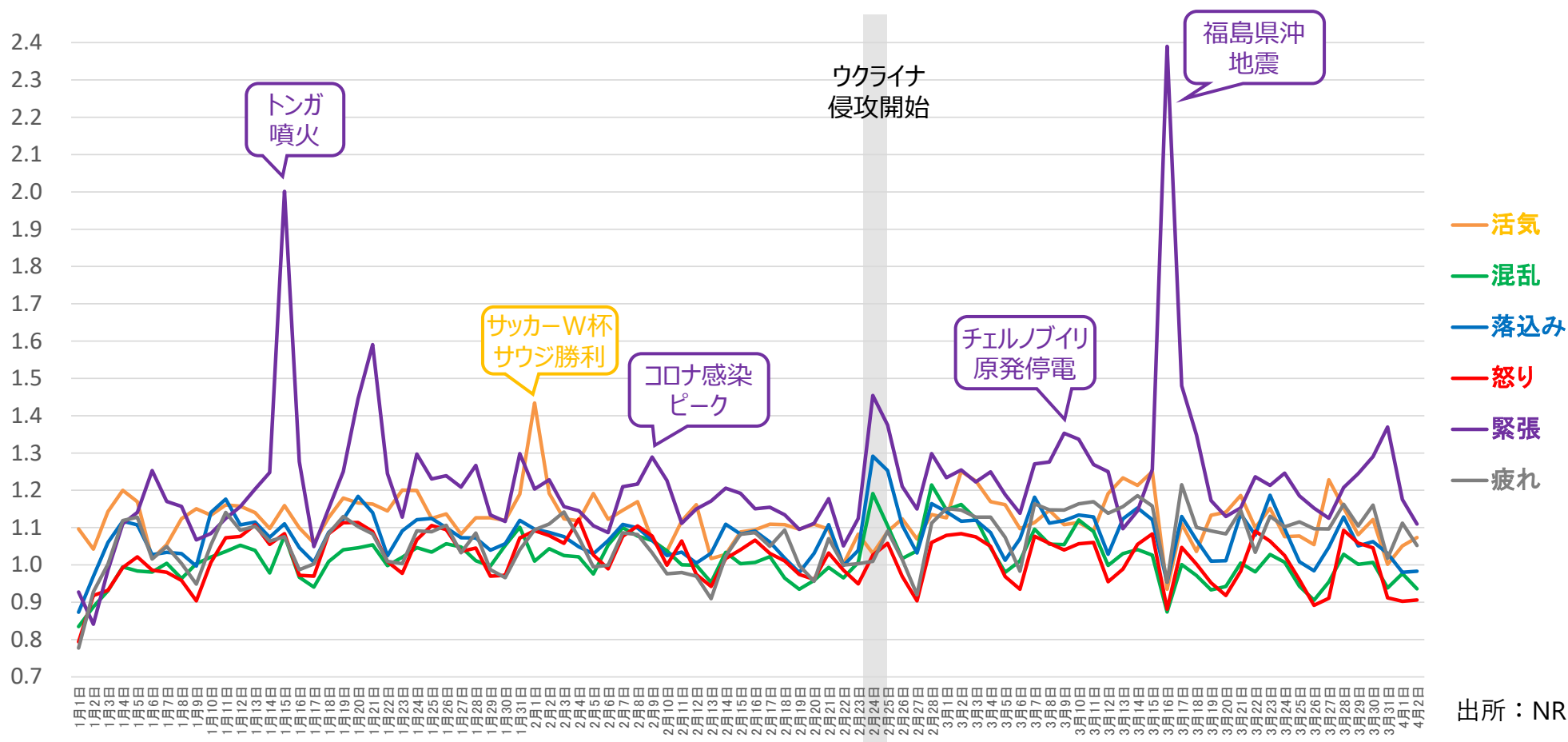
黒い点が実際の値、青色の線が数理解析モデルによる予測幅（2022年は予測値）、赤点線は変曲点（トレンドが変わったタイミング）

「空気感指数」の特徴：日別の空気感指数の変化（2022年1月1日～3月22日）

今年に入り、「トンガ噴火」「ウクライナ侵攻」「福島県沖地震」などが空気感指数に影響

- 空気感指数はデیلیーで観測することで、短期的な日本の空気感の変化を把握することができる
- 2月24日の「ウクライナ侵攻開始」では、「緊張」だけではなく「落込み」や「混乱」も高める

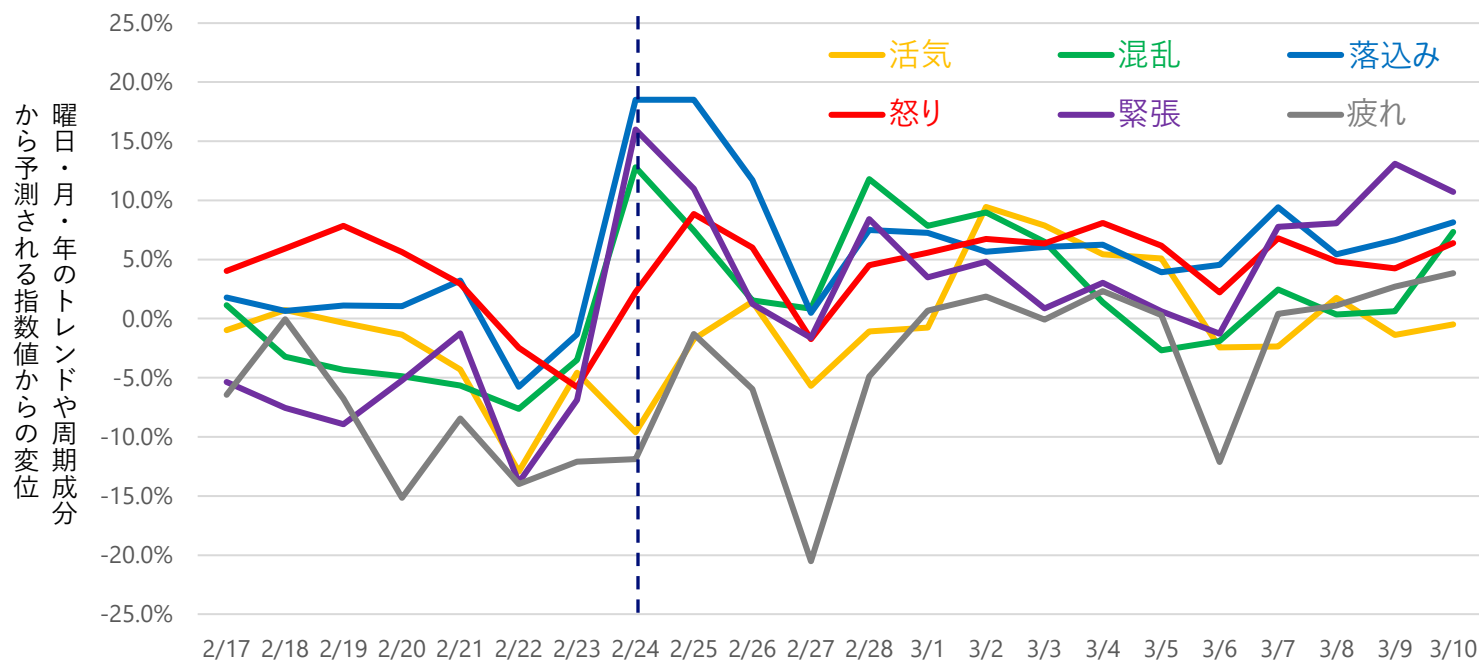
各指標のデیلیーの推移（2022年）



周期性などを排除すると、2月24日のウクライナ侵攻で「落込み」の拡大が明らかに

- 日本の空気感指数では、周期性や長期トレンドの影響を排除して、何もない場合（平常時）の水準と比較することで、特定の事象が、各指標の変動に影響したかどうかを把握可能
- 2月24日のロシアによるウクライナ侵攻の開始に伴い、「混乱」「落込み」「緊張」が大幅に増加していることがわかる

周期やトレンドを除去した変位の成分（点線は侵攻開始とされる2月24日）



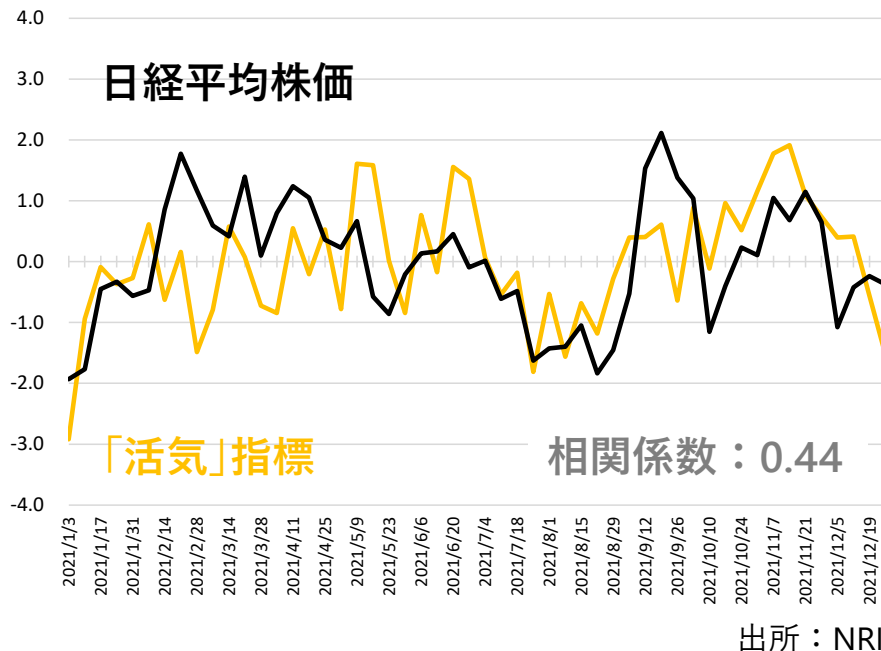
出所：NRI

外部変数との連動

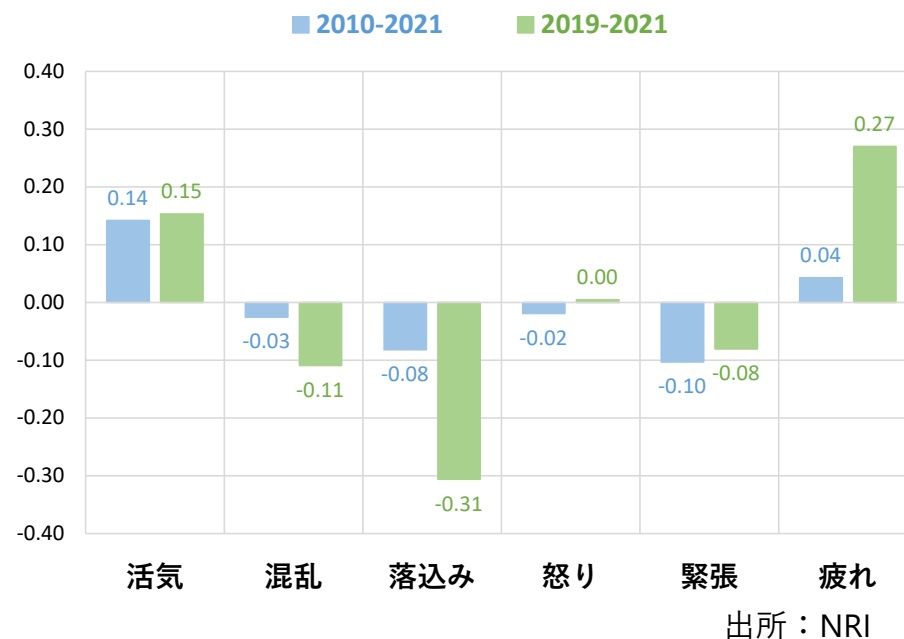
2021年のデータをみると、「活気」指標と、「日経平均株価」は相関している

- 時期や時勢により経済社会を駆動する空気感は異なるが、直近は、日経平均株価は「活気」指数と連動（過去の研究でも、ポジティブな感情値と株価指数に統計的に有意な関係があることが知られている）
- 過去のデータをみても、「活気」は安定的に相関があり、コロナ禍では、「落込み」や「疲れ」が大きく株価に影響
- 各指標と株価の関係をみることで、現在の日本経済がどのような背景にあるかを把握することも期待できます

2021年週次の日経平均と「活気」指数を
スケールを合わせて表示



各指標と日経平均株価の月次差分の相関係数
(2010～2021年、2019～2021年)

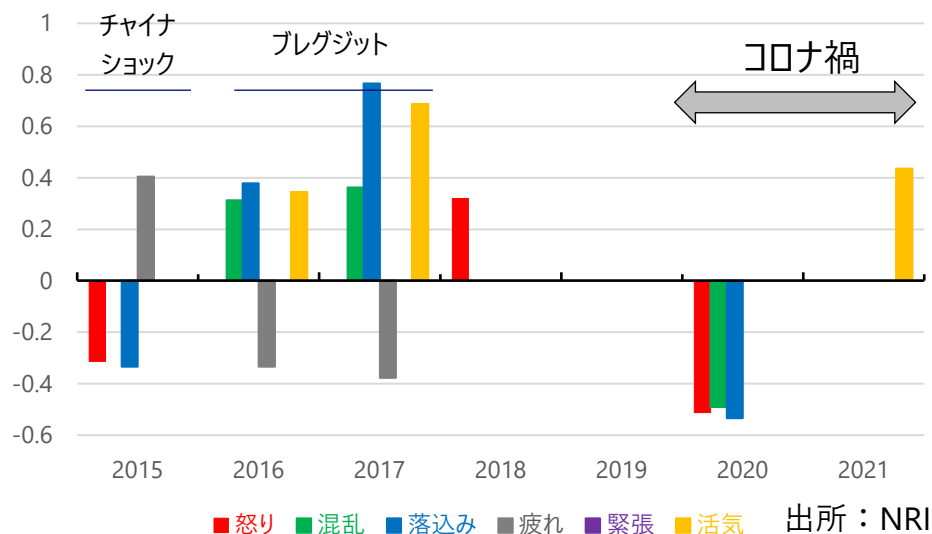


各時期別に、株価と相関のある指標は異なっており、6指標の定点観測が重要

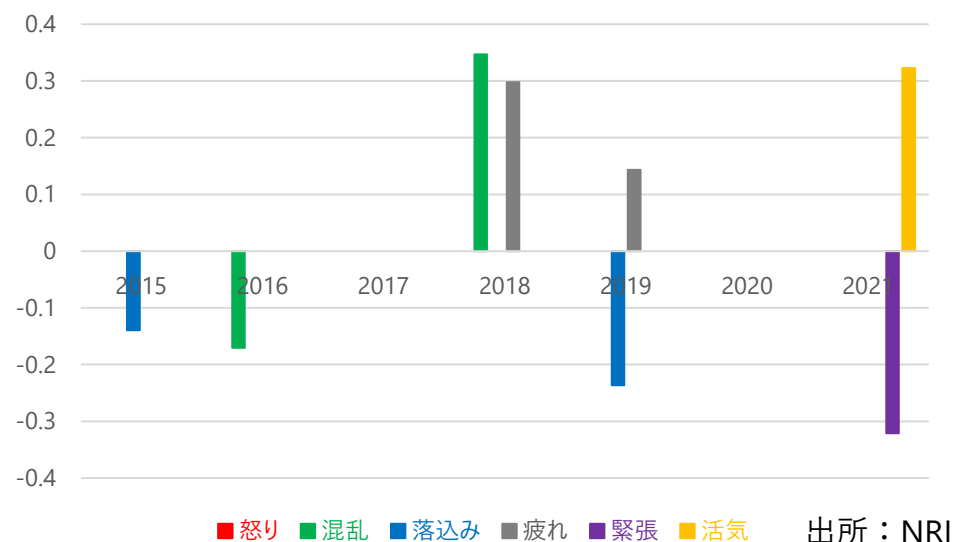
- 年次別に、各指標と日経平均株価の相関をみると、コロナ禍以前は、さまざまな社会的ニュースにより、株価が影響を受けており、相関のある指標や、相関関係（正・負）が変動している
- 近年では、コロナ関連のニュースに影響が強く、2020年は、混乱、落込み、怒りの指標と負の相関があり、株価の低迷に影響している
- 2021年は、活気指標と正の相関があり、活気が株価の上昇に影響し、緊張指標とは負の相関で、緊張の空気感が収まれば、株価は上昇する傾向にある

各指標と日経平均株価の相関係数

週次データの相関係数



週次対数差分データの相関係数



注）統計的に有意とされた相関係数のみ表示（p値<0.05の場合）

今後の展望

今後の「日本の空気感指数」に関する研究開発について

テーマ	研究開発の内容	活用の可能性
指数の精緻化	- 最新研究などを反映した新しい「指標」の追加（友好など） - 過去研究との整合比較を行い、指標の一般化	新しい指標を開発することで、日本の空気感を、より細かく表現することができる
指数の将来予測	- 短期的な予測値を算出するための時系列解析 - 各指標の変動に影響する事象の数理解析モデルの構築	過去のデータだけではなく「将来予測値」も公表することで、将来の戦略立案に反映できる
生活者関連の指標との相関分析	- 各種経済指標だけではなく、消費者嗜好データとの相関分析（NRIの生活者調査結果など） - 消費動向に結びつく要因の抽出	経済的な視点以外にも、各指標が影響を与えている項目を明確にし、政策提言などに反映できる
海外における指数開発	- 日本国内で開発したものと同一手法を海外にも適応可能 - 各国の価値観の違いを考慮することが課題	日本と海外各国の空気感を同じ指数として比較でき、国際社会における相対的な位置づけを確認できる

Appendix

(参考) SNSを利用した社会の空気感・感情の定量化に関するこれまでの研究

■ SNSを利用した社会の空気感・感情の定量化技術については、広く研究が行われている。

● 個別事象に着目した研究

- 気候に関するツイートのみを抽出し、幸福度(センチメント)付与。石油採掘ニュースで幸福度が落ちるなどが観測できる。
 - Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PloS One*, 10(8), 1–18.
- 日本語のツイート：コロナ関連語を用いて感情変動を抽出(ML-Ask)。恐れが最初の段階で多かった。
 - Toriumi, F., Sakaki, T., & Yoshida, M. (2020). ソーシャルメディアを用いた新型コロナ禍における感情変化の分析 Social Emotions Under the Spread of COVID-19 Using Social Media. *人工知能学会論文誌*, 35(4), F – K45_1–7.

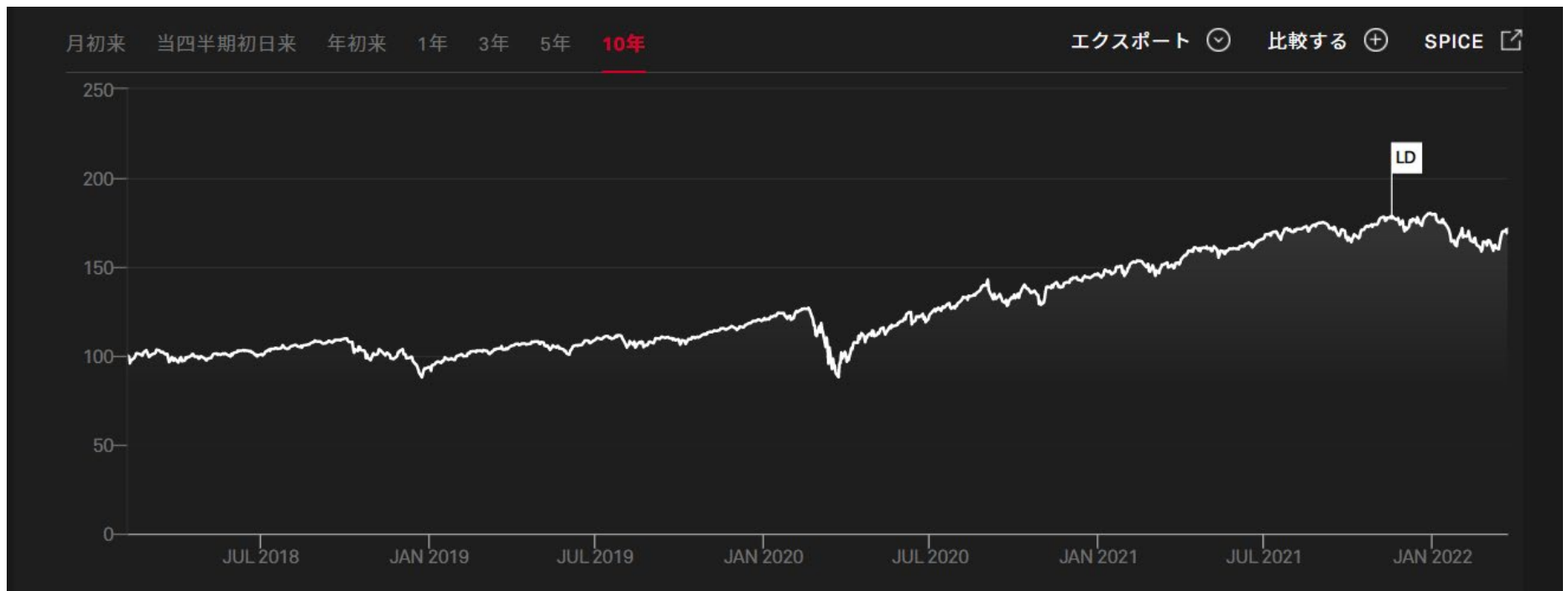
● インデクス化として検討された研究

- Twitterの大規模なデータを用いて、社会の幸福度を観測する。
 - Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS One*, 6(12), e26752.
- POMSを使ってTwitterからmoodを検出。POMSを拡張したGPOMSを使って、DJIAを予測。Social Sentimentによる経済指標分析の第1論文。
 - Bollen, J., Mao, H., & Pepe, A. (2011). Modeling Public Mood and Emotion : Twitter Sentiment and Socio-Economic Phenomena. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 450–453
 - Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.

(参考) SNSを利用した社会の空気感・感情の定量化に関するこれまでの研究

- Twitter社とS&P Dow Jones Indicesが提携し開発したS&P500ツイッター・センチメント指数は、株式投資での企業スコアにつながる指数。
 - Twitterの投稿からS & P500を構成する企業をスコア付けし、センチメントスコアの高い企業のパフォーマンスを測定する。

S&P500ツイッター・センチメント指数（算出開始日は2021年11月18日）



出所：<https://www.spglobal.com/spdji/jp/indices/strategy/sp-500-twitter-sentiment-index/>

The text is framed by two decorative swooshes. The top swoosh is a gradient bar transitioning from blue on the left to red on the right. The bottom swoosh is a solid blue bar.

Share the Next Values!