BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Suplement No. 1, Vol. 6, No. 12, 2010

STRATEGIE POSZUKIWAŃ NAJLEPSZEJ ŚCIEŻKI PRZEZ GRAF REPREZENTUJĄCY HIPOTEZY SŁÓW

BARTOSZ ZIÓŁKO, DAWID SKURZOK

Department of Electronics, AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Kraków, Poland
bziolko@agh.edu.pl

Automatyczne rozpoznawanie mowy ciągłej jest nadal poważnym wyzwaniem, zwłaszcza dla języka polskiego, będącego językiem wysoce fleksyjnym. Rozpoznawanie izolowanych słów jest rozwiązaniem technicznym niewystarczającym dla większości zastosowań. Aby umożliwić bardziej skomplikowane zadania, na przykład dyktowanie, konieczne jest budowanie zdań z dużej liczby możliwych słów. Hipotezy słów są zwykle podobne akustycznie do siebie, co oznacza, że często mogą być różnymi odmianami tego samego słowa. Wybranie spośród nich poprawnej hipotezy zdania jest skomplikowanym zadaniem.

Istnieją dwie typowe struktury, które mogą być użyte do modelowania zdania z hipotez słów: lista najlepszych propozycji lub graf skierowany, będący siatką. Lista *n* najlepszych hipotez jest prostsza w realizacji, ale nie zapewnia takiej jakości jak graf, umożliwiający rozpatrzenie większej liczby kombinacji bardzo podobnych do siebie akustycznie słów. Dlatego w systemie rozpoznawania mowy Akademii Górniczo - Hutniczej zdecydowaliśmy się na użycie grafu.

Klasyfikator mowy szuka słów z nagrań różnych długości, w przybliżeniu równych czasowi wymawiania słowa. Wszystkie hipotezy są oceniane przez porównywanie ze słowami występującymi w słowniku, używając metryki edycyjnej. Wybierana jest najlepsza hipoteza dla danego słowa (pod kątem czasu trwania). Wówczas algorytm przechodzi do nagrań rozpoczynających się bezpośrednio po końcu wybranej hipotezy. Występuje zawsze kilka równoległych hipotez z różnymi słowami. Algorytm łączy je, jeżeli ich początki i końce sobie odpowiadają. W ten sposób powstaje siatka do dalszego użycia.

Typową strategią poszukiwania najlepszej ścieżki przez siatkę słów jest zastosowanie algorytmu Viterbiego. W przypadku naszego systemu, chcemy przede wszystkim zmniejszyć ilość krawędzi w grafie, poprzez rozcięcie połączeń między słowami niewystępującymi w statystykach 2-słów. Dysponujemy statystykami zebranymi z ponad 10 gigabajtów tekstu. Uważamy, że są wystarczająco reprezentatywne. W większości przypadków, można założyć, że jeśli nie ma odnotowanego połączenia dwóch słów, to nie mogą one po sobie występować w poprawnym zdaniu. Ta strategia umożliwi znaczne zmniejszenie stopnia komplikacji siatki. Pozwoli to przeprowadzać obliczenia w czasie rzeczywistym, nawet z zastosowaniem dużego słownika.

Słowa kluczowe: rozpoznawanie mowy, modelowanie języka, szukanie najlepszej ścieżki w grafie

Praca była finansowana z grantu MNiSW OR00001905.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Suplement No. 1, Vol. 6, No. 12, 2010

STRATEGIES OF WALKS THROUGH A WORD LATTICE IN AIM OF FINDING THE BEST SENTENCE HYPOTHESIS

BARTOSZ ZIÓŁKO, DAWID SKURZOK

Department of Electronics, AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Kraków, Poland
bziolko@agh.edu.pl

Automatic recognition of continuous speech is still a serious challenge, especially for Polish, as a highly inflective language. Recognition of isolated words is not enough for several applications. Sentences have to be build from a large number of word hypotheses for more complicated tasks like dictating. These hypotheses are typically acoustically similar, which means they can be often different inflective versions of a same word. Choosing a correct sentence hypothesis from them is complicated.

There are two typical structures which can be used to model a sentence from word hypotheses: n-best list and a lattice. N-best list is easier in implementation, while a lattice outperforms it thanks to ability of taking more similar combinations into account. This is why in the AGH speech recognition system, we decided to use a lattice.

Speech classifier searches for words for several speech recordings of different lengths, approximately equal to a time necessary to pronounce a word. All hypotheses are evaluated by comparing to words from a dictionary using Levenshtein distance. The strongest version of a particular word (according to its length) is chosen. Then the algorithm proceeds for the recording exactly following the end of the chosen hypothesis. There are always several, parallel hypothesis with different words. The algorithm connects them, if their beginnings and endings fit each other. Then a lattice is ready for further use.

A typical strategy to search for a best path through a word lattice is by applying Viterbi algorithm. In our case, we want first, to reduce the number of edges in the lattice by cutting off the connections between words which do not appear in 2-grams. These are word statistics, collected by us from over 10 GB of text. We think that they are representative enough and in the very most of cases it can be assumed that if there is no 2-gram, the two words cannot appear one after the other in a correct sentence. This strategy will allow us to reduce a lattice substantially, allowing to conduct calculation in real time, even with a large vocabulary.

Keywords: speech recognition, Language modeling, searching for a best path in a graph.

This work was supported by MNISW grant OR00001905.