

Review Paper

14.8429

Problem

(Wu, Chen, and Han, n.d.)

1. Membuat Unified Framework dari ukuran-ukuran *interestingness* yang sifatnya null-variant (rangennya $[0 - 1]$), kemudian mengusulkan ukuran tambahan – Imbalanced Ratio – sebagai ukuran komplemen *interestingness*.
2. Mengusulkan algoritma GAMiner dan membandingkannya dengan algoritma FP-GROWTH untuk menggali (to mine) ukuran Kulc dan Cosine.

Method

Percobaan algoritma GAMiner dan FP-GROWTH dilakukan pada komputer Pentium dengan CPU 3.0 GHz, 1GB RAM, dan 160G hard disk. Semua kode sumber (source code) ditulis dalam bahasa C ++ dan dikompilasi dengan menggunakan Microsoft Visual C ++ di Windows XP. Dataset yang digunakan adalah Dataset sintesis pada Han, dkk (2000).

Bagaimana GAMiner menjalankan algoritma dijelaskan pada gambar berikut.

Algorithm: GAMiner: Mining interesting association patterns through measure upper bounding.

Input: A database DB represented by a FP-Tree $Tree$, support threshold θ , association threshold γ , and an initially empty pattern P .

Output: The complete set of frequent and strongly associated patterns \mathbb{S}_{kule} .

Procedure: GAMiner ($Tree, \theta, \gamma, P$)

```

1  FOR each event  $a_i$  in the header of  $Tree$  DO
2    Generate pattern  $Q = P \cup a_i$ ;
3    IF  $Kule(Q) \geq \gamma$  THEN  $\mathbb{S}_{kule} = \mathbb{S}_{kule} \cup Q$  ;
4    Get  $Q$ -projected database including a set  $I_Q$  of events;
5    Calculate  $min\_sup_b$ ,  $c$ , and  $tree\_depth$  from the  $Q$ -projected
    database;
6    FOR each  $b_j$  in  $I_Q$  DO
7      IF  $sup(Qb_j) < \theta$  THEN
8        Remove  $b_j$  from  $I_Q$ ; //support pruning
9      IF  $length(Qb_j) + tree\_depth < \lceil \frac{c}{1-\gamma} \rceil$  THEN
10       Remove  $b_j$  from  $I_Q$ ; //association pruning
11    Construct  $Q$ -conditional FP-Tree  $Tree_Q$  with events in  $I_Q$ ;
12    IF  $Tree_Q \neq \emptyset$  THEN
13      GAMiner ( $Tree_Q, \theta, \gamma, Q$ );

```

Figure 1: Algoritma GAMiner

Result

1. Paper ini adalah paper pertama yang membuat Unified Framework untuk null-variant interesting measures yang telah ada.
2. GAMiner lebih efisien daripada FP-GROWTH. Untuk support threshold yang sama, GAMiner membutuhkan waktu yang lebih sedikit daripada untuk FP-GROWTH.

Future Work

Dengan paper ini, penulis yakin bahwa penting untuk memilih *interestingness* measures yang tepat ketika *me-mining* data set yang memiliki banyak kejadian (many events) yang memiliki probabilitas yang kecil. Oleh karena itu, pemeriksaan ulang terhadap interestingness measure pada banyak aplikasi lanjutan, seperti analisis jaringan sosial, penambahan data biomedis, dan klasifikasi berbasis pola dan pengelompokan, merupakan tugas penting untuk penelitian selanjutnya.

Referensi

Wu, Tianyi, Yuguo Chen, and Jiawei Han. n.d. “Re-Examination of Interestingness Measures in Pattern Mining: A Unified Framework.” *Data Mining and Knowledge Discovery*.