

Web 上の感性情報に基づく料理画像からの意味理解支援

堂土 奨[†] 佐野 睦夫[‡]大阪工業大学大学院情報科学研究専攻[†] 大阪工業大学情報科学研究科[‡]

1 はじめに

毎日の食事の栄養推定は、健康サポートや食生活の管理に重要な情報源である。最近では、健康などのアプリケーションツールとして Food Log や食事管理のインターネットサービスが実装されている。そのツールでは、ユーザが何を食べたかをログとして残し食生活の改善が行える。食生活の改善には食事内容の記録が効果的であるということが実証されていることもあり、食事の記録は摂取した栄養バランスを知ることができ、摂取不足や過剰摂取を防ぐことができる。これを実装するために、食材認識が必要とされている。従来の方式では、Bag of Features (BoF) を使用し、画像から局所特徴量を抽出しクラスタリングすることで、辞書を作成し、認識系を構成している[1]。このような方式では、認識精度に限界がある。一方で、認識精度を向上させるために食材学習を用いて食事画像認識している研究もある[2]。しかしながら、これらの研究ではクラスタ数が与える認識精度に与える影響はあまり考慮されていない。今までに k-means 法の手法を取り入れ、最適なクラスタ数を推定し、認識精度の向上を試みてきた。しかしながら、クラスタ数を自分で決める必要がある。この問題を打開する為に自動でクラスタ数を推定する混合ディリクレ過程の手法を本研究では取り入れてきた[3]。その結果、クラスタ数を自動で推定しかつ、手動で入力したクラスタ数の平均値の認識率が自動で推定したクラス多数から得られた。つまり、ある程度の精度があることが分かった。本研究では、この混合ディリクレ過程を用いた方針を従来研究とし、新たに Caffe を用いた深層学習を用いた方式と比較する[4][5]。認識率の違いを検討し、混合ディリクレ過程の必要生を検討する。さらに、画像認識に加え、言語認識も行う。感性情報に関しては、入力画像からどの料理名かを推定するだけではなく、感性情報を言語でユーザに日時場所、季節感、感触、味覚などをワードとして大きく定義し、細分化した内容を Web 上からタグを抽出し統計的に分析する。これを画像に付加することで、画像に言語的な意味付けを行う。

2 画像認識部

k-means 法の手法では、k の値によって結果が大きく変わってしまう。混合ディリクレ過程を用いたクラスタリングを行うことにより、統計的により尤もらしい k の値を自動で判断する。

混合ディリクレ過程のクラスタリングにおける入出力を以下に示す。

- ・入力データ：n 個のデータ ($X_1 \dots X_n$) とする。
- ・出力データ：各データの所属クラスタを表す潜在変数の値は、 $s_1 \dots s_n$ とする。

混合ディリクレ過程によるクラスタリングのアルゴリズムを以下に示す。

STEP1 初期設定

- ・潜在変数 $s_1 \dots s_n$ を初期化する。このときのクラスタ数を c とし、クラスタ ω_i に更新するパターン数を n_i とする ($i = 1 \dots c$)。
- ・この初期結果と各クラスタに所属するパターンを用いて、クラスタのパラメータ $\theta_1 \dots \theta_c$ を初期化する。
- ・さらに、事後確率の最大値 P_{max} を 0 とし初期化する。

STEP2 所属クラスタの更新

- ・ここでは、 $k = 1 \dots n$ に対して実行する。パターン X_k の所属クラスタ $s_k (= \omega_j$ と仮定する) を更新するため、 X_k を現在のクラスタから除外し、 n_j を $n_j - 1$ とする。
- ・これにより、 $n_j = 0$ となり空きクラスタが発生した場合は、この空きクラスタを除去すべく、 c を $c - 1$ とするとともに、j 以降のクラスタインデックスを全て 1 つ減らす。また、それに対応するパラメータのインデックスも更新する。
- ・次に、既存クラスタに所属するパターン数 n_i' 、新規クラスタ数を α 、 $G_0(\theta_{new})$ を事前確率分布とした場合、 s_k の値を確率的に決定する。
 $s_k = \omega_i$ ($i = 1 \dots c$) に対して、計算する。また、 $s_k = \omega_{new}$ に対しても計算する。また、 $c + 1$ 個の値の比の確率で s_k の値を決定する。もし $s_k = \omega_i$ と更新されたとき、 n_i を $n_i + 1$ とする。それ以外で $s_k = \omega_{new}$ と更新されたとき、 n_{c+1} を 1 とするとともに c を $c + 1$ とし総クラスタ数を更新する。

STEP3 各クラスタのパラメータ更新

- ・上記 STEP2 で得た $s = \{s_1 \dots s_n\}$ の値に基づいて、事前確率及びクラスタに所属するパターンに対する尤度を $p(\{k: X_k \in \omega_i\} | \theta_i)$ とし、各クラスタのパラメータ θ_i を確率的に決定し、更新する。

Expression Understanding Support from Food Images by Using Kansei Information on the web

† Dodo Sho ‡ Mutsuo Sano

†Osaka Institute of Technology University

‡Osaka Institute of Technology University

STEP4 事後確率最大化

- ・現時点での $s=\{s_1 \dots s_n\}$ 及び $\theta=\{\theta_1 \dots \theta_c\}$ の値を用い、 $P(S)$ をイーウェンスの公式を用いて求め、ベイズの定理を用いて事後確率 γ を計算する。
- ・事後確率最大化において以下の処理を行う。
 - (1) $\gamma > P_{max}$ ならば以下の更新を行う。
 - ・ P_{max} を γ とする
 - ・ $s=\{s_1 \dots s_n\}$ とする
 - (2) それ以外ならば、これまでの P_{max} と s を保存

STEP5 終了判定

- ・以上のことを繰り返すことにより、 P_{max} の更新されない。状態が継続された場合、実行を終了し、現時点でのクラス数やクラスターの割り振られた情報を出力する。そうでない場合は、STEP2 の工程まで戻り再び実行を繰り返す[4]。

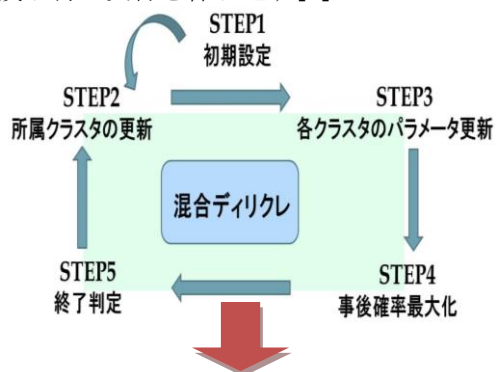


図1 混合ディリクレ過程の構成図

今までの方法と DeepLearning との違いとして特徴量を機械学習で設定してくれるのが DeepLearning である。BoF では特徴量を色と Sift 情報から行なった分手間がかかる。その点の違いからどちらが実用的なのかという点とこの混合ディリクレ過程 (DP-means) 法と Caffe を用いた DeepLearning との認識率の違いを測る。

3 言語処理部

言語情報を提供するに部分では辞書のワードを X 個準備する。さらに、出現頻度を計測し各料理にどのような言語情報が含まれているかを統計的に調べる。ここでは、各料理名を検索し出現した内容を web 上・Twitter 上・CookPad から抽出する[6][7]。方針として、以下のように行う。

1. 各料理名を含む web 上・Twitter 上・CookPad からテキストデータ一文を抽出する。
2. 形態素解析を行う
3. X 個の定義ワード(日時場所、季節感、感触、味覚、イベントなど)がどれくらい含まれているのか出現頻度をカウントする
4. 最も多く出現したワードを各料理の単語とする。
5. ネットワーク分析により関係性をみる。

例えば、「ちらし寿司」(場所：家 季節：3月 感触：しっとり 味覚：甘い イベント：祝い)のように表現する。

4 全体の構成

画像認識部と言語処理部の両サイドから検証を行う。画像を認識し、言語情報を出力する側と言語情報から画像を推定する側から全体構成を担っている。トップダウン・ボトムアップの手段から以下のような構成で行う。

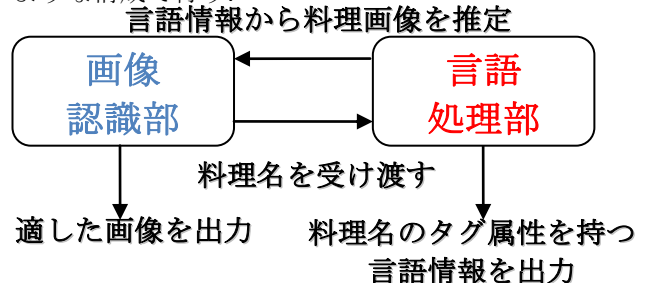


図2 全体の構成図

5 実験方法

料理画像には洋・日本菓子 10 種類×100 枚を学習画像とし、入力画像 10 種類×10 枚に対して何枚料理名を認識できたかで認識率を計算する。混合ディリクレ過程を用いた認識については、自動でクラスター数が推定されるので出力されたクラスター数に対して認識を行う。DeepLearning も同様の枚数で行。この2つの手法を比較し、結果を出力する。

言語処理に関しては、web 上・Twitter 上・CookPad それぞれの 100 件を統計データとして使用する。また、それぞれの環境での違いを検討する。

また、インタフェースとしての観点からも評価を行う。



図3 ケーキを推定した時の一例

文献

- [1] 田中雄翔, 滝口哲也, 有木康雄” ウェブ画像を用いたカテゴリ別 Visual Words による未知物体判別,” 画像の認識・理解シンポジウム(MIRU2012), 2012
- [2] 数藤恭子, 村崎和彦, 神谷叔季, 谷口行信,” 料理画像の素材セグメンテーションに基づく成分推定,” 電子情報通信学会誌, No. 468 (IMQ), No. 469 (IE), No. 470 (MVE), pp. 73-76, 2014.
- [3] 上田修功, 山田武士,” ノンパラメトリックベイズモデル,” 電子情報通信学会技術研究報, pp. 81-86, 2007.
- [4] 上田修功, 石田健一郎,” 続・わかりやすいパターン認識—教師なし学習入門,” pp. 260-261, 2014, オーム社.
- [5] 石橋 崇司” Caffe をはじめよう 深層学習による画像解析の実践,” 2015
- [6] 加藤 大介, 宮部真衣, 荒牧英治, 瀧本明代,” インターネット上のメディア毎の「おいしさ」表現比較分析,” DEIM Forum2015, 2015.
- [7] Willi Richert, Luis Pedro Coelho,” 実践機械学習システム,” 2014