IF4071 Pemrosesan Suara

Tugas Besar II
Dessi Puji Lestari

Pembangunan Model Pengenal Ucapan untuk Bahasa Daerah (Bahasa Jawa)

Dikerjakan berkelompok sebanyak 5-6 orang

Tujuan

1. Melatih mahasiswa dalam membuat korpus ucapan khususnya bahasa Jawa.
2. Melakukan eksplorasi berbagai teknik pemodelan berbasis DNN dengan membaca makalah-makalah akademik dan sumber di web.
3. Melakukan eksperimen pembangunan model pengenal ucapan berbasis DNN (mis. Conformer, Transformer, RNN, dsb.) baik dari awal maupun fine-tune dari pretrained model yang ada.
4. Mengevaluasi secara kuantitatif menggunakan WER (Word Error Rate) dan PER (Phoneme Error Rate / Phone Error Rate/ Character Error Rate) serta analisis kualitatif.
5. Mempresentasikan hasil eksplorasi makalah dan hasil eksperimen.

Langkah Pengerjaan

Langkah 1: Pengumpulan Data

Setiap mahasiswa diminta

1. Membuat 30 kalimat berbahasa Jawa formal sehari-hari (mis. sapaan, permintaan sederhana, kalimat kegiatan harian).
   - Tidak menggunakan kata kasar/sensitif.
   - Transkrip harus lowercase, tanpa punctuation kecuali apostrof bila diperlukan.
   - JIka ada angka tuliskan dalam bentuk kata.
   - Tuliskan kalimat tersebut dalam **file csv bersama** dengan kolom id kalimat | transkrip.
     - Setiap mahasiswa memiliki speaker id unik (mis. speaker01, speaker02, ...).
     - Id kalimat: speaker<id>_[m|f]_[n|nn]_utt<xx>

- Speaker id: urutan di daftar presensi. Jangan gunakan NIM.
- [m|f] m untuk male dan f untuk female
- [n|nn] n untuk native dan nn untuk non-native bahasa Jawa

Contoh:

Speaker01_f_nn_utt01 **`Sugeng enjang, kulo badhe dhahar`**

Setiap mahasiswa mendapatkan jatah 30 baris untuk menuliskan kalimatnya.

2. Baca kalimat tersebut dan direkam.
    - Format file: WAV, PCM 16-bit, mono, sample rate 16 kHz (jika memungkinkan 16 kHz cukup; bila ada 44.1/48 kHz, ubah ke 16 kHz saat preprocessing).
    - Nama file: speaker<id>_[m|f]_[n|nn]_utt<xx>.wav
    - Cari lingkungan yang cukup sunyi (minim kebisingan latar).
    - Gunakan alat perekam yang sebisa mungkin memiliki kualitas yang baik.
    - Jarak mulut ke mikrofon ~10–20 cm.
    - Hindari clipping (Jangan mengatur level input terlalu tinggi sehingga sinyal terdengar terdistorsi).
    - Jika menggunakan smartphone, gunakan mode perekaman default dan sebutkan perangkat di metadata (tambahkan kolom di file csv)

- **Pengiriman:**

  Semua file audio dan `transcripts.csv` dikumpulkan ke satu repository kelas.

Langkah 2: Pembagian Data & Split

Speaker-disjoint split

- Alokasikan 20% speaker sebagai set uji, 10% native dan 10% non native (acak tapi reproducible dengan seed).
- Gunakan 70% data untuk training dan 10% untuk validation (per-utterance split).

Langkah 3: Eksplorasi Teknik Pemodelan

- Arsitektur: Setiap kelompok boleh memilih arsitektur berbeda (satu arsitektur dikerjakan oleh 2-3 kelompok). Contoh: Conformer, Transformer, RNN / BiLSTM, CTC-based, Seq2Seq with attention.
- Pelajari dan pahami makalah atau referensi terkait.

Langkah 4: Presentasi Eksplorasi Teknik Pemodelan

Tuliskan hasil pemahaman ke dalam slide yang akan dipresentasikan di kelas di minggu ke-13. Masing-masing kelompok 15 menit presentasi.

Langkah 5: Eksperimen Pembangunan Model Pengenal Ucapan

- Dua Pendekatan:
  1. Train from scratch: inisialisasi model dari awal dengan data yang tersedia.
  2. Fine-tune pretrained: gunakan model pretrained *jika tersedia* (mis. pretrained ASR/encoder dari sumber publik) lalu fine-tune pada data Jawa.
  Untuk setiap model catat hyperparameter, jumlah epoh, batch size, learning rate, dan waktu pelatihan.

- Lakukan optimasi model dengan menggunakan data validasi.
- Lakukan evaluasi dengan menggunakan data uji. Bandingkan kedua pendekatan menggunakan metrik WER dan PER/CER.

- Bisa menggunakan framework yang direkomendasikan: PyTorch, TensorFlow atau framework ASR spesifik seperti ESPnet, Kaldi, NeMo, dll.

Langkah 6: Lakukan evaluasi kualitatif

Analisis perbandingan (misal kenapa fine-tune lebih baik/lebih buruk, tipe error, kelebihan kekurangan).

Langkah 7: Buat dokumentasi eksperimen.

1. Proses pengumpulan data (cara pembuatan kalimat, instruksi perekaman, jumlah speaker, ringkasan metadata).
2. Struktur dataset (jumlah utterance, rata-rata durasi, distribusi gender/age, naive dan non native.
3. Preprocessing yang dilakukan (resampling, trimming, VAD, normalisasi).
4. Arsitektur model & justifikasi pemilihan hyperparameter.
5. Detail pretrained model yang dipakai (nama, sumber, lisensi).
6. Setup pelatihan (hardware, seed, dependency).
7. Hasil eksperimen (WER & PER table + grafik convergence).
8. Analisis Kualitatif
9. Kesimpulan & saran pengembangan lanjut.
10. Lampiran: link ke repository, perintah untuk reproduce (runbook), skrip evaluasi.

Langkah 8: Presentasi di kelas

Presentasi (di kelas) minggu ke-15 dan 16

- Durasi: 15 menit presentasi + 5 menit tanya jawab per kelompok.
- Isi slide: ringkas dataset & preprocessing, arsitektur, hasil utama (tabel WER/PER), demo singkat (opsional), kesimpulan.

---

10. Penilaian (rubric)

- Pengumpulan data & kepatuhan format : 15% (nilai individu)
- Pemahaman terhadap teknik yang digunakan : 20% (nilai kelompok)
- Implementasi model & eksperimen (dari-awal) : 20% (nilai kelompok)
- Implementasi model & eksperimen (fine-tune) : 20% (nilai kelompok)
- Pengujian (WER, PER) : 10% (nilai kelompok)
- Analisis Kualitatif + Saran : 10% (nilai kelompok)
- Presentasi : 5% (nilai individu)

Timeline

- Minggu 11 :Pembuatan data: Penyusunan kalimat & pengumpulan audio, pembersihan data, pembuatan transcripts.csv + upload repo
- Minggu 12 :Eksplorasi teknik pemodelan
- Minggu 13 :Presentasi hasil eksplorasi pemodelan. Eksperimen train from scratch (baseline).
- Minggu 14 : Fine-tuning pretrained models. Buat Dokumentasi dan slide.
- Minggu 15 dan 16 : Kumpulkan hasil dan Presentasi Hasil

Pengumpulan

- Slide hasil eksplorasi teknik : Minggu, 23 November 2025 Pukul 21.00 di Edunex.
- Dokumentasi Final : Minggu, 7 Desember 2025 Pukul 21.00 di Edunex. Kumpulkan hanya dokumentasi.

Selamat Mengerjakan