

1. PyClustering

Pustaka PyClustering adalah pustaka data mining Python dan C++ yang berfokus pada analisis kluster. Ini sangat relevan untuk algoritme yang didasarkan pada jaringan osilasi, yang dinamikanya diatur oleh sistem persamaan diferensial. Jika dukungan untuk kompiler C++ tidak terdeteksi, PyClustering kembali ke implementasi Python murni dari semua kernel. Untuk meningkatkan kinerja implementasi Python, PyClustering menggunakan pustaka NumPy untuk manipulasi arraynya.

Fungsionalitas visualisasi mencakup plot 2D dan 3D dari penyisipan cluster, segmen gambar, dan, dalam kasus jaringan osilasi, grafik proses sinkronisasi. Pustaka PyClustering tersedia di PyPi dan dari repositori github. PyClustering menyediakan kode yang terdokumentasi lengkap untuk setiap versi pustaka, termasuk contoh, deskripsi matematika dan algoritme, serta petunjuk penginstalan.

2. The density-based clustering method for privacy-preserving data mining

Karena penambahan data yang menjaga privasi adalah tugas non-sepele, yang juga berkaitan dengan masalah NP-hard, beberapa algoritma evolusioner disajikan untuk menemukan solusi yang dioptimalkan tetapi kebanyakan dari mereka fokus pada mempertimbangkan fungsi tujuan tunggal dengan pra- nilai berat yang ditentukan dari tiga efek samping. Dalam makalah ini, kami bertujuan untuk merancang metode optimasi segerombolan partikel objektif ganda untuk menyembunyikan informasi sensitif berdasarkan pendekatan pengelompokan kepadatan. CMPSO yang disajikan lebih fleksibel untuk memilih solusi yang paling tepat untuk menyembunyikan informasi sensitif berdasarkan preferensi pengguna. Meskipun teknik data mining dapat menemukan informasi dan pengetahuan yang berguna, namun mengungkapkan pesan rahasia atau pribadi dari informasi yang ditemukan, yang dapat menyebabkan ancaman keamanan. Data mining yang menjaga privasi telah menjadi isu penting dalam beberapa dekade terakhir karena tidak hanya dapat menyembunyikan informasi pribadi tetapi juga menemukan informasi yang diperlukan dengan berbagai teknik data mining. Sanitasi data adalah salah satu penelitian utama PPDM yang mengganggu database untuk

menyembunyikan informasi sensitif. Selama proses sanitasi, biasanya menyebabkan tiga efek samping, misalnya, kegagalan persembunyian, biaya yang hilang, dan biaya buatan.

3. A Survey of Distance Metrics in Clustering Data Mining Techniques

Teknik

Akhir-akhir ini, seiring dengan bertambahnya ukuran database, beberapa aspek telah dipelajari secara mendetail, seperti pengelompokan, pencarian tetangga terdekat dan metode identifikasi lainnya. Telah ditemukan bahwa dalam ruang multidimensi, konsep jarak tidak menawarkan kinerja tinggi. Tujuan dari makalah ini adalah menggunakan analisis kluster untuk mengidentifikasi dampak metrik jarak dalam pendeteksian gaya arsitektur menggunakan teknik Data Mining. Ada dua alasan utama untuk menggunakan pemrosesan gambar.

Kedua, memungkinkan ekstraksi informasi dari gambar, yang merupakan data input untuk sistem pengenalan bentuk dan klasifikasinya.

Dalam studi ini kami menganalisis dampak metrik jarak untuk mengidentifikasi gaya arsitektur dalam sebuah gambar, yang bertujuan untuk mencapai akurasi yang lebih tinggi. Metrik jarak Manhattan dan Euclidean.

Dalam clustering, kita mulai dari sekumpulan data yang tidak harus terstruktur dan berdasarkan ukuran kemiripan/ketidakmiripan yang akan digunakan antar data yang membangun matriks kesamaan/ketidakmiripan. Sebuah model yang paling menggambarkan bagaimana pengelompokan data dikelompokkan sehingga data milik cluster yang sama lebih mirip dengan satu milik cluster yang berbeda.