

食事ログに向けた画像処理による料理の味推定

宮崎 達也¹ デシルヴァ, ガムヘワゲ チャミンダ² 山崎 俊彦³ 相澤 清晴^{2,3}

¹東京大学工学部電子情報工学科

²東京大学大学院情報学環

³東京大学大学院情報理工学系研究科電子情報学専攻

1. はじめに

近年, 料理メディアや食事画像への関心が高くなってきている. 例えば, 写真アップロードサイトの Flickr[1]には「food」というタグの付いた写真が 350 万枚以上もアップロードされている. 研究の分野では, 料理メディアを専門に研究する, 料理メディア研究会という研究会が 2007 年に発足している. また, Web 上にアップロードされた写真から食事画像を選び分け, その食事の食事バランスを推定する FoodLog[2][3]というサービスが 2009 年に開始されたり, 画像に写った料理を 50 種類のメニューの中から推定する研究[4]がなされたりしている.

一方, 食事の中において, 味は非常に重要な要素である. 例えば, 日本国内にも「味と匂学会」という, 物の味や匂いに関する研究の進展を目的とした学会が存在する.

しかし, このような中にあっても食事画像と味とを結び付ける研究は未だなされていない. そこで, 本稿では料理の画像から, その料理の持つ味を予測することの可能性について検討する. 本研究では, 最終的には食事ログ(FoodLog)[2]で食事バランスと共に表示することを目的とする. さらに, 予測した味の表現方法から味の表現方法の構造化をすすめることも目標としている.

2. 味について

人間が感じる味は, 味覚による刺激だけでなく, 「香ばしい」などの嗅覚由来の刺激や, 歯触り, 舌触りなど触覚による刺激の影響も受けている. さらに, 「おいしい」「まずい」などの主観的な感想も味を表す要素のひとつであると同時に, それらの感想は, 周囲の環境や, 個人の記憶, 経験に左右される. 以上のような要因から, 味を完璧に再現し, 表現するのは非常に困難であると考えられる.

そこで, 本研究ではその中から基本五味と辛味とを使用して味を表すことにした. 基本五味とは, 甘味, 苦味, 酸味, 塩味, うま味の 5 つで, 理論上は基本五味の組み合わせで全ての味覚を表現できるとされている. 辛味は, JIS 規格[5]で基本味として基本五味以外に定められている味のひとつで, 基本五味のいずれとも大きく異なる刺激であり,

Taste Estimation by Image Processing for “FoodLog”
Tatsuya Miyazaki¹, Gamhewage Chaminda De Silva²,
Toshihiko Yamasaki³ and Kiyoharu Aizawa^{2,3}

¹Department of Information and Communication Engineering, The University of Tokyo

²The University of Tokyo Interfaculty Initiative in Information Studies

³Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo

一般に広く用いられていることから, 基本五味と共に用いた.

全ての味の有無を判定することは困難であるため, 本稿では最も代表的な味を 1 つ推定することを目的とする. ただし, 2 つ以上の味が同じスコアを持つ場合はその全てを代表の味とする.

3. ユーザスタディ

食事画像から味が予測できるかを調べるため, 事前に人間がユーザスタディを行った. 使用した画像は 41 枚, 回答者は 4 名で, それぞれの画像について, 今回用いる 6 つの要素それぞれの有無を 5 段階で予測評価してもらった. それぞれの回答は独立になっている.

事前に著者らが二値で設定した 6 つの要素それぞれの「あり」「なし」を正解とし, 各回答者の正解率を示したものが表 1 である. 設定した解答に対する正解率が全員 70%以上となったことから, 画像から得られる情報を元に味を予測することは可能であると推測した.

ただし, ユーザスタディで回答を行うのは人間であるため, 画像内の特徴から直接味を予想しているのではなく, メニューの特定など, 機械による予測では難しい過程を経た予測結果であると考えられる.

表 1 各回答者の正解率

回答者	正解率
A	0.73
B	0.78
C	0.77
D	0.70

※それぞれの要素に対して, 5 段階評価の下 3 段階を「なし」上 2 段階を「あり」として正解, 不正解を評価した.

4. 実験

食事画像から特徴量を抽出し, 機械学習器を用いて料理の味予測を行う実験を行った. 食事画像として, 事前に料理部分のみ手動で切り取った画像を使用した.

事前に, 各味の要素の予測にはどのような特徴量が向いているか実験した. その結果は表 2 である. 表 2 中の R+G+B とは, R, G, B それぞれの BoK によるヒストグラムを連結したものである. この事前実験では, 500 枚の料理部分切り抜き済み画像を使用し, 機械学習器として Support Vector Machine (SVM) を使用した. そして, それぞれの特徴量, 味の要素に対して, 再現率を求めた. 一連の予測を, 甘味, 苦味, 酸味, 塩味, うま味, 辛味の 6 つ全てに行った.

表 2 の結果に基づき行った機械学習による味の予測実験の処理の流れを, 図 1 に示す. 特徴量はグレースケール画像, R 値画像, G 値画像, B 値画像の SIFT 特徴量を抽出し, それぞれの画像群

から Bag of Keypoints (BoK)[6]により求めた特徴ベクトルのヒストグラムを用いた. SVMを用いた予測をグレースケール, R, G, B, R+G+B それぞれの画像に対して行い, それらの予測結果をまとめ, 最終的な結果を求めた. この際使用した特徴量は味の要素により異なり, それを示しているのが表 2 の太字部分である.

図 1 から求まるものは, 味のうちのある 1 つの要素の有無の度合いを表すスコアであり, 0~1 の小数とした. 学習時は, それぞれの味について「あり」「なし」がほぼ同数になるように, データベースからランダムに選択した. それぞれ甘味は 227 枚, 苦味は 161 枚, 酸味は 116 枚, 塩味は 156 枚, うま味は 146 枚, 辛味は 200 枚の画像を使用した. また, テストデータは全ての要素で, 共通のもの 185 枚を使用した.

以上のようにして 6 つの要素に対して予測を行い, それらの予測結果の中で値が最も大きいものをその食事の味とした. 同じスコアを持つものが 2 つ以上あった場合には, 全てをその食事の味の予測結果とした.

表 2 各味の要素ごとの特徴量別正解率

	甘味	苦味	酸味	塩味	うま味	辛味
SIFT(Gray) + BoK	0.61	0.64	0.53	0.64	0.54	0.68
SIFT(R+G+B) + BoK	0.64	0.68	0.55	0.67	0.57	0.71
SIFT(R) + BoK	0.62	0.67	0.56	0.59	0.58	0.70
SIFT(G) + BoK	0.60	0.64	0.54	0.63	0.60	0.71
SIFT(B) + BoK	0.64	0.68	0.50	0.64	0.60	0.68
HoG	0.52	0.53	0.50	0.50	0.51	0.50
ColorCorrelogram	0.63	0.67	0.50	0.56	0.60	0.50
HSV Histogram	0.50	0.50	0.50	0.50	0.50	0.50
RGB Histogram	0.50	0.50	0.50	0.50	0.50	0.50

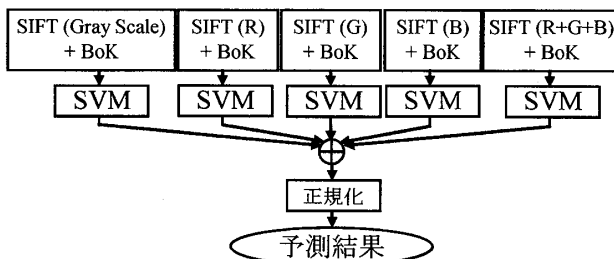


図 1 味の要素ごとの予測の流れ

5. 結果・考察

ユーザスタディからもわかるように, 食事の味を完全に当てるのは非常に難しいため, 正解データに対して予測結果が正解しているということを次のように定義した. ただし, 代表味候補が 3 つ以上でた場合は不正解とした.

- 正解データと予測とが完全に一致している場合は正解とする.
- 正解データよりも 1 つだけ余計に要素を予測している場合も正解とする (例: 正解: うま味, 予測: うま味, 辛味).
- 正解データに含まれる味の要素が 2 つの場合, 片方を予測できていれば正解とする (例: 正解: 塩味, うま味 予測: 塩味)

実験による予測の結果, 正解率は 43% となった. ランダムに 6 つの要素の中から 1 つまたは 2 つを選んだ場合, 正解率は 17%~33% の間になる. 最も高く見積もった場合でも 33% であるので, 予測結果はランダムの場合よりは高く, ある程度意味を持った結果であると言える.

予測に成功した画像の例を図 2 に示す. 各食品名と共に括弧付きで書いてある味がその食品に設定した味である.

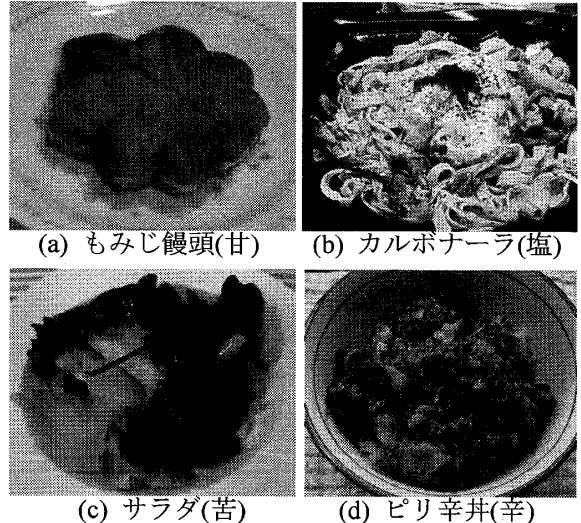


図 2 予測に成功した食事画像の例

6. まとめ

本稿では, 食事の味予測について検討するため, 味の要素の予想に向けた特徴量の調査を行い, それを元に機械学習を用いて食事の味予測を行った. その結果, 43% の精度で正解を出すことができた.

今後はデータセットの改良や, 抽出する特徴量の工夫などを行い, 精度の向上を目指したい.

参考文献

- [1] Flickr <http://www.flickr.com/>
- [2] FoodLog <http://www.foodlog.jp/>
- [3] 北村 圭吾, 山崎 俊彦, 相澤 清晴: “食事ログの取得と処理 : 画像処理による食事記録支援” 映像情報メディア学会誌, Vol.63, No.3 pp. 376-379 (2009)
- [4] 上東 太一, 甫足 創, 柳井 啓司: “Multiple Kernel Learning による 50 種類の食事画像の認識” MIRU, pp. 111-118 (2009)
- [5] 日本工業標準調査会: JIS Z 8144: 2004 官能評価分析——用語 (2004)
- [6] Csurka, G., Bray, C., Dance, C. and Fan, L.: “Visual categorization with bags of keypoints,” European Conference on Computer Vision Workshop on Statistical Learning in Computer Vision, pp. 1-22 (2004)