

unifiedml : Une Interface Unifiée pour l'Apprentissage Automatique dans R

T. Moudiki (Auteur et mainteneur du package, thierry.moudiki@gmail.com)

L'écosystème R pour l'apprentissage automatique (ML) est riche mais fragmenté, avec des interfaces hétérogènes (formule $y \sim .$ vs. matrice x, y) selon les packages (`glmnet`, `randomForest`, `e1071::svm`, etc.). Cette diversité peut créer une surcharge cognitive, et ralentir le processus d'apprentissage automatique.

Pour unifier ce paysage, je présente `unifiedml`, un package R offrant une interface cohérente et inspirée de `scikit-learn` pour n'importe quel modèle de ML dans R.

`unifiedml` introduit une classe R6 `Model` qui encapsule n'importe quelle fonction de modèle. Un mécanisme de détection automatique identifie l'interface appropriée (formule ou matrice) et le type de tâche (régression ou classification) à partir de la variable réponse. Cela permet une syntaxe unique et intuitive :

```
Model$new(fonction_modele)$fit(X, y, ...).
```

Le package standardise également (à ce jour) des étapes clés du workflow : validation croisée universelle, métriques automatiques, interprétabilité via dérivées numériques et graphiques de dépendance partielle.

Mots-clefs (3 à 5) :

Package R – Apprentissage Automatique – Interface Unifiée – Science des Données – Reproductibilité

Développement

1. Contexte et problématique

Le langage R dispose d'une collection impressionnante de packages pour l'apprentissage automatique, chacun avec ses forces spécifiques (régularisation, forêts aléatoires, machines à vecteurs de support, etc.). Cependant, cette richesse a un coût en termes de complexité d'utilisation.

Les signatures de fonctions, les formats de données et les conventions diffèrent selon les packages, ce qui ralentit l'expérimentation. Par exemple :

- `glmnet(x, y, ...)` utilise une interface matricielle
- `randomForest(y ~ ., data, ...)` utilise une interface par formule comme `lm`
- `e1071::svm(x, y, ...)` utilise une interface matricielle

Des solutions comme `caret` ou `mlr3` existent pour réunifier ainsi l'écosystème, mais elles reposent sur un grand nombre de dépendances et nécessitent parfois l'apprentissage d'un nouveau cadre conceptuel.

2. Solution proposée : `unifiedml`

`unifiedml` agit comme une couche d'abstraction légère, inspirée (mais plus légère) de l'API de `scikit-learn`, sans imposer de nouveau formalisme.

L'utilisateur interagit toujours avec la même interface `fit / predict`, quel que soit le modèle sous-jacent.

3. Impact pour la communauté R francophone

- Débutants : réduction de la courbe d'apprentissage
- Utilisateurs avancés : prototypage rapide et comparaisons rigoureuses
- Formateurs : enseignement du ML sans surcharge syntaxique

4. Plan de la présentation

1. Fragmentation des interfaces ML en R
2. Démonstration de unifiedml
3. Comparaisons avec caret et mlr3
4. Limites et perspectives