

## GENÉRATEUR DE DONNÉES ET SUPERVISEUR D'EXPERIENCES

Gilles Verley et Eric Ramat<sup>1</sup>

Laboratoire d'Informatique

Ecole d'Ingénieurs en Informatique pour l'Industrie

64, avenue Jean Portalis

37200 - TOURS

☎ 02 47 36 14 14

e-mail: ramat,verley@univ-tours.fr

### Résumé

Générer des jeux de données complexes conformes à un plan d'expériences et faciliter la mise en oeuvre de ces données sur les programmes informatiques que l'on veut expérimenter sont les deux objectifs que nous nous sommes donnés pour réaliser l'atelier informatique d'expérimentation sur des programmes que nous présentons d'abord au travers d'une expérimentation sur les classifieurs supervisés et, ensuite, de manière théorique.

### 1. Introduction

Disposer de données numériques générées artificiellement est une nécessité chaque fois que l'on veut observer le comportement externe d'un programme informatique qui est censé résoudre un problème particulier, c'est-à-dire produire une sortie aussi proche que possible d'une sortie désirée lorsqu'on lui fournit certaines données en entrée. Ces simulations ou expériences sont justifiées de différentes manières selon la nature des problèmes à résoudre:

- lorsqu'il existe des solutions déterministes, les programmes informatiques qui les implémentent sont, en général et en pratique, très difficilement prouvables. Il est donc nécessaire de les vérifier empiriquement avant de les mettre en situation réelle de fonctionnement. De plus, même lorsqu'ils sont formellement équivalents, certaines de leurs performances, comme le temps de calcul, peuvent varier.
- lorsqu'il n'existe pas de solution déterministe, les heuristiques qui en font office sont, par nature, gourmandes en paramètres d'optimisation. Ceci nécessite de vastes plans d'expérience pour estimer et comparer les performances sur des ensembles de données suffisamment représentatifs. Cette problématique est très courante dans des domaines tels que la reconnaissance des formes et l'ordonnancement.

Face à cette situation, le concepteur d'un programme doit réaliser un second programme, sur mesure, permettant de valider ou optimiser le premier. Ce second programme, outre le temps et le coût que son développement occasionne, est une nouvelle source potentielle d'erreurs. De plus, la reproduction éventuelle des expériences par des tierces personnes suppose qu'elles disposent également de ce second programme ou, pour le

---

<sup>1</sup> Allocataire de Recherche du Conseil Régional du Centre - Moniteur CIES