Theoretical basis of the rule analysis methods implemented in the RuleXAI package

Rule definition

The RuleXAI package operates on rule-based representation and is applicable to classification, regression and survival tasks. Each rule r forming a rule representation has the form:
IF w_1 and w_2 and . . . and w_n THEN *decision*
The premise of a rule is a conjunction of elementary conditions $w_i \equiv a_i \odot x_i$, with x_i, being an element of the a_j domain and $\odot$ representing a relation (= for symbolic attributes; $<$, $\leq$, $>$, $\geq$ for ordinal and numerical ones).

In a classification rule the decision part has the form *decision* = v. The value $v \in V_d$ points to one of decision classes (concepts) under which the examples are classified. The meaning of the rule is as follows: if an example fulfils all conditions specified in the conditional part of the rule, then it belongs to the decision class specified in the rule conclusion. An example satisfying the conditions specified in the rule premise is stated to be covered by the rule. The examples whose labels are the same as the conclusion of r are called positive examples, while the others are called negative examples.

In a regression rule the decision attribute is real-valued. Therefore, the decision part of the rule has the form $decision = v, v \in \boldsymbol{R}$. The value v can be calculated on the basis of some function f (e.g. linear), whose values depend on the set of examples covering the rule r (and denoted further as $[r]$) and values of conditional attributes (e.g. f can be a linear combination of conditional attribute values). During the regression rule induction the formula of f is fixed, but its parameters must be determined in the rule induction process (for example, see [1]). This increases the computational complexity of the rule induction process. However, it turns out that good prediction results can be obtained with a simple form of a regression rule r with the conclusion $v \in \boldsymbol{R}$, calculated as an average or a median of decision values of examples from $[r]$ [2,3]. This form of a rule is used in RuleXAI because it can be very easily interpreted by a human. Interpretation of a regression rule is similar to that of classification rule: the decision points to the value of the decision attribute for the examples covering the rule.

In survival rules the decision part has the form *decision* = $S([r])$. The symbol $S([r])$ means the Kaplan-Meier estimate calculated on the basis of examples from $[r]$. The survival rule meaning is as follows: if an example fulfils all constraints specified in the conditional part of the rule, then for its survival time estimation the Kaplan-Meier estimate from the rule conclusion is used.

Rule evaluation

To induce a rule, it is necessary to define four values: *p*, *n*, *P*, *N*. For a classification rule *r*, *P*= |Pos(*r*)|, where Pos(*r*) is a set of all training examples whose decisions are equal to the rule decision part. The value of *N* is calculated similarly: *N*=|Neg(*r*)|, where Neg(*r*) is a set of all remaining training examples (that do not belong to Pos(*r*)). The value of *p* is the number of examples from the positive decision class satisfying the conditions in rule premise (i.e. positive examples covered by the induced rule), *n* is the number of examples from the negative class satisfying the conditions in rule premise (i.e. negative examples covered by the induced rule), which can be formally stated as follows:

$p = |\mathrm{Pos}(r) \cap [r]|$

$n = |\mathrm{Neg}(r) \cap [r]|$

In case of regression rules the continuous form of the decision variable should be considered. Let us consider a regression rule $r \equiv \varphi \rightarrow v, v \in \boldsymbol{R}$. The set Pos(*r*) is the set of those training examples whose decision attribute values belong to $[v-\delta, v+\delta]$, where δ is a standard deviation of decision attribute values of all examples from $[r]$ (i.e. examples covering the rule *r*). The Neg(*r*) set contains all training examples that do not belong to Pos(*r*). With these assumptions, values of *p* and *n* are defined by analogy as in the case of the classification rule.

Two measures are typically used in rule-based exploratory analysis:

$precision = \frac{p}{p+n}$,

$coverage = \frac{p}{P}$.

Other measures may also be used, in particular in RuleXAI the *C2* measure was used as the measure of rule quality in rule induction and elementary condition importance evaluation. The *C2* measure uses *p*, *n*, *P*, *N* values defined above and it has the following form:

$$C2 = \left(\frac{Np - Pn}{N(p+n)}\right)\left(\frac{P+p}{2p}\right) . \quad (1)$$

The *C2* measure can be viewed as the weighted version of Kappa statistics. During the condition evaluation, the quality of the rule – with and without the evaluated condition – is calculated.

Importance of elementary conditions

The importance of elementary conditions is considered from the point of view of the precision of rules containing them. In the classical approach, the analysis considers the precision of rules containing the evaluated conditions and rules not containing them. Validity is not assessed from the point of view of statistical significance, therefore the term condition significance is not used but importance. In the definition of indicators designed to assess the importance of elementary conditions, a key role is played by the Shapley index [4] which in game theory is used to assess the strength of players and coalitions. This index – after some modifications – can be used to create indices assessing the importance of sets of elementary

conditions. Positive values of these indices mean that the evaluated conditions increase the accuracy of the rules, while negative values mean that the evaluated conditions are (at least partially) redundant.

Let us assume that a decision rule r and a set of elementary conditions $Ec(r) = \{w_1, w_2, \dots, w_m\}$ are given. In the standard form, the values of the Shapley index (2) are calculated based on information about the average impact of the evaluated elementary condition on the precision of rules that are all possible generalizations of the rule *r*:

$$\phi_S(w,r) = \sum_{Y \subseteq Ec(r)-\{w\}} \frac{(m-|Y|-1)!\,|Y|!}{m!} [precision(Y \cup \{w\}, r) - precision(Y,r)] \qquad (2)$$

In formula (2), *precision*(*Y*,*r*) denotes the precision of the rule *r*, whose premise is built only from the conditions contained in the set *Y*. In addition, the following assumption is made: $precision(\emptyset, r) = 0,\ precision(Ec(r), r) = precision(r)$.

Let us denote by RUL_X the set of all rules pointing to the decision class *X*. The validity of an elementary condition in the set RUL_X is calculated based on the validity of this condition in all rules belonging to RUL_X, and also based on the validity of this condition in all rules pointing to a decision class other than *X*. The methods implemented in RuleXAI are based on an approach in which the evaluation of the importance of the condition *w* for the decision class *X* is expressed by formula (3):

$$G(w, RUL_X) = \sum_{r \in RUL_X} (\phi_S(w,r)\ \ coverage(r)) \qquad (3)$$

According to (3), in the global evaluation of an elementary condition, its contribution to the precision of each rule containing it and all generalizations of such a rule is taken into account, and the coverage of the rules containing the evaluated condition is also taken into account.

Coverage rule induction algorithms perform an ongoing evaluation of elementary conditions in the growth and pruning phases. This evaluation is usually done by means of a quality measure. A natural generalization of formulas (2-3) is therefore to replace the *precision* measure with a quality measure, in this case the *C2* measure. This change means that each generalization of *r* will examine the effect of a given condition not only on the precision of the rule, but also on its coverage. Formula (2) then takes the form (4), and the evaluation of the elementary condition in the rule set is expressed by formula (5):

$$\phi_{C2S}(w,r) = \sum_{Y \subseteq Ec(r)-\{w\}} \frac{(m-|Y|-1)!\,|Y|!}{m!} [C2(Y \cup \{w\}, r) - C2(Y,r)] \qquad (4)$$

$$G_{C2}(w, RUL_X) = \sum_{r \in RUL_X} \phi_{C2S}(w,r) \qquad (5)$$

In formula (5), the evaluation of the coverage of rules containing the elementary condition being evaluated has been removed, as it is already performed during the determination of the value of $\phi_{C2S}(w,r)$.

Determining the values of the index (2) or (4) for the rule consisting of elementary conditions, on the whole set of examples we have to determine the value of the *precision* measure or *C2* quality measure 2^{m-1} times, because this is how many generalizations of the rule *r* we get. When the number of elementary conditions is large or the set of examples on the basis of

which the rule quality is calculated is large, such an operation will be time-consuming. In order to reduce the time of calculations connected with determining the values of indices, their simplified forms can be presented, consisting only of those components of the sums occurring in formulas (2) and (4), which contribute most to the assessment of the validity of an elementary condition. The simplified index for elementary condition importance evaluation (inspired by the Shapley index) assumes that the most information about the validity of the elementary condition is contributed by:

- base rules, i.e. those containing the elementary condition being evaluated,
- base rules from which only the elementary condition being evaluated has been removed,
- rules whose premises contain only the elementary condition being evaluated.

This simplified index is expressed by formulas (6-7):

$$\phi_{sS}(w,r) = \frac{1}{m}[precision(r) - precision(Ec(r) - \{w\}, r) + precision(\{w\}, r)] \quad (6)$$

$$\phi_{sC2S}(w,r) = \frac{1}{m}[C2(r) - C2(Ec(r) - \{w\}, r) + C2(\{w\}, r)] \quad (7)$$

It is worth noting that in the case of basic (2) and modified Shapley (4) indices the greatest weight is assigned to those components that appear in the formulas defining simplified forms of the index (6-7).

It cannot be proved that the orders of the elementary conditions formed by indices (2) and (6) and (4), (7) will be identical. However, one can empirically test whether and to what extent these orders are correlated. The correlation of the ordering of the elementary conditions for the original and simplified indices is shown in Tables 1 to 3 in the Appendix A below. For details on the assessment of the validity of conditions, see the work [5].

In the RuleXAI package the simplified index expressed by formulas (5) and (7) is used to evaluate the elementary conditions for classification tasks. In the case of regression rules, the elementary conditions are evaluated similarly, using the C2 measure, which is determined taking into account the continuous form of the decision variable. In the case of survival rules, the log-rank test was used to assess the elementary condition importance [6].

In order to validate the approach provided by RuleXAI, a comparison of it and SHAP was performed. Such an analysis is reasonable when comparing the attribute rankings generated by both methods (SHAP does not automatically generate the attribute value ranges (rule elementary conditions) as it is done by RuleXAI). SHAP was chosen as a reference method because it is widely recognised and used throughout the community. A Python implementation of SHAP was used in the experiments (https://shap-lrjball.readthedocs.io/en/docs_update/generated/shap.TreeExplainer.html). The analysis carried out included training the decision tree model, generating the importance of attributes by each method and comparing the three most important attributes selected by each method. The results of the analysis for the classification and regression tasks are presented in Appendix B in Tables 4 and 5 respectively.

References

[1] Quinlan JR. Learning with continuous classes. In: Proceedings of the Australian Joint Conference on Artificial Intelligence. Singapore: World Scientific; 1992. p. 343–348.

[2] Fürnkranz J. Separate-and-conquer rule learning. Artificial Intelligence Review. 1999;13(1):3–54. https://doi.org/10.1023/A:1006524209794.

[3] Sikora M, Skowron A, Wróbel Ł. In: Ramsay A, Agre G, editors. Rule Quality Measure-Based Induction of Unordered Sets of Regression Rules. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 162–171. https://doi.org/10.1007/978-3-642-33185-5_18.

[4] Sikora M. Selected methods for evaluation and pruning of decision rules (Wybrane metody oceny i przycinania reguł decyzyjnych). Studia Informatica, 2012;33(3B):5-331.

[5] Sikora M. Redefinition of decision rules based on the importance of elementary conditions evaluation. Fundamenta Informaticae, 2013;123(2):171-197. https://doi.org/10.3233/FI-2013-806

[6] Wróbel Ł, Gudyś A, Sikora M. Learning rule sets from survival data. BMC Bioinformatics 2017;18:285. https://doi.org/10.1186/s12859-017-1693-x

Appendix A

	Data set	Correlation
1	autos	0.854
2	balance-scale	0.918
3	breast-cancer	1.000
4	breast-w	0.835
5	car	0.956
6	contact-lenses	1.000
7	credit-a	0.877
8	echocardiogram	0.821
9	ecoli	0.813
10	flag	0.853
11	glass	0.867
12	hayes-roth	1.000
13	heart-c	0.901
14	heart-statlog	0.937
15	hepatitis	0.912
16	horse-colic	0.875
17	hungarian-heart-disease	0.894
18	hypothyroid	0.917
19	ionosphere	0.811
20	iris	0.944
21	kdd-synthetic-control	1.000
22	kr-vs-kp	0.723
23	labor	1.000
24	lymph	0.939
25	mammographic-masses	0.983
26	mushroom	0.912
27	primary-tumor	1.000
28	prnn-synth	1.000
29	sonar	0.830
30	soybean	0.867
31	splice	1.000
32	tic-tac-toe	0.824
33	titanic	1.000
34	vote	1.000
35	wine	1.000
36	zoo	1.000
	Mean	0.918
	Std. dev.	0.074

Table 1. Correlation of the ordering of the elementary conditions for the original and simplified indices – application to classification task on 36 data sets.

	Data set	Correlation
1	auto-mpg	0.841
2	auto-price	0.852
3	auto93	0.796
4	bodyfat	0.825
5	bolts	0.905
6	boston-housing	0.740
7	breasttumor	0.801
8	cholesterol	0.769
9	cloud	1.000
10	concrete	0.782
11	cpu	0.885
12	diabetes	1.000
13	echomonths	0.923
14	ele-1	1.000
15	ele-2	1.000
16	elusage	1.000
17	fishcatch	0.916
18	fruitfly	0.937
19	gascons	0.786
20	housing	0.773
21	kidney	1.000
22	laser	0.874
23	lowbwt	0.824
24	machine	0.770
25	mbagrade	1.000
26	meta	0.820
27	methane	0.863
28	mortgage	0.767
29	pharynx	0.876
30	pollution	0.877
31	pwlinear	0.736
32	pyrim	0.835
33	servo	1.000
34	triazines	0.813
35	veteran	0.751
	Mean	0.867
	Std. dev.	0.088

Table 2. Correlation of the ordering of the elementary conditions for the original and simplified indices – application to regression task on 35 data sets.

	Data set	Correlation
1	actg320	0.817
2	BHS	1.000

3	biecek-o	0.810
4	BMT-ch	0.911
5	cancer	0.758
6	cost	0.835
7	DLBCL	0.818
8	echomonths	0.824
9	follic	1.000
10	GBSG2	0.886
11	grace1000	1.000
12	halibut	0.709
13	hd	1.000
14	kidney	1.000
15	lung	1.000
16	Melanoma	0.912
17	mgus	0.939
18	nursing	0.882
19	pbc	1.000
20	pharynx	0.944
21	Rossi	0.944
22	stanford	1.000
23	sTRACE	1.000
24	toy	0.752
25	uis	1.000
26	unemployed	0.933
27	veteran	1.000
28	wcgs	0.767
29	whas1	1.000
30	whas500	1.000
	Mean	0.915
	Std. dev.	0.093

Table 3. Correlation of the ordering of the elementary conditions for the original and simplified indices – application to survival task on 30 data sets.

Appendix B

	Data set	**Number of common features present in the top 3 ranking positions**
1	autos	2.17
2	balance-scale	2.00
3	breast-cancer	2.00
4	breast-w	2.50
5	car	2.00
6	contact-lenses	2.00
7	credit-a	1.50
8	echocardiogram	2.50
9	ecoli	2.63
10	flag	2.25
11	glass	2.33
12	hayes-roth	3.00
13	heart-c	2.50
14	heart-statlog	2.50
15	hepatitis	1.50
16	horse-colic	2.00
17	hungarian-heart-disease	2.00
18	hypothyroid	2.00
19	ionosphere	2.50
20	iris	3.00
21	kdd-synthetic-control	1.83
22	kr-vs-kp	2.00
23	labor	2.00
24	lymph	2.00
25	mammographic-masses	3.00
26	mushroom	1.00
27	primary-tumor	0.62
28	prnn-synth	3.00
29	sonar	1.50
30	soybean	0.89
31	splice	2.33
32	tic-tac-toe	2.00
33	titanic	2.50
34	vote	2.00
35	wine	2.33
36	zoo	2.57
	mean value	2.124

Table 4. Comparison of the attribute importance rankings generated by the RuleXAI and SHAP methods for classification data sets.

	Data set	**Number of common features present in the top 3 ranking positions**
1	auto-mpg	3
2	auto-price	2
3	auto93	2
4	bodyfat	2
5	bolts	2
6	boston-housing	3
7	breasttumor	3
8	cholesterol	3
9	cloud	3
10	concrete	3
11	cpu	2
12	diabetes	3
13	echomonths	2
14	ele-1	3
15	ele-2	3
16	elusage	3
17	fishcatch	2
18	fruitfly	2
19	gascons	3
20	housing	3
21	kidney	3
22	laser	2
23	lowbwt	3
24	machine	3
25	mbagrade	3
26	meta	1
27	methane	3
28	mortgage	3
29	pharynx	3
30	pollution	3
31	pwlinear	2
32	pyrim	2
33	servo	3
34	triazines	1
35	veteran	3
	mean value	2.571

Table 5. Comparison of the attribute importance rankings generated by the RuleXAI and SHAP methods for regression data sets.