

KLASIFIKASI SECARA EFISIEN PADA DATABASE MULTI RELASI DENGAN ALGORITMA CROSSMINE

Sarwosri, Darlis Herumurti, Indri Sulistyowati

Jurusan Teknik Informatika, Fakultas Teknologi Informasi,
Institut Teknologi Sepuluh Nopember
e-mail: sri@its-sby.edu , darlis@its-sby.edu

Abstract

Multi-relation classifications can be widely used in many disciplines, such as financial decision making, medical research, and geographical applications, and information stored in multiple relations needs to be used in decision making. Crossmine is an efficient and scalable approach for multi-relation classification. Crossmine algorithm has three step, first is find-rules, the rule has been gotten from find a rule process than remove all positif tuples satisfying rule while there are more than ten percent positif tuple left. The second is find a rule, this step has input from the result of find best predicate process, that is the complex predicate with most foilgain. If foilgain value is more than mingain, the predicate is added with rule, and max rule length less than six. Third is find best predicate, in this step we find the best predicate with definition, if the foilgain value more than the max gain value, the predicate will be saved and the bigger gain value will replace the last gain value for next comperative process. In other side, the accuracy is computed from each rule that produce in find rules process. The test for this application use the sum tuple of 200, 500, 1000, 5000 for measuring the level of accuracy from rule which is produced by crossmine algoritm.

Keyword: crossmine, positive tuple, negative tuple, rule, predicate, accuracy.

Abstrak

Klasifikasi multi-relasi dapat digunakan secara luas dalam banyak disiplin, seperti pembuatan keputusan keuangan, penelitian medis, dan aplikasi geografi, informasi yang disimpan dalam multi-relasi diperlukan dalam pembuatan keputusan. CrossMine, suatu pendekatan yang efisien dan dapat diskala untuk klasifikasi multi-relasi. Algoritma crossmine memiliki tiga langkah, pertama dimulai dari pencarian sekumpulan rule, rule didapatkan dari proses pencarian sebuah rule yang kemudian diambil tuple positifnya sampai tersisa lebih dari sepuluh persen. Langkah kedua adalah pencarian sebuah rule, pada langkah ini memiliki input dari hasil proses pencarian predikat terbaik, yaitu predikat komplek dengan nilai foilgain, nilai foil gain dibandingkan dengan min gain jika hasilnya lebih besar min gain maka predikat tersebut ditambahkan dengan rule, dan maksimal kedalaman relasi adalah kurang dari enam. Langkah ketiga adalah find best pred, pada langkah ini dilakukan pencarian predikat terbaik dengan ketentuan jika nilai foilgain lebih besar dari nilai max gain maka predikat akan disimpan dan nilai gain yang lebih besar akan menggantikan nilai gain sebelumnya untuk proses perbandingan selanjutnya. Sedangkan untuk accuracy dihitung dari masing masing rule yang dihasilkan pada proses pencarian sekumpulan rule. Uji coba aplikasi ini menggunakan jumlah tuple 200, 500, 1000, 5000 untuk mengukur tingkat efisiensi dan akurasi dari rule yang dihasilkan algoritma crossmine.

Kata kunci : crossmine, tuple positif, tuple negative, rule, predikat, accuracy

1. PENDAHULUAN

Kebanyakan data terstruktur saat ini disimpan dalam relasi database. Database ini terdiri dari banyak relasi yang berhubungan secara konseptual melalui *link-link entity-relation* dalam desain skema-skema relasi database. Klasifikasi multi-relasi dapat digunakan secara luas dalam banyak disiplin, seperti pembuatan keputusan keuangan, penelitian medis, dan