

ODOR RECOGNITION DENGAN MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS DAN NEAREST NEIGHBOUR CLASSIFIER

Anies Hannawati, Thiang, Yudi Prasetyo

Fakultas Teknologi Industri, Jurusan Teknik Elektro, Universitas Kristen Petra
Email : anies@petra.ac.id; thiang@petra.ac.id

Abstrak

Makalah ini menjelaskan tentang penerapan metode *principal component analysis* pada sistem pengenalan jenis zat berdasarkan aroma atau bau yang dikeluarkan oleh zat tersebut. Dalam penelitian ini terdapat 15 jenis zat dengan masing-masing zat terdiri atas 20 konsentrasi.

Sistem yang telah didisain menerapkan metode *principal component analysis* (PCA) untuk mereduksi jumlah data yang berdimensi besar. Hasil reduksi ini akan menjadi input dalam sistem pengenalan. Metode sistem pengenalan yang diterapkan adalah metode *nearest neighbour*.

Pengujian sistem telah dilakukan dengan melihat tingkat reduksi data yang berhasil dilakukan dengan menghasilkan tingkat pengenalan yang cukup signifikan. Dari pengujian didapat bahwa dengan menggunakan PCA, data dapat direduksi sebesar 93,75% dengan tingkat keberhasilan untuk mengenali zat 99,56%.

Kata kunci : *Principal Component Analysis, Odor recognition, Nearest Neighbour.*

Abstract

This paper describes about implementation of principal component analysis method on odor recognition system. There are 15 essences used in this research and each essence has 20 concentrations.

Principal Component Analysis (PCA) method was used in this system to reduce the data's dimension. The result of PCA would be input of recognition system. The recognition system used in this research is nearest neighbour method.

Experiments were done for testing the designed system. Although the data were reduced until 93,75%, the system could recognize the essences well and achieve a recognition rate of 99.56%.

Keywords : *Principal Component Analysis, Odor Recognition, Nearest Neighbour.*

Pendahuluan

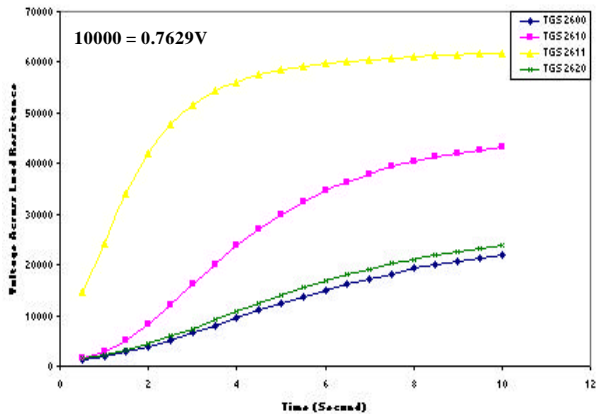
Dalam dunia teknologi, sensor yang dapat mendeteksi aroma atau bau tertentu termasuk teknologi yang masih baru. Dengan ditemukannya sensor pendeteksi bau atau sensor gas tentunya akan sangat membantu dalam dunia industri. Sebagai contoh, dengan sensor gas, kita dapat mengetahui bagaimana kadar polusi udara di suatu daerah, menentukan kualitas makanan, mendeteksi kebocoran gas dan lain-lain. Aplikasi dari teknologi ini biasanya menggunakan lebih dari sebuah sensor gas yang mempunyai karakteristik yang berbeda-beda. Data dari sensor-sensor gas tersebut akan dikombinasikan untuk diaplikasikan.

Makalah ini memaparkan hasil penelitian tentang aplikasi dari sensor gas untuk pengenalan jenis zat berdasarkan aroma atau bau yang dikeluarkan oleh zat tersebut. Dalam penelitian ini digunakan data 15 jenis zat dengan variasi 20 jenis konsentrasi untuk setiap zat. 15 jenis zat tersebut adalah acetone, air, amonia, champor, coffee espresso, coffee ground, coffee mocca, ethanol, eucalyptus, red oil, tea cinnamon, tea peppermint, vegemite, wood dan yellow oil.

Data-data setiap zat tersebut diambil dari empat jenis sensor yaitu TGS2600, TGS2610, TGS2611 dan TGS2620. Tiap sensor mempunyai sensitivitas berbeda terhadap jenis zat yang berbeda. Data empat sensor tersebut akan membentuk suatu kombinasi yang dapat digunakan untuk membedakan suatu jenis zat dengan yang lain. Data yang didapat dari satu sensor berjumlah 20 data. Jadi untuk data satu zat dengan 20 konsentrasi terdapat sebanyak

Catatan: Diskusi untuk makalah ini diterima sebelum tanggal 1 November 2003. Diskusi yang layak muat akan diterbitkan pada Jurnal Teknik Elektro volume 4, nomor 1, Maret 2004.

1600 data. Gambar 1 menunjukkan contoh grafik respon empat sensor TGS untuk jenis zat ethanol.



Gambar 1. Grafik Respon Sensor TGS untuk Ethanol

Karena jumlah data yang besar, maka dalam penelitian ini digunakan metode principal component analysis untuk mereduksi data tersebut sehingga dapat mempercepat proses pengenalan. Metode pengenalan yang digunakan adalah metode nearest neighbour.

Selanjutnya makalah ini akan diorganisasi sebagai berikut: bagian kedua akan membahas tentang metode principal component analysis, bagian ketiga akan menjelaskan metode pengenalan yang digunakan yaitu nearest neighbour, bagian keempat akan memaparkan hasil pengujian sistem dan makalah ini akan ditutup dengan kesimpulan yang didapatkan dari penelitian ini.

Principal Component Analysis

Dalam sistem yang didisain, metode Principal Component Analysis (PCA) diterapkan untuk mereduksi jumlah data yang besar. Prinsip dasar dari algoritma PCA adalah mengurangi dimensi suatu set data namun tetap mempertahankan sebanyak mungkin informasi dalam set data tersebut. Secara matematis PCA mentransformasikan sejumlah variabel yang berkorelasi ke dalam bentuk yang bebas tidak berkorelasi. Principal Component satu dengan yang lain tidak saling berkorelasi dan diurutkan sedemikian rupa sehingga Principal Component yang pertama memuat paling banyak variasi dari data set. Sedangkan Principal Component yang kedua memuat variasi yang tidak dimiliki oleh Principal Component pertama.

Bila satu set data disajikan dalam matrik X, maka algoritma PCA meliputi tahap-tahap berikut:

1. Mencari komponen rata-rata dari matrik X yang dirumuskan dengan persamaan berikut:

$$\bar{x}_l = \frac{1}{n} \sum_{k=1}^n x_{l,k} \quad (1)$$

2. Mencari matrik covariance dengan menggunakan persamaan berikut:

$$C = X * X^T \quad (2)$$

3. Mencari eigenvalue dan eigenvector dari matrik covariance dengan menggunakan persamaan berikut:

$$CQ = IQ \quad (3)$$

dimana **I** adalah eigenvalue dan **Q** adalah eigenvector.

4. Mencari Feature Principal Component dengan mentransformasi data set ke dalam ruang eigen dengan menggunakan persamaan 4.

$$f = \sum_{i=1}^m Q \times (I_i - \bar{X}_i) \quad (4)$$

i menyatakan data ke-*i*, *m* adalah jumlah data dan **Q** menyatakan matriks eigenvector.)

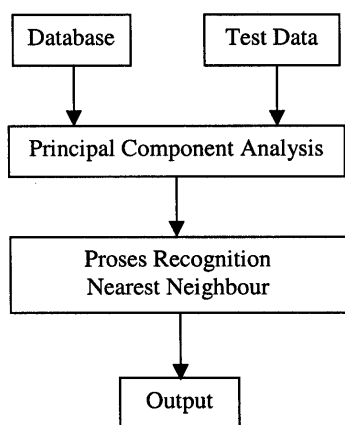
Proses Recognition

Dalam sistem yang telah didisain, metode recognition yang digunakan adalah metode Nearest Neighbour. Metode ini adalah metode yang paling sederhana tetapi mempunyai akurasi yang cukup tinggi. Prinsip dari metode ini adalah mencari selisih jarak yang paling minimum antara data yang akan dikenali dengan database. Jarak tersebut dapat diperoleh dengan menggunakan persamaan berikut:

$$D = \sqrt{\sum (I_x - I_d)^2} \quad (5)$$

Sebuah data yang tidak diketahui akan dikenali sebagai database yang punya jarak terdekat dengan data tersebut.

Secara umum blok diagram dari sistem recognition yang telah didisain dapat dilihat pada gambar 2.



Gambar 2. Blok Diagram Sistem Odor Recognition

Baik database ataupun test data akan diproses dengan menggunakan algoritma PCA untuk mereduksi jumlah data. Jumlah data hasil reduksi bergantung pada berapa principal component yang akan digunakan. Kemudian proses recognition test data dilakukan dengan menggunakan metode nearest neighbour.

Pengujian Sistem

Sistem odor recognition yang telah didisain menggunakan data 15 zat dengan variasi 20 konsentrasi untuk masing-masing zat. Semua data tersebut tersedia dalam bank data. Nama-nama zat yang digunakan seperti yang telah disebutkan pada bagian pendahuluan dari makalah ini.

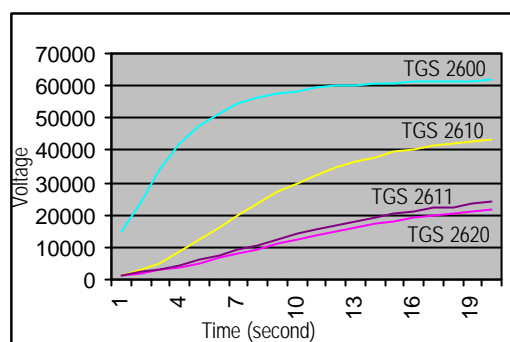
Pengujian sistem odor recognition yang telah didisain meliputi beberapa pengujian yaitu:

1. Pengujian feature atau principal component.
2. Pengujian jumlah feature atau principal component.
3. Pengujian tingkat keberhasilan recognition.

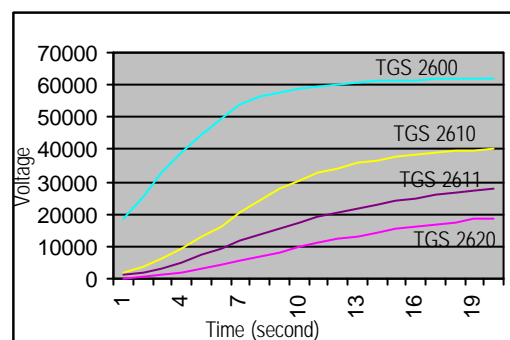
1. Pengujian Feature atau Principal Component

Pengujian ini bertujuan untuk memeriksa *feature* atau *principal component* yang mana yang membawa informasi data asli paling banyak. Pengujian ini dilakukan dengan merekonstruksikan kembali data yang telah diproses dengan PCA ke bentuk semula. Dalam pengujian ini digunakan sejumlah *Principal Component* dengan urutan mulai bagian pertama dan urutan mulai bagian terakhir.

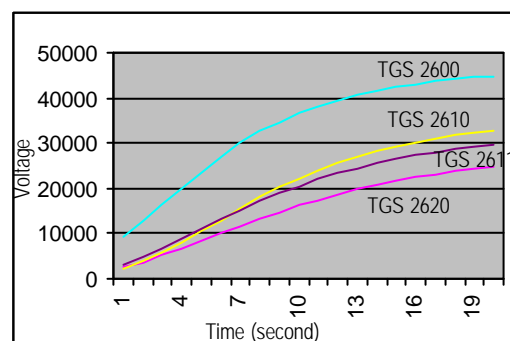
Gambar 3 menunjukkan hasil pengujian yang telah dilakukan dengan menggunakan data Acetone. Gambar 3a menunjukkan grafik data asli dari Acetone, gambar 3b menunjukkan grafik data acetone dari hasil rekonstruksi dengan mengambil empat principal component yang pertama dan gambar 3c menunjukkan grafik data Acetone dari hasil rekonstruksi dengan mengambil empat principal component yang terakhir.



(a)



(b)



(c)

Gambar 3. (a) Grafik Data Asli Acetone, (b) Hasil Rekonstruksi dari 4 Principal Component Pertama, (c) Hasil Rekonstruksi dari 4 Principal Component Terakhir

Bila gambar 3b dan 3c dibandingkan dengan gambar 3a maka terlihat bahwa gambar 3b lebih mendekati gambar 3a dibandingkan dengan gambar 3c. Dengan demikian dapat dikatakan

bahwa *Principal Component* pertama membawa paling banyak informasi dari data asli dan semakin jauh dari *Principal Component* pertama informasi yang dibawa semakin berkurang.

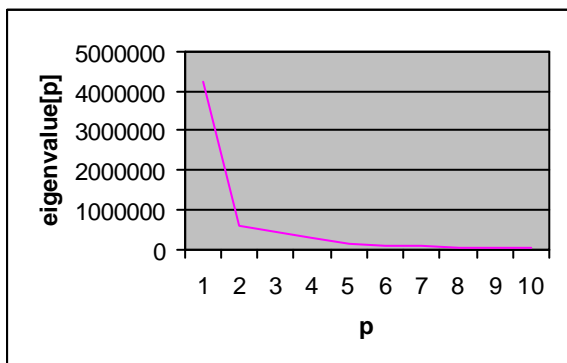
2. Pengujian Jumlah Feature atau Principal Component

Pengujian ini dilakukan untuk menentukan berapa jumlah feature atau principal component yang dianggap cukup untuk mewakili data asli. Penentuan jumlah principal component dilakukan dengan menerapkan aturan yaitu aturan berdasarkan prosentase variasi dan scree plot.

Hasil pengujian yang telah dilakukan dapat dilihat pada tabel 1 dan gambar 4. Tabel 1 menunjukkan eigenvalue ke-1 sampai ke-10 dari data pengujian dan gambar 4 adalah scree plot dari hasil pengujian.

Tabel 1. Eigenvalue Data Pengujian

λ_1	4224746,7247
λ_2	579565,7460
λ_3	449085,2009
λ_4	319681,3700
λ_5	150390,4025
λ_6	116539,4342
λ_7	94153,8264
λ_8	65487,6914
λ_9	29414,4124
λ_{10}	28263,6060



Gambar 4. Scree Plot Hasil Pengujian

Dari hasil pengujian didapatkan bahwa jumlah total dari 80 *eigenvalue* adalah 6.274.737.676. Sehingga untuk mencapai prosentase variasi sebesar 90% maka cukup diambil batas eigenvalue sampai urutan ke-5. Jika dihitung prosentase variasi dengan batas sampai eigenvalue ke-5 maka akan didapat bahwa

prosentase variasinya adalah 91.21%. Hal ini juga terlihat jelas pada scree plot dari tabel 1 yang ditunjukkan oleh gambar 4. Dari gambar terlihat bahwa titik curam terjadi antara titik ke-1 sampai titik ke-5 dan perubahan grafik menjadi landai terjadi secara mencolok mulai titik ke-5. Hal ini menunjukkan bahwa jumlah *Principal Component* yang mencukupi adalah 5.

3. Pengujian Tingkat Keberhasilan Recognition

Tujuan dari pengujian ini adalah untuk mencari tingkat keberhasilan sistem odor recognition yang telah didisain. Ada 15 jenis zat yang digunakan dengan 20 konsentrasi. Tabel 2 menunjukkan hasil pengujian sistem dengan variasi jumlah principal component yang digunakan sedangkan tabel 3 menunjukkan hasil pengujian secara keseluruhan.

Dari tabel 2 terlihat bahwa penggunaan 5 *Principal Component* yang pertama sudah mencukupi walaupun hasil yang terbaik didapatkan pada sistem yang menggunakan 50 *Principal Component* yang pertama. Penggunaan di atas 50 *Principal Component* tidak memberikan perbaikan yang berarti.

Tabel 2. Hasil Pengujian Sistem Odor Recognition dengan Variasi Jumlah Principal Component yang Digunakan

Jumlah <i>Principal Component</i> yang digunakan	Tingkat Keberhasilan (%)
1	23,59
2	71,15
3	94,85
4	97,99
5	99,56
6	99,56
10	99,61
20	99,61
30	99,67
50	99,73
60	99,70
70	99,64
80	99,67

Dengan reduksi data dari 80 menjadi 5, dimana prosentase reduksi data mencapai 93.75% sistem masih dapat melakukan pengenalan zat dengan baik. Secara keseluruhan, sistem dapat mencapai tingkat pengenalan sebesar 99,56% seperti yang ditunjukkan pada tabel 3.

Tabel 3. Hasil Pengujian Sistem Odor Recognition Secara Keseluruhan

Nama zat	Tingkat Keberhasilan (%)
Acetone	100
Air	100
Amonia	100
Champor	100
Coffee espresso	100
Coffee ground	98.99
Coffee mocca	100
Ethanol	100
Eucalyptus	100
Red oil	96.24
Tea cinnamon	100
Tea peppermint	98.16
Vegemite	100
Wood	100
Yellow oil	100
Total Rata-rata	99.56

Kesimpulan

Dari hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa sistem yang telah didisain dapat mengenal zat dengan baik. Dengan metode PCA, sistem masih dapat mengenal zat dengan baik walaupun data direduksi sampai 93,75%. Secara keseluruhan sistem dapat mencapai tingkat pengenalan sebesar 99,56%

Daftar Pustaka

- [1]. Hollmen, Jaakko, Principal Component Analysis, (20 Juni 2001): 1 pp. 10 Januari 2003, <http://www.cis.hut.fi/~jhollmen/dippa/node30.html>
- [2]. Jennings, Alan, Matrix Computation for Engineering and Scientists, Belfast: Queens University, 1980.
- [3]. Borsam, Paul and Weenink, David, Principal Component Analysis, (23 Maret 1999): 2 pp. 10 Januari 2003, http://www.fon.hum.uva.nl/praat/manual/Principal_component_analysis.html
- [4]. Hyvarinen, Aapo, Classical Linear Transformations, (Oktober 2002): 1 pp. 10 Januari 2003, <http://www.cis.hut.fi/~aapo/papers/NCS99web/node5.htm>
- [5]. Thermo Electron Business, Principal Component Analysis Method, (Maret 1996): 2 pp. 10 Januari 2003, <http://www.galactic.com/algorithms/pca.htm>