

LECTURE NOTES

COMP8014

Knowledge Data Discovery

Cluster Analysis: Basic Concepts

LEARNING OUTCOMES

LO4: Mahasiswa diharapkan mampu untuk menentukan metode data mining yang sesuai dengan sifat kealamiahannya permasalahan yang dihadapi.

LO5: Mahasiswa diharapkan mampu untuk mengimplementasikan metode data mining.

OUTLINE MATERI :

1. Analisa Kluster: Suatu Tinjauan
2. Pendekatan Umum
3. Metode Hierarki
4. Metode Partisi: K-Means dan variasinya
5. Density-Based methods: DBSCAN
6. Metode Berbasis Grid
7. Evaluasi Klustering

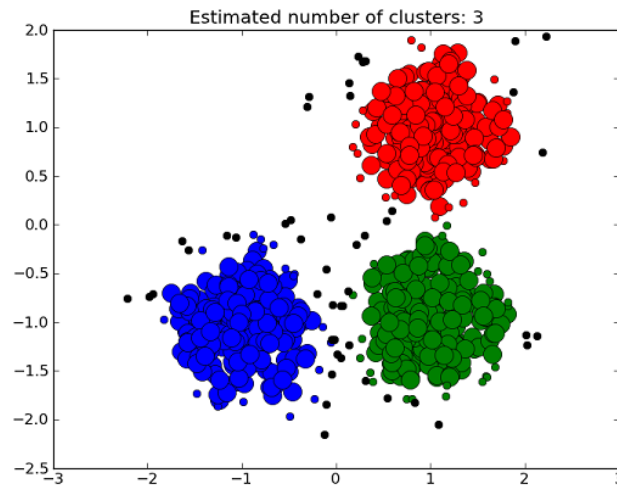
ISI MATERI

Apakah Analisis Klaster?

- Cluster: Kumpulan objek data
 - Serupa (atau terkait) satu sama lain dalam kelompok yang sama
 - Berbeda (atau tidak terkait) dengan objek pada kelompok lain
- Analisis klaster (atau clustering, segmentasi data, ...)
Menemukan kesamaan antara data sesuai karakteristik yang terdapat pada data dan pengelompokan objek data yang serupa ke dalam cluster
- Pembelajaran tanpa pengawasan (unsupervised learning): tidak ada kelas yang telah ditentukan sebelumnya (yaitu, belajar dengan pengamatan vs. belajar dengan contoh: diawasi)
- Gambar di bawah adalah ilustrasi klustering dari dataset orang berdasarkan warna kulit.



- Gambar dibawah adalah contoh clustering dari suatu dataset menjadi 3 kelompok atau kluster.



Aplikasi Umum Klustering?

- Sebagai alat yang berdiri sendiri untuk mendapatkan wawasan tentang distribusi data
- Sebagai langkah preprocessing untuk algoritma lainnya

Aplikasi in riil?

- Biology: taxonomy of living things: kingdom, phylum, class, order, family, genus and species
- Information retrieval: document clustering
- Land use: Identification of areas of similar land use in an earth observation database
- Marketing: Help marketers discover distinct groups in their customer bases, and then use this knowledge to develop targeted marketing programs

- City-planning: Identifying groups of houses according to their house type, value, and geographical location
- Earth-quake studies: Observed earth quake epicenters should be clustered along continent faults
- Climate: understanding earth climate, find patterns of atmospheric and ocean
- Economic Science: market research

Metode Klastering?

- K-Means → lihat di slide ppt mengenai prosedur atau algoritma K-Means.
- Fuzzy Clustering
- Gaussian Mixture
- Self-Organizing Map SOM

SIMPULAN

1. Kelompok cluster menganalisis objek berdasarkan kemiripannya dan memiliki aplikasi yang luas
2. Ukur kemiripan bisa dihitung untuk berbagai jenis data
3. Algoritma Clustering dapat dikategorikan ke dalam metode partisi, metode hirarkis, metode berbasis kerapatan, metode berbasis grid, dan metode berbasis model.
4. Algoritma K-means dan K-medoids adalah algoritma pengelompokan berbasis partisi yang populer
5. Birch dan Chameleon adalah algoritma clustering hirarkis yang menarik, dan ada juga algoritma clustering hirarkis probabilistik
6. DBSCAN, OPTICS, dan DENCLU adalah algoritma berbasis kepadatan yang menarik
7. STING dan CLIQUE adalah metode berbasis grid, dimana CLIQUE juga merupakan algoritma clustering subruang
8. Kualitas hasil pengelompokan dapat dievaluasi dengan berbagai cara

DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, Y. (2006). "Data Mining: Concepts and Technique". Edisi 3. Morgan Kaufman. San Francisco
2. Tan, P.N., Steinbach, M., & Kumar, V. (2006). "Introduction to Data Mining". Addison-Wesley. Michigan
3. Witten, I. H., & Frank, E. (2005). "Data Mining : Practical Machine Learning Tools and Techniques". Second edition. Morgan Kaufmann. San Francisco