

РОЗПІЗНАВАННЯ ГОЛОСОВИХ КОМАНД МЕТОДОМ ДИНАМІЧНОГО ПРОГРАМУВАННЯ З ВИКОРИСТАННЯМ РАНГОВИХ КОДІВ

Вінницький національний технічний університет

Анотація

Розглядається задача розпізнавання голосових команд за допомогою методу динамічного програмування. Запропоновано новий підхід до визначення відстаней між марками звуків еталонного слова і слова, що розпізнається, заснований на рангових кодах (кодах, що зберігають ранги відстаней) і з опорю на реперні точки.

Ключові слова: динамічне програмування, розпізнавання образів, ефективність розпізнавання, ознаковий простір, рангові коди, помилка розпізнавання.

Abstract

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

Keywords: dynamic programming, pattern recognition, recognition efficiency, feature space, rank codes, 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. Автори в даній роботі запропонували використовувати в динамічному програмуванні опорні точки, які відповідають положенню центрів складів у часовій послідовності звуків слова

Результати дослідження

Мовні сигнали, які представляються у вигляді набору команд — слів, подаються на вхід пристрою розпізнавання, в пам'яті якого зберігаються еталони всіх слів, що розпізнаються.

В цьому випадку розрахунок міри подібності між невідомим словом і одним із еталонів складається з двох етапів. Першим етапом є розрахунок міри подібності між мілкими елементами мови, тобто розрахунок локальних мір подібності. А наступним — 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 в даній роботі запропонували метод співставлення з опорю на реперні (надійні) точки, в якості яких використовуються центри складів. Основна ідея цього методу полягає в розбитті матриці відстаней на сукупність прямокутних під матриць, 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 під матриці.

Розроблена система пошуку використовує акустичний, силабічний (складовий) та фонетичний аналіз для генерування фонетичного опису еталонів ключових слів та фрагментів вхідної мови. 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 різницеvими кодами, які усувають необхідність запам'ятовування кодів відстаней. Даному методу притаманні недоліки, зумовлені можливістю виникнення помилок через властивості запропонованих кодів, а також обмеженням швидкодії через необхідність обчислення відстаней Хемінга між послідовностями.

В даній роботі авторами пропонується метод розпізнавання, що ґрунтується на представленні марок кодами, що зберігають ранги відстаней (DRP-codes) [3]. DRP — кодом є відображення $i \rightarrow B_i$ множини $M = \{1, 2, \dots, m\}$ в множини $\{0, 1\}^n$ двійкових послідовностей довжини n таке, що

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

У 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 “Г”, 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 “Г з самоблокуванням”.

Висновки

Запропонований метод дозволив 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 ІАВ-126, факультет комп'ютерних систем та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: alekseenko1719@gmail.com;

Науковий керівник: *Микола Максимович Биков* — професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця.

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

Supervisor: *Mykola M. Bykov* — Professor of Computer Control System, Vinnytsia National Technical University, Vinnytsia.