

ANALISIS DAN IMPLEMENTASI CLIQUE PARTITION UNTUK KLASTERISASI DATA

Fitriliani Hayu Puspasari¹, Adiwijawa², Moch Arif Bijaksana³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

-

Kata Kunci : -

Abstract

-

Keywords : -



Telkom
University

1. Pendahuluan

1.1 Latar belakang

Dewasa ini, ledakan data terjadi hampir di segala aspek kehidupan baik dalam bidang telekomunikasi, internet, ekonomi, biologi, sosiologi, dll. Dengan semakin banyak tumpukan data, semakin banyak pula informasi yang dapat diperoleh. Namun ada juga permasalahan yang timbul dari tumpukan data yang semakin banyak, yaitu proses untuk menggali data mentah menjadi informasi yang dibutuhkan. Permasalahan itu muncul karena tumpukan data tersebut tidak dapat digali dengan cara manual. Salah satu alternatif yang dapat digunakan untuk mengatasi masalah tersebut adalah dengan menggunakan klasterisasi.

Klasterisasi yaitu mengelompokkan data berdasarkan banyaknya kemiripan yang dimiliki data mentah ke dalam sebuah kelas. Kelas yang dihasilkan dari proses klasterisasi disebut dengan klaster. Objek - objek yang terdapat dalam satu klaster hendaknya memiliki banyak kemiripan antara satu dengan yang lainnya, sedangkan antara objek- objek yang terdapat dalam sebuah klaster dengan klaster yang lainnya haruslah memiliki sedikit kemiripan. Pada tugas akhir ini, akan diimplementasikan metode *clique partition* untuk mengklasterkan data. Proses klasterisasi yang dilakukan yaitu dengan membagi data yang telah direpresentasikan kedalam graf menjadi beberapa sub-graf, berdasarkan *clique* atau *independent-set* yang diperoleh dari graf tersebut.

Untuk mengubah data menjadi suatu graf, langkah pertama yang dilakukan yaitu menentukan elemen dari data yang menjadi simpul dan sisinya. Dimana setiap simpul akan merepresentasikan objek yang berbeda sedangkan sisi akan merepresentasikan keterhubungan antar objek tersebut. Dua buah objek dapat dikatakan terhubung jika nilai *cross-correlation* antara dua objek tersebut lebih besar atau sama dengan batas *threshold* θ .

Graf yang telah terbentuk dipecah menjadi beberapa sub-graf berdasarkan *clique* atau *independent set*. *Clique* itu sendiri yaitu suatu sub-graf yang lengkap, dimana semua simpulnya saling terhubung dengan semua simpul yang ada pada sub-graf tersebut. Sedangkan *independent-set* yaitu suatu sub-graf dimana simpul – simpulnya tidak ada yang terhubung. Setiap *clique* yang diperoleh bukan merupakan himpunan bagian dari *clique* yang lain yang lebih besar, dimana sering disebut dengan *maximal clique*. Menemukan *maximal clique* pada suatu graf G akan sama dengan menemukan *maximal independent-set* pada *complement* graf G tersebut.

1.2 Perumusan masalah

Pada tugas akhir ini, akan dicoba mengimplementasikan *clique partition* untuk menklasterkan data. Seperti yang telah dijelaskan diatas, bahwa sekumpulan data yang besar akan direpresentasikan terlebih dahulu menjadi suatu graf dimana setiap objek yang berbeda akan dipetakan menjadi sebuah simpul, dan setiap pasangan objek akan dihubungkan oleh sebuah sisi jika memiliki nilai korelasi lebih besar atau sama dengan nilai *threshold*. Klasterisasi dilakukan dengan

memecah graf yang telah terbentuk menjadi beberapa sub-graf yang lengkap atau dapat disebut dengan *clique*.

Batasan masalah pada penyusunan tugas akhir ini yaitu pada segi data yang akan diklasterkan, karena tidak semua data dapat direpresentasikan menjadi sebuah graf. Data yang digunakan hendaknya memiliki objek – objek yang saling berkorelasi sehingga dapat direpresentasikan menjadi sebuah graf.

Pada tugas akhir ini, akan dianalisis pengaruh perubahan *threshold* terhadap hasil klasterisasi.

1.3 Tujuan

Tujuan penyusunan tugas akhir ini meliputi :

- Mengimplementasikan *clique partition* pada sekumpulan data untuk klasterisasi.
- Menganalisis dampak perubahan *threshold* terhadap hasil klasterisasi data.
- Memberikan *summarizing* dan visualisasi dari data yang telah diklasterkan

1.4 Metodologi penyelesaian masalah

- Studi literatur
Mengumpulkan dan mempelajari semua bahan materi yang berkaitan dengan data mentah, proses mengubah data mentah menjadi suatu graf, proses mengklasterkan data dengan cara menemukan *clique* pada graf tersebut karena data akan dibagi berdasarkan data yang diperoleh.
- Pencarian dan pengumpulan data
Pencarian dan pengumpulan data ditujukan untuk memperoleh data mentah yang akan digunakan sebagai bahan penelitian.
- Implementasi metode terhadap data dan pembangunan perangkat lunak
Pada tahap ini, mengimplementasikan *clique partition* pada sekumpulan data mentah, kemudian menganalisa sesuai dengan tujuan yang telah disebutkan diatas.
- Analisa hasil
Untuk tahap akhir, akan dilakukan analisa terhadap hasil yang diperoleh. Salah satunya, analisis perubahan *threshold* terhadap hasil klasterisasi.
- Pembuatan laporan

5. Penutup

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari penelitian kali ini yaitu

- Perubahan threshold memiliki pengaruh terhadap hasil klasterisasi dimana jumlah klasternya akan semakin besar jika nilai thresholdnya semakin besar. Jika nilai threshold semakin besar maka objek yang ada di dalam klaster semakin banyak yang berjumlah satu buah. Dan jika dibutuhkan informasi mengenai objek – objek yang saling berkorelasi maka nilai threshold sama dengan satu tidak disarankan. Sedangkan pengaruh perubahan threshold terhadap *maximum clique* berbanding terbalik, semakin besar nilai threshold maka semakin kecil jumlah objek yang memiliki korelasi satu dengan yang lainnya.
- Hasil klasterisasi yang dilakukan pada data penjualan PT. Anugrah Sentosa berupa klaster – klaster yang di dalamnya terdapat item – item yang memiliki rata – rata fluktuasi penjualan yang tidak jauh berbeda. Informasi mengenai item – item yang dikelompokkan menjadi satu klaster dapat digunakan untuk acuan dalam *take place* item – item tersebut.

5.2 Saran

Perlu dilakukan penelitian lebih lanjut mengenai kualitas dari hasil klasterisasi data yang telah diperoleh.

Daftar Pustaka

- [1] Bomze IM, Budinich M, Pardalos PM, Pelillo M., 1999, "The maximum clique problem", In: Du D-Z, Pardalos PM, editors. Handbook of combinatorial optimization. Dordrecht: Kluwer Academic Publishers. p. 1–74.
- [2] Camilo Rostoker, 2006, "Interactive Visualization of the Market Graph", University of British Columbia, Department of Computer Science.
- [3] Daniel Baum, 2003, "Finding All Maximal Cliques of a Family of Induced Subgraphs".
- [4] Eric R Haley, 2003, "Graph Algorithms for Assembling Integrated Genome Maps", University of Toronto, Departement of Computer Science.
- [5] Iko Pramudiono, 2003, "Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data ", Kuliah Umum IlmuKomputer.Com.
- [6] Pang-Ning Tan, Michael Steinbach, and Kumar Vipin, 2005, "Introduction To Data Mining", Boston, Pearson Education Inc.
- [7] Panos M. Pardalos, 2006, "Data Mining and Knowledge Discovery in Dynamic Networks", McKnight Brain Institute, Computer & Information Science & Engineering Departement.
- [8] Rinaldi Munir, 2006, "Matematika Diskrit", Bandung, Informatika.
- [9] Sergiy Butenko, 2003, "Maximum Independent Set and Related Problems with Applications", University of Florida
- [10] Vladimir Boginski, Sergiy Butenko, and Panos M. Pardalos, 2005, "Mining market data: A network approach".
- [11] Volker Stix, 2001, " Finding all maximal cliques in dynamic graphs", Vienna University of Economics, Department of Information Business.
- [12] [http://en.wikipedia.org/wiki/Clique_\(graph_theory\)](http://en.wikipedia.org/wiki/Clique_(graph_theory)) , tanggal 9 Juli 2007
- [13] <http://en.wikipedia.org/wiki/Correlation> , tanggal 9 Juli 2007
- [14] http://en.wikipedia.org/wiki/Maximal_clique , tanggal 9 Juli 2007

Telkom
University