

## **ДВОРІВНЕВИЙ АНАЛІЗ МОВНОЇ КОМАНДИ**

*Соколов Д. Ю., студент; Дюжаєв Л. П., к.т.н., доцент  
Національний технічний університет України  
«Київський політехнічний інститут», м. Київ, Україна*

Одним з перспективних способів організації людино-машинної взаємодії є передача комп'ютерній системі інструкцій користувача у форматі мовних команд.

В якості методу розпізнавання більшість сучасних систем використовують алгоритми розпізнавання ключового слова [1, 2], побудовані на основі прихованих марківських моделей [3, 4]. Типовий метод розпізнавання голосової команди полягає в застосуванні алгоритму розпізнавання ключового слова до всього мовного сигналу для кожної можливої команди зі словника команд. Такий підхід має два істотні недоліки: велика обчислювальна складність; команди можуть включати слова, які погано розпізнаються за допомогою алгоритму розпізнавання ключового слова.

Для вирішення проблем, описаних вище, запропоновано наступний підхід: додавання в команду заздалегідь визначеного ключового слова та застосування алгоритму розпізнавання ключового слова тільки для цього слова, а для решти команди використання стандартного алгоритму розпізнавання злитого мовлення.

Недоліком запропонованого підходу розпізнавання команди є некоректні спрацьовування системи, якщо після ключового слова йдуть команди або фрази не зі словника команд. Це вносить наступне обмеження на використання системи: користувач повинен знати набір підтримуваних системою команд і використовувати лише їх.

Архітектурною відмітною особливістю підходу є об'єднання двох різних алгоритмів розпізнавання мови в єдиний блок.

Одним з популярних алгоритмів розпізнавання ключового слова є метод ковзного вікна. Аналіз методу показує, що метод містить істотний недолік - велику обчислювальну складність, що створює незручності і заважає застосуванню системи на практиці.

Для усунення даного недоліку пропонується алгоритм визначення ключового слова, який використовує ті ж моделі, що і алгоритм розпізнавання команди, але не враховує ймовірності переходу. В кожний дискретний момент ( $t$ ) цього алгоритму відбувається наступне:

1. Стартує новий шлях з поточної позиції сигналу ( $y_t$ ) в початковий стан ключового слова (в початковий стан прихованої марківської моделі першої фонемі ключового слова,  $s_1$ ).

2. Кожен існуючий шлях дублюється: один залишається в поточному стані (перехід  $s_i \rightarrow s_i$ ), а другий переходить в наступний стан ключового слова ( $s_i \rightarrow s_{i+1}$ ).

3. Відбувається скорочення шляхів на основі функції правдоподібності.

Як показник відповідності розроблений метод використовує не тільки самі ймовірності спостереження  $b_j(y_t)$ , але і відношення ймовірності найбільш відповідного стану з усього набору моделей і поточного стану ключового слова  $b_j(y_t) / b_{best}(y_t)$ . Функція ж правдоподібності заснована на показниках гіршої моделі в ключовому слові, а не на загальних показниках всього слова.

Експериментальні дані (табл.1.), показали, що такий підхід зменшує помилку, пов'язану з неправильним спрацьовуванням на слові, близькому за звучанням, а також підтверджують ефективність такого підходу.

Відмінною рисою запропонованого дворівневого методу розпізнавання голосової команди є поділ мовної команди на дві частини: ключового слова і безпосередньо

Таблиця 1

| Критерії \ Методи   | Метод ковзного вікна | Запропонований метод |
|---|----------------------|----------------------|
| Час роботи алгоритму, для вхідної ділянки тривалістю 7с         | 20 с                 | 1 с                  |
| Відсоток спрацьовувань алгоритму на ключових/не ключових словах | 86,1% / 5%           | 96% / 1%             |

команди. Експериментальні дані підтверджують ефективність такого підходу: спрацьовування на ключовому слові - 96%, не на ключовому слові - 1%. Запропонований алгоритм розпізнавання ключового слова і проведене експериментальне порівняння цього алгоритму з методом ковзного вікна показують перевагу запропонованого методу як в плані швидкості, так і в плані якості.

### Література

1. Сорокин В. Н. Сегментация и распознавание гласных / В. Н. Сорокин, А. И. Цыплихин // Информационные процессы. — 2004. — №2. — С. 202 — 220. — Бібліогр. : с. 220.
2. Hazen T. Recognition confidence scoring and its use in speech understanding systems / T. Hazen // Computer Speech and Language. — 2002. — №1. — С. 49 — 67. — Бібліогр. : с. 67.
3. Rosti I. Linear gaussian models for speech recognition / I. Rosti; University of Cambridge. — Cambridge, 2004. — 142с. : обкл., іл., табл. — Бібліогр. : С. 133 — 142.
4. Couvreur Chr. Hidden Markov Models and Their Mixtures / Chr. Couvreur; Department of Mathematics, Catholic University of Louvain. — Louvain, 1996. — 147с. обкл., іл., табл. — Бібліогр. : С. 131 — 147.