Family of Relief algorithms

Alicja Gosiewska

Warsaw University of Technology

22.10.2018

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of relieff and rrelieff. Machine Learning 53, 23–69.

Introduction

Relief is an algorithm for estimation of features quality. Feature scoring is based on the identification of feature value differences between nearest neighbor instance pairs.

Family of Relief algorithms have commonly been viewed as a:

- feature subset selection methods,
- feature weighting method,
- method for selecting splits in the building phase of decision tree learning.

Family of Relief algorithms

Relief (Kira & Rendell, 1992) is limited to classification problems with two classes. **ReliefF** (Kononenko, 1994) can

deal with multiclass problems. It is more robust and also able to deal with incomplete and noisy data.

RReliefF (Robnik-Sikonja & Kononenko, 1997) is adapted for continuous regression problems.

Contents [hide]

If a feature value difference is observed in a neighboring instance pair with the same class (a 'hit'), the feature score decreases. Alternatively, if a feature value difference is observed in a neighboring instance pair with different class values (a 'miss'), the feature score increases.

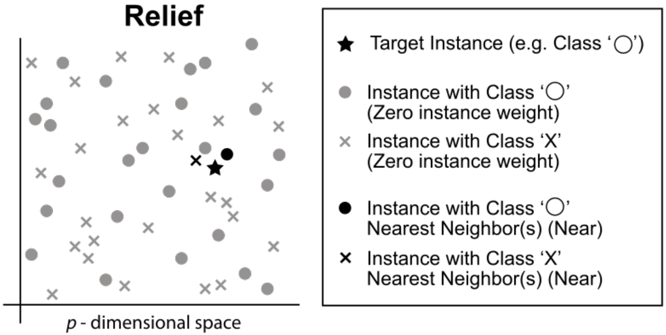


Figure 1: Illustration of Relief neighbor selection for scoring. Source: Relief Wiki.pdf Author: Docurbs

Notation

We assume that examples $I_1, I_2, ..., I_n$ in the instance space are described by a vector of attributes $A_i, i = 1, ..., a$, where a is the number of explanatory attributes and they are labelled with the target value τ_j.

$$\begin{array}{cc} & \begin{array}{cccc} A_1 & A_2 & \ldots & A_a \end{array} \\ \begin{array}{c} I_1 \\ I_2 \\ \vdots \\ I_n \end{array} & \left[\begin{array}{cccc} & & & \\ & & & \\ & & & \\ & & & \end{array}\right] \end{array} \qquad \begin{array}{c} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_n \end{array} \left[\begin{array}{c} \\ \\ \\ \\ \end{array}\right]$$

Difference function

Function $diff(A, I_1, I_2)$ calculates the difference between the values of the attribute A for two instances I_1 and I_2.
For nominal attributes it is defined as:

$$diff(A, I_1, I_2) = \begin{cases} 0, & value(A, I_1) = value(A, I_2) \\ 1, & \text{otherwise.} \end{cases}$$

and for numerical attributes as:

$$diff(A, I_1, I_2) = \frac{|value(A, I_1) - value(A, I_2)|}{max(A) - min(A)}$$

The function diff is used also for calculating the distance between instances to find nearest neighbours. The total distance is the sum of distances over all attributes.

Relief Algorithm

Algorithm Relief

Input: for each training instance a vector of attribute values and the class value

Output: the vector W of estimations of the qualities of attributes

1. set all weights $W[A] := 0.0$;
2. **for** i := 1 **to** m **do begin**
3. randomly select an instance R_i;
4. find nearest hit H and nearest miss M;
5. **for** $A := 1$ **to** a **do**
6. $W[A] := W[A] - \text{diff}(A, R_i, H)/m + \text{diff}(A, R_i, M)/m$;
7. **end**;

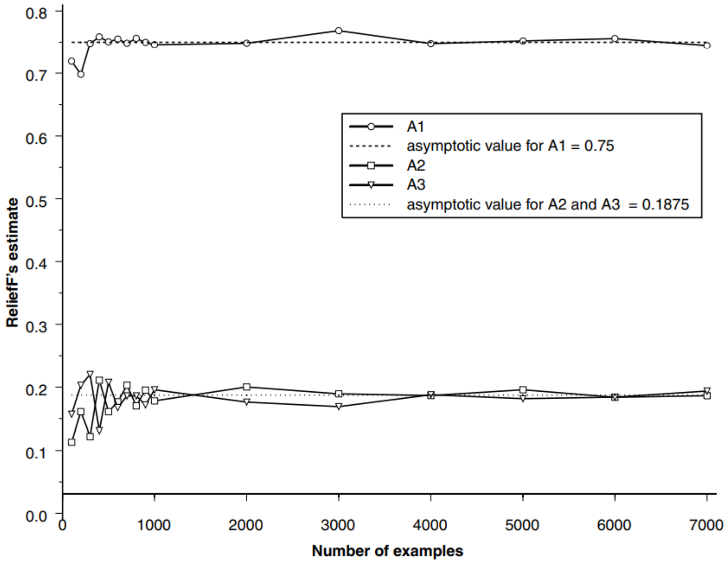
Table 1. Tabular description of the concept $\tau = (A_1 \wedge A_2) \vee (A_1 \wedge A_3)$ and the responsibility of the attributes for the change of the predicted value.

Line	A_1	A_2	A_3	τ	Responsible attributes
1	1	1	1	1	A_1
2	1	1	0	1	A_1 or A_2
3	1	0	1	1	A_1 or A_3
4	1	0	0	0	A_2 or A_3
5	0	1	1	0	A_1
6	0	1	0	0	A_1
7	0	0	1	0	A_1
8	0	0	0	0	(A_1, A_2) or (A_1, A_3)

$$A_1 = \frac{4 + 2 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2}}{8} = \frac{6}{8} = 0.75$$

$$A_2 = \frac{2 \cdot \frac{1}{2} + \frac{1}{2}}{8} = \frac{3}{16} = 0.1875$$

Relief Algorithm



The original Relief can deal with nominal and numerical attributes.
However, it is limited to two class problems and cannot deal with incomplete data.

ReliefF

Algorithm ReliefF
Input: for each training instance a vector of attribute values and the class value
Output: the vector W of estimations of the qualities of attributes

set all weights $W[A] := 0.0$;
for i := 1 **to** m **do begin**
    randomly select an instance R_i;
    find k nearest hits H_j;
    for each class $C \neq class(R_i)$ **do**
        from class C find k nearest misses $M_j(C)$;
for A := 1 **to** a **do**

$$\boxed{\begin{aligned} W[A] &:= W[A] - \sum_{j=1}^{k} \text{diff}(A, R_i, H_j)/(m \cdot k) + \\ &\sum_{C \neq class(R_i)} \Big[\frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^{k} \text{diff}(A, R_i, M_j(C))\Big]/(m \cdot k); \end{aligned}}$$

end;

Missing values

To deal with incomplete data we change the diff function. Missing values of attributes are treated probabilistically.
If one instance (e.g. I1) has unknown value:

$$diff(A, I_1, I_2) = 1 - P(value(A, I_2) | class(I_1))$$

If both have unknown value:

$$diff(A, I_1, I_2) = 1 - \sum_{V}^{\#values(A)} [P(V | class(I_1)) P(V | class(I_2))]$$

RReliefF - in regression

Relief's estimate $W[A]$ of the quality of attribute A is the estimation of the following difference of probabilities:

$$W[A] = P(\text{diff. value of A}|\text{nearest inst. from diff. class}) - P(\text{diff. value of A}|\text{nearest inst. from same class})$$

If we rewrite

$$P_{diffA} = P(\text{different value of A}|\text{nearest instances})$$

$$P_{diffC} = P(\text{different prediction}|\text{nearest instances})$$

$$P_{diffC|diffA} = P(\text{diff. prediction}|\text{diff.value of A and nearest instances})$$

We obtain (Bayes rule):

$$W[A] = \frac{P_{diffC|diffA}P_{diffA}}{P_{diffC}} - \frac{(1 - P_{diffC|diffA})P_{diffA}}{1 - P_{diffC}}$$

RReliefF

Algorithm RReliefF
Input: for each training instance a vector of attribute values $\mathbf{x}$ and predicted value $\tau(\mathbf{x})$
Output: vector W of estimations of the qualities of attributes

set all N_{dC}, $N_{dA}[A]$, $N_{dC\&dA}[A]$, $W[A]$ to 0;
for i := 1 **to** m **do begin**
  randomly select instance R_i;
  select **k** instances I_j nearest to R_i;
  for j := 1 **to** k **do begin**
    $N_{dC} := N_{dC} + \mathrm{diff}(\tau(\cdot), R_i, I_j) \cdot d(i, j)$; P_{diffC}
    for A := 1 **to** a **do begin**
      $N_{dA}[A] := N_{dA}[A] + \mathrm{diff}(A, R_i, I_j) \cdot d(i, j)$; P_{diffA}
      $N_{dC\&dA}[A] := N_{dC\&dA}[A] + \mathrm{diff}(\tau(\cdot), R_i, I_j) \cdot$ $P_{diffC|diffA}$
        $\mathrm{diff}(A, R_i, I_j) \cdot d(i, j)$;
    end;
  end;
end;
for A := 1 **to** a **do**
  $W[A] := N_{dC\&dA}[A]/N_{dC}$ - $(N_{dA}[A] - N_{dC\&dA}[A])/(m - N_{dC})$;

Applications

- Feature selection
- Feature weighting
- Building tree based models
- Discretization of attributes
- Use in ILP and with association rules

Bibliography

- Kira, K., & Rendell, L. A. (1992b). A practical approach to feature selection. In D. Sleeman, & P. Edwards (Eds.), Machine Learning: Proceedings of International Conference (ICML'92) (pp. 249–256). Morgan Kaufman.
- Kononenko, I. (1994). Estimating attributes: Analysis and extensions of Relief. In L. De Raedt, & F. Bergadano (Eds.), Machine Learning: ECML-94 (pp. 171–182). Springer Verlag
- Robnik Šikonja, M., & Kononenko, I. (1997). An adaptation of relief for attribute estimation in regression. In D. H. Fisher (Ed.), Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97) (pp. 296–304). Morgan Kaufman.
- Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of relieff and rrelieff. Machine Learning 53, 23–69.