

# 決定木および決定ネットワークの画像分類過程の可視化

小林 雅幸<sup>†</sup>

長尾 智晴<sup>††</sup>

<sup>†</sup> 横浜国立大学 理工学部

<sup>††</sup> 横浜国立大学 大学院環境情報研究院

## 1 はじめに

近年、機械学習による画像分類器の自動生成の研究が盛んに行われている。画像分類において高い分類精度が求められる一方、機械学習で作られた分類器を安心して使いたいという人間の要求がある。そのため、画像分類器は高い分類精度と同時にその分類過程を利用者に分かりやすく説明をする必要がある。

分類器の中でも昔から比較的に理解がしやすい構造として、if-then 形式で分類を行う決定木がある。また、我々の研究グループでは、決定木を拡張し、高精度かつコンパクトで可読性のある構造を自動構築する進化的条件判断ネットワーク(Evolutionary Decision Network; EDEN)[1] とその言語化の手法[2] を先に提案している。図1に EDEN の構造、図2に言語化の例を示す。

EDENでは進化計算を用いて、画像の特徴量としきい値の組み合わせをネットワーク構造状に最適化することで分類器の自動構築をする。EDENの言語化では、特徴量やしきい値に対応する語句、その後続く語尾を変化させることで説明過程を説明する文章を自動生成する。しかし、この言語化の手法では適切な表現方法がない複雑な特徴量は語句による説明が困難であり、利用者への分かりやすさが低下するという問題がある。

本研究ではEDENや決定木において視覚情報を用いて複雑な特徴量に対しても分類過程を分かりやすく説明する手法を提案する。

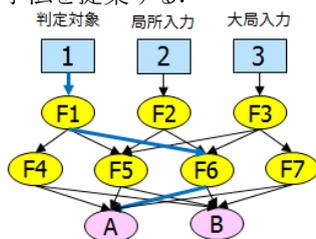


図1 EDEN の構造



黄色っぽい部分が多くて、ある程度明るくて、色数が少ないため、ひまわりである

図2 EDEN の言語化例

## 2 提案手法

本手法は分類データを流した際に通過したパスに対する分類画像例の提示と、画像中で各特徴量が強く反応する部分の可視化の2つを用いて分類過程の説明をする。

### 2.1 分類画像例の提示

ここでは、分類データを流した際に通過したパスについて、データの分布を考慮してパスの代表画像を決定する。はじめに、通過したパス上の各ノードで、流したデータの特徴量に対してクラスタリング手法のX-means法[3] を適用し、分岐ノードごとにデータのクラスタリングを行う。次に各データが属するクラスタとの中心距離の大きさからスコアを算出する。このスコアをデータ集合の性質度とし、これをパス内の各ノードについて算出し、性質度のヒートマップと分類画像例を得る。

### 2.2 特徴量の強く反応する部分の可視化

EDENの言語化では分類データが通過したパス上のノードで使用されている特徴量としきい値から、あらかじめ作成した辞書を用いて対応する語句に変換して説明文を生成する。本提示手法では、自然言語では説明が難しい複雑な特徴量に対して、画像内のどの部分が強く反応するかを提示することで説明を行う。具体的には、画像内の全ての画素についてある画素に注目したときに、ノードごとの分岐から特徴量ごとに説明文中の複雑な表現が画像内のどの部分を示しているのかを提示する。

## 3 視覚情報を用いた分類過程の説明の実験

### 3.1 実験設定

本手法を2クラス分類の画像に適応した。そして、生成された説明文および提示画像について主観評価実験を行った。

### 3.2 実験結果および考察

図3に分類画像例、図4に特徴量マップの例を示す。生成された説明文について6段階の主観評価を行った結果、説明文のみでは分かりやすさが平均1.88であったのに対して、画像を用いて説明した場合には分かりやすさが平均4.18という結果を得ることができた。より理解がしやすい分類過程の説明には提示方法などの検討が必要である。



図3 分類画像例



図4 特徴量マップ

## 4 まとめ

本稿では決定木およびEDENの分類過程を、視覚情報を用いて説明する手法を提案し、得られた結果について主観評価実験を行い評価した。今後、さらに直感的で分かりやすい分類証拠の提示方法の検討が必要である。

## 参考文献

- [1] 中山史郎, 穂積知佐, 矢田紀子, 長尾智晴“進化的条件判断ネットワークEDENによる画像分類”, 映像情報メディア学会誌, Vol.67, No.7, pp.J278-J285(2013)
- [2] 崎津実穂, 菅沼雅徳, 土屋大樹, 長尾智晴“決定木及び決定ネットワークによる画像分類過程の説明文の自動生成”, 情報処理学会論文誌: 数理モデル化と応用(TOM) (2015)
- [3] Dan Pelleg and Andrew Moore “X-means: Extending K-means with Efficient Estimation of the Number of Clusters”, ICML-2000