

RECURRENT NEURAL NETWORK UNTUK PENGENALAN UCAPAN PADA BAHASA SUNDA SELATAN DIALEK GARUT

RECURRENT NEURAL NETWORK FOR SPEECH RECOGNITION ON SOUTH SUNDANESE GARUT DIALECT

¹Luthfi Abdul Hakim, ²Andrew Brian Osmond, ³Randy Erfa Saputra

¹²³Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹luthfi111213@gmail.com, ²abosmond@telkomuniversity.ac.id, ³resaputra@telkomuniversity.ac.id

Abstrak

Bahasa adalah alat yang digunakan manusia untuk berkomunikasi. Pada sebuah bahasa memiliki bermacam-macam dialek, salah satunya adalah bahasa sunda. Bahasa sunda ini memiliki 6 dialek, yaitu dialek Barat (bahasa banten), dialek Utara, dialek Selatan (priangan), dialek Tengah Timur, dialek Timur laut (bahasa Cirebon), dialek Tenggara. Dengan kata lain dialek ini juga merupakan pembeda (ciri khas) dalam sebuah bahasa. Bisa dikatakan bahwa dialek memiliki pola yang membedakan antara satu dialek dengan dialek yang lain.

Pembedaan pola ini bisa dilakukan dengan menggunakan *recurrent neural network* dengan memfokuskan pada pengambilan sampel yang baik, dan penggunaan fungsi utama *recurrent neural network* yang membaca data secara sekuensial (sejalan dengan waktu), metode *recurrent neural network* yang digunakan disini adalah *long-short term memory*.

Pada penelitian ini didapat hasil epoch terbaik adalah 250, minibatch size terbaik adalah 10 dan hidden size terbaik adalah 50, dan akurasi kecocokan antara data uji dan data latih yang di dapat adalah saat diuji dengan bahasa sunda dialek garut sebesar 100%, diuji dengan bahasa sunda yang dialeknnya bukan garut sebesar 77,778%, dan untuk orang yang bukan sunda sebesar 66,667%. dari penelitian ini dapat disimpulkan bahwa pengaruh besar nilai epoch, minibatch size, hidden size dan banyak data latih berpengaruh terhadap nilai akhir.

Kata kunci : Bahasa sunda, dialek, recurrent neural network, long-short term memory.

Abstract

Language is a tool that humans use to communicate. In a language has a variety of dialects, one of which is the language of Sundanese. This Sundanese language has 6 dialects, which are Western dialect (banten language), Northern dialect, Southern dialect (Priangan), Middle Eastern dialect, Northeastern dialect (Cirebon language), Southeastern dialect. In other words this dialect is also a differentiator in a language. It can be said that dialect has a pattern that distinguishes between one dialect with another dialect.

The differentiation of this pattern can be done using a recurrent neural network by focusing on good sampling, and the use of the main recurrent neural network function that reads data sequentially (over time), the recurrent neural network method used here is long-short term memory.

In this research, the best epoch result is 250, the best minibatch size is 10 and the best hidden size is 50, and the accuracy of the match between the test data and the train data can be tested with Sundanese dialect garut 100%, tested with Sundanese language the dialect is not garut of 77.778%, and for people who are not Sundanese 66.667%. from this research can be concluded that the influence of the value of epoch, minibatch size, hidden size and a lot of training data affect the final value.

Keywords: language of Sundanese, dialect, recurrent neural network, long-short term memory

1. Pendahuluan

1.1 Latar Belakang

Saat ini banyak aplikasi yang berfungsi untuk mengubah ucapan menjadi teks, metode yang digunakan adalah *recurrent neural network*, akan tetapi penulis tidak menemukan adanya penelitian mengenai *speech recognition* terhadap dialek. Maka dari itu untuk mengetahui bagaimana *recurrent neural network* jika melakukan pengambilan pola seseorang dalam berbicara terhadap dialeknnya.

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 selatan dengan dialek Garut 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 Garut

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.

2. Dasar Teori

1.1 Audio Digital

Audio atau suara merupakan gelombang yang mengandung sejumlah komponen penting (amplitudo, panjang gelombang dan frekuensi) yang dapat menyebabkan suara yang satu berbeda dari suara lain. Audio digital merupakan versi digital dari suara analog. Pengubahan suara analog menjadi suara digital membutuhkan suatu alat yang disebut Analog to Digital Converter (ADC). ADC akan mengubah amplitudo sebuah gelombang analog ke dalam waktu interval (sampel) sehingga menghasilkan penyajian digital dari suara.

1.2 Speech Recognition

Speech Recognition atau yang dikenal sebagai *Automatic Speech Recognition* (ASR) adalah kemampuan mesin atau program untuk mengidentifikasi kata dan frase dalam bahasa lisan [1]. Input ucapan yang di keluarkan makhluk hidup akan diolah menjadi sebuah sinyal analog lalu sinyal analog tersebut di ubah menjadi sebuah menjadi sinyal digital yang kemudian disesuaikan dengan pola tertentu untuk mengidentifikasi informasi apa yang ada pada sinyal tersebut. [2] Teknologi ini memungkinkan suatu perangkat dapat mengenali kata-kata dengan cara menganalisis spesifikasi kata yang disebutkan lalu mendigitalisasi kata dan mencocokkan sinyal digital tersebut dengan pola tertentu yang tersimpan [3].

1.3 Mel-Frequency Cepstral Coefficient (MFCC)

MFCC merupakan salah satu metode ekstraksi ciri untuk sinyal akustik terbaik [4]. Analisis suara pada *mel frequency* didasarkan pada persepsi pendengaran manusia, karena telinga manusia telah diamati dapat berfungsi sebagai filter pada frekuensi tertentu. *Mel Frequency Cepstrum Coefficients* (MFCC) merupakan 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. MFCC memiliki dua jenis filter yang spasi linear pada frekuensi rendah di bawah 1000 Hz dan logaritmik di atas 1000 Hz [5].

1.4 Recurrent Neural Network

Recurrent neural network atau jaringan saraf tiruan ini dikenal sebagai teknik yang memiliki kemampuan untuk mengklasifikasi masalah. Sebuah sistem yang memproses data secara berulang, menggunakan kembali data yang telah digunakan pada proses sebelumnya untuk memproses data input yang selanjutnya. [7]

1.5 Long Short-Term Memory

Asal mula LSTM adalah karena RNN memiliki kekurangan dalam prosesnya yaitu ketika file keluaran akan dimasukan kembali sebagai masukan, keluaran dari proses RNN tersebut memungkinkan mengeluarkan hasil yang sama dengan nilai sebelumnya. Maka dibuatlah memori agar bisa menyimpan keluaran sebelumnya agar tidak terjadi keluaran yang sama seperti sebelumnya.

Ketika keluaran akan dimasukan kembali sebagai masukan yang baru akan disimpan pada memori tetapi pada tahap ini memori hanya bisa menyimpan dan menghapus hasil rekaman dari RNN saja maka ditambah lah selection.

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

Neural network yang digunakan pada penelitian ini *recurrent neural network*. Langkah utama yang dilakukan untuk penelitian ini adalah pencarian sampel berupa rekaman seseorang yang berbicara dengan bahasa sunda dialek garut, yang sebagiannya akan digunakan sebagai data latih dan sebagiannya lagi digunakan untuk data uji serta rekaman seseorang yang berbicara dengan bahasa sunda dialek bogor sebagai data uji.

Setelah data suara terkumpul dilakukan proses menghilangkan gangguan dalam suara yang direkam (*noise*) menggunakan program yang sudah disediakan lalu setelah proses menghilangkan *noise* dilakukan, proses selanjutnya adalah membuat *database* dengan menggunakan sebagian dari rekaman seseorang yang berbicara dengan bahasa sunda dialek garut untuk dilatih, dan *database* dengan menggunakan sisa rekaman seseorang yang berbicara dengan bahasa sunda dialek garut sebagai data uji ke-1, serta *database* rekaman orang sunda yang berbicara dengan bahasa sunda tetapi bukan dialek garut sebagai data uji ke-2, dan *database* orang yang bukan orang sunda berbicara bahasa sunda sebagai data uji ke-3.

3. Pengujian

3.1. Pengujian Epoch

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 Jumlah *Epoch*

Pengujian ke-	<i>Epoch</i>	<i>Hidden unit</i>	<i>Mini Batch Size</i>	Akurasi
1	50	50	20	61.111%
2	100	50	20	63.889%
3	150	50	20	69.444%
4	200	50	20	72.222%
5	250	50	20	75%

Berdasarkan hasil pengujian pada Tabel 3.1 didapatkan hasil akurasi terbesar ketika *epoch*=250 dengan nilai *Mini Batch Size* sebesar 20 dan *Hidden* sebesar 50. Dari hasil pengujian akan diuji kembali dengan menggunakan parameter *minibatchsize* yang berbeda untuk membuktikan apakah ada perubahan nilai pada akurasi.

3.2. Pengujian Hidden Size

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

Tabel 3.2 Hasil Pengujian Nilai *Hidden Size*

Pengujian ke-	<i>Epoch</i>	<i>Hidden size</i>	<i>Mini Batch Size</i>	Akurasi
1	250	10	20	72.222%
2	250	30	20	75%
3	250	50	20	77.778%
4	250	60	20	77.778%
5	250	70	20	77.778%

Berdasarkan hasil pengujian pada Tabel 3.2 didapatkan bahwa hasil Akurasi setelah nilai *hidden size* yang dihasilkan sebesar 50 keatas akurasi tidak bertambah baik dan tidak bertambah buruk yang artinya ketika *hidden size* mencapai titik jenuh maka nilai akurasi tidak akan bertambah baik maupun bertambah buruk.

3.3 Pengujian Nilai Batch Size

Pengujian dilakukan untuk mengetahui pengaruh nilai *Mini Batch Size* terhadap performansi jaringan yang telah dibuat, dan mencari nilai *Mini Batch Size* terbaik yang akan digunakan untuk proses pelatihan. Berikut ini adalah hasil pengujian yang dilakukan berdasarkan perubahan nilai *Mini Batch Size* yang berbeda-beda dengan *database* yang jumlah *hidden size* sebanyak 50 dan nilai *epochs* sebesar 250.

Tabel 3.3 Hasil Pengujian Nilai Batch Size

Pengujian ke-	<i>Epoch</i>	<i>Hidden size</i>	<i>Mini Batch Size</i>	Akurasi
1	250	50	10	77.778%
2	250	50	20	75%
3	250	50	30	75%
4	250	50	40	75%
5	250	50	50	75%

Berdasarkan hasil pengujian pada Tabel 3.3 didapatkan bahwa hasil akurasi terbaik yaitu pada saat nilai *Mini Batch Size* sebesar 10 dengan besar akurasi 77.778%. tidak bisa disimpulkan jika *Mini Batch Size* kecil bisa menaikkan akurasi karena *mini batch size* ketika sudah mencapai angka di bawah 10 maka akan menemui hasil acak karena dalam *mini batch size* ketika nilainya dibawah 10 sistem akan dengan acak menambahkan iterasi yang nilainya akan berpengaruh kepada banyaknya *epoch*.

3.4 Pengujian Data latih Terhadap Data Uji

Pengujian ini dilakukan dengan menggunakan bahasa sunda dialek garut sebagai data latih dan data uji yang digunakan disini ada 3 yaitu dengan bahasa sunda dialek garut, bahasa sunda dialek selain garut, bukan orang sunda, dalam percobaan ini *epoch*, *minibatch size*, *hidden size* yang digunakan adalah yang mendapatkan akurasi terbaik pada percobaan sebelumnya yaitu dengan *epoch* 250, *hidden size* sebesar 50, dan *minibatch size* sebesar 10.

Tabel 3.4 Hasil Perbandingan data latih dengan data uji

Data Train yang digunakan	Data latih yang digunakan	akurasi
Bahasa sunda dialek garut	Orang sunda berdialek garut	100%
Bahasa sunda dialek garut	Orang sunda bukan dialek garut	77,778%
Bahasa sunda dialek garut	Bukan Orang Sunda	66,667%

Pada tabel 3.4 didapatkan bahwa hasil akurasi terbaik adalah Orang Garut yaitu sebesar 100%

3.5 Pengujian metode terhadap kalimat

Untuk mengetahui berapa persentase kecocokan dari tiap-tiap kalimat maka dilakukan pengujian perkalimat, karena pada pengujian sebelumnya hanya mencek antara database utama dengan data latih, dengan menggunakan parameter terbaik dari pengujian sebelumnya yaitu *epoch* sebesar 250 *hidden size* sebesar 50 dan *MiniBatch Size* sebesar 10. Hasil dari percobaan ini terdapat pada tabel 3.5.

Tabel 3.5 Hasil Pengujian PerKalimat

Kalimat ke-	Kalimat yang digunakan	Akurasi		
		Garut	Bukan Garut	Bukan Orang Sunda
1	Tempat tuang caket dieu palih mana?	100%	50%	33,333%
2	Ngawulang murang kalih mapah	100%	50%	33,333%
3	Jalan ka bandung palih mana	100%	100%	66,6667%
4	Mun murangkalih gering panas make obat naon	100%	50%	37,5%
5	Ngadangukeun lagu bari nyetiran teh picilakaeun komo deui bari nelepon	100%	66,6667%	57,1429%

4. Kesimpulan

Berdasarkan proses-proses yang telah dilakukan pada penelitian tugas akhir ini, dimulai dari perancangan sampai dengan pengujian dan analisis sistem, dapat disimpulkan beberapa hal diantaranya :

1. Recurrent neural network dapat digunakan untuk program pengenalan suara khususnya untuk bahasa sunda selatan dialek garut.
2. Pelatihan untuk mendapatkan tingkat akurasi yang tinggi dalam proses ini adalah sebagai berikut :
 - a. *MiniBatch size* : 10
 - b. *Hidden size* : 50
 - c. Epoch : 250
3. Akurasi yang dihasilkan dari perbandingan data base pada setiap data uji adalah
 - a) Orang sunda berdialek Garut. : 100%
 - b) Orang sunda selain dialek Garut : 77,78%
 - c) Bukan orang sunda : 66,66%

Daftar Pustaka:

- [1] Margaret Rouse, "Speech Recognition" [online] (<http://searchcrm.techtarget.com/definition/speech-recognition> diakses tanggal 3 Juni 2017).
- [2] John Coleman, "Introducing Speech and language processing", Cambridge university press, 2005.
- [3] Dr.R.L.K.Venkateswarlu, "Speech Recognition By Using Recurren Neural Networks", International Journal of Scientific & Engineering Research Volume 2, 2011 .
- [4] 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.
- [5] Muhammad Nashih Rabbani, "Implementation of Voice Recognition Based Key Using Mel Frequency Cepstral Coefficient (MFCC)" e-Proceeding of Engineering : Vol.3, No.3 Desember 2016.
- [6] Angga Setiawan, "Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer" Journal TRANSMISI, 13 (3), 2011, 82-86.
- [7] L. Mangu and A. Emami, "Empirical study of neural network language models for arabic speech recognition," in Proc. IEEE Workshop Autom. Speech Recognit. Understanding, Kyoto, Japan, 2007, pp. 147-152.
- [8] Jasha droppo "exploiting LSTM Structure IN deep neural network for speech" shanghai education commision.
- [9] Prof. R. Hersch, "long short term memory in recurrent neural network",Lausanne EPFL, 2001.