

IncFDs: un nouvel algorithme d'inférence incrémentale des dépendances fonctionnelles

Ghada Gasmi*

* Département des Sciences de l'Informatique
Faculté des sciences de Tunis
Campus universitaire, 1060 Tunis, Tunisie
ghada.gasmi@gmail.com

Résumé. L'inférence des dépendances fonctionnelles est l'une des problématiques les plus étudiées en bases de données. Elle a fait l'objet de plusieurs travaux qui ont proposé des algorithmes afin d'inférer, efficacement, les dépendances fonctionnelles pour les utiliser dans différents domaines : administration de bases de données, ré-ingénierie, optimisation des requêtes, etc. Toutefois, pour les applications réelles, les bases de données sont évolutives et les relations sont fréquemment augmentées ou diminuées de tuples. Par conséquent, afin de s'adapter à ce cadre dynamique, une solution consiste à appliquer l'un des algorithmes, disponibles dans la littérature, pour inférer les dépendances fonctionnelles, après chaque mise à jour. Cette solution étant coûteuse, nous proposons, dans cet article, d'inférer les dépendances fonctionnelles d'une manière incrémentale. À cet effet, nous introduisons un nouvel algorithme, appelé INCFDS, et nous évaluons ses performances par rapport à l'approche classique d'inférence des dépendances fonctionnelles à partir d'une relation dynamique.

1 Introduction

Les dépendances fonctionnelles (DF) ont été largement étudiées dans le cadre des bases de données relationnelles. À l'origine, l'étude des DF a été motivée par le fait qu'elles peuvent être utilisées pour exprimer des contraintes vérifiées pour un schéma de relation indépendamment d'une instance particulière du schéma (Abiteboul et al. (1995)).

Un peu plus tard, l'étude des DF a été motivée par une autre problématique. En effet, lors de la conception des systèmes d'information ou de bases de données, la connaissance de dépendances fonctionnelles était supposée connue. Toutefois, durant la vie d'une base de données, le problème se pose dans d'autres termes : quelles sont, à partir des ensembles des données stockées, les dépendances qui sont effectivement vérifiées ? Ce problème est connu comme celui de *l'inférence de dépendances fonctionnelles*, et s'il a fait l'objet de nombreux travaux (Wyss et al. (2001); Lopes et al. (2000); Yao et al. (2002); Novelli et Cicchetti (2001); Flach et Savnik (1999); Huhtala et al. (1999); Mannila et Raiha (1987))¹, c'est parce que la réponse apportée à

1. Une étude détaillée portant sur ces algorithmes est présentée dans (Lopes et Novelli)