

вона може бути використана для розв'язання важливих для технології гідротермічної обробки задач визначення релаксаційно-деформівних і тепломасообмінних процесів всередині та на межі неруйнівної області деформування гігроскопічних матеріалів.

1. Подстригач Я.С., Павлина В.С. Диффузионные процессы в упруговязком деформируемом теле // Прикл. механика. – 1974. – 11, №5. – С.47–53. 2. Дей У.А. Термодинамика простых сред с памятью. – М.: Мир, 1974. – 191 с. 3. Никитенко Н.И. Сопряжение и обратные задачи тепломассопереноса. – К.: Наук. думка, 1988. – 240с. 4. Акулич И.В. Милитцер К.Е. Моделирование неизотермического влагопереноса и напряжений в древесине при сушке // Инженерно-физический журнал. – 1998. – №3. – С. 404–414. 5. Лыков А.В. Теория сушки. – М.: Энергия, 1968. 6. Луцык Р.В. Уравнения теории сушки деформируемых твердых изотропных тел // Промышл. теплотехника. – 1985. – 7, №6. – С. 8–20. 7. Соколовский Я.И. Взаимосвязь деформационно-релаксационных и тепломассообменных процессов при сушке капиллярно-пористых тел // Прикл. механика. – 1998. – 34. – №7. – С. 101–107. 8. Корн Г.К., Корн Т.К. Справочник по математике. – М., 1977. – 831 с. 9. Победря Б.Е. Термодинамический критерий прочности композитов // Механика композитных материалов. – 1993. – Т. 29, №3. – С.302–310. 10. Бленд Д. Теория линейной вязкоупругости (перевод с англ.). – М.: Мир, 1965. – 156 с. 11. Хухрянский П.Н. Прочность древесины. – М.: Гослесбуиздат, 1955. – 152 с.

УДК 534.4:621.391

Г. Зірнєєва

Дніпропетровський національний університет

ПОРІВНЯЛЬНИЙ АНАЛІЗ АЛГОРИТМІВ ПРИЙНЯТТЯ РІШЕНЬ В ЗАДАЧІ РОЗПІЗНАВАННЯ МОВИ

© Зірнєєва Г., 2007

Побудовані алгоритми прийняття рішень в задачі розпізнавання мови ґрунтовані на методах локальних екстремумів, за принципом максимуму Понтрягіна та на основі нейронних мереж. Створена програмна реалізація цих алгоритмів та проведено порівняльний аналіз створених алгоритмів між собою та із стандартним методом динамічного програмування.

The algorithms of making decision in the task of speech recognition based on methods of local extreme, a principle of Pontryagin's maximum are constructed and on the basis of neural networks. Program's realization of these algorithms is created and analysis of the created algorithms is lead among themselves and with a standard method of dynamic programming.

Постановка проблеми. Проблема прийняття рішення в задачі розпізнавання мови є сьогодні досить актуальною. Алгоритми, на основі яких приймаються рішення при побудові пристроїв розпізнавання мови, є або досить повільними за швидкістю, або не дають високої якості розпізнавання. У зв'язку з цим постає задача розроблення алгоритмів прийняття рішення в задачі розпізнавання мови, які б забезпечували необхідну якість розпізнавання, а час виконання цих алгоритмів був меншим за існуючі на сьогодні, провести порівняльний аналіз розроблених алгоритмів з вже існуючими.

Аналіз останніх досліджень та публікацій. Вперше прийняття рішення в задачі розпізнавання мови на основі методу динамічного програмування було запропоновано J. S. Bridle. Використання алгоритму локальних екстремумів та алгоритму на основі принципу максимуму Понтрягіна при прийнятті рішень в задачі розпізнавання мови вперше запропонував О.М. Карпов. Уперше теорія нейронних мереж як новий напрямок була позначена в роботі Маккаллока і Піттса. Великий внесок у розпізнаванні мовного сигналу за допомогою нейронних мереж зробили такі учені: П.Е. Овчинников, Ю.А. Семін, О.И. Федяєв, Д. Разуміхін, О.М. Карпов, П.А. Лалетін та ін.

Цілі статті. Розробити такі алгоритми прийняття рішення :

- Алгоритм на основі пошуку локальних екстремумів;
- Алгоритм на основі принципу максимуму Понтрягіна;
- Алгоритм на основі нейронних мереж.

Створити програмну реалізацію цих методів та порівняти розроблені алгоритми між собою та стандартним алгоритмом прийняття рішення в задачі розпізнавання мови – методом динамічного програмування.

Основний матеріал. Метод на основі локальних екстремумів. Розглянемо алгоритм розпізнавання мовного сигналу на основі локальних екстремумів.

Будуємо матрицю відстаней. Нам треба з точки А потрапити в В. З кожної точки можемо вирушати в двох напрямках. Переберемо всі точки так:

1. Рух відбувається за діагоналями. Для кожної точки перевіряються всі можливі шляхи.
2. Для кожної точки зберігаються:

- її значення;
- всі можливі шляхи до неї;
- для кожного шляху середнє арифметичне всіх точок, що йому належать.

Так, нехай одним з шляхів до точки d_{24} є $d_{00} \rightarrow d_{01} \rightarrow d_{11} \rightarrow d_{12} \rightarrow d_{22} \rightarrow d_{23} \rightarrow d_{24}$.

Тоді для цього шляху зберігаємо її значення, d_{24} ; шлях до неї, починаючи з d_{00} , $d_{00} \rightarrow d_{01} \rightarrow d_{11} \rightarrow d_{12} \rightarrow d_{22} \rightarrow d_{23} \rightarrow d_{24}$; середнє арифметичне за всіма точками на шляху до неї, починаючи з d_{00} та закінчуючи самою точкою d_{24} :

$$\bar{d}_{24} = \frac{d_{00} + d_{01} + d_{11} + d_{12} + d_{22} + d_{23} + d_{24}}{7}.$$

3. На кожному шляху перевіряється, чи є середнє арифметичне за кожним із шляхів локальним екстремумом.

4. Якщо отримано локальний екстремум, для кожної точки обнулюються шляхи, що вели до неї, та середнє арифметичне для них; відкидаються усі точки зліва та згори від знайденого екстремуму; повернення на пункт 1. Загальний шлях дорівнюватиме сумі усіх локальних екстремумів та середнього значення, отриманого в останній точці.

Метод на основі принципу максимуму Понтрягіна

Нехай Y_i , $i=1,m$ – параметри мовного сигналу еталонної послідовності, X_j , $j=1,k$ – параметри мовного сигналу вхідної послідовності. a_{ij} – схожість i -ї ознаки еталонної і j -ї вхідної. Динаміка переходу від однієї чарунки до іншої описується:

$$x_1(n+1) = x_1(n) + u_1(n) \quad x_2(n+1) = x_2(n) + u_2(n)$$

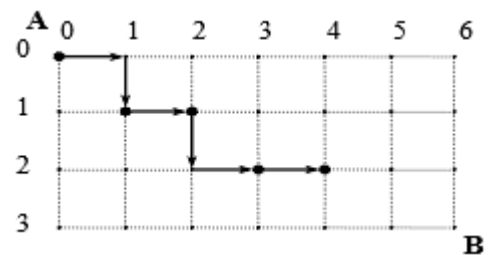
x_1 – описує положення у рядку, x_2 – у стовпці.

Відомі початковий та кінцевий стан системи та інтервал керування $n = \overline{0, N}$.

$$x_1(0) = 0, x_1(N) = k-1 \quad u_1(n) = 0 \vee 1$$

$$x_2(0) = 0, x_2(N) = m-1 \quad u_2(n) = 0 \vee 1$$

$$N \leq k + m \quad u_1(n) + u_2(n) \geq 1$$



Побудувати таке керування u_1, u_2 , щоб на інтервалі керування $n = \overline{0, N}$ виконувалась умова мінімуму цільової функції.

$$J = \sum_{n=0}^N \alpha(x_1(n), x_2(n)) \rightarrow \min$$

Вираз для Гамільтоніана

$$H = p_1(x_1 + u_1) + p_2(x_2 + u_2) - \alpha(x_1, x_2)$$

Різницеві рівняння, за якими підраховуються функції p_i , визначаються з виразу:

$$p_i(n) = \left. \frac{\partial H}{\partial x_i} \right|_{p=p(n+1)}, i = \overline{1, 2}$$

У результаті обчислень отримаємо:

$$x_1(n+1) = x_1(n) + \frac{p_1(n+1) - x_1(n) \cdot \Delta_1(n) - \Delta_1(n)}{2\Delta_1(n)};$$

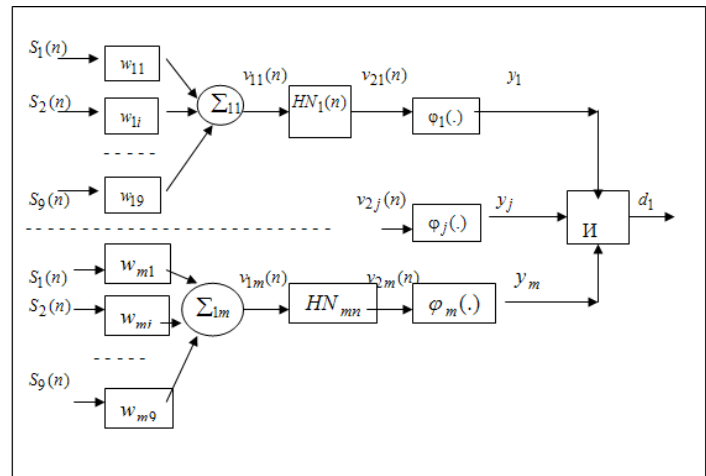
$$x_2(n+1) = x_2(n) + \frac{p_2(n+1) - x_2(n) \cdot \Delta_2(n) - \Delta_2(n)}{2\Delta_2(n)}$$

Рівняння для p_i описують динамічну систему від кінця до початку, а рівняння x_1 – від початку до кінця. Приведемо їх до однакового вигляду від кінця до початку. У цьому випадку вибір напрямку ні на що не впливає, тому що граничні умови змінних x_i визначені і в початку, і в кінці.

$$\begin{aligned} x_1(n) &= 2x_1(n+1) - \frac{p_1(n+1)}{\Delta_1(n)} + 1 & x_2(n) &= 2x_2(n+1) - \frac{p_2(n+1)}{\Delta_2(n)} + 1 \\ p_1(n) &= \frac{p_1(n+1)}{2} + \frac{\Delta_1(n)}{2}(x_1(n) + \Delta_1(n)) & p_2(n) &= \frac{p_2(n+1)}{2} + \frac{\Delta_2(n)}{2}(x_2(n) + \Delta_2(n)) \\ x_1(N) &= k & x_2(N) &= m \end{aligned}$$

Метод на основі нейронних мереж

Припустімо, для мовного сигналу $s(t)$ знайдене спектрально-часове представлення $S(\omega_k, T_l)$, де ω_k – частота k -ї гармоніки перетворення Фур'є, T_l – l -й інтервал аналізу при $T_l = 12,5$ Мсек, $k = \overline{1, 128}$, $l = \overline{1, L}$, $L = T_0 / T_l$, T_0 – тривалість мовного висловлення. В часі функція представлена деякою сукупністю сегментів, кожен сегмент – сукупність спектральних зрізів. Кожен спектральний зріз – вектор спектральних параметрів [1]. Будується нейронна мережа для 9 спектрально-смужових параметрів. Кожен спектральний зріз є варіантом образу, та нейрон повинен навчатися на сукупності спектральних зрізів, які на інтервалі сегмента утворюють образ.



Нейронна мережа повинна бути багатошаровою. Перший шар – налаштування на спектральний зріз з вихідним значенням $v_{1j}(n)$, де n – номер спектрального зрізу, j – номер сегмента (фонем) в слові. Другий шар – послідовність $v_{1j}(n)$ на вході сегмента j . Вихід другого шару – $v_{2j}(j^*n, j^*n - p_j)$, де $(j^*n, j^*n - p_j)$ відповідає перебору значень $v_{1j}(n)$ для кожного сегмента $j = \overline{1, m}$, m – кількість сегментів, HN_{mm} – FIR-фільтр.

Розглянемо приклад роботи наведених алгоритмів. Нехай на вході слово “п’ять”. Проведемо первісний аналіз мовного сигналу та апроксимацію на основі функції: друга похідна функції Гаусса.

Початковий сигнал наведено на рис.1. Тривимірне зображення початкового сигналу наведено на рис.3, результат апроксимації – на рис.4. На рис.2. наведені лінії рівня після апроксимації мовного сигналу. Вилучення складових мовного сигналу відбувається за схемою SET – послідовного вилучення складових. Докладніше цей метод розглянуто в роботі [1,3].

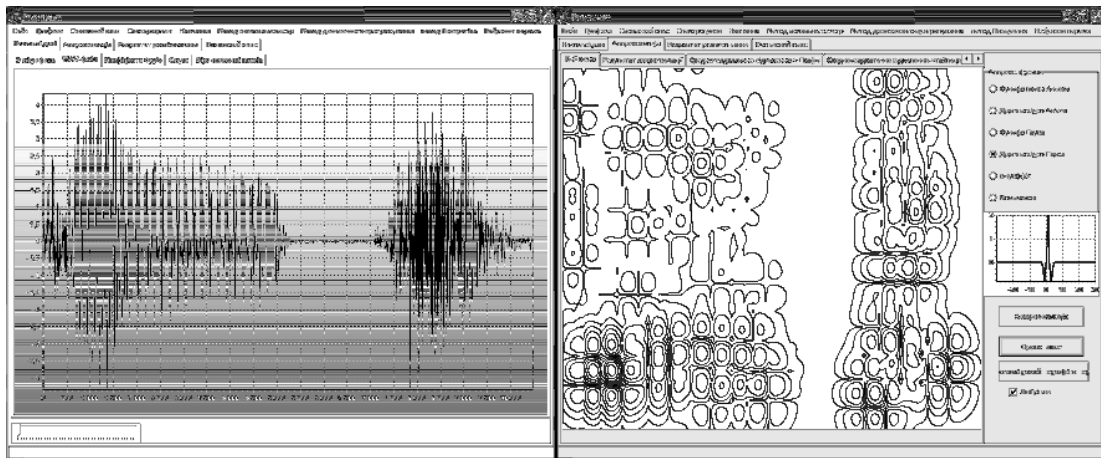


Рис. 1. Початковий сигнал. Слово “н’ять”

Рис.2. Лінії рівня

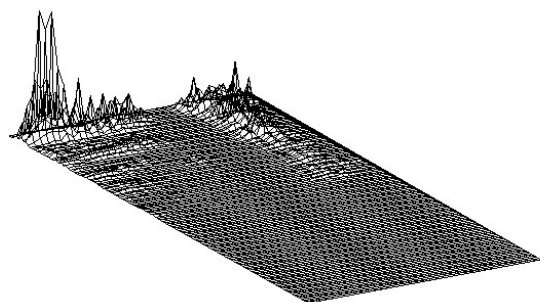
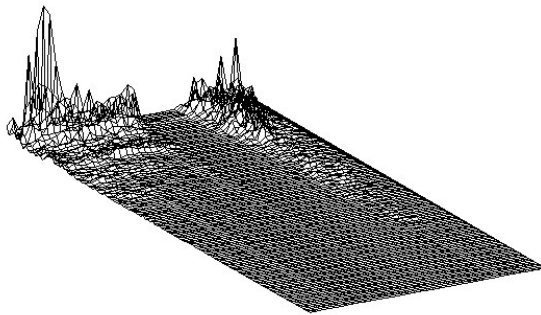


Рис. 3. Тривимірне зображення. Слово “н’ять”

Рис. 4. Результат апроксимації

Program

Метод локального екстремуму: Метод динамічного програмування: Принцип максимуму Понтріяка: Нейронна мережка

Кількість ітерацій: 2410 Кількість ітерацій: 2512 Кількість ітерацій: 40 Кількість ітерацій: 135

Результат розпізнавання

Відстані між словами

Слово	Відстання	Слово	Відстання	Слово	Відстання	Слово	Відстання
нуль	0,244685785387	нуль	0,151638112611	нуль	0,1623900804210	нуль	0,017453241626047
єдин	0,201742010405	єдин	0,1623904900308	єдин	0,138240364597	єдин	0,0194949437323159
два	0,229070550267	два	0,169777977101	два	0,153742635230	два	0,02725759786263
три	0,211301508105	три	0,141424179324	три	0,142365039606	три	0,017723060573673
чотири	0,218227839048	чотири	0,137577128739	чотири	0,123626646816	чотири	0,025487845388898
п’ять	0,126362270662	п’ять	0,008975046716	п’ять	0,116502475323	п’ять	0,006271991902132
шість	0,215954713959	шість	0,113348172771	шість	0,139968042964	шість	0,02906497235046
сімь	0,208126640198	сімь	0,131393178645	сімь	0,152500540311	сімь	0,021970125986165
вісьмь	0,207852020482	вісьмь	0,138528868723	вісьмь	0,148373036735	вісьмь	0,019240994110082
дев’ять	0,221547170235	дев’ять	0,140072911896	дев’ять	0,154993309042	дев’ять	0,025044868653306

Рис. 5. Результат роботи методів

Висновки. Проведемо порівняльний аналіз наведених методів.

Порівняльний аналіз методів

Метод	Складність методу	Якість розпізнавання
Метод динамічного програмування	$n*m$	95–98 %
Метод на основі локальних екстремумів	$n*m-k$	90 %
Метод на основі принципу максимуму Понтрягіна	$(n+1)$ або $(m+1)$	95 %
Метод на основі нейронних мереж	$P*F*S$	70 %

де n, m – розміри вхідних послідовностей, $k \approx 3*(m+n)$, P – кількість смуг (дорівнює 9), F – кількість фонем у слові, S – кількість часток у фонемі (дорівнює 5).

Проаналізувавши ці дані, доходимо висновку, що алгоритм на основі принципу максимуму Понтрягіна дає високу якість розпізнавання, але швидкість роботи алгоритму є нижчою ніж за інших методів.

1. Карпов О.Н., Габович А.Г., Марченко Б.Г., Хорошко В.А., Щербак Л.Н. Компьютерные технологии распознавания речевых сигналов: Монография. – К.: ООО “Полиграф-Консалтинг”, 2005. – 138 с. 2. Хайкин Саймон. Нейронные сети: Полный курс, 2-е изд.: Пер. с англ. – М.: Издательский дом “Вильямс”, 2006. – 1104 с.: ил. 3. Карпов О.Н. Технология построения устройств распознавания речи: Монография. – Д.: Изд-во ДНУ, 2001. – 190 с. 4. Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф. Математическая теория оптимальных процессов. – М.: Наука, 1976. – 392 с.