

P-11

サポートベクトルマシンによる海苔の品質判定における HSV 色空間の有効性 Effectiveness of HSV Color Space in Grading Dried Seaweed Quality Using Support Vector Machine

○鈴木開渡¹, 平石秀史²*Kaito Suzuki¹, Hidefumi Hiraishi²

Abstract: The purpose of this research is to automatically grade dried seaweed using support vector machine (SVM). Previous research tackles this problem using deep learning. However, deep learning tends to be haunted by typical obstacles: high-level requirements of computational power and a scarcity of datasets. We exploited SVM and obtained the desired accuracy with low computational costs and a small amount of data.

1. はじめに

海苔の市場価格は、品質によって決定されるため、生産者には重要な関心事である。海苔の品質判定は、生産地ごとの有資格者の検査員によって、目視で実施される。しかし検査員の高齢化により、後継者不足深刻な問題となっており、海苔の等級判定の自動化が要望されている。

自動化の先行研究として、カラーリーダーを用いたもの [1] や深層学習 (特に転移学習) を用いたもの [2] [3] がある。深層学習は画像の分類を得意とする一方で、高精度の分類機を作るには、大規模な計算資源と大量のデータセットを必要とする場合が多い。しかし、海苔は季節性の海産物で、一度に大量のデータを入手することが難しいため、我々は、少ないデータセットでも、高精度の分類を実現しうるサポートベクトルマシン(SVM)に着目した。

2. 実験手法

実験に用いる海苔の画像データを事前に解析した結果、等級が低くなるにつれ、全体的に赤みがかかるなど色味に関して、いくつかの特徴がみられることが分かった。この特性に着目し、海苔の画像データの等級間の色味の違いに基づいて判定器を作成する。本実験では画像の各ピクセルの色を特徴量として用いる、一つのデータにつき約 33 万の特徴量を用いて実験を行う。

2.1 実験環境

本研究では、1から5等級の5つのクラスを分類する多クラス分類を行う。以下では、簡単のため、2クラス分類を行うソフトマージンSVMについて説明をする。ソフトマージンSVMは以下の最適化問題として定式化される。詳細は [4] を参照のこと。

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$

$$\text{subject to } t_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad (i = 1, 2, \dots, n) \\ \xi_i \geq 0 \quad (i = 1, 2, \dots, n)$$

ここで、 $n, t_i, \mathbf{x}_i, C, \xi_i$ はそれぞれ、学習データの数、学習データのクラスを表すラベル、学習データの特徴量ベクトル、コストパラメータ、誤分類の度合いを表現する人工パラメータである。ソフトマージンSVMは学習データを誤差を許容した上で超平面により分類するモデルである。コストパラメータは、誤分類に対するペナルティを制御する値で、これが大きければ誤分類に対し厳格であり、小さければ誤分類を許容する。本実験ではコストパラメータは1に設定する。また、SVMではデータを高次元の特徴空間に写像するための関数 (カーネル関数) を設定することもできる。カーネル関数により分類の精度が左右されうるが、今回は最も単純な線形カーネルを用いる。

2.2 画像データ

用いた画像は1等級45枚、2等級45枚、3等級45枚、4等級44枚、5等級46枚の計225枚である。画像サイズは3400×3250 pixel、画像形式はjpegである。これらは [2] [3] のデータセットと同一のものである。

2.3 前処理

縮小 オリジナルの画像では特徴量の数 (ピクセル数×色チャンネル数) が約3300万となり計算量が膨大になるため、340×325 pixelまで縮小変換を行う。

色空間の変換 海苔の画像の色空間をRGB色空間からHSV色空間に変換して学習を行う。

表1: RGB色空間における等級判定テストの正答率(小数点以下第3位を四捨五入)

| セット | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 計 |
|------|-------|------|-------|-------|------|------|------|------|-------|------|-------|-------|-------|------|-------|---------|
| 正解枚数 | 11/15 | 9/15 | 11/15 | 14/15 | 9/15 | 6/15 | 9/15 | 8/15 | 12/15 | 5/15 | 11/15 | 11/15 | 10/15 | 9/15 | 12/15 | 147/225 |
| 正答率 | 0.73 | 0.60 | 0.73 | 0.93 | 0.60 | 0.40 | 0.60 | 0.53 | 0.80 | 0.33 | 0.73 | 0.73 | 0.67 | 0.60 | 0.80 | 0.65 |

表2: HSV色空間における等級判定テストの正答率(小数点以下第3位を四捨五入)

| セット | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 計 |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| 正解枚数 | 15/15 | 15/15 | 15/15 | 15/15 | 14/15 | 15/15 | 14/15 | 13/15 | 14/15 | 15/15 | 15/15 | 14/15 | 15/15 | 14/15 | 14/15 | 217/225 |
| 正答率 | 1.00 | 1.00 | 1.00 | 1.00 | 0.93 | 1.00 | 0.93 | 0.87 | 0.93 | 1.00 | 1.00 | 0.93 | 1.00 | 0.93 | 0.93 | 0.96 |

HSV色空間は、色相 (Hue)、彩度 (Saturation)、明度 (Value) の3つのパラメータで表現され、人間の色知覚に近いモデルである。

色相 (Hue) は、色の種類を表し、 0° から 360° の範囲で表され、赤、青などの色味を決定するパラメータである。

彩度 (Saturation) は、色の鮮やかさや純度を表すパラメータで、0%から100%の範囲で表される。彩度が0%であることは、鮮やかさがないことを示し、黒色を表現する。

明度 (Value) は、色の明るさを表すパラメータで、0%から100%の範囲で表される。明度が0%であることは、色は明るくないことを示し、黒色を表現する。

一方、RGB色空間では黒色の表現が単純であり、色相と明度が同時に表現されるため、黒色を明るさが低い色であると捉えられてしまう。したがって、HSV色空間の方が、黒色の表現方法がより多様で、色相と彩度を保ったまま明度を調整でき、異なる黒色のバリエーションを表現することができる。ゆえに、HSV色空間は海苔の画像データを表現するに適した色空間と言える。

2.4 学習データとテストデータの分割

各等級3枚の画像をテストデータ、それ以外の画像を学習データとし、乱数を用いてランダムに分ける。このデータセットで学習からテストまで実施し、SVMの予測の正答率を求める。これを1セットとし、15セット繰り返す。なおどのデータセットにおいても他のデータセットとのテストデータの重複はなく、全ての画像に対して1回のテストが行われる。

ただし4等級の画像が44枚、5等級の画像が46枚であることより、15セット目のテストデータは4等級の画像が2枚、5等級の画像が4枚である。

3. 結果と結論

RGB色空間で実験を行なったものを表1、RGB色空間からHSV色空間に変換し実験を行なったものを表2に示す。

実験結果より、RGB色空間においては学習データ、テストデータによって正答率が大幅に上下していて不安定である。一方HSV色空間においては十分な判定器を作成できており、海苔の画像に対しては、RGB色空間よりもHSV色空間の方が適していることが、わかる。このことより、SVMで分類を行う際は、分類対象の特性に合わせて色空間を選択することで、分類性能を大幅に向上できることがわかった。

また、15セット分の学習からテストまでの計算時間は、5分程度であった。深層学習の場合、高性能なマシンを使用した場合、1時間程度、個人の持つPCでは20時間程度であったことから大幅に計算時間を短縮することができた。しかし、今回作成した判定器では、人間の眼では認識することが難しいレベルの微小なノイズをテストデータにかけた場合、正答率が85%~90%まで減少した。そのため、今後は微小なノイズがかかった海苔の画像に対しても安定して高い精度で判定できる判定器を作成したい。

4. 参考文献

- [1] 安保康治, 野村 実咲, 大内 美佳, カラーリーダーによる板海苔の品質評価, 久留米信愛女学院短期大学研究紀要, No.32, pp.29-31, 2009.
- [2] 鈴木大智, 栗野俊一, 深層学習による海苔の品質判定のデータ分割の有効性, 令和3年度日本大学理工学部学術講演会予稿集, P-1, pp.831-832, 2021.
- [3] 阿部舜耶, 平石秀史, 転移学習による海苔の品質判定における Data Augmentation の有効性, 令和4年度日本大学理工学部学術講演会予稿集, P-1, pp.830-831, 2022.
- [4] 田村孝廣, やさしく学べるサポートベクトルマシン, オーム社, 2022.