

Wekaにおける属性選択

2013年1月21日

阿部 秀尚

属性選択(Attribute Selection)とは

- ▶ 分類にとって適切な関係にある(relevant)属性を選択すること
- ▶ Feature selectionとも呼ばれる

- ▶ アプローチは2通り
 - ▶ フィルターアプローチ
 - ▶ 属性毎に評価値を算出して、閾値以上の属性を残す
 - ▶ ラッパーアプローチ
 1. サンプルした訓練データに対して、実際に分類学習アルゴリズムを適用する
 2. 別の部分属性集合に対して、1.を繰り返す
 3. 分類精度(あるいはその期待値)が最高となる部分集合を得る



Wekaで属性選択を行うために

- ▶ 結果自体を確認→「属性選択」タブ
- ▶ 所与のデータについて属性選択処理を実行→「前処理」タブの”AttributeSelection”フィルターを利用
 - ▶ “選択” → ”filters” → ”supervised” → ”AttributeSelection”
- ▶ 属性選択の各アプローチの実行 (Attribute Selectionフィルタの設定でも共通)
 - ▶ フィルターアプローチ
 - ▶ 属性検証(evaluator): *AttributeEvalとつくEvaluator
 - ▶ 検索方法(search): Ranker
 - ▶ ラッパーアプローチ
 - ▶ 属性検証(evaluator): *SubsetEvalとつくEvaluator
 - ▶ 検索方法(search): Ranker以外の探索方法を指定



代表的な属性選択アルゴリズムとWekaでの組み合わせ

▶ Relief.F

- ▶ 属性検証(evaluator): ReliefFAttributeEval
- ▶ 検索方法(search): Ranker

▶ ラッパー(決定木など)による属性選択

- ▶ 属性検証(evaluator): WrapperSubsetEval
 - Classifierの“選択”でJ4.8などを指定
- ▶ 検索方法(search): BestFirst, Greedy/Stepwise (classifierの正解率に基づくヒューリスティック探索), ExhaustiveSearch (全組み合わせ探索)

▶ 遺伝的アルゴリズムによる属性選択

- ▶ 属性検証(evaluator): WrapperSubsetEval
 - Classifierの“選択”でJ4.8などを指定
- ▶ 検索方法(search): GeneticSearch



AttributeSelectionフィルターの実行(1)

例: ラッパー+最優先探索(バックトラック付)による属性選択

前処理 分類 クラス

※クラス(属性)が最も右の列になるようファイルを準備しておくこと

ファイルを開く URLを開く DBを開く

フィルター

weka

- filters
 - AllFilter
 - MultiFilter
 - supervised
 - attribute
 - AddClassification
 - AttributeSelection
 - ClassOrder
 - Discretize

正規表

weka.gui.GenericObjectEditor

weka.filters.supervised.attribute.AttributeSelection

ケパビリティについて

A supervised attribute filter that can be used to select attributes.

evaluator 選択 WrapperSubsetEval -B weka.classifiers.rules.ZeroR -F 5 -T

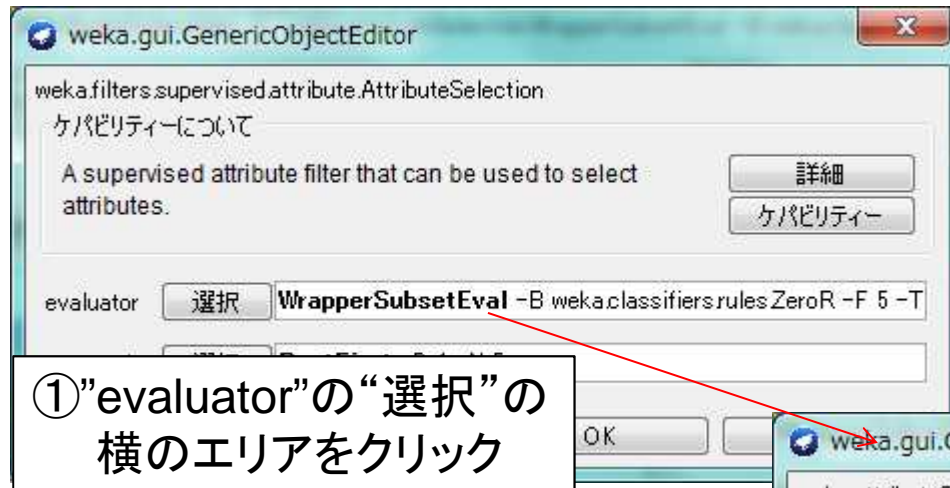
search 選択 BestFirst -D 1 -N 5

開く... 保存... OK キャンセル

*evaluateに対する「検証」とsearchに対する「検索」は誤り

AttributeSelectionフィルターの実行(2)

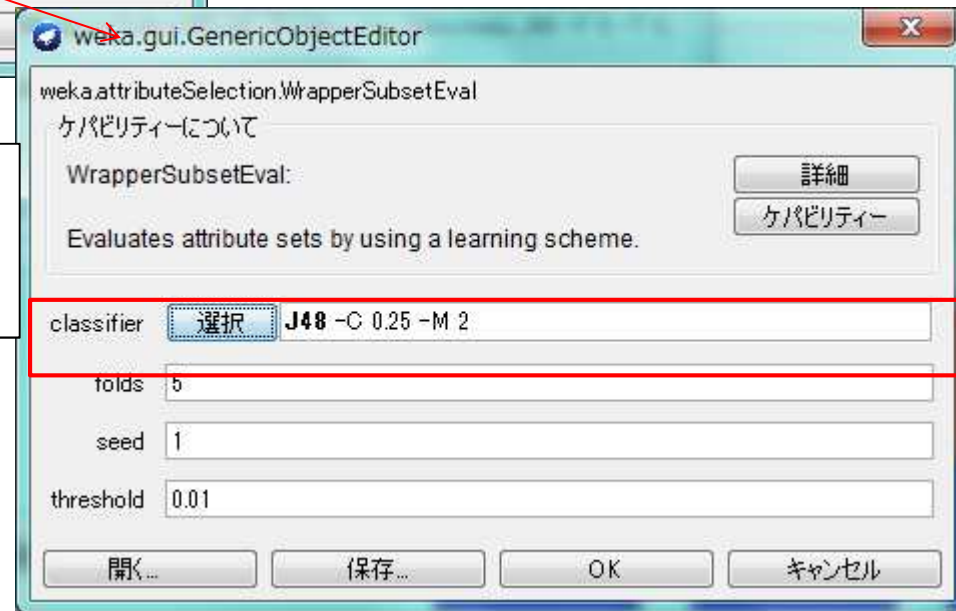
例: ラッパー+最優先探索(バックトラック付)による属性選択



① "evaluator" の "選択" の横のエリアをクリック

② "classifier" の横の "選択" をクリックして、分類学習アルゴリズムを指定 (例では、決定木J4.8)

③ "OK" を押して、「前処理」パネルに戻ったら、「適応」* のボタンを押す



*Applyに対する「適応」は誤り→「適用」

AttributeSelectionフィルターの実行(3)

Weka Explorer

前処理 分類 クラスタ アンソエート 属性選択 ビス

選択した結果は保存ボタンを押してファイルに保存

ファイルを開く URLを開く DBを開く 生成 元に戻す 編集 保存

フィルター
選択 AttributeSelection -E "weka.attributeSelection WrapperSubsetEval -B weka.classifiers.trees.J48 -F 5 -T 0.01 -R 1 -- -C 0. 適応

現在のデータ
データ名: example1-weka.filters.supervised.attribute.AttributeSel...
インスタンス数: 30 属性数: 2

属性
全選択 非選択 内容反転 正規表現

番号	名称
1	<input checked="" type="checkbox"/> GainRatio
2	<input type="checkbox"/> GainRatioCat

選択属性
属性名 : GainRatio 属性型 : 数値型
欠損数 : 0 (0%) 識別数 : 30 一意数 : 30 (100%)

統計値	値
最小値	-2.14
最大値	60.27
平均値	10.966
標準偏差値	13.542

cat (Nom) ビジュアル化

フィルタで実行された属性選択アルゴリズムが
選択した属性が残される

属性	値
GainRatio	4
GainRatioCat	2
その他	1