

Compteurs et aléas en théorie des automates

Nathanaël Fijalkow

L'informatique fondamentale s'attaque à la question suivante : qu'est-ce que l'information ? Dans une société où l'information se crée, s'échange et se transforme se posent de nombreuses questions : comment représenter, analyser, protéger, diffuser et pérenniser les données ? Le terme "big data" est dans toutes les bouches ; que signifie t-il ? L'informatique fondamentale nous donne des clés pour appréhender ces grands sujets d'actualité.

L'algorithmique est la branche de l'informatique fondamentale qui explore la notion de calcul, au cœur de ces problématiques : comment automatiser le traitement de l'information ? Il s'agit de bâtir les fondements mathématiques de l'automatisation, afin d'en étudier la force, d'en estimer les limites, et de capturer son essence.

Les travaux fondateurs de Turing ont jeté un pont entre l'univers de l'informatique et celui des mathématiques : l'idée de machine venant du premier est formalisée par la notion d'automate, appartenant au second.

La force de la notion d'automate est l'abstraction. Sa formalisation en tant qu'objet mathématique abstrait permet de raisonner sur cet objet, d'en étudier les propriétés, de construire des algorithmes et d'évaluer leurs mérites. La théorie des automates fascine les mathématiciens par la profondeur de ses problèmes, impliquant par exemple l'algèbre, la combinatoire, la logique et la topologie, ainsi que les informaticiens par l'élégance des solutions qu'elle propose. Développée conjointement par des mathématiciens et des informaticiens depuis presque un siècle, la notion d'automate est à la fois un élément clé dans le développement industriel de l'informatique et une pièce centrale dans notre compréhension de l'informatique fondamentale.

Vers une modélisation des données quantitatives

Les automates ont un fonctionnement booléen, "vrai" ou "faux". Ils traitent des données encodées par des suites de 0 et de 1, ce que l'on appelle des données qualitatives.

Or, certaines données sont de manière inhérentes quantitatives : elles représentent des mesures physiques, des statistiques ou bien sont le résultat de calculs numériques. Afin de prendre en compte ces spécificités la notion d'automate s'est enrichie : selon les applications, un automate peut être équipé d'attributs quantitatifs lui permettant de modéliser des données quantitatives.

L'objet de ma thèse a été d'étudier deux modèles quantitatifs d'automates : ceux équipés de compteurs, et ceux équipés d'aléas. Dans les deux cas, il s'agit de modèles très répandus, abondamment étudiés en informatique fondamentale depuis plusieurs décennies.

L'originalité de mon travail réside dans son approche : afin d'étudier les automates à compteurs et probabilistes, j'ai introduit de nouveaux outils issus des mathématiques, apportant un nouveau point de vue sur ces modèles. En particulier, ces nouvelles techniques ont permis de créer des liens entre ces deux modèles, a priori non reliés.

Analyse des automates à compteurs à l'aide de la théorie des jeux.....

La première partie de ma thèse étudie les automates à compteurs, c'est-à-dire des machines disposant d'un attribut permettant de compter. Le principal objet d'étude est un formalisme logique appelé la logique monadique de coût, qui permet de modéliser la notion de ressources. Ce formalisme a été étudié depuis une dizaine d'années, et a permis la construction de nombreux algorithmes. Son développement culmine par l'établissement d'une conjecture par Colcombet et Löding en 2010, qui implique une solution au problème de l'index de Mostowski, un des plus importants problèmes ouverts en théorie des automates et en logique.

Ma contribution principale sur le sujet est de donner un contre-exemple à la conjecture de Colcombet et Löding. Loin de couper tout espoir, en montrant des limites il ouvre des perspectives, et permet d'énoncer une conjecture plus précise, ayant les mêmes implications. Ma deuxième contribution est de prouver un cas particulier de la nouvelle conjecture, en m'appuyant sur des outils de théorie des jeux et de décomposition de graphes.

Analyse des automates probabilistes par l'algèbre et la topologie.....

La deuxième partie de ma thèse étudie les automates probabilistes, c'est-à-dire des machines capables d'effectuer des choix aléatoires. La question que l'on se pose ici est de déterminer algorithmiquement la fiabilité d'un tel automate, formalisée par la notion de valeur issue de la théorie des jeux. Il est connu qu'il n'existe pas d'algorithme permettant de calculer, ou même d'approximer, la valeur d'un automate probabiliste : ce problème est indécidable. Néanmoins, il existe de nombreux algorithmes donnant des résultats partiels sur cette question.

Ma contribution principale sur le sujet est de construire un algorithme issu de techniques algébriques et combinatoires, à savoir la théorie des semigroupes et les théorèmes de Ramsey, et de prouver au moyen de techniques topologiques qu'il est optimal. Cette nouvelle approche topologique, inspirée de la théorie profinie, donne un nouveau point de vue sur les automates probabilistes.

D'irréalisable à praticable

La finalité des résultats présentés dans cette thèse est la construction d'algorithmes efficaces permettant de manipuler des données quantitatives. L'étude des propriétés de ces algorithmes, l'introduction de nouvelles techniques amenant différents points de vue, amènent à construire des algorithmes toujours plus efficaces.

Une réalisation concrète est l'implantation des algorithmes construits dans cette thèse dans les logiciels libres ACME et Stamina. Grâce aux progrès théoriques obtenus depuis plusieurs décennies, ACME et Stamina résolvent des problèmes réputés inattaquables il y a vingt ans.