

Extraction de Dépendances Fonctionnelles Approximatives: une Approche Incrémentale

Ekaterina Simonenko* et Noël Novelli**

* LRI-CNRS UMR 8623

Université Paris-Sud XI; F-91405 Orsay Cedex
ekaterina.simonenko@lri.fr

** LIF-CNRS UMR 6166 – Case 901

Université d'Aix-Marseille; Faculté des Sciences de Luminy;
F-13288 Marseille Cedex 9
noel.novelli@lif.univ-mrs.fr

Résumé

La découverte de dépendances fonctionnelles (DF) à partir d'une relation existante est une technique importante pour l'analyse de Bases de Données. L'ensemble des DF exactes ou approximatives extraites par les algorithmes existants est valide tant que la relation n'est pas modifiée. Ceci est insuffisant pour des situations réelles où les relations sont constamment mises à jour.

Nous proposons une approche incrémentale qui maintiens à jour l'ensemble des DF valides, exactes ou approximatives selon une erreur donnée, quand des tuples sont insérés et supprimés. Les résultats expérimentaux indiquent que lors de l'extraction de DF à partir d'une relation continuellement modifiée, les algorithmes existants sont sensiblement dépassés par notre stratégie incrémentale.

Summary

Functional Dependency (FD) inference from existing relations is an important data base analysis technique. The problem has been treated by using data mining approach, by the algorithms TANE, DEPMINER, FASTFDSet FUN. Since the FD set, inferred by these algorithms is valid as long as the relation remains unchanged, they become efficient in the real life situations when the relation is constantly updated. It is particularly important to be able to extract approximate FDs, allowing to take into account errors and exceptions in the given data. We propose an incremental approach, allowing to update the valid set of approximate FDs, while tuples are inserted or deleted. The algorithm employs a level-wise strategy to keep the internal representation of the relation up-to-date. Experimental results show that during FDs extraction from a constantly updated relation, existing algorithms are significantly outperformed by our incremental approach.