

Міністерство освіти і науки України
Вінницький національний технічний університет

М. М. Биков, Т. В. Грищук

**МОДЕЛЮВАННЯ ПРОЦЕСУ
АНАЛІЗУ І КЛАСИФІКАЦІЇ
ГОЛОСОВИХ КОМАНД**

Монографія

Вінниця
ВНТУ
2009

УДК 681.3.06
Б 60

Рекомендовано до видання Вченою радою Вінницького національного технічного університету Міністерства освіти і науки України (протокол № 2 від 25.09.2008 р.)

Рецензенти:

Р. Н. Квєтний, доктор технічних наук, професор

В. М. Лисогор, доктор технічних наук, професор

Биков, М. М.

Б 60 Моделювання процесу аналізу і класифікації голосових команд : монографія / М. М. Биков, Т. В. Гришук. – Вінниця : ВНТУ, 2009. – 128 с.

ISBN 978-966-641-322-5

В монографії проведено ґрунтовний аналіз процесу аналізу і класифікації голосових команд. Для моделювання процесу аналізу і класифікації голосових команд розроблено математичну модель оптимізації цього процесу у вигляді дерева класифікації. Представлено математичну модель мовного сигналу на рівні попередньої обробки та математичну модель синтаксичного аналізу і класифікації голосових команд на ґраматичних марковських мережах. Запропонована технологія розпізнавання узагальнює процеси синтаксичного аналізу формальної ґраматички та класифікації вхідних голосових команд, що дає можливість скорочувати варіанти в процесі класифікації без втрат достовірності класифікації.

Монографія розрахована на науковців, фахівців з питань розпізнавання мови, викладачів, аспірантів, студентів.

УДК 631.3.06

ISBN 978-966-641-322-5

© М. Биков, Т. Гришук, 2009

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	5
ВСТУП	7
РОЗДІЛ 1. АНАЛІЗ СУЧАСНОГО СТАНУ МОДЕЛЮВАННЯ ПРОЦЕСУ АНАЛІЗУ І КЛАСИФІКАЦІЇ ГОЛОСОВИХ КОМАНД	10
1.1. Класифікація систем розпізнавання мови	10
1.2. Аналіз математичних моделей класифікації мовних сигналів	13
1.2.1. Акустичне моделювання.....	13
1.2.2. Динамічне програмування	14
1.2.3. Статистична класифікація мовних образів	16
1.2.4. Приховані марковські моделі	18
1.2.5. Штучні нейронні мережі.....	23
1.3. Аналіз математичних моделей мовоутворення і мовосприйняття.....	25
1.3.1. Лінійна модель мовоутворення	26
1.3.2. Модель мовоутворення на основі лінійного прогнозу	27
1.3.3. Математична модель мовосприйняття	28
1.4. Визначення способів представлення мови на синтаксичному рівні	29
Висновки та постановка задач досліджень	34
РОЗДІЛ 2. МАТЕМАТИЧНІ МОДЕЛІ ПРОЦЕСУ АНАЛІЗУ І КЛАСИФІКАЦІЇ ГОЛОСОВИХ КОМАНД.....	36
2.1. Вибір критерію ефективності систем розпізнавання мови.....	36
2.2. Метод оцінки ефективності систем розпізнавання мови.....	37
2.2.1. Оцінка інформативності систем розпізнавання мови на лексичному рівні.....	37
2.2.2. Оцінка інформативності систем розпізнавання мови на синтаксичному рівні	44
2.3. Математична модель оптимізації процесу класифікації голосових команд.....	49
2.4. Математичні моделі процесу аналізу і класифікації голосових команд на різних рівнях ієрархії.....	55
2.4.1. Математична модель мовного сигналу на рівні попереднього аналізу мовної інформації	56
2.4.2. Механізм виведення фраз формальної граматики	62
2.4.3. Математична модель процесу аналізу і класифікації голосових команд на граматичних марковських мережах	69
РОЗДІЛ 3. ЗАСТОСУВАННЯ МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ОПТИМІЗАЦІЇ ПРОЦЕСУ КЛАСИФІКАЦІЇ ГОЛОСОВИХ КОМАНД	76

3.1. Реалізація процедури ефективної стратегії класифікації голосових команд для випадку моделювання дискурсу	76
3.2. Метод класифікації голосових команд на базі прихованих марковських мереж	81
3.2.1. Якість класифікації фонемі	81
3.2.2. Оптимізація процесу класифікації голосових команд за критерієм якості класифікації фонем	84
3.2.3. Використання інформації про тривалість звучання фонем для збільшення швидкодії процесу класифікації	87
3.2.4. Динамічне скорочення варіантів перебору в процесі класифікації голосових команд	93
3.2.5. Скорочення варіантів перебору N-Best-Node	94
3.3. Метод кодування мовних образів ранговими кодами та математичної моделі нейронної мережі, призначеної для класифікації	95
РОЗДІЛ 4 ПРОГРАМНЕ СЕРЕДОВИЩЕ ПРОЕКТУВАННЯ СИСТЕМ РОЗПІЗНАВАННЯ ГОЛОСОВИХ КОМАНД ТА ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ	102
4.1. Методика проектування та оптимізації систем розпізнавання голосових команд	102
4.2. Інструментарій для розробки KB-граматик	104
4.2.1. Загальна функціональність	104
4.2.2. Синтаксис запису граматик	105
4.3. Інструментарій для розробки граматичних марковських мереж	106
4.4. Дослідження адекватності “квазічастотної” моделі мовного сигналу	109
4.5. Дослідження адекватності математичної моделі та методу класифікації голосових команд на базі прихованих марковських мереж	114
4.5.1. Класифікація голосових команд з врахуванням спільного граматичного майбутнього граматичних ланцюжків	114
4.5.2. Класифікація голосових команд з використанням інформації про тривалість звучання фонем	115
4.5.3. Динамічне скорочення варіантів перебору	116
4.5.4. Оцінка достовірності прийнятого рішення за критерієм ефективності систем розпізнавання мови	117
4.5.5. Оцінка достовірності прийнятого рішення за критерієм якості класифікації послідовності фонем прихованої марковської мережі	118
ВИСНОВКИ	119
ЛІТЕРАТУРА	120

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

a_{ij}	ймовірність переходу з i -ї вершини в j -ту вершину прихованої марковської моделі;
$a_i(t)$	коефіцієнти лінійного прогнозу;
$b_j(t)$	ймовірність появи вектора спостереження o_t в вершині S_j ;
$c_q^m(t)$	кепстральні коефіцієнти;
d_{ij}^W	відстань між i -м та j -м словами;
$e(t)$	миттєва потужність сигналу в момент часу t ;
$e_{f_i}(t)$	миттєві спектральні потужності сигналу;
$F_t(jw, t)$	коефіцієнти Фур'є;
$O = \{o_1, \dots, o_T\}$	послідовність векторів спостереження (обсервацій мовного сигналу);
$o_t = \{x_{t1}, \dots, x_{tm}\}$	вектор спостереження в момент часу t ;
$q_j(t)$	якість класифікації в вершині j в момент часу t ;
$P(w_i / w_j)$	ймовірність подібності i -го слова словника на j -те слово;
$S = \{S_1, \dots, S_V\}$	множина вершин моделі;
$Seq_{in}(\varphi, k)$	множина вхідних граматичних ланцюжків та їх оцінок Вітербі для фонему φ на k -му граматичному етапі класифікації;
$Seq_{out}(\varphi, k)$	множина вихідних граматичних ланцюжків та їх оцінок Вітербі для фонему φ на k -му граматичному етапі класифікації;
$\tau(V, t) = \{< \alpha, \phi_V(t, \alpha) >\}$	множина граматичних ланцюжків та їх оцінок Вітербі для вершини V в момент часу t ;
$\tau\lambda_i = [\tau \min(\lambda_i), \tau \max(\lambda_i)]$	інтервал тривалості звучання ПММ λ_i ;

$V_N = \{V_{N1}, V_{N2}, \dots, V_{NK}\}$	множина нетермінальних вершин ГММ;
$V_T = \{V_{T1}, V_{T2}, \dots, V_{TM}\}$	множина термінальних вершин ГММ;
$W = \{W_i\}$	множина слів, що класифікуються;
$\lambda = \{\lambda_i\}$	множина ПММ терміналів ГММ;
r_d	рівень шуму навколишнього середовища (дБ);
$\xi_j(O_t)$	якість класифікації вектора обсервації O_t в вершині j в момент часу t ;
\sim	
μ_k	середня величина відстані для слова w_k ;
$\phi_V(t, \alpha)$	оцінка Вітербі граматичного ланцюжка α в вершині V в момент часу t ;
ГМ	граматична мережа;
ГММ	граматична марковська мережа;
ПММ	прихована марковська модель;
САРМ	система автоматичного розпізнавання мови;
СРМ	система розпізнавання мови.

ВСТУП

Класифікація образів (об'єктів, сигналів, ситуацій, явищ або процесів) – одна з найбільш розповсюджених задач, яку людині доводиться вирішувати майже щосекунди від першого до останнього дня свого існування.

Найбільш простим і природним для людини засобом обміну інформацією та подання команд служить мова. Проте автоматичне розпізнавання і розуміння мови є дуже складним. 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. Однак впровадження такого 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 цього інформаційного процесу.

Робота над розв'язанням проблеми автоматичного розпізнавання мови почалась ще у 50-х роках 20 століття. Серед закордонних фірм, що займаються вирішенням проблеми створення систем розпізнавання мови, можна назвати такі як Nuance, IBM, Kolvox, Kurzweil, Voice Connection, Phillips. До найбільш відомих існуючих систем можна віднести: Dragon NaturallySpeaking, IBM ViaVoice Gold, Philips FreeSpeech, Office Talk та інші.

Дослідження в галузі машинного розпізнавання мови проводяться вітчизняними науковцями Інституту кібернетики АН України, Національного технічного університету України “Київський політехнічний інститут”, Національного університету “Львівська політехніка”, Харківського національного університету радіоелектроніки, Донецького інституту проблем штучного інтелекту, Дніпропетровського національного університету, Вінницького національного технічного університету.

Не зважаючи на значні результати, одержані багатьма науковими колективами в розв'язанні проблеми розпізнавання мови, залишається нерозв'язаною значна кількість задач, розв'язання яких дозволило б підвищити ефективність систем розпізнавання мови, що дало б можливість розширити коло користувачів таких систем. Зокрема, актуальними є задачі оцінки ефективності систем розпізнавання мови, спрощення процедури їх розробки під конкретні застосування, розробки математичних моделей процесу аналізу і класифікації голосових команд з метою підвищення достовірності та швидкодії систем розпі-

знання мови. Саме тому актуальними є дослідження, описані в монографії, що направлені на розв'язання цих задач.

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 цього 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.

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 швидкодію і достовірність класифікації, тому їх використання не є ефективним.

Викладене вище і визначило зміст наукових досліджень, описаних в монографії.

В другому розділі монографії пропонуються математичні моделі процесу аналізу і класифікації голосових команд. Так на основі узагальненого функціонально-статистичного критерію описується задача оптимізації процесу класифікації голосових команд, основою якого є декомпозиція різних стратегій розпізнавання мови. Для опису мовного сигналу на рівні попередньої обробки пропонується “квазічастотна” модель, що має властивості дикторонезалежності і дає можливість виконувати попередню сегментацію мовного сигналу.

Існують різні класи систем розпізнавання мови, що розрізняються за призначенням: голосові інтерфейси, системи диктування тексту, системи розшифрування записів, здійснених попередньо на цифрові носії, тощо. В монографії основна увага приділяється системам

першого типу. Основна відмінність таких систем полягає у тому, що на етапі проектування закладається певний набір команд, за допомогою яких користувач може контролювати роботу комп'ютерної системи. У монографії розглядається синтаксис запису набору команд СРМ та математичні моделі синтаксичного аналізу та класифікації голових команд, що використовують запропонований синтаксис.

В третьому розділі монографії описується метод розпізнавання голових команд та його модифікації, метою яких є скорочення часу прийняття рішення про розпізнавання.

В четвертому розділі монографії наводяться дані про практичні експерименти, що підтверджують адекватність наведених у монографії моделей та методів розпізнавання мови.

Ця монографія написана в рамках наукового напрямку, започаткованого М.М. Биковим у роботах [10–11], всі наведені математичні моделі є результатами кандидатської дисертації Т.В. Гришук.

Автори вдячні к. м. н. М.З. Грузману за постійну підтримку та цінні консультації.

РОЗДІЛ 1. АНАЛІЗ СУЧАСНОГО СТАНУ МОДЕЛЮВАННЯ ПРОЦЕСУ АНАЛІЗУ І КЛАСИФІКАЦІЇ ГОЛОСОВИХ КОМАНД

1.1. Класифікація систем розпізнавання мови

Робота над вирішенням проблеми автоматичного розпізнавання мови почалась ще у 50-х роках. В 1952 році Девід, Бідалф і Балашек з фірми Bell Telephone Laboratories розробили перший завершений зразок автоматичної системи розпізнавання [39]. В подальшому багато наукових центрів, в тому числі і в нашій країні, займалися вирішенням цієї проблеми. Фундаментальні дослідження теорії мови, які проводилися в 70-х роках минулого сторіччя в СРСР, стали основою багатьох сучасних програмних продуктів. Але перший серйозний прорив в галузі мовних технологій вдалось зробити лише в 1986 році у відомому американському агентстві перспективних досліджень Міністерства оборони DARPA [8].

В нашій країні дослідження в галузі автоматичного розпізнавання образів проводяться вже протягом 60 років. Сьогодні важливий внесок у розвиток даного напрямку робить Українська асоціація з оброблення інформації та розпізнавання образів (УАсОІРО) [59], членами якої є провідні фахівці вітчизняних технічних вузів. Необхідно відзначити особливе значення робіт Т.К. Вінцюка [12-14, 33] та М.І. Шлезінгера [66].

Існує міжнародний стандарт в галузі програмування систем розпізнавання мови, який прийнято зараз практично усіма відомими розроблювачами систем розпізнавання мови (СРМ) - Microsoft Speech API, згідно з яким в усьому світі прийнято проводити класифікацію систем розпізнавання мови [90]. Відповідно до згаданого стандарту, СРМ розрізняють за такими ознаками:

- інтервал між окремими словами;
- залежність від диктора;
- ступінь деталізації при заданні еталонів;
- розмір словника.

Системи розпізнавання, яким властива відносна незалежність від диктора, дозволяють користувачу працювати без попередньої настрійки, однак поліпшують надійність розпізнавання після навчання. Незалежність від диктора таких систем звичайно досягається за рахунок збереження звукових еталонів для усіх найбільш типових голосів носіїв мови. Це, безумовно, вимагає в кілька разів більшої продуктивності й обсягу пам'яті. Налаштування на голос диктора дикторозалежних систем займає звичайно від 30 хвилин до декількох годин. Це

складає головну незручність для користувача. Звичайно дикторозалежні системи дозволяють працювати з відносним ступенем надійності без попереднього настроювання на голос конкретного користувача. Третім різновидом систем за цією ознакою є системи, що автоматично настроюються на голос диктора в міру їх використання. Системи останнього типу мають дві основні особливості. По-перше, їм потрібно знати, чи зробив користувач помилку, вимовивши конкретне слово (інакше навчання буде невірним). А по-друге, після настроювання на одного диктора такі системи перестають надійно працювати з іншими дикторами.

На сьогодні найбільш популярними є системи голосового керування комп'ютером і системи дискретного диктування текстів. Системи голосового керування комп'ютером ("Command and Control" у термінології Microsoft Speech API [8, 90]) – це системи дикторонезалежного розпізнавання неперервних команд, складених зі слів обмеженого (до декількох сотень слів) словника. Для подібних систем, якщо користувач вимовляє команду, що не входить у список, система або видає відмову від розпізнавання, або видає як відповідь схожу "на слух" команду. Список команд звичайно інтуїтивно зрозумілий у кожній конкретній ситуації. Відповідно до стандарту Microsoft Speech API, системи голосового керування повинні працювати успішно на комп'ютерах 486/66 МГц при наявності 1 МБ вільної оперативної пам'яті.

На сьогоднішній день в світі найбільшого визнання та поширення здобули такі програмні продукти з розпізнавання мови [8, 28, 67], як Dragon Naturally Speaking, IBM Voice, OfficeTALK, KVWin, Micro-Introvoice, Speech Magic. Ці програмні продукти використовуються переважно в медичних закладах, юриспруденції та для роботи в офісах.

Системи дискретного диктування текстів ("Discrete Dictation" [90]) – це системи дикторозалежного розпізнавання дискретно вимовлених слів з великих за обсягом (десятки тисяч слів) словників. Подібні системи звичайно вимагають процесора Pentium/60 МГц і 8 МБ вільної оперативної пам'яті.

Системи диктування текстів є на сьогодні привабливими для покупців у силу новизни можливостей, що надаються користувачу, ефективної рекламної кампанії і здорового дослідницького інтересу до задачі. Однак реальні системи диктування повинні відповідати таким вимогам: час набору тексту з голосу, з урахуванням часу виправлення помилок, повинен бути меншим часу набору того ж тексту з клавіатури; користувач не повинен втомлюватися від читання тексту голосом більше, ніж від набору того ж тексту з клавіатури; вартість системи диктування повинна окуповуватись виграшем у часі диктування за порівняно короткий період.

Аналіз існуючих систем показує, що вони у більшості випадків не задовольняють жодну з цих вимог. Тому ці системи поки є не більш ніж дорогими мультимедіа-іграшками, як і переважна більшість інших мультимедіа-систем, які зараз продаються, (наприклад, систем навчання іноземним мовам).

Варто, однак, згадати, що системи диктування текстів на Заході знайшли своє практичне застосування в медицині. Це пов'язано в першу чергу з тим, що галузь наукових розробок для використання в медицині на Заході добре фінансується. Крім того, задача тут спрощується тим, що словники медичних термінів у вузькій предметній галузі мають менший обсяг, ніж словники повсякденного спілкування, а синтаксис і семантика речень, які надиктовуються, надзвичайно жорсткі, що підвищує надійність розпізнавання. Застосування розпізнавання мови значно підвищило ефективність роботи лікарів.

Сьогодні велике поширення здобувають портативні комп'ютерні засоби Pocket PC. Створення систем розпізнавання мови для таких пристроїв є актуальною задачею, оскільки використання інших типів введення інформації (через клавіатуру, мишу тощо) ускладнюється через малі розміри пристроїв. Однак, якість СРМ для Pocket PC не задовольняє очікування користувачів. Користувачі вимушені повторювати по декілька разів свої команди та звикнути до того, що час від часу комп'ютер буде помилятися і виконувати інші команди. При створенні СРМ для портативних комп'ютерів розробники повинні враховувати такі основні моменти [78]:

1. Програмне забезпечення розпізнавання мови вимагає великих витрат пам'яті – потрібно приблизно 4 Мб дискової пам'яті та від 4 до 5 Мб оперативної пам'яті.

2. Погана якість вбудованої акустичної системи – мікрофони, що вбудовані в портативні комп'ютери, не призначені для повночасотного запису з високою точністю відтворення.

3. Недоліки операційної системи Pocket PC 2002 від Microsoft.

Найбільш відомими на сьогодні системами розпізнавання для портативних комп'ютерів є [76]:

- Voice LookUp від HandHeld Speech [51];
- PDSay від ScanSoft [53];
- Fonix VoiceCentral від Fonix [50];
- Microsoft Voice Command від Microsoft Software [52].

Дане програмне забезпечення дозволяє переглядати контакти, а також запускати та перемикати програми. Дані системи є дикторозалежними з функціональністю адаптації до конкретного користувача.

1.2. Аналіз математичних моделей класифікації мовних сигналів

В перших пристроях автоматичного розпізнавання мови (АРМ), які з'явилися у 1948 році, для грубої обробки аналогових напруг акустичних образів, які задавались на відносно великих часових проміжках, з метою розпізнавання слів чи коротких висловів одного диктора, використовувалась порогова логіка. Вже тоді було встановлено, що ніякі прості виміри не будуть достатніми для розпізнавання мови. В цьому розділі будуть розглядатися математичні моделі процесу класифікації голосових команд, які можуть використовуватись в машинних АРМ.

Всі існуючі системи розпізнавання мови ґрунтуються на двох ідеях, які прийняті більшістю існуючих кількісних і якісних теорій мови [34, 60, 61].

Перша фундаментальна гіпотеза: безпосередня форма перенесення інформації мовним сигналом реалізована в часових змінах його миттєвого амплітудного спектра. Також загально визнано, що частина інформації передається часовими інтервалами між мовними актами і іншими просодичними ознаками, наприклад висотою і інтенсивністю звуку.

Друга фундаментальна гіпотеза: мова – це складний сигнал, ієрархічно організований таким чином, що більш прості образи нижчого рівня за однозначними правилами об'єднуються в більш складні образи вищого рівня. Отже, згідно з деяким складним кодом, будується багаторівнева ієрархічна система з завадостійкою структурою, що переносить інформацію.

В якості найпростіших вихідних образів в лінгвістичних теоріях звичайно обираються спектральні чи просодичні характеристики сигналу, з яких тим чи іншим чином формуються фонemi, основні фонетичні категорії, дифони, напівсклади, склади, супрасегментні фрази.

1.2.1. Акустичне моделювання

Найпростішим методом розпізнавання ізольованих слів є акустичне порівняння звукових сигналів. У цьому методі ідентифікують окрему фонему за допомогою її акустичних властивостей. Приміром, відомо, що фонема $|c|$ має довжину 50 мс і головна частота її прояву близько 44 кГц. Такі знання переводяться в правила і використовуються в статистичному методі класифікації.

Акустичні ознаки, які використовуються в машинному розпізнаванні мови:

- 1) частота першої форманти – $F1$;
- 2) частота другої форманти – $F2$;
- 3) частота третьої форманти – $F3$;
- 4) частота основного тону – $F0$;
- 5) частота першого антирезонансу – F_{z1} ;

- 6) частота другого антирезонансу – F_{z2} ;
- 7) середня потужність сигналу – E ;
- 8) миттєва потужність сигналу – ρ ;
- 9) тривалість голосних чи приголосних – t .

В якості різних ознак для опису мовного сигналу на акустичному рівні використовуються в основному ознаки, які описують властивості моделей мовоутворення. Деякі ознаки, отримані на основі неспектральних перетворень, наприклад, кількість нульових перетинів сигналу, описує його частотні властивості і також відповідає моделі мовоутворення неявним чином.

В переважній більшості робіт з акустичного аналізу використовується послідовна стратегія сегментації і фонетичної ідентифікації акустичних сигналів, згідно з якою спочатку визначається спосіб утворення звуку, а потім місце його утворення, після чого визначається фонемна належність звуку [42].

1.2.2. Динамічне програмування

Розпізнавання ізолюваних слів вимагає від методів класифікації володіння технікою аналізу, яка б використовувалась для випадків, коли еталони слів і відповідні їм мовні сигнали мають різну часову структуру. Для цього виконують часове вирівнювання послідовностей векторів. На рис. 1.1 показано приклад такого вирівнювання.

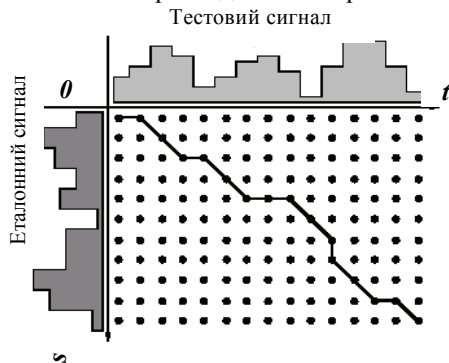


Рис. 1.1. Нелінійне часове вирівнювання двох різних сигналів

На рис. 1.1 горизонтальна та вертикальна послідовності ознак двох реалізацій одного і того ж слова. Різна часова і артикуляційно-геометрична структура обох слів вирівнюється за допомогою відносного викривлення часової шкали. Імовірне співвідношення між часовими точками сигналів зображено у вигляді шляху в решітці.

Ранні системи розпізнавання мали архітектуру класифікатора, яка зображена на рис.1.2.

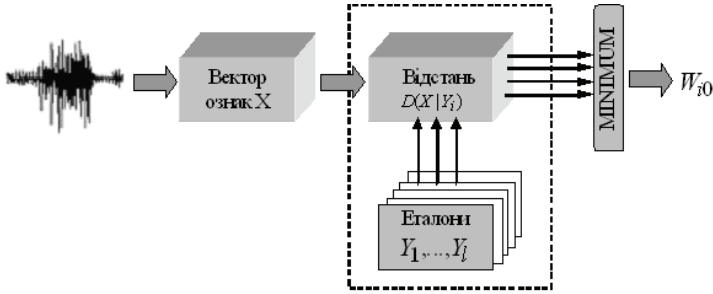


Рис. 1.2. Структура автоматичного класифікатора

Кожне слово W_l репрезентується за допомогою еталона Y_l (прототип, або шаблон). Кожний еталон Y_l представляє собою послідовність векторів ознак різних реалізацій слова W_l . Для того, щоб віднести невідоме слово до одного з відомих класифікатору класів, необхідно послідовність векторів ознак цього слова порівняти з усіма еталонами $Y_{l,1}, \dots, Y_{l,M_l}$. Рішення приймається на користь слова W_{l^*} , яке має найкращий відгук еталона, на основі правила

$$l^* = \arg \min_l \min_m D(X, Y_{l,m}).$$

Відстань $d_{x,y}^W$ між вхідною послідовністю і еталонною послідовністю $Y = y_1, \dots, y_s$, які мають різну довжину $S \neq T$, розраховується як сума локальних відстаней $d_{i,j} = d(x_i, y_j)$ вздовж шляху вирівнювання між послідовностями векторів; локальна функція відстані $d(\cdot, \cdot)$ реалізується через евклідову метрику.

Функція вирівнювання, яка використовується для відображення X на Y , є монотонною і неперервною як на t – шкалі, так і на s – шкалі, і повинна показувати мінімальну відстань між двома послідовностями ознак. Цю складну задачу дискретної оптимізації – зниження експоненційно зростаючого числа комбінаторно можливих шляхів вздовж довжини еталонної послідовності – можна розв'язати за допомогою динамічного програмування. Для обчислення кумулятивної відстані $D_{i,j} = D(x_1, \dots, x_i, y_1, \dots, y_j)$ між двома послідовностями векторів X, Y використовується рекурсивна формула

$$D_{i,j} = \begin{cases} 0 & i = j = 0 \\ \min\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\} + d_{ij} & i > 0, j > 0 \\ \infty & \text{в іншому випадку} \end{cases} \quad (1.1)$$

і загальна відстань між векторами $D(X, Y) = D_{ST}$ розраховується за $O(S \cdot T)$ обчислювальних операцій.

Кількість і форма мінімізованого результату в рекурсивному виразі (1.1) залежить від дозволених локальних переходів функції вирівнювання. З використанням даної структури було запропоновано велику кількість модифікованих методів обмежень шляху в решітці відстаней [79, 82, 87, 92, 112].

Цей ефективний алгоритм для розрахунку відстані між послідовностями незалежно від часу має назву динамічного вирівнювання і використовується з 1970 року в численних варіантах автоматизованих систем для розпізнавання слів [89, 112].

1.2.3. Статистична класифікація мовних образів

Велике поширення на сьогодні здобув статистичний метод розпізнавання мови [89]. На теперішній час в сфері автоматичного розпізнавання мови статистичний підхід де-факто є стандартом.

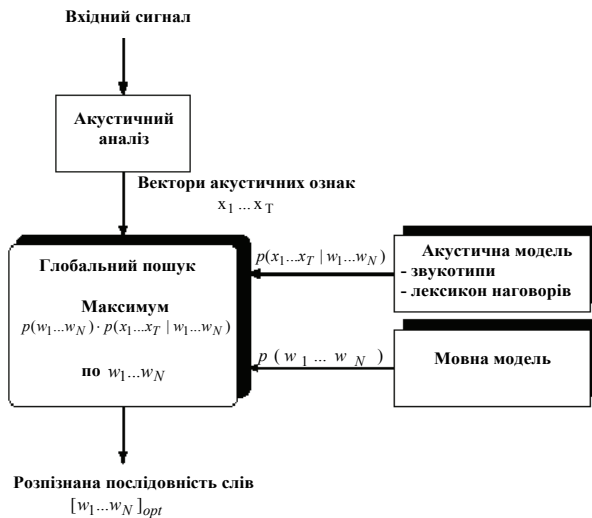


Рис. 1.3. Архітектура системи автоматичного розпізнавання мови

Як видно з рис. 1.3, архітектура системи автоматичного розпізнавання мови в структурі правил прийняття рішення Байєса складається з чотирьох основних компонентів:

1. Виділення акустичних ознак, тобто параметризація аналогового мовного сигналу.

2. Акустична модель. Для заданої послідовності слів ця модель визначає ймовірність обсервації відповідної тимчасової послідовності акустичних ознак. Акустичну модель можна поділити на два джерела знань:

- акустична модель для найменших звукотипів, які можуть бути розпізнані, а саме: фонемі, склади чи навіть моделі цілих слів;
- лексикон наговорів, який задає розподіл слів на звукотипи.

3. Мовна модель, статистична модель синтаксису, семантики і прагматики мови, що розпізнається.

4. Процедура пошуку послідовності слів з найвищою апостеріорною ймовірністю для даного мовного сигналу.

Ці складові будуть описані нижче. Надалі будуть виділені стандартні процедури для визначення відповідних стохастичних моделей.

Опис цього методу має такий вигляд.

Нехай

$$W = \{w_1, w_2, \dots, w_n\} \quad (1.2)$$

– множина, яка складається з n слів і нехай $O = \{o_1, o_2, \dots, o_T\}$ – акустичні ознаки (набір векторів спостереження), на основі яких пристрій розпізнавання приймає рішення про те, які слова були вимовлені. Якщо $P(W/O)$ є ймовірністю того, що при спостереженні ознак O вимовлені слова W , то пристрій повинен прийняти рішення на користь W при збереженні умови

$$P(\hat{W}/O) = \max_W P(W/O). \quad (1.3)$$

Критерій рішення (1.3) є природним і загальноприйнятим. Для мови, система писемності якої хоча б наближено є фонетичною, пристрій розпізнавання, що ґрунтується на (1.3), зводить до мінімуму кількість правок, необхідних для перетворення фрази, яка вимовлена диктором, в текст.

Перетворення мовного сигналу в дані O називається акустичною обробкою. Для обробки мовної інформації в системах автоматичного розпізнавання мови використовуються непараметричні та параметричні методи.

Необхідність в непараметричних методах виникла при переході від “миттєвого” спектру чи інших безпосередніх вимірів мовного сигналу, що допускають порівняння незалежно від їх положення в часі, до послідовності таких вимірювань, які необхідні для опису мовних сигналів більшої часової тривалості. Так, для порівняння вимов довжиною в одне слово необхідно сконструювати таку міру відстані, що

залишається інваріантною стосовно локальних змін часової шкали, яка не має відношення до тотожності висловлення. І навпаки, міра відстані, корисна для класифікації, має бути чутливою до таких змін сигналу в часі, у яких закодоване повідомлення.

1.2.4. Приховані марковські моделі

Разом з непараметричними методами обробки мовної інформації з наступним її розпізнаванням використовуються також і параметричні методи [34]. Стохастичні моделі, які використовуються на сьогоднішній момент, – це моделі, що описують двічі стохастичні процеси. Термін “двічі” використовується для позначення такої пари процесів, один із яких є основним, але прихованим від нас і спостерігається тільки через інший стохастичний процес. Моделі такого типу особливо зручні для опису мовного сигналу, оскільки в дійсності звукова хвиля тиску, яку ми вимірюємо, представляє тільки деякий код основного символічного процесу, що протікає в повністю недоступних ділянках мозку. Передбачається, що основний процес є марковським у тому сенсі, що у довільний момент часу його значення залежать тільки від деякого кінцевого інтервалу його найближчої передісторії. Це обмеження задає визначення моделі прихованого марковського процесу (МППМ) [71, 72, 105].

Задача розпізнавання ізольованих слів полягає в тому, що необхідно розпізнати слово, яке належить до множини слів $W = \{w_1, w_2, \dots, w_n\}$.

Для розв’язання цієї задачі необхідно обрати тип класифікатора, на вхід якого буде подаватися інформація про ознаковий опис фонем, а на виході ми будемо отримувати інформацію про ступінь приналежності невідомої фонемі до того чи іншого класу первинних елементів. Приховані марковські моделі (ПММ) відображають стохастичну природу мовного сигналу. Математичний апарат ПММ дозволяє виконувати класифікацію нестационарних образів, до яких і відноситься мовний сигнал.

Для опису ПММ введемо низку позначень її параметрів [41, 105]:

$S = \{S_1, S_2, \dots, S_N\}$ – множина вершин моделі;

$O = \{o_1, o_2, \dots, o_T\}$ – послідовність векторів спостереження (обсервацій мовного сигналу);

$o_t = \{x_{t1}, x_{t2}, \dots, x_{tm}\}$ – вектор спостереження, що представляє собою ознаковий опис мовного сигналу на певному стаціонарному проміжку;

N – кількість станів моделі;

T – кількість різних векторів спостереження;

$t = 1, 2, \dots$ – дискретні моменти часу;

A – матриця перехідних ймовірностей;

$a_{ij} = p(q_{t+1} = S_j | q_t = S_i)$ – імовірність переходу з вершини S_i у вершину S_j , ці ймовірності повинні задовольняти нормальні статистичні умови, тобто:

$$a_{ij} \geq 0, \quad i, j = 1, 2, \dots, N \quad \text{і} \quad \sum_{i=1}^N a_{ij} = 1, \quad j = 1, 2, \dots, N;$$

B – матриця ймовірностей появи k -го вектора спостереження при реалізації j -ї вершини;

$b_j(t) = p(o_t | q_t = S_j)$ – ймовірність появи вектора спостереження o_t у вершині S_j .

Вершини марковської моделі поділяються на емісійні та неемісійні. Кожна з емісійних вершин моделі навантажена певним законом розподілу. В неемісійних вершинах ПММ ймовірність появи векторів обсервації не розраховується і залежить від її розташування в моделі.

Марковська модель представляє собою “пристрій” для генерації (породження) обсервації спостереження. Кожна обсервація може бути отримана з деякою ймовірністю.

Визначення:

Прихована марковська модель – це імовірнісний граф, з кожною з емісійних вершин якого пов’язана функція розподілу векторів обсервації:

$$M = \langle G, \{b_j\} \rangle.$$

Марковська модель може бути представлена так, як це показано на рис. 1.4.

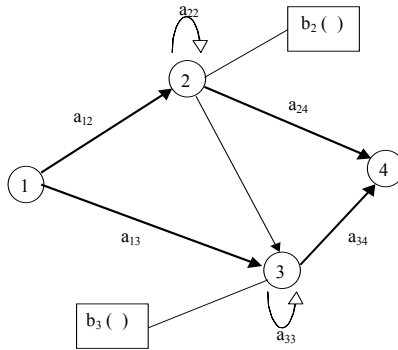


Рис. 1.4. Приклад ПММ

Для розв’язання задачі розпізнавання мови вершини ПММ, в загальному випадку, навантажуються нормальним законом розподілу.

Одиницею лінгвістичного представлення терміналів граматики прийнято обирати фонему природної мови. Опис фонему мови доцільно здійснювати за допомогою ПММ лінійного типу з трьома емісійними станами (рис. 1.5).

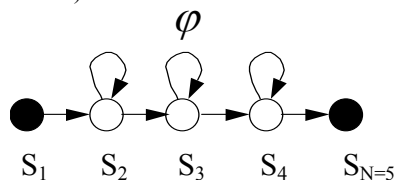


Рис. 1.5. ПММ лінійного типу

Це пов'язано з тим, що в неперервній мові фонему не спостерігається чітких акустичних переходів від однієї фонему до іншої, тобто між кожними двома фонемами є перехідні проміжки. Коли ми обираємо триемісійну структуру моделі, то обумовлюємо те, що в першій вершині мають породжуватися з великою ймовірністю вектори лівої перехідної зони, в середній вершині – власні вектори фонему, а в третій – вектори правої перехідної зони. Такий підхід повністю відповідає фонетичній класифікації фонему на трифону.

Для знаходження ймовірності породження вектора обсервації марковською моделлю використовують оцінки Вітербі [34, 105, 110].

Отримання марковської моделі для ланцюжка граматики відбувається шляхом їх конкатенації, як показано на рис. 1.6.

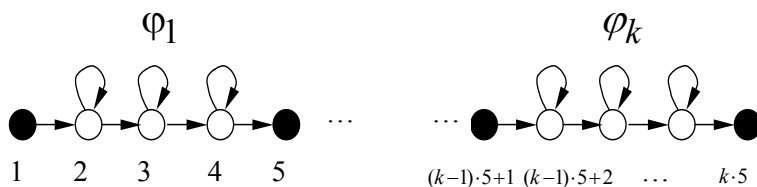


Рис. 1.6. Модель послідовності з k фонему

Якщо відбувається одночасна генерація граматичних ланцюжків, то марковські моделі терміналів об'єднуються в мережу, в якій існує тільки перехід від єдиного вихідного стану одного терміналу до єдиного вхідного стану іншого терміналу. Повторення такої перехідної структури породжує ґратчасту структуру.

Для класифікації невідомого вислову, складеного з послідовності слів заданого словника, необхідно знайти таку послідовність станів

ЛІТЕРАТУРА

1. Ахо А. Теория синтаксического анализа, перевода и компиляции “Синтаксический анализ” / А. Ахо, Дж. Ульман. – М. : Мир, 1978. – 224 с.
2. Биков М. М. Використання нейронних мереж для розпізнавання звуків мови / М. М. Биков, Т. В. Гришук, А. А. Раїмі // Оптикоелектронні інформаційно-енергетичні технології. – 2001. – №2. – С. 92–97. ISSN: 1681-7893.
3. Биков М. М. Ієрархічна стратегія розпізнавання мови / М. М. Биков, Т. В. Гришук // Вісник Технологічного університету Поділля. – 2004. – № 2. – Т. 2. – С. 58–61.
4. Биков М. М. Оптимізація процесу пошуку в задачі розпізнавання мовних образів : Тези доповідей восьмої міжнародної науково-технічної конференції [“Контроль і управління в складних системах”] (КУСС – 2005) / М. М. Биков, Т. В. Гришук. – Вінниця : УНІВЕРСУМ – Вінниця, 2005. – 263 с. – ISBN: 966-641-063-X
5. Биков М. М. Підвищення швидкодії розпізнавання мови прихованими марковськими моделями / М. М. Биков, Т. В. Гришук. // Комп’ютерні технології друкарства. – 2005. – №13. – С. 99–107. – ISBN: 966-322-013-9.
6. Биков М. М. Розробка методів оцінки ефективності автоматизованих систем розпізнавання мови / М. М. Биков, Т. В. Гришук. // Вісник Технологічного університету Поділля. – 2003. – Т. 1, № 3. – С. 122–125.
7. Биков М. М. 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) / М. М. Биков, І. В. Кузьмін, Т. В. Гришук. – Харьков : АН ПРЭ, ХНУРЭ, 2002. – С. 134–138. – ISBN 966-659-038-7.
8. Борисов М. ПО для распознавания речи / М. Борисов // Мир ПК. – 1999. – № 3. – С. 124–29.
9. Быков Н. М. Анализ эффективности средств распознавания в системах принятия решений : Эффективность и надежность автоматизированных систем принятия решений / Н. М. Быков, Ю. С. Данилюк, Н. П. Ильницький. – Киев : ИК АН УССР, 1982. – С. 24–29.
10. Быков Н. М. Методы и средства измерения и преобразования информации в системах машинного распознавания речи : дис. на соискание уч. ст. канд. техн. наук : 19.05.85 / Быков Николай Максимович. – Винница, ВПИ, 1985. – 243 с.
11. Быков Н. М. Модель потенциальной системы для распознавания речи в СЧМ : Исследование и проектирование систем “человек-машина” / Н. М. Быков, А. С. Агеев. – Киев : ИК АН УССР, 1989. – С. 57–61.
12. Винцюк Т. К. Анализ, распознавание и интерпретация речевых сигналов / Винцюк Т. К. – К. : Наукова думка, 1987. – 261с.
13. Винцюк Т. К. О математических моделях речевого сигнала, используемых в распознавании речи / Т. К. Винцюк // Автоматическое распознавание слуховых образов. – Киев: ИК АН УССР, 1982. – С. 34–37.

14. Винцюк Т. К. Поэлементное распознавание непрерывной речи, составленной из слов заданного словаря / Т. К. Винцюк // Кибернетика. – 1972. – № 2. – С. 35–38.
15. Волков Е. А. Численные методы / Е. А. Волков – М. : Наука, 1987. – 248 с.
16. Галайчук Г. Л., Перваченко В. С. Об одной модели слухового аппарата человека / Г. Л. Галайчук, В. С. Перваченко // Вестник КПИ. Серия электроакустики и звукотехники. – 1977. – Т. 1. – С. 10–11.
17. Гинзбург С. Математическая теория контекстно-свободных языков / С. Гинзбург – М. : Мир, 1977. – 328с.
18. Гришук Т. В. Методи зменшення перебору в процесі розпізнавання мови на граматичних марковських мережах / Т. В. Гришук // Вісник Вінницького політехнічного інституту. – 2003. – №6. – С. 250–255.
19. Гришук Т. В. Побудова голосових інтерфейсів на основі методу поетапного розпізнавання Вітербі : матеріали науково-практичної конференції [“Стан та перспективи розвитку новітніх науково-освітніх комп’ютерних технологій”], 14–15 лист. 2003 р. / Т. В. Гришук. – Миколаїв, 2003. – С. 72–73.
20. Гришук Т. В. Розпізнавання природної мови на граматичних марковських мережах / Т. В. Гришук // Наукові праці Донецького національного технічного університету. серія: “Обчислювальна техніка та автоматика”. – Донецьк : ДонНТУ, 2005. – С. 181–187. – ISBN: 966-377-001-5.
21. Гришук Т. В. Новый підхід до підвищення ймовірності розпізнавання в дикторонезалежних системах : Четвертый российский-украинский научный семинар [“Интеллектуальный анализ информации ИАИ–2004”], (Киев, 19–21 мая 2004 г.) / Т. В. Гришук, Ю. В. Дубова – К. : Просвіта, 2004. – С. 88–94.
22. Гросс М. Теория формальных грамматик / М. Гросс, А. Лантен. – М. : Мир, 1971. – 296 с.
23. Гудонавичюс Р. В. Распознавание речевых сигналов по их структурным свойствам / Гудонавичюс Р. В., Кемешис П. П., Гитавичюс А. Б. – Л. : Энергия, 1977. – 148 с.
24. Деклараційний патент 55863А України, МПК G10L15/00. Пристрій для виділення ознак мовних сигналів / Биков М.М., Гришук Т. В., Ковтун В. В. – № 2002075729 ; заявл. 11.07.02 ; опубл. 15.04.03, Бюл. № 4. – 3 с.
25. Деклараційний патент 66184А України, МПК G06E1/04. Спосіб розпізнавання мовних образів / Биков М. М., Гришук Т. В. – № 2003087546 ; заявл. 11.08.03 ; опубл. 15.04.04, Бюл. № 4. – 3 с.
26. Джелинек Ф. Разработка экспериментального устройства, распознающего раздельно произносимые слова / Ф. Джелинек // ТИИЭР, 1995. – №11. – С. 91–100.
27. Дуда Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – М. : Мир, 1976. – 512 с.
28. Електронні сторінки фірми 21st Century Eloquence [Електронний ресурс] – Режим доступу : <http://www.voicerecognition.com/>.

29. Загоруйко Н. Г. Методы распознавания и их применение / Загоруйко Н. Г. – М. : Сов. радио, 1972. – 207 с.
30. Карпов О. Н. Текстовый редактор с речевым управлением / О. Н. Карпов // Проблемы управления и информатики. – 1997. – №3. – С. 132–142. – ISSN : 0572-2691.
31. Кельманов А. В. Решения проблем выделения спектральных корреляционных признаков речевого сигнала / Кельманов А. В. – М. : Мир, 1991. – 170 с.
32. Книппер А. В. К описанию речевых сигналов : Речевое общение в автоматизированных системах / А. В. Книппер, В. А. Махони. – М. : Наука, 1975. – С. 45–59.
33. Конструктивный подход к распознаванию речи. Применение математического программирования в распознавании слуховых образов : тр. всесоюз. шк.-семинара АРСО–4 / Т. К. Винцюк. – К. : Канев, 1968. – С. 21–42.
34. Левинсон С. Е. Структурные методы автоматического распознавания речи / С. Е. Левинсон // ТИИЭР, 1985. – №11. – С. 100–129.
35. Лингвистический энциклопедический словарь. – Москва : Советская энциклопедия, 1990. – 685 с.
36. Льюис Ф. Теоретические основы проектирования компиляторов / Ф. Льюис, Д. Розенкранц, Р. Стирнз – М. : Мир, 1979. – 654 с.
37. Маркел Д. Д. Линейное предсказание речи / Д. Д. Маркел, А. Х. Грей. – М. : Связь, 1980. – 380 с.
38. Методи підвищення дикторонезалежності опису і розпізнавання мовної інформації в мережі INTERNET: Третя міжнародна конференція “Інтернет-Освіта-Наука-2002” / М. М. Биков, Т. В. Гришук. – Вінниця : УНІВЕРСУМ – Вінниця, 2002. – Т. 2. – С. 329–332.
39. Методы автоматического распознавание речи : пер. с англ. / [под. ред. У. Ли]. – М. : Мир, 1983. – 716 с.
40. Оброблення сигналів і зображень та розпізнавання образів : Третя Всеукраїнська міжнародна конференція 26–30 листопада 1996 року, Україна, Київ : УкрОБРАЗ'96. – К. : НАН України, Ін-т кібернетики, 1996. – 291 с.
41. Определение динамических состояний для скрытых моделей Маркова в задачах распознавания речи: сб. тр. X сессии Российского акустического общества / А. В. Аграновский, Д. А. Леонов, Б. А. Телеснин. – М. : ГЕОС, 2000. – Т. 2. – С. 289–291.
42. Основи розпізнавання мови [Електронний ресурс] – Режим доступу : <http://www.spracherkennung.de>.
43. Острем К. Введение в стохастическую теорию управления : пер. с англ. / К. Острем – М. : Мир, 1973. – 322 с.
44. Параллельная обработка информации / [ред. А.Н. Свенсона]. – Киев : Наук. думка, 1985. – (Параллельные методы и средства распознавания образов). Т. 2. – 1985. – 280 с.

45. Природа обработки естественного языка [Электронный ресурс] / А. А. Попов – Режим доступа : <http://prof9.narod.ru/library/lib005/doc067.html>.
46. Проблемы сегментации речевого сообщения при построении систем автоматического распознавания речи [Электронный ресурс] / А. И. Шевченко – Режим доступа: <http://www.dialog-21.ru/Archive/2000/Dialogue%202000-2/370.htm>.
47. Рабинер Л. Р. Цифровая обработка речевых сигналов : пер. с англ. / Л. Р. Рабинер, Р. В. Шафер. – М. : Радио и связь, 1981. – 496 с.
48. Распознавание, классификация, прогноз. / Отв. ред. Ю. И. Журавлев // Математические методы и их применение. Вып. 4. – М. : Наука, 1992. – 224 с.
49. Рейурд-Смит Дж. Теория формальных языков. Вводный курс : пер. с англ. / Дж. Рейурд-Смит – М. : Радио и связь, 1988. – 129 с.
50. Розпізнавання мови для IPAQ [Електронний ресурс] // Офіційний сайт компанії Fonix. – Режим доступу : www.fonix.com.
51. Розпізнавання мови для IPAQ [Електронний ресурс] // Офіційний сайт компанії HandHeld Speech. – Режим доступу : www.handheldspeech.com.
52. Розпізнавання мови для IPAQ [Електронний ресурс] // Офіційний сайт компанії Microsoft Software. – Режим доступу : www.windowsmobile.com.
53. Розпізнавання мови для IPAQ [Електронний ресурс] // Офіційний сайт компанії ScanSoft. – Режим доступу : www.lhsl.com/PDSay.
54. Розпізнавання мовних образів з використанням нейромережевого підходу : Праці міжнародної конференції з індуктивного моделювання (МКІМ – 2002) / М. М. Биков, Т. В. Гришук. – Львів : Державний НДІ інформаційної інфраструктури, 2002. – Т. 1., Ч. 2. – С. 203–207.
55. Розробка трансляторів [Електронний ресурс] : конспект лекцій – Режим доступу : <http://softcraft.ru/translat/lect/t02-01.shtml>.
56. Ротштейн А. П. Интеллектуальные технологии идентификации : нечеткие множества, генетические алгоритмы, нейронные сети / А. П. Ротштейн. – Винница : УНИВЕРСУМ-Вінниця, 1999. – 320 с.
57. Сорокин В.Н. Сегментация и распознавание гласных / В. Н. Сорокин, А. И. Цыплихин // Информационные процессы. – 2002. – Т. 4, № 2. – С. 202–220.
58. Стандарт запису ABNF. W3C Recommendation 16 March 2004 [Електронний ресурс] – Режим доступу : <http://www.w3.org/TR/2004/REC-voicexml20-20040316>.
59. Українська асоціація з оброблення інформації та розпізнавання образів. Офіційний сайт [Електронний ресурс] – Режим доступу: <http://www.uasoiro.org.ua>.
60. Фант Г. Акустическая теория речеобразования / Г. Фант – М. : Наука, 1964. – 284с.

61. Фланаган Дж. Анализ, синтез и восприятие речи: пер. с англ. / Дж. Фланаган – М. : Связь, 1968. – 396 с.
62. Харкевич А. А. Спектры и анализ / А. А. Харкевич – М. : Физматиз, 1962. – 320 с.
63. Хомский Н. О некоторых формальных свойствах грамматики / Н. Хомский – Кибернетический сборник, вып.5. – М. : ИЛ, 1962. – С. 279–311.
64. Хомский Н. Синтаксические структуры / Н. Хомский // Новое в лингвистике – М. : ИЛ, 1962. – С. 412–527.
65. Хомский Н. Три модели для описания языка / Н. Хомский // Кибернетический сборник – М. : ИЛ, 1961. – С. 237–266.
66. Шлезингер М. И. Десять лекций по статистическому и структурному распознаванию / М. И. Шлезингер, В. Главач. – Киев : Наук. думка, 2004, 546 с. – ISBN 966-00-0341-2.
67. Шмелёва А. Речевые технологии: по следам CeBIT / А. Шмелёва // “PC WEEK”. – 2000. – № 17.
68. Albesano D. Word recognition with neural network / D. Albesano, R. Gemello, F. Mana // CSELT Techn. Repts. – 1992. – № 6. – P. 553–559.
69. Alleva P. Improvements on the Pronunciation Prefix Tree Search Organization / P. Alleva, X. D. Huang, M.-Y. Hwang // Proc. 1996 Int. Conf. on Acoustics, Speech and Signal Processing. – Atlanta: GA, May 1996. – V. 1. – P. 133–136.
70. Auditory physics. Physical principles in hearing theory / E. De Boer // Physics Reports. – 1980. – V.1, № 10. – P. 88–174.
71. Bahl L. Decoding for Channels with Insertions, Deletions and Substitutions with Applications to Speech Recognition / L. Bahl, F. Jelinek // IEEE Trans. on Information Theory. – 1975. – V.21, №4. – P. 404–411.
72. Bahl L. Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate / L. Bahl, J. Cocke, F. Jelinek, J. Raviv // IEEE Trans. on Information Theory. – 1974. – V.20. – P. 284–287.
73. Bahl L. R. A maximum likelihood approach to continuous speech recognition / L. R. Bahl, F. Jelinek, R. L. Mercer // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1983. – V. 5(2). – P. 179–190.
74. Bahl L. Speech Recognition with Continuous Parameter Hidden Markov Models / L. Bahl, L. Brown, P. De Souza, R. Mercedes // Computer Speech and Language. – 1987. – V. 2. – P. 219–234.
75. Barron A. Statistical Properties of Artificial Neural Networks. in Proc. / A. Barron // IEEE Conf. on Decision and Control. – Tampa : FL, 1989. – P. 45–49.
76. Best Software Awards 2004 [Электронный ресурс] / Pocket PC Magazine – Режим доступа : www.PocketPCmag.com.
77. Bocchieri E. Vector Quantization for the Efficient Computation of Continuous Density Likelihoods / E. Bocchieri // Proc. 1993 Int. Conf. on Acoustics, Speech and Signal Processing. – Minneapolis : MN. – April 1993. – V. 2. – P. 692–695.

78. Bogue R. Speech Recognition for the Pocket PC [Электронный ресурс] / Bogue R., Pocket PC Magazine, May 2002. – Режим доступа : www.PocketPCmag.com.
79. Bridle J. An Efficient Elastic Template Method for Detecting Given Words in Running Speech / J. Bridle // Proc. British Acoustical Society Meeting. – P. Paper 73SHC3. – 1973. – P. 224–228.
80. Brown P. The Acoustic-Modeling Problem in Automatic Speech Recognition / Brown P. – Pittsburgh, Pennsylvania : Carnegie Mellon University, 1987. – 223 p.
81. Fritsch J. ACID/HNN: A Framework for Hierarchical Connectionist Acoustic Modeling / J. Fritsch //in S. Furui, B.-H. Juang, W. Chou (eds.). – Proc. 1997 IEEE Workshop on Automatic Speech Recognition and Understanding. – Santa Barbara, CA. – December 1997. – P. 164–171.
82. Glassman M. Hierarchical DP for Word Recognition / M. Glassman // Proc. Int. Conf. on Acoustics, Speech and Signal Processing. – Florida: Tampa, 1985. – P. 886–889.
83. Glave R. D. The David speech recognition system / R. D. Glave, Giet Vander. – Proc. IEEE Int. Conf. ASSP. – Tulsa, 1978. – P. 429–432.
84. Haeb-Umbach R. Improvements in Beam Search for 10000-Word Continuous-Speech Recognition / R. Haeb-Umbach, H. Ney // IEEE Transactions on Speech and Audio Processing. – April, 1994. – V. 2, № 2. – P. 29–40.
85. Huang X. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. Chapter 12 / Huang X., Acero A, and Hon H.– Prentice Hall PTR, 2001. – 210 p.
86. Irie B. Capabilities of Three-Layered Perceptrons / B. Irie, S. Miyake // Proc. IJCNN. – San Diego : CA,1988. – P. 641–648.
87. Itakura F. Minimum Prediction Residual Principle Applied to Speech Recognition / F. Itakura // IEEE Trans. on Acoustics, Speech and Signal Processing. – 1975. – V.23. – P. 67–72.
88. Jelinek F. A Fast Sequential Decoding Algorithm Using a Stack / F. Jelinek // IBM Journal of Research and Development. – November, 1969. – V. 13. – P. 675–685.
89. Kitazume Y. LSI Implementation of a Pattern Matching Algorithm for Speech Recognition / Y. Kitazume, E. Ohira, T. Endo // IEEE Trans. on Acoustics, Speech and Signal Processing. – 1985. – V.33, № 1. – P. 1–4.
90. Microsoft Speech SDK 3.0. Beta version. Documentation [Электронный ресурс] – Режим доступа: <http://www.microsoft.com/speech/speech2007/>
91. Murveit H. 1000-Word Speaker-Independent Continuous-Speech Recognition Using Hidden Markov Modells / H. Murveit, M. Weintraub // Int. Conf. On Acoustics, Speech and Signal Processing. – New York, 1988. – P. 115–118.
92. Myers C. Performance Tradeos in Dynamic Time Warping Algorithms for Isolated Word Recognition / C. Myers, L. Rabiner, A. Rosenberg // IEEE Trans. on Acoustics, Speech, and Signal Processing. – 1980. – V.28, № 6. – P. 623–635.

93. Nene S. A. Closest Point Search in High Dimensions / S. A. Nene, S. K. Nayar // Proc. IEEE Conf. on Computer Vision and Pattern Recognition. – San Francisco, CA, June 1996. – P. 859–865.
94. Ney H. A Data-Driven Organization of the Dynamic Programming Beam Search for Continuous Speech Recognition / H. Ney, D. Mergel, A. Noll, A. Paeseler // Proc. 1987 Int. Conf. on Acoustics, Speech and Signal Processing. – Dallas: TX, April 1987. – P. 833–836.
95. Ney H. Improvements in Beam Search for 10000-Word Continuous Speech Recognition / H. Ney, R. Haeb-Umbach, B.-H. Tran, M. Oerder // Proc. 1992 Int. Conf. On Acoustics, Speech and Signal Processing. – San Francisco : CA, March 1992. – V. 1. – P. 9–12.
96. Ney H. Search Strategies for Large-Vocabulary Continuous-Speech Recognition / H. Ney // NATO Advanced Studies Institute, Bubion, Spain, June/July 1993, pp. 210-225, in : A. J. Rubio Ayuso, J. M. Lopez Soler (eds.), Speech Recognition and Coding. – Springer, Berlin : New Advances and Trends, 1995. – P. 120–131.
97. Ney H. The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition / H. Ney // IEEE Trans. on Acoustics, Speech and Signal Processing. – April 1984. – V. ASSP – 32, № 2. – P. 263–271.
98. Ney H. Word Graph Algorithm for Large Vocabulary Continuous Speech Recognition / H. Ney, X. Aubert // Int. Conf. on Spoken Language Processing. – Yokohama, September 1994. – V. 3. – P. 1355–1358.
99. Ney H. Word Graphs: An Efficient Interface Between Continuous Speech Recognition and Language Understanding / H. Ney, M. Oerder // Proc. 1993 Int. Conf. on Acoustics, Speech and Signal Processing. – Minneapolis, MN, April 1993. – V. 2. – P. 119–122.
100. Ortmanns S. An Experimental Study of the Search Space for 2000-Word Speech Recognition / S. Ortmanns, H. Ney // Proc. 1995 Europ. Conf. on Speech Communication and Technology. – Madrid, September 1995. – V. 2. – P. 901–904.
101. Ortmanns S. Language-Model Look-Ahead for Large Vocabulary Speech Recognition / S. Ortmanns, H. Ney, A. Eiden // Proc. 1996 Int. Conf. on Spoken Language Processing. – Philadelphia : PA, October 1996. – P. 2095–2098.
102. Ortmanns S. Look-Ahead Techniques for Improved Beam Search / S. Ortmanns, H. Ney, A. Eiden, N. Coenen // Proc. CRIM-FORWISS Workshop. – Montreal, October 1996. – P. 10–22.
103. Ortmanns S. Word Graph Algorithm for Large Vocabulary Continuous Speech Recognition / S. Ortmanns, H. Ney, X. A. Aubert // Computer Speech and Language. – January 1997. – V. 11, № 1. – P. 43–72.
104. Preparata F. P. Difference-Preserving Codes / F. P. Preparata, J. Nivergelt – IEE Trans. Information Theory, IT-20, 1974. – P. 643–649.
105. Rabiner L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition / L. Rabiner // Proceedings of the IEEE. – 1989. – V.77, № 2. – P. 257–285.

106. Rosenberg A. Effect of Glottal Pulse Shape on the Quality of Natural Vowels / A. Rosenberg // Journal Acoust. – Soc. Amer. – 1971. – V.49, № 2. – P. 583–590.
107. Schwartz R. A Comparison of Several Approximate Algorithms for Finding Multiple (N-Best) Sentence Hypotheses / R. Schwartz, S. Austin // Proc. 1991 Int. Conf. on Acoustics, Speech and Signal Processing. –Toronto, May 1991. – P. 701–704
108. Schwartz R. The N-Best Algorithm: An Efficient and Exact Procedure for Finding the N most likely Sentence Hypotheses / R. Schwartz, Y. L. Chow // Proc. 1990 Int. Conf. on Acoustics, Speech and Signal Processing. – Albuquerque : NM, April 1990. – P. 81–84.
109. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus [Электронный ресурс] – Режим доступа: http://www.ldc.upenn.edu/Catalog/readme_files/timit.readme.html.
110. The HTK Book [Электронный ресурс]: S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valthev, P. Woodland – Режим доступа: <http://htk.eng.cam.ac.uk/>.
111. Vintsyuk T. K. Elementwise Recognition of Continuous Speech Composed of Words from a Specified Dictionary / T. K. Vintsyuk // Cybernetics. – March/April 1971. – V. 7. – P. 133–143.
112. Weste N. Dynamic Time Warp Pattern Matching Using an Integrated Multiprocessing Array / N. Weste, D. Burr, B. Ackland // IEEE Trans. on Computers. – 1983. – V.32. – P. 413–417.

Наукове видання

**Биков Микола Максимович
Грищук Тетяна Вікторівна**

**МОДЕЛЮВАННЯ ПРОЦЕСУ
АНАЛІЗУ І КЛАСИФІКАЦІЇ
ГОЛОСОВИХ КОМАНД**

Монографія

Редактор С. Малішевська

Оригінал-макет підготовлено Т. Грищук

Підписано до друку 15.09.2009 р.
Формат 29,7×42¼. Папір офсетний.
Гарнітура Times New Roman.
Друк різнографічний. Ум. друк. арк. 7,39.
Наклад 100 прим. Зам № 2009-157.

Вінницький національний технічний університет,
комп'ютерний інформаційно-видавничий центр,
21021, м. Вінниця, Хмельницьке шосе, 95,
ВНТУ, ГНК, к. 114
Тел. (0432) 59-85-32
Свідоцтво суб'єкта видавничої справи
серія ДК № 3516 від 01.07.2009 р.

Віддруковано у Вінницькому національному технічному університеті,
в комп'ютерному інформаційно-видавничому центрі,
21021, м. Вінниця, Хмельницьке шосе, 95,
ВНТУ, ГНК, к. 114
Тел. (0432) 59-81-59
Свідоцтво суб'єкта видавничої справи
серія ДК № 3516 від 01.07.2009 р.