

ANALISIS PERBANDINGAN ALGORITMA FP-GROWTH DAN ALGORITMA TREE PROJECTION DALAM PEMBANGKITAN FREQUENT PATTERN COMPARISON ANALYSIS OF FP-GROWTH AND TREE PROJECTION ALGORITHMS IN FREQUENT PATTERN GENERATING

Sri Maryeti¹, -²

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Sebuah basisdata berukuran besar biasanya berpotensi menyimpan banyak informasi penting. Informasi tersebut dapat diperoleh dari pola-pola yang muncul secara berulang. Salah satu jenis pola berulang dalam data mining adalah frequent pattern atau frequent itemset. Untuk mendapatkan frequent pattern pada basisdata yang besar, diperlukan cost yang cukup besar. Oleh karena itu dibutuhkan suatu algoritma yang efisien. FP-growth dan Tree Projection adalah algoritma mining frequent pattern yang memiliki performansi yang baik. FP-growth merupakan algoritma yang efisien karena tidak membutuhkan pembangkitan kandidat frequent pattern, sedangkan Tree Projection merupakan algoritma yang membangkitkan kandidat frequent pattern yang memiliki performansi yang baik. Untuk mengetahui kemampuan serta kelebihan dan kekurangan masing-masing algoritma tersebut, perlu dilakukan studi perbandingan kompleksitas dan performansi algoritma. Dalam tugas akhir ini dilakukan studi perbandingan kompleksitas dan performansi algoritma FP-growth dan Tree Projection. Untuk kebutuhan analisis performansi kedua algoritma ini, maka dibangun suatu perangkat lunak yang mengimplementasikan algoritma Tree Projection dan algoritma FP-growth. Berdasarkan hasil penelitian, kompleksitas algoritma FP-growth dan Tree Projection dipengaruhi oleh jumlah frequent items. Performansi kedua algoritma berbanding terbalik dengan minimum support. Performansi FP-growth lebih baik dari pada Tree Projection ketika minimum support semakin kecil, dan performansi Tree Projection semakin mendekati bahkan lebih baik dari pada FP-growth ketika minimum support semakin besar.

Kata Kunci : data mining, association analysis, frequent pattern, minimum support , FP-growth, TreeProjection

Abstract

Generally, large databases potent to hide a lot of high valuable information. The information can be obtained by paying attention to repeatedly accuring patterns. Frequent Pattern is one of important recurring pattern type in data mining. Getting frequent pattern from large databases needs big enough cost. Therefore an efficient algorithm is necessity. FP-growth and Tree Projection are mining frequent pattern algorithms which have good performance. FP-growth is an efficient algorithm since not require generation of candidate frequent pattern, while Tree Projection is generating candidate frequent pattern algorithm which have good performance. to know ability and also the advantage and disadvantage of the algorithm each other, need study of comparison of complexity and performance algorithm. In this final project will be conducted comparison study of complexity and performance FP-growth and Tree Projection algorithms. For performance analyse requirement of both algorithms, then would be builded a software for implementing Tree Projection and FP-growth algorithms. Based on experiment result, the complexity of FP-growth and Tree Projection effected by frequent items number. Both algorithm performance opposite ratio with minimum support threshold. FP-growth performance is better than Tree Projection when the minimum support getting smaller, and Tree Projection getting closer even better then FP-growth when the minimum support threshold is higher.

Keywords : data mining, association analysis, frequent pattern, minimum support , FP-growth, TreeProjection

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Data mining bertujuan untuk menemukan pola-pola yang valid, baru, mempunyai nilai guna, dan mudah dipahami dari data yang ada. Jenis pola yang dihasilkan ditentukan oleh fungsionalitas *data mining* yang digunakan.

Penambangan *frequent pattern* merupakan salah satu bagian yang sangat penting dalam *data mining*. *Frequent pattern* pertama kali diperkenalkan oleh R.Agrawal [2] sebagai bagian dari fungsionalitas *association*. Hampir keseluruhan proses penambangan *association rule* didominasi oleh proses penambangan *frequent pattern*. Sehingga dibutuhkan suatu algoritma *mining frequent pattern* yang efektif dan efisien.

Salah satu algoritma *mining frequent pattern* yang paling banyak dikembangkan adalah Apriori. Apriori melakukan penambangan *frequent pattern* dengan cara membangkitkan sejumlah kandidat *frequent pattern* kemudian menguji setiap kandidat tersebut dengan cara mencocokkannya dengan data yang ada. Dengan cara kerja tersebut Apriori mampu menemukan semua *frequent pattern*. Akan tetapi cara kerja seperti ini sangat mahal, baik dari segi ruang *memory* maupun waktu.

Beberapa algoritma telah dikembangkan untuk mengatasi kekurangan yang ada pada Apriori, diantaranya adalah FP-growth [8] dan Tree Projection [1]. Hasil penelitian dalam jurnal [1,8] menunjukkan bahwa kedua algoritma ini memiliki kemampuan lebih baik dari Apriori. Untuk mengetahui performansi serta kelebihan dan kekurangan kedua algoritma tersebut dibutuhkan suatu studi analisis perbandingan performansi.

1.2 Perumusan Masalah

Dengan mengacu pada latar belakang masalah di atas, maka permasalahan yang akan dibahas dan diteliti adalah:

- 1) Cara kerja algoritma FP-growth dan Tree Projection
- 2) Penerapan algoritma FP-growth dan Tree Projection untuk menambang *frequent pattern*.
- 3) Performansi algoritma FP-growth dan Tree Projection dalam menghasilkan *frequent pattern*.

1.3 Maksud Tugas Akhir

Berdasarkan rumusan masalah di atas, penyusunan Tugas Akhir ini dimaksudkan untuk :

- 1) Melakukan studi tentang dua algoritma *mining frequent pattern* yaitu FP-growth dan Tree Projection.
- 2) Menganalisa kompleksitas algoritma FP-growth dan Tree Projection
- 3) Membangun perangkat lunak untuk mengimplementasikan algoritma FP-growth dan Tree Projection pada suatu perangkat lunak untuk kebutuhan analisa.
- 4) Melakukan studi perbandingan performansi terhadap kedua algoritma di atas berdasarkan parameter analisa yang ada pada batasan masalah dengan menggunakan perangkat lunak pada poin 3.

1.4 Tujuan Tugas akhir

Dari Tugas Akhir ini diharapkan hal-hal sebagai berikut :

- 1) Mengetahui kelebihan dan kekurangan algoritma FP-growth dan Tree Projection.
- 2) Mengetahui kompleksitas algoritma FP-growth dan Tree Projection.
- 3) Mengetahui performansi kedua algoritma berdasarkan parameter-parameter yang ada pada batasan masalah poin 1.

1.5 Batasan Masalah

Batasan masalah untuk Tugas Akhir ini adalah sebagai berikut :

- 1) Parameter-parameter yang digunakan untuk mengetahui kinerja kedua algoritma tersebut yaitu :
 - a. *Minimal support* dan waktu untuk menghasilkan *frequent pattern*.
 - b. Jumlah transaksi dan waktu untuk menghasilkan *frequent pattern*.
- 2) Analisis kompleksitas algoritma hanya dilakukan untuk kompleksitas waktu.
- 3) Pembahasan Tugas Akhir ini tidak sampai menghasilkan *association rules*.
- 4) Jenis *dataset* yang digunakan adalah *dataset* transaksi, dan nilai setiap atribut telah disesuaikan dengan kebutuhan aplikasi.
- 5) Lexicographic tree pada Tree Projection dibangun menggunakan metode *breadth first* dan pemroyeksian transaksi menggunakan metode *depth first*.
- 6) Analisis performansi FP-growth dan Tree Projection hanya meliputi performansi FP-growth dan Tree Projection pada struktur data yang bisa dimuat di *main memory* sehingga perangkat lunak yang dibangun tidak menangani struktur data FP-tree dan Lexicographic tree yang melebihi kapasitas *main memory*.

1.6 Metodologi Penyelesaian Masalah

Metode - metode yang digunakan dalam pelaksanaan tugas akhir ini adalah:

- 1) Studi literatur :
 - a. Pencarian referensi, mencari referensi yang berhubungan dengan *data mining*, *frequent pattern*, algoritma FP-growth dan Tree Projection, dan kompleksitas algoritma.
 - b. Pendalaman materi, mempelajari dan memahami materi yang berhubungan dengan Tugas Akhir ini.
- 2) Analisis kompleksitas algoritma FP-growth dan algoritma Tree Projection.

- 3) Perancangan dan pembuatan perangkat lunak *mining frequent pattern* menggunakan algoritma FP-growth dan Tree Projection sebagai media pengumpul data untuk analisis.
- 4) Melakukan berbagai percobaan dengan nilai *minimal support* dan ukuran *dataset* yang bervariasi pada kedua algoritma.
- 5) Melakukan studi perbandingan performansi kedua algoritma.
- 6) Penyusunan laporan dalam bentuk tertulis sebagai laporan Tugas Akhir.

1.7 Sistematika Penulisan

Tugas Akhir ini disusun berdasarkan sistematika sebagai berikut :

BAB I PENDAHULUAN

Bab ini meliputi latar belakang masalah, perumusan masalah, maksud dan tujuan Tugas Akhir, batasan masalah, metodologi penyelesaian masalah, dan sistematika penulisan.

BAB II DASAR TEORI

Bab ini memuat berbagai dasar teori yang mendukung dan mendasari penulisan Tugas Akhir ini, yaitu mengenai konsep dari *data mining*, *frequent pattern*, FP-growth, dan Tree Projection.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Dalam bab ini diuraikan tentang analisis algoritma FP-growth dan Tree Projection dari segi teori dan kompleksitas, analisis dan perancangan perangkat lunak sebagai alat bantu dalam proses analisa.

BAB IV IMPLEMENTASI DAN HASIL PENGUJIAN

Bab ini memuat tentang analisis terhadap performansi algoritma FP-growth dan algoritma Tree Projection, membandingkan kinerja kedua algoritma berdasarkan

parameter pembandingan yang telah disebutkan pada batasan masalah.

BAB V

KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang diambil dari pembahasan bab-bab sebelumnya serta saran-saran untuk pengembangan selanjutnya.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

- 1) FP-growth mampu menambang *frequent pattern* dengan efisien karena: (1) menggunakan struktur data yang padat yaitu FP-tree, (2) metode *pattern fragment growth* untuk menghindari biaya pembangkitan kandidat *frequent pattern*, (3) metode *divide-and-conquer* sebagai teknik pencarian pada proses *mining*. Performansi FP-growth mengalami penurunan ketika FP-tree sangat rimbun.
- 2) Tree Projection efisien karena: (1) proyeksi transaksi membatasi proses perhitungan *support*, (2) Lexicographic tree memfasilitasi manajemen dan perhitungan *support* kandidat *frequent itemset*, (3) proses *pruning* menjamin bahwa *nodes* yang ada pada Lexicographic tree adalah *nodes* yang bisa diperluas (aktif) menjadi *frequent itemset* yang lebih panjang. Performansi Tree Projection mengalami penurunan ketika jumlah *frequent items* sangat banyak.
- 3) Kompleksitas algoritma FP-growth dan Tree Projection dipengaruhi oleh jumlah *frequent itemset* dan *support* masing-masing *frequent itemset*. Semakin banyak *frequent itemset* dan semakin besar *supportnya* maka semakin kompleks algoritma FP-growth dan algoritma Tree Projection yang digunakan, demikian pula sebaliknya.
- 4) Algoritma Tree Projection lebih kompleks dari pada algoritma FP-growth karena algoritma Tree Projection harus memproyeksikan transaksi pada setiap *node* Lexicographic tree.
- 5) Performansi algoritma FP-growth dan Tree Projection berbanding terbalik terhadap nilai *minimal support*. Semakin kecil nilai *minimal support* maka performansi kedua algoritma semakin menurun, demikian pula sebaliknya.
- 6) Performansi algoritma FP-growth pada saat *minimal support* semakin kecil lebih baik dari pada algoritma Tree Projection. Sedangkan pada

saat *minimal support* semakin besar performansi algoritma Tree Projection mendekati dan bahkan lebih baik dari pada algoritma FP-growth.

- 7) Pertambahan waktu pembangkitan *frequent pattern* untuk algoritma FP-growth dan Tree Projection cenderung linier. Semakin banyak jumlah transaksi maka waktu pembangkitan *frequent pattern* semakin lama, demikian pula sebaliknya.

5.2 Saran

- 1) Karena performansi algoritma dipengaruhi oleh kebutuhan ruang dan waktu, maka perlu penelitian lebih lanjut untuk analisis kedua algoritma mengenai kompleksitas ruang.
- 2) Menganalisa performansi algoritma pada data yang memiliki banyak *itemset* pada setiap transaksi.
- 3) Menerapkan kedua algoritma pada fungsionalitas *data mining* selain asosiasi seperti *clustering*.
- 4) Menerapkan *multiple minimum support* pada algoritma FP-growth dan Tree Projection dan menganalisa performansinya.

DAFTAR PUSTAKA

- [1] Agarwal, Ramesh C.; Aggarwal Charu C., Prasad V.V.V., 2000, *A Tree Projection algorithm for Generation of Frequent Itemset*, Yorktown Heights: IBM T. J. Watson Research Center.
- [2] Agrawal, Rakesh; Imielinski Tomasz, Swami Arun, 1993, *Mining Association Rules between Sets of Items in Large Databases*, San Jose: IBM Almaden Research Center.
- [3] Agrawal, Rakesh; Srikant Ramakrishnan, 1994, *Fast Algorithms for Mining Association Rules*, San Jose: IBM Almaden Research Center.
- [4] Aho, Hopcroft Ullman, 1974, *The Design and Analysis of Computer Algorithms*, Addison-Wasley.
- [5] A. Kusters, Walter; Pijls Wim, Popova Viara, 2003. *Complexity Analysis of Depth First and FP-growth Implementations of APRIORI*, Netherlands: Leiden Institute of Advanced Computer Science Universiteit Leiden.
- [6] Booch, Grady; Rumbaugh James, Jacobson Ivar, 1999, *The Unified Modeling Language User Guide*, United States of America: Addison Wesley Longman, Inc
- [7] Grahne, Gösta Grahne; Zhu Jianfei. *Efficiently Using Prefix-trees in Mining Frequent Itemsets*, Canada : Concordia University Montreal
- [8] Han, Jawei; Pei Jian, Yin Yiwen, 2000, *Mining Frequent Patterns without Candidate Generation*, ACM SIGMOD Conference Proceedings.
- [9] Han, Jiawei; Kamber Micleline, 2001, *Data Mining : concepts and techniques*, United States of America: Morgan Kaufmann Publisher.
- [10] K. H. Tung, Anthony, *Towards a Unifying Framework for Frequent Pattern Mining Algorithm*, Singapore: Departement of Computer Science National University of Singapore.
- [11] Munir, Rinaldi, 2003, *Matematika Diskrit Edisi Kedua*, Bandung: Penerbit Informatika.
- [12] N.Popova, Viara, 2004, *Knowledge Discovery and Monotonicity*, Thesis, Rotterdam: Erasmus University Rotterdam.

- [13] Pei, Jian, 2002, *Pattern Growth Method for Frequent Pattern Mining*, Thesis, Computing Science, Simon Fraser University.
- [14] Pranata, Antony, 2003, *Pemrograman Borland Delphi 6 Edisi 4*, Yogyakarta: Penerbit Andi Yogyakarta.
- [15] Saswita, Yossy, 2004, *Penggunaan Data Mining Association Rules dengan Algoritma Frequent Pattern tree(FP-tree)*, Tugas Akhir S1, Bandung: Sekolah Tinggi Teknologi Telkom.
- [16] http://www.cse.cuhk.edu.hk/~kdd/data_collection.html, IBM_VC++.

