

PENGELOMPOKKAN DATA WAJAH MENGGUNAKAN METODE AGGLOMERATIVE CLUSTERING DENGAN ANALISIS KOMPONEN UTAMA

Altien J. Rindengan¹⁾ dan Deiby Tineke Salaki¹⁾

¹⁾Program Studi Matematika FMIPA Universitas Sam Ratulangi Manado 95115

ABSTRAK

Pada penelitian ini dilakukan analisis pengelompokkan data wajah dengan analisis komponen utama untuk mengambil beberapa akar ciri yang cukup mewakili data tersebut dan pengelompokkannya menggunakan metode *agglomerative clustering*. Dengan menggunakan program Matlab, data wajah yang terdiri dari 6 orang dengan 10 *image* dapat dikelompokkan sesuai data aslinya. Pengelompokkannya cukup menggunakan 3 akar ciri pada selang 68 %.

Kata kunci: *agglomerative clustering*, analisis komponen utama, data wajah

FACE DATA CLUSTERING USING AGGLOMERATIVE CLUSTERING METHODS WITH PRINCIPAL COMPONENT ANALYSIS

ABSTRACT

In this research, face data is grouped using principal component analysis by getting some of its eigenvalues which are representative enough to describe the data and then by using agglomerative clustering the data is clustered. By running the Matlab program, face data which is consist of 6 people with 10 images can be clustered to fit the original data. The clustering is enough using 3 eigenvalues with 68 % of interval.

Keywords: agglomerative clustering, principal component analysis, face data

PENDAHULUAN

Data berupa gambar, saat ini semakin banyak ditemui dalam sistem basis data. Data ini dapat dianalisis secara statistika dengan terlebih dahulu mengonversinya ke data yang berbasis angka. Data berbasis angka ini juga memperkecil ruang penyimpanan dalam komputer.

Pada penelitian ini, data gambar yang berupa wajah dikonversi ke data berbasis angka dalam bentuk matriks dengan menggunakan bantuan program Matlab. Data wajah ini selanjutnya dianalisis secara statistik. Dengan memasukkan beberapa contoh gambar wajah dari sejumlah orang yang diambil sampelnya, gambar ini dikelompokkan sehingga setiap orang mempunyai matriks data wajah yang spesifik.

TINJAUAN PUSTAKA

Analisis Peubah Ganda

Analisis Peubah Ganda (APG), dapat dikatakan mempelajari sistem-sistem dari peubah-peubah acak yang berkorelasi atau contoh acak dari system tersebut. Analisis Eksplorasi Data Peubah Ganda berkaitan dengan upaya mengungkap keterangan dari sekumpulan dari sekumpulan data peubah ganda dan keterkaitan antar peubah (Siswadi & Suharjo, 1997).

Awalnya penggunaan APG masih berkembang di kalangan tertentu. Namun perkembangan yang sangat pesat di bidang komputer, mendukung pesatnya perkembangan APG di kalangan luas. Analisis ini diharapkan dapat memberi keterangan dengan memanipulasi data, meringkasnya, memeragakannya sehingga lebih mudah memahaminya dan mengenal adanya hubungan atau pola tidak acak dalam data serta kemungkinan penyimpangannya. (Siswadi & Suharjo, 1997).

Penggunaan analisis peubah ganda dalam mereduksi dimensi data, meliputi Analisis Komponen Utama, Biplot, Faktor, Gerombol, Penskalaan Dimensi Ganda, Korelasi Kanonik, Diskriminan dan Analisis Korespondensi.

Analisis Komponen Utama

Analisis Komponen Utama (AKU), merupakan analisis tertua dalam APG yang diperkenalkan oleh Karl Pearson tahun 1901, yang biasanya digunakan untuk: (1) identifikasi peubah baru yang mendasari data peubah ganda, (2) mereduksi jumlah himpunan peubah yang banyak dan saling berkorelasi menjadi peubah-peubah baru yang tidak berkorelasi dengan mempertahankan sebanyak mungkin keragaman data tersebut, dan (3) menghilangkan peubah-peubah asal yang tidak memberi informasi yang penting (Johnson & Wichern, 2002).

Peubah baru yang terbentuk adalah peubah yang (1) merupakan kombinasi linear dari peubah asal, (2) jumlah kuadrat koefisien dalam kombinasi linear tersebut bernilai satu, (3) tidak saling berkorelasi dan (4) ragamnya terurut dari yang terbesar ke yang terkecil. Peubah-peubah baru ini memanfaatkan informasi dari peubah-peubah asal dan nilai yang nantinya diperoleh dari masing-masing objek merupakan ordinat objek-objek tersebut dalam peubah yang baru yang merupakan suatu sumbu-sumbu koordinat.

Sebelum mencari KU dilakukan penguraian pada peubah-peubah asal dengan penguraian nilai singular (PNS), yaitu $nX_p = nU_r L_r A_p'$; $p(X) = r$. Matriks L adalah matriks diagonal yang unsur-unsur diagonalnya merupakan akar kuadrat dari akar ciri $X'X$ dan A merupakan matriks yang kolom-kolomnya adalah vektor ciri dari $X'X$ yang berpadanan dengan akar ciri λ dimana $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$. Matriks A dan U merupakan matriks dengan kolom ortonormal dimana $U'U = A'A = I$. Sedangkan matriks

$$U = \left[\frac{Xa_1}{\sqrt{\lambda_1}}, \frac{Xa_2}{\sqrt{\lambda_2}}, \dots, \frac{Xa_r}{\sqrt{\lambda_r}} \right] \text{ dimana } a_i$$

merupakan vektor ciri yang berkaitan dengan λ_i .

Perhatikan kombinasi linear berikut:

$$Y_1 = a_1'X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a_2'X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

$$Y_p = a_p'X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

Dan $\text{Var}(Y_i) = a_i' \Sigma a_i, I = 1, 2, \dots, p$

$\text{Cov}(Y_i, Y_k) = a_i' \Sigma a_k, I, k = 1, 2, \dots, p.$

Secara umum, komponen utama ke- I adalah kombinasi linear $a_i'X$ yang memaksimalkan $\text{Var}(a_i'X)$ terhadap kendala $a_i'a_i = 1$ dan $\text{Cov}(a_i'X, a_k'X) = 0$ untuk $k < I$ (Johnson & Wichern, 2002).

Analisis Gerombol (Clustering Analysis)

Analisis gerombol merupakan suatu metode dalam analisis peubah ganda yang bertujuan untuk mengelompokkan n satuan pengamatan ke dalam k gerombol dengan ($k < n$) berdasarkan p peubah, sehingga unit-unit pengamatan dalam satu kelompok mempunyai ciri-ciri yang lebih homogen dibandingkan unit pengamatan dalam kelompok lain (Dillon & Goldstein, 1984).

Metode *agglomerative* melakukan penggerombolan/pengklasteran dari kecil ke besar. Klaster kecil adalah klaster dengan jumlah anggota sedikit dan sebaliknya untuk klaster besar. Metode ini mengawali pengklasteran untuk klaster yang terdiri dari 1 anggota lalu hasilnya disajikan dalam bentuk dendrogram.

Beberapa metode dalam *agglomerative* adalah: metode *Single Linkage* atau pautan tunggal dimana jarak antara dua cluster (kelompok) dapat ditentukan dari dua obyek berpasangan yang memiliki kesamaan atau memiliki jarak terdekat dalam klaster berbeda, metode *Complete Linkage* dimana jarak antar klaster ditentukan dari jarak terjauh antara dua obyek pada klaster yang berbeda. Metode lain adalah *Average Linkage* yang jarak antar klaster ditentukan dari nilai rata-rata jarak seluruh obyek suatu klaster terhadap seluruh obyek pada klaster lainnya.

Jarak antar objek, dinyatakan dalam beberapa jenis ukuran jarak yang dikenal seperti jarak Euclid, Mahalanobis, Manhattan dan Minkowski.

METODOLOGI PENELITIAN

Data

Data yang digunakan dalam penelitian ini adalah sampel data 60 *image/gambar* wajah yang diperoleh dari 6

orang masing-masing 10 *image* yang berukuran 50 x 40 pixel.

Metode

Pada penelitian ini digunakan metode *agglomerative clustering* dengan *single linkage method* dan konsep jarak Euclid untuk mengelompokkan data wajah yang telah direduksi dengan analisis komponen utama. Analisis ini dilakukan dengan bantuan software Matlab.

HASIL DAN PEMBAHASAN

Data gambar wajah memuat gambar enam wajah yang masing-masing diambil 10 gambar dalam *pose* yang sama dengan ukuran 50 x 40 piksel. File gambar berekstensi JPG-JPEG. Data yang awalnya berwarna diubah menjadi *grayscale* dengan Matlab.

Berdasarkan gambar tersebut dikelompokkan dengan komponen utama dengan beberapa selang kepercayaan. Dengan menggunakan selang kepercayaan 95% yang hanya mewakili 3 akar ciri, telah menghasilkan menghasilkan pengelompokkan data wajah sesuai kelompok wajah sebenarnya sebagaimana terlihat pada Gambar 1. Data 1-10 mewakili wajah orang pertama, data 11-20 mewakili wajah orang kedua, data 21-30 mewakili wajah orang ketiga, data 31-40 mewakili wajah orang keempat, data 41-50 mewakili wajah orang kelima dan 51-60 mewakili wajah orang keenam. Pengujian dengan selang

kepercayaan 85 % dan 95 % juga menunjukkan pengelompokkan yang sama dengan 68%.

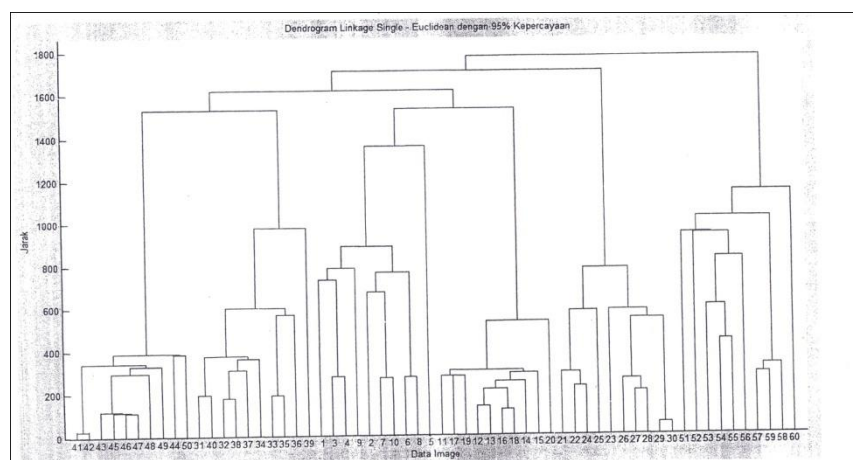
KESIMPULAN

Berdasarkan hasil penelitian ini diperoleh bahwa data wajah dapat dibaca sebagai matriks untuk dikelompokkan dengan terlebih dahulu melakukan analisis komponen utama untuk mengambil beberapa akar ciri yang representatif bagi data asal. Dari data 6 wajah masing-masing dengan 10 *image*, memberikan hasil pengelompokkan yang sesuai dengan data asal.

Penelitian ini dapat dikembangkan untuk data wajah dengan *pose* yang berbeda dan juga dapat digunakan analisis dengan konsep jarak yang berbeda.

DAFTAR PUSTAKA

- Dillon, W.R and M. Goldstein. 1984. *Multivariate Analysis, Method and Application*. John Wiley & Sons. New York.
- Johnson R.A and D.W. Wichern. 2002. *Applied Multivariate Analysis*. 5th Ed. Prentice Hall Inc. New Jersey.
- Siswadi dan Suharjo. 1997. Analisis Eksplorasi Data Peubah Ganda. Jurusan Matematika FMIPA IPB. Bogor.



Gambar 1. Dendrogram Linkage Single-Euclidean 95% Kepercayaan