

Extraction itérative de motifs intéressants et représentations condensées

Arnaud Giacometti
Université de Tours

Dans le cadre des bases de données inductives, l'extraction itérative de motifs intéressants pose principalement deux problèmes :

1. Le stockage des réponses aux requêtes d'extraction déjà calculées. Le volume des réponses aux requêtes d'extraction pouvant être très important, ce sont en général des représentations condensées de ses réponses qui sont calculées et stockées.
2. L'optimisation du calcul de la réponse à de nouvelles requêtes d'extraction en utilisant les représentations condensées des réponses aux requêtes déjà calculées.

Dans nos travaux, nous nous intéressons au calcul de motifs intéressants où les prédicats de sélection considérés sont des conjonctions de prédicats de sélection monotones et anti-monotones (ce qui implique que les réponses aux requêtes d'extraction considérées sont des espaces convexes). Dans ce cadre, nous montrons :

1. Comment ne pas stocker de manière indépendante les représentations condensées des réponses aux requêtes déjà calculées. Plus précisément, nous montrons comment fusionner ces représentations condensées en introduisant la notion de représentation condensée d'un ensemble de réponses à des requêtes d'extraction.
2. Comment utiliser ce nouveau type de représentation condensée pour optimiser le calcul de nouvelles requêtes d'extraction.
3. Comment mettre à jour la représentation condensée d'un ensemble de réponses à des requêtes d'extraction lorsqu'une nouvelle requête d'extraction a été calculée.

Nous présentons également quelques résultats expérimentaux préliminaires démontrant l'intérêt de notre approche. Nous concluons en ouvrant quelques perspectives concernant ces travaux, et plus généralement, les recherches dans les bases de données inductives.