

特集 「パーソナルデータに基づく気付きの創発」

個人の生活管理のためのソーシャルデータからの行動パターンの発見

Discovering Behavior Patterns from Social Data for Managing Personal Life

潘 睿
Rui Pan
ZS アソシエイツ
ZS Associates, Inc.
rui.pan@zsassociates.com

松尾 豊
Yutaka Matsuo
東京大学大学院工学系研究科
School of Engineering, The University of Tokyo.
matsuo@weblab.t.u-tokyo.ac.jp. <http://weblab.t.u-tokyo.ac.jp/>

Keywords: social data, personal data, data mining, behavior patterns.

1. 生活管理の方法

どうすれば、毎日をより生産的に過ごすことができるだろうか。あるいは、毎日をより楽しく、充実して過ごすことができるだろうか。

こういった問いはおそらく誰しもがもっており、それぞれの人生経験や知識から、自分なりの生活スタイル、仕事のスタイル、自己規律の方法や信条など、何らかの「生活の質を上げるための方法」（本稿では生活管理の方法と呼ぶ）を身につけているのではないだろうか。これは、自分自身の行動をメタに観察し、その行動と成功体験、失敗した経験、快い感情や不快な感情と結びつけて学習され、蓄積されているものと考えることができる。

こうした生活管理の方法は、自身に関わることだけに主観的になりがちである。研究者であれば、あまり科学的思考になじみのない人が、迷信や占いを信じる、験（げん）を担ぐなど、一見すると非科学的なルールを見つけ出していることに疑問を抱くこともあるかもしれない。一般に、確率的な事象を人間が正しく把握することは困難であり、多くの人が「科学的には正しくなさそうに見える」ルールを見つけ出しているのも、ある程度やむを得ないことかもしれない。

しかし、近年では、人間の行動を観測できるさまざまなデータがある。現在注目を集めているのは、個人に特化しながらも客観性を保っているパーソナルデータである。パーソナルデータを活用すると、科学的な分析手法に基づき、その人自身に良い気付きをもたらす結論を出すことができ、それによって行動の変化を促しやすくなる。さらに、他人と比べることで自身の特性や差異をより客観的に理解することができる。

なかでも、Web、特にソーシャルメディア上のデータが増えてきている。さまざまな Web サービスやスマー

トフォンの普及によって、Web を使用する時間が増え、Web 上に存在するデータが我々の生活とますます密接に関わるようになった。毎朝起きて、Gmail をチェックし、オンラインのニュースを見て、Facebook を確認する。おもしろいことがあれば、つぶやいたり、いいねしたりする。仕事の報告はメールで行い、友人や同僚との連絡もさまざまなソーシャルメディアで行う。したがって、Web にあるパーソナルデータによって、個人をますます正確に、多面的に理解できるようになってきているのかもしれない。各種のサービスの API が整備されてきている状況も、こういった可能性を後押ししている。

本解説では、Web 上にあるパーソナルデータを用いて、生活管理の方法を見つけ出す著者らの研究を事例として紹介する。同様の研究は世界でもほとんど行われておらず、こういった研究を広く紹介できないのは不本意であるが、Web 上の情報からのマイニングに関して関連した研究については少し紹介する。

Web からパーソナルデータをとる方法として、大きく分けると当人から発信した一般公開された内容を取る方法と、承認を取ったうえで制限された内容を取る方法がある。例えば、twitter では、一般公開したツイートである限り、ユーザ名を特定できれば個人のツイート内容や時間を取得することが可能である。Gmail は、ユーザに承認してもらくと、当人が設定した期間内のみデータを取得することができる。Facebook は、ユーザに承認してもらくと、当人と合意した項目（例えばフィード）のみ取得することができる。本稿では、こうして得られる Web 上のデータをソーシャルデータと呼ぶことにする*1。

*1 厳密には、Gmail はソーシャルメディアでないで、ソーシャルデータという言い方は適切ではないが、Facebook や twitter に代表されるデータという意味でソーシャルデータと呼ぶ。

2. Webからの生活管理に関するマイニングの関連研究

ソーシャルメディア上のデータを活用し、人の特性を捉える研究にはさまざまなものがある。例えば、Golbeck らが行った Facebook [Golbeck 11a] や twitter [Golbeck 11b] に関する研究では、ソーシャルメディアから取得したデータを分析し、ユーザの性格を Big Five という五つのカテゴリー（開放性、外向性、協調性、誠実性、情緒不安定性）に高い精度で分類できることを示した。これは、ソーシャルメディアのデータがネット上の行動だけでなく、現実世界でのユーザ特性を反映していることを表している。

O'Connor らは、twitter のつぶやきをポジティブ・ネガティブに分類し、それをユーザのポジティブ・ネガティブな感情を表す指標として用いた [O'Connor 10]。Go の研究 [Go 09] と、それに続く研究 [Davidov 10, Marchett-Bowick 12] では、スマイル「:」のような感情を表す絵文字や、「#happy」のようなハッシュタグを、手動ラベルに代わる Noisy Label として使用し、Distant Supervision と呼ばれる方法で分類器を訓練し、twitter に表される感情を分析している。

仕事をどのくらい効率的に進められるかという「能率」を定量化する研究としては、Bulkley と Alstyne の E メール研究 [Bulkley 06] がある。ここでは、E メールを用いる行為が仕事の能率に強い関係性をもつことを示している。例えば、同僚など近い関係の人に返信するスピードが速い、もしくは、関係の遠い人にメールを送る場合に、仕事の能率が高い。後に述べる著者らの研究事例でも、この研究を参考に、Eメールの行為から日常の能率を表す変数をつくっている。

行動以外にも、天気や温度などの環境要素が日常の質に影響を与える研究として、Denissen ら [Denissen 08] の研究がある。冬には気温が高いほど、人の機嫌にポジティブな影響を与えるが、夏ではネガティブに働く。後述の分析でも、天気という環境要素が日常の能率や機嫌へ与える影響が見られ、同じ結果が得られた。

こうした研究は、常に、一般的な知見なのか、その人個人に適用できる知見なのかという葛藤がある。Williams らは、一般的な規則だけを個人に適用することの有効性に対する疑問を提示した [Williams 08]。また、個人に自身のデータを見せることで、気付きを与え、行動を変える有効性に関する研究として、Allcott ら [Allcott 11] の研究がある。著者らの研究事例でも、一般的なデータからパターンを抽出するだけでなく、ある一人のデータからパターンを抽出することで、より有効的に個人に気付きを与えることを意図している。

3. 日常生活の質と行動

本稿で紹介する著者らの研究事例は、Web から得られたソーシャルデータをパーソナルデータと考え、それを収集、分析し、生活管理の方法を見つける研究である。複数のソーシャルメディアやメールのデータから、日常生活の質に影響する可能性のある要因を素性として取り出し分析する。

本研究で生活の質として扱うのは、能率と機嫌である。仕事をしている人の多くにとって、仕事があまくできたかどうかは日々の生活の中で大きな関心事であろう。能率の高い日もあれば、なかなか仕事はかどらない日もある。したがって、能率が高い要因を知ることは、仕事をする人にとって重要な知見となり得る。一方、日々の生活において仕事はかどるだけで良いかというもちろんそれだけではない。仕事とは無関係に、充実した楽しい人生を送れているかはまた別の極めて重要な軸である。楽しい日々を送ることができているかを計るための、その日の感情の善し悪しをここでは機嫌と呼ぶことにする。

本研究を行うことで、例えば、当日の機嫌や能率にポジティブ・ネガティブな影響を与えるソーシャルメディアの使用行動を見つけることができる。また、そうした要因の相互の影響関係を取り出すことで、ソーシャルメディア使用の行動を促進・抑制する環境や行動といったパターンを発見することができる。こうしたパターンを抽出することによって、個人により深く自身の行動の理由や結果を理解させ、生活の質を向上させるような行動を促進することができる。

ここで重要なことは、こうした能率や機嫌といった変数が、自分のコントロール可能な行動と結びついているかである。環境や他人の行動は自分でコントロールすることは難しいが、自分の行動との影響関係をつかむことができれば、自ら意識して能率を上げたり、毎日を楽しむことができるかもしれない。したがって、毎日をより楽しく、能率の良い日にするためには、まず自身の行動をはじめとするコントロールできる要素がいかに1日の質に影響を与えているかを理解する必要がある。

図1で示すとおり、本研究において対象とする問題は主に以下の二つである。

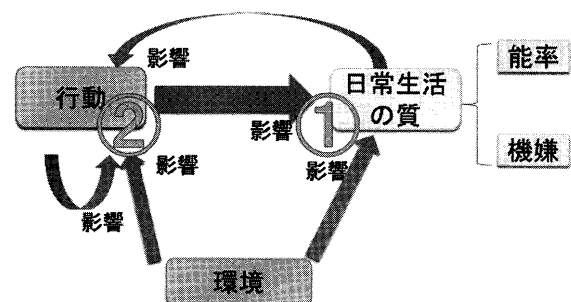


図1 対象とする二つの主な問題

1. 自身の行動や環境といった要素が、いかに1日の生活の質（能率と機嫌）に影響を与えているか
2. どんな行動や環境が、自身の行動に影響を与えているか

すなわち、1は、能率および機嫌という目的変数を決めたうえで、これに影響する要素を知ることが目的であり、2は、さまざまな行動が、ほかの行動や環境からどのような影響を受けているかを見るのが目的である。

4. データと素性の構築

本研究では主に Gmail, Facebook, twitter からデータを収集する。Gmail からは、図 2 で示すように、送信アドレス、受信アドレス、送信時間、メール ID、返信メール ID など、E メール相手とのインタラクションに関する情報を収集する。Facebook からは、図 3 で示すように、ユーザの行動を示すフィードデータを収集する。Story は行動の概要、from は行動を起こしたユーザ、type は行動のカテゴリである。twitter からは、図 4 に示すとおり、ツイートの内容とツイートした時間を収集した。

```
{
  "From":
  "dbc0ce9052996550862ae27b2403202f83d09bf9",
  "To":
  ["07d8e5d4ca258a9d54e3cdcecae7a3b543bf73d2"],
  "Message-ID":
  "da3236f342dfab46cf602bb4505663c126953e1f"
  "References":
  ["3e8b54ee1bac37a265e1d1bb0f1252bd4a101e5a"],
  "In-Reply-To":
  "3e8b54ee1bac37a265e1d1bb0f1252bd4a101e5a",
  "Date": "Thu, 30 Jun 2011 00:00:00 +0900", }
```

図 2 Gmail データ

```
{
  "story": "FBUser likes ABC Solutions.",
  "from":
  {
    "name": "FBUser",
    "id": "0123456789"
  },
  "comments": {
    "count": 0,
    "updated_time":
    "2012-06-30T00:00:00+0000",
    "story_tags": {
      "0":
      [
        {
          "length": 5,
          "offset": 0,
          "type": "user",
          "id":
          "0123456789",
          "name": "FBUser"
        },
        {
          "length": 38,
          "offset": 12,
          "id":
          "987654321012",
          "name": "ABC Solutions: Solutions of ABC problem"
        }
      ]
    },
    "created_time":
    "2012-06-30T00:00:00+0000",
    "type": "status",
    "id": "0123456789_01234567890123"
  }
}
```

図 3 Facebook データ

```
{
  Sat Jun 30 00:00:00 +0000 2012
  RT @DEF What a beautiful day! I just have to took a picture. Check it out! http://a.b/cdefgh #ABC
}
```

図 4 twitter データ

本研究では主に 3 種類の素性を設計する。まず 1 日の質、すなわち能率と機嫌を表す素性である。次に 1 日の中の行動を表す素性を、さまざまなソーシャルデータから構築する。さらに、行動以外に生活の質に影響し得る環境面の要素として、ここでは天気を表す素性を用いた。以下では、それぞれをどのように構築したかを述べる。

能率を表す素性 (TODAY_PERFORM) の設計には、Bulkley と Alstyne の研究 [Bulkley 06] を参考にした。彼らの研究では、近い関係をもつ人からももらった E メールに早く返事する、あるいは関係の遠い人に E メールを送るといった行為が、売上や収入で計った仕事の能率と正の関係性をもつことを示している。したがって、本研究では、E メールをやり取りする相手を関係の近さによって二つのクラスに分ける。クラス A は関係の遠い人達であり、クラス B は関係の近い人達とする*2。そして、以下の式で能率を表す素性を作成した。MsgA, MsgB はクラス A とクラス B への送信数であり、送信数が多いほど能率が高くなるように作成した。さらに、関係の近い人への返信時間 (RES) が短いほど能率が高いように、平均返信時間 μRES で標準化した数値を加えた。

$$P = Msg_A + \sum_{j=1}^{Msg_B} \left(1 + \frac{(RES_j - \mu_{RES})}{\sigma_{RES}} \right)$$

機嫌を表す素性 (TODAY_SENTI) の設計においては、Davidov らの研究と同じ手法により、ツイートをポジティブ・ネガティブに分類した。1 日の中でポジティブ感情をもつツイート数 (TP)、ネガティブ感情をもつツイートの数 (TN) の差を機嫌の値として用いた。

1 日の行動の素性に関しては、まず Gmail や twitter, Facebook の使用頻度を表す素性をつくった。1 日といっても、その始まり (起床) と終わり (就寝) を捉える必要がある。そのため、行動がある程度連続した期間を 1 日として、1 日の最初に計測できた行動タイミング (例えば最初に Gmail で E メールを送信した時間) を起きる時間、最後の行動のタイミング (例えば 1 日最後のツイートを送った時間) を寝る時間を示す素性とした。そして 1 日 (起床時間から就寝時間) の中での Gmail, twitter, Facebook の使用頻度を測定した。

行動が生活の質に及ぼす影響は、その日だけでなく、次の日以降にも引き継がれる可能性 (例えば、前日の寝不足が次の日の能率を損なうなど) があるので、観測する各対象日において、当日、前日、前々日の 3 日分の素性を作成した。

また、能率が高い、機嫌が良いというのは、その人自

*2 この二つのクラスを得るために、送信数、受信数、返信までの平均時間、スレッドの長さ、相互のメッセージ、重要度の比率などを素性としたクラスタリングを事前に行う。返信の確率が高く、重要度の高いクラスがクラス A であり、返信の確率が低く、短いスレッドが多いクラスがクラス B である。

表1 素性リスト

	当日	前日	前々日	個人平均
能率	TODAY_PFORM	LD_PERFORM	LLD_PERFORM	AVER_PERFORM
機嫌	TODAY_SENTE	LD_SENTE	LLD_SENTE	AVER_SENTE
行動 (ネット)	TODAY_PRES TODAY_FBCOUNT TODAY_TTCOUNT	LD_PRES LD_FBCOUNT LD_TTCOUNT	LLD_PRES LLD_FBCOUNT LLD_TTCOUNT	AVER_PRES AVER_FBCOUNT AVER_TTCOUNT
行動 (睡眠)	TODAY_BEDTIME TODAY_UPTIME TODAY_SLEEP	LD_BEDTIME LD_UPTIME LD_SLEEP	LLD_BEDTIME LLD_UPTIME LLD_SLEEP	AVER_BEDTIME AVER_UPTIME AVER_SLEEP
環境 (天気)	DAYLIGHT HIGH_TEMPER HUMIDITY LOW_TEMPER			

身が能率が高い、もしくは機嫌が良い人であるかよりも、ある1日がほかの日と比べて高いか低いかにか意味がある。したがって、行動や能率、機嫌などの各素性の平均値も素性とした。

環境に関する素性としては、天気、湿度、最高気温および最低温度を用いた。用いた素性の全体をまとめたものが表1である。

5. パターンの抽出

本研究では、次のような分析を行う。複数のユーザーから得られたデータを集合データと呼び、まず、集合データを用いて、行動をはじめとする要素が能率や機嫌にどのような影響を与えるかという一般的なパターンを抽出する。各要素が能率（もしくは機嫌）に対する影響を、2クラスの分類問題として解くことで算出する。すなわち、行動や環境などを表す素性をもって、TODAY_PFORM、TODAY_SENTEが普段より高いかどうかの{+1, -1}の2クラスに分類する問題を解く。得られたモデルでの各素性の重みを見ることで、重要度の高い要素を見つけることができる。以下では、重みの符号を+または-で表し、能率（もしくは機嫌）への影響がポジティブかネガティブかを示すとする。

次に、行動に影響する要素を調べるために、能率や機嫌ではなく、各行動自身を予測する分類問題を構成することで、各行動がどのような要素に影響を受けているかを分析する。

さらに、集合データではなく、一人分のデータだけ使用し、個人のパターンを抽出する。また、全期間のデータではなく、一時期のデータのみを使用した分析も示す。これによって、個人や時期特有のパターンを抽出することができることを示す。

6. 実験と結果

本研究で扱うデータは、Gmailをはじめとする個人情報なので、実験の協力者を得るのが難しいが、50人の協力者を募り、1年分（2011年7月1日から2012年

6月30日まで）のGmail, Facebook, twitterデータを収集した。合計Gmailの送受信Eメール数は74246件、Facebookのフィード6388件、twitterのツイート29175件である。

6.1 能率に関連するパターン

まず50人すべての集合データを使用し、その日の能率を予測した。交差検定を行い、83.7%の精度でその日の能率の良し悪し（平均より高い・低いこと）を予測することができた。

重みが上位である素性、すなわち能率により大きな影響を与える要素は表2が示すとおりであり、以下のいくつかのパターンが見られる。

1. Eメール関連の行動が能率へ及ぼす影響：当日のEメールによるプレッシャー（TODAY_PRES、すなわち関係の近い人から受信したEメール数）が高い場合、能率が普段より高い傾向がある。
2. ソーシャルメディアにおける行動：前日にFacebook（LD_FBCOUNT）を多く使用すると、当日の能率が低くなる傾向がある。当日のFacebook, twitter使用（TODAY_TTCOUNT, TODAY_FBCOUNT）はそうではない。
3. 睡眠の能率への影響：前々日の就寝時間が早く（LLD_BEDTIME）、前日の寝る時間（LD_SLEEP）が長いと、当日の能率が高くなる傾向がある。
4. 前日の能率の引継ぎ：前日の能率（LD_PERFORM）が高いと、当日の能率も高くなる傾向がある。
5. 環境が能率への影響：1日の最高気温（HIGH_TEMPER）が高いと、当日の能率が低くなる傾向がある。

表2 能率に影響する主な要素（集合データ1年分）

素性	重み
TODAY_PRES	8.043
TODAY_TTCOUNT	1.231
TODAY_FBCOUNT	1.030
LLD_BEDTIME	-0.766
LD_SLEEP	0.712
LLD_PRES	-0.662
LD_FBCOUNT	-0.659
TODAY_BEDTIME	0.637
HIGH_TEMPER	-0.547
AVER_UPTIME	-0.540
LD_PERFORM	0.526

6.2 機嫌に関連するパターン

次に、その日の機嫌を予測した結果、機嫌の良し悪しを73.0%の精度で予測することができた。能率と比べるとこの値は少し低くなっている。

機嫌により大きな影響を与える要素、すなわち予測における重みが上位の素性は、表3に示すとおりである。その日のものが少なく、主に平均的な指標である。すな

表3 機嫌に影響する主要素 (集合データ1年分)

素性	重み
AVER_SENTI	0.933
AVER_BEDTIME	-0.133
AVER_SLEEP_TIME	0.067

わち機嫌は、能率ほどその日の行動に左右されず、その人自身のもつ特性を表す指標が大きく寄与する。

1. 平均的な機嫌：機嫌の良い人 (AVER_SENTI) のほうが、その日の機嫌も良い。
2. 睡眠時間：より早く眠って (AVER_BEDTIME)、睡眠時間が長い場合 (AVER_SLEEP_TIME)、その日の機嫌も良くなる傾向がある。

6.3 行動のクロスパターン

1日の質 (能率・機嫌) を影響する要素パターンがわかったので、次に、それらの要素を促進・抑止する要素のパターンを抽出する。同じ手法を用いて、各要素でほかの要素を予測させることで、ポジティブ・ネガティブの関係性と強さに関わるパターンを抽出した。その結果を表4に示す。

表4 主な関連する要素 (集合データ1年分)

*重みが0.1以下のものは空欄にした

	TODA TIME	TODAY _BED _E	TODA _SLE _EP	TODA _PRE _SS	TODAY_ TTCOUN _T	TODAY_ FBCOUN _T
TODAY_ UPTIME	0.725					
TODAY_ SLEEP	-1.140	1.566				
TODAY_ BEDTIME		1.361	-0.870			
LLD_ UPTIME			0.634			
LLD_ TTCOUNT				0.029		
LLD_ PRES				0.148		
LLD_ FBCOUNT					0.491	
LLD_ BEDTIME				0.172		
LD_ TTCOUNT				0.066		
LD_ FBCOUNT					0.508	
TODAY_ SENTI					-0.330	
DAY LIGHT					0.430	

表4の中には、左側 (表側) に書いてある要素を用いて、トップ行 (表頭) に書いてある各要素を予測させるときの重みを記載している。要素の重みが0.1以下のものは空白としている。表の中から以下のようなパターンを見つけることができる。

1. 機嫌とツイート頻度の関係：当日の機嫌 (TODAY_SENTI) に関し、当日のツイート数 (TODAY_TTCOUNT) の重みが負であるので、機嫌が悪い日のほうが、ツイート数が多い傾向がある。

2. 日照時間とツイートの関係：当日の日照時間の長さ (DAYTIME) と当日のツイート数 (TODAY_TTCOUNT) の関係を見ると、日照時間が長い日のほうが、ツイート数が多い傾向がある。

3. ソーシャルネットワーク：前々日 (LLD_) や前日 (LD_) の Facebook の使用頻度 (FBCOUNT) と twitter の使用頻度 (TTCOUNT) の、当日 (TODAY_) の使用頻度に対する影響を見ると、前々日や前日の twitter 使用頻度は当日とあまり関係がないが、Facebook の使用頻度は前々日、前日から当日にプラスの影響を与えている。したがって、twitter に比べ、Facebook のほうが、その日の使用が次の日の使用につながる。

6.4 個人のパターン

今までは、集合データを使用してきたが、より有効に個人に気付きを与えるために、一人だけのデータを使用し、能率や機嫌、そして行動に影響を与えるパターンを抽出した。以下の結果は、ある特定の人物のデータから得られたパターンである。

能率を予測することで得られたモデルは、表5のとおりである。個人のパターンと集合データのパターンの違いが見られることがわかる。例えば、当日のメールの受信数 (TODAY_PRES) は、集合データに比べて低いので、メールの受信が多くても気にならない傾向にあるといえる。このように、自身のパターンと一般的なパターンを比較することによって、さらに自身を深く理解することができるだろう。

表5 個人パターンと集合パターンの比較

素性	個人データにおける重み	集合データにおける重み
TODAY_PRES	5.738	8.043
LLD_TTCOUNT	-0.936	-0.225
TODAY_FBCOUNT	0.933	1.030
LLD_PRES	-0.915	-0.662
TODAY_TTCOUNT	0.824	1.231
LD_FBCOUNT	-0.546	-0.659
DAYLIGHT	0.365	0.257

6.5 時期特有のパターン

さらに、1年ではなく、限られた時期のデータを使用しパターンを抽出することで、個人の変化や環境の変化による時期特有なパターンを発見することができる。一例であるが、例えば表6のように、梅雨を含む時期のデータでは、湿度 (HUMIDITY) が上がると能率が下がる傾向は顕著になることがわかる。

表6 一時期のパターンと1年間のパターンの比較

素性	6か月データに おける重み	1年データにお ける重み
TODAY_PRES	4.736	5.738
LLD_TTCOUNT	-1.207	-0.936
TODAY_FBCOUNT	0.981	0.933
LLD_PRES	-0.964	-0.915
HUMIDITY	-0.691	-0.290
TODAY_TTCOUNT	0.669	0.824
LD_TTCOUNT	-0.481	-0.268

7. 考察とまとめ

本稿では, Gmail, Facebook, twitterといったデータ(ソーシャルデータ)を用いて, 生活管理の方法を見つけるといった研究を紹介した. その結果, 例えば, 「前日 Facebook を多く使用すると, 当日の能率が低い傾向がある」, 「一般的に早く寝て, 長く眠ったほうが, 当日の機嫌が良い」, 「機嫌が悪い日に, 普段より多くツイートする傾向がある」などの知見を導き出すことができた. これらは, 多くの人が何となく感じていたことかもしれないが, それをデータから示したことに意義があると考えている.

もちろん, 本稿で示した方法には限界がある. まず, ソーシャルメディアを使用していない人は分析できない. また, ツイートや Eメールの返信の自動設定をしている人, 所属する組織の PR など特殊な用途に用いている人などについては, 結果が実際の生活を反映しない可能性がある. 本研究では, 各個人の行動を1日を単位として区切っているが, それには, 一定のリズムがある生活を送っていることを前提にしている. 例えば, 医師や看護師のような夜勤の多い人, あるいはキャビンアテンダントのようなタイムゾーンをよく変更する人のデータは正確に分析することが難しいかもしれない.

ここで紹介した方法は, ソーシャルデータを用いているが, それ以外にも, 生体データ, GPS データ, モバイルアプリデータなどを活用することにより, さらに詳細に精度良く, 自らの生活管理に役立つパターンを抽出することができるだろう. また, リアルタイムに分析結果をフィードバックすることで, 「このままだと今日は仕事の能率が悪そうなので何か工夫しよう」というように, 行動を変えることにつながるかもしれない. また, 集合のパターンと個人のパターンを比較することで, 自身の特性を理解し, 仕事を選ぶなど人生の大きな決断をするときの判断材料にもなるかもしれない. ソーシャルデータから我々の行動パターンを捉え, より良く理解するこ

とで, セルフコントロール, 合理的な人生決断, そして充実した1日につながっていくことを期待している.

◇ 参考文献 ◇

- [Allcott 11] Allcott and Hunt: Rethinking real-time electricity pricing, *Resource and Energy Economics*, Vol. 33, No. 4, pp. 820-842 (2011)
- [Bulkeley 06] Bulkeley, N. and V Alstyne, M. W.: An empirical analysis of strategies and efficiencies in social network, *Proc. Int. Sunbelt Social Network Conference XXVI* (2006)
- [Davidov 10] Davidov, D., Tsur, O. and Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys, *Proc. 23rd Int. Conf. on Computational Linguistics (COLING)* (2010)
- [Denissen 08] Denissen, J. J. A., Butalid, L., Penke, L. and van Aken, M. A. G.: The effects of weather on daily mood: A multilevel approach, *Emotion*, Vol. 8, No. 5, pp. 662-667 (2008)
- [Go 09] Go, A., Bhayani, R. and Huang, L.: Twitter sentiment classification using distant supervision, Technical Report, Stanford Digital Library Technologies Project (2009)
- [Golbeck 11a] Golbeck, J., Robles, C. and Turner, K.: Predicting personality with social media, *Proc. 29th ACM Conf. on Human Factors in Computing Systems (CHI)* (2011)
- [Golbeck 11b] Golbeck, J., Robles, C., Edmondson, M. and Turner, K.: Predicting personality from twitter, *Proc. IEEE Int. Conf. on Privacy, Security, Risk, and Trust, and IEEE Int. Conf. on Social Computing* (2011)
- [Marchetti-Bowick 12] Marchetti-Bowick, M. and Chambers, N.: Learning for microblogs with distant supervision: Political forecasting with twitter, *Proc. European Chapter of the Association for Computational Linguistics (EACL)* (2012)
- [O'Connor 10] O'Connor, B., Balasubramanyan, R., Routledge, B. R. and Smith, N. A.: From tweets to polls: Linking text sentiment to public opinion time series, *Proc. 4th AAAI Conference on Weblogs and Social Media* (2010)
- [Williams 08] Williams, B., Myerson, J. and Hale, S.: Individual differences, intelligence, and behavior analysis, *J. Experimental Analysis of Behavior*, Vol. 90, No. 2, pp. 219-231 (2008)

2013年10月4日 受理

著者紹介



潘 睿

2010年東京大学システム創成学科卒業. 2012年同大学院工学系研究科技術経営戦略学専攻修士課程修了. 主な研究分野は, ソーシャルデータのマイニング, 仕事の動機付け. 現在, ZS アソシエイツにてデータ解析を利用した経営コンサルティングの仕事に従事.



松尾 豊 (正会員)

1997年東京大学工学部電子情報工学科卒業. 2002年同大学院工学系研究科博士課程修了. 博士(工学). 産業技術総合研究所研究員, スタンフォード大学客員研究員を経て, 2007年より東京大学大学院工学系研究科准教授. 専門は, 人工知能と Web マイニング. 2012年より本誌編集委員長.