

SNS の投稿を用いた感情記録ライフログシステム ～ Emote ～

三井健史^{1, a)} 伊藤智也¹ 中西勇人¹ 濱川礼¹

本論文では, SNS の投稿に含まれた感情を記録するライフログシステム「Emote」について述べる. 近年, 様々なライフログシステムが流行している. ライフログのために画像情報, 音声情報, 位置情報, 感情情報などを記録している. 画像情報, 音声情報, 位置情報はデバイス機器の発達により自動化が発達した. だが, 感情取得の自動化は中々進んでいない. そこで Twitter, Facebook の投稿文の感情を分析し, 分析結果を自動的に記録していくライフログシステム「Emote」を開発した.

「Emote」では Twitter, Facebook 各投稿から感情分析を行い, 各感情を可視化し手軽に利用できるライフログシステムの提供を目的とする. 事前に SNS (Twitter) の投稿データ, 約 3 万件を収集し, これらの投稿から自然言語処理とナイーブベイズ分類器を用いて単語ごとの感情を学習させた. 「Emote」では, その学習データを用いて Twitter, Facebook に投稿されている各文章の感情を決定している. それらを用いて, 各投稿を自動的に 6 つの感情に当てはめ感情をカレンダーとともに色で表現している.

1. 背景・目的^a

現在, 情報化社会と言われるほど日常に大量の情報が溢れている[1]. [2]によると, 「インターネットの普及に伴って, 一般消費者・個人が発信・受信する情報は, 量・種類共に増大の一途をたどっている. 情報化社会の中で, インターネットサイトにおける検索, 購買に関する記録や, スマートフォン等に搭載される GPS 等の位置情報に関する情報など, 行動の履歴情報ともいえるべき「ライフログ」がインターネット上に蓄積され, それを収集・活用するサービスが注目を集めている」と述べている.

ライフログシステムの例として twilog[3]や miin[4]がある. twilog は Twitter の tweet をブログ形式で保存, 閲覧が出来るサービスである. miin はプライベートなクラウドサーバー上に文字, 画像, 動画を記録することができ, SNS とアプリ連携を行うことで, 他人と共有することが出来るシステムである.

ライフログを記録するためには, ユーザはそのためだけに行う手間はできるだけ少ないほうが良い. 画像情報, 音声情報, 位置情報等はデバイス機器の発達により自動化が加速した[5]. しかし感情取得の自動化は中々進んでいない. 計測装置[6]はあっても 24 時間つけるのは現在では現実的ではない. そこで, 我々は SNS に着目した. SNS では, 日々に文字を書いている人が増加している. これを活用すれば, 余計な手間を取らせない. 実際にこれに着目して感情を抽出している研究もある[7].

それらの事実を踏まえて, twitter[8]と facebook[9]を用いて感情を自動的に記録し, それらをカレンダーとして見ることでユーザ自身がカレンダーの感情を見た際にその日の出来事を思い出せるよう支援する本システム「Emote」を提案する.

2. 関連研究

「Emote」では, ユーザ自身が過去のカレンダーの感情を見た際その日何があったか思い出すことを支援することを目的とする. 以下に関連研究を述べる.

2.1 感情分析

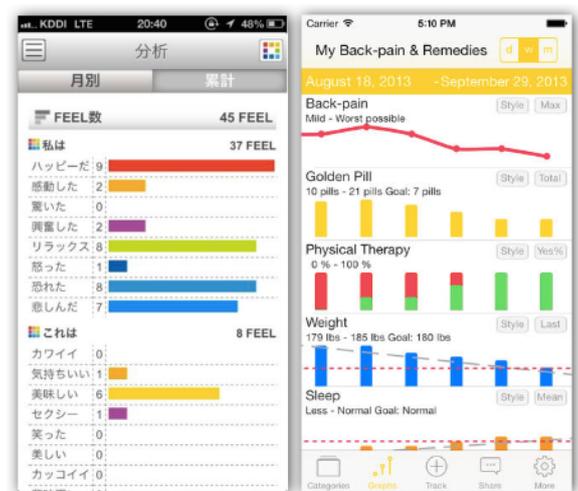
加藤らは「Twitter から得られる自然言語情報を用いて行う単語への感情付加手法」を行なっている. Twitter から得られる言語を分析し, 一つひとつの単語の感情値を計算することで, 全ての単語に感情情報を付加する手法である[10].

Johan Bollen らは, 「Twitter のムードを使用した株式予測」を行っている. Twitter の Tweet から感情に関わる情報を引き出すことで, ダウ平均株価を予測する手法である[11].

2.2 ライフログ

feelpress[12]では, 興奮した, 感動した, 美しい, 笑った, 等の予め用意された 16 種のフィーリングの中から今のフィーリングにマッチするモノを手動で選択して記録する. 図 2.1 が感情を表示する画面である.

track&share[13]では, ユーザが記録したいアイテムを設定する. 例えば, 幸福度や体調, やる気などである. そのアイテムを段階で分け, 記録する際にユーザが段階を選択する. これらの動作を行い, 図 2.2 のグラフ表示機能を使用すると今まで蓄積されたログを見ることが出来る.



左図 2.1 feelpress の感情表示画面

右図 2.2 track&share のグラフ表示画面

¹ 中京大学情報理工学部情報システム工学科

^{a)} h111108@st.chukyo-u.ac.jp

2.3 「Emote」との比較

「Emote」では、ユーザが SNS で投稿した際の感情分析を「自動的」に行い、カレンダーに表示する。ユーザが自身のカレンダーに表示されている感情を見ることで、当時の記憶を思い出せるように支援することが目的である。加藤らは各単語にユーザが持つ主観的な単語に対する感情を取得する。Twitter から tweet を取得し、tweet を品詞などに分解し、品詞ごとに感情値を計算する。そしてそれらをデータベースに格納する。Johan Bollen らの「Twitter のムードを利用した株式予測」は、取得した Tweet を OpinionFinder と GPOMS にかけて 7 種類の感情に関する時系列データを作成する。これらのシステムでは、文章としての感情ではなく単語毎の感情値が出るため、総合的な文章の感情値は取得できない。

ライブログでは、特別な手間がないことが望ましいとされている。feelpress では、予め用意された 16 種類の感情の中から手動で記録を行うため、手間が増える。track&share では、ユーザが記録したいアイテムを自分で設定を行い段階を手動で記録していく。これらのシステムは手間がかかってしまう。その点「Emote」では、感情分析からカレンダー表示まで自動で行うため特別な手間が必要なくシステムを使用することが出来る。

2.4 感情

2.4.1 感情の種類

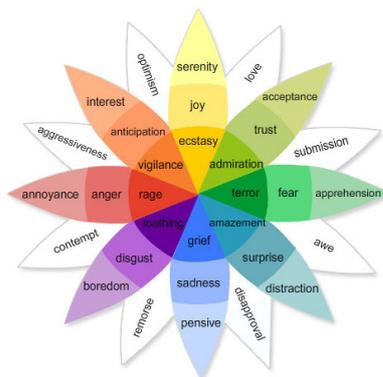


図 2.3 Plutchik の感情の輪

「Emote」では、心理学者 Robert Plutchik 考案した感情モデルである感情の環を元に感情を選択した[14][15]。「Emote」はライブログであるため、感情は人間が抱くシンプルで身近なものを選択しようと考えた。そこで Plutchik の感情の環は人間が抱くシンプルな感情で構成されているため、「Emote」に最適だと考えた。

感情の環は、「喜び」、「悲しみ」、「嫌悪」、「憧れ」、「怒り」、「恐怖」、「驚き」、「期待」の 8 個の感情から構成されている。これらは、喜び-悲しみ、受容-憧れ、怒り-恐怖、驚き-期待、のように感情が対になって配置され、縦断面は各感情の強度を表されている。

「Emote」では、「喜び」、「悲しみ」、「怒り」、「恐怖」、「驚き」、「期待」の 6 個の感情を使用した。

感情データを学習している際、「嫌悪」と「信頼」という感情に分類されることが少なく、学習データが他に偏ってしまった。例

えば「勉強が嫌だ」という文章があるとする。この場合、「嫌悪」に分類するより「悲しみ」に分類される傾向がある。次に「あなたを信じている」という文があるとする。この場合、「信頼」に分類するより「喜び」に分類される傾向がある。

これらの例より、一般的に「嫌悪」と「信頼」の文は隣接する感情に分類される傾向があるため除外した。

2.4.2 色と感情の関係

感情と色は密接な関係にある。各色から連想される感情は色彩心理学によって、関連付けられている。

心理学において、感情とは「広い意味では情動・情熱・気分・情操などを含む概念ですが、狭い意味では強度や身体表出の少ない、快-不快の次元で捉えられる心の作用を示すもの。」とされている[16]。そして、色彩心理学とは、「色彩と人間の関係性を心理学的に解明する学問である。」とされている[17]。

色彩心理学によって、色に対して感情が対応付けされている。例えば、「赤」は怒り・危険・情熱・興奮・攻撃的、「青」は悲しみ・冷静・冷たい・冷淡・不安、「黄」は愉快・光・明朗・幸福、「緑」は成長・新鮮・疑い・嫉妬、「青緑」は感性・創造性・混乱・無力感、「橙」は洞察力・活力・ポジティブ・調和とされている[18][19]。これらのように一つひとつの色に対して人間の心理が関連付けられている。以上を用いて「Emote」では、怒りを「赤」、悲しみを「青」、楽しみを「黄」、恐怖を「緑」、驚きを「青緑色」、期待を「橙」に設定している。

3. 感情処理手法

3.1 感情分類手法

ユーザの投稿情報から感情を抽出する処理にはナイーブベイズ分類器を用いる。ナイーブベイズ分類器とは、ベイズ推定を適応することで実現した確率的分類器であり、未知文書に対して事後確率が最大となるクラスを出力することで分類を行う。

「Emote」ではこのナイーブベイズ分類器を用いて、投稿情報に含まれる感情の推定を行う。ナイーブベイズ分類器では、投稿文章を特徴的な単語のリストに変換した投稿文章ベクトルを

$message = (word_1, \dots, word_n)$, 文章の感情カテゴリを

$emo \in \{angry, fear, joy, sadness, surprise, anticipation\}$

とした場合、事後確率 $P(emo | message)$ の値が最大となる感情カテゴリ emo を投稿文章の感情カテゴリとする処理を行う。実際には、事後確率にベイズの理論を用いて、

$P(message)$ の値は感情カテゴリに左右されずに一定となるので無視すると最終的には式(1)の通り、感情カテゴリの出現率 $P(emo)$ と感情カテゴリごとの文章ベクトルの出現確率 $P(message | emo)$ の積が最大となる感情カテゴリを求めれば良い事となる。

$$\begin{aligned}
 emo &= \arg \max_{emo} P(emo | message) \\
 &= \arg \max_{emo} \frac{P(message | emo)P(emo)}{P(message)} \\
 &= \arg \max_{emo} P(message | emo)P(emo) \quad (式 1)
 \end{aligned}$$

式(1)において,文章ベクトルはさまざまな値を取りうるので,事後確率 $P(message | emo)$ の値を直接算出するのは非常に難しい,そこで感情カテゴリが与えられた時の文章ベクトルの出現確率の間に独立性を仮定すると以下のように計算できる.

$$\begin{aligned}
 P(message | emo) &= P(word_1, \dots, word_n | emo) \\
 &= \prod_{i=1}^n P(word_i | emo)
 \end{aligned}$$

従って,投稿文章の感情分類には次の式の値が最大となる感情カテゴリ emo を計算すればよい.

$$emo = \arg \max_{emo} P(emo) \prod_{i=1}^n P(word_i | emo)$$

例として 100 個の投稿文章の感情分類を行うとする.

100 個の投稿文章のうち,各感情の文章が出現した個数を図 3.1 の様であったとする.感情の出現確率は,出現回数を総投稿総数である 100 で割った値である.

感情	個数	感情の出現確率
怒り(angry)	6	0.06
恐れ(fear)	4	0.04
喜び(joy)	39	0.39
悲しみ(sadness)	20	0.20
驚き(surprise)	9	0.09
期待(anticipation)	22	0.22

図 3.1. 投稿文章における感情の出現回数

次にナイーブベイズ分類器では,単語の出現頻度をもとに,分類をおこなうので,図 3.2 に 100 個の投稿文章中に出現した単語とその出現個数の一例を示す.

感情	いらいら	天気	地震
怒り(angry)	4	0	0
恐れ(fear)	0	0	2
喜び(joy)	0	2	0
悲しみ(sadness)	2	1	1
驚き(surprise)	0	1	0
期待(anticipation)	1	2	0

図 3.2 感情ごとの単語出現回数

この表を元に感情分類を行う.最初に単語の出現回数について調整を行う.この表を元にして忠実に確率を算出してしまうと,例えば“いらいら”が含まれる文章は絶対に“恐れ”の感情カテゴリには分類されないことになってしまう.これを解決するために,まず全ての単語の出現回数に 1 を足し合わせるラプラス

スムージングを適応する.

感情	いらいら	天気	地震
怒り(angry)	5	1	1
恐れ(fear)	1	1	3
喜び(joy)	1	3	1
悲しみ(sadness)	3	2	2
驚き(surprise)	1	2	1
期待(anticipation)	2	3	1

図 3.3 ラプラススムージング適応後の単語出現表

ここでは $message=(“いらいら”, “天気”)$ という文章ベクトルについて各確率を示す.

$$P(“いらいら”) = (5+1+1+3+1+2) / 100 = 0.13$$

$$P(“天気”) = (1+1+3+2+2+3) / 100 = 0.12$$

$$P(“いらいら” | angry) = 5 / 6 = 0.833$$

$$P(“天気” | angry) = 1 / 6 = 0.167$$

$$P(angry | message)$$

$$= (0.833 \times 0.06 / 0.13) \times (0.167 \times 0.04 / 0.12)$$

$$= 0.0214$$

同様に各感情カテゴリの事後確率を算出すると図 3.4 のようになる.

感情	$P(emo message)$
怒り(angry)	0.0214
恐れ(fear)	0.0642
喜び(joy)	0.0194
悲しみ(sadness)	0.0354
驚き(surprise)	0.0385
期待(anticipation)	0.0385

図 3.4 $P(emo | message)$ の算出結果

よって, $message=(“いらいら”, “天気”)$ という文章ベクトルは,もっとも事後確率の値が大きい“恐れ”の感情カテゴリに分類される.以上の手順で,1 つの投稿情報をナイーブベイズ分類器により,感情カテゴリに分類していく.

3.2 感情可視化処理

次に,蓄積したユーザの感情情報を可視化していく処理を行う.感情の可視化はカレンダーで表示することを意識しているので一日を一纏めにして表現する.「Emote」では感情データを通常のカレンダーとともに,感情バルーンとして表示し,感情バルーンは 5 段階の大きさで表現される,その日の投稿文章の中でより大きな割合を占めるものほど大きなバルーンとして表示される.各感情バルーンの表示位置は図 3.5 の通り日付の左上に怒り,左に悲しみ,左下に期待,右上に喜び,右に恐れ,右下に驚きのバルーンが表示される.



図 3.5 感情バルーンの表示位置

例えば 1 日の総投稿数が 30 件あり感情カテゴリごとの出現個数は図 3.6 であったとする。

感情	個数
怒り(angry)	1
恐れ(fear)	5
喜び(joy)	14
悲しみ(sadness)	2
驚き(surprise)	0
期待(anticipation)	8

図 3.6 一日分の投稿文章内での各感情の出現個数

感情バルーンのサイズは、その日の投稿総数を S 、感情カテゴリ emo に分類された投稿数 $N(emo)$ とすると、 $\text{floor}(5 \times N(emo) / S)$ として算出される。図 3.5 の場合、各感情バルーンの大きさはそれぞれ図 3.7 の結果となる。

感情	個数
怒り(angry)	0
恐れ(fear)	1
喜び(joy)	2
悲しみ(sadness)	0
驚き(surprise)	0
期待(anticipation)	1

図 3.7 各感情における感情バルーンの大きさ

また実際のカレンダー上では図 3.8 のように表示される。



図 3.8 感情バルーン

以上の動作を一ヶ月分処理し、感情カレンダーとしてユーザに表示する。以下図 3.9, 図 3.10 に、実際のユーザ A, B のある月の感情カレンダーを表示する。

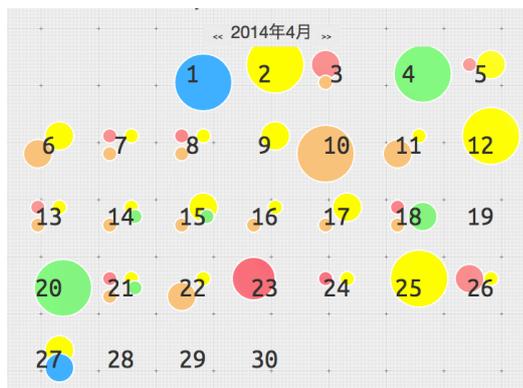


図 3.9 登録ユーザ A のカレンダー

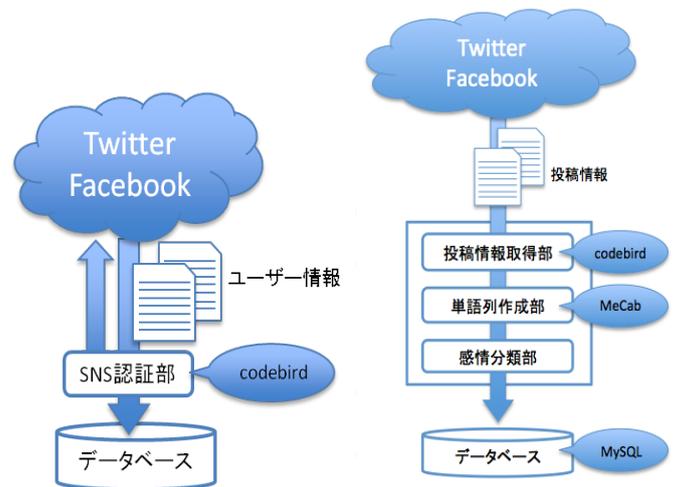


図 3.10 登録ユーザ B のカレンダー

表示されるカレンダーから、そのユーザの一月の感情のデータが一览でき、感情の移ろいがひと目で見る事ができる。

4. システム概要

「Emote」の処理の流れを以下図 4.1, 図 4.2, 図 4.3 に示す。また、図中の吹き出し内には「Emote」を作成する上で使用したライブラリやソフトウェア等を記した。



左図 4.1 SNS 認証部 右図 4.2 投稿情報取得

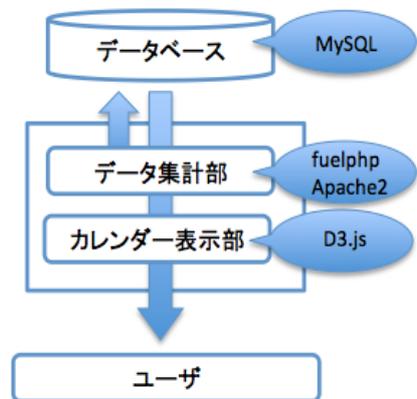


図 4.3 カレンダー表示

図 4.1 はユーザの SNS 認証処理を示したものである。ユーザに Twitter, Facebook へのアクセスを許可してもらうことでユー

ザ情報を取得しデータベースに保存する。

図 4.2 は登録ユーザの Twitter, Facebook の投稿情報を取得する処理群を示したものである。最初に投稿情報取得部でユーザが Twitter, Facebook に投稿した文章を取得する。次に取得された文章に対し単語列作成部で形態素解析を行い、文章から特徴的な単語のリストへ変換する、そして最後に得られた単語リストを感情分類部にて、学習データで訓練したナイーブベイズ分類器を用いて各感情に分類しデータベースに保存する。なお学習データについては 5 章で記述する。

図 4.3 はユーザのカレンダーを表示する処理群を示したものである。データ集計部でそのユーザの指定月の感情情報を集計し、そのデータを元にカレンダー表示部で、カレンダーと感情バルーンを生成する。以上の処理について順に述べる。

4.1 SNS 認証部

SNS 認証部ではユーザのアカウント情報を取得する。Twitter, Facebook の認証を承諾してもらうことでユーザの ID, プロフィール画像, アクセストークンを取得しデータベースに保存する。Twitter のアプリ認証は codebird[20] ライブラリを使用した。

4.2 投稿情報取得部

投稿情報取得部では Twitter, Facebook に投稿された情報をデータベースに保存されているユーザごとに取得する。取得した情報の中からユーザの ID, 投稿された文章, 投稿時間をデータベースに保存する。Twitter の投稿取得は codebird ライブラリを使用した。

4.3 単語列作成部

文章を単語列へと分解する処理にはオープンソースの形態素解析エンジン MeCab[21]を用いる。文章表現の丸め込みの為に、半角カナを全角カナへ、全角英数記号を半角英数記号へ、それぞれ変換後、MeCab を用いて、投稿文章から名詞, 動詞, 形容詞, 形容動詞, 感動詞を抽出し、文章ベクトルする。

4.4 感情分類部

感情分類部では、先述のナイーブベイズ分類器を用いて、単語列作成部から受け取った文章ベクトルを感情分類し、その結果をデータベースに保存する。

4.5 データベース

データベースにはオープンソースのデータベースソフトである MySQL[22]を用いた。

4.6 データ集計部

データ集計部では、データベースに保存されたユーザの投稿情報とその感情データから、一ヶ月分のデータを集計し、それを Web API として提供する機能を有する。

4.7 カレンダー表示部

カレンダー表示部では、データ集計部の Web API から受け取るデータを元に、カレンダーと感情バルーンを表示する処理を行う。Web ブラウザ上で感情バルーンを表示するには、JavaScript のデータビジュアライゼーションライブラリである D3.js[23]を用いた。

5. 学習データ

今回ナイーブベイズ分類器を用いるにあたり、文章とその文章に含まれる感情情報とが対になった学習データが必要となった。その為、学生 15 名に協力して貰い、SNS (Twitter) から収集した投稿情報を見て、その文章から客観的に受け取る感情を記録したデータ、総計約 3 万件を作成した。このデータを「Emote」ではナイーブベイズ分類器の学習データとして用いる。また、その作業を円滑に行う為、学習データ収集ページ図 5.1 を作成した。



図 5.1 感情データ収集ページ

6. 評価

6.1 システムを利用した効果

Twitter, Facebook のアカウントを持っている人の中で日頃、投稿をしている人を対象に「Emote」を使用してその日の出来事を思い出せるか、「カレンダーに正しい感情が反映されていると思うか」の 2 項目を 7 人に回答してもらった。最初の項目に対しては約 57% の人が思い出せたと回答した。思い出せなかった理由はカレンダーの表示が毎日同じような結果になってしまうことが原因で「Emote」は思い出す助けにならなかった。カレンダーに正しい感情が反映されていると思うかの項目に関しては約 67% の人が反映されていると思うと回答を得た。

6.2 システムの使用感

「Emote」の使用感を評価するため、学生 13 名、中京大学の OB1 名に「Emote」を使用してもらい「見て面白い楽しいと思うか」、「色を見て直感的に感情がわかるか」、「操作は簡単か」、「Emote」を今後使いたいかなの 4 項目に 5 段階で回答してもらった。

「Emote」によって作成されたカレンダーを「見て面白い楽しいと思うか」の回答は 8 名が思う、5 名がまあまあ思うという結果で図 6.1 のように全体の約 93% と高い評価を得た。

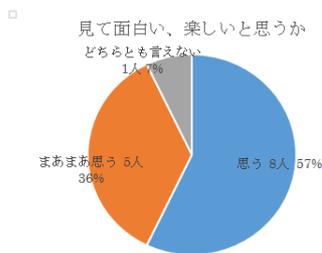


図 6.1 評価”見て面白い、楽しいと思うか”

“色を見て直感的に感情が分かるか”という質問については 3 名がわかる, 5 名がまあまあわかると回答し図 6.2 のように全体の約 57%であった。

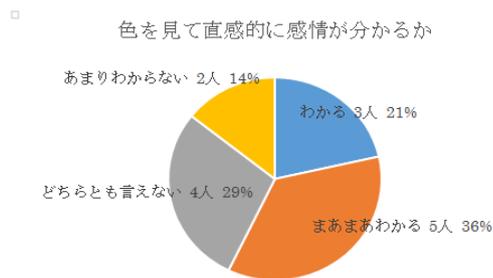


図 6.2 評価”色を見て直感的に感情が分かるか”

“操作は簡単か”と, “「Emote」を今後も使いたか”の項目には Twitter, Facebook のアカウントを持っている 11 名に回答してもらい, “操作の簡単さ”についてはすべての人から簡単であると回答を得た。ライフログとして毎日続けるためには非常に良い結果であると考えられる。“「Emote」を今後も使いたか”については 3 名が思う, 4 名がまあまあ思うという結果になり図 6.3 のように全体の約 64%であった。

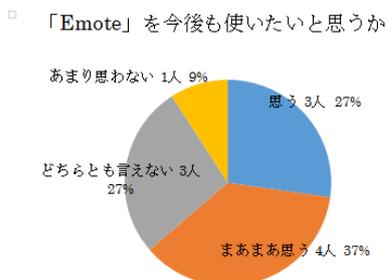


図 6.3 評価”Emote を今後も使いたか”

また使用しての意見に感情の色の決定に関連した投稿の表示を希望するものが多数あり改善を検討している。

6.3 感情分類精度

感情分類の精度を調べる為に, 5 章で記述した学習データの投稿文章をランダムで 8:2 に分け, ナイーブベイズ分類器の訓練に 8 割を使用し, 残りの 2 割のデータの分類結果が正しいか判定するテストを 1000 回試行した。

その結果, 正解の感情カテゴリに分類される確率は 34.8%であった。以上の結果になった理由としては主に 2 点挙げられる。

まず 1 点目に, 人によって同じ文章を見ても, 識別する感情カテゴリは異なる場合があること。それにより学習データの精度が落ちている可能性が挙げられる。

2 点目に, 単語列作成時の問題である。「Emote」では投稿文章の中から, 名詞, 動詞, 形容詞, 形容動詞, 感動詞のみを抽出し, 文章ベクトルとして処理している。しかし, 投稿文章が短文であったり, 抽出できる単語がない場合には, 単語列作成部にて空の文章ベクトルであったり, 学習データとして有効な文章ベクトルが生成できないため, 現状の分類方法では精度が挙げられないという点がある。

7. 今後の展望

今後の展望としてより多くの人々が感情をイメージできる工夫を考えている。また 6.1 で述べたカレンダー表示の問題は投稿数が多いと毎日同じような結果が並ぶ問題があり解決するために特徴的なものを表示することや, 感情の色に加えて一定の感情を想起させることができるものの追加などの改善案を模索している。

8. 参考文献

- 喜連川優, “情報爆発のこれまでとこれから”, <http://www.ieice.org/jpn/books/kaishikiji/2011/201108.pdf>, 電子情報通信学会誌 vol.94, No.8, 2011
- 和田恭, (2010), “米国におけるライフログを巡る最近の動向” <http://www.ipa.go.jp/files/000006063.pdf>
- twilog, <http://twilog.org>
- miin, <http://miin.me>
- RBBTODAY, 日本における SNS 利用者数, <http://www.rbbtoday.com/article/img/2013/05/30/108575/305426.html>, (2014/5/1)にアクセス
- trinity 株式会社, jawbone up, <http://trinity.jp/products/jawbone/up/>
- ヤフーリアルタイム検索「つぶやき感情分析」, <http://search.yahoo.co.jp/realtime>
- Twitter, <http://twitter.com>
- Facebook, <http://facebook.com>
- 加藤慎一郎, 濱川 礼, “Twitter から得られる自然言語情報を用いて行う単語への感情付加手法”, 情報処理学会第 148 回ヒューマンコンピュータインタラクション研究発表会 (HCI) (2012.6)
- Johan Bollen, (2011), “Twitter mood predicts the stock market”, <http://arxiv.org/pdf/1010.3003&>
- feelpress, <http://www.feelpress.com/ja/>
- track&share, <http://www.trackandshareapps.com/>
- Robert Plutchik: Emotion: Theory, Research, and Experience, New York: Academi.
- Robert Plutchik, “The nature of emotions,” American Scientist, Vol. 89, Iss. 4.
- 心理学講座 感情心理学, http://www1.info8s.com/post_11.html, (2014/4/24 アクセス)
- 日本色彩心理学研究所, <http://www.nihon-shikisai-shinrigaku.com/>, (2014/4/26 アクセス)
- アトリエアニメート web 構築 色相心理学, <http://www.animato-jp.net/contents/3/shikisou.html>, (2014/4/26 アクセス)
- 深瀬啓介, カラーリーディング, <http://profile.ameba.jp/exmind/>, (2014/4/26 アクセス)
- codebird <https://github.com/jublonet/codebird-php>
- MeCab: Yet Another Part-of-Speech and Morphological Analyzer <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- MySQL <http://www.jp.mysql.com/>
- Data-Driven Documents <http://d3js.org>