

Analisa Data Menggunakan Algoritma ACO (*Ant Colony Optimization*)

Harun

Politeknik Ganesha Medan
Jl. Veteran No. 190 Pasar VI Manunggal

Abstrak — Perlu disadari pada saat ini, banyak perusahaan yang memiliki data dalam jumlah yang besar. Data dalam jumlah besar tersebut ternyata dapat dimanfaatkan untuk meningkatkan kinerja perusahaan. Untuk itu diperlukan proses Data Mining. Salah satu metode dalam Data Mining adalah klasifikasi. Klasifikasi bertujuan untuk memperoleh pola tertentu, dalam bentuk tree, aturan klasifikasi atau model matematis. Untuk memperoleh pola tersebut diperlukan algoritma tertentu. Salah satunya adalah dengan ACO (*Ant Colony Optimization*) atau dengan nama lain *Ant Miner*. Pembahasan ini dianalisa pengaruh perubahan yang dilakukan pada *Ant_Miner2* dan *Ant_Miner3* terhadap tingkat akurasi dan simplisitas aturan yang dihasilkan, serta parameter sistem yang mempengaruhinya. Hasilnya tingkat akurasi *Ant_Miner* lebih baik daripada C5.0, sementara simplisitas aturan yang dihasilkan tidak jauh berbeda.

Kata Kunci — *Data Mining, Ant Colony Optimization, Ant_Miner, Ant_Miner3*

I. PENDAHULUAN

Melihat dari perkembangannya pada saat ini setiap perusahaan sudah memiliki data dalam jumlah yang besar. Data tersebut biasanya hanya disimpan di dalam tempat penyimpanan dan digunakan untuk keperluan tertentu. Data tersebut dapat digunakan lagi, terutama untuk mengevaluasi kinerja perusahaan atau bahkan membantu mengambil keputusan pihak manajemen perusahaan. Semua itu dapat dilakukan dengan menerapkan Data Mining.

Data mining adalah sebuah proses mengolah data yang didalamnya menerapkan suatu metode tertentu untuk mendapatkan suatu pola atau knowledge dari suatu data. Data mining terbagi ke dalam beberapa bagian yaitu klasifikasi, clustering, regresi, asosiasi dan lain-lain. Klasifikasi bertujuan untuk mencari aturan klasifikasi dari suatu data. Aturan klasifikasi yang dihasilkan dari proses klasifikasi berupa pernyataan if-then seperti :

IF <kondisi> THEN <class>

Pada bagian IF (antecedent) terdiri atas kumpulan kondisi. Setiap kondisi disebut term. Setiap term memiliki bentuk <atribut,operator,nilai> misalnya <Gender=male>. Bagian THEN (consequent) menyatakan class yang ditujukan untuk semua case yang memenuhi semua kondisi pada bagian IF.

Banyak sekali algoritma yang dapat diterapkan didalam klasifikasi. Setiap algoritma dapat kita bandingkan dari beberapa hal, diantaranya akurasi prediktif dan simplisitas aturan.

Algoritma semut sering dipakai dalam menyelesaikan masalah kombinatorial seperti TSP (*Travelling Salesman Problem*) dan penentuan jadwal kuliah. Namun algoritma semut masih tergolong baru dalam penggunaannya untuk data mining, khususnya klasifikasi. Alex F. Freitas dan kawan-kawannya merupakan yang pertama kali menerapkan algoritma semut di dalam data mining metode klasifikasi. Mereka menyebut algoritmanya dengan *Ant_Miner*. Algoritma *Ant_Miner* ini kemudian diperbaiki oleh Bo Liu dan kawan-kawan. Mereka mengubah rumus heuristic function menjadi lebih sederhana. Algoritma ini disebut dengan *Ant_Miner2*

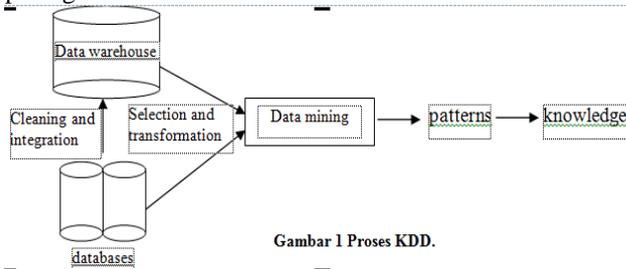
Untuk mengetahui apakah perubahan yang dilakukan mampu meningkatkan akurasi dan simplisitas aturan yang dihasilkan, maka kita merasakan perlu untuk menganalisa pengaruh perubahan yang terjadi didalam *Ant_Miner2* dan pengembangan *Ant_Miner3* terhadap tingkat akurasi dan simplisitas aturan yang dihasilkan. Selain itu, untuk melihat apakah *Ant_Miner2* mampu bersaing dengan algoritma klasifikasi lain yang telah ada, maka dilakukan perbandingan dengan algoritma C5.0 yang sering dipakai di dalam Data Mining khususnya klasifikasi.

II. LANDASAN TEORI

A. Data Mining

Data mining adalah suatu proses mengolah data-data yang ada dengan metode dan algoritma tertentu, yang bertujuan mengekstrak pola yang berguna dari data yang besar tersebut [1]. Banyak

sekali yang menyamakan data mining dengan KDD (Knowledge Discovery in Database). Padahal sebenarnya data mining hanyalah sebuah langkah di dalam proses KDD [1]. Seperti yang diperlihatkan pada gambar 1 dibawah ini :



Gambar 1 Proses KDD.

Data Mining memiliki beberapa metode, diantaranya adalah klasifikasi, asosiasi, clustering, regresi dan lain-lain. Masing-masing metode tersebut menggunakan algoritma-algoritma tertentu dalam mengolah data dan menghasilkan pola tertentu

B. Klasifikasi

Klasifikasi adalah proses menemukan sekumpulan model (atau fungsi) yang menjelaskan dan membedakan data ke dalam kelas-kelas tertentu, dengan tujuan menggunakan model tersebut dalam menentukan kelas dari suatu objek yang belum diketahui kelasnya. Model tersebut didapatkan dengan melakukan analisa dari sekumpulan data yang disebut **training data**. Model dapat direpresentasikan dalam berbagai bentuk, seperti aturan klasifikasi (IF-THEN), decision tree, rumus matematis, atau jaringan syaraf.

Tabel 1. Contoh Ttraining Data

Blood type	Can fly	Lives in water	Hibernates	Class
Warm	No	No	No	Mammal
Cold	No	No	Yes	Reptile
Cold	No	Yes	No	Fish
Cold	No	Sometimes	Yes	Amphibian
Warm	Yes	No	Yes	Mammal

C. Ant Colony Optimization

Algoritma Ant Colony Optimization (ACO) ialah sebuah sistem yang berdasarkan agen-agen yang mensimulasikan perilaku dari semut-semut, termasuk mekanisme kerja sama dan adaptasi.

Algoritma ACO dibuat berdasarkan gagasan berikut :

1. Setiap jalur yang dilalui oleh semut diasosiasikan dengan kandidat solusi dari suatu masalah.

2. Jika suatu semut mengikuti suatu jalur, jumlah pheromone pada jalur tersebut sebanding dengan kualitas kandidat solusi yang bersangkutan.
3. Jika suatu semut diharuskan untuk memilih antara dua jalur, jalur yang memiliki jumlah pheromone lebih banyak memiliki peluang lebih besar untuk dipilih semut tersebut.

Hasilnya, semut-semut akan memusat ke jalur yang pendek, dengan harapan menghasilkan solusi yang optimal atau mendekati optimal.

Intinya, desain algoritma ACO berdasarkan hal-hal di bawah ini :

1. Dapat merepresentasikan masalah dengan baik, dimana semut-semut secara bertahap membangun dan mengubah solusi dengan menggunakan aturan
2. transisi probabilistik, berdasarkan jumlah pheromone pada suatu jalur dan sebuah fungsi heuristik.
3. Sebuah fungsi heuristik yang sesuai dengan masalah (η) yang menyatakan kualitas item yang akan ditambahkan ke solusi sementara.
4. Aturan dalam melakukan update jumlah pheromone, yang menjelaskan bagaimana mengubah jumlah pheromone (τ) suatu jalur.
5. Aturan transisi probabilistik berdasarkan nilai fungsi heuristik (η) dan jumlah pheromone (τ) yang akan digunakan dalam membangun suatu solusi

D. Ant_Miner : Algoritma ACO untuk Data Mining

Algoritma ACO yang diterapkan dalam Data Mining disebut dengan Ant_Miner. Ant_Miner sampai saat ini telah dikembangkan menjadi Ant_Miner2 dan Ant_Miner3. Ketiga algoritma tersebut akan dibahas pada bagian ini.

E. Ant_Miner

Dalam algoritma ACO setiap semut secara bertahap membangun / emngubah sebuah solusi dari masalah yang diberikan, pada kasus ini adalah pencarian aturan klasifikasi. Aturan klasifikasi memiliki bentuk sebagai berikut :

$$IF <term1 AND term2 AND ...> THEN <class>$$

Masing-masing term merupakan gabungan <atribut,operator,nilai>, dimana nilai menyatakan nilai dari suatu atribut. Operator yang digunakan disini selalu “=”.

F. Analisa Sistem

Pada bagian ini akan dilakukan analisa sistem yang akan dibangun. Ada 2 aspek yang akan dibahas yaitu mengenai kebutuhan sistem dan batasan perancangan sistem.

III. ANALISA HASIL

Pada bagian ini akan dijelaskan parameter pengujian yang dipakai, karakteristik dataset yang digunakan dalam pengujian dan analisis hasil pengujian yang didapatkan.

A. Parameter Pengujian

Pada bagian ini akan dilakukan pengujian sistem yang telah dibuat. Dari hasil pengujian akan diperoleh aturan klasifikasi dari setiap Trainset serta parameter performansinya. Adapun parameter yang digunakan sebagai pengukur performansi adalah :

1. Akurasi prediktif
Merupakan parameter untuk mengukur ketepatan aturan klasifikasi yang dihasilkan dalam mengklasifikasikan data di dalam Testset berdasarkan atribut yang ada ke dalam kelasnya. Dinyatakan dalam persentase (0-100 %). Semakin tinggi nilainya, maka semakin baik aturan yang dihasilkan dalam mengklasifikasikan data baru ke dalam kelas yang benar.
2. Sederhananya aturan
Merupakan parameter yang menyatakan kesederhanaan aturan yang dihasilkan. Parameter ini dinyatakan dalam jumlah rata-rata tiap term pada bagian antisenden setiap aturan. Semakin kecil nilainya maka semakin sederhana aturan yang dihasilkan, sehingga semakin mudah untuk dipahami.

Selain itu, akan dilakukan analisa juga mengenai parameter sistem yang digunakan didalam Ant_Miner2, yaitu

1. *No_of_ants* : jumlah maksimal semut yang digunakan, berarti juga jumlah maksimal kandidat aturan yang dihasilkan
2. *Min_cases_per_rule* : jumlah case terkover yang harus dipenuhi oleh setiap kandidat aturan yang dihasilkan
3. *Max_uncovered_cases* : jumlah maksimal case yang tidak terkover pada Trainingset setelah dijalankan algoritma Ant_Miner2
4. *No_rules_converg* : jumlah aturan yang menyatakan bahwa semut-semut Ant_Miner2 telah konvergen ke satu jalur.

5. ρ (*pheromone evaporation*) : menyatakan tingkat penguapan pheromone.

B. Karakteristik Dataset

Dataset yang digunakan disini adalah dataset dari repository UCI (University of California). Dataset yang dipilih adalah Breast Cancer, Tic-tac-toe, dan House Votes. Ketiga dataset ini dipilih karena atributnya tidak ada yang kontinu, dan tidak ada nilai yang hilang (*missing values*) atau null di dalamnya, serta data didalamnya acak. Karena aplikasi See5 yang digunakan sebagai pembanding masih berupa versi Demo, maka penggunaan dataset masih dibatasi yaitu hanya sebanyak 400 case

Tabel 2. Karakteristik Dataset

Dataset	Jumlah kasus	Jumlah atribut	Jumlah kelas
Breast Cancer	390	10	2
Tic-tac-toe	400	10	2
House Votes	232	17	2

C. Analisa Hasil Pengujian

Pada bagian ini akan dijelaskan mengenai hasil analisa akurasi prediktif dan simplisitas Ant_Miner3 dan See5, pengaruh penggunaan pruning dan penggunaan pheromone serta pengaruh parameter sistem terhadap akurasi dan simplisitas aturan yang dihasilkan.

D. Analisa Sederhananya Aturan

Sederhananya aturan didapat dengan menghitung aturan yang dihasilkan dan rata-rata term yang terdapat di bagian antisenden aturan

Tabel 3. Sederhananya Aturan Ant_Miner3 dan See5

Dataset	Ant_Miner3		See5	
	Jml aturan	Jml term	Jml aturan	Jml term
reast Cancer	15.00	1.02	14.40	1.15
Tic-tac-toe	9.90	1.06	21.70	2.89

House Votes	5.70	1.25	2.40	1.15
-------------	------	------	------	------

Jumlah aturan yang dihasilkannya lebih banyak daripada See5. Ini membuktikan bahwa Ant_Miner3 juga tidak jauh berbeda dengan See5 dalam hal simplisitas aturan.

Dari tabel berikut terlihat bahwa proses pruning tidak memiliki pengaruh yang cukup besar pada akurasi prediktif Ant_Miner3. Hal ini dapat dipahami sebab tujuan utama pruning adalah menghilangkan / menyederhanakan aturan yang dihasilkan, dengan mempertahankan akurasi prediktifnya.

Tabel 4. Perbandingan Simplisitas Aturan Ant_Miner3 dengan Pruning dan Tanpa Pruning

Dataset	Jml aturan		Jml term	
	Pruning	Tanpa pruning	pruning	Tanpa pruning
Breast Cancer	15.00	15.90	1.02	1.06
Tic-tac-toe	9.90	23.50	1.06	1.85
House Votes	5.70	7.90	1.25	1.87

Dari tabel 4 terlihat bahwa penggunaan pruning dapat menurunkan jumlah term pada aturan yang ditemukan. Hal ini sangat penting sebab semakin sederhana aturan yang ditemukan, maka semakin mudah bagi kita untuk memahaminya.

Sehingga dari kedua tabel diatas dapat ditarik kesimpulan bahwa pruning diperlukan untuk menyederhanakan

Aturan yang dihasilkan tanpa terlalu mempengaruhi tingkat akurasi yang dihasilkan.

Analisa Pengaruh Pheromone

Untuk menganalisa pengaruh pheromone terhadap performansi Ant_Miner3, jumlah pheromone diset konstan 1. dengan demikian setiap semut hanya dipengaruhi oleh nilai heuristik term.

Tabel 5. Perbandingan Akurasi Prediktif Ant_Miner3 dengan Pheromone dan Tanpa Pheromon

Dataset	Akurasi Prediktif (%) dengan pheromone	Akurasi Prediktif (%) tanpa pheromone
Breast Cancer	97.18	95.89
Tic-tac-toe	96.50	95.75
House Votes	98.69	96.52

Tabel 6. Perbandingan Simplisitas Aturan Ant_Miner3 dengan Pheromone dan Tanpa Pheromone

Dataset	Jml aturan		Jml term	
	Pheromone	Tanpa pheromone	Pheromone	Tanpa pheromone
Breast Cancer	15.00	15.00	1.02	1.02
Tic-tac-toe	9.90	8.50	1.06	1.08
House Votes	5.70	5.30	1.25	1.19

Dari kedua tabel diatas terlihat bahwa simplisitas aturan tidak terlalu terpengaruh oleh penggunaan pheromone. Namun, penggunaan pheromone dapat meningkatkan akurasi prediktif. Hal ini dapat dipahami sebab dengan digunakannya pheromone, maka proses pembangunan kandidat aturan dipengaruhi oleh dua hal, yaitu nilai heuristik dan nilai pheromone, sehingga akan meningkatkan nilai akurasi prediktif.

Analisa Parameter Sistem

Ant_Miner3 menggunakan lima buah parameter sistem yaitu :

1. *No_of_ants*
2. *Min_cases_per_rule*
3. *Max_uncovered_case*
4. *No_rules_converg*
5. *Pheromone evaporation* (ρ)

Pertama akan dianalisis pengaruh `no_of_ants` dan `no_rules_converg`. Dua parameter ini saling berkaitan dalam menentukan banyaknya kandidat aturan yang akan dihasilkan oleh semut `Ant_Miner3`. Apabila nilai `no_rules_converg` terlalu rendah, maka semut `Ant_miner3` akan terlalu cepat konvergen ke satu jalur, sehingga tidak semua semut akan menghasilkan kandidat aturan. Sehingga dengan memberikan nilai `no_rules_converg` yang tinggi, akan menyebabkan semut `Ant_Miner3` tidak terlalu cepat konvergen.

Untuk menganalisa pengaruh `no_of_ants` dan `no_rules_converg` digunakan empat kombinasi yang berbeda yaitu `no_of_ants` (1000 dan 3000) dan `no_rules_converg` (5 dan 10).

Tabel 7. Siplisitas Aturan `Ant_Miner3` untuk kombinasi pheromone evaporation

Paramet er Dataset	Jml aturan (ρ)					Jm term (ρ)				
	0	0	0	0	0	0	0	0	0	0
	1	3	5	7	9	1	3	5	7	9
Breast Cancer	7	7	7	7	7	1	1	1	1	1

	2	6	6	8	5	0	0	0	0	0
Tic-tac- toe	8	7	8	8	8	1	1	1	1	1

	2	9	1	3	0	0	0	0	0	0
House Votes	3	3	3	3	3	1	1	1	1	1

	6	5	8	4	8	1	1	1	1	1
	0	0	0	0	0	3	8	9	6	6

Dengan melihat hasil yang diperoleh pada tabel diatas, maka dapat ditarik kesimpulan bahwa simplisitas `Ant_Miner3` sangat dipengaruhi adanya pruning. Dengan adanya pruning maka simplisitas aturan meningkat. Sementara untuk parameter sistem tidak ada yang mempengaruhi simplisitas secara signifikan.

IV. KESIMPULAN

Algoritma `Ant_Miner3` dapat digunakan untuk menemukan aturan klasifikasi pada dataset Breast Cancer, Tic-tac-toe dan House Votes.

1. Algoritma `Ant_Miner3` mengurangi kasus terhentinya algoritma `Ant_Miner` dengan penggunaan bilangan random.
2. Dalam hal akurasi prediktif dan simplistas aturan, `Ant_Miner3` kompetitif dengan algoritma C5.0 yang banyak digunakan pada aplikasi data mining.
3. Akurasi `Ant_Miner3` dapat ditingkatkan dengan beberapa hal, diantaranya menggunakan pheromone, memberikan nilai yang semakin besar untuk parameter `no_of_ants` dan `no_rules_converg` serta pheromone evaporation.
4. Siplisitas aturan `Ant_Miner3` dapat ditingkatkan dengan beberapa hal, diantaranya menggunakan pruning, memberikan nilai yang semakin besar untuk parameter `max_uncovered_case` dan `min_cases_per_rule`.

REFERENSI

- [1] Han, Jiawei dan Micheline Kamber 2001. *Data Mining Concepts and Techniques*. San Diego : Academic Press.
- [2] Parpinelli, Rafael S. dan Heitor S. Lopes dan Alex A. Freitas. *Data Mining with an Ant Colony Optimization*. Brazil.
- [3] Maniezzo, Vittorio dan Luca Maria Gambardella dan Fabio de Luigi. 2001. *Ant Colony Optimization*. European Commission.
- [4] Effendi, Arya Bima. 2004. *Klasifikasi pada Data Mining Menggunakan Algoritma Ant Colony Optimization*. Bandung : STT Telkom.
- [5] Berry, Michael J.A. dan Gordon S. Linoff. 2000. *Mastering Data Mining*. USA : John Wiley & Sons, Inc.
- [6] Middendorf, Martin dan Frank Reischle dan Hartmut Schmeck. *Information Exchange in Multi Colony Ant Algorithm*. Karlsruhe, Germa