

H-041

処理コストを考慮した決定木学習に基づく高速ショット境界検出

Fast Shot Boundary Detection Method Based on Decision Tree Learning
Considering Processing Cost河合 吉彦† 住吉 英樹† 柴田 正啓†
Yoshihiko Kawai Hideki Sumiyoshi Masahiro Shibata

1 まえがき

ショット境界検出は、映像解析における最も基本的な処理のひとつであり、高速かつ高精度な処理が求められる。特にNHKアーカイブスのような大量のデータを処理する場合には、計算コストは非常に重要な要素となる。近年では、フレーム画像から複数の画像特徴を算出し、SVMなどの機械学習によってショット境界を判定する手法が多く提案されているが、精度向上のために様々な画像特徴を利用すると計算コストが高くなるという問題があった。そこで我々は、画像特徴の算出とショット境界の判定を逐次的に実行し、ショット境界の可能性が低いフレームについては早い段階で処理を打ち切ることによって計算量を削減する手法を提案した[1]。しかし、どのような画像特徴をどのような順番で判定するかについては経験的に決定しているという問題が残されていた。そこで本稿では、利用する画像特徴や判定順序を決定木学習に基づいて決定する検出手法を提案する。提案手法では、データの識別精度に加えて、画像特徴の算出に要する計算コストも考慮できる新たな決定木学習アルゴリズムを提案し、処理速度と検出精度を両立したショット境界検出器の構築を試みる。実験では、実際のテレビ番組映像に提案手法を適用し、手法の有効性を検証する。

2 従来の決定木学習

決定木では、事前にすべての特徴量を計算する必要がなく、ノード分岐において必要となった特徴量のみを逐次的に算出すればよい。そこで本手法では、決定木をベースに、逐次判定に利用する特徴量と判定順序の学習を試みる。まず、従来の決定木学習について説明する。

代表的な従来手法としては、ID3[2]、C4.5[3]、CART[4]などがある。基本的には、いずれの手法も不純度を最も減少させる特徴量を探索して学習データを再帰的に分割していき、データがそれ以上分割できない、あるいは不純度を減少させる特徴量が見つからなければ分割を終了する。最後に、必要に応じて決定木の枝を剪定(枝刈り)する。各手法の大きな違いは不純度の計算方法である。以降では、不純度としてエントロピーを用いた場合を例に、具体的な処理の流れを説明する。

学習データの集合を D 、葉の集合を L とすると、決定木における条件付エントロピーは以下のようにかける。

$$H(D|L) = \sum_i P(l_i)H(D|l_i) \quad (1)$$

$$H(D|l_i) = - \sum_j P(d_j|l_i) \log P(d_j|l_i) \quad (2)$$

ここで、 l_i は決定木におけるそれぞれの葉を表す。また、 $P(l_i) = C(l_i)/N$ 、 $P(d_j|l_i) = C(d_j|l_i)/C(l_i)$ であり、 N は学習データの総数、 $C(l)$ は葉 l に属するデータの総数、 $C(d|l)$ は葉 l に属するデータのうちクラスが d であるデータの総数を表す。

特徴量 q_k によって、葉 l_i の下に2つの葉 $l_{i,s}$ ($s = 0, 1$) を生成したとすると、新しい葉における条件付エントロピーは次のように書ける。

$$H(D|l_i, q_k) = \sum_s P(i_{i,s})H(D|l_{i,s}) \quad (3)$$

よって、新しい葉の生成による決定木全体のエントロピーの減少量は、下式で求められる。

$$\Delta H(D|l_i, q_k) = H(D|l_i) - H(D|l_i, q_k) \quad (4)$$

木を成長させる手順は以下のとおりである。

1. 根のみの決定木を作成する。
2. $\Delta H(D|l_i, q_k)$ が最大となる \hat{l}_i と \hat{q}_k を選ぶ。
3. $\Delta H(D|\hat{l}_i, \hat{q}_k) < \epsilon$ なら終了。 ϵ は閾値。
4. 葉を追加して2.に戻る。

最後に、枝刈りを実施して学習を終了する。

3 処理コストを考慮した決定木学習

従来の決定木は、不純度の減少量のみに基づいて特徴量を選択しており、生成された決定木を用いて実際にデータを分類する際の計算コストについては全く考慮されていない。例えば、特徴量の算出に数百倍の計算時間が必要だとしても、不純度の減少がわずかでも大きければ、その特徴量が選択されてしまう。特に根に近いノードは、入力データの大部分が通過するため、木全体の計算コストの増加は非常に大きなものとなる。

そこで本手法では、特徴量を選択する際に、不純度の減少量に加えて計算コストの増加量も考慮することにより、精度と計算コストの両立を図る。さらに、計算コストを考慮した枝刈りによって、より効率的な木への変換を試みる。以降では、各学習手順の詳細を説明する。

3.1 手順1: 枝の成長

特徴量 q_k によって葉 l_i を分岐するとき、計算コストの増加量は以下のように計算できる。

$$\Delta T(l_i, q_k) = P(l_i)T(l_i, q_k) \quad (5)$$

ここで、 $T(l, q)$ は葉 l において特徴量 q の計算に要する計算コストを表し、実際の処理時間や演算数などに基づいて与える。また、計算コスト T は0.0から1.0の範囲に正規化しておく。式(5)では計算コスト T と、ノード

† NHK 放送技術研究所

l_i にデータが到達する確率 $P(l_i)$ を乗ずることによって計算コストの増加量の期待値を算出している。

特徴量を選択するための評価値は、次式で定義する。式中の $\alpha (0 \leq \alpha \leq 1)$ は、計算コストをどの程度、考慮するかを調整するパラメータである。

$$\delta(l_i, q_k) = \Delta H(D|l_i, q_k) - \log \Delta T^\alpha(l_i, q_k) \quad (6)$$

提案手法では、 $\delta(l_i, q_k)$ が最大となる \hat{l}_i と \hat{q}_k を選択して木を成長させていく。

3.2 手順 2: 枝刈り

次に枝刈りを実施する。本手法における枝刈りの考え方は、あるノードに対して、子孫ノードにおける計算コストの総和と不純度の減少量の総和を基準として、それと同程度の性能が得られる別の特徴量が見つければ差し替えるというものである。大局的な視点から特徴量を選択することで、より効率的な木を構成することが可能となる。提案手法では、この処理を子孫ノードの総数が多いノードから順に実施していく。

具体的に、ノード l_i の子孫ノードによる計算量の増加の総和は次式で算出できる。式中の l_n は l_i の子孫ノード (l_i 自身も含む) を表す。

$$\Delta T^*(l_i) = \sum_{l_n \in \text{descendant}(l_i)} P(l_n) T(l_n, q_n) \quad (7)$$

また、 l_i の子孫ノードによる不純度は次のように計算できる。 l_t は l_i の子孫ノードのうち葉ノードを表す。

$$H^*(D|l_i) = \sum_{l_t \in \text{leaf}(l_i)} P(l_t) H(D|l_t) \quad (8)$$

よって、子孫ノードによる決定木全体の不純度の減少量は、以下のように計算できる。

$$\Delta H^*(D|l_i) = H(D|l_i) - H^*(D|l_i) \quad (9)$$

枝刈りの条件を示す。式 (10) を満たす特徴量 \hat{q}_k が見つければノードを置換する。 $\varepsilon_1, \varepsilon_2$ は閾値を表す。

$$\begin{aligned} \Delta T^*(l_i) - \Delta T(l_i, \hat{q}_k) &> \varepsilon_1 \quad \text{and} \\ \Delta H(D|l_i, \hat{q}_k) - \Delta H^*(D|l_i) &> \varepsilon_2 \end{aligned} \quad (10)$$

4 決定木に基づくショット境界検出

決定木に基づくショット境界検出器について説明する。提案手法では、フレーム間の類似度を特徴量としてショット境界の判定を試みる。これは、同一のショットに含まれるフレーム同士は類似性が高く、その間にショット境界が含まれる場合には類似性が低くなるという特性に基づく。各フレーム画像 f_i に対する特徴量として、以下に示す 5 種類を利用する。これらの中から、決定木の各ノードにおける特徴量を選択していく。

- (a) 画素値の絶対差分和 $D_{sad}(f_i, f_{i-k})$
- (b) 色ヒストグラム差分 $D_{hist}(f_i, f_{i-k})$
- (c) エッジ方向ヒストグラム差分 $D_{edge}(f_i, f_{i-k})$
- (d) 余弦類似度 $D_{cos}(f_i, f_{i-k})$
- (e) ブロックマッチング差分 $D_{bm}(f_i, f_{i-k})$

表 1 ショット境界 (カット) の検出結果

α	再現率	適合率	処理時間
0.00	97%(1021/1052)	99%(1021/1032)	462 秒
0.25	98%(1026/1052)	97%(1026/1057)	255 秒

ここで、 f_i は入力映像における i 番目のフレーム画像を表す。本論文では、各特長について $k = 1, \dots, 5$ の 5 種類を利用した。隣接フレーム間の比較のみではなく、離れたフレームも考慮することによって、フラッシュなどの瞬間的な激しい変化による誤検出の軽減を図る。

なお、これらの特徴量は、瞬時にショットを切り替えるカットの検出を目的としたものである。複数フレームに渡るディゾルブなどに対応する場合には、それらを識別できる他の特徴量 [1] を追加する必要がある。

5 実験結果

提案手法の有効性を確認するため実験を実施した。実験には、ドラマ、スポーツ、ドキュメンタリなど合計で約 180 分のテレビ番組を使用し、半分を学習データ、半分をテストデータとして用いた。識別精度のみを考慮して学習した場合 ($\alpha = 0.00$) と、計算コストも考慮した場合 ($\alpha = 0.25$) について、カットの検出精度と計算コストを比較した。実験結果を表 1 に示す。表では、比較を簡単にするため枝刈り前の結果を示している。また、表中の処理時間は、映像のデコード処理を除いた計算時間を表す。実験の結果、計算コストを考慮した場合は、精度のみを考慮した場合と比較して、同程度の検出精度を維持したまま、計算時間を約半分に削減することができた。

6 あとがき

本稿では、決定木学習を利用した高速かつ高精度なショット境界検出手法を提案した。提案手法では、データの分類精度に加え、特徴量の算出に要する計算コストも考慮した新たな学習アルゴリズムを提案し、検出速度と精度の両立を図った。また、決定木の子孫ノードにおける計算コストの総和と不純度の減少量の総和を基準とした枝刈りの手法も提案した。実験では、分類精度のみに基づく通常の決定木と比較して、同程度の検出精度を維持したまま、計算コストを約半分に削減することができた。今後は、枝刈りの有効性について検証したい。また、ディゾルブなど複数フレームに渡るショット境界に対しても特徴量を追加した上で実験を実施したい。

参考文献

- [1] 河合, 住吉, 八木, “逐次的な特徴算出によるディゾルブ, フェードを含むショット境界の高速検出手法”, 信学論 (D), Vol. J91-D, No.10, pp.2529–2539, 2008.
- [2] Quinlan, “Induction of Decision Trees”, Machine Learning, Vol.1, pp.81–106, 1986.
- [3] J.R.Quinlan, “C4.5 Programs for Machine Learning”, Morgan Kaufmann, 1993.
- [4] L.Breiman, J.Friedman, R.A.Olshen and C.J.Stone, “Classification and Regression Trees”, Wadsworth, 1984.