Titre: liens entre génération aléatoire et génération exhaustives

Thématique: Complexité, algorithmique

Laboratoire : LIP6 (Équipe RO), Université Pierre et Marie Curie

Encadrant : Yann Strozecki(http://www.prism.uvsq.fr/~ystr/)

Contexte:

En complexité on s'intéresse généralement au temps nécessaire pour décider si un problème a une solution et éventuellement la produire.

La complexité d'énumération étudie les algorithmes qui produisent toutes les solutions d'un problème, par exemple l'ensemble des tris topologiques d'un graphe, ou un ensemble d'objets combinatoires, par exemple les arbres d'une taille donnée. Dans ce contexte on étudie la dynamique de l'algorithme c'est-à-dire le délai entre la production de deux solutions. Un problème d'énumération est résolu efficacement si il existe un algorithme avec délai et mémoire polynomiale en l'entrée qui produit son ensemble de solutions.

Les problèmes d'énumération apparaissent naturellement quand on s'intéresse à l'optimisation, aux requêtes sur les bases de données ou au data-mining. Il s'agit ici d'étudier leur complexité d'un point de vue théorique. On obtient parfois des résultats étonnants, par exemple trouver un couplage parfait dans un graphe est facile, compter leur nombre est difficile et tous les produire est facile. Énumérer les ensembles indépendants maximaux d'un graphe dans l'ordre lexicographique est facile mais est dur dans l'ordre antilexicographique.

Objectifs:

Dans ce stage on voudrait décrire un cadre pour obtenir automatiquement des algorithmes efficaces d'énumération.

L'idée est d'essayer de combiner des problèmes qu'on sait résoudre efficacement pour obtenir des problèmes plus compliqués qu'on saurait ainsi automatiquement résoudre. Par exemple le produit cartésien ou l'union de deux ensembles faciles à énumérer sont faciles à énumérer mais ça n'est pas le cas de leur intersection ou de leur différence [1].

Ce genre de méthode est connue dans le domaine du comptage de structures combinatoires et de la génération uniforme de ces structures. Certains objets très simples admettent des générateurs uniformes et une grammaire dont les opérations comprennent justement l'union et le produit cartésien permettent de construire de nombreuses classe d'objets ainsi que le générateur uniforme associé.

Ces générateurs s'appellent des Boltzman samplers [2]. On voudrait répliquer cette approche pour l'énumération exhaustive d'objets, éventuellement en l'implémentant et en évaluant automatiquement le délai.

En général, il serait intéressant d'approfondir les relations entre génération uniforme, comptage approchée

et énumération exhaustive ou approchée. Le lien est connu [3] mais a été peu exploré. Un travail récent [4] a montré que l'existence d'un générateur uniforme implique l'existence d'un algorithme d'énumération randomisé avec un bon délai et une mémoire exponentielle. On peut également obtenir une mémoire polynomiale, au prix d'un délai plus grand (incrémental). Il serait intéressant de trouver des propriétés de certains générateurs aléatoires qui permettent d'obtenir à la fois délai et mémoire polynomiale et de déterminiser ces alorithmes, par exemple en utilisant des séquences universelles qui permettent de faire un parcours du graphe des solutions sans utiliser de mémoire.

Ce travail sera fait dans le cadre du projet ANR AGGREG.

- [1]:Yann Strozecki. Enumeration complexity and matroid decomposition. PhD thesis
- [2]:Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer.
- Boltzmann samplers for the random generation of combinatorial structures

[3]: Leslie Ann Goldberg. Efficient algorithms for listing combinatorial structures. PhD thesis, University of Edinburgh, UK, 1991.
[4]:Florent Capelli, Yann Strozecki. On the Complexity of Enumeration

Compétences désirées (non obligatoire): théorie de la complexité, probabilité discrète/combinatoire, algorithmique