

IMPLEMENTASI SISTEM PENGENALAN SUARA PADA APLIKASI ASISTEN PRIBADI KOMPUTER DENGAN BERBASIS HIDDEN MARKOV MODEL DAN VITERBI SEARCH

Dhilla Rahadian Maharantau¹, Suyanto², Agung Toto Wibowo³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Memodelkan sinyal model untuk pengenalan suara adalah suatu tugas yang menantang. Pemodelan tersebut akan menuntut kita sejumlah besar informasi tentang masalah yang akan dimodelkan. Sistem pengenalan suara umumnya mengasumsikan bahwa sinyal suara adalah realisasi dari beberapa pesan yang terkodekan dalam satu atau lebih sekuen simbol-simbol. Simbol-simbol ini disebut dengan fitur suara (sekuen vektor suara yang teramati). Dalam pengenalan suara berbasis HMM, biasanya diasumsikan bahwa sekuen vektor suara yang teramati tersebut akan berkoresponden dengan setiap kata, yang direpresentasikan oleh suatu model suara yang disebut Markov model. Dalam tugas akhir ini, fitur-fitur suara akan dianalisa dengan MFCC (Mel-Frequency Cepstral Coefficient), dan digunakan algoritma Viterbi untuk mencari jalur rangkaian terbaik dari sekuen simbol-simbol yang berkoresponden dengan suatu kata yang akan dikembalikan sebagai hasil pengenalan.

Kata Kunci : fitur, MFCC, HMM, algoritma Viterbi

Abstract

Modeling signal model for speech recognition is challenging task. It gives us great deal of information about problem being modeled. Speech recognition systems generally assume that the speech signal is a realization of some message encoded as a sequence of one or more symbols. These symbols called speech features (sequence of observed speech vectors). In HMM based speech recognition, it is assumed that the sequence of observed speech vectors corresponding to each word, represented by a speech model called Markov model. In this final project, speech features analyzed with MFCC (Mel-Frequency Cepstral Coefficient), and Viterbi algorithm is used to find the best path of symbols sequence corresponding to a word and return it as recognition result.

Keywords : feature, MFCC, HMM, Viterbi algorithm

Telkom
University

BAB I

PENDAHULUAN

1. Latar belakang masalah

Sistem pengenalan suara (*speech recognition system*) pada dasarnya adalah kumpulan proses pengubahan sinyal suara akustik yang diterima oleh alat input suara semisalnya *microphone*, menjadi serangkaian data digital (*byte, binary code, dsb*), untuk kemudian diambil karakteristik-karakteristik khususnya, dan dikenali dengan cara membandingkan karakteristik-karakteristik khusus tersebut dengan data pengetahuan yang dimiliki [3]. Sistem pengenalan suara pada umumnya mengasumsikan bahwa sinyal suara adalah suatu realisasi dari sekuen pesan yang terkodekan kedalam satu atau lebih simbol-simbol [11].

Selayaknya ciri-ciri biometrik unik yang ada pada manusia, bahwa pada suara manusia terdapat informasi dan karakteristik-karakteristik unik yang memungkinkan suara tersebut untuk dikenali [11]. Sehingga, apabila informasi dan karakteristik-karakteristik suara tersebut dapat diperoleh dengan suatu proses ekstraksi, dan sistem memiliki pengetahuan akan suara tersebut, maka memungkinkan suara tersebut untuk dikenali oleh sistem.

Dalam sistem pengenalan suara, sinyal suara perlu dimodelkan ke dalam bentuk matematik sehingga dapat dikomputasi oleh sistem. Hidden Markov Model merupakan model statistik yang umum digunakan untuk memodelkan sinyal suara dalam sistem pengenalan suara. Pada sistem pengenalan suara dengan HMM, sinyal suara input akan mengalami *front-end processing*. Hasil dari *front-end processing/frequency analysis* berupa segmen-segmen suara dan fitur-fitur yang kemudian akan dikirimkan ke *decoder* untuk dilakukan pengenalan dengan menghitung *likelihood* terhadap model pengetahuan HMM yang terdiri atas *acoustic model, language model, dan dictionary*. Proses pengenalan dapat diartikan sebagai suatu proses memaksimalkan nilai probabilitas data pengamatan terhadap model suara (*likelihood*), dinyatakan dalam $P(O|\lambda)$, dari suatu sekuen pengamatan signal suara $O=O_1, O_2, \dots, O_T$ terhadap model yang diberikan $\lambda=(A, B, \pi)$ [2].

Fokus pembahasan dan pengerjaan tugas akhir ini akan dilakukan pada *speech recognizing* dan aplikasi pemanfaatannya. Dalam penerapannya, dilakukan dengan memanfaatkan komponen-komponen serta *tool-kit* yang dapat diperoleh secara gratis seperti *Java SAPI* sebagai program antarmuka, serta template dataset model yang diperlukan untuk sebagai sumber pengetahuan sistem.

2. Perumusan masalah

Permasalahan yang timbul dari latar belakang pembuatan tugas akhir ini diantaranya sebagai berikut:

1. Bagaimana memanfaatkan dan mengimplementasikan *Hidden Markov Model* pada suatu *speech recognition system*.
2. Membuat aplikasi *speech personal asisten* pada komputer dengan memanfaatkan *speech engine* yang telah dibuat untuk menangani konteks *command and control Windows* pada *isolated word recognition*.
3. Bagaimana mengukur tingkat akurasi berdasar jumlah kebenaran respon yang diberikan aplikasi pada input sejumlah user, serta performansi dalam kaitannya dengan waktu respon dan penggunaan memori.

3. Tujuan

Beberapa tujuan yang ingin dicapai dari pengerjaan tugas akhir ini nantinya dapat dijelaskan sebagai berikut:

1. Membuat suatu sistem pengenalan suara dengan berbasis *Hidden Markov Model*. Suatu sistem pengenalan suara terdiri dari modul-modul yang mendefinisikan dan melakukan pemrosesan pengenalan berdasarkan input yang diterima dan model pengetahuan yang ada.
2. Menerapkan sistem pengenalan suara yang telah dibuat ke dalam suatu aplikasi *Speech Computer Personal Assistant*, yang mampu menjalankan sejumlah aksi pada sistem operasi Windows dengan perintah yang diberikan melalui ucapan.

3. Melakukan pengukuran akurasi terhadap hasil pengerjaan dengan menghitung *word error rate*, performansi dalam kaitannya dengan waktu respon, dan penggunaan sumber daya memori, serta gambaran implementasi sistem pada platform komputer dan jaringan telepon.

4. Batasan masalah

Adapun batasan-batasan masalah yang diberikan pada tugas akhir ini adalah :

- Input suara yang dikenali adalah pengucapan dalam bahasa Inggris serta menggunakan template data pengetahuan yang telah tersedia.
- Tipe aplikasi pengenalan suara adalah *speaker independent* (tidak memperhatikan/tergantung pada pembicara tertentu)
- Implementasi pada konteks sistem *command and control Windows* dengan kata perintah yang dikenali telah ditentukan dan dibatasi.
- Perintah yang diterima sistem pengenalan suara untuk menjalankan aksi di Windows hanya terbatas pada beberapa kata perintah berikut: *start, windows, open note, open word, open excel, open music, press enter, press escape, press up, press down, press left, press right, switch dan release.*

5. Metodologi penyelesaian masalah

Metodologi penelitian yang dilakukan dalam penyusunan tugas akhir ini meliputi:

1. Studi literatur

Studi literatur dilakukan untuk memahami teori dasar mengenai *Java SAPI, Hidden Markov Model, Viterbi algorithm*, optimasi fungsi $P(O|\lambda)$, serta teori-teori dasar lain tentang teknologi *speech recognition* itu sendiri.

2. Analisis Perancangan Perangkat Lunak

Dilakukan proses analisis *requirement* dari sistem yang akan dibangun sehingga didapat gambaran mengenai sistem yang akan dibuat.

3. Implementasi Desain Sistem

Melakukan implementasi terhadap hasil desain sistem yang telah dilakukan dengan menggunakan aplikasi pembangun NetBeans IDE 6.7. Implementasi dimulai dengan pemrosesan terhadap input sinyal suara. Selanjutnya pembuatan *speech engine* dengan menerapkan HMM sebagai pemodelan suaranya, yang akan menghasilkan sebuah output text dari input yang berupa sinyal suara digital. Output dari *speech engine* akan menjadi input bagi aplikasi asisten untuk melakukan suatu aksi pada Windows.

4. Analisa Data dan Pelaporan

Menganalisis hasil implementasi aplikasi sehingga didapat data-data mengenai performansi dan akurasi dari metode yang diimplementasikan, serta mendokumentasikannya dalam bentuk laporan tugas akhir.

6. Jadwal kegiatan

Tabel 1: jadwal pengerjaan

| Kegiatan (2009) | April | | | | Mei | | | | Juni | | | | Juli | | | | Agustus | | | | September | | | |
|-----------------------------------|-------|---|---|---|-----|---|---|---|------|---|---|---|------|---|---|---|---------|---|---|---|-----------|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Studi Literatur | | | | | | | | | | | | | | | | | | | | | | | | |
| Analisa dan Desain | | | | | | | | | | | | | | | | | | | | | | | | |
| Implementasi | | | | | | | | | | | | | | | | | | | | | | | | |
| Analisa Hasil | | | | | | | | | | | | | | | | | | | | | | | | |
| Kesimpulan dan Penyusunan Laporan | | | | | | | | | | | | | | | | | | | | | | | | |



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Beberapa kesimpulan yang dapat diambil dengan memperhatikan hasil pengujian di Bab IV, diantaranya adalah sebagai berikut:

1. Sistem pengenalan suara yang dibuat telah berfungsi cukup baik dalam mengenali suara tanpa membedakan pembicara (*speaker independent*) dalam batasan 20 kata *grammar*.
2. Performansi sistem pengenalan ucapan dengan Hidden Markov Model, pada penerapannya di sistem komputer diperoleh dengan tingkat akurasi 76.47% dan pada sistem telepon diperoleh dengan tingkat akurasi 44.125%, dimana membuka kemungkinan untuk pengembangan dan pemanfaatan lebih lanjut.
3. Pemodelan suara dengan Hidden Markov Model cukup efektif untuk diterapkan pada sistem pengenalan suara sederhana *command and control*, dapat dilihat dari tingkat akurasi yang relatif baik dan penggunaan sumber daya memori yang rendah.
4. Mekanisme pencarian Viterbi memberikan efektifitas pencarian sekuen fonem penyusun kata dengan mengurangi ruang pencarian dimana pencarian hanya dilakukan pada fonem-fonem dengan nilai terbesar.
5. *Threshold* pada pencarian Viterbi berperan untuk mengurangi efek *noise* lingkungan pada sistem pengenalan suara.

5.2 Saran

1. Penggunaan sistem ini sebaiknya dilakukan di lingkungan dengan tingkat gangguan suara (*noise*) yang rendah, sehingga tingkat akurasi pengenalan diharapkan semakin bagus.
2. Untuk pengembangan ke depan, diharapkan dikembangkan aplikasi pengenalan suara yang tidak hanya terbatas pada beberapa kata saja (*continuous speech recognition*) yang mendukung dikte langsung melalui ucapan (*dictation*).
3. Perlu pengembangan lebih lanjut dalam rangka meningkatkan performansi aplikasi yang salah satunya dengan cara menggabungkan metode *hidden markov model* dengan metode lain seperti *genetic algorithm* (*Hybrid HMM/GA*).



Daftar pustaka

- [1] Rabiner, Lawrence, B. H. Juang. 1993. *Fundamentals of Speech Recognition*. Englewood Cliffs, New Jersey: Prentice-Hall.Inc.
- [2] Rabiner, Lawrence. 1989. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, IEEE, Vol. 77, No. 2, pp. 257-285, Februari 1989.
- [3] *Speech recognition*. http://en.wikipedia.org/wiki/Speech_Recognition. Tanggal akses: 1 juni 2008, 20.00 WIB
- [4] Zue, Cole, Ward. 2007. *Speech recognition*. www.speech_reg_cslu.cse.ogi.edu-HLTsurvey-ch1node.html . 1 juni 2009: 20.00 WIB
- [5] Suyanto. 2007. *An Indonesian Phonetically Balanced Sentence Set for Collecting Speech Database*. *Jurnal Teknologi Industri Vol. XI No. 1 Januari 2007: 59-68*
- [6] Rozak, Mike. 2008. *Talk to Your Computer and Have It Answer Back with the Microsoft Speech API*. <http://www.microsoft.com/msj/> . 11 Desember 2008: 10.00 WIB
- [7] Uchat, Nirav S. 2007. *Hidden Markov Model and Speech Recognition*. Indian Institute of Technology, Mumbai . India
- [8] Liao, Chun-Feng. 2008. *CMU Shpinx Speech Recognition Engine*. NCCU Dept. of Computer Sceince. Intelligent Media Lab
- [9] Laydrus, N.C., Ambikairajah, E.,Celler, B. 2007. *Automated Sound Analisis System For Home Telemonitoring Using Shifted Delta Cepstral Features*. School of Electrical Engineering and Telecommunications,The University of New South Wales, Sydney NSW 2052 Australia
- [10] Deller, JR, Proakis, JG, Hansen, JHL. 1993. *Discrete-time Processing of Speech Signals*, Macmillan Publishing Company
- [11] Steve Young and friends. 2006. *The HTK Book (for HTK Version 3.4)*. Cambridge University Engineering Department.
- [12] Maharani, Warih. 2001. *Perbandingan Metode Hybrid HMM/MLP Dan Metode HMM Dalam Proses Pengenalan Ucapan Berbahasa Indonesia*. Tugas Akhir. STT Telkom.
- [13] Paul, Dougla P., Janet, Baker M. 1991. *The Design for Wall Street Journal-based CSR Corpus*. MIT Lincoln laboratory, Dragon System Inc.