

Pengenalan Dokumen Jawa Menggunakan Jaringan Syaraf Tiruan Feedforward Elman Type Algorithm

Hans Christian Indrayana¹, Gregorius Satia Budhi², Liliana³
Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra
Jl. Siwalankerto 121-131 Surabaya
+62 31 8439040, 8394830-31
Hanschristian85@yahoo.com; greg@petra.ac.id; lilian@petra.ac.id

ABSTRAK

Proses pelatihan merupakan bagian yang terpenting untuk dapat mengenali sesuatu, karena tanpa pelatihan, tidak mungkin data – data yang ada dapat mengenali sesuatu hal yang ingin dikenali. Proses pelatihan untuk mengenali huruf Jawa memiliki tingkat kesulitan sendiri untuk mengumpulkan data – data yang dapat digunakan untuk proses pelatihan pengenalan huruf Jawa, serta butuh waktu yang lama untuk melakukan proses pelatihan. Dalam skripsi ini dikembangkan aplikasi yang dapat melakukan proses pelatihan pengenalan huruf Jawa dengan data – data ekstraksi fitur yang ada.

Adapun proses yang dilakukan adalah sebagai berikut: gambar huruf Jawa yang sudah disegmentasi sebelumnya, akan di ekstraksi fiturnya, kemudian data – data hasil ekstraksi fitur dijadikan input untuk melakukan proses pelatihan dalam Elman Type. Hasil proses pelatihan berupa data - data bias dan weight yang nantinya bila sudah dirasa dapat menghasilkan output yang baik, digunakan untuk dapat mengenali huruf Jawa dari data – data ekstraksi fitur huruf Jawa lainnya. *Input* berupa 88 atribut dari data ekstraksi fitur, yaitu berdasar teori *area-based* ekstraksi fitur dari gambar sebuah huruf Jawa. *Output* berupa bias, weight, dan hasil pengenalan huruf Jawa disimpan dalam file.txt. Aplikasi ini dibuat dengan bahasa pemrograman C# dengan Microsoft Visual Studio 2010 sebagai IDE-nya.

Hasil pengujian menunjukkan bahwa akurasi pengenalan huruf Jawa dengan menggunakan metode Elman Type terbaik mencapai 85.16% .

Kata Kunci: Huruf Jawa, *Backpropagation*, Pelatihan, Pengenalan, Elman Type.

ABSTRACT

The training process is the most important thing to recognize something, because without the process training, it's impossible to recognize something with what datas you have. The training process of Javanese letter has its own difficulty level for collecting the extraction feature datas that can be used for training process and it took a long time to do the training process. In this paper developed an application that can make the process training to recognize of Javanese Letter with the extration feature datas.

The process is carried out as follows: image that has been previously segmented, then extract the feature datas and used it as input to begin the training process with Elman Type. The Inputs are 88 attribute from data extraction features, that is the result of extraction features based on area-based theory from a

picture of Javanese letter. The result of the training process are some datas, bias, weight datas, when the data is deemed to able produce good output, then it can be used for recognize the Javanese letter from the other source extraction feature datas . The output is some datas, like bias, weight, and Java letter recognition results stored in file.txt. This application is made to the programming language C # with Microsoft Visual Studio 2010 as the IDE.

The results show that javanese letter recognition accuracy by using Elman reach 85.16 %.

Keywords: *Javanese letter, Backpropagation, Training, Recognize, Elman Type.*

1. LATAR BELAKANG

Pembuatan aplikasi pengenalan huruf Jawa bertujuan untuk melestarikan budaya Jawa yang saat ini mulai terkikis oleh adanya budaya atau trend yang lebih modern. Budaya yang diangkat dalam pembuatan aplikasi ini adalah huruf Jawa, yang saat ini sedang dilakukan suatu penelitian untuk membuat aplikasi pengenalan huruf Jawa.

Aplikasi pengenalan huruf Jawa membutuhkan 3 tahap yaitu segmentasi gambar, ekstraksi fitur, dan pelatihan pengenalan huruf Jawa. Segmentasi gambar adalah proses pengambilan tiap huruf dari dokumen huruf Jawa. Setelah semua huruf Jawa tersegmentasi, proses selanjutnya adalah ekstraksi fitur jumlah garis, jumlah lengkung, jumlah *loop* huruf Jawa yang merupakan proses pengambilan ciri huruf Jawa pada setiap huruf Jawa yang telah disegmentasi, dan tahap terakhir adalah pengenalan huruf Jawa, yaitu proses pelatihan huruf Jawa berdasar hasil ekstraksi fitur huruf Jawa yang didapatkan. Pembuatan aplikasi pengenalan huruf Jawa ini sudah sampai pada tahap ekstraksi fitur huruf Jawa, dan selanjutnya ini adalah proses pengenalan huruf Jawa. Proses pengenalan huruf Jawa ini membutuhkan waktu yang cukup lama untuk melatih komputer dapat mengenali huruf Jawa, dikarenakan butuh banyak pelatihan, percobaan dan pengujian untuk mendapatkan hasil yang terbaik.

Proses pelatihan pengenalan huruf Jawa menggunakan Elman Type dipilih dalam pelatihan huruf Jawa dikarenakan memiliki arsitektur yang hampir sama dengan *Backpropagation*, tetapi ada tambahan *memory* yang membuat pelatihan dalam Elman Type mungkin bisa menjadi lebih baik. Metode inilah yang nantinya akan dicoba untuk pelatihan huruf Jawa.

2. LANDASAN TEORI

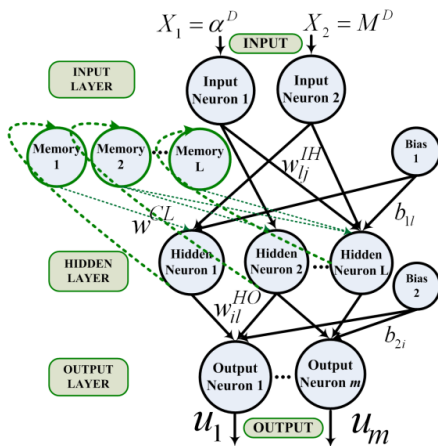
2.1 Review Penelitian Sebelumnya

Pada penelitian sebelumnya telah dilakukan proses segmentasi [4] dan ekstraksi fitur huruf Jawa[3] dari suatu dokumen huruf Jawa. Ekstraksi fitur yang digunakan adalah dengan memberikan label kombinasi matrix 3x3 pixel pada suatu huruf Jawa yang telah dinormalisasi ukurannya menjadi 33x24 pixel, sehingga akan didapat 88 atribut untuk setiap huruf Jawa, dengan teori ekstraksi fitur area based[1].

2.2 Elman Type

Jaringan syaraf tiruan merupakan suatu sistem pemrosesan informasi yang mempunyai karakteristik menyerupai jaringan syaraf biologi dan tercipta sebagai suatu generalisasi model matematis dari pemahaman manusia. Jaringan syaraf tiruan memiliki arsitektur layar tunggal (*single layer*) dan banyak lapisan, layar tunggal terdiri dari 1 input dan 1 output, sedangkan banyak lapisan (*multi layer*) memiliki *input layer*, *hidden layer* dan *output layer*.

Elman Type Neural Network [5] merupakan salah satu metode jaringan syaraf tiruan yang komposisi adalah multilayer, yang terdiri dari *input layer*, *hidden layer* dan *output layer*, namun bedanya terdapat suatu fungsi *Feedback* yang berasal dari *output hidden layer* menuju ke input *hidden layer* (*Memory*). Contoh arsitektur *Elman Type* dapat dilihat pada Gambar 1



Gambar 1. Contoh arsitektur Elman Type

2.3 Chi²

Algoritma Chi² merupakan algoritma yang menggunakan χ^2 statistik untuk mendiskritkan atribut yang memiliki nilai numerik[2]. Algoritma ini dinilai efektif untuk penyeleksian fitur-fitur penting dari sekelompok atribut numerik. Bila fitur-fitur tersebut relevan, maka algoritma ini dapat mempercepat proses pelatihan dan meningkatkan akurasi prediksi dari algoritma pelatihan yang digunakan.

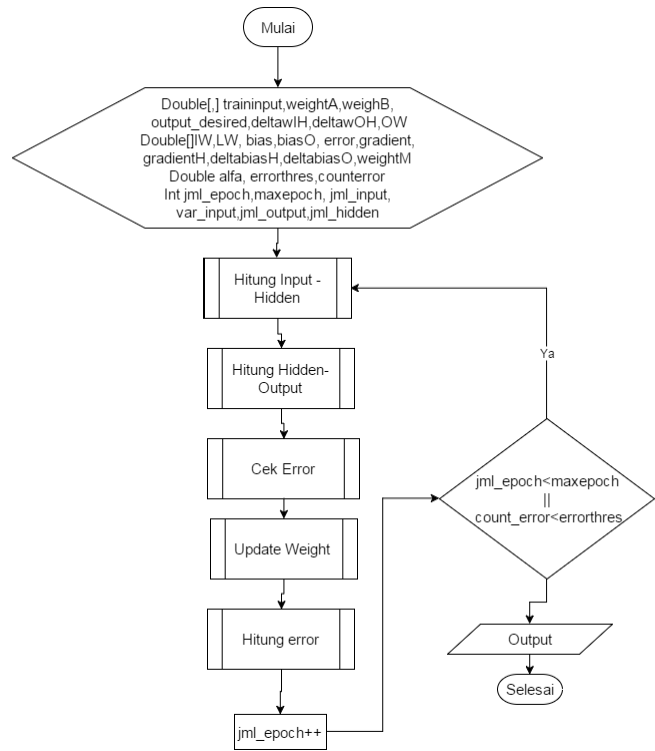
3. DESAIN SISTEM

3.1 Garis Besar Sistem Kerja Perangkat Lunak

Dalam Aplikasi pengenalan huruf Jawa ini terdapat fitur utama, yaitu pelatihan menggunakan Elman Type.

3.2 Garis Besar Elman Type

Proses pelatihan Elman Type memerlukan beberapa tahapan. Adapun rancangan sistem kerja perangkat lunak untuk Elman Type secara garis besar ditunjukkan oleh Gambar 2. Untuk Elman Type terdapat proses Memory yang berada pada *input* ke *hidden layer*, sebelum ke *output layer*, hasil perhitungan sebelumnya akan dijadikan memory.



Gambar 2. Diagram alur proses Elman Type

4. IMPLEMENTASI SISTEM

Pemrograman aplikasi ini menggunakan prinsip prosedural. Terdapat tiga lingkup fungsi yang digunakan dalam aplikasi ini, yaitu *File*, *Training*, dan *Testing*.

Pada lingkup File berisi fungsi – fungsi yang digunakan seputar file, seperti save data dalam bentuk file text, load data dari file text. Fungsi yang penting dalam segmen program file ini adalah fungsi load fitur yang digunakan untuk mengambil data training, menghitung jumlah data, serta memberikan tambahan parameter suatu nilai untuk memudahkan pelatihan dengan memberikan tanda berupa suatu nilai pada huruf Jawa yang dimaksud. Contohnya untuk huruf Ha diberikan nilai 0, kemudian Na diberi nilai 1, dan seterusnya.

Pada lingkup Training berisi fungsi – fungsi yang digunakan untuk pelatihan dalam metode Elman Type. Fungsi *Start*

digunakan untuk memberi inisialisasi nilai - nilai awal untuk pelatihan Elman, meliputi nilai random weight yang dibutuhkan, jumlah hidden, jumlah input, banyaknya *data training*, jumlah output, nilai *output desired* yang diinginkan, nilai *bias*, *errorthreshold* dan lain - lain.

Fungsi trainelman digunakan untuk proses pelatihan dengan menggunakan arsitektur Elman. Terdapat 5 tahap yaitu perhitungan input ke hidden dan memory, perhitungan hidden ke output, cek error, update weight, dan hitung error. Tahap pertama adalah proses perkalian data training input dengan weight yang ada pada hidden, dikurangi dengan bias, kemudian diaktifasi dengan fungsi sigmoid. Pada Tahap ini terdapat memory, yaitu difungsikan untuk menampung hasil perhitungan yang sebelumnya dan dikalikan dengan weight yang ada pada memory.

Tahap Kedua adalah perhitungan dari hidden layer ke output. Dengan mengalikan hasil perhitungan input ke hidden pada weight yang ada pada output layer.

Tahap ketiga adalah cek error, yaitu proses menghitung error yang dijadikan nilai perhitungan dalam mengupdate bias dan juga weight.

Tahap keempat adalah update weight, yaitu menambahkan weight yang lama dengan weight yang baru.

Tahap terakhir adalah menghitung error dengan memproses Feedforward(perhitungan input ke hidden, hidden ke hidden dan hidden ke output) kemudian *output desired* dikurangi dengan hasil perhitungan feedforward untuk dapat nilai errornya.

5. PENGUJIAN SISTEM

Dalam Pengujian sistem, data pengujian diambil dari 10 dokumen beraksara Jawa yang berasal dari beberapa buku / dokumen. Input yang digunakan adalah hasil ekstraksi fitur berdasar teori area-based dengan 88 atribut. Pembagian data training dan data testing dilakukan dengan 2 macam cara untuk pengujiannya. Cara pertama, dari 10 dokumen yang ada, digunakan 7 dokumen yang tiap hurufnya diambil 5 macam variasi data dengan segmentasi terbaik untuk tiap hurufnya, kemudian sisanya digunakan untuk data testing. Cara ini untuk pengujian χ^2 dan pengujian *learning rate*, kemudian cara kedua adalah dengan menggunakan 10 dokumen yang ada kemudian diambil 1/3 bagian sebagai data training dengan 10 variasi macam data untuk tiap huruf Jawa, dan sisanya dijadikan data testing. Cara kedua ini untuk pengujian jumlah *hidden neuron* dan jumlah *hidden layer*.

5.1 Pengujian Pengenalan Huruf Jawa Dengan dan Tanpa χ^2

Pengujian ini akan dimulai dengan menggunakan data training dan data testing dengan dan tanpa menggunakan χ^2 . Hasil yang didapat hasil dengan χ^2 menghasilkan hasil pengenalan data testing yang lebih baik daripada dengan tanpa χ^2 untuk data testing. Hasil Pengujian dapat dilihat pada Tabel 1 dan Tabel 2

Tabel 1 Hasil uji tanpa χ^2

Jenis Data	Jumlah Hidden Layer	Hidden Neuron	Learning Rate	Akurasi (dalam%)
Data Training	2	50	0.05	98.63
Data Testing				44.30

Tabel 2 Hasil uji dengan χ^2

Jenis Data	Jumlah Hidden Layer	Hidden Neuron	Learning Rate	Akurasi (dalam%)
Data Training	2	50	0.05	99.54
Data Testing				46.91

5.2 Pengujian Pengenalan Huruf Jawa Elman dengan Variasi *Learning rate*

Pengujian ini dilakukan dengan menggunakan variasi *learning rate*, sebesar 0.1, 0.01, 0.05, 0.005. Hasil pengujian yang telah dilakukan, didapati learning rate dengan 0.05 menghasilkan pengenalan yang terbaik untuk data testing. Hasil pengujian learning rate dapat dilihat pada Tabel 2.

Tabel 2 Hasil uji learning rate

Jenis Data	Jumlah Hidden Layer	Hidden Neuron	Learning Rate	Akurasi (dalam%)
Data Training	2	50	0.05	99.54
Data Testing				46.91

5.3 Pengujian Pengenalan Huruf Jawa Dengan Menggunakan Variasi Jumlah Neuron Pada Tiap Hidden Layer

Pengujian ini akan menggunakan variasi jumlah hidden neuron sebesar 25, 50, 88, dan 100. Dalam pengujian kali ini menggunakan 2 jumlah hidden layer, 1 dan 2, dan didapati dengan hidden neuron 100 dan jumlah hidden layer 1 mendapat hasil pengenalan terbaik untuk data testing. Hasil pengujian jumlah hidden neuron dapat dilihat pada Tabel 3

Tabel 3 Hasil uji jumlah hidden neuron

Jenis Data	Jumlah Hidden Layer	Hidden Neuron	Learning Rate	Akurasi (dalam%)
Data Training	1	88	0.05	95.64
Data Testing				85.16

5.4 Pengujian Pengenalan Huruf Jawa Backpropagation dan Elman Dengan Variasi Jumlah Hidden Layer

Pengujian ini akan menggunakan variasi jumlah hidden layer sebesar 1, 2, 3, dan 4. Hasil pengujian ini didapati dengan hidden layer 1 menghasilkan data pengenalan data testing yang paling baik. Secara keseluruhan hasil yang terbaik dapat dilihat pada Tabel 4.

Tabel 4 Hasil terbaik Elman Type

Jenis Data	Jumlah Hidden Layer	Hidden Neuron	Learning Rate	Akurasi (dalam%)
Data Training	1	88	0.05	95.64
Data Testing				85.16

6. KESIMPULAN

Berdasarkan hasil pengujian, dapat ditarik beberapa kesimpulan sebagai berikut :

- Hasil Pengenalan huruf Jawa dengan menggunakan Chi^2 hasilnya lebih baik daripada yang tidak menggunakan Chi^2
- Hasil pengujian Elman dengan *setting*-an yang terbaik adalah dengan jumlah *hidden neuron* 100, *learning rate*-nya 0.05, serta jumlah *hidden layer* nya adalah 1 dengan keakuratan 85.16% pada *data testing*
- Dalam pengujian Elman Type hasil yang paling baik untuk pengenalan secara keseluruhan adalah sebesar 85.16% pada *data testing*.

7. REFERENSI

- [1] Adipranata, R., Satia G., Liliana, Sebastian B. 2015. Comparison Between Shape-Based And Area-Based Features Extraction For Java Character Recognition, Materi dipresentasikan dalam *Proceeding of the 2nd Management and Innovation Technology International Conference*, 16-18 November 2015, Miticon, Bangkok.p.2-3
- [2] Hartanto, J. 2013. *Peningkatan Kemampuan Pengenalan Pola Dari Jaringan Syaraf Tiruan Dengan Menggunakan Diskritisasi Chi^2* . Skripsi. 01021226/INF/2013.Universitas Krsiten Petra:Surabaya
- [3] Indrawijaya, M. 2015. *Aplikasi Ekstraksi Fitur Citra Huruf Jawa Berdasarkan Morfologinya*. Skripsi. 01021378/INF/2015, Universitas Kristen Petra
- [4] Mardianto, S. 2015. *Aplikasi Segmentasi Huruf Jawa*. Skripsi. 01021385/INF/2014/ Universitas Kristen Petra.
- [5] Pasila, F.,Lesmana T., Ferdinando H. 2009. Elman Neural Network with accelerated LMA Training for East Java-Bali Electrical Load Time Series Data Forecasting. *Proceeding of the 5th International Conferences of Information, Communication, Technology, and Systems (ICTS)*, 4 Agustus 2009. Surabaya, Indonesia, p.141-148.