

Extreme Learning Machine を用いた ハフフォレストによる物体検出

岩本 恵太[†] 中山 英樹[†]
東京大学大学院情報理工学系研究科[†]

1 はじめに

画像認識分野において物体検出とは、画像から人や車や動物のような一般的なクラスを推測し、さらにその領域を矩形等で特定するタスクである。このタスクは、車の自動運転や監視カメラの物体追跡など実世界への応用が広い。

実世界の物体検出に適した手法の一つとしてハフフォレスト [2] が挙げられる。ハフフォレストとは木構造の識別・回帰モデルである決定木を複数用いてアンサンブル学習させるランダムフォレスト [4] の応用である。このモデルは実世界で起こりうるような物体の遮蔽に頑健で、高速な物体検出が可能である。

ハフフォレストはランダムフォレストをベースにした手法であるため、与えられた入力を適切にクラスタリングするような決定木の分割関数が定義される事で精度が向上することが期待できる。そのため本研究では、ハフフォレストの持つ決定木の分割関数について提案する。具体的には分割関数に非線形な関数を利用し、その学習に Extreme Learning Machine [1] と呼ばれる学習モデルを活用することでハフフォレストの精度向上を図る。

2 関連研究

2.1 ハフフォレストによる物体検出

ハフフォレストが持つ決定木は 2 分木構造であり、2 つの子ノードと分割関数を持つような内部ノードと子ノードを持たない末端ノードで構成される。

ハフフォレストによる物体検出では、画像から取り出されるパッチと呼ばれる矩形の小領域を入力とする。取り出したパッチは C 種類の画像特徴量のマップを持つ。始めにパッチ $I = \{I^1, I^1, \dots, I^C\}$ を親ノードを持たない根ノードに与え、各内部ノードでは以下の分割関数 t の出力結果によって子ノードのどちらに遷移させるかを決定する。

$$t(I) = \begin{cases} 0 & \text{if } I^f(p) - I^f(q) < \tau \\ 1 & \text{otherwise.} \end{cases}$$

全ての分割関数はチャンネル f の 2 変数 p と q と閾値 τ による超平面により、パッチが遷移する子ノードを決定させる。



(a) 入力画像 (b) 尤度マップ

図 1 : Weizmann Horses [5]

このような操作を末端ノードに到達するまで繰り返す。そして到達した末端ノードが持つ訓練パッチによって、入力したパッチの位置から検出対象の中心までの相対位置を投票ベースで推定する。これを画像からラスタ状に取り出した全てのパッチに対して行うことで、図 1(a) の入力画像から図 1(b) のような尤度マップが得られる。この尤度マップは白の濃さが物体の中心であるという確信度を表す。そのため、尤度マップの適切な位置にて値が大きくなるような学習が精度向上の糸口となる。

2.2 ハフフォレストの学習

ハフフォレストの学習では、検出対象を含むパッチ (ポジティブパッチ) と含まないパッチ (ネガティブパッチ) を集めたパッチ群を訓練データ集合とする。訓練データの各パッチ I_i には入力とは別に、以下の 2 つの教師情報を含む。

- 検出対象を含むか否かのクラス情報 : $c_i \in \{0, 1\}$
- 検出対象の中心からのオフセット情報 : $d_i \in \mathbb{R}^2$

ハフフォレストの学習の主な目的は、それぞれの決定木における内部ノードの分割関数と末端ノードが持つ訓練パッチの決定である。内部ノードが保有する分割関数は、同じクラスのパッチ及びオフセットのベクトル情報が近いパッチ同士が同じ子ノードに遷移されるようになるのが望ましい。そのため学習では、まず分割関数の組み合わせ $\{f, p, q, \tau\}$ をランダムに複数生成する。そして各分割関数によって二分された訓練データ集合の評価し、最も良い組み合わせを内部ノードの分割関数として決定する。

評価値には、ポジティブパッチとネガティブパッチの分離度を評価する Class-label Uncertainty とパッチのオフセットのばらつきを評価する Offset Uncertainty のどちらかを内部ノード毎にランダムで用いる。

ランダムに生成された分割関数のうち選ばれるのは 1 つであったが、これは残りの分割関数の候補が未知パッチに頑健であった場合は機会損失となる。そこで本研究ではランダムに生成した複数の分割関数を活かして、パッチからできるだけ多くの情報が取り出せるような分割関数を提案する。

Object Detection by Hough Forest using Extreme Learning Machine

[†] Keita Iwamoto and Hideki Nakayama

[†] Graduate School of Information Science and Technology, The University of Tokyo

3 提案手法

提案手法では従来の分割関数の組み合わせを、 E 個ずつランダムに生成したパラメータ $\{f_j, p_j, q_j, \tau_j\}_{j=1}^E$ を用いる。パッチ I についての分割関数 $t(I)$ を以下の式で表す。

$$t(I) = \begin{cases} 0 & \text{if } o(I) < 0 \\ 1 & \text{otherwise.} \end{cases}$$

$$o(I) = \sum_{j=1}^E g(h_j(I))\beta_j$$

$$h_j(I) = I^{f_j}(p_j) - I^{f_j}(q_j) - \tau_j$$

分割関数 $t(I)$ は目的変数 $o(I)$ の正負によってバイナリを出力する。さらにパッチから取り出される E 個の説明変数 $h_j(I)$ と目的変数 $o(I)$ の関係は、非線形なシグモイド関数 $g(x) = 1/(1 + e^{-x})$ と重み $\{\beta_j\}_{j=1}^E$ によって決まる。

学習では説明変数と目的変数の関係を表す重み $\{\beta_j\}_{j=1}^E$ を決定させなければならない。提案手法では、分割関数に入力層・中間層間の重みがスパースな中間層次元 E ・出力層次元 1 の Extreme Learning Machine (ELM) [1] を構築して、中間層・出力層間に対応する重み $\{\beta_j\}_{j=1}^E$ を解析的に求める [1]。

また ELM の学習にはパッチ I ごとに、 $o(I)$ を決定するようなラベル付けが必要となる。そこで内部ノードの評価値が Class-label Uncertainty の時はパッチが持つクラス情報 c , Offset Uncertainty の時は訓練データのオフセット $D = \{d_1, d_2, \dots, d_n\}$ の第一主成分とパッチが持つオフセット d の比較から望ましい $o(I)$ を決定させるようにラベル付けを行う。

4 評価実験

提案手法と従来のハフフォレストの比較を行うために評価実験を行う。実験では Weizmann Horses [5] と呼ばれる馬検出のためのデータセットを利用した。このデータセットは訓練画像としてポジティブパッチ・ネガティブパッチ生成用の画像が合わせて 200 枚とテスト画像の 454 枚からなる。訓練・学習ではデータセットの RGB 画像から Lab 色情報や HoG 特徴量などによる豊富な特徴量 (チャンネル数 $C=32$) を抽出した。実験では学習のために訓練画像からポジティブパッチとネガティブパッチをそれぞれ 25000 個生成した。

ハフフォレストのパラメータとして、決定木は 15 本、深さの最大は 15、パッチサイズは 16×16 、生成する分割関数のパラメータ候補数は 20000 とする。さらに提案手法における説明変数の個数 (ELM の中間層次元数) は $E=\{100,150,200\}$ とする。また、ELM のデータへの過学習と性能の関係を見るために正則化項 $t=\{5000, 0.005\}$, ELM の学習に用いるパッチの最大は 50 とする。また、計算コスト削減のため深さ 10 までの分割関数は従来の線形な分割関数を利用し、それ以降の深さでは提案手法の非線形な分割関数を利用する。

尤度マップの極大点が閾値よりも大きい場合にその点を中心

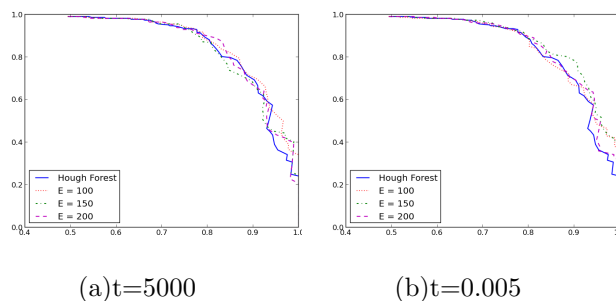


図 2 : Precision Recall Curve

とする矩形が物体の領域であると学習モデルが推測したとみなす。予想した矩形と正解の矩形の重複度が 50% を越えていればその検出は正解とし、Precision = (推測の正解数) / (推測した矩形数), Recall = (推測の正解数) / (正解の矩形数) という 2 つの評価指標を定義する。この 2 つの指標をとりうる閾値で走査して Precision-Recall Curve を各パラメータから出力した。

以上の設定で従来手法・提案手法ともに実験を行った結果が図 2 である。この曲線は右上に近い曲線が描けるほうが優れた尤度マップを出力するモデルであることを示している。図 2(a) は $t=5000$ の実験結果で、図 2(b) $t=0.005$ の実験結果である。どちらの画像に対しても概ね提案手法の隠れ層次元数が大きいハフフォレストの方が精度が良いという結果が得られた。

5 今後の展望

提案手法は一定の深さまでは従来の線形な分割関数を利用し、それ以降の深さでは非線形な分割関数を利用する手法であり、実験によって精度が向上したことがわかった。今後は未知の入力パッチについて分割関数の汎化性能が高くなれば良い判別ができることを仮定し、ELM の学習方法についても工夫する。例えば Batch Intrinsic Plasticity [3] のように、入力層・中間層間の重みを改善する初期化手法を用いることでさらに精度の向上が期待できる。また提案手法の根幹は分割関数に関するアプローチであり、分割関数以外でハフフォレストを改良させる手法との併用が可能である。将来的にはそれらを組み合わせることで、更なるハフフォレストの性能向上を期待したい。

参考文献

- [1] G.-B. Huang et al., "Extreme Learning Machine: Theory and Applications," *Neurocomputing*, Volume 70, no 1, pages 489-501, 2006.
- [2] J. Gall et al., "Class-Specific Hough Forests for Object Detection," In Proc. of CVPR, pages 1022-1029, 2009.
- [3] K. Neumann et al., "Batch Intrinsic Plasticity for Extreme Learning Machines," In Proc of ICANN, pages 339-346, 2011.
- [4] L. Breiman, "Random Forests," *Machine Learning*, volume 45, pages 5-32, 2001.
- [5] E. Borenstein et al., "Class-Specific, Top-Down Segmentation," In Proc. of ECCV, pages 639-641, 2002.