SEGMENTASI SINYAL SUARA DENGAN MENGGUNAKAN JARINGAN SYARAF TIRUAN LONG SHORT-TERM MEMORY (LSTM) SPEECH SEGMENTATION BY USING LSTM NEURAL NETWORK

Prima Febriansyah[1], Suyanto[2], -[3]

[1]Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Segmentasi sinyal suara adalah upaya untuk melakukan pengklasifikasian sinyal suara berdasarkan suatu hal tertentu. Segmentasi dapat dilakukan berdasarkan kata, berdasarkan suku kata, atau berdasarkan fonem. Proses segmentasi ini dapat menjadi proses awal dari proses pengenalan sinyal suara. Dalam Tugas Akhir ini, segmentasi sinyal suara akan dilakukan berdasarkan fonem. Langkah awal dalam proses segmentasi ini adalah pembacaan dan pensegmentasian sinyal suara secara manual.. Kemudian dilanjutkan dengan proses feature extraction terhadap masing – masing segmen. Hasil dari proses feature extraction ini adalah ciri dari masing – masing segmen. Ciri ini kemudian akan digunakan sebagai parameter masukan Jaringan Syaraf Tiruan (JST). Selanjutnya JST akan melakukan proses learning terhadap ciri-ciri tersebut. Penelitian yang telah ada sebelumnya menggunakan 15 parameter dengan JST Propagasi Balik. Sedangkan dalam Tugas Akhir ini akan digunakan JST LSTM (Long Short-Term Memory) dengan 3 parameter masukan JST, yaitu Zero Crossing Rate (ZCR) , Signal Energy Rate (SER) , dan Maximum Frequency Index (MFI).

Kata Kunci : segmentasi sinyal suara, zero crossing rate, signal energy rate, maximum frequency index, feature extraction.

Abstract

Speech Segmentation is an effort to classify the speech signal base on particular thing. The segmentation can be done base on word, syllable, or phonem. The segmentation process can be a preprocessing of speech recognition. In this final project, the segmentation of speech is done base on phonem. First step in segmentation process is getting the signal and do the manual segmentation. Then, feature extraction proses is done to each segmen. Feature extraction process result the characteristics of segmen. This characteristics are used as input of Neural Network later and then learning process is done. The last research used 15 input parameter and Back Propagation was used as the Neural Network. While this final project used LSTM (Long Short-Term Memory) Neural Network with 3 input parameter, namely Zero Crossing Rate (ZCR), Signal Energy Rate (SER), and Maximum Frequency Index (MFI).

Keywords : segmentation of speech, zero crossing rate, signal energy rate, maximum frequency index, feature extraction.

1. Pendahuluan

1.1 Latar belakang

Terdapat dua macam pendekatan dalam metode pengenalan kata bahasa Indonesia, yang pertama adalah mengenali suku kata sebagai satu kesatuan pola dan yang lain dengan mengenali kata sebagai kumpulan dari fonem. Pendekatan pertama kurang menguntungkan karena jumlah kata dalam bahasa Indonesia sangat besar. Sedangkan pendekatan ke-dua merupakan metode yang sulit dilakukan, tetapi lebih menarik untuk pengembangan lebih lanjut. Pendekatan kedua membutuhkan proses segmentasi untuk memisahkan fonem vokal dan fonem konsonan.

Penelitian yang telah ada sebelumnya tentang segmentasi sinyal ucapan, menggunakan JST Back Propagation dan metode Hybrid, ternyata memberikan hasil yang kurang baik.

Terdapat arsitektur baru dalam JST, yaitu LSTM (*Long Short-Term Memory).* Jika dilihat dari arsitekturnya, LSTM diharapkan memiliki performansi yang lebih baik dari Back Propagation dalam melakukan segmentasi sinyal ucapan.

1.2 Perumusan masalah

Dari latar belakang masalah diatas dapat dirumuskan masalah berikut :

1. Bagaimana kemampuan LSTM dalam melakukan segmentasi sinyal ucapan ?
2. Bagaimana mensimulasikan metode LSTM untuk segmentasi sinyal ucapan ?

1.3 Tujuan

Tujuan tugas akhir ini adalah membuat simulasi dari JST LSTM untuk melakukan segmentasi sinyal suara, menentukan fitur dari sinyal suara sebagai masukan ke LSTM, melihat kemampuan JST LSTM untuk segmentasi sinyal ucapan.

1.4 Batasan masalah

Dalam penelitian ini, objek penelitian dibatasi dengan ruang lingkup sebagai berikut :

1. Data yang dijadikan pembelajaran dan pengujian bagi LSTM telah ditentukan sebelumnya, yaitu ucapan *isolated digit.*
2. Jaringan Syaraf Tiruan yang dipakai adalah JST LSTM.
3. Pengambilan data suara dilakukan di ruangan sunyi dengan *noise* rendah, berfrekwensi 16000 Hz, dan berlevel 16 bit.
4. Lebar segmen yang digunakan adalah 480 sampel.

1.5 Metodologi

Metodologi yang dilakukan dalam pelaksanaan tugas akhir ini :

1. Studi literatur, dilakukan dengan mencari informasi dan referensi yang berkaitan dengan topik ini
2. mempelajari fitur sinyal ucapan, segmentasi sinyal ucapan, dan LSTM,
3. perancangan dan implementasi,
4. analisis dan evaluasi,
5. serta pengambilan kesimpulan.

1.6 Sistematika Penulisan

Untuk kemudahan dalam penyusunan tugas akhir ini maka diperlukan sistematika penulisan yang dibagi dalam beberapa bab, dan masing-masing bab menjelaskan permasalahan yang saling berkaitan. Adapun susunan penulisannya dirinci sebagai berikut :

BAB I : Pendahuluan
Dalam bab ini akan diurai tentang latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi, sistem penulisan, jadwal pengerjaan.

BAB II : Landasan Teori
Dalam bab ini akan diurai tentang teori yang menunjang pembahasan tentang segmentasi sinyal suara dan LSTM.

BAB III : Perancangan Sistem
Dalam bab ini akan dibuat rancangan sistem meliputi diagram blok dan implementasinya.

BAB IV : Pengujian dan Implementasi Sistem
Bab ini akan menguji sistem dalam melakukan segmentasi sinyal suara

BAB V : Kesimpulan dan Saran
Penutup meliputi kesimpulan dan saran atas penelitian ini.

5. Pengujian dan Implementasi Sistem

5.1 Kesimpulan

Kesimpulan yang didapat dari pengujian sistem adalah :

1. JST LSTM berhasil mengklasifikasikan fonem – fonem bahasa Indonesia dengan baik.
2. Kombinasi *Zero Crossing Rate*, *Signal Energy Rate*, dan *Maximum Index Frequency* sebagai parameter masukan JST dari hasil *feature extraction* ternyata menghasilkan hasil sebesar 84.62% untuk akurasi tiap segmen, 30.2479 (ms) untuk waktu rata-rata pergeseran, dan 54 % untuk akurasi per kelompok segmen, 103.4405 % untuk *insertion*, dan 5.275 % untuk *deletion*.

5.2 Saran

Saran penulis untuk meningkatkan performansi sistem klasifikasi ini diantaranya :

1. Melakukan proses *filtering* terhadap sinyal suara yang akan diklasifikasi. Hal ini bertujuan untuk menghilangkan *noise-noise* yang bisa mempengaruhi hasil dari ekstraksi fitur, sehingga mempengaruhi hasil pelatihan JST.
2. Menambah atau menggunakan fitur yang lain dari sinyal suara sebagai parameter masukan ke JST.
3. Pemotongan huruf dilakukan dengan cara lebih obyektif yaitu dilakukan oleh beberapa orang sehingga didapatkan hasil yang obyektif pula.

Daftar Pustaka

[1] A. Saputra Eka Salehudin, Haryatno Joko Ir, Ad Dairi Hakim S, Ir,(1998). **"Aplikasi Jaringan Syaraf Tiruan Probabilistik Untuk Mengkonversikan Sinyal Suara Ke String Pada Suku Kata Bahasa Indonesia Terbatas Dengan Pradeteksi Histogram".**

[2] Puspitaningrum Diyah, ST,M.Kom, (2006). "**Pengantar Jaringan Syaraf Tiruan**". Andi Offset. Yogyakarta.

[3] Hochreiter, Sepp and Schmidhuber, Juergen, (1997) "**Long Short-Term Memory**", *Neural Computation*, Vol 9 (8), pp: 1735-1780.

[4] Jacobson Anders,Gustavsson Christian, (1995)."**Prediction Of Residue Number Contacts in Protein Using LSTM Neural Network**".

[5] Graves Alex, Eck Douglas, Beringer Nicole, Schmidhuber Juergen. **"Biologically Plausible Speech Recognition with LSTM Neural Nets"**.

[6] Jek, Jong Siang Drs, M.Sc, (2005). "**JST & Pemrogramannya Menggunakan Matlab"**. Andi Offset. Yogyakarta.

[7] Schmidhuber Jurgen."**Long Short-Term Memory : Tutorial On LSTM Recurrent Network**". [http://www.idsia.ch]

[8] Putra Husni Muhammad,(1998). "**Sistem Pengenalan Pembicara Dengan Teks Bebas Menggunakan Metode Pengukuran Gaussian Likelihood**". Tugas Akhir pada Jurusan Teknik Informatika STTTelkom Bandung.

[9] Suyanto ST, (1992). "**Buku Ajar AI – Suyanto IF.pdf**". Bandung : STT Telkom.

[10] Suyanto,(1998). "**Penerapan Jaringan Syaraf Tiruan Probabilistik dan Transformasi Wavelet Untuk Mengkonversi Sinyal Suara ke String Ucapan Suku Kata Bahasa Indonesia**". Tugas Akhir pada Jurusan Teknik Informatika STTTelkom Bandung.

[11] Maharani Warih, (1999). "**Perbandingan Metofe Hybrid HMM/MLP dan Metode HMM Dalam Proses Pengenalan Ucapan Berbahasa Indonesia**". Tugas Akhir pada Jurusan Teknik Informatika STTTelkom Bandung.

[12] Matlab Help.