

Analyse de la complexité d'algorithmes sur les automates finis et les langages

Ce sujet de thèse porte sur l'analyse de la complexité en moyenne des algorithmes sur les automates finis et les langages rationnels. À tout langage rationnel est associé de manière unique son automate minimal, qui est l'automate le plus petit (ayant le moins d'états) parmi tous les automates déterministes qui reconnaissent ce langage. Ainsi la complexité en espace d'un langage rationnel peut être interprétée comme le nombre d'états de son automate minimal. Pour cette mesure, la complexité dans le pire cas des algorithmes sur les automates finis est connue dans la plupart des cas [5]. En revanche, très peu de résultats ont été obtenus concernant l'analyse en moyenne de ces algorithmes.

L'analyse en moyenne repose sur une étude combinatoire des objets manipulés tant d'un point de vue analytique qu'algébrique, ainsi que sur une analyse fine des algorithmes considérés.

Génération aléatoire

Un premier axe de ce projet est lié au développement théorique et logiciel d'algorithmes de génération aléatoire. La génération aléatoire d'automates permet en effet de tester les propriétés d'un automate aléatoire et d'étudier expérimentalement la complexité d'algorithmes complexes. Les résultats empiriques et la possibilité de tester de manière exhaustive des conjectures sur les objets de petite taille aide à la conception d'algorithmes efficaces et à l'élaboration de résultats théoriques.

Un générateur aléatoire d'automates déterministes a déjà été conçu et développé [2]. Expérimentalement cette librairie permet également d'engendrer de manière efficace des automates minimaux. La complexité de cette génération reste à établir et dépend du nombre d'automates minimaux de taille fixée.

La conception de générateurs aléatoires d'automates non-déterministes reste un sujet ouvert. Plusieurs classes d'automates non-déterministes peuvent être considérées : tous les automates non-déterministes, les automates co-déterministes, les automates obtenus par application d'un algorithme (celui de Glushkov par exemple), des automates déterministes auxquels on ajoute un nombre fixé de transitions non-déterministes...

Ce premier axe s'inscrit dans le cadre d'un projet ANR Blanc en cours d'évaluation intitulé "Génération Aléatoire : Modèles, Méthodes et Algorithmes" et dont je suis porteuse.

Minimisation

Un deuxième axe de ce projet concerne l'étude de la complexité d'algorithmes de minimisation.

Un premier problème est celui du calcul de la borne inférieure de la complexité dans le pire des cas de la minimisation. En d'autres termes, quelle est la complexité minimale d'un algorithme de minimisation ? Nous conjecturons que cette complexité pour la classe des

algorithmes reposant sur des raffinements de partitions est proportionnelle à $n \log n$ où n est la taille des automates traités. Ce résultat établirait l'optimalité dans le pire des cas de l'algorithme de Hopcroft, qui atteint cette borne [3].

Un autre problème est celui de la complexité en moyenne de l'algorithme de Moore. Algorithme très simple, sa complexité dans le pire cas est en n^2 . Experimentalement, sa complexité semble sous quadratique et son efficacité en moyenne meilleure que celle de l'algorithme d'Hopcroft.

L'étude en moyenne de la complexité de l'algorithme de Brzozowski permettrait de compléter cette étude de la minimisation. Cet algorithme repose sur la déterminisation et non plus sur le raffinement de partitions. Sa complexité dans le pire est exponentielle, mais il se révèle efficace dans certains cas. Une étude de sa complexité sur la classe des automates co-déterministe serait intéressante.

Déterminisation

De nombreux problèmes par exemple en recherche de motifs sont facilement décrits par des automates non-déterministes. Parallèlement, la plupart des algorithmes sont basés sur la représentation des langages par un automate déterministe, ou mieux par leur automate minimal.

La complexité de la déterminisation est dans le pire exponentielle. une étude de la complexité moyenne de la déterminisation de classes d'automates non-déterministes permettrait de compléter l'analyse d'algorithmes classiques.

Opérations sur les langages

Si la complexité dans le pire cas des opérations sur les langages rationnels est souvent connue, la complexité en moyenne des opérations de base restent à étudier. Quelle est par exemple la taille moyenne de l'intersection de deux langages rationnels ?

Bibliographie

- [1] F. Bassino, C. Nicaud, Enumeration and random generation of accessible automata, *Theoretical Computer Science*, 381. (2007), 86–104.
- [2] F. Bassino, J. David, C. Nicaud, REGAL: a Library to randomly and exhaustively generate automata, volume 4783 of LNCS, p. 303-305, Springer-Verlag, 2007.
- [3] J. Berstel, O. Carton, On the complexity of Hopcroft's state minimization algorithm, In *CIAA '2004*, volume 3317 of LNCS, p. 35-44, Springer-Verlag, 2004.
- [4] J.-M. Champarnaud, T. Paranthoën, Random generation of DFAs, *Theoret. Comput. Sci.*, 330 (2005), 221–235.
- [5] S. Yu, Q. Zhuang and K. Salomaa, The state complexities of some basic operations on regular languages, *Theoret. Comput. Sci.*, 125 (1994), 315–328.