

Abstract

Scopul principal al lucrării este elaborarea unui sistem robust de recunoaștere a vorbirii în mediul cu zgomot pentru limba română, dependent de vorbitor, vocabular limitat și putere de calcul redusă. Acest obiectiv a necesitat analizarea, design-ul și evaluarea unor metode și algoritmi robusți pentru mediul cu zgomot. Sunt propuși algoritmi de detecție VAD și de normalizare cepstrală (CMN) – aceștia se bazează pe DTW și VQ. Detecția cât mai precisă a momentului de început al unui cuvânt (în prezența zgomotului ambiental) precum și a momentului de sfârșit de cuvânt permite reducerea la minim a volumului de date ce trebuie prelucrat ulterior.

A fost propus și un set de algoritmi DTW hibridi; setul de algoritmi este utilizat în faza de antrenare și combină avantajele DTW cu cele ale cuantizării vectoriale (VQ); la finalul aplicării fiecărui algoritm stocăm un singur model pentru un cuvânt, iar acel model utilizează clase (exact ca în metoda cuantizării vectoriale), fiecare clasă fiind reprezentată prin centroidul său. În faza de recunoaștere sunt comparați parametrii rostirii necunoscute cu centroizii modelului de referință. Algoritmii propuși sunt asemănători, combinând câte 2 modele ale aceleiași etichete, rezultând un model nou, sintetic; sunt aplicați iterativ până rămâne un singur model pentru o etichetă și sunt aplicați în etapa de învățare neintroducând timp suplimentar în etapa de recunoaștere. Diferențele dintre acești algoritmi apar la modul de calcul al parametrilor corespunzători unei celule de pe drumul optim. Setul de algoritmi este format din: algoritmul DTWVQ cu clase, cel cu lungime minimă sau maximă și algoritmul DTWVQ de mediere aritmetică.

Pentru scăderea timpului de comparație este descris un algoritm de compresie bazat pe matricea de adiacență (din teoria grafurilor). Pe baza acestui algoritm a fost propusă o metodă originală de recunoaștere (algoritmul rotor). Metoda produce un număr de clase pe baza cărora se definesc template-urile din dicționar. Este creată și o tabelă a tranzițiilor dintr-o clasă în alta. Tabela de tranziții se folosește alături de un set de numărătoare asociate template-urilor din dicționar. În faza de recunoaștere, pronunția necunoscută este împărțită în ferestre și parametrizată, apoi comparată cu fiecare centroid din dicționar rezultând apartenența la o anumită clasă. Folosind tabela de tranziții, clasa anterioară și cea curentă, se incrementează unul sau mai multe contoare asociate template-urilor. Odată cu ultima fereastră se alege valoarea maximă a numărătoarelor. Template-ul asociat acestui numărător devine alegerea algoritmului pentru pronunția necunoscută.

The main purpose of this work is the development of a robust ASR for Romanian language, of small dictionary, speaker dependent and using small computational systems. This goal needs the analysis, design and evaluation of robust methods and algorithms for noise environment. We proposed VAD and CMN algorithms – based on DTW and VQ. The precise detection of start and end of a word (in the presence of noise) minimizes the computing time and data.

We proposed also, a set of DTW hybrid algorithms used in the training phase, which combines the advantages of DTW with VQ; at the end of each algorithm we save only one template for each word; the template uses classes (like in VQ method), and each class is represented by his centroid. In the recognition phase we compare the unknown utterance parameters with template centroids. The proposed algorithms are very similar, combining 2 models of same label (word), generating one new, synthetic template; the algorithms run iteratively until only one template is obtained. These algorithms run in the training phase and they don't add time in the recognition phase. The difference between these algorithms is the computing function of parameters of the cells on the optimal path. These algorithms are: DTWVQ based on classes, minimum length, maximum length and average.

We describe a compression algorithm based on adjacency matrix (from graph theory) and an original method of recognition. The method produces a number of classes for defining templates from dictionary. A state transition table is created. Supplementary, we use a set of counters associated to each template from dictionary. In the recognition phase, the unknown utterance is divided into windows, then parameterised and compared with each centroid from dictionary. Using the transition table, the previous and current class, we increment one or more template counters. After the last window, we compute the maximum of the counters. The template associated with this counter is considered to be the recognised word.