

# ОБГРУНТУВАННЯ ВИБОРУ МЕТОДУ РОЗПІЗНАВАННЯ ІНФОРМАЦІЇ В СИСТЕМАХ ГОЛОСОВОГО УПРАВЛІННЯ

Вінницький національний технічний університет

## **Анотація**

*В даній роботі проаналізовано методи обробки та розпізнавання мовних сигналів, що можуть бути використаними у системах голосового управління, проведена порівняльна характеристика та обґрунтовано вибір статистичного методу.*

**Ключові слова:** голосове управління, мовний сигнал, марковська модель.

## **Abstract**

*This paper analyzes the methods of processing and recognition of speech signals that can be used in voice control systems, comparative characteristics and substantiates the choice of statistical method*

**Keywords:** voice control, speech signal, Markov model.

Побудова системи голосового управління є в даний час актуальною, оскільки завданням таких систем є виділення і розпізнавання з потоку звукового сигналу (як мовного, так і не мовного) заздалегідь певного набору мовних команд, що може широко застосовуватись в цифрових пристроях. Прикладом такої команди може служити фраза "Включити світло". При цьому система не повинна реагувати на інші ділянки звукового сигналу, включаючи і ті, які містять окремі слова зумовлених команд. Такі системи здатні істотно полегшити взаємодію користувача з комп'ютерною системою. Особливо ця ідея розвинена в концепції так званих «Розумних будинків», сучасних автомобілів та мобільних пристроїв. Більш того, іноді голосовий інтерфейс є необхідним компонентом для людей з різними вадами та допомагають їм краще комунікувати в суспільстві [1].

Більшість сучасних систем автоматизованого розпізнавання команд голосового управління використовують модульну архітектуру з використанням блоку підсилення мовлення (speech enhancement), детектора голосу (VAD), перетворювача сигналу в вектори особливостей (Frontend) і головного модуля (searchengine), що взаємодіють за таким алгоритмом розпізнавання ключового слова:

- 1) цифровий сигнал надходить в модуль підсилення мовлення [2], де підвищується якість сигналу внаслідок видалення шумів і внесеного каналом спотворення;
- 2) детектор голосу виділяє ділянки сигналу, що містять мову;
- 3) мовні ділянки за допомогою модуля перетворення сигналу в вектори особливостей перетворюються у вектори коефіцієнтів, які надходять в головний модуль;
- 4) в головному модулі відбувається безпосереднє визначення наявності і розпізнавання команди.

Таким чином, на виході головного модуля сформується інформація щодо наявності команди.

Серед інших, найбільшої популярності набули такі методи як статистичний та метод ковзного вікна

Статистичний метод [1] передбачає побудову таблиці ймовірностей використання слів в кожному граматичному значенні. Це завдання можна вирішити на основі тестових текстів, проаналізованих вручну. В англійській мові завдання на цьому етапі називається Part-Of-Speech tagging [1]. Розглянемо пропозицію англійською мовою: «The can will rust». Отже, the - визначення артикль або частка «тем»; can - може одночасно бути і модальним дієсловом, і іменником, і дієсловом; will - модальне дієслово, іменник і дієслово; rust - іменник або дієслово. Слово «can» в більшості випадків використовується як дієслово, але іноді воно може бути і іменником. З огляду на цей недолік, була створена модель, яка бере до уваги той факт, що після артикля піде прикметник або іменник [3]:

$$\arg \max \prod_{t=1}^n p(w_t | t_t) p(t_t | t_{t-1}),$$

- де  $t$  – tag (іменник, прикметник і т.д.);
- $w$  - слово в тексті (rust, can ...);
- $p(w | t)$  - ймовірність того, що слово  $w$  відповідає tag;
- $p(t_1 | t_2)$  - ймовірність того, що  $t_1$  йде після  $t_2$ .

З запропонованої формули видно, що потрібно підібрати теги, щоб слово підходило tag, і tag підходив попередньому tag [4]. Даний метод дозволяє визначити, що «can» виступає в ролі іменника, а не як модального дієслова. Ця статистична модель може бути описана як ергодична прихована марковська модель (ПММ)[3], що представлена на рисунку 1. Саме таку модель доцільно використовувати для ефективного розпізнавання мовлення, зокрема для кращої ідентифікації структури речення, як описано в прикладі вище.

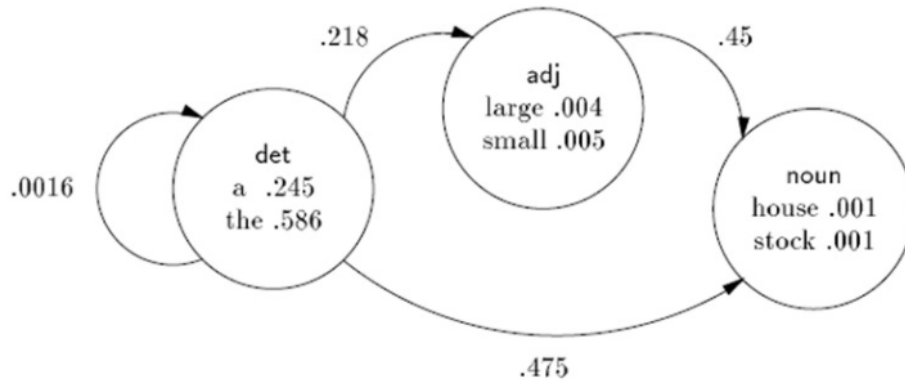


Рисунок 1 – Ергодична марковська модель

Метод ковзного вікна – алгоритм трансформації, що дозволяє сформувати з членів тимчасового ряду набір даних, який може служити навчальною виборкою для побудови моделі прогнозування [5]. Під вікном у даному випадку розуміється часовий інтервал, що містить набір значень, які використовуються для формування навчального прикладу. У процесі роботи алгоритму вікно зміщується в часі на одиницю спостереження, а кожне положення вікна утворює один приклад. Суть методу ковзного вікна при розпізнаванні мовлення полягає у визначенні входження ключового слова за допомогою алгоритму Вітербо [4], який широко застосовується для розпізнавання мовлення.

Цей алгоритм вирішує таку задачу: дано вектор спостережень ( $o$ ), потрібно визначити найбільш підходящу послідовність ПММ ( $s$ ) і переходів між їх станами для цього вектора спостережень (рис. 2). Так, на рис. 2 зображені всі можливі шляхи для даної ділянки сигналу і певної послідовності ПММ. Часто для оцінки шляху використовується значення ймовірності, отримане за допомогою алгоритму Вітербо.

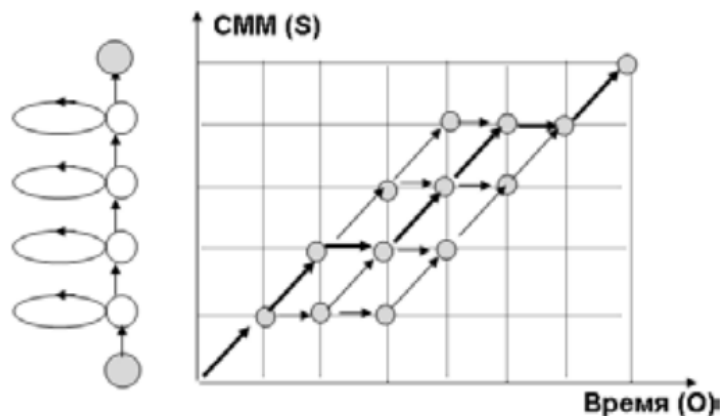


Рисунок 2 - Приклад роботи алгоритму Вітербо

Алгоритм Вітербо дозволяє зробити найбільш ймовірне припущення про послідовність станів прихованої марковської моделі на основі послідовності спостережень.

Порівняльна характеристика обох описаних методів наведена в таблиці 1.

Таблиця 1 – Порівняльна характеристика методів

Метод	Переваги	Недоліки	Основні вимоги до методу
Статистичний метод	Мала ймовірність появи помилок	Велика обчислювальна складність, мала швидкодія, необхідність додаткового моделювання	Необхідність у великих обчислювальних ресурсах
Метод ковзного вікна	Швидкодія	Велика обчислювальна складність, необхідність додаткового моделювання, велика ймовірність появи помилок (розпізнавання ключових слів як загальних)	Необхідність у великих обчислювальних ресурсах, необхідна збалансована мовна база, якісний підбір ключового слова

Таким чином, дані методи розпізнавання мовних сигналів, що можуть бути використаними у системах голосового управління, мають такі спільні недоліки, як велика обчислювальна складність та необхідність додаткового моделювання. Проте, статистичний метод є кращим варіантом для оптимізації та усунення означених недоліків, зокрема, через імплементацію алгоритмів розпізнавання ключового слова. Оптимізація даного методу покращить швидкодію та зменшить обчислювальну складність.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Т.К. Винцюк. Анализ, распознавание и смысловая интерпретация речевых сигналов. — Киев. Наукова думка, 1987.
2. Методы автоматического распознавания речи: В 2-х книгах. Пер. с англ./Под ред. У. Ли. — М.: Мир, 1983. — Кн. 1. 328 с., ил.
3. Синтез и распознавание речи. Современные решения: А.В. Фролов, Г.В. Фролов.
4. G. David Forney, Jr. The Viterbi Algorithm: A Personal History /MIT
5. Armagnac, Alden P. "Tell It to Sceptron!" // Popular Science. — April 1963. — Vol. 182 — No. 4 — P. 120.
6. Davies, K.H., Biddulph, R. and Balashek, S. (1952) Automatic Speech Recognition of Spoken Digits, J. Acoust. Soc. Am. 24 (6) pp. 637—642

**Савчук Тамара Олександрівна** — к.т.н., професор кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця, Хмельницьке шосе, 95.

**Пупко Олександр Валерійович** — аспірант кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, Хмельницьке шосе, 95, e-mail: [salexpk@gmail.com](mailto:salexpk@gmail.com).

**Tamara O. Savchuk** — Cand. Sc. (Eng), Assistant Professor, Professor of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia, Khmelnytske Shose, 95.

**Oleksandr V. Pupko** — Postgraduate Student of Computer Science Department, Vinnytsia National Technical University, Vinnytsia, Khmelnytske Shose, 95, e-mail: [salexpk@gmail.com](mailto:salexpk@gmail.com).