

RECURRENT NEURAL NETWORK UNTUK PENGENALAN UCAPAN PADABAHASA SUNDA DIALEK UTARA

RECURRENT NEURAL NETWORK FOR SPEECH RECOGNITION ON SUNDANESE LANGUAGE OF THE NORTHERN DIALECT

¹Arianto Anggoro, ²Andrew Brian Osmond, ³Randy Erfa Saputra

¹²³Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom
¹ariantoanggoro@rocketmail.com, ²purbadaru@telkomuniversity.ac.id,
³casisetianingsih@telkomuniversity.ac.id

Abstrak

Bahasa merupakan sebuah keahlian yang dipakai oleh kita manusia untuk berkomunikasi antar satu dengan lainnya, bahasa juga mempunyai sistemnya sendiri yang ada agar proses komunikasi antar satu individu dengan individu lainnya berjalan dengan lancar. Setiap bahasa tentunya memiliki logat ataupun dialektanya masing-masing mungkin dari cara berbicaranya maupun sistem penulisannya. Untuk mengenali bahasa yang digunakan pada setiap daerah dibutuhkan sistem yang bisa mengenali dialek-dialek yang ada pada daerah tersebut.

Pada penelitian ini menggunakan metode Deep Learning Recurrent Neural Network. Dari hasil penelitian yang dilakukan, dengan nilai parameter tertentu didapatkan akurasi sebesar 72.9%. Setelah didapatkan akurasi tersebut, dilakukan pengujian dengan mengganti salah satu dari nilai parameter yang terdapat pada sistem, sehingga mendapatkan hasil akurasi sebesar 74%. Dari pengujian tersebut dapat disimpulkan bahwa semakin besar parameter epoch yang dimasukan dan data latih yang banyak akan menghasilkan akurasi yang lebih baik

Kata kunci : Neural Network, Deep Learning

Abstract

Language is a skill used by us humans to communicate with each other, the language also has its own system that exist for the communication process between one individual with other individuals running smoothly. Each language must have its own dialect or as much as possible from the way it speaks and the writing system. To recognize the language used in each region required a system that can recognize dialects that exist in the area.

In this study case using Deep Learning Recurrent Neural Network method. From the results of research conducted, with the value of certain parameters obtained accuracy of 72.9%. After obtaining the accuracy, tested by replacing one of the parameter values found in the system, so get the accuracy of 74%. From these tests it can be concluded that the larger the epoch and the many training data will produce better accuracy.

Keywords: Artificial Intelligence, Neural Network, Artificial Neural Networks

1. Pendahuluan

1.1 Latar Belakang

Bahasa merupakan salah satu cara kita manusia berkomunikasi antar satu dengan yang lainnya, setiap daerah mempunyai bahasanya masing-masing dan dialektanya masing – masing. Tingkat akurasi merupakan salah satu masalah dalam aplikasi Speech Recognizer salah satunya kata dalam Bahasa Sunda. Didalam Bahasa Sunda terdapat perbedaan dialek dari setiap daerah yang menggunakannya. Para pakar bahasa biasanya membedakan enam dialek yang berbeda [1].

Automatic Speech Recognition merupakan kemampuan mesin untuk mengenal kata yang diucapkan dan mengubahnya menjadi suatu variabel yang bisa dibaca oleh mesin. Recurrent Neural

Network merupakan salah satu cabang dari Artificial Neural Network untuk memproses data sekuensial, cara yang dilakukan RNN adalah dengan menggunakan looping sehingga input yang kita masukan bisa dipanggil kembali dikemudian harinya.

1.2 Tujuan

Adapun tujuan yang akan dicapai pada Tugas Akhir ini :

Tujuan dari penelitian tugas akhir ini yaitu Membuat sebuah sistem pengenalan suara dengan mengimplementasikan metode Recurrent Neural Network pada pengenalan Bahasa Sunda dialek Bogor dengan inputan suara (audio) yang akan diproses sehingga menghasilkan sebuah output berupa klasifikasi dari input tersebut, sehingga input bisa dinyatakan memenuhi bobot atau tidak pada Bahasa Sunda dialek Bogor.

1.3 Identifikasi Masalah

Beberapa identifikasi masalah yang akan dibahas dalam Tugas Akhir ini adalah sebagai berikut :

Bagaimana kemampuan dari Recurrent Neural Network dalam pengenalan ucapan. Kemudian pengujian pengaruh nilai parameter terhadap hasil akurasi.

1. Dasar Teori

1.1 *Speech Recognition*

Speech Recognition atau Automatic Speech Recognition (ASR) merupakan teknologi dimana mesin dapat mengenal apa yang diucapkan oleh kita, mesin akan menerima masukan berupa suara yang diproses menjadi suatu pola yang bisa dibaca oleh suatu mesin [6], teknologi ini memungkinkan untuk mesin memahami dan mengenal apa yang kita ucapkan. Input ucapan yang kita keluarkan akan diolah menjadi sebuah sinyal dengan cara mengubah gelombang suara menjadi sekumpulan angka yang kemudian disesuaikan dengan pola tertentu untuk mengidentifikasi kata-kata tersebut. Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan, dan bisa menjadi sebuah perintah agar perangkat tersebut melakukan apa yang kita inginkan

1.2 Neural Network

Neural Network merupakan representasi otak manusia. Dimana sebuah model yang terdiri dari elemen-elemen pengolahan serta beberapa neuron atau node berfungsi berdasarkan cara kerja otak manusia [9]. Neural Network biasanya dilatih sebagai klasifikasi frame-level pada speech recognition. Neural Network memodelkan beberapa aspek dari otak manusia, dimana proses berfikir didapatkan dari hubungan sinapsis setiap neuron. Struktur ini berlapis dan mampu menggunakan struktur paralel yang tinggi. Neural Network berguna untuk klasifikasi, aproksimasi fungsi dan masalah rumit [2], yang tidak memerlukan solusi yang akurat. Neural Network harus dilatih sebelum digunakan agar tau konsep bagaimana manusia belajar. Biasanya unit tersebut bersifat homogen, tetapi ada juga yang heterogen. Biasanya dilatih sebagai pengklasifikasi tingkat frame dalam pengenalan suara [10].

1.3 Mel-Frequency Cepstral Coefficient (MFCC)

MFCC merupakan salah satu metode ekstraksi ciri untuk sinyal akustik terbaik [13]. Mel Frequency Cepstrum Coefficients (MFCC) merupakan salah satu metode yang banyak dipakai dalam bidang speech recognition. Metode ini digunakan untuk melakukan feature extraction, sebuah proses yang mengkonversikan sinyal suara menjadi beberapa parameter. Filter ini digunakan untuk menangkap karakteristik fonetis penting dari sebuah ucapan. Analisis suara pada mel-frequency didasarkan pada persepsi pendengaran manusia, karena telinga manusia telah diamati dapat berfungsi sebagai filter pada frekuensi tertentu.

1.4 Recurrent Neural Network

Secara prinsip Recurrent Neural Network dapat menggunakan koneksi feedback mereka untuk memunculkan kembali input yang baru saja digunakan untuk mengaktifkan memori [7]. Memprediksi sequential data dianggap oleh banyak orang sebagai masalah utama dalam Machine Learning dan Artificial Intelligence [8]. Recurrent Neural Network (RNN) merupakan salah satu bagian dari keluarga Neural Network untuk memproses data sekuensial. RNN menyimpan informasi dari proses sebelumnya sehingga dapat memprediksi data dari inputan selanjutnya. Cara yang dilakukan RNN agar dapat menyimpan informasi dari output sebelumnya adalah menggunakan looping sehingga output dari percobaan sebelumnya tetap tersimpan

1.5 Long Short-Term Memory

Long Short-Term Memory (LSTM) Merupakan arsitektur Recurrent Neural Network yang mampu mempelajari ketergantungan jangka panjang. LSTM adalah RNN yang mempunyai mekanisme gate spesial yang mengontrol memory sel [5]. LSTM adalah sel memori yang dapat mempertahankan nilainya dari waktu ke waktu dan unit gerbang non linear yang mengatur arus informasi keluar dan masuk pada sel [3]. LSTM mempunyai unit khusus yang disebut blok memori di recurrent hidden layer. Blok memori berisi sel memori dengan self-connections yang menyimpan keadaan sementara dari jaringan di samping unit khusus yang disebut gate untuk mengendalikan arus informasi. Setiap blok memori dalam arsitektur berisi input gate dan output gate. Input gate mengendalikan aliran input activation ke dalam sel memori. Output gate mengendalikan arus keluaran cell activations ke seluruh jaringan. Nantinya, forget gate ditambahkan ke blok memori [4].

2. Perancangan

2.1 Gambaran Sistem Secara Umum

Sistem yang dibuat merupakan sistem yang dapat mengenali dan mengklasifikasikan input yang berupa rekaman suara (audio) dengan Bahasa Sunda dialek Bogor. Gambaran umum sistem yang akan dibuat dapat dilihat pada gambar 2.1



Gambar 2.1 Gambaran Umum Sistem

Metode yang digunakan pada penelitian ini adalah recurrent neural network. Langkah utama yang dilakukan untuk penelitian ini adalah pencarian sampel, sampel berupa rekaman suara dari narasumber yang berbicara bahasa sunda dialek bogor sebagai data latih, data uji menggunakan narasumber yang berbicara Bahasa Sunda dialek bogor, Bahasa Sunda Lain, dan Bukan Orang Sunda berbicara Bahasa Sunda.

Setelah data terkumpul dilakukan preprocessing, dimana data masukan yang berupa sinyal diolah agar dapat menghasilkan bentuk dan nilai yang diharapkan, setelah itu dilakukan proses ekstrasi ciri dengan Mel-Frequency Cepstral Coefficient (MFCC) agar mendapatkan nilai ciri dari masing-masing data masukan.

3. Pengujian

3.1. Skenario Pengujian

Pada pengujian ini dilakukan perubahan pada nilai epoch dan minibatch size untuk mencari nilai akurasi tertinggi, pengujian ini dilakukan sepuluh kali pada nilai epoch, dan lima kali pada nilai minibatchsize.

3.2. Pengujian MiniBatch

Pada tahap ini dilakukan pengujian dengan mengubah nilai epoch yang ada pada sistem. Pengujian Epoch bisa dilihat pada Tabel 3.1 berikut.

Tabel 3.1 Hasil Pengujian Epoch

Pengujian	Epoch	MiniBatchSize	Akurasi
1	10	27	55%
2	20	27	57%
3	30	27	60%
4	40	27	62%
5	50	27	60%
6	100	27	62.2%
7	150	27	64.15%
8	200	27	66.66%
9	250	27	72.9%
10	300	27	83.58%

Pada tabel 3.1 diatas dilakukan 10 pengujian dengan nilai epoch yang berbeda, didapatkan nilai akurasi tertinggi dengan tingkat akurasi sebesar 83.58% pada pengujian ke-10 menggunakan nilai epoch sebesar 300. Nilai tersebut akan diuji kembali dengan menggunakan nilai MiniBatchSize yang berbeda.

3.3. Pengujian MiniBatch

Pada tahap ini dilakukan pengujian dengan mengubah nilai MiniBatchSize menggunakan nilai epoch yang mendapatkan hasil nilai akurasi tertinggi untuk melihat perbedaan nilai akurasi. Pengujian MiniBatchSize bisa dilihat pada Tabel 3.2 berikut.

Tabel 3.2 Pengujian MiniBatch

Pengujian	Epoch	MiniBatchSize	Akurasi
1	300	10	85,7%
2	300	20	65.6%
3	300	30	67.18%
4	300	40	67.1%
5	300	50	63%

Pada tabel 4.2 diatas dilakukan lima kali pengujian dengan nilai MiniBatchSize yang berbeda, didapatkan nilai akurasi tertinggi sebesar 85.7% pada pengujian pertama dengan nilai MiniBatchSize 10.

3.4. Pengujian Datalatih terhadap Data Uji

Setelah dilakukan pengujian diatas, didapatkan nilai terbaik dari masing-masing parameter yaitu 300 pada nilai epoch dan 10 pada nilai MiniBatch, untuk selanjutnya dilakukan pengujian terhadap data latih terhadap data uji, untuk hasil dari pengujian ini bisa dilihat pada Tabel 4.3 berikut

Tabel 3.3 Pengujian Datalatih terhadap Data Uji

Data Latih	Data Uji	Akurasi
DATA BASE	BOGOR	100%

	SUNDA LAIN	74.19%
	BUKAN SUNDA	48%

3.5 Pengujian Datalatih terhadap Data Uji

Pada tahap ini dilakukan pengujian untuk setiap kalimat yang digunakan pada penelitian ini. Pengujian ini menggunakan nilai terbaik dari masing-masing parameter yang sudah ditemukan pada pengujian sebelumnya.

Kalimat Ke-	Kalimat yang digunakan	Akurasi		
		Bogor	Sunda Lain	Bukan Sunda
1	Tempat tuang caket dieu palih mana?	100%	100%	100%
2	Ngawulang murang kalih mapah	100%	100%	66%
3	Jalan ka bandung palih mana	100%	100%	83%
4	Mun murangkalih gering panas make obat naon?	100%	100%	75%
5	Kumaha cara neda nu sae	100%	100%	83%

4. Kesimpulan

Bedasarkan dari hasil Tugas Akhir ini, dapat disimpulkan bahwa ::

1. Nilai Epoch sangat berpengaruh terhadap tingkat akurasi karena epoch mempengaruhi banyaknya pembelajaran yang dilakukan. Semakin besar nilai epoch maka, semakin banyak pembelajaran yang dilakukan oleh system.
2. Nilai akurasi tertinggi yang didapatkan oleh sistem yaitu 85.7% dengan nilai setiap parameter adalah sebagai berikut.
 - a. MiniBatch size : 10
 - b. Epoch : 300
3. Banyaknya data latih berpengaruh, karena semakin banyak data latih, maka semakin banyak ciri yang didapatkan dan hasil akurasi yang didapatkan menjadi lebih baik.
4. Akurasi yang dihasilkan dari perbandingan database pada setiap data uji adalah sebagai berikut.
 - a. Orang Sunda Bogor : 100%
 - b. Orang Sunda lain : 74.19%
 - c. Bukan Sunda : 48%
5. Nilai epoch dan MiniBatch mempengaruhi hasil akurasi yang didapatkan, semakin tinggi nilai epoch yang dimasukan, semakin baik nilai akurasi yang didapatkan sedangkan pada MiniBatch semakin kecil nilai yang dimasukan, semakin baik nilai akurasi yang didapatkan.

Daftar Pustaka:

- [1] Misalkan Wurm dan ShirôHattori dalam Language Atlas of Asia-Pacific (1983)
- [2] Venkateswarlu, Dr.R.L.K. Kumari Dr.R.Vasantha, JayaSri G.Vani. "Speech Recognition By Using Recurrent Neural Networks" (2011).
- [3] Klaus Greff; Rupesh Kumar Srivastava; Jan Koutn ĩ; Bas R. Steunebrink; Jürgen Schmidhuber. "LSTM: A Search Space Odyssey"(2015)
- [4] Hasim Sak, Andrew Senior, Francoise Beaufays. "Long Short-Term Memory Recurrent Neural 9
- [5] Nal Kalchbrenner, Ivo Danihelka, Alex Graves. "Grid Long Short-Term Memory". (2015).
- [6] Margaret Rouse, __. Speech Recognition [online] (<http://searchcrm.techtarget.com/definition/speech-recognition> diakses tanggal 4 Oktober.
- [7] Sepp Hochreiter, Jurgen Schmidhuber. " Long Short-Term Memory" (1997)
- [8] Tomas Mikolov, Martin Karfiat, Lukas Buget, Jan Honza Cernocky, Sanjeev Khundapur. "Recurrent Neural Network based language model" (2010)
- [9] Hu Yu Hen dan Jeng-Neng Hwang. "Hardbook of Neural Network Signal Processing" (2002)..
- [10] Alex Graves, Navdeep Jaitly . "Toward end-to-end speech recognition with recurrent neural network" (2014)
- [11] R. S. Chavan dan G. S. Sable, "An Overview of Speech Recognition Using HMM," International Journal of Computer Science and Mobile Computing, vol. 2, no. 6, 2013.