

## Penerapan Multiple Unsupervised Artificial Neural Network pada Pengenalan Suara Vokal

Leo Willyanto Santoso

Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra

Jl. Siwalankerto 121-131 Surabaya 60236

Email: leow@petra.ac.id

### Abstract

*Teknologi Pengenalan suara memungkinkan komputer yang dilengkapi dengan microphone untuk menginterpretasikan suara manusia. Hal ini sangat membantu seseorang yang mempunyai kelemahan fisik yang tidak dapat menggunakan metode pengaksesan lainnya. Berbagai metode telah dikembangkan dalam pengenalan suara, salah satunya adalah metode neural network.*

*Pengenalan suara vokal dari data suara mentah dapat dilakukan dengan menggunakan multiple unsupervised artificial neural network, yang ditraining dengan suara vokal secara individu untuk mengenal inputan berupa suara. Pembelajaran yang kompetitif digunakan untuk memilih yang terbaik yang mengindikasikan jenis dari suara vokal yang telah diucapkan. Dalam penelitian ini, kita akan mengetahui bagaimana menggunakan neural network untuk mengklasifikasikan suara vokal.*

*Dari penelitian ini, diketahui bahwa untuk suara dengan bentuk gelombang yang unik akan relatif mudah dikenali daripada suara yang lainnya. Training neural network dengan berbagai jenis suara akan membuat perangkat lunak ini speaker independent.*

**Kata kunci:** *Pengenalan suara, artificial neural network*

### 1. Pendahuluan

Salah satu penerapan artificial neural network adalah untuk pemetaan data input dengan suatu pola yang kita inginkan, seberapa besar tingkat kekerabatannya. Perbedaan antara proses mapping dan mapping kohonen yang sudah banyak dipakai adalah adanya scale invariant. Jika kita secara sederhana membuat sebuah topologi yang selalu menjaga transformasi dari input yang nantinya akan di-mapping ke output yang sesuai, maka secara simultan

Menjamin bahwa hanya input yang mirip yang akan di-mapping ke setiap output neuron. Jadi neural network akan membuat topologi yang akan menjaga mapping antara input dan output. Apabila kita menggunakan network Kohonen, secara otomatis membuat perkiraan untuk sebuah *Voronoi Tessilation* dari data input ketika mapping fitur scale invariant mengabaikan informasi yang ada dalam magnitudo input dan mengekstrak informasi tentang magnitudo relatif dari komponen vektor input.

Neural network terdiri dari sebuah single layer dari neuron, dimana aktifasinya ditentukan oleh penjumlahan bobot dari setiap input. Dapat kita tuliskan formulanya yaitu:

$$s_i = \sum_{j=1}^N w_{ij} x_j \quad (1)$$

Dimana  $s_i$  adalah aktivasi dari output neuron ke- $i$ ,  $x_j$  adalah komponen ke- $j$  dari vektor input berdimensi  $N$  dan  $w_{ij}$  adalah bobot dari input neuron ke- $i$  untuk komponen dari input vektor ke- $j$ . Firingnya kemudian dikirimkan kembali melewati bobot yang sama, untuk memberikan nilai:

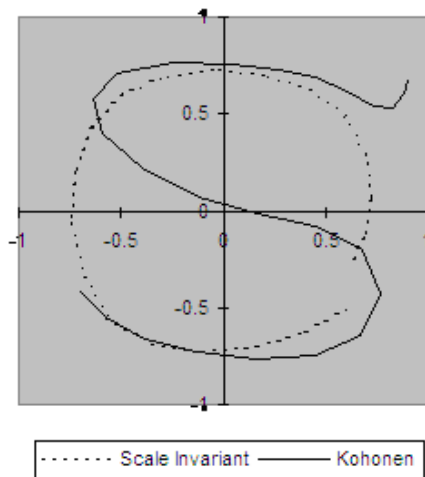
$$x_j(t+1) \leftarrow x_j(t) - w_{pj} \text{ untuk semua } j \quad (2)$$

Dimana  $x(t)$  adalah input vektor pada waktu  $t$ . Neuron dengan aktivasi tertinggi terpilih sebagai neuron terbaik atau pemenang dan outputnya akan di-set dengan nilai aktivasi maksimum (1). Semua output neuron lainnya diberi dengan nilai 0, dan aktifasinya tergantung stimulasi dari neuron pemenang melalui fungsi kekerabatan  $\Lambda(p, j)$ , dimana  $p$  adalah posisi dari neuron pemenang dan  $j$  adalah posisi dari neuron lainnya. Untuk mengubah bobot, kita menggunakan pembelajaran Hebbian sederhana yang ditunjukkan berikut ini.

$$\Delta w_{ij} = \eta_t \Lambda(p, i) \cdot x_j(t+1) \quad (3)$$

Untuk menunjukkan properties dari map fitur scale invariant, kita dapat melakukan training terhadap data artificial dan melakukan observasi pada vektor bobotnya. Hasil training data artificial pada interval

$\{(x,y):-1 \leq x \leq 1, -1 \leq y \leq 1\}$  dan kemudian dilakukan plotting vektor bobot 2D yang hasilnya dapat kita lihat pada Gambar 1.



**Gambar 1.** Matriks bobot dari Scale Invariant dan Map Kohonen.

Perkiraan lingkaran diamati yang cukup berbeda dari pola yang telah dihasilkan oleh Self Organising Map (SOM) Kohonen. Neural network Kohonen menghasilkan matriks bobot yang akan menutup ruang data dalam shape “S”.

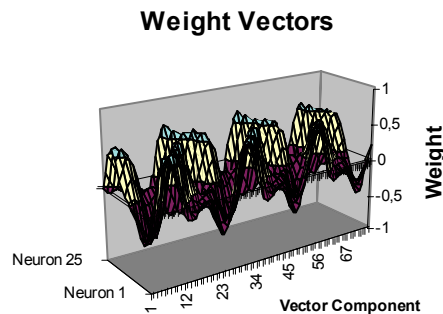
## 2. Training Network

Traing terhadap neural network dilakukan dengan mengucapkan beberapa kata-kata yang mengandung huruf vokal, antara lain (“bat”, “bed”, “beep”, “boot”, “but”) yang diucapkan masing-masing 8 kali pada frekuensi 8 kHz oleh penulis. Suara yang tidak dipakai kemudian dipotong sehingga diperoleh suara vokal dimana nantinya digunakan untuk training data pada neural network. Sebelum dilakukan training, kita tidak melakukan apapun terhadap data yang telah diperoleh.

Fungsi kekerabatan eksponensial  $\frac{x^2}{e^{-r}}$  digunakan, dimana  $x$  adalah jarak antara neuron dan  $r$  adalah radius dari fungsi. Nilai  $r$  atau radius diinisialisasi 30 dan akan dikurangi selama proses training pada nilai 1 untuk setiap seribu putaran training, dan mempunyai nilai minimum 4.

Lima map scale invariant dilakukan training, dimana semua mengandung huruf vokal. Vektor input mempunyai 150 komponen yang merepresentasikan 150 contoh yang berurutan dan mereka semua diletakkan secara random dalam data suara. Dari percobaan ini, kita mengetahui bahwa bobot dapat untuk mengatur vektor untuk setiap neuron output sehingga dapat diperoleh pola seperti gelombang.

Pola yang dihasilkan dari proses training ini dapat dilihat pada Gambar 2.



**Gambar 2.** Pola dari Training untuk bobot.

## 3. Uji Coba

Selama klasifikasi suara vokal, suara dikirimkan terus ke lima neural network, sehingga nantinya akan menghasilkan lima neuron pemenang/terbaik. Neural network yang terbaik mempunyai aktivasi tertinggi yang dipilih untuk semua terbaik. Meskipun langkah ini biasanya membuat klasifikasi yang benar, tetapi tidak selalu menghasilkan output yang diinginkan. Solusi untuk masalah ini adalah dengan mengulang langkah tersebut beberapa kali, dimana setiap proses mengambil input dari tempat yang berbeda dalam data suara. Neural network yang menang, biasanya dapat dipastikan akan menang untuk semuanya.

Pemilihan secara acak vektor input menghasilkan pendeteksian suara vokal yang tidak reliable, dan hasil yang lebih baik diperoleh dengan memilih vektor input secara berurutan dari data suara, sehingga dapat dipastikan terbentuk pola gelombang suara dengan fase yang diinginkan. Sebagai contoh, jika sebuah suara vokal tersusun dari  $n$  contoh, dan kita ingin menggunakan  $x$  komponen dalam input vektor, dan panjang fase dari  $p$ , kemudian kita harus memilih contoh  $s$  secara acak, dimana interval  $0 \rightarrow n-x-p$ . Interval berikut harus dikirimkan kembali.

- 1: interval  $s \rightarrow s + x - 1$
- 2: interval  $s + 1 \rightarrow s + x$
- 3: interval  $s + 2 \rightarrow s + x + 1$
- 4: interval  $s + 3 \rightarrow s + x + 2$

$$p: \text{interval } s + (p-1) \rightarrow s + x + (p-2)$$

Dari sini dapat kita ketahui bahwa beberapa suara vokal mempunyai amplitudo lebih tinggi daripada lainnya. Hal ini dapat menyebabkan neural network yang sama untuk keluar menjadi pemenang beberapa kali. Untuk mengatasi masalah ini, semua vektor bobot dilakukan normalisasi sehingga mempunyai magnitudo Euclidean 1, nantinya semua network mempunyai kesempatan yang sama untuk keluar sebagai pemenang.

Pada percobaan berlangsung, panjang fase yang dipilih adalah 100, sehingga dipastikan bahwa paling sedikit satu gelombang penuh dapat dikirim kembali. Frekuensi pemenang dari setiap suara vokal dicatat, tetapi data ini mempunyai tingkat variasi yang tinggi. Setiap 100 tes yang dilakukan, akan diulang 10 kali, dengan pemilihan posisi start secara random. Dari sini diperoleh, setiap suara vokal direpresentasikan total sebanyak 1000 kali dan hasilnya dapat dilihat pada Tabel 1.

Tabel 1. Hasil Uji Coba

	Suara Vokal Yang dihasilkan				
	bat	bed	beep	boot	but
bat	817				
bed		934			
beep			973		
boot				621	
but					478

Semua hasil yang diperoleh dihasilkan dari proses testing dari neural network pada data yang sama seperti yang dilakukan waktu proses training. Pada tabel 2, dapat dilihat hasil uji coba yang dilakukan dari pembicara yang sama dengan melakukan proses training ulang, tetapi dengan perekaman yang berbeda.

Tabel 2. Hasil Uji Coba 2

	Suara Vokal Yang dihasilkan				
	bat	bed	beep	boot	but
bat	823				
bed		971			
beep			987		
boot				675	
but					528

#### 4. Kesimpulan

Dari percobaan yang dilakukan dapat diambil kesimpulan, yaitu:

1. Multiple unsupervised neural network dapat digunakan untuk mengenali suara vokal.
2. Suara “beep” dapat dikenali dengan nilai yang sangat tinggi, yaitu 973 dan 987, hal ini menandakan bahwa suara ini mempunyai bentuk gelombang yang unik apabila dibandingkan dengan suara lainnya.
3. Program ini masih sangat bergantung dengan speaker, sehingga masih perlu dilakukan training dengan jenis suara yang berbeda-beda, sehingga dapat menjadi pengenalan suara yang tidak bergantung kepada speaker.

#### References

- [1] Demuth, Howard, “*Neural Network Toolbox*”, The MathWorks, Inc., 1998
- [2] Fausett, Laurene, “*Fundamental of Neural Networks*”, Prentice Hall, 1994.
- [3] Fyfe, C., “*A Scale invariant Feature Map*” Network: Computation in Neural Systems 7, pp.269-275, 1996
- [4] Gonzales, Avelino J., dan Douglas D. Daniel, “*The Engineering of Knowledge-based Systems*”, Prentice Hall, 1993
- [5] Kohonen, T., “*Self Organizing Maps*”, Berlin: Springer, 1995
- [6] Russel, Stuart J., dan Peter Norvig, “*Artificial Intelligence*” A Modern Approach, Prentice hall, 1995