Context-dependent feature analysis with random forests
Supplementary materials

Antonio Sutera[1] **Gilles Louppe**[2] **Vân Anh Huynh-Thu**[1] **Louis Wehenkel**[1] **Pierre Geurts**[1]
[1] Dept. of EE & CS, University of Liège, Belgium [2] New York University, USA

A Details of Example 1

Table 1: Values of X_1, X_2, X_c and Y.

X_1	X_2	X_c	Y
0	0	0	0
0	0	0	0
0	0	1	2
0	0	1	3
0	1	0	2
0	1	0	3
0	1	1	0
0	1	1	0
1	0	0	1
1	0	0	1
1	0	1	2
1	0	1	3
1	1	0	2
1	1	0	3
1	1	1	1
1	1	1	1

B Proof of Theorem 1

Theorem. *X_c is irrelevant to Y with respect to V iff all variables in V are context-independent to Y with respect to X_c (and V) and $I(Y;X_c) = 0$.*

Necessary condition.

Proof. If X_c is irrelevant to Y w.r.t. V, we have, by definition, that $I(Y;X_c|B) = 0$ for all subset $B \subseteq V$. Hence, we have $I(Y;X_c) = 0$ as a special case.

A variable $X_m \in V$ is context-independent if for all $B \subseteq V^{-m}$ and for all $x_c \in \mathcal{X}_c$, $b \in \mathcal{B}$, we have

$$I(Y;X_m|B=b,X_c=x_c) - I(Y;X_m|B=b) = 0.$$

Let us proof this:

$$\begin{aligned}
&I(Y;X_m|B=b,X_c=x_c) - I(Y;X_m|B=b)\\
&= H(Y|B=b,X_c=x_c) - H(Y|X_m,B=b,X_c=x_c)\\
&\qquad \hookrightarrow -H(Y|B=b) + H(Y|X_m,B=b)\\
&= H(Y|B=b) - H(Y|X_m,B=b)\\
&\qquad \hookrightarrow -H(Y|B=b) + H(Y|X_m,B=b)\\
&= 0,
\end{aligned}$$

where $H(Y|B=b,X_c=x_c) = H(Y|B=b)$ and $H(Y|X_m,B=b,X_c=x_c) = H(Y|X_m,B=b)$ are consequences of $I(Y;X_c|B) = 0$ for all B if we assume that $p(B=b) \neq 0$ ($\forall b \in \mathcal{B}$) and $p(X_c=x_c,B=b) \neq 0$ ($\forall x_c \in \mathcal{X}_c$ and $\forall b \in \mathcal{B}$). □

Sufficient condition.

Proof. If all variables are context-independent, we have that for all $X_m \in V$, $B \subseteq V^{-m}$, $b \in \mathcal{B}$, and $x_c \in \mathcal{X}_c$:

$$I(Y;X_m|B=b,X_c=x_c) = I(Y;X_m|B=b).$$

By averaging the left- and right-hand sides of this equality over $P(B,X_c)$, we get:

$$I(Y;X_m|B,X_c) = I(Y;X_m|B).$$

From this, one can derive (Louppe et al., 2013):

$$I(Y;X_c|B,X_m) = I(Y;X_c|B).$$

Since this equality is valid for all B, including $B=\emptyset$, and all X_m, we have that for all $B' \subseteq V$, $I(Y;X_c|B')$ can be reduced to $I(Y;X_c)$, which is equal to zero by hypothesis. The variable X_c is thus irrelevant to Y with respect to V. □

C Proof of Theorem 2

Theorem. *A variable $X_m \in V$ is context-independent to Y with respect to X_c iff $Imp^{|x_c|}(X_m) = 0$ for all x_c.*

Necessary condition.

Proof. By definition of context-independence, we have

$$I(Y; X_m|B = b, X_c = x_c) - I(Y; X_m|B = b) = 0 \quad \forall B \subseteq V^{-m}, \forall x_c \in \mathcal{X}_c, \forall b \in \mathcal{B}. \tag{1}$$

Given that each term

$$|I(X_m; Y|B = b) - I(X_m; Y|B = b; X_c = x_c)|$$

of $Imp^{|x_c|}(X_m)$ (Equation (**??**)) is equal to 0, the sum is thus also equal to 0. □

Sufficient condition.

Proof. Given the definition of $Imp^{|x_c|}(X_m)$:

$$Imp^{|x_c|}(X_m) = \sum_{k=0}^{p-1} \frac{1}{C_p^k} \frac{1}{p-k} \sum_{B \in \mathcal{P}_k(V^{-m})} \sum_{b \in \mathcal{B}} P(B = b) \hookrightarrow |I(X_m; Y|B = b) - I(X_m; Y|B = b; X_c = x_c)|, \tag{2}$$

appears to be a sum of positive terms (because of the absolute value). As in Theorem 1, we assume that probabilities are non-null and therefore, we have that the only way to have the sum equal to zero is to have each term of the sum equal to 0. Hence, we have $|I(X_m; Y|B = b) - I(X_m; Y|B = b; X_c = x_c)| = 0$ for all x_c, B and b which verifies the definition of context-independence for X_m.

□

D Proof of Theorem 3

Theorem. *If* $|Imp^{x_c}(X_m)| = Imp^{|x_c|}(X_m)$ *for a context-dependent variable* X_m*, then* X_m *is context-complementary if* $Imp^{x_c}(X_m) < 0$ *and context-redundant if* $Imp^{x_c}(X_m) > 0$.

Proof. The absolute value of a sum is less than or equal the sum of the absolute value of each terms. The equality is only verified when all terms are of the same sign. Therefore, the sign of $Imp^{x_c}(X_m)$ indicates the sign of all terms and thus verify either the context-complementarity if all terms are negative or the context-redundancy if all terms are positive. □

E Results for Problem 3.

Table 2: Importances as computed analytically using asymptotic formulas. The context is defined by the binary context feature *Sex* ($Sex = 0$ denotes *female* and $Sex = 1$ denotes *male*).

		$Imp(X_m)$	$Imp(X_m \mid X_c = x_c)$		$Imp^{\lvert x_c \rvert}(X_m)$		$Imp_s^{x_c}(X_m)$	
m		-	$x_c = 0$	$x_c = 1$	$x_c = 0$	$x_c = 1$	$x_c = 0$	$x_c = 1$
0	age	0.2958	0.3386	0.2885	0.1382	0.1505	-0.0095	-0.0156
1	histologic-type	0.3522	0.1389	0.4366	0.2087	0.114	0.1988	-0.0569
2	degree-of-diffe	0.4413	0.4175	0.4208	0.1653	0.158	0.0561	0.0157
3	bone	0.2429	0.2502	0.2367	0.0933	0.0755	-0.0043	0.0165
4	bone-marrow	0.0192	0.0201	0.0148	0.0126	0.0101	0.0009	0.0041
5	lung	0.1627	0.2059	0.1370	0.1038	0.0949	-0.0259	0.0172
6	pleura	0.1485	0.1496	0.1015	0.0590	0.09	0.0313	0.0234
7	peritoneum	0.3184	0.3459	0.1979	0.0861	0.138	0.0147	0.0956
8	liver	0.2285	0.2138	0.2630	0.0786	0.1279	0.0375	-0.0602
9	brain	0.0465	0.0349	0.0548	0.0378	0.0254	0.0114	-0.0104
10	skin	0.0677	0.0362	0.0923	0.0314	0.0403	0.0252	-0.0133
11	neck	0.2215	0.0690	0.2582	0.1466	0.0692	0.1316	-0.0081
12	supraclavicular	0.1676	0.1915	0.1448	0.0845	0.067	-0.0198	0.0269
13	axillar	0.1393	0.1457	0.1068	0.0655	0.0629	-0.0067	0.0447
14	mediastinum	0.1838	0.2050	0.1716	0.1016	0.0806	-0.0059	0.0140
15	abdominal	0.2553	0.3296	0.1372	0.1346	0.1379	-0.0330	0.0898

References

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In *Advances in Neural Information Processing Systems*, pages 431–439.