Supplementary Material for:

Échantillonnage actif pour la découverte de règles classification via des comparaisons par paires

Tudor M. Opran* Samir Loudni†

February 4, 2025

Remark 1. In this proof, we only consider rules with distinct scores. Since the rule space is finite, we order the rules by their score, i.e.,

$$\forall i > j \in \mathcal{D}, \quad g(r_i) > g(r_j),$$

where $g(\cdot)$ is a rule scoring function. $\mathcal{D}$ is the set of all possible rule indices.

Definition 2. Let $\theta(\cdot, \cdot)$ be a rule out-ranking certainty function satisfying the following properties:

- $\forall i, j \in \mathcal{D}, \quad \theta(r_i, r_j) + \theta(r_j, r_i) = 1.$
- $\forall k \in \mathcal{D}, \forall i > j, \quad \theta(r_k, r_j) > \theta(r_k, r_i).$

Lemma 3.

$$\forall i \in \mathcal{D}, \quad \theta(r_i, r_i) = 0.5 \quad and \quad \forall i > j, \quad \theta(r_i, r_j) > \theta(r_j, r_i) \quad and \quad \forall k \in \mathcal{D}, i > j, \ \theta(r_i, r_k) > \theta(r_j, r_k)$$

Proof. The first point is trivial from the first property of Def. 2:

$$\forall i \in \mathcal{D}, \ \theta(r_i, r_i) + \theta(r_i, r_i) = 1 \Leftrightarrow \theta(r_i, r_i) = 0.5$$

The second point is proven as a direct consequence of the second property of Def. 2:

$$\forall k \in \mathcal{D}, \forall i, j \in \mathcal{D}, \ \theta(r_k, r_j) > \theta(r_k, r_i)$$

We apply this property with $k = i$ and j respectivelly:

$$\theta(r_i, r_j) > \theta(r_i, r_i) = 0.5 = \theta(\underbrace{r_j}_{r_k}, r_j) > \theta(r_j, r_i)$$

The third point is direct:

$$\theta(r_k, r_j) > \theta(r_k, r_i) \Leftrightarrow 1 - \theta(r_k, r_j) < 1 - \theta(r_k, r_i) \Leftrightarrow \theta(r_i, r_k) > \theta(r_j, r_k)$$

□

*tudor-matei.opran@imt-atlantique.net
†samir.loudni@imt-atlantique.fr

Definition 4 (Stochastic Matrix T)**.** Let $T \in \mathcal{M}_n([0,1])$ be the stochastic matrix associated with the Markov chain whose states are the decision rules. For all i, j, $T_{i,j}$ represents the probability of transitioning from state i to state j. The matrix T satisfies the following conditions:

1. $\forall r_j \notin N(r_i), \quad T_{r_i,r_j} = 0$.
2. $\forall r_j \in N(r_i), \quad T_{r_i,r_j} = \frac{1}{|N(r_i)|}\theta(r_j, r_i)$.
3. $T_{r_i,r_i} = \frac{1}{|N(r_i)|}\sum_{r_j \in N(r_i)} \theta(r_i, r_j)$.

This ensures that the probabilities of exiting state i sum up to one.

Theorem 5. *For an irreducible chain, there is an equilibrium distribution¿*

Proof. Let $f_{i,i}(k)$ be the probability of the first return in i after exactly k steps. By definition, the periodicity of the Markov chain is $\mathcal{P}_i = \gcd\{k : f_{i,i}(k) \neq 0\}$. Since the Markov chain is ireducible, all states have the same periodicity. Let $i \in \mathcal{D}$, for all $r_j \in N(r_i)$, $\theta(r_i, r_j) > 0$ otherwise we wouldn't be able to access the state r_i contradicting ireducibility. Thus the period of the stater_i is one, showing the aperiodicity of the Markov chain. Since the chain is finite, we can conclude that it is ergodic and thus has an equilibrium distribution. □

Theorem 6. *Let π be the probability vector associated with the equilibrium distribution of the Markov chain of transition matrix T. In the general case, we have:*

$$\forall i, j \in \mathcal{D}, \quad i > j \implies \frac{\pi_i}{|N(r_i)|} > \frac{\pi_j}{|N(r_j)|}.$$

Proof. Suppose, for contradiction, that there exist $i, j \in \mathcal{D}$ such that $i > j$ but $\pi_j \geq \pi_i$. Since the chain is aperiodic and irreducible, the balance equation holds:

$$\forall i \in \mathcal{D}, \quad \sum_{k=1, k\neq i}^{n} [\pi_k \theta(r_i, r_k) - \pi_i \theta(r_k, r_i)] = 0.$$

Applying this to the supposed i and j, we obtain:

$$\begin{aligned}
&[\pi_j\theta(r_i, r_j) - \pi_i\theta(r_j, r_i)] + \sum_{k=1,k\neq i,j}^{n} [\pi_k\theta(r_i, r_k) - \pi_i\theta(r_k, r_i)] \\
&- \left([\pi_i\theta(r_j, r_i) - \pi_j\theta(r_i, r_j)] + \sum_{k=1,k\neq i,j}^{n} [\pi_k\theta(r_j, r_k) - \pi_j\theta(r_k, r_j)] \right) \\
&= 2[\underbrace{\pi_j\theta(r_i, r_j) - \pi_i\theta(r_j, r_i)}_{>0}] + \sum_{k=1,k\neq i,j}^{n} [\pi_k(\underbrace{\theta(r_i, r_k) - \theta(r_j, r_k)}_{>0}) + \underbrace{\pi_j\theta(r_k, r_j) - \pi_i\theta(r_k, r_i)}_{>0}] > 0.
\end{aligned}$$

This contradiction establishes that $\pi_i \geq \pi_j$. □

Remark 7. Let $N(\cdot)$ be a neighborhood such that $|N(r_i)|$ increases with the rule's score $g(r_i)$. The equilibrium probability of a rule is increasing with its score: $\forall i, j \in \mathcal{D}, \quad i > j \Rightarrow \pi_i > \pi_j$. Here lies the inherent difficulty of an algorithm. The easiest way to satisfy this condition is to ensure that all of the rules have the same number of neighbors.