

KLASIFIKASI PENGGUNA WEB DALAM WEB USAGE MINING UNTUK BUSINESS INTELLIGENCE DENGAN ALGORITMA ANT COLONY OPTIMIZATION

Abdurrahman¹, Bambang Riyanto T.², Rila Mandala³, Rajesri Govindaraju⁴

¹STMIK BANDUNG

^{2,3}Sekolah Teknik Elektro & Informatika, Institut Teknologi Bandung

⁴Fakultas Teknologi Industri, Institut Teknologi Bandung

¹mr.indonesian@gmail.com, ²briyanto@lskk.ee.itb.ac.id, ³rila@informatika.org,

⁴rajesri_g@mail.ti.itb.ac.id

Abstrak

Web Usage Mining (WUM) merupakan pemanfaatan metode *data mining* untuk melakukan penemuan pengetahuan dari data penggunaan web. Salah satu fungsi WUM adalah untuk kebutuhan *Business Intelligence* (BI), dimana informasi penting yang dibutuhkan untuk BI adalah adanya klasifikasi pengguna web yang dapat digunakan untuk aktifitas akuisisi, penetrasi, dan retensi pengguna. **Permasalahan utama yang dihadapi dalam melakukan klasifikasi pengguna web penentuan atribut sebagai term anteseden kaidah klasifikasi, dimana hal ini merupakan permasalahan utama dalam fungsi klasifikasi data mining pada umumnya.** Metode klasifikasi pengguna web yang diusulkan adalah pengembangan metode klasifikasi berbasis metode *ant colony optimization* (ACO) yang merupakan sistem cerdas terdistribusi dengan menggunakan fungsi heuristik sesuai dengan ranah permasalahannya. **Metode yang dikembangkan adalah dengan melakukan modifikasi fungsi heuristik algoritma Ant-Miner.** Fungsi heuristik yang diusulkan yaitu heuristik berbasis penggunaan web dengan menggunakan *gain ratio* atribut jumlah akses dan rata-rata durasi akses pengguna web. Pengujian dilakukan dengan membandingkan performansi algoritma yang diusulkan dengan algoritma *Ant-Miner*. **Dari hasil pengujian diperoleh bahwa performansi algoritma yang diusulkan mempunyai tingkat akurasi kaidah dan simplifikasi kaidah yang lebih tinggi dibandingkan dengan algoritma Ant-Miner.**

Kata Kunci : web usage mining, klasifikasi pengguna web, ant colony optimization, algoritma Ant-Miner, fungsi heuristik, tingkat akurasi kaidah, simplifikasi kaidah.

Abstract

Web Usage Mining (WUM) is an implementation of data mining method for knowledge discovery of web usage data. One of WUM functions is applied for Business Intelligence (BI) to provide web user classification for acquisition, penetration, and retention of web user. The challenges in web user classification is how to determine attributes as term candidates in rule antecedent, that it was a big problem in classification task in data mining. We proposed a development of ant colony optimization (ACO) method as distributed intelligent system that uses problem dependent heuristic to solve the problem. We proposed a heuristic functions for web user classification, based on web usage data by using of gain ratio from attribute of total number of web user access and mean of access duration of web user. The testing is conducted to compare performance of proposed method with Ant-Miner algorithm and proposed algorithm. The testing indicated that proposed method generates rule set with accuracy rate higher and simpler than rule set that generated by Ant-Miner algorithm.

Keywords : web usage mining, web user classification, ant colony optimization, Ant-Miner algorithm, heuristic function, preprocessing, accuracy rate, rule simplification.

1. Pendahuluan

Interaksi pengguna (*user*) dengan web menghasilkan data akses web yang sangat banyak dalam periode waktu tertentu yang tersimpan dalam file *web access log* di server. Data hasil interaksi ini diharapkan dapat memberikan informasi yang bernilai bagi pengelola web dalam rangka memasarkan keberadaan webnya dan produk yang dijualnya. Dalam konteks inilah *web usage mining* (WUM) mempunyai peran dalam menemukan pengetahuan (*knowledge discovery*) dari data pemakaian web tersebut. Disamping data *web access*

log, data yang terbentuk dari interaksi pengguna dengan web e-commerce adalah data profil pengguna dan data transaksi.

1.1 Overview Web Usage Mining

Web mining merupakan aplikasi teknik *data mining* untuk mengekstrak pengetahuan (*knowledge*) dari data web [4]. Ada dua pendekatan yang digunakan untuk mendefinisikan *web mining*, yaitu pendekatan berbasis proses (*process-centric view*) yang mendefinisikan *web mining* sebagai

sekumpulan aktivitas (*sequence of tasks*). Yang kedua adalah pendekatan berbasis data (*data-centric view*) yang mendefinisikan *web mining* sebagai terminologi tipe data web yang digunakan untuk proses *data mining*. Dalam paper ini pendekatan yang digunakan adalah pendekatan kedua.

1.2 Penelitian Terkait

Penelitian terkait dengan riset ini adalah....

2. Pembahasan

Algoritma *Ant-WUM* merupakan pengembangan algoritma *Ant-Miner* dari sisi fungsi heuristik. Dalam pembahasan ini kami akan menjelaskan secara utuh mengenai tahapan-tahapan algoritma *Ant-WUM* yang mengadopsi algoritma *Ant-Miner*.

2.1 Metodologi

Metodologi yang digunakan adalah.....

2.2 Pembahasan Algoritma yang diusulkan

Algoritma ini menggunakan pendekatan sekuensial untuk menemukan sejumlah kaidah klasifikasi untuk melingkupi data training (*training data*). Iterasi pada pengulangan REPEAT-UNTIL pada Algoritma *Ant-Miner* (Gambar.1) terdiri dari tiga tahapan, yaitu pembuatan kaidah, *rule pruning* (pembuangan kaidah yang tidak sesuai) dan updating *pheromone*.

$$P_{ij} = \frac{\eta_{ij} \cdot \tau_{ij}(t)}{\sum_{i=1}^a x_i \cdot \sum_{j=1}^{b_i} (\eta_{ij} \cdot \tau_{ij}(t))} \quad (1)$$

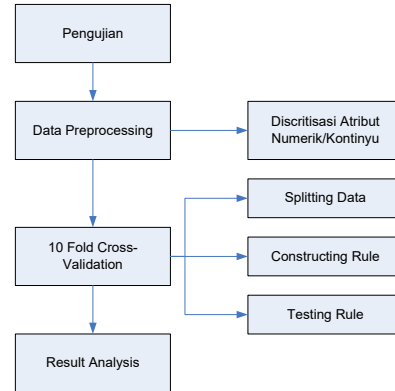
dimana:

- η_{ij} = nilai fungsi heuristik untuk $term_{ij}$ yang dihasilkan oleh persamaan (14). Semakin besar nilai η_{ij} maka semakin relevan untuk klasifikasi $term_{ij}$ dan semakin besar probabilitasnya untuk dipilih.
- $\tau_{ij}(t)$ = jumlah *pheromone* yang diasosiasikan dengan $term_{ij}$ pada iterasi t berkorespondensi dengan jumlah *pheromone* eksisting yang tersedia pada posisi jalur i,j yang diikuti oleh agen semut eksisting.
- a = jumlah atribut
- x_i = bernilai 1 jika atribut A_i belum digunakan oleh agen semut eksisting, dan bernilai 0 jika sebaliknya
- b_i = jumlah nilai dalam domain atribut ke- i
- W adalah kelas atribut (atribut yang domainnya *-WUM*, permasalahan yang akan dipecahkan adalah sama dengan permasalahan fungsi klasifikasi dalam domain *data mining* yaitu mencari pasangan atribut-nilai sebagai kandidat *term* dalam anteseden kaidah klasifikasi. Fungsi heuristik berbasis penggunaan web ini hanya dikhususkan untuk domain WUM dalam fungsi klasifikasi, dan tidak

dapat digunakan untuk dataset yang tidak mengandung atribut “jumlah akses” dan “rata-rata akses”. Berbeda dengan fungsi heuristik lainnya, seperti heuristik entropi menggunakan teori informasi yang digunakan oleh *Ant-Miner*

1.3 Pengujian

Pengujian dilakukan dengan skenario sebagai berikut:



Gambar 2. Skenario Pengujian

Pada bagian ini akan dijelaskan mengenai hasil pengujian algoritma *Ant-WUM* dalam membangun kaidah klasifikasi terhadap data penggunaan web. Pengujian dilakukan dengan melakukan perbandingan antara *Ant-WUM* dengan algoritma *Ant-Miner*.

Setting parameter yang digunakan dapat dideskripsikan dalam Tabel 1. sebagai berikut :

Tabel 1. Setting Parameter Pengujian

Nama Parameter	Deskripsi	Nilai
Folds	Jumlah <i>run</i> algoritma	10
Number of Ants	Jumlah semut	5
Min-cases Per Rule	Jumlah minimal data yang diingkupi satu kaidah	5
Max-uncovered cases	Jumlah maksimal data yang tidak dilingkupi oleh kaidah	10
Rules of convergence	Jumlah konvergensi semut dalam membangun kaidah	10
Number of Iterations	Jumlah iterasi	100

3. Kesimpulan dan Saran

Ant-WUM merupakan pengembangan ACO untuk melakukan klasifikasi pengguna web dalam domain WUM yang dengan fungsi kheuristik yang spesifik untuk permasalahan *web usage mining*. Heuristik berbasis penggunaan web, dimana heuristik ini menggunakan atribut entropi kandidat *term* dan *gain ration* jumlah akses dan rata-rata akses sebagai atribut dominan dalam mengidentifikasi tingkat frekuensi dan potensial pengguna web. Fungsi

heuristik ini bersifat khusus yang hanya ditujukan untuk klasifikasi pengguna web dalam domain WUM. Dari hasil pengujian diperoleh bahwa pemanfaatan heuristik ini untuk klasifikasi pengguna web menghasilkan kaidah klasifikasi dengan tingkat akurasi yang lebih tinggi dan jumlah kaidah yang lebih sedikit dibandingkan dengan pemanfaatan heuristik berbasis teori informasi maupun prinsip kedekatan.

Beberapa tindakan lanjutan diusulkan baik sebagai bahan penelitian lanjutan maupun untuk meningkatkan kesempurnaan penelitian ini, antara lain:

- a. Perlu dilakukan pengoleksian dataset latih untuk WUM. Dengan tersedianya berbagai dataset ini, maka pengujian terhadap metode klasifikasi *Ant-WUM* dapat memberikan hasil penelitian yang lebih akurat.
- b. Pengembangan *Ant-WUM* dari fungsi heuristik dan updating *pheromone* dan dalam menghadapi data numeric.

Daftar Pustaka

- [1] Parpinelli R.S., 2002, *Data Mining with an Ant Colony Optimization Algorithm*, IEEE Transaction on Evolutionary Computation, special issue on Ant Colony Algorithm, v.6, p.321-332
- [2] Ramadhan H., 2005, *A Classification on Techniques for Web Usage Analysis*, Journal of Computer Science 1(3), 413-418, Science Publication.