

РОЗРОБКА І ДОСЛІДЖЕННЯ МОДУЛЯ ЕФЕКТИВНОГО РОЗПІЗНАВАННЯ ЗВУКІВ В СИСТЕМАХ ГОЛОСОВОГО УПРАВЛІННЯ

Вінницький національний технічний університет

Анотація

Розглядається задача розробки модуля ефективного розпізнавання голосових команд за допомогою методу динамічного програмування. Запропоновано новий підхід до визначення відстаней між марками звуків еталонного слова і слова, що розпізнається.

Ключові слова: динамічне програмування, розпізнавання образів, ефективність розпізнавання, форманта, помилка розпізнавання.

Abstract

The problem of development of the voice recognition by dynamic programming techniques is considered. New approach to defining of distances between the speech marks of two words is proposed.

Keywords: dynamic programming, pattern recognition, recognition efficiency, formant, recognition error.

Вступ

Застосування методів розпізнавання образів в системах управління дозволяє розширити їх можливості і розв'язати низку важливих задач, пов'язаних з діагностикою і автоматизацією прийняття рішень в різних сферах. Традиційні методи оброблення інформації в підсистемах розпізнавання образів передбачають їх перетворення в послідовність дискретних елементів, маркування з наступним розпізнаванням шляхом порівняння з еталонними послідовностями [1]. Рішення про розпізнавання приймається за мінімумом відстані між реалізацією і однією з еталонних стрічок. Недоліком таких методів є те, що для визначення вказаної відстані потрібно, окрім кодів марок, зберігати ще й коди відстаней між ними, що збільшує апаратні ресурси і зменшує швидкість розпізнавання. Тому актуальною є задача розробки методів оброблення і представлення інформації в підсистемах розпізнавання, які б дозволили усунути вказані недоліки. В запропонованому модулі розпізнавання звуків для усунення вказаних недоліків використовується метод динамічного програмування з опорою на реперні точки з представленням марок звуків кодами, що зберігають ранги відстаней.

Результати дослідження

Традиційна модель автоматичного розпізнавання мови припускає, що шляхом відстеження акустичних параметрів і застосування одного із засобів пошуку по набору еталонів фонематических сегментів можна встановити фонематичні ряди. Потім ці ряди можуть застосовуватися для виділення слів, фраз і сенсу висловів. При автоматичному розпізнаванні мови великі труднощі представляють собою процеси виявлення та ідентифікації деяких груп фонем. Тому для більш ефективного розпізнавання голосових команд запропоновано розподілити цей процес на декілька етапів.

Першим етапом є сегментація, тобто кожне слово чи група слів спочатку поділяється на менші одиниці мови — склади. Також дозволяє проводити розпізнавання або на складовому, або на фонемному рівні і одночасно зберігати в пам'яті параметри (тривалість, енергію тощо), які в подальшому будуть використовуватися. Оскільки основним параметром сигналу є дзвінкий звук, тобто рознос гармонік основного тону, а характеристики голосу з достатньою повнотою визначаються частотами формант. За ознаку виділення складів у мовному потоці вибрано кореляцію енергій в вибраних формантних діапазонах. Наступним і кінцевим етапом є розрахунок міри подібності між цілими словами, тобто розрахунок інтегральних мір подібності. Рішення про

розпізнавання приймається за мінімумом відстані між реалізацією і однією з еталонних стрічок [1]. В покращених алгоритмах для рядків однакової довжини обчислювальна складність має оцінку $O(n^2/\log n)$.

Для зменшення обчислювальної складності, а отже, підвищення швидкості розпізнавання, автори даній роботі запропонували метод співставлення з опорою на реперні (надійні) точки, в якості яких використовуються центри складів. Основна ідея цього методу полягає в розбитті матриці відстаней на сукупність прямокутних підматриць, розмір яких визначається розміщенням опорних точок в еталонній послідовності звуків слова і в послідовності, що розпізнається. Це дозволяє отримати значний вигравш у складності, яка в даному випадку буде оцінюватися, як $O(m+k)$, де m — кількість реперних точок, k — розмір підматриці.

Аналіз складової інформації здійснюється згідно алгоритму надійної сегментації неперервного мовного сигналу на склади, запропонованого авторами раніше, і виділення таких ознак складів, як їх кількість у слові, тривалість і часові інтервали центрів складів. Використання для фонетичного опису параметрів сигналу в центрах складів дозволяє мінімізувати контекстну варіацію і, таким чином, створити надійні опорні точки для процедури порівняння еталонних описів і описів фрагментів вхідної мови.

Іншим недоліком методу динамічного програмування є те, що для визначення вказаної відстані потрібно, окрім кодів марок, зберігати ще й коди відстаней між ними, що збільшує апаратні ресурси і зменшує швидкість розпізнавання. В роботі [2] для усунення цього недоліку був запропонований метод розпізнавання, що ґрунтується на представленні марок різницевиими кодами, які усувають необхідність запам'ятовування кодів відстаней.

В даній роботі авторами описується модуль розпізнавання, оснований на методі динамічного програмування з використанням представлення марок кодами, що зберігають ранги відстаней (DRP-codes) [3]. DRP — кодом є відображення $i \rightarrow B_i$

множини $M = \{1, 2, \dots, m\}$ в множину $\{0, 1\}^n$ двійкових послідовностей довжини n таке, що

$$\forall_{i,j} (R(d_{ij}) = r \Rightarrow R(h_{ij}) = r), \quad r = \overline{1, m_r}, \quad i, j \in M. \quad (1)$$

виразі (1) $R(d_{ij})$ — ранг відстаней d_{ij} між об'єктами i і j в просторі об'єктів; $R(h_{ij})$ — ранг відстані h_{ij} в просторі двійкових кодів; r — ціле число, конкретне значення рангу; m_r — максимальна величина рангу. Далі під кодованими об'єктами розумітимемо звукові образи.

Відстань між марками, закодованими DRP-кодом, під час їх порівняння визначається за допомогою порозрядної логічної операції “I”, однак при цьому можуть виникати помилки, зумовлені невиконанням аксіоми ідентичності під час визначення відстані між однаковими марками. Тому було запропоновано модифікувати дану логічну операцію таким способом, щоб для неї виконувалась аксіома ідентичності. Для цього було сформульовано математичний опис операції та розроблена відповідна логічна схема, яка скидає результат порівняння однакових марок до “0”. Модифікована операція була названа логічною операцією “I з самоблокуванням”.

Висновки

Використання в розробленому модулі розпізнавання звуків методу динамічного розпізнавання з опорою на реперні точки і їх маркування кодами, що зберігають ранги відстаней, дозволило підвищити надійність і швидкість розпізнавання та розширити склад словника мовних образів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Методы автоматического распознавания речи: в 2-х книгах. Пер. с англ. / Под ред. У. Ли. — М.: Мир, 1983. — Кн. 1. 328 с.
2. Glave R.D., Vander Giet. The David speech recognition system. - Proc. IEEE Int. Conf. ASSP.-Tulsa, 1978, p.429-432.
3. Биков М.М. Універсальний метод представлення інформації в інтелектуальних еволюційних системах / Биков М.М. // Відбір і обробка інформації. - 2006.- Вип. 24(100). С. 35-42.

Анна Сергіївна Алексєєнко — студентка групи 2АКІТ-16м (іт), факультет комп'ютерних систем та автоматички, Вінницький національний технічний університет, м. Вінниця, e-mail: alekseenko1719@gmail.com;

Микола Максимович Биков — професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця, e-mail: nkbykov@vntu.edu.ua.

Anna S. Aliksieienko — student of Computer System and Automation Department, Vinnytsia National Technical University, Vinnytsia, e-mail: alekseenko1719@gmail.com;

Mykola M. Bykov — professor of Computer Control System Department, Vinnytsia National Technical University, Vinnytsia.