

アノテーション付き画像の感情分析における深層表現を伴った教師付きトピックモデル

茂庭 綾香[†] 坂田 洋介^{††} 江口 浩二^{††}

[†] 神戸大学工学部情報知能工学科 〒657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学大学院システム情報学研究科情報科学専攻 〒657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]moniaya@cs25.scitec.kobe-u.ac.jp, ^{††}sakata@cs25.scitec.kobe-u.ac.jp, ^{††}eguchi@port.kobe-u.ac.jp

あらまし 近年、画像共有サイトなどの普及により、利用可能な画像データが増大の一途をたどっているが、利用者が自由に発信できるそのような画像データの多くは、統一的な基準で分類されていない点や、利用者により定義され付与されたタグを伴う点などの特徴を有する。それらのタグ付き画像データの分析において筆者らは感情に基づく分類に着目する。すなわち、分析対象のタグ付き画像がポジティブな感情を印象付けるものであるか、あるいはネガティブな感情を想起させるものであるかという観点での自動分類を行う。提案手法は、画像データに対して畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) の全結合層の一つに着目することによって得た多項分布からサンプリングを行うことによって離散表現を抽出し、タグを構成するテキストデータとともに用いることで、教師付きマルチモーダルトピックモデルを推定する。推定された潜在トピックは、画像とタグの特徴を反映しつつ、教師信号として与えた感情極性ラベルを反映したものとなり、ラベルが未知であるタグ付き画像データに対してラベル予測を実現する。画像共有サイト Flickr から抽出されたデータセットを用いて有効性の評価を行った結果、約66%の正解率を得た。

キーワード 感情分析, マルチモーダルデータ分析, 教師付きトピックモデル, CNN

1. はじめに

スマートフォンやソーシャルメディアの普及に伴い、多くの人々がテキストによるメッセージとともに画像を公開・共有するようになった。そこには、投稿者が感じた気持ちや伝えたい感情が込められている。大量に存在するそれらの画像を感情に基づいて分析することができれば、世論調査やマーケティング、表出している感情をキーとした画像検索などへの応用が期待できる。

このような画像の感情分類課題に対して、これまで様々な研究が行われてきた。画像の特徴量の表現方法として、色ヒストグラムや SIFT 特徴量を使う手法 [10] や、心理学や芸術理論から設計した特徴量 [3] が提案されている。また、形容詞と名詞から成るフレーズ (Adjective Noun Pairs: ANP) [1] や画像中に存在する顔の表情 [4] を中間表現とする手法も提案されている。畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) をファインチューニングする手法 [9] [11] は特に良い結果を出しており、CNN における物体認識課題から感情分類課題へのドメイン転移学習も有効であることが示されている [11]。

以上の研究では分類対象を画像のみのユニモーダルデータと捉えているが、画像だけでなくそれに付随するテキスト情報を判断材料に加えた感情分類手法も成果を挙げている [8]。このことから、本稿でも分類対象を画像とテキストから成るマルチモーダルデータと捉え、それをポジティブな感情を表しているものとネガティブな感情を表しているものの2種類に分類する。

マルチモーダルデータを扱う際には、異なるモード間の相互依存性をいかに表現するかが課題となる。文献 [8] では、各モードの特徴量をベクトルで表し共通の潜在空間に射影することで対応している。本稿ではこの課題に対し、マルチモーダルデータに対応した教師ありトピックモデルの一種である Symmetric Correspondence MedLDA (SCMLDA) を用いた手法を検証する。画像モードの特徴表現には CNN 特徴量を用いる。ただし、実数値ベクトルのままではトピックモデルで扱うことができないため、CNN の全結合層の一つを多項分布のパラメータとみなし、サンプリングを行うことで離散表現に変換する。トピックモデルにおける未知パラメータや潜在変数の推定には、周辺化ギブスサンプリングを用いる。

2. 関連研究

2.1 マルチモーダルデータとしての画像の感情分類

これまで、感情に基づいて画像を分類する様々な試みがなされてきたが、その多くは画像をユニモーダルデータとして扱い、画像に付随したテキスト情報を活用する議論は十分でなかった。しかし、画像共有サイト等のソーシャルメディア上に投稿された画像には、タグや説明文等のテキスト情報が付随していることが多い。文献 [8] ではこの点に着目し、分類対象を画像そのものとテキストから成るマルチモーダルデータと捉えて、予測精度を向上させることに成功した。本節では文献 [8] の概略を述べる。

文献 [8] では、画像そのもの・画像に付随したテキスト・テキストから抽出した感情語の3つのモードを利用して感情分類が

行われた。異なるモードの特徴を同一空間上で扱うために、各モードの特徴ベクトルが共通の埋め込み空間へ射影される。分類の手順としては、(1) 始めに各モードごとに特徴量を算出し、(2) それらを共通の埋め込み空間に射影した後、(3) 埋め込み空間における新たな特徴表現に対して線形 SVM を適用し、ポジティブな感情・ネガティブな感情のいずれかへ分類される。

画像の特徴量には CNN 特徴量が用いられている。具体的には、ILSVRC 2012 データセットを用いて学習された 8 層 CNN に画像を入力した上で、7 層目の出力が画像の特徴量とされる。感情語特徴には、SentiWordNet という感情語辞書のポジティブスコア・ネガティブスコアが利用されている。SentiWordNet とは、WordNet で定義されている synset (同義語のグループ) に対して、そこに含まれる単語がどれだけポジティブであるか・ネガティブであるかを点数で表した辞書である。

2.2 マルチモーダルトピックモデル

Symmetric Correspondence LDA (SymCorrLDA) [6] はトピックモデルの一種であり、マルチモーダルデータに対してモード間の相互依存性を捉えながら潜在トピックを推定することができる。

トピックモデルでは文書を、そこに含まれる単語の多重集合 (Bag-of-Words) で表現する。例として、「国会で来年度の予算が可決されました」というニュース記事を考えると、この文書は「国会」「来年度」「予算」「可決する」という単語の集合として表される。それぞれの単語の背後にはトピックと呼ばれる潜在変数があり、個々の単語は背後にあるトピックから生成されていると考える。単語の出現確率はトピックごとに異なり、単語はその確率分布に従って生成される。例えば、トピック「政治」における単語「可決する」の出現確率は 3%、トピック「学校」における単語「可決する」の出現確率は 0.2%、などと想像してほしい。トピックの出現確率は、文書ごとに異なる値が与えられる。

画像とテキスト情報といった複数のモードを持つ文書では、トピックの出現確率 $\theta_d^{(\ell)}$ がモードごとに存在する。しかし、同一の文書であるにもかかわらずモード間でトピック分布が大きく異なるという状況は、現実的ではない。同じ文書であるならば、同じ話題 (トピック) について語っているはずである。そこで、同一文書におけるモード間の相互依存性を表現するために、基準となるモード (ピボットモード) を一つ選択し、他のモードはピボットモードに現れたトピックから自分のモードのトピックを選択する。どのモードをピボットモードとするかは単語ごとに異なり、確率分布に従って決定される。

SymCorrLDA ではピボットモードの割り当ておよび単語やトピックの出現確率を多項分布で与え、多項分布の事前分布にはディリクレ分布を導入する。SymCorrLDA における文書集合の生成過程を以下に示す^(注1)。ここで、Dir はディリクレ分布、Mult は多項分布、Unif は一様分布を表す。

(注1)：文献 [6] ではピボットモードフラグが自身のモードと異なる場合、それを明確に区別するために、トピック z ではなく y を定義しているが、本稿では文献 [13] にならってトピックを表す変数を z に統一した。

1. D 個の文書に対し：
 - (a) L 個のモードに対して $\theta_d^{(\ell)} \sim \text{Dir}(\alpha^{(\ell)})$ を選択する。
 - (b) $\pi_d \sim \text{Dir}(\gamma)$ を選択する。
2. K 個のトピック及び L 個のモードに対して $\phi_k^{(\ell)} \sim \text{Dir}(\beta^{(\ell)})$ を選択する。
3. 文書 d のモード ℓ における $N_d^{(\ell)}$ 個の単語 $w_{di}^{(\ell)}$ に対し：
 - (a) ピボットモードフラグ $x_{di}^{(\ell)} \sim \text{Mult}(\pi_d)$ を選択する。
 - (b) $(x_{di}^{(\ell)} = \ell)$ の場合、トピック $z_{di}^{(\ell)} \sim \text{Mult}(\theta_d^{(\ell)})$ を選択する。
 - (c) $(x_{di}^{(\ell)} = m (m \neq \ell))$ の場合、トピック $z_{di}^{(\ell)} \sim \text{Unif}(\mathbf{z}_d^{*(m)})$ を選択する。
 - (d) 単語 $w_{di}^{(\ell)} \sim \text{Mult}(\phi_k^{(\ell)})$ を選択する。 ($k = z_{di}^{(\ell)}$)

ここで、 $\theta_d^{(\ell)} = \{\theta_{dk}^{(\ell)}\}$ であり、 $\theta_{dk}^{(\ell)}$ は文書 d のモード ℓ においてトピック k が出現する確率である。 $\phi_k^{(\ell)} = \{\phi_{kv}^{(\ell)}\}$ であり、 $\phi_{kv}^{(\ell)}$ はモード ℓ においてある単語のトピックが k である時、その単語が語彙 v である確率を表す。 $\alpha^{(\ell)}$ 、 $\beta^{(\ell)}$ はそれぞれ $\theta_d^{(\ell)}$ 、 $\phi_k^{(\ell)}$ に対応するディリクレ分布のハイパーパラメータである。 $x_{di}^{(\ell)}$ は、文書 d のモード ℓ における i 番目の単語のピボットモードを表す潜在変数である。 $\pi_d = \{\pi_{d\ell}\}$ であり、 $\pi_{d\ell}$ は文書 d においてモード ℓ がピボットモードとなる確率である。 γ は π_d に対応するディリクレ分布のハイパーパラメータである。 $z_{di}^{(\ell)}$ は、文書 d のモード ℓ における i 番目の単語のトピックを表す潜在変数である。3.-(c) に表記されている $\mathbf{z}_d^{*(m)}$ は、文書 d のモード m におけるトピック集合 $\{z_{d1}^{(m)}, \dots, z_{dN_d^{(m)}}^{(m)}\}$ の内、ピボットモードが m であるトピックの集合を表している。 $w_{di}^{(\ell)}$ は、文書 d のモード ℓ における i 番目の単語の語彙を表す変数である。

以降の本稿中では、 $\mathbf{W} = \{\mathbf{W}_d\}$ 、 $\mathbf{W}_d = \{w_d^{(\ell)}\}$ 、 $w_d^{(\ell)} = \{w_{di}^{(\ell)}\}$ 、 $\mathbf{Z} = \{\mathbf{Z}_d\}$ 、 $\mathbf{Z}_d = \{z_d^{(\ell)}\}$ 、 $z_d^{(\ell)} = \{z_{di}^{(\ell)}\}$ 、 $\Theta = \{\theta_d\}$ 、 $\theta_d = \{\theta_d^{(\ell)}\}$ 、 $\Phi = \{\phi^{(\ell)}\}$ 、 $\phi^{(\ell)} = \{\phi_k^{(\ell)}\}$ 、 $\Pi = \{\pi_d\}$ とする。また、 $\mathbf{W}_{-d\ell i}$ は \mathbf{W} から $w_{di}^{(\ell)}$ を除いた集合、 $\mathbf{Z}_{-d\ell i}$ は \mathbf{Z} から $z_{di}^{(\ell)}$ を除いた集合、 $\mathbf{X}_{-d\ell i}$ は \mathbf{X} から $x_{di}^{(\ell)}$ を除いた集合とする。

周辺化ギブスサンプリングを用いて SymCorrLDA の潜在変数を推定する際の完全条件付き確率を、以下に示す。SymCorrLDA では、潜在トピック $z_{di}^{(\ell)}$ とピボットモードフラグ $x_{di}^{(\ell)}$ を同時に推定する。

- (i) ピボットモードフラグが自身のモード ℓ である場合

$$p(z_{di}^{(\ell)} = k, x_{di}^{(\ell)} = \ell | w_{di}^{(\ell)} = v^{(\ell)}, \mathbf{W}_{-d\ell i}, \mathbf{Z}_{-d\ell i}, \mathbf{X}_{-d\ell i}, \alpha^{(\ell)}, \beta^{(\ell)}, \gamma)$$

$$\propto \frac{C_{d,-d\ell i}^{(\ell)} + \gamma}{C_{d,-d\ell i}^{(\ell)} + \sum_{\ell' \neq \ell} C_{d,-d\ell i}^{(\ell')} + L\gamma} \cdot \frac{C_{d,-d\ell i}^{k^{(\ell)}} + \alpha^{(\ell)}}{\sum_{k'} C_{d,-d\ell i}^{k'^{(\ell)}} + K\alpha^{(\ell)}} \times \frac{C_{k,-d\ell i}^{v^{(\ell)}} + \beta^{(\ell)}}{\sum_{v' \in \ell} C_{k,-d\ell i}^{v'^{(\ell)}} + V^{(\ell)}\beta^{(\ell)}} \quad (1)$$

- (ii) ピボットモードフラグが自身のモード l ではない ($m \neq l$) 場合

$$p(z_{di}^{(\ell)} = k, x_{di}^{(\ell)} = m | w_{di}^{(\ell)} = v^{(\ell)}, \mathbf{W}_{-dli}, \mathbf{Z}_{-dli}, \mathbf{X}_{-dli}, \alpha^{(\ell)}, \beta^{(\ell)}, \gamma)$$

$$\propto \frac{C_{d,-dli}^{(m)} + \gamma}{C_{d,-dli}^{(m)} + \sum_{m' \neq m} C_{d,-dli}^{(m')} + L\gamma} \cdot \frac{C_{d,-dli}^{k(m)}}{\sum_{k'} C_{d,-dli}^{k'(m)}} \times \frac{C_{k,-dli}^{v^{(\ell)}} + \beta^{(\ell)}}{\sum_{v' \neq v} C_{k,-dli}^{v'(\ell)} + V^{(\ell)}\beta^{(\ell)}} \quad (2)$$

ここで、 $C_d^{k(m)}$ は文書 d のモード m においてトピック k が割り当てられた回数、 $C_k^{v(m)}$ はモード m においてトピック k が語彙 v に割り当てられた回数、 $C_d^{(m)}$ は文書 d の全モードにおいてピボット m が割り当てられた回数とすると、 $C_{d,-dli}^{k(m)}$ 、 $C_{k,-dli}^{v(m)}$ 、 $C_{d,-dli}^{(m)}$ はそれぞれ文書 d のモード l 中の i 番目の単語の情報を除いた $C_d^{k(m)}$ 、 $C_k^{v(m)}$ 、 $C_d^{(m)}$ である。 K はトピック数、 L はモード数、 $V^{(\ell)}$ はモード l における語彙数である。

2.3 教師ありマルチモーダルトピックモデル

マージン最大化マルチモーダルトピックモデル (Symmetric Correspondence MedLDA: SCMLDA) [13] は、マルチモーダルデータに付与されているラベルを予測することができるトピックモデルである。SCMLDA では、SymCorrLDA と同様にモード間の相互依存性を考慮しながら単語の背後にある潜在トピックを推定した上で、その潜在トピックに基づいてラベルを予測する。

SCMLDA における文書集合の生成過程を以下に示す。記号の定義は 2.2 節を参照されたい。

1. D 個の文書に対し：
 - (a) L 個のモードに対して $\theta_d^{(\ell)} \sim \text{Dir}(\alpha^{(\ell)})$ を選択する。
 - (b) $\pi_d \sim \text{Dir}(\gamma)$ を選択する。
2. K 個のトピック及び L 個のモードに対して $\phi_k^{(\ell)} \sim \text{Dir}(\beta^{(\ell)})$ を選択する。
3. 文書 d のモード l における $N_d^{(\ell)}$ 個の単語 $w_{di}^{(\ell)}$ に対し：
 - (a) ピボットモードフラグ $x_{di}^{(\ell)} \sim \text{Mult}(\pi_d)$ を選択する。
 - (b) ($x_{di}^{(\ell)} = \ell$) の場合、トピック $z_{di}^{(\ell)} \sim \text{Mult}(\theta_d^{(\ell)})$ を選択する。
 - (c) ($x_{di}^{(\ell)} = m (m \neq \ell)$) の場合、トピック $z_{di}^{(\ell)} \sim \text{Unif}(\mathbf{z}_d^{*(m)})$ を選択する。
 - (d) 単語 $w_{di}^{(\ell)} \sim \text{Mult}(\phi_k^{(\ell)})$ を選択する。 ($k = z_{di}^{(\ell)}$)
4. D 個の文書に対してラベル r_d を以下の式に従って選択する。

$$r_d \sim F \left(z_{d1}^{(1)}, \dots, z_{dN_d^{(1)}}^{(1)}, z_{d1}^{(2)}, \dots, z_{dN_d^{(L)}}^{(L)}, \boldsymbol{\eta} \right) = \arg \max_{r \in \{1, \dots, R\}} \left(\boldsymbol{\eta}^{rT} \left(\sum_{\ell=1}^L \bar{\mathbf{z}}_d^{(\ell)} \right) \right) \quad (3)$$

SymCorrLDA と同様の生成手順をたどった後、トピック集合 $\{z_{d1}^{(1)}, \dots, z_{dN_d^{(1)}}^{(1)}, z_{d1}^{(2)}, \dots, z_{dN_d^{(L)}}^{(L)}\}$ に基づいて文書 d に付与されるラベル r_d を決定する。 $\boldsymbol{\eta} = \{\boldsymbol{\eta}^r\}$ 、 $\boldsymbol{\eta}^r = \{\eta_k^r\}$ であり、 $\boldsymbol{\eta}^{rT}$ は $\boldsymbol{\eta}^r$ の転置行列を表す。 η_k^r はラベル r の評価値を計算する際のトピック k に対する重み係数であり、値が大きいほどトピック k がラベル r に強く関わっていることを表している。最終的に、評価値の最も高いラベルをその文書のラベルとする。(一つの文書に複数のラベルが付く場合は、一文書当たり a 個のラベルが付くと仮定した上で上位 a 個を選出する。) 4. において、 $\bar{\mathbf{z}}_d^{(\ell)} = \frac{1}{N_d^{(\ell)}} \sum_i \mathbf{v}_{di}^{(\ell)}$ であり、 $\mathbf{v}_{di}^{(\ell)}$ は $z_{di}^{(\ell)}$ が示すトピックに対応する成分が 1、その他の成分が 0 となる K 次元のベクトルである。

2.4 教師ありマルチモーダルトピックモデルにおける周辺化ギブスサンプリング

以下では、SCMLDA において周辺化ギブスサンプリングを行う際の完全条件付き確率について述べる。

まず、全ての未知パラメータと潜在変数の同時確率分布を以下に示す。

$$p(\boldsymbol{\eta}, \boldsymbol{\Theta}, \mathbf{Z}, \boldsymbol{\Phi}, \boldsymbol{\Pi}, \mathbf{X} | \mathbf{W}, \mathbf{r}) \propto p_0(\boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\Pi}) p(\mathbf{Z} | \boldsymbol{\Theta}, \mathbf{X}) p(\mathbf{X} | \boldsymbol{\Pi}) p(\mathbf{W} | \mathbf{Z}, \boldsymbol{\Phi}) \psi(\mathbf{r} | \mathbf{Z}, \boldsymbol{\eta}) \quad (4)$$

$p_0(\boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\Pi})$ は予め定義された事前分布であり、 $p(\mathbf{W} | \mathbf{Z}, \boldsymbol{\Phi})$ は尤度である。 $\psi(\mathbf{r} | \mathbf{Z}, \boldsymbol{\eta})$ はラベルの擬尤度であり、 $\psi(\mathbf{r} | \mathbf{Z}, \boldsymbol{\eta}) = \sum_{d=1}^D \sum_{j=1}^R \psi_j(r_d^j | \mathbf{z}_d^{(1)}, \dots, \mathbf{z}_d^{(L)}, \boldsymbol{\eta})$ である。 r_d^j は、文書 d にラベル j が付与されている場合は 1、そうでない場合は -1 をとる 2 値変数である。

ここで、 ψ_j を以下のように定義する。 [5]

$$\psi_j(r_d^j | \mathbf{z}_d^{(1)}, \dots, \mathbf{z}_d^{(L)}, \boldsymbol{\eta}) := \exp(-2c \max(0, \zeta_d^j)) \quad (5)$$

ここで、 $c > 0$ は正則化パラメータであり、

$$\zeta_d^j = l - r_d^j \sum_{\ell=1}^L \boldsymbol{\eta}^{jT} \bar{\mathbf{z}}_d^{(\ell)} \quad (6)$$

である。 l は損失パラメータである。 $\sum_{\ell=1}^L \boldsymbol{\eta}^{jT} \bar{\mathbf{z}}_d^{(\ell)}$ は式 (3) にも見られるように、ラベル j の評価値を表しており、この値が大きいほどラベル j が付与されやすいと判定される。しかしそれに反してラベル j が付与されなかった場合、 $r_d^j = -1$ となるため、評価値 $\sum_{\ell=1}^L \boldsymbol{\eta}^{jT} \bar{\mathbf{z}}_d^{(\ell)}$ が大きければ大きいほど ζ_d^j の値が大きくなる。すなわち、 ζ_d^j の値は誤分類が甚だしいほど大きな値をとるように定義されている。

$c > 0$ であるので、誤分類が多いすなわち ζ_d^j の値が大きいほど、ラベルの擬尤度 $\psi_j(r_d^j | \mathbf{z}_d^{(1)}, \dots, \mathbf{z}_d^{(L)}, \boldsymbol{\eta})$ は小さくなる。また、 ψ_j はこのように定義することで以下の式 (7) のように変形することができる。 [5]

$$\psi_j(r_d^j | \mathbf{z}_d^{(1)}, \dots, \mathbf{z}_d^{(L)}, \boldsymbol{\eta}) = \int_0^\infty \frac{1}{\sqrt{2\pi\lambda_d^j}} \exp\left(-\frac{(\lambda_d^j + c\zeta_d^j)^2}{2\lambda_d^j}\right) d\lambda_d^j \quad (7)$$

λ は新たに拡張されたデータ拡張変数であり、 λ_d^j が r_d^j に対応している。 $\boldsymbol{\lambda} = \{\lambda_d\}$, $\boldsymbol{\lambda}_d = \{\lambda_d^j\}$ とすると、 λ を含んだ同時確率分布は以下ようになる。

$$\begin{aligned} p(\boldsymbol{\eta}, \boldsymbol{\Theta}, \mathbf{Z}, \boldsymbol{\Phi}, \boldsymbol{\Pi}, \mathbf{X}, \boldsymbol{\lambda} | \mathbf{W}, \mathbf{r}) \\ = p_0(\boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\Pi}) p(\mathbf{Z} | \boldsymbol{\Theta}, \mathbf{X}) p(\mathbf{X} | \boldsymbol{\Pi}) p(\mathbf{W} | \mathbf{Z}, \boldsymbol{\Phi}) \psi(\mathbf{r}, \boldsymbol{\lambda} | \mathbf{Z}, \boldsymbol{\eta}) \end{aligned} \quad (8)$$

ここで、

$$\psi(\mathbf{r}, \boldsymbol{\lambda} | \mathbf{Z}, \boldsymbol{\eta}) = \prod_{j=1}^R \prod_{d=1}^D \frac{1}{\sqrt{2\pi\lambda_d^j}} \exp\left(-\frac{(\lambda_d^j + c\zeta_d^j)^2}{2\lambda_d^j}\right) \quad (9)$$

である。

ここから $(\boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\Pi})$ を周辺化することで、以下の式が得られる。ここで、 $p_0(\boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\Pi}) = p_0(\boldsymbol{\eta}) p_0(\boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\Pi})$ と仮定されている。

$$p(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{Z}, \mathbf{X} | \mathbf{W}, \mathbf{r}) \propto p_0(\boldsymbol{\eta}) p(\mathbf{W}, \mathbf{Z}, \mathbf{X} | \alpha, \beta, \gamma) \psi(\mathbf{r}, \boldsymbol{\lambda} | \mathbf{Z}, \boldsymbol{\eta}) \quad (10)$$

さらに周辺化することで、周辺化ギブスサンプリングで用いる各パラメータの完全条件付き確率を求めることができる。

2.4.1 重みベクトル $\boldsymbol{\eta}$ の完全条件付き確率

$p(\boldsymbol{\eta} | \boldsymbol{\lambda}, \mathbf{Z}, \mathbf{X})$ は、

$$p(\boldsymbol{\eta} | \boldsymbol{\lambda}, \mathbf{Z}, \mathbf{X}) = \prod_{j=1}^R p(\boldsymbol{\eta}^j | \boldsymbol{\lambda}, \mathbf{Z}, \mathbf{X}) \quad (11)$$

と変形できる。 $\boldsymbol{\eta}^j$ のガウス事前分布を

$$p_0(\boldsymbol{\eta}^j) = \sum_{k=1}^K \mathcal{N}(\boldsymbol{\eta}_k^j; 0, \nu^2) \quad (12)$$

と仮定すると、式 (10) から $(\boldsymbol{\lambda}, \mathbf{Z}, \mathbf{X})$ を周辺化することで以下の式が得られる。

$$\begin{aligned} p(\boldsymbol{\eta}^j | \boldsymbol{\lambda}, \mathbf{Z}) &\propto p_0(\boldsymbol{\eta}^j) \prod_{d=1}^D \exp\left(-\frac{(\lambda_d^j + c\zeta_d^j)^2}{2\lambda_d^j}\right) \\ &\propto \exp\left(-\sum_{k=1}^K \frac{\eta_k^{j2}}{2\nu^2} - \sum_{d=1}^D \frac{(\lambda_d^j + c\zeta_d^j)^2}{2\lambda_d^j}\right) \\ &= \mathcal{N}(\boldsymbol{\eta}^j; \boldsymbol{\mu}^j, \boldsymbol{\Sigma}^j) \end{aligned} \quad (13)$$

ここで、

$$\boldsymbol{\mu}^j = \boldsymbol{\Sigma}^j \left(\sum_{d=1}^D \left(\frac{c\boldsymbol{\eta}_d^j}{\lambda_d^j} (\lambda_d^j + cl) \left(\sum_{\ell=1}^L \bar{z}_{d\ell}^{(\ell)} \right) \right) \right) \quad (14)$$

であり、 $\boldsymbol{\Sigma}^j$ の逆行列の要素 $(\boldsymbol{\Sigma}^{j-1})_{ab}$ は

$$(\boldsymbol{\Sigma}^{j-1})_{ab} = \frac{\Delta(a=b)}{\nu^2} + c^2 \sum_{d=1}^D \left(\frac{(\sum_{\ell=1}^L \bar{z}_{da}^{(\ell)}) \cdot (\sum_{\ell=1}^L \bar{z}_{db}^{(\ell)})}{\lambda_d^j} \right) \quad (15)$$

となる。 $\Delta(\cdot)$ は、 (\cdot) 内の式が真ならば 1、そうでなければ 0 を返す関数である。 $\bar{z}_{dk}^{(\ell)}$ は、文書 d のモード ℓ の中でトピック k が割り当たっている割合を表している。以上より、 $p(\boldsymbol{\eta}^j | \boldsymbol{\lambda}, \mathbf{Z})$ は平均 $\boldsymbol{\mu}^j$ 、分散共分散行列 $\boldsymbol{\Sigma}^j$ の多変量正規分布となり、この多変量正規分布から $\boldsymbol{\eta}^j$ の各要素 η_k^j をサンプリングすること

2.4.2 トピック \mathbf{Z} 及びピボット \mathbf{X} の完全条件付き確率

SCMLDA では、トピック $z_{di}^{(\ell)}$ とピボット $x_{di}^{(\ell)}$ を同時に推定する。

ピボットモードフラグ $x_{di}^{(\ell)}$ が自身のモード ℓ である場合のトピック割り当て $z_{di}^{(\ell)}$ の完全条件付き確率は、式 (10) から $(\boldsymbol{\eta}, \boldsymbol{\lambda})$ を周辺化することで

$$\begin{aligned} p(z_{di}^{(\ell)} = k, x_{di}^{(\ell)} = \ell | w_{di}^{(\ell)} = v^{(\ell)}, \mathbf{W}_{-d\ell i}, \mathbf{Z}_{-d\ell i}, \mathbf{X}_{-d\ell i}, \\ \alpha^{(\ell)}, \beta^{(\ell)}, \gamma, \boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{r}) \\ \propto \frac{C_{d,-d\ell i}^{(\ell)} + \gamma}{C_{d,-d\ell i}^{(\ell)} + \sum_{\ell' \neq \ell} C_{d,-d\ell i}^{(\ell')} + L\gamma} \cdot \frac{C_{d,-d\ell i}^{k(\ell)} + \alpha^{(\ell)}}{\sum_{k'} C_{d,-d\ell i}^{k'(\ell)} + K\alpha^{(\ell)}} \\ \times \frac{C_{k,-d\ell i}^{v(\ell)} + \beta^{(\ell)}}{\sum_{v'(\ell)} C_{k,-d\ell i}^{v'(\ell)} + V^{(\ell)}\beta^{(\ell)}} \\ \times \prod_{j=1}^R \exp\left(\frac{c\eta_k^j}{N_d^{(\ell)}\lambda_d^j} \left(r_d^j (cl + \lambda_d^j) \right. \right. \\ \left. \left. - \frac{c}{2N_d^{(\ell)}} \left(\eta_k^j + 2 \sum_{k'=1}^K \eta_{k'}^j C_{d,-d\ell i}^{k'(\ell)} \right) \right) \right) \end{aligned} \quad (16)$$

となる。また、ピボットモードフラグ $x_{di}^{(\ell)}$ が自身とは別のモード m である場合 ($m \neq \ell$) の、トピック割り当て $z_{di}^{(\ell)}$ の完全条件付き確率は、

$$\begin{aligned} p(z_{di}^{(\ell)} = k, x_{di}^{(\ell)} = m | w_{di}^{(\ell)} = v^{(\ell)}, \mathbf{W}_{-d\ell i}, \mathbf{Z}_{-d\ell i}, \mathbf{X}_{-d\ell i}, \\ \alpha^{(\ell)}, \beta^{(\ell)}, \gamma, \boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{r}) \\ \propto \frac{C_{d,-d\ell i}^{(m)} + \gamma}{C_{d,-d\ell i}^{(m)} + \sum_{m' \neq m} C_{d,-d\ell i}^{(m')} + L\gamma} \cdot \frac{C_{d,-d\ell i}^{k(m)}}{\sum_{k'} C_{d,-d\ell i}^{k'(m)}} \\ \times \frac{C_{k,-d\ell i}^{v(\ell)} + \beta^{(\ell)}}{\sum_{v'(\ell)} C_{k,-d\ell i}^{v'(\ell)} + V^{(\ell)}\beta^{(\ell)}} \\ \times \prod_{j=1}^R \exp\left(\frac{c\eta_k^j}{N_d^{(\ell)}\lambda_d^j} \left(r_d^j (cl + \lambda_d^j) \right. \right. \\ \left. \left. - \frac{c}{2N_d^{(\ell)}} \left(\eta_k^j + 2 \sum_{k'=1}^K \eta_{k'}^j C_{d,-d\ell i}^{k'(\ell)} \right) \right) \right) \end{aligned} \quad (17)$$

となる。

2.4.3 データ拡張変数 $\boldsymbol{\lambda}$ の完全条件付き確率

$\boldsymbol{\lambda}$ は、 r_d^j に対応する λ_d^j から成る RD 次元のベクトルであり、 $p(\boldsymbol{\lambda} | \boldsymbol{\eta}, \mathbf{Z}) = \prod_{j=1}^R \prod_{d=1}^D p(\lambda_d^j | \boldsymbol{\eta}, \mathbf{Z})$ と変形できる。式 (10)

から $(\boldsymbol{\eta}, \mathbf{Z}, \mathbf{X})$ を周辺化することにより, $\boldsymbol{\lambda}$ の完全条件付き確率は下記の式のような一般化逆ガウス (Generalized inverse Gaussian: GIG) 分布に従うことが分かる.

$$\begin{aligned} p(\lambda_d^j | \boldsymbol{\eta}, \mathbf{Z}) &\propto \frac{1}{\sqrt{2\pi\lambda_d^j}} \exp\left(-\frac{(\lambda_d^j + c\zeta_d^j)^2}{2\lambda_d^j}\right) \\ &= \text{GIG}\left(\lambda_d^j; \frac{1}{2}, 1, c^2(\zeta_d^j)^2\right) \end{aligned} \quad (18)$$

一般化逆ガウス分布 $\text{GIG}(x; p, a, b)$ の確率密度関数は以下の通りである. $e(p, a, b)$ は正規化定数である.

$$\text{GIG}(x; p, a, b) = e(p, a, b)x^{p-1} \exp\left(-\frac{1}{2}\left(\frac{b}{x} + ax\right)\right) \quad (19)$$

周辺化ギブスサンプリングでは, 2.4.1 節 ~ 2.4.3 節で示した完全条件付き確率を反復させることで各パラメータを推定する.

3. 提案手法

本稿では, テキストアノテーション付き画像の感情分類課題に対して, SCMLDA を用いたフレームワークを提案する. 画像・テキストの 2 つのモードを Bag-of-Words の形に表現した上で SCMLDA に入力し, ポジティブな感情・ネガティブな感情のいずれかへ分類する. 提案手法のイメージ図を Fig.1 に示す.

3.1 画像特徴量の Bag-of-Words 化

本稿では, 画像の特徴量として CNN 特徴量を用いる. 具体的には, ILSVRC 2012 データセットを用いて学習された 8 層 CNN [2] に画像を入力した上で, 全結合層の 2 層目 (以下, 「fc7 層」) の出力を画像の特徴量とする. この出力は 4096 次元の実数値ベクトルである. SCMLDA へ入力する際には単語の多重集合 (Bag-of-Words) の形をとる必要があるため, fc7 層の実数値ベクトルを Bag-of-Words 表現に変換する.

文書 d の画像を入力した際の fc7 層の出力ベクトルを $\mathbf{y}_d \in \mathbb{R}_+^{4096}$ とし, \mathbf{y}_d の全要素の合計が 1 になるよう正規化したものを $\hat{\mathbf{y}}_d$ とする. $\hat{\mathbf{y}}_d$ の i 番目要素 \hat{y}_{di} は以下の式によって導かれる.

$$\hat{y}_{di} = \frac{y_{di}}{\sum_{j=1}^{4096} y_{dj}} \quad (20)$$

$\hat{\mathbf{y}}_d$ を多項分布のパラメータとみなし N 回サンプリングすることで, 4096 個の語彙の中から選ばれた N 個の単語集合を得る. すなわち, 語彙 i が出現する確率が \hat{y}_{di} であるとして N 回サンプリングを行う. このようにして得られた単語集合を文書 d の画像モードの Bag-of-Words とする. fc7 層の出力から Bag-of-Words を生成する手順を以下に示す.

1. D 個の文書に対し:

- (a) 画像を 8 層 CNN に入力して, fc7 層の出力ベクトル \mathbf{y}_d を得る.

- (b) \mathbf{y}_d を式 (20) により正規化したベクトル $\hat{\mathbf{y}}_d$ を得る.

2. 文書 d の N 個の単語 w_{di} に対して単語 $w_{di} \sim \text{Mult}(\hat{\mathbf{y}}_d)$ を選択する.

3.2 テキスト情報の Bag-of-Words 化

画像に付随したテキストからは, stopwords と punctuation を除去した上で, スペースを単語の区切りと見なして分割し Bag-of-Words とする. アルファベットの大文字と小文字は区別しない.

3.3 モデルの学習とテスト画像のラベル予測

3.1 節, 3.2 節に記した手順で Bag-of-Words 化した画像およびテキストを, SCMLDA への入力とする.

初めにモデル学習として, 教師ラベル付きデータ (学習用データ) を用いてモデルの未知パラメータを推定する. (この操作を以下「学習」と呼ぶ.) 具体的には, 2.4.1 節 ~ 2.4.3 節の式を用いて周辺化ギブスサンプリングを行う. 次に, 学習により得られたトピック-単語分布 Φ を用いて, ラベル予測をしたい対象 (テストデータ) の潜在トピック \mathbf{Z} を推定する. (この操作を以下「リサンプリング」と呼ぶ.) テストデータはラベル r_d が未知の状態潜在トピックを推定するため, SCMLDA の完全条件付き確率 (式 (16), (17)) を利用することはできない. そのため, リサンプリングでは SymCorrLDA の完全条件付き確率 (式 (1) および式 (2)) を用いて潜在トピックを推定する.

最後に, 学習時に得られたラベル評価用の重み係数 $\boldsymbol{\eta}$ とリサンプリング時に推定したトピック分布 \mathbf{Z} を用い, 式 (3) に従って各文書に付与されるラベルを予測する.

4. 実験

4.1 データセット

実験に用いるデータセットとして, 文献 [7] [8] において教師ラベルを付与された Flickr データセット [12] を用いる. このデータセットは, 画像共有サイト Flickr において収集されたテキストアノテーション付き画像から成っている. 文献 [7] [8] では 3 名の被験者により意見極性ラベルを付与しており,

- (i) 3 名とも意見極性がポジティブまたはネガティブで一致する場合
- (ii) 2 名がポジティブまたはネガティブで一致して残り 1 名はニュートラルである場合

について, 感情分類ラベルとして採用している. 本稿でもこれと同じ条件の感情分類ラベルを用いる. データセットに含まれる感情分類ごとの画像の枚数を Table 1 に示す. 実験では, ポジティブな画像・ネガティブな画像ともに同じ枚数を使用する.

Table 1 The number of positive and negative images in dataset.

Label	3-agree	2-agree	Total
Positive	21549	20001	41550
Negative	3988	4914	8902

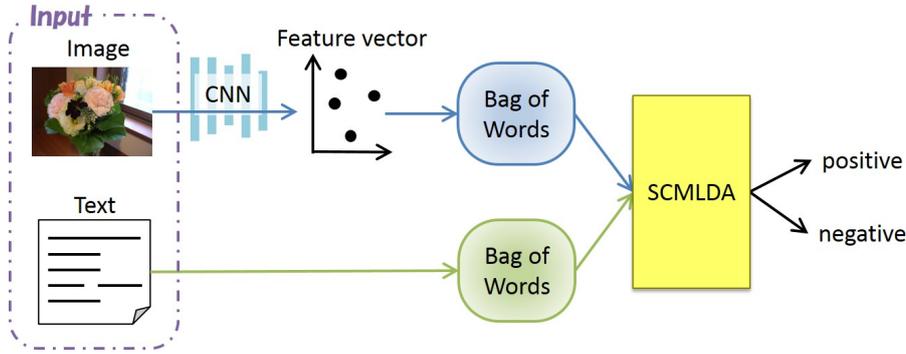


Fig. 1 An outline of proposed method.

画像に付随するテキスト情報には、タイトルやタグ、説明文、他のユーザからのコメントなどの項目がある。本稿では文献 [8] と実験条件を合わせるために、タグと説明文のみを使用し、それらを一括して一つのモードとして扱う。

実験に際して、Table 1 の画像からランダムに画像を選択することで、学習用セットとテスト用セットの組を5組作成した。具体的には、ポジティブな画像・ネガティブな画像それぞれから 6000 枚ずつ画像をランダムに選出し、これを学習用セットとした。同様に、ポジティブな画像・ネガティブな画像それぞれから 2500 枚ずつランダムに選出した画像を、テスト用セットとした。これを5回繰り返すことで、学習用セット・テスト用セットの組を5通り作成した。

4.2 実験条件と評価方法

公開されている CaffeNet ライブラリ^(注2) [2] を用いて各画像から CNN 特徴量を抽出し、一画像当たりの特徴数を $N = 1000$ として Bag-of-Words に変換した。

その上で、学習用セットに対して SCMLDA の周辺化ギブスサンプリングを行い未知パラメータを推定した。これにより得られたトピック-単語分布 Φ を用いてリサンプリングを行い、テスト用セットの潜在トピック \mathbf{Z} を推定した。最後に、学習時に得られたラベル評価用の重み係数 η とリサンプリング時に推定したトピック分布 \mathbf{Z} を用い、式 (3) に従ってテスト用セットの各文書に付与されるラベルを予測した。これを学習用セット・テスト用セットの組それぞれに対して行い、予測したラベルの正解率を各組ごとに算出する。

$$(\text{正解率}) = \frac{(\text{テスト画像の内、正しくラベル予測できた枚数})}{(\text{テスト画像の総数})} \quad (21)$$

なお、本稿で使用するデータセットの中には、テキスト情報が全く無い（すなわち、テキストモードの単語数が 0 個である）画像がいくつか存在する。モード m の単語数が 0 である場合、 $\bar{z}_{dk}^{(m)} = \frac{0}{0}$ となり、式 (14) および式 (15)、また式 (6) を通じて式 (18) に影響する。そのため、モード m の単語数が 0 の場合は $\bar{z}_{dk}^{(m)} = 0$ とする例外処理を設けた^(注3)。

学習およびサンプリング時の各ハイパーパラメータの値は、全てのモード ℓ において $\alpha^{(\ell)} = 0.1$, $\beta^{(\ell)} = 0.01$, $\gamma = 1$ とした。損失パラメータ $l = 1$, 正則化パラメータ $c = 1$ とした。 η^j のガウス事前分布（式 (12)）における標準偏差 ν は、 $\nu = 1$ とした。周辺化ギブスサンプリングの終了条件は、反復回数 200 回とした。周辺化ギブスサンプリングの初期設定は、トピックおよびピボットの割り当てについては全てランダムに選択し、 λ の要素は全て 1 とした。SCMLDA の周辺化ギブスサンプリングでは、 $\eta, \mathbf{Z}, \lambda$ の順にパラメータを推定した。トピック数は、いずれのデータセットに対しても {10,20,40,60,80,100} の 6 通りの実験を行った。

比較対象として、各テスト用セットに対してランダムに分類する実験も行った。

4.3 実験結果

5通り作成した学習用セット・テスト用セットの組それぞれに対して正解率を算出し、それらの平均と標本標準偏差を Table 2 に示す。

提案手法の正解率がランダムに分類した場合を上回っていることから、与えられたテキストや画像の情報が分類に役立てられていることが分かる。また、提案手法におけるトピック数は、トピック数 10 ~ 60 程度まではトピック数が多い方が結果が良くなり、その後は正解率が収束している。

Table 2 Averaged accuracy over five datasets and its sample standard deviation.

Method	Average	Standard deviation
Random	0.5008	0.0053
Proposed method (with 10 topics)	0.5039	0.0058
Proposed method (with 20 topics)	0.5408	0.0583
Proposed method (with 40 topics)	0.6064	0.0420
Proposed method (with 60 topics)	0.6420	0.0047
Proposed method (with 80 topics)	0.6599	0.0265
Proposed method (with 100 topics)	0.6557	0.0410

の中でトピック k が割り当たっている割合を表しており、この値を参考にラベルの評価値 $\sum_{\ell=1}^L \eta^{jT} \bar{z}_d^{(\ell)}$ が算出される。単語数が 0 のモードの情報はラベル評価に加味しないことが妥当であるので、モード m の単語数が 0 の場合は $\bar{z}_{dk}^{(m)} = 0$ とする。

(注2) : <http://caffe.berkeleyvision.org/>

(注3) : すべての $\bar{z}_{dk}^{(m)}$ は式 (6) に由来している。 $\bar{z}_{dk}^{(m)}$ は文書 d のモード m

5. おわりに

本稿では、マルチモーダルデータとしての画像の感情分類課題に対して、教師付きトピックモデルの一種である SCMLDA を用いた手法を提案した。画像の特徴表現には CNN 特徴量を採用し、これを多項分布のパラメータとみなしてサンプリングを行うことで Bag-of-Words の形式に変換した。提案手法の性能を評価するために、画像共有サイト Flickr から収集されたテキストアノテーション付き画像データセットを用いて実験を行った。提案手法が一定の効果を持つことが示され、トピック数が多いほど分類性能が向上することが分かった。

今後はさらにトピック数を増やした実験や、画像の CNN 特徴量を離散化する際の特徴数 N や正則化パラメータ c を変動させた実験を行うことで、提案手法が持つ性能の上限を探る予定である。なお、本稿では文献 [13] の方針にならってトピックモデルの式を導出した。ここでは、式 (16) と式 (17) の導出の過程で用いられる変数 $\bar{z}_{d,-d\ell i}^{(\ell)}$ を $\bar{z}_{d,-d\ell i}^{(\ell)} = \frac{1}{N_d^{(\ell)}} \sum_i \mathbf{v}_{di,-d\ell i}^{(\ell)}$ と解釈していると見られる。ここで $\bar{z}_{d,-d\ell i}^{(\ell)}$ は、文書 d のモード ℓ における i 番目の単語を除外した $\bar{z}_d^{(\ell)}$ である。 $\bar{z}_{d,-d\ell i}^{(\ell)}$ の意味から、 $\bar{z}_{d,-d\ell i}^{(\ell)} = \frac{1}{N_d^{(\ell)} - 1} \sum_i \mathbf{v}_{di,-d\ell i}^{(\ell)}$ という解釈も考えられるため、どちらの解釈がより良い性能をもたらすかについても調査する予定である。

また、現時点では SCMLDA のみを学習対象とし、画像の特徴量を抽出する CNN に対してファインチューニングは行っていない。今後は、CNN のファインチューニングも含めた学習方法を検討することで、さらなる分類性能の向上を図りたい。

なお、SCMLDA は多クラス分類を前提に設計されたモデルであるため、ポジティブ・ネガティブの 2 分類にとどまらず多種類の感情への分類能力も期待できる。

謝 辞

本研究に使用したデータを提供して頂いた同志社大学理工学部 桂井麻里衣助教に感謝する。本研究の一部は科学研究費補助金 基盤研究 (B) (15H02703) の援助による。

文 献

- [1] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs." *ACM MM* (2013)
- [2] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, "Caffe: Convolutional Architecture for Fast Feature Embedding" *arXiv preprint*, arXiv:1408.5093 (2014)
- [3] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory." *ACM MM* (2010)
- [4] J. Yuan, S. Mcdonough, Q. You, and J. Luo, "SentrIBUTE: Image sentiment analysis from a mid-level perspective." *Proc. Int. Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM)*, pp. 10:1-10:8 (2013)
- [5] J.Zhu, N.Chen, H.Perkins and B.Zhang, "Gibbs Max-margin Topic Models with Data Augmentation" *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1073-1110 (2014).
- [6] Kosuke Fukumasu, Koji Eguchi, and Eric P.Xing, "Symmetric Correspondence Topic Models for Multilingual Text Analysis." *Advances in Neural Information Processing Systems*, pp. 1286-1294 (2012).
- [7] M. Katsuragi and S. Satoh, "Image Sentiment Analysis Using Latent Correlations Among Visual, Textual, and Sentiment Views" *Proc. ICASSP* (2016)
- [8] 桂井麻里衣, 佐藤真一, "画像・テキスト・感情語の潜在的な相関に基づく画像の感情分類" *DEIM Forum 2016* (2016)
- [9] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks", *The Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, pp. 381-388 (2015)
- [10] S. Siersdorfer, E. Minack, F. Deng, and J. Hare, "Analyzing and predicting sentiment of images on the social web." *Proc. Int. Conf. Multimedia (MM)*, pp. 715-718 (2010)
- [11] V. Campos, A. Salvador, X. Giro-i Nieto, and B. Jou. "Diving deep into sentiment: Understanding fine-tuned cnns for visual sentiment prediction." *Proc. Int. Workshop on Affect & Sentiment in Multimedia (ASM)*, pp. 57-62 (2015)
- [12] Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Juanzi Li, and Jie Tang, "How Do Your Friends on Social Media Disclose Your Emotions?" *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI'14)*, pp. 306-312 (2014)
- [13] 坂田洋介, 江口浩二, "マージン最大化マルチモーダルトピックモデルによるマルチラベル分類" *DEIM Forum 2016* (2016)

- [1] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs." *ACM MM* (2013)
- [2] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, "Caffe: Convolutional Architecture for Fast Feature Embedding" *arXiv preprint*, arXiv:1408.5093 (2014)
- [3] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory." *ACM MM* (2010)
- [4] J. Yuan, S. Mcdonough, Q. You, and J. Luo, "SentrIBUTE: Image sentiment analysis from a mid-level perspective." *Proc. Int. Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM)*, pp. 10:1-10:8 (2013)
- [5] J.Zhu, N.Chen, H.Perkins and B.Zhang, "Gibbs Max-