

ВІННИЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Грищук Тетяна Вікторівна

УДК 681.3.06

**МОДЕЛЮВАННЯ ПРОЦЕСУ АНАЛІЗУ І КЛАСИФІКАЦІЇ
ГОЛОСОВИХ КОМАНД**

01.05.02 – Математичне моделювання
та обчислювальні методи

Автореферат

дисертації на здобуття наукового ступеня
кандидата технічних наук

Вінниця – 2006

Дисертацією є рукопис.

Робота виконана у Вінницькому національному технічному університеті Міністерства освіти і науки України.

Науковий керівник: кандидат технічних наук, доцент

Биков Микола Максимович,
Вінницький національний технічний університет,
професор кафедри комп'ютерних систем управління

Офіційні опоненти:

доктор технічних наук, професор
Карпов Олег Миколайович,
Дніпропетровський національний університет,
професор кафедри математичного забезпечення ЕОМ

доктор технічних наук, професор
Кветний Роман Наумович,
Вінницький національний технічний університет,
завідувач кафедри автоматики та
інформаційно-вимірювальної техніки

Провідна установа: Харківський національний
університет радіоелектроніки,
кафедра системотехніки,
Міністерство освіти і науки України, м. Харків

Захист відбудеться "17 березня 2006 р. о 9³⁰ годині на засіданні спеціалізованої вченої ради Д 05.052.01 у Вінницькому національному технічному університеті за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95.

З дисертацією можна ознайомитись у бібліотеці Вінницького національного технічного університету за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95.

Автореферат розісланий "14 лютого 2006 р.

Вчений секретар
спеціалізованої вченої ради

Захарченко С.М.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Одним із факторів, що визначають науково-технічний прогрес України, є ступінь розвитку автоматизованих систем управління і телекомунікації. Як показує досвід їх використання, найбільш життєздатними є людиномашинні системи, в яких висока швидкість і точність обробки інформації поєднуються з перевагами людського мислення на етапі прийняття рішення. Ефективність таких систем суттєво залежить від спілкування людини з машиною.

Найбільш простим і природним для людини засобом обміну інформацією та подання команд служить мова. Мовний сигнал має беззаперечні переваги, які роблять його ефективним засобом передачі інформації в різноманітних областях. Мова є достатньо швидким засобом спілкування, вона легко передається по каналах зв'язку, залишаючи при цьому свободу рухів очам і рукам. Проте автоматичне розпізнавання і розуміння мови є дуже складним, і впровадження даного способу обміну інформацією між людиною і машиною обмежується на сьогоднішній день недостатнім вивченням процесу аналізу і класифікації голосових команд, що проявляється в недостатній якості існуючих на сьогодні математичних моделей даного інформаційного процесу. Тому проблема розробки адекватних і ефективних моделей процесу аналізу і класифікації голосових команд є на сьогодні достатньо актуальною.

Зв'язок роботи з науковими програмами, планами, темами. Основний зміст роботи складають результати наукових розробок та експериментальних досліджень, які проводилися протягом 2001-2005 років на кафедрі комп'ютерних систем управління Вінницького національного технічного університету. Протягом 2001 і 2002 років наукові розробки проводилися в рамках написання магістерської дисертації, яка була захищена восени 2002 р. Протягом 2004-2005 років наукова робота проводилася відповідно до науково-дослідної держбюджетної теми "Розробка теорії, методів, моделей та алгоритмів для оцінки ефективності та оптимізації систем прийняття рішень" (НДР № держ. реєстрації 0104UO00741).

Мета і задачі дослідження. Метою дослідження є покращення характеристик процесу аналізу і класифікації голосових команд, що полягає в збільшенні швидкодії та достовірності процесу класифікації за рахунок розробки і використання його адекватних моделей.

Для досягнення вказаної мети в роботі розв'язуються такі основні задачі:

- аналіз відомих підходів до моделювання процесу аналізу і класифікації голосових команд;
- розробка математичної моделі оптимізації процесу класифікації голосових команд;
- розробка математичної моделі мовного сигналу на рівні попередньої обробки;
- розробка математичної моделі процесу аналізу і класифікації голосових команд на рівні прийняття рішення;
- розробка методів класифікації голосових команд, що базуються на розроблених математичних моделях процесу аналізу і класифікації голосових команд;
- розробка методики проектування та оптимізації систем розпізнавання голосових команд;
- розробка алгоритмів та програмного забезпечення для моделювання процесу аналізу і класифікації голосових команд;

– дослідження адекватності розроблених математичних моделей та обчислювальної ефективності запропонованих алгоритмів.

Об'єкт дослідження – процес аналізу і класифікації голосових команд.

Предмет дослідження – математичні моделі процесу аналізу і класифікації голосових команд.

В дисертаційній роботі використані такі *методи дослідження*:

– положення і методи теорії графів, теорії оцінювання, теорії розпізнавання, теорії оптимізації, обчислювальні методи для розробки математичної моделі оптимізації процесу класифікації голосових команд;

– методи теорії обробки сигналів, теорії спектрального аналізу для розробки математичної моделі мовного сигналу на рівні параметричної обробки;

– методи теорії оцінювання, теорії ймовірностей і теорії прийняття рішень для розробки математичної моделі синтаксичного аналізу і класифікації голосових команд на базі прихованих марковських мереж;

– методологія імітаційного моделювання на ЕОМ для дослідження адекватності розроблених математичних моделей та обчислювальної ефективності запропонованих алгоритмів.

Наукова новизна одержаних результатів. В ході розв'язання поставлених задач були отримані нові наукові результати:

1. Вперше розроблено математичну модель оптимізації процесу класифікації голосових команд, яка, на відміну від існуючих, враховує ієрархічну будову мовного сигналу та дозволяє оптимізувати параметри модульних систем розпізнавання голосових команд для досягнення більшої швидкодії та достовірності роботи таких систем.

2. Вперше запропоновано математичну модель мовного сигналу на рівні параметричної обробки, в якій процес мовоутворення базується на принципі “квазічастотної” модуляції голосового тракту. На відміну від існуючих, модель дозволяє підвищити дикторонезалежність ознак, що виділяються, та збільшити швидкість і достовірність процесу класифікації голосових команд.

3. Вперше розроблено математичну модель синтаксичного аналізу і класифікації голосових команд, яка дозволяє врахувати інформацію про синтаксичну структуру голосових команд і представити її у вигляді орієнтованого графа. На відміну від відомих моделей, запропонована модель дає можливість збільшити швидкодію процесу аналізу і класифікації голосових команд на синтаксичному рівні без втрат в достовірності класифікації.

4. Вперше розроблено метод класифікації голосових команд, що використовує математичну модель синтаксичного аналізу і класифікації на базі прихованих марковських мереж, в якому, на відміну від існуючих, враховується інформація про тривалість звучання окремих фонем дискурсу, що дозволяє зменшити час класифікації.

Практичне значення одержаних результатів дисертаційної роботи полягає в наступному:

1. У результаті математичного моделювання отримані аналітичні залежності достовірності класифікації звуків української мови від впливу шуму навколишнього середовища.

2. Розроблено алгоритми класифікації голосових команд, які ґрунтуються на використанні розроблених методів та моделей аналізу і класифікації голосових команд.

3. Розроблено модифікований спосіб запису контекстно-вільних граматик та процедуру граматичного виводу фраз граматики.

4. Отримано аналітичний вираз функції динамічного скорочення варіантів перебору в процесі класифікації голосових команд на базі прихованих марковських мереж.

5. Розроблено алгоритми сегментації та маркування мовного сигналу на звукотипи, що використовують “квазічастотну” модель мовоутворення.

6. Розроблено спосіб кодування мовних образів ранговими кодами та математичну модель нейронної мережі, призначеної для класифікації, на основі рангових структур, які враховують інформацію про ранги відстаней між елементами класифікації і дозволяють підвищити швидкість і достовірність процесу класифікації.

7. Розроблено методику проектування та оптимізації систем розпізнавання голосових команд.

8. На базі розроблених моделей, методів і алгоритмів процесу аналізу і класифікації голосових команд були створені програмні середовища для розробки контекстно-вільних граматик та систем розпізнавання голосових команд.

Розроблені в дисертації моделі, методи і алгоритми опису, навчання і класифікації голосових команд можуть бути використані при розробці систем розпізнавання мови для підвищення їх ефективності.

Результати дисертаційних досліджень впроваджені на підприємстві “Комп’ютерні Мультимедіа Системи” (м. Вінниця) та у навчальний процес кафедри комп’ютерних систем управління Вінницького національного технічного університету. Впровадження результатів дисертаційних досліджень підтверджені відповідними актами.

Особистий внесок здобувача. Всі результати, які складають основний зміст дисертації, отримані здобувачем самостійно. В роботах, опублікованих у співавторстві, здобувачу належить розробка ієрархічної стратегії розпізнавання мови та модифікація критерію оцінки ефективності для задачі класифікації голосових команд [1, 3, 12], модифікація моделі потенціальної системи розпізнавання та дослідження якості класифікації від рівня шуму навколишнього середовища [8], розробка “квазічастотної” моделі мовоутворення та алгоритмів маркування і сегментації мовного сигналу [7, 10], розробка пофонемного методу класифікації голосових команд [2], розробка методики проектування модульних систем розпізнавання мови з використанням нейромережових класифікаторів [4, 9], розробка структурної схеми пристрою для розпізнавання мовних образів ранговими кодами [11].

Апробація результатів дисертації. Основні положення і результати дисертаційної роботи доповідалися і обговорювалися на 11 науково-технічних конференціях: на Міжнародній конференції з індуктивного моделювання MKIM-2002 (м. Львів, 2002 р.); на 1-му Міжнародному радіоелектронному форумі “Прикладна радіоелектроніка. Стан та перспективи розвитку” (м. Харків, 2002 р.); на третій міжнародній конференції “Інтернет. Освіта. Наука” (м. Вінниця, 2002 р.); на VII та VIII Міжнародних науково-технічних конференціях “Контроль і управління в складних системах” (м. Вінниця, 2003, 2005 рр.); на Міжнародних конференціях “Мікропроцесорні пристрої та системи в автоматизації виробничих процесів” (м. Хмельницький, 2002-2003 рр.); на Четвертому російсько-українському науковому семінарі “Інтелектуальний аналіз інформації IAI-2004” (м. Київ, 2004); на щорічних науково-технічних конференціях професорсько-викладацького складу, співробітників і студентів ВНТУ (м. Вінниця, 2002-2005 рр.).

Публікації. За темою дисертації опубліковано 12 праць, в тому числі 6 статей надруковано у фахових виданнях, затверджених ВАК України, 4 роботи

опубліковано в збірниках матеріалів конференцій, а також надруковано 1 тези доповідей, отримано 1 деклараційний патент України на винахід.

Структура та обсяг дисертації. Дисертація складається з вступу, чотирьох розділів, висновків, списку використаних джерел, додатків. Повний обсяг дисертації складає 185 сторінок, з яких основна частина складає 149 сторінок. Дисертація містить 50 рисунків, 11 таблиць, 6 додатків, список використаних джерел із 157 найменувань.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** до дисертаційної роботи обґрунтовано актуальність теми, зазначено зв'язок роботи з науковими програмами, планами, темами, сформульовано мету та задачі досліджень. Також наведено характеристики об'єкта і предмету досліджень, викладено основні наукові та практичні результати, наведено відомості про впровадження результатів роботи, їх апробацію та публікацію.

У **першому розділі** на основі аналізу літературних джерел та останніх публікацій за напрямом дисертаційних досліджень вводяться основні вихідні положення, обґрунтовується вибір предмету досліджень і формулюються задачі досліджень.

До вихідних положень відносяться наступні. Всі математичні моделі процесу аналізу і класифікації голосових сигналів можна розділити на такі основні групи: математичні моделі мовного сигналу на рівні попередньої обробки, математичні моделі процесу прийняття рішення, математичні моделі синтаксичного аналізу. Розглянуто основні принципи моделювання процесу аналізу і класифікації голосових команд та роль математичних моделей різних рівнів ієрархії опису даного процесу. Наявність чітко виражених структурних елементів, а також складність систем розпізнавання мови (СРМ) стали основною причиною того, що процес аналізу і класифікації голосових команд описується на трьох незалежних рівнях ієрархії: параметричний рівень, на якому здійснюється виділення акустичних ознак мовного сигналу для опису голосових команд); лексичний рівень - для заданої послідовності слів визначається ймовірність обсервації відповідної тимчасової послідовності акустичних ознак; синтаксичний рівень - виконується пошук послідовності слів з найвищою апостеріорною ймовірністю для даного мовного сигналу.

В результаті виконаного аналізу існуючих на сьогодні підходів до моделювання процесу аналізу і класифікації голосових команд були виявлені головні проблеми і недоліки, що закладені в самі моделі класифікації, а також проблеми, що виникають в процесі роботи систем розпізнавання мови. Зокрема, було виявлено, що на сьогодні актуальними є задачі розробки СРМ для голосового керування персональними та портативними комп'ютерами, приладами та автоматами. Однак через складну ієрархічну будову мовного сигналу на сьогодні існують математичні моделі, що дають можливість описати процес аналізу і класифікації голосових команд на окремих рівнях, не враховуючи взаємозв'язку між усіма рівнями ієрархії. Через це ускладнюється задача оптимізації процесу класифікації в цілому. До того ж, існуючі моделі мовного сигналу не забезпечують достатню швидкодю і достовірність класифікації, тому їх використання не є ефективним.

Викладене вище і визначило зміст наукових досліджень, вказаний в задачах досліджень.

Другий розділ присвячено розробці математичних моделей процесу аналізу і класифікації голосових команд.

Відповідно до мети дисертаційної роботи сформульовано поняття ефективності та запропоновано критерій оцінки ефективності систем розпізнавання мови, а також формалізовано задачу оптимізації процесу класифікації голосових команд.

В роботі за загальносистемний критерій оптимізації використовується узагальнений функціонально-статистичний критерій, модифікований належним чином до системи мовного спілкування шляхом належного вибору потенційної і реальної CPM [3]:

(1)

де i - функціонально-статистичні критерії для реальної і потенційної CPM відповідно; δ - задана точність класифікації.

Задача побудови моделі потенційної CPM може бути описана наступним чином [8]:

Дано:

Інформаційний канал передачі мовного повідомлення з характеристиками:

- словник голосових команд, що класифікуються;
- імовірності подібності слів словника W ;
- рівень шуму навколишнього середовища (дБ).

Потрібно:

Визначити кількість інформації, яку отримує потенціальна CPM.

Центральним поняттям в цій моделі є інформація, яка спостерігається на виході дискретного чи неперервного випадкового процесу. В дисертації подаються формули для розрахунку ентропії мовного повідомлення до та після його передавання з урахуванням синтаксичних конструкцій дискурсу, що моделюється. За модель потенційної системи розпізнавання може слугувати матриця $\|d_{ij}^{m_\alpha}\|$ відстаней між первинними елементами класифікації множини m_α і оператор перетворення $\Phi(r)$ матриці відстаней, який визначає вплив шуму на цю матрицю:

(2)

(3)

Для визначення залежності впливу шумів навколишнього середовища на достовірність розпізнавання звуків мови в дисертації проведено відповідні експериментальні дослідження. В роботі наводяться рівняння для розрахунку оператора $\Phi(r)$ для голосних і приголосних звуків, а також значення

достовірності апроксимації R^2 . Встановлено, що для опису моделі потенційної системи розпізнавання мови голосні звуки можуть описуватись лінійним рівнянням, а приголосні звуки для рівня шуму r , такого що $r > 5$, описуються рівнянням другого порядку, а для рівня шуму $r < 5$ - рівнянням третього порядку. Таким чином, в дисертації отримано модифікований вигляд критерію ефективності процесу класифікації голосових команд з врахуванням впливу шумів навколишнього середовища та синтаксичної складності їх опису.

Формальна постановка задачі оптимізації процесу класифікації голосових команд може бути подана в наступному вигляді [12]:

(4)

де \tilde{S}_{Gi} - одна із стратегій розпізнавання із замкнутої відносно доступної інформації множини стратегій розпізнавання \tilde{S}_G ; W_d, r_d, E_d - задані умовою задачі класифікації словник, рівень завад і точність класифікації відповідно.

Декомпозицію цих змінних в критерії (1) можна здійснити шляхом зображення стратегії розпізнавання в вигляді покрокової процедури класифікації.

Опишемо довільний модуль ієрархії процесу класифікації. Складовими частинами i -го функціонального модулю є:

S^i - алфавіт, яким описуються голосові команди, що класифікуються;

$\tilde{I}(S^i)$ - параметричний опис мовного сигналу (математична модель на рівні попередньої обробки),

$C_i(x_l^i)$ - обчислювальна складність l -ї ознаки, що використовується для опису мовних образів;

$A_i(\tilde{I}(S^i))$ - математична модель прийняття рішення (класифікатор);

τ_{cp}^i - середній час класифікації.

Процес розпізнавання полягає в знаходженні оптимальної послідовності кроків класифікації:

(5)

Процес розпізнавання можна представити у вигляді дерева класифікації, в якому є один кореневий вузол, а всі інші вузли - термінальні. Кількість гілок такого дерева (кроків класифікації) дорівнює кількості термінальних вузлів (образів, що класифікуються). На основі ентропійних властивостей множини звукотипів, які використовуються для опису голосових команд, в дисертації обґрунтовано вибір оптимальної структури дерева класифікації, яка дає можливість мінімізувати загальну помилку класифікації і одночасно підвищити швидкість класифікації.

Розв'язання задачі побудови моделі оптимізації процесу розпізнавання голосових команд у вигляді дерева класифікації можна здійснити з допомогою оптимізаційної процедури "керованого пошуку вперед з поверненням". В цій процедурі критерій (1) керує пошуком такої структури дерева рішень серед всіх можливих, в якій на кожному кроці пошуку вибирається та конфігурація вузлів, яка має найвище значення критерію [1]. Таким чином, застосування математичної моделі оптимізації процесу класифікації голосових команд у вигляді ієрархічного дерева дозволяє здійснити декомпозицію критерію ефективності та максимізувати його значення на кожному з ієрархічних рівнів процедури класифікації.

В другому розділі дисертації також розроблено математичну модель мовного сигналу на рівні попередньої обробки. Недоліком відомих методів розпізнавання мови, які використовують принцип частотного опису, є те, що спектральний опис проводиться у вузьких смугах спектра. Такий опис нестійкий до зміни диктора, у зв'язку з чим проводиться оновлення еталонів шляхом навчання. Тому в дисертаційній роботі запропонована модель квазічастотної модуляції мовного сигналу [7, 10], яка дозволяє уникнути вказаних недоліків.

Мовний тракт вважається джерелом інформаційного (мовного) сигналу, кодування повідомлення в якому проводиться шляхом модуляції трьох несучих частот - частоти I форманти, частоти II форманти і частоти III форманти. Спочатку спектральний аналіз проводиться у критичних смугах слуху, потім

шляхом відповідного об'єднання виходів фільтрів і зважування їх сигналів визначають стани частотних моментів M_{kf} сигналу:

(6)

де A_f – спектральна густина мовного сигналу для смуги частот df ; f – поточне значення частоти сигналу; k – номер частотного каналу, $k = 1, 2$.

Закодувавши декілька положень частотного моменту в кожній смузі частот, можна перейти від опису мовного сигналу в неперервному тривимірному просторі частотних моментів до дискретного опису моментів у просторі двійкових значень частотно-детектуючої функції. Аналіз спектральних властивостей звуків показує, що в кожному з вибраних частотних каналів необхідно розглядати такі форми спектра (положення частотних моментів): девіація в бік збільшення несучої частоти, девіація в бік зниження несучої частоти та нульова девіація. Для двійкового кодування цих положень частотного моменту необхідно розбити кожен канал на 3 піддіапазони, при цьому частотно-детектуюча функція θ_δ визначається таким чином:

(7)

де σ – одинична функція; α – відношення домінування, якщо i дорівнює 0 в протилежному випадку.

При такому визначенні частотно-детектуючої функції для першого зазначеного випадку девіації частоти $\theta_\delta = \{0, 0\}$, для другого – $\theta_\delta = \{1, 1\}$, для третього – $\theta_\delta = \{0, 1\}$.

При зміні положення органів артикуляції форма голосового тракту і його довжина змінюються, що відповідає зміні параметрів модуляції; положення формант і форма спектру вихідного сигналу змінюються відповідно тому чи іншому звуку мови. Таким чином, отриманий початковий опис мовного сигналу з допомогою частотно-детектуючої функції має вигляд восьмибітового двійкового слова. Значення цієї функції обчислюється для кожного τ -го первинного сегмента сигналу мови, тривалість якого вибрана рівною $t_s = 10$ мсек. Для кожної пари суміжних в часі значень частотно-детектуючої функції обчислюються значення сегментуючої функції за формулою:

(8)

де символ \oplus означає логічну операцію “сума по модулю два”.

Значення сегментуючої функції використовуються для процесу сегментації сигналу мови на окремі звукотипи за правилом:

(9)

де τ_s^j – j -та сегментна мітка; ρ_i – вага i -го двійкового розряду сегментуючої функції; δ_n – величина порогу, які визначаються в процесі навчання класифікатора з “учителем”.

Таким чином, розроблена математична модель дає можливість виконувати сегментацію мовного повідомлення на звукотипи паралельно з їх класифікацією, завдяки чому збільшується швидкодія процесу класифікації голосових команд. З експериментальних досліджень даної математичної моделі встановлено, що її застосування є доцільним для грубого опису мовного сигналу на рівні звукотипів “голосний”, “сонорний”, “шумний” та “вибуховий”. Параметричний опис мовного сигналу, що отримується, є більш стійким до зміни диктора за рахунок проведення спектрального аналізу мовного сигналу в більш широких смугах спектру та за рахунок використання відношення енергії сигналу в цих діапазонах при переході до опису сигналу в Хемінговому просторі, що усуває вплив амплітудних варіацій на достовірність класифікації.

Невід'ємною частиною процесу розпізнавання природної мови є процес синтаксичного виводу фраз дискурсу СРМ. Фрази дискурсу описуються граматикою. В другому розділі дисертації також розроблено математичну модель синтаксичного аналізу і класифікації голосових команд [6]. Основою даної моделі є модифікований синтаксис запису формальних граматик. Існуючі на сьогодні механізми виводу фраз граматик є неефективними, оскільки вимагають повного перебору граматичних правил та постійної "прив'язки" до правил граматик. В дисертаційному дослідженні пропонується механізм виводу фраз граматик, що базується на модифікованому синтаксисі запису контекстно-вільних граматик.

Розглянемо довільне правило граматик:

де $\langle M \rangle$ - нетермінальний символ; x_i - термінальний чи нетермінальний символ граматик.

В модифікованому синтаксисі кожний нетермінальний символ $\langle X \rangle$ розбиваємо на два символи: $\langle X \mid X \rangle$, які будемо називати, відповідно, лівою і правою дужкою нетерміналу.

Наступним етапом є розробка методу генерування фраз дискурсу СРМ.

Введемо поняття граматичної мережі [6].

Визначення 1:

Граматична мережа (ГМ) – це зважений граф, вершинами якого є символи дужкового синтаксису граматик G , а дуги R проводяться згідно правил P .

Визначення 2:

Навантаження – це правило граматик в модифікованому синтаксисі без першого входу в нетермінал і останнього виходу з нетерміналу.

Таким чином, ГМ містить $2n+k$ вершин, де n - кількість нетермінальних символів граматик, а k - кількість термінальних символів граматик. В дисертації наводяться правила побудови графа формальної граматик та правила переходу по ребрам графу в процесі генерації граматичних ланцюжків.

Для поєднання процесу виводу граматичних ланцюжків та процесу розрахунку ймовірності приналежності вхідної обсервації до дискурсу СРМ, введемо поняття граматичної марковської мережі.

Визначення 3:

Нехай $G\text{-net}$ – граматична мережа деякої граматик. Поставимо у відповідність кожній термінальній вершині даної мережі її приховану марковську модель (ПММ).

Тоді отриману таким чином мережу будемо називати граматичною марковською мережею (ГММ).

На рис.1 представлено приклад ГММ для граматик

$\langle m \rangle : \langle a \rangle z \mid z$

$\langle a \rangle : x;$

Вершини марковської мережі терміналу X будемо позначати:

$1_x, 2_x, \dots, (N-1)_x, N_x.$

Частина ребер ГММ – це тонкі ребра (стрілки), а частина – товсті. Це пов'язано з правилами перебігу часу в ГММ. Перехід по тій або іншій стрілці мережі або збільшує значення часу на 1, або залишає значення часу незмінним.

Рис. 1. Приклад ГММ

У початковий момент часу $t = 0$ “ми знаходимося” у вершині $< \text{main}$. Перехід по стрілці, що веде з деякої нетермінальної або термінальної вершини в початкову вершину деякої марковської моделі, супроводжується збільшенням часу на 1. Перехід по стрілці з емісійної в емісійну вершину деякої марковської моделі також супроводжується інкрементом часу. В усіх інших випадках “час не змінюється”.

Математичний апарат ГММ дозволяє розробити новий алгоритм класифікації голосових команд, що є переносом на граматичні марковські мережі класичного алгоритму Вітербі. В кожний момент часу t з кожною вершиною V граматичної марковської мережі пов’язується деяка множина граматичних ланцюжків і їх оцінок Вітербі:

де α - ланцюжок; $\phi_V(t, \alpha)$ - оцінка Вітербі граматичного ланцюжка α в вершині V в момент часу t .

Загальний алгоритм процесу класифікації голосових команд:

1. В момент часу t виконується асоціювання граматичних ланцюжків з вершинами мережі.
2. Розраховуються оцінки Вітербі для кожної вершини.
3. Визначаються групи ланцюжків мережі, що мають спільне граматичне майбутнє.
4. В кожній групі ланцюжків залишається тільки найкращий кандидат.
5. Якщо $t = T$, то переходимо до п. 6, інакше повторюємо п. 1-4.
6. За максимумом оцінки Вітербі визначається найкращий кандидат-ланцюжок, який оголошується розпізнаним.

Математична модель синтаксичного аналізу та класифікації на основі граматичних марковських мереж дозволяє об’єднати процес граматичного виводу фраз граматики та процес розрахунку ймовірності належності фрази, яку вимовляє користувач, до дискурсу системи розпізнавання мови, завдяки чому збільшується швидкодія процесу класифікації голосових команд. Описаний в дисертації алгоритм класифікації дозволяє проводити неризиковані скорочення варіантів перебору, що не приводять до втрати потенціальних кандидатів на розпізнавання.

У **третьому розділі** розглянуто питання розробки методу класифікації голосових команд.

Оскільки під час класифікації за допомогою ПММ реально обчислюється не ймовірність генерації мовного сигналу, а щільність такої ймовірності, яка не є нормованою величиною, то виникає потреба розробки критерію оцінки якості класифікації. Даний критерій повинен не суперечити природі ПММ і його значення повинно бути нормованою величиною.

В роботі пропонується використовувати критерій абсолютного відхилення [5]:

(10)

Виконавши потрібні підстановки, кінцево отримуємо вираз:

(11)

В дисертації наводяться формули для розрахунку якості класифікації в довільній фонемі ГММ та в групі фонем. На основі запропонованого підходу до обчислення якості класифікації в дисертації розроблено метод зменшення перебору в процесі класифікації голосових команд на граматичних марковських

мережах, що полягає в порівнянні в кожний момент часу значень критерію якості класифікації в усіх вершинах ГММ з пороговими значеннями, завдяки чому зменшується множина вершин, в яких будуть проводитися розрахунки в наступні моменти часу, що скорочує час класифікації.

Згідно з класичним підходом до класифікації голосових команд прихованими марковськими моделями в процесі класифікації в кожний момент часу $t < T$ розраховуються оцінки Вітербі в усіх вершинах ПММ дискурсу. При цьому не враховується порядок розташування певної фонемі у фразі, а також нехтується інформація про тривалість звучання фонем. В результаті виникає ситуація, коли розраховується ймовірність породження останніх векторів обсервації в початкових ПММ фраз, що, безумовно, позбавлено сенсу. Для усунення вказаного недоліку в дисертації розроблено метод класифікації, що використовує інформацію про тривалість звучання фонем [2]. Дана інформація знаходиться статистично, шляхом обробки представницької вибірки звукового матеріалу. В результаті з кожною ПММ фонемі ми асоціюємо дані про її інтервал обсервування:

(12)

де $\tau_{\min}(\lambda_i)$ - мінімальна тривалість звучання фонемі, якій відповідає марковська модель λ_i ; $\tau_{\max}(\lambda_i)$ - максимальна тривалість звучання фонемі.

Постановка задачі формулюється наступним чином:

Дано :

- $O = \{O_1, \dots, O_T\}$ - послідовність векторів обсервації мовного сигналу;
- граматику СРМ в синтаксисі розширених форм Бекуса-Наура, згідно до міжнародного стандарту;
- $\lambda = \{\lambda_i\}$ - множина ПММ, що описує термінали граматики;
- $\tau = \{\tau_{\alpha_i}\}$ - множина інтервалів обсервування кожної з марковських моделей.

Необхідно: визначити ймовірність належності обсервації O до дискурсу G .

На вхід прихованої марковської моделі фонемі φ надходить множина граматичних ланцюжків фонем

(13)

де $name$ - ім'я ланцюжка, яке складається з конкатенації імен фонем, які передують фонемі φ , що розглядається.

Кожний ланцюжок характеризується наступною множиною наборів параметрів:

(14)

де t_g - час закінчення обсервування ланцюжка (час виходу з попередньої фонемі); ϕ - оцінка Вітербі вхідного ланцюжка $name$ в момент часу t_g ; Q - оцінка якості класифікації вхідного ланцюжка $name$ в момент часу t_g .

В процесі класифікації на "виході" моделі фонемі отримаємо:

(15)

(16)

де $name' = name + \varphi$; t_g' - час закінчення обсервування в фонемі φ ; ϕ' - оцінка Вітербі ланцюжка $name'$; Q' - оцінка якості класифікації ланцюжка $name'$.

Після проведення класифікації фонемі отримано множину $Seq_{out}(\varphi, k)$, яка містить більшу кількість елементів, ніж вхідна множина $Seq_{in}(\varphi, k)$. Для

зменшення кількості елементів вихідної множини фонем розроблено алгоритм зменшення перебору (алгоритм TimePerplexity):

1. Для кожної фонем за граматичною схемою класифікації розраховується вихідна множина.

2. Для всіх наборів параметрів $\langle t', \phi', Q' \rangle$ множини $Seq_{out}(\phi, k)$, що мають однакові t' , залишають тільки одного кандидата за максимумом оцінки Вітербі.

На рис.2 показана залежність швидкості класифікації від кількості граматичних конструкцій, якими користувач може відповідати альтернативно на запити системи.

Рис. 2. Залежність швидкості класифікації від обсягу голосових команд

З рис.2 видно, що використання модифікованого методу класифікації є доцільним навіть при невеликих обсягах словника.

Також у третьому розділі дисертації розроблено метод динамічного скорочення варіантів перебору граматичних ланцюжків в процесі класифікації. Аналітичний вигляд функції динамічного скорочення варіантів перебору можна записати як суму двох функцій з додатнім коефіцієнтом асиметрії:

(17)

Коефіцієнти C_1 і C_2 знаходяться за методом найменшої квадратичної похибки. Нелінійні параметри λ_1 і λ_2 можуть бути знайдені за допомогою симплекс-методу. Параметри підбираються таким чином, щоб результуюча функція “покривала” дискретну функцію. Таким чином, для кожного дискретного моменту часу визначається кількість граматичних ланцюжків, в яких необхідно проводити розрахунок ймовірності генерації векторів мовної обсервації, що класифікується.

Четвертий розділ дисертаційної роботи присвячено практичній реалізації результатів, отриманих в другому і третьому розділах.

Розроблено методику проектування та оптимізації систем розпізнавання мови [4, 9]. Враховуючи необхідність забезпечення максимальної ефективності процесу розпізнавання, висунуто рекомендації щодо вибору технічних засобів, на основі яких можна побудувати систему розпізнавання мови.

На основі математичної моделі аналізу і класифікації голосових команд та методу класифікації голосових команд на граматичних марковських мережах розроблено програмне забезпечення, яке працює під управлінням операційних систем сімейства Windows, і призначене для моделювання та оптимізації процесу класифікації голосових команд.

Проведено дослідження адекватності розроблених математичних моделей та обчислювальної ефективності запропонованих алгоритмів. Зокрема, були проведені експерименти з оцінки ефективності існуючих та розроблених методів класифікації голосових команд. Проведені дослідження показали, що застосування синтаксичної моделі голосових команд збільшує швидкодію процесу класифікації на 12%. Встановлено, що використання пофонемного методу класифікації голосових команд на граматичних марковських мережах є доцільним для СРМ, в яких паралельно класифікуються до 500 слів. Проведено експерименти з оцінки достовірності прийнятого рішення на основі модифікованого узагальненого функціонально-статистичного критерію та розробленого критерію якості класифікації на ПММ. Встановлено, що метод оцінки за критерієм якості класифікації доцільно використовувати для граматик з середньою довжиною фраз в 2-3 слова і більше.

В **додатках** наведено алгоритм маркування та сегментації мовного сигналу, що базується на “квазічастотній” моделі мовоутворення, список фраз тестового матеріалу та лістинг граматики, які використовувались в експериментальних дослідженнях, відомості про впровадження результатів дисертації та деклараційний патент України на винахід.

ВИСНОВКИ

У дисертаційній роботі наведено теоретичне узагальнення і нове вирішення актуальної наукової задачі, яка полягає в розробці математичних моделей та методів, призначених для моделювання процесу аналізу і класифікації голосових команд, на основі яких можна підвищити ефективність даного процесу.

Основні наукові та практичні результати дисертаційної роботи такі.

1. Виконано аналіз відомих підходів до моделювання процесу аналізу і класифікації голосових команд.

2. Вперше розроблено математичну модель оптимізації процесу класифікації голосових команд, яка, на відміну від існуючих, враховує ієрархічну будову мовного сигналу та дозволяє оптимізувати параметри модульних систем розпізнавання голосових команд. За критерій оптимізації було обрано узагальнений функціонально-статистичний критерій оцінки ефективності складних систем, який було модифіковано для задачі класифікації голосових команд. В модель потенціальної системи розпізнавання була введена оцінка впливу шумів навколишнього середовища на достовірність класифікації звукотипів мови. Адекватність розробленої моделі забезпечуються коректністю та строгістю постановки задачі і коректністю використаних методів при доведенні наукових положень.

3. У результаті математичного моделювання отримані аналітичні залежності достовірності класифікації звукотипів української мови від впливу шуму навколишнього середовища.

4. Вперше запропоновано математичну модель мовного сигналу на рівні параметричної обробки, в якій, на відміну від існуючих, процес мовоутворення базується на принципі “квазічастотної” модуляції голосового тракту, що дозволяє підвищити дикторонезалежність ознак, що виділяються, та збільшити швидкість і достовірність класифікації. Адекватність “квазічастотної” моделі мовного сигналу підтверджується збігом теоретичних результатів, що отримані в дисертації, з результатами, які були отримані з використанням відомих моделей сегментації мовних сигналів на звукотипи, та з експертними даними. Проведені дослідження “квазічастотної” моделі мовного сигналу показали, що ефективність процесу сегментації мовного повідомлення на основі даної моделі є більшою на 2,6% в порівнянні з існуючими математичними моделями сегментації мовного сигналу.

5. Вперше розроблено математичну модель синтаксичного аналізу і класифікації голосових команд, яка, на відміну від існуючих, дозволяє врахувати інформацію про синтаксичну структуру голосових команд і представити її у вигляді орієнтованого графа. Запропонована модель дає можливість збільшити швидкість класифікації на синтаксичному рівні без втрат в якості класифікації. Адекватність даної математичної моделі підтверджується збігом теоретичних результатів, що отримані в дисертації, з результатами, які були отримані з використанням відомих методів класифікації голосових команд. Експериментальні дослідження розробленої моделі показали, що застосування даної моделі до процесу класифікації голосових команд дає можливість збільшити швидкість класифікації на 12% в порівнянні з класичними моделями процесу класифікації голосових команд.

6. Запропоновано модифікований спосіб запису контекстно-вільних граматик та процедуру граматичного виводу фраз граматики. На відміну від існуючих підходів до генерації фраз контекстно-вільної граматики,

запропонований підхід відрізняється простотою реалізації та забезпечує можливість паралельного виводу всіх фраз формальної граматики.

7. Вперше розроблено метод класифікації голосових команд, використання якого дозволяє збільшити швидкодію процесу класифікації на прихованих марковських мережах. На відміну від існуючих, в запропонованому методі враховується інформація про тривалість звучання окремих фонем дискурсу, що дозволяє зменшити час класифікації навіть для СРМ з невеликими за обсягом словниками.

8. Отримано аналітичний вираз функції динамічного скорочення варіантів перебору в процесі класифікації голосових команд на базі прихованих марковських мереж.

9. Вперше запропоновано спосіб кодування мовних образів ранговими кодами та математичну модель нейронної мережі, призначеної для класифікації, на основі рангових структур, які, на відміну від існуючих, враховують інформацію про ранги відстаней між елементами класифікації і дозволяють підвищити швидкість і достовірність класифікації.

10. Розроблено методику проектування та оптимізації систем розпізнавання голосових команд.

11. Розроблено алгоритми сегментації та маркування мовного сигналу на звукотипи, що використовують “квазічастотну” модель мовоутворення, та алгоритми класифікації голосових команд, які ґрунтуються на використанні розроблених методів та моделей класифікації голосових команд на граматичних марковських мережах.

12. На базі розроблених моделей, методів і алгоритмів процесу аналізу і класифікації голосових команд були створені програмні середовища для розробки контекстно-вільних граматик та систем розпізнавання голосових команд.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Биков М.М., Грищук Т.В. Ієрархічна стратегія розпізнавання мови // Вісник Технологічного університету Поділля. - 2004. - №2. - Т. 2. - С.58-61.

2. Биков М.М., Грищук Т.В. Підвищення швидкодії розпізнавання мови прихованими марковськими моделями // Комп'ютерні технології друкарства. - Львів: Українська академія друкарства, 2005. - №13. - С. 99-107.

3. Биков М.М., Грищук Т.В. Розробка методів оцінки ефективності автоматизованих систем розпізнавання мови // Вісник Технологічного університету Поділля. - 2003. - №3. Т. 1 - С.122-125.

4. Биков М.М., Грищук Т.В., Раїмі А.А. Використання нейронних мереж для розпізнавання звуків мови // Оптикоелектронні інформаційно-енергетичні технології. - 2001. - №2. - С.92-97.

5. Грищук Т.В. Методи зменшення перебору в процесі розпізнавання мови на граматичних марковських мережах // Вісник Вінницького політехнічного інституту. - 2003. - №6. - С.250-255.

6. Грищук Т.В. Розпізнавання природної мови на граматичних марковських мережах // Наукові праці Донецького національного технічного університету. Серія: “Обчислювальна техніка та автоматика”. - Донецьк: ДонНТУ, 2005. - С. 181-187.

7. Биков М.М., Грищук Т.В. Методи підвищення дикторонезалежності опису і розпізнавання мовної інформації в мережі INTERNET // Третя міжнародна конференція “Інтернет-Освіта-Наука-2002” (ІОН - 2002). - Вінниця: УНІВЕРСУМ - Вінниця, 2002. - Т. 2.- С.329-332.

8. Биков М.М., Грищук Т.В. Оцінка впливу рівня шумів на ефективність систем розпізнавання слів української мови // Матеріали сьомої міжнародної науково-технічної конференції “Контроль і управління в складних системах” (КУСС – 2003). – Вінниця: УНІВЕРСУМ - Вінниця, 2003. – С. 65-70.

9. Биков М.М., Грищук Т.В. Розпізнавання мовних образів з використанням нейромережевого підходу // Праці міжнародної конференції з індуктивного моделювання (МКІМ-2002.). – Львів: Державний НДІ інформаційної інфраструктури, 2002. – Т. 1., Ч. 2.- С. 203-207.

10. Биков М.М., Кузьмін І.В., Грищук Т.В. Структурування фонетичної інформації в акустичному сигналі в задачах дикторонезалежного розпізнавання мови // Сборник научных трудов 1-го международного радио-электронного форума “Прикладная радиоэлектроника. Состояние и перспективы развития” (МРФ-2002). – Харьков: АН ПРЭ, ХНУРЭ, 2002. – С.134-138.

11. Деклараційний патент 66184А України, МПК G06E1/04. Спосіб розпізнавання мовних образів / Биков М.М., Грищук Т.В. (Україна) - №2003087546; Заявлено 11.08.2003; Опубл. 15.04.2004, Бюл.№4. – 3 с.

12. Биков М.М., Грищук Т.В. Оптимізація процесу пошуку в задачі розпізнавання мовних образів // Тези доповідей восьмої міжнародної науково-технічної конференції “Контроль і управління в складних системах” (КУСС – 2005). – Вінниця: УНІВЕРСУМ - Вінниця, 2005. – С. 263.

АНОТАЦІЯ

Грищук Т.В. Моделювання процесу аналізу і класифікації голосових команд. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 – математичне моделювання та обчислювальні методи. – Вінницький національний технічний університет, Вінниця, 2006.

Дисертація присвячена розв’язанню задачі моделювання процесу аналізу і класифікації голосових команд.

Для моделювання процесу аналізу і класифікації голосових команд розроблено математичну модель оптимізації даного процесу у вигляді дерева класифікації. В якості критерію оптимізації обрано узагальнений функціонально-статистичний критерій, модифікований для задачі класифікації голосових команд, що можуть формально описуватись складними синтаксичними структурами. Для моделювання процесу аналізу мовного сигналу на рівні попередньої обробки розроблено “квазічастотну” модель мовоутворення, в якій мовний тракт вважається джерелом інформаційного сигналу, передача повідомлення в якому здійснюється шляхом модуляції трьох несучих частот. Розроблено метод класифікації голосових команд та математичну модель синтаксичного аналізу і класифікації голосових команд на граматичних марковських мережах. Запропонована математична модель узагальнює процеси синтаксичного аналізу формальної граматики та класифікації вхідних голосових команд, що дає можливість проводити скорочення граматичних ланцюжків в процесі класифікації без втрат в достовірності класифікації. В розробленому методі класифікації голосових команд використовується інформація про тривалість звучання окремих фонем мови, за рахунок чого досягається підвищення швидкодії процесу класифікації голосових команд. Запропоновано спосіб кодування мовних образів ранговими конфігураціями та модель нейтронної мережі, призначеної для класифікації.

На основі запропонованих моделей та методів розроблено методичні, алгоритмічні та програмні засоби для моделювання та оптимізації систем розпізнавання мови.

Ключові слова: моделювання, аналіз і класифікація, голосові команди, ефективність, оптимізація процесу класифікації, марковські моделі, розпізнавання мови.

АННОТАЦИЯ

Гришук Т.В. Моделирование процесса анализа и классификации голосовых команд. - Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.02 - математическое моделирование и вычислительные методы. - Винницкий национальный технический университет, Винница, 2006.

Диссертация посвящена решению задачи моделирования процесса анализа и классификации голосовых команд.

Проведен анализ процесса анализа и классификации голосовых команд, принципов создания современных систем распознавания речи, а также методов и моделей оптимизации таких систем.

Для моделирования процесса анализа и классификации голосовых команд разработана математическая модель оптимизации данного процесса в виде дерева классификации. В качестве критерия оптимизации выбран обобщенный функционально-статистический критерий, модифицированный для задачи классификации голосовых команд. Процесс распознавания представляется в виде дерева классификации. На основе энтропийных свойств множества звукотипов, используемых для описания голосовых команд, в диссертации обоснован выбор оптимальной структуры дерева классификации, которая дает возможность минимизировать общую ошибку классификации и одновременно повысить скорость классификации.

Для моделирования процесса анализа речевого сигнала на уровне предварительной обработки разработана "квазичастотная" модель речеобразования, в которой голосовой тракт считается источником информационного сигнала, передача сообщения в котором происходит путем модуляции трех несущих частот. Сначала спектральный анализ проводится в критических полосах слуха, а затем путем соответствующего объединения выходов фильтров и взвешивания их сигналов определяют положения частотных моментов, что увеличивает дикторнезависимость описания речевого сигнала. Разработанная модель позволяет выполнять параллельно сегментацию и классификацию выделенных звукотипов речи, что увеличивает скорость классификации голосовых команд.

Разработана модель синтаксического анализа и классификации голосовых команд. Основой данной синтаксической модели является модифицированный синтаксис записи формальных грамматик, который дает возможность представлять контекстно-независимую грамматику в виде нагруженного ориентированного графа, вершины которого соответствуют терминалам и нетерминалам моделируемой грамматики. Каждой терминальной вершине ставится в соответствие ее скрытая марковская модель. Таким образом, предложенная математическая модель позволяет совмещать процессы синтаксического анализа и классификации голосовых команд, что дает возможность увеличить скорость классификации голосовых команд.

В разработанном методе классификации голосовых команд на скрытых марковских сетях используется информация про длительность звучания отдельных фонем языка, за счет чего достигается увеличение быстродействия процесса классификации голосовых команд.

Разработан метод динамического сокращения вариантов перебора грамматических цепочек в процессе классификации. Получен аналитический вид функции динамического сокращения вариантов перебора в виде суммы двух функций с положительным коэффициентом асимметрии.

Предложен способ кодирования речевых образов ранговыми конфигурациями и модель распознающей нейронной сети.

На основе предложенных моделей и методов разработаны методические, алгоритмические и программные средства для моделирования и оптимизации систем распознавания речи.

Ключевые слова: моделирование, анализ и классификация, голосовые команды, эффективность, оптимизация процесса классификации, марковские модели, распознавание речи.

ANNOTATION

Gryshchuk T.V. Modeling of voice commands analysis and classification process. – Manuscript.

Thesis for a candidate's degree of technical sciences by specialty 01.05.02 – mathematical modeling and computational methods. – Vinnytsia National Technical University, Vinnytsia, 2006.

The thesis is devoted to solving the task of voice commands analysis and classification process modeling.

The new mathematical model of voice commands analysis and classification process in the form of a classification tree is created. The generalized functional-statistical criterion that was modified for classification of voice commands, which can be described with complex syntax structures, is chosen as an optimization criterion. The new quazifrequency model of speech production is created for speech signal analysis modeling on the preprocessing level. According to this model, the vocal tract is considered as a source of information signal, in which a message transfer is performed by three carrier frequencies modulation. The method of voice commands classification and voice commands analysis and classification are developed on the basis of the grammatical Markov networks. The offered model combines parsing and input voice commands classification processes, so it's possible to reduce grammatical chains set not decreasing the classification quality. In the offered method of voice commands classification, the information about phonemes length is used, so in this way voice commands classification performance is raised. The method of speech patterns encoding with rank configurations and the model of neural network classification are offered.

On the basis of the offered models and methods methodical, algorithmic and program means for speech recognition systems modeling and optimization are developed.

Keywords: modeling, analysis and classification, voice commands, effectiveness, optimization of the classification process, Markov models, speech recognition.

Підписано до друку 02.02.2006 р. Формат 29.7×42 1/4
Наклад 100 прим. Зам. № 2006-018
Віддруковано в комп'ютерному інформаційно-видавничому центрі
Вінницького національного технічного університету
м. Вінниця, вул. Хмельницьке шосе, 95. Тел.: 44-01-59