

АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ МОВЛЕННЯ

Вінницький національний технічний університет

Анотація

Здійснено аналіз основних методів розпізнавання мовлення, зокрема метод розпізнавання мовних сигналів на основі ознак, приховані Марківські моделі та нейромережевий підхід до розпізнавання мови. Розглянуто категорії систем розпізнавання та підходи до розпізнавання мовного сигналу.

Ключові слова: *розпізнавання мовлення, мовний сигнал, DTW, прихована Марківська модель, нейромережа.*

Abstract

The analysis of basic methods of speech recognition, in particular, the method of recognition of speech signals on the basis of signs, the hidden Markov models and the neural network approach to speech recognition are analyzed. Categories of recognition and approaches to speech recognition are considered.

Keywords: *speech recognition, speech signals, DTW, Hidden Markov Model, neural network.*

Проблема розпізнавання мовлення на сьогоднішній день вважається серйозною і відіграє велику роль у спілкуванні людини з машиною. Звукова мова є важливим і природним засобом комунікації між людьми, а також однією із основних форм подання інформації в людино-машинних системах обробки інформації і управління.

Автоматизовані системи розпізнавання мови можуть бути класифіковані за багатьма ознаками: за типом мови, за множиною дикторів, за об'ємом та повнотою словника, що необхідно розпізнавати. За типами мову поділяють на дискретну і неперервну. Дискретною називають таку мову в якій паузи між словами значно більші за природні паузи всередині слів, наприклад, надиктування окремих команд. Неперервна мова в свою чергу не має значної паузи між словами. Природний людський режим спілкування є неперервною мовою. За ознакою множини дикторів системи розпізнавання мови поділяють на дикторозалежні, тобто такі, якість розпізнавання яких залежать від індивідуальних особливостей вимовляння диктора, та дикторонезалежні, у яких ймовірність вдалого розпізнавання слова однакова для усіх дикторів. За об'ємом словника системи розпізнавання мови класифікуються на дві категорії: системи з малими та великими словниками. Ці системи значно відрізняються одна від одної. Так, систему з малим словником можна навчити послідовно, вимовляючи кожне слово зі словнику. У системах розпізнавання мови, що виділяють слова, розпізнавання вимагає порівняння між вхідним словом і різними словами в словнику. Повнота словника полягає в тому, що кожне слово вхідного мовленнєвого сигналу має бути присутнім в словнику. Як правило повний словник мають лише системи розпізнавання мови з малим словником[1].

Сучасні системи розпізнавання мови будуються на основі принципів визнання форм розпізнавання. Методи й алгоритми можуть бути розділені на чотири більші класи: методи дискримінантного аналізу, засновані на Байєсовській дискримінації; приховані моделі Маркова; динамічне програмування – часові динамічні алгоритми (DTW); нейронні мережі.

Процес розпізнавання мови включає наступні етапи: запис і оцифрування мови, обчислення часових і спектральних характеристик зразків, класифікацію часових фреймів по фонетичних категоріях, підбір відповідних категорій слів, вимір ступеня довіри до ухваленого рішення.

Найпростіший підхід до розпізнавання – зіставлення вхідного образу з еталонами, що зберігаються в пам'яті. Розпізнаваний образ (невідомого класу) рівняється з еталоном кожного класу. Класифікація ґрунтується на заздалегідь обраному критерієві подоби, тобто якщо вхідний образ краще відповідає еталону і-го класу, ніж будь-якому іншому, то вхідний образ класифікується як приналежний і-ому класу. Для порівняння необхідно зробити виміри на вхідному образі. Подання мовних елементів через систему ознак складно. Розпізнавання слова починається з розкладання по тимчасових інтервалах фонем. БД еталонів фонем формується в процесі навчання. Потім, використовуючи фонемні параметри розрізнення, обчислюються ступені подоби з еталонами, що зберігаються в базі даних для

всіх фонем. По максимальному ступені подоби невідомих фонем еталонам виконується їхня оцінка або ідентифікація параметрів. Слідом за розпізнаванням фонем оцінюється окреме слово на основі баз даних. При розпізнаванні з використанням подібних фонемних одиниць досить підготувати базу даних з декількох десятків фонем основних видів. Подальше розпізнавання безлічі слів уже виконується досить легко.

Одним із основних методів розпізнавання мови є розпізнавання мовних сигналів на основі ознак. Такий метод припускає, що кожний об'єкт, що підлягає класифікації, можна охарактеризувати вектором ознак та безліччю значень ознак. Критерій відбору ознак заснований на важливості їх для характеристик образу або важливості їхнього впливу на якість розпізнавання. Систему ознак вибирають інтуїтивно. Рішення, що вимірювати, є суб'єктивним і залежить від конкретного розв'язуваного завдання – у системах розпізнавання мови використовуються вейвлет-коефіцієнти, спектральний опис, коефіцієнти лінійного проформування й інші параметри. Завдання розпізнавання формулюється як завдання статистичного ухвалення рішення [2].

Нейромережевий підхід до розпізнавання має певні переваги. В основі процесу виділення звуків, літер, побудови слів та речень лежать алгоритми класифікації та прогнозування. Саме нейронні мережі з кожним роком показують усе кращі результати в подібних задачах. Нейромережі вирішують завдання: генерації, навчання й тестування нейромереж, визначення значимості вхідних параметрів, контрастування нейромережі, пошук мінімального вирішального набору вхідних даних. Нейронна мережа з часовою затримкою у випадку розпізнавання обмеженої кількості слів дає кращі результати ніж метод прихованого Марківського моделювання.

Приховані Марківські моделі в даний час є одним з найефективніших підходів до побудови систем автоматичного розпізнавання мови. Алгоритм дозволяє розглядати мову як деякий код, а мовний потік – як послідовність деяких кодових пакетів. Неважливо, що є елементом цього коду – фонемі, склади або цілі слова, значення має лише те, що імовірність появи будь-якого елемента коду залежить від деякого числа попередніх елементів. Таким чином, мова породжується Марківським джерелом, а мовний код є випадковим. Для реалізації цього підходу застосовується так звана «прихована Марківська модель» (Hidden Markov Model), в якій дозволяються переходи тільки у наступний чи поточний стан [3].

Алгоритм DTW обчислює оптимальну послідовність трансформації (деформації) часу між двома часовими рядами. Алгоритм обчислює обидва значення деформації між двома рядами й відстанню між ними. Алгоритм починається з розрахунків локальних відхилень між елементами двох послідовностей, що використовують різні типи відхилень. Визначення слова може здійснюватися шляхом порівняння числових форм сигналів або шляхом порівняння спектрограми сигналів. Алгоритм застосовується для прямого порівняння сигналу, коли для кожної числової послідовності створюється нова послідовність, розміри якої значно менші за рахунок зменшення числових значень шляхом їх видалення між кутовими крапками. Це може призвести до зменшення точності розпізнавання, але через збільшення швидкості, точність, по суті, підвищується за рахунок збільшення слів у словнику [4].

На сьогоднішній день існує багато методів розпізнавання мови людини, але ні один метод не є ідеальним. Проте вони мають також свої переваги. Так, DTW алгоритм є дуже корисним для розпізнавання окремих слів в обмеженому словнику, хоч він і має певні слабкі сторони. Для розпізнавання швидкої мови часто використовують приховані моделі Маркова. Зазвичай спостерігається залежність, коли ймовірність розпізнавання покращується зі збільшенням розміру словника, але разом з тим і збільшується час розпізнавання.

Отже, проаналізувавши основні методи та підходи, для розпізнавання вимови під час вивчення іноземної мови було обрано нейромережевий підхід на основі рекурентних нейронних мереж (RNN). Такий підхід передбачає використання комірок довгої коротко-тривалої пам'яті у складі стандартних RNN. Основними перевагами такої моделі є висока швидкість роботи, більша точність розпізнавання, коректність роботи в умовах підвищеного шуму.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Плотников В. Н. Речевойдиалог в системах управления / В. Н. Плотников, В. А. Суханов, Ю. Н. Жигулевцев. — М. : Машиностроение, 1988. — 223 с. — ISBN 5-217-00148-8
2. Петров А. О. Моделі та методи розпізнавання мови [Текст] / А. О. Петров // Сучасний захист інформації. – 2012. – № 1. – С. 25-33.

3. M. Sazhok, V. Robeiko, D. Fedoryn. Distinctive features for Ukrainian real-time speech recognition system – УкрОбраз 2014.
4. Кузнецов В., Отт А. Автоматический синтез речи. – Таллинн: Валгус, 1989. – 135 с.

Миколюк Ірина Олександрівна – студентка групи ІКН-14б, факультет інформаційних технологій і комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: 2kn14b.mykoliuk@gmail.com;

Суприган Олена Іванівна – к. т. н., доцент кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця.

Iryna O. Mykoliuk – Student, ComputerScienceDepartment, VinnytsiaNationalTechnicalUniversity, Vinnytsia, e-mail: 2kn14b.mykoliuk@gmail.com;

Olena I. Suprygan – Associate Professor of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia.