Résumé

L'évaluation de la qualité de logiciel s'avère d'une importance fondamentale dans le domaine de développement de logiciel. La plupart des qualités ne peuvent, cependant, pas être mesurées d'avance (à priori), c'est-à-dire avant que le logiciel ne soit utilisé pour une certaine période de temps. Cependant, ces qualités peuvent être estimées à partir de propriétés qui sont mesurables. Dans ce but, sont construits les modèles d'estimation de la qualité de logiciel, qui trouvent un grand usage. La plupart de ces modèles sont construits à partir de techniques statistiques ou d'apprentissage automatique. Toutefois, dans ce domaine, il est difficile d'obtenir des ensembles de données pouvant être utilisés pour produire de tels modèles. En effet, la plupart du temps, ces ensembles de données sont des propriétés privées et ceux qui sont publics sont de tailles très limitées ou non représentatifs. Ceci implique une détérioration assez importante dans la prédiction que les modèles performent sur des données nouvelles.

Cette thèse explore l'utilisation des algorithmes génétiques pour résoudre le problème d'optimisation des modèles, à base de règles, d'estimation de la qualité de logiciel. La contribution principale de cette oeuvre consiste en deux approches évolutionnistes à ce probléme d'optimisation. La première consiste à combiner différents modèles existants, et de les adapter à un ensemble de données. La deuxième consiste à optimiser un seul modèle. L'idée principale de cette thèse est de prendre des modèles qui existent, ayant été déjà construits à partir d'un certain ensemble de données, et de les adapter à d'autres données. En pratique, ceci pourrait être vu comme l'adaptation de modèles, construits à partir des données générales, à des données reliées à un contexte bien spécifique. Notre technique maintient la nature "boîte blanche" des modèles, ces derniers pouvant être utilisés comme normes dans des procédures futures de développement de logiciel.