УДК 681.3

Л.В. Чирун
Національний університет "Львівська політехніка",
кафедра "Інформаційні системи та мережі"

МОДЕЛЬ АДАПТИВНОГО СИНТЕЗУ МОВИ В ЦИФРОВИХ СИГНАЛЬНИХ ПРОЦЕСОРАХ НА ОСНОВІ НЕПЕРЕРВНИХ ДРОБІВ

In this article adaptive model for speech synthesis on digital signal processor is described. Application of continued fractions in digital signal processor is proposed. Direct realization of continued fractions by multimesh structures is given. This procedure is used for the design of the zero-pole model of human vocal tract.

Описано адаптивну модель синтезу голосових сигналів у цифровому сигнальному процесорі. Запропоновано застосування неперервних дробів у цифровому сигнальному процесорі. Наведено реалізацію неперервних дробів за допомогою багатокоміркових структур. Ця процедура використовується для реалізації моделі голосового апарата людини.

ВСТУП

Сучасні системи розпізнавання голосових сигналів об'єднують в собі технології з таких галузей сучасної науки, як обробка сигналів, розпізнавання образів, природна мова, лінгвістика. Такі системи, що мають широке застосування в проблематиці обробки сигналів, зчинили справжній бум в галузі цифрової обробки сигналів *(Digital Signal Processing (DSP))*. Раніше в цій галузі домінували векторно-орієнтовані процесори і алгебраїчний математичний апарат, натомість теперішнє покоління DSP покладається на софістичні статистичні моделі і використовує комплексне програмне забезпечення для практичної реалізації. Сучасні моделі розпізнавання голосових сигналів спроможні розуміти неперервну мову на вході для словників, що складаються з сотень тисяч слів в операційних середовищах.

Лінійний предиктивний аналіз голосових сигналів історично є найбільш важливим в технологіях аналізу голосових сигналів. Основою цього є модель фільтра-джерела, що являє собою ідеальний лінійний фільтр.

Лінійне предиктивне кодування найчастіше використовується в мовному аналізі і синезі чи в передачі чи збереженні мовних сигналів. Для цієї мети зазвичай використовують ідеальні коміркові структури моделювання голосового тракту людини. Вперше ці структури з коефіцієнтами відображення були сформульовані Маркелом, Греєм [1] і Макхоулом [2]. Модель в просторі станів неідеальної коміркової структури з двома і чотирма множниками на секцію для цифрових сигнальних процесорів було проаналізовано в [3]. Загальну систему голосового синтезу, наведену в [4], зображено на рис. 1.

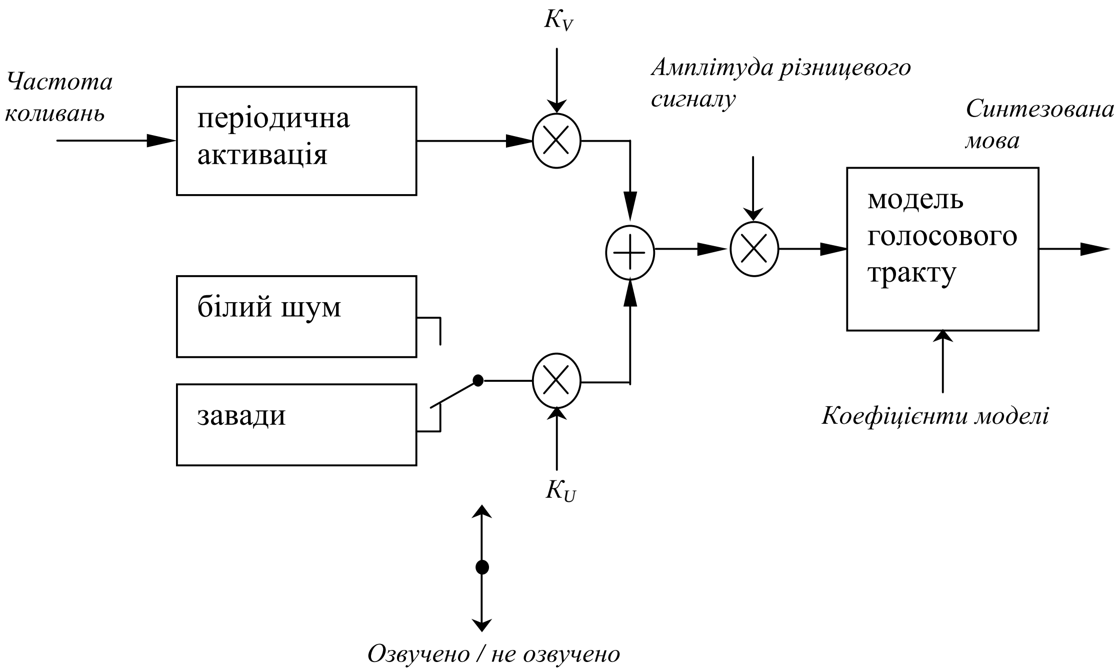

Рис. 1. Система синтезу мови

У загальному випадку задача лінійного предиктування полягає в наступному [9–12]. Нехай ми маємо голосовий сигнал $s(n)$, і нехай $\tilde{s}(n)=\sum_{k=1}^{p}\alpha_k s(n-k)$ – предиктована величина. Похибка предиктування в цьому випадку задається так:

$$e(n)=s(n)-\tilde{s}(n)=s(n)-\sum_{k=1}^{p}\alpha_k s(n-k).$$

Ми, звичайно, хочемо мінімізувати похибку для знаходження найкращих, чи оптимальних, значень $\{\alpha_k\}$. Визначимо короткочасову середню похибку:

$$E=\sum_n e^2(n)=\sum_n\left\{s(n)-\sum_{k=1}^{p}\alpha_k s(n-k)\right\}^2=\sum_n s^2(n)-\sum_n\left\{2s(n)\sum_{k=1}^{p}\alpha_k s(n-k)\right\}+$$

$$+\sum_n\left\{\sum_{k=1}^{p}\alpha_k s(n-k)\right\}^2=\sum_n s^2(n)-2\sum_{k=1}^{p}\alpha_k\sum_n s(n)s(n-k)+\sum_n\left\{\sum_{k=1}^{p}\alpha_k s(n-k)\right\}^2$$

Ми можемо мінімізувати похибку α_l для кожного $1\le l\le p$ за допомогою диференціювання E і прирівнювання результату до нуля

$$\frac{\partial E}{\partial\alpha_l}=0=-2\sum_n s(n)s(n-l)+2\sum_n\left\{\sum_{k=1}^{p}\alpha_k s(n-k)\right\}s(n-l)$$

У випадку коваріантного методу ми для початку дещо перевизначимо терміни

$$\sum_n s(n)s(n-l)=\sum_{k=1}^{p}\alpha_k\left(\sum_n s(n-k)s(n-l)\right)\quad\text{або}\quad c(l,0)=\sum_{k=1}^{p}\alpha_k c(k,l).$$

Це рівняння ще відоме як рівняння лінійного предиктування (*рівняння Юле-Волкера*). $\{\alpha_k\}$ називають коефіцієнтами лінійного предиктування, або предикторними коефіцієнтами. При обчисленні рівнянь для всіх значень l ми можемо записати їх в матричній формі

$$\bar{c} = \underline{C}\bar{\alpha},$$

де

$$\bar{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_p \end{bmatrix} \qquad \underline{C} = \begin{bmatrix} c(1,1) & c(1,2) & \dots & c(1,p) \\ c(2,1) & c(2,2) & \dots & c(2,p) \\ \dots & \dots & \dots & \dots \\ c(p,1) & c(p,2) & \dots & c(p,p) \end{bmatrix} \qquad \bar{c} = \begin{bmatrix} c(1,0) \\ c(2,0) \\ \dots \\ c(p,0) \end{bmatrix}.$$

Для розв'язання цього рівняння необхідно знайти обернену матрицю:

$$\bar{c} = \underline{C}^{-1}\bar{\alpha}.$$

Цей метод називають коваріаційним методом. Зауважимо, що коваріаційна матриця є симетричною. Швидким методом для знаходження розв'язку цього рівняння є метод Холецького (коваріаційна матриця розбивається на нижню і верхню трикутні матриці).

Використовуючи дещо інший підхід до мінімізації похибки, ми можемо знайти розв'язок рівняння лінійного предиктування, використовуючи автокореляціійний метод

$$\bar{\alpha} = \underline{R}^{-1}\bar{r},$$

де

$$\bar{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_p \end{bmatrix} \qquad \underline{R} = \begin{bmatrix} r(0) & r(1) & \dots & r(p-1) \\ r(1) & r(0) & \dots & r(p-2) \\ \dots & \dots & \dots & \dots \\ r(p-1) & r(p-2) & \dots & r(0) \end{bmatrix} \qquad \bar{r} = \begin{bmatrix} r(1) \\ r(2) \\ \dots \\ r(p) \end{bmatrix}.$$

Матриця системи є симетричною, і всі діагональні елементи однакові, що означає, що обернена матриця завжди існує і розв'язки системи знаходяться в лівій півплощині.

Авторегресивне моделювання з використанням предикту найменших квадратів, або лінійне предиктування, формує основу широкого спектру задач обробки сигналів і комунікаційних систем, що включають адаптивне фільтрування і контроль, моделювання мовних систем і систем кодування, адаптивне канальне вирівнювання, оцінку параметричного спектру і системи ідентифікації.

Для реалізації лінійного предиктування даних чи цілей моделі необхідно визначити величини коефіцієнтів лінійного предиктування, а також порядок. Деякі методи моделі вибору порядку, що часто застосовуються на практиці, містять методи інформаційного критерію, запропонованого Акайке, метод мінімуму довжини опису, запропонованого Шварцом, і прицип предикту найменших квадратів Рісанена. В оригінальній формі перші два критерії містять явний баланс між подібністю вхідних даних моделі і поняття штрафу для складності моделі. Інтуїтивно в методі інформаційного критерію основною метою є мінімізація кількості бітів, що буде необхідна для опису даних. Коли ж дані вже можливо змоделювати параметрично і потім кодувати блоками, використовують підхід виділення блоків подібних даних, а потім модель штрафується додатковою кількістю бітів, необхідною для кодування її параметрів.

Однак голосова модель, основана на косинусному перетворені Фур’є для синтезу мови має кращі властивості і застосування, меншу чутливість до ефектів квантування і в результаті дає більш природну синтезовану мову. Параметрами цієї моделі є коефіцієнти косинусного перетворення Фур’є. Дана модель базується на косинусному розкладі логарифмічного короткочасового голосового діапазону і синтез реалізований наближеним оберненим косинусним перетворенням Фур’є з використанням неперервних ланцюгових дробів. Цей підхід є параметричним і не оснований на ніяких спрощуючих припущеннях про голосову модель, тому що полюси, так як і нулі голосової моделі, є обґрунтовані.

ГОЛОСОВА МОДЕЛЬ, ОСНОВАНА НА КОСИНУСНОМУ ПЕРЕТВОРЕННІ ФУР’Є

Нехай маємо логарифмічний діапазон $\ln\left|S\left(e^{j\omega T}\right)\right|$ сегмента голосових даних $\{s(n)\}$, де T – вибірковий інтервал, $f_s = \dfrac{1}{T}$ – вибіркова частота, і ω – кутова частота. Ця функція може бути виражена з використанням дійсних коефіцієнтів косинусного перетворення Фур’є $\{c_n\}$

$$\ln\left|S\left(e^{j\omega T}\right)\right| = \sum_{n=-\infty}^{\infty} c_n e^{-jn\omega T} \tag{1}$$

Комплексні коефіцієнти косинусного перетворення Фур’є $\{g_n\}$ дискретної системи з мінімальною стабільністю фази є випадкові і можуть бути зв’язані з $\{c_n\}$ такими співвідношеннями

$$\begin{aligned} g_n &= c_n, \quad n = 0, N_F/2, \\ g_n &= 2c_n, \ 0 < n < N_F/2, \\ g_n &= 0, \quad n < 0 \end{aligned} \tag{2}$$

де N_F – розмірність застосованої FFT. Цифровий фільтр, чия логарифмічна величина відповідності апроксимує функцію $\ln\left|S\left(e^{j\omega T}\right)\right|$ визначається системою трансферних функцій

$$\tilde{S}(z) = e^{c_0} \exp \sum_{n=1}^{N_0-1} 2c_n z^{-n} = e^{c_0} \prod_{n=0}^{N_0-1} \exp\left[2c_n z^{-n}\right], \tag{3}$$

де $0 < N_0 < N_F/2$. Коефіцієнт e^{c_0} дорівнює значенню RMS моделі косинусного перетворення Фур’є для різнецевого сигналу. В наших експериментах над голосовою моделлю ми використовували $f_s = 8kHz$, $N_F = 512$, довжина голосового сегмента 25 *ms* з 12 *ms* перекриванням і $N_0 = 25$.

З (3) випливає, що система трансферних функцій $\tilde{S}(z)$ є добутком трансцендентних трансферних функцій

$$H_n(z) = e^{2c_n z^{-n}}, \quad 0 < n \le N_0 - 1. \tag{4a}$$

Відповідна імпульсна характеристика задається

$$h_n(m)=\begin{cases}\dfrac{(2c_n)^i}{i!}, & m=ni, \quad i=0,1,2,\dots \\ 0, & m\neq ni\end{cases} \tag{4b}$$

Це означає, що система трансферних функцій $\tilde{S}(z)$ має такий вигляд (рис. 2)

$$\tilde{S}(z)=e^{c_0}\prod_{n=0}^{N_0-1}H_n(z) \tag{5}$$

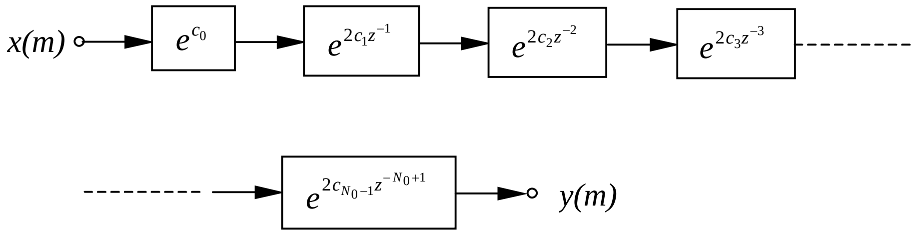

Рис. 2. Голосова модель косинусного перетворення Фур'є

Для реалізації трансферної функції $H_n(z)$ за допомогою ціфрового фільтра необхідно знайти апроксимацію $H_n(z)$, яку можливо практично реалізувати. Одним з варіантів апроксимації експоненціальної функції в (4а) є неперервні ланцюгові дроби [4]. Іншою можливістю реалізації апроксимації експоненціальної функції є використання апроксимації Паде. Тоді система трансферних функцій в практичній голосовій моделі на основі косинусного перетворення Фур'є буде мати вигляд

$$\tilde{\tilde{S}}(z)=e^{c_0}\prod_{n=0}^{N_0-1}\tilde{H}_n(z). \tag{6}$$

Апроксимація за допомогою розкладу з використанням неперервних ланцюгових дробів.

Експоненціальна функція, виражена розкладом в неперервний дріб, може набувати такого вигляду [5, 6, 7]:

$$e^x=\frac{1}{1-}\,\frac{x}{1+}\,\frac{x}{2-}\,\frac{x}{3+}\dots+\frac{x}{2-}\,\frac{x}{2s-1},\dots \tag{7}$$

де параметр $x=2c_nz^{-n}$. Точність наближення голосової моделі залежить не лише від кількості коефіцієнтів косинусного перетворення Фур'є в (3), але і від кількості членів неперервного дробу в (7), тобто від довжини неперервного дробу, котра визначається s. Скінченний ланцюговий дріб для функції e^x також може бути виражений множиною дійсних функцій, котрі апроксимують експоненціальну функцію з все зростаючою точністю

$$e^x\cong\frac{1}{1},\frac{1}{1-x},\frac{2+x}{2-x},\frac{6+2x}{6-4x+x^2},\frac{12+6x+x^2}{12-6x+x^2},\dots \tag{8}$$

Ці функції відомі як апроксимації Паде експоненціальної функції. Рекомендовано використовувати непарну кількість елементів неперервного дробу в (7). Це призводить до апроксимації експоненціальної функції раціональною функцією з однаковими степенями поліномів у чисельнику та знаменнику в (8). Цими вибраними апроксимаціями є

$$\tilde{H}_1(z) = \frac{2+x}{2-x}, \quad \tilde{H}_2(z) = \frac{12+6x+x^2}{12-6x+x^2},$$
$$\tilde{H}_3(z) = \frac{120+60x+12x^2+x^3}{120-60x+12x^2-x^3}, \tag{9}$$
$$\tilde{H}_4(z) = \frac{1680+840x+180x^2+20x^3+x^4}{1680-840x+180x^2-20x^3+x^4}$$

де z є змінна z-перетворення і $x = 2c_n z^{-n}$.

$$\nu_{2s+1}(m) = -\frac{c_n}{2s-1}\nu_{2s}(m-n),$$
$$\nu_{2s}(m) = -\frac{c_n}{2s-1}\nu_{2s-1}(m-n) + \nu_{2s+1}(m),$$
$$\vdots$$
$$\nu_5(m) = -\frac{c_n}{3}\nu_4(m-n) + \nu_6(m),$$
$$\nu_4(m) = -\frac{c_n}{3}\nu_3(m-n) + \nu_5(m),$$

$$\begin{aligned} \nu_3(m) &= -c_n\nu_2(m-n) + \nu_4(m), \\ \nu_2(m) &= 2c_n\nu_1(m-n) + \nu_3(m), \\ \nu_1(m) &= x(m) + \nu_2(m), \\ y(m) &= \nu_1(m). \end{aligned} \tag{10}$$

Граф потоку сигналу (*Signal Flow Graph* (SFG)) раціональної трансферної функції, котра апроксимує часткову нераціональну трансферну функцію в (4а), наведена на рис. 3. Граф потоку сигналу визначає різнецеві рівняння в просторі стану.

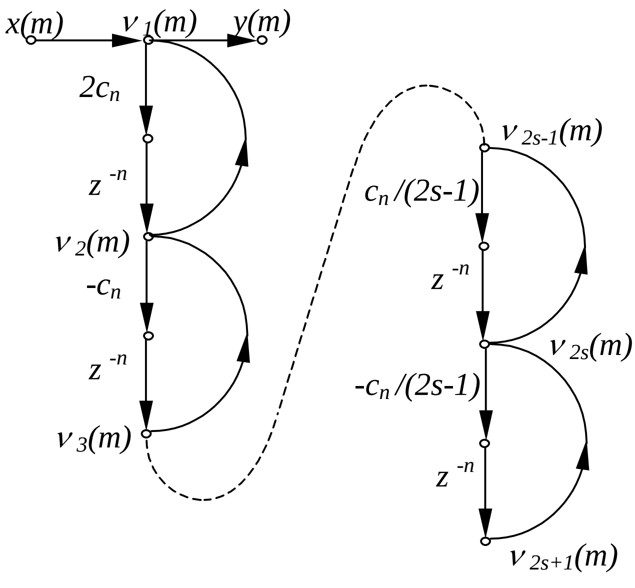

Рис. 3. Структура розкладу в неперервний дріб

СТРУКТУРА АДАПТИВНОГО СИНТЕЗУ

Як було зазначено вище, похибка апроксимації для e^x визначається кількістю елементів неперервного дробу для розкладу експоненціальної функції в неперервний дріб. Ця похибка в подальшому також залежить від величин модулів дійсних коефіцієнтів косинусного перетворення Фур'є $|c_n|$. На основі статистичного аналізу коефіцієнтів косинусного перетворення Фур'є для опису голосової моделі чоловічого гучномовця було зроблено таку оцінку відносно стабільності системи і точно визначено межу безпеки для трансферних функцій $H_n(z)$ в рівнянні (4а). З вищесказаного випливає, що функції $H_n(z)$ можна апроксимувати так:

$$n = 1, 2, 3 \quad \Rightarrow \quad \tilde{H}_3(z)$$
$$n = 4, 5 \quad \Rightarrow \quad \tilde{H}_2(z)$$
$$n = 6, 7, \ldots, 25 \quad \Rightarrow \quad \tilde{H}_1(z)$$

Більш ефективним відносно до загальної похибки апроксимації та економії кількості арифметичних операцій, що необхідні для практичної реалізації голосового моделювання, є використання адаптивної структури неперервного дробу. Кількість відповідних комірок (рис. 3) може вибиратися відповідно до величин коефіцієнтів косинусного перетворення Фур'є. Наступне адаптивне емпіричне правило може використовуватися:

для $|c_n| \le 0.3$ *дві комірки – відповідають* $\tilde{H}_1(z)$,

для $|c_n| \le 0.5$ *чотири комірки – відповідають* $\tilde{H}_2(z)$,

для $|c_n| \le 1$ *шість комірок – відповідають* $\tilde{H}_3(z)$,

для $|c_n| > 1$ *вісім комірок – відповідають* $\tilde{H}_4(z)$.

Для прикладу, голосову модель стаціонарної частини (24 *ms*) голосного звуку "е" використовують

$c_0 = -0.491$ – логарифм величини різнецевого сигналу

$c_1 = 0.700 \quad c_5 = 0.159$

$c_2 = 0.354 \quad c_6 = -0.027$

$c_3 = -0.026 \quad c_7 = -0.310$

$c_4 = 0.205 \quad |c_{8\ldots25}| < 0.3$

Використовуючи зазначене емпіричне правило, можна побудувати голосову модель косинусного перетворення Фур'є, представлену на рис. 4.

При практичній реалізації трансцендентних трансферних функцій $H_n(z)$ було отримано числові результати (див. таблицю).

Значення трансцендентної трансферної функції $H_1(z) = e^{z^{-1}}$

n	$H_n(z)$ – точні значення	$H_n(z)$ – наближені значення
1	2	3
1	1.00000000000000	0.99993896484375
2	1.00000000000000	0.99993896484375
3	0.50000000000000	0.49999648242188

Значення трансцендентної трансферної функції $H_1(z)=e^{z^{-1}}$ (*продовження*)

1	2	3
4	0.16666666666667	0.16666549414062
5	0.08333333333333	0.06092749023438
6	0.00138888888889	0.00003051757812
7	0.00019841269841	0.00030517578125
8	0.00002480158730	-0.0012207031250
9	0.00000271636432	-0.00003051757812

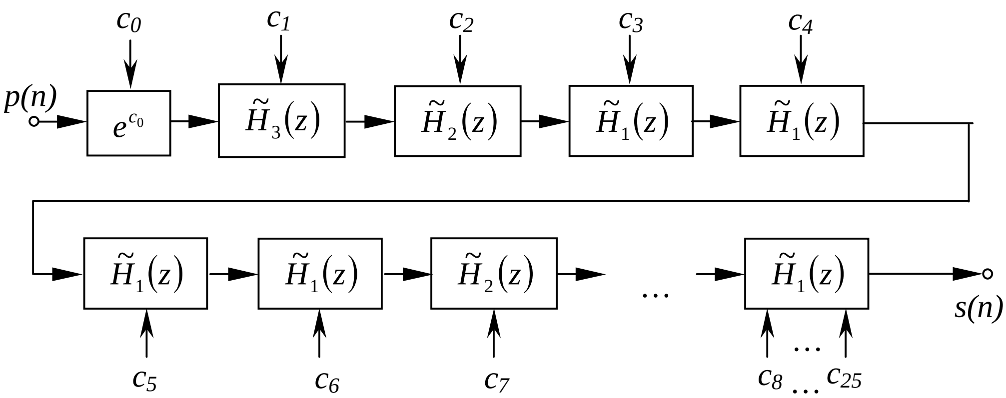

Рис. 4. Голосова модель косинусного перетворення Фур'є стаціонарної частини голосного звуку "е":
$p(n)$ *– активований сигнал;* $s(n)$ *– синтезований голосовий сигнал*

ВИСНОВКИ

Голосове моделювання на основі косинусного перетворення Фур'є насправді належить до спектрального синтезу голосових сигналів і не основане на будь-яких спрощуючих апріорних міркуваннях про систему відтворення мови. Воно також містить інформацію про діапазон активуючого голосового тракту.

Процедура є параметричною, і моделі є як формантні, так і антиформантні. Якщо використовувати синхронну висоту тону звуку, згенерована синтезована мова не відрізняється від природної мови. Прцедура голосового моделювання на основі косинусного перетворення Фур'є вимагає виконання більше арифметичних операцій, ніж підходи, основані на лінійному предиктивному кодуванні, але структуру цифрового фільтра можна оптимізувати.

Неперервні дроби пропонують цікавий інструмент не лише в синтезі мови. Апроксимація високого порядку алгебраїчних трансцендентних функцій може бути використана в системах біологічного і виробничого моделювання. Пряма реалізація неперервних дробів надалі надає можливість реалізації багатокоміркових структур.

1. Markel J.D. and Gray A.H. linear Prediction of Speech. – Berlin; Springer Verlag, 1976. 2. Makhou J. Stable and Efficient Lattice Methods for Linear Prediction // IEEE Trans.

Acoustics, Speech and signal Processing. – October 1977. – Vol. ASSP-25, No. 5. – P. 423–428. 3. Vich R. and Smekal Z. Continued Fractions in Digital Filter Synthesis, Proc. of Inter. Scient. Colloquium. – Ilmenau, Germany, 18–21 September 1995. – P. 353–356. 4. Vich R. and Smekal Z. Digital Filter Realization of Nonrational Transfer Functions, Proc. of the First European Conference on Signal Analysis and Prediction, ECSAP-97. – Prague, Czech Republic, 24–27 June 1997. – P. 179–182. 5. Шмойлов В.И. Периодические цепные дроби. – Львов: Академический Экспресс, 1998. – 219 с. 6. Шмойлов В.И., Слобода М.З. Расходящиеся непрерывные дроби. – Львов: Меркатор, 1999. – 820 с. 7. Шмойлов В.И., Чирун Л.В. Комплексные числа и непрерывные дроби. – Львов: Меркатор, 2001. – 564 с. 8. Strum R.D. and Kirk D.E. First Principles of Discrete Systems and Digital Signal Processing. – Massachusetts; Addison-Wesley Publishing Company, 1988. 9. Ваврук Є.Я., Рашкевич Ю.М. Особливості реалізації пристроїв обміну інформацією в системах цифрової обробки сигналів // Моделювання та інформаційні технології: Зб. наук. праць ІПМЕ НАНУ. – 1999. – Вип. 4. – С. 119–123. 10. Ваврук Є.Я, Рашкевич Ю.М., Цмоць І.Г. Оцінка основних характеристик процесорів управління та обробки інформації на НВІС // Вісн. Держ. ун-ту “Львівська політехніка”. – 1999. – № 386. – С. 5–11. 11. Рашкевич Ю. М. Перетворення часового масштабу мовних сигналів. – Львів: ТзОВ НВТ “Акад. Експрес”, 1997. – 140 с. 12. Ваврук Є.Я., Рашкевич Ю.М., Цмоць І.Г. Підходи до побудови та вибору елементної бази процесорів управління та обробки сигналів // Моделювання та інформаційні технології: Зб. наук. праць ІПМЕ НАНУ. – 1999. – Вип. 3. – С. 160–168.

УДК 681.3

Н.Б. Шаховська
Національний університет “Львівська політехніка”,
кафедра “Інформаційні системи та мережі”

МЕТОДИ УСУНЕННЯ НЕВИЗНАЧЕНОСТЕЙ У БАЗАХ ЗНАНЬ, ПОБУДОВАНИХ НА ОСНОВІ РЕЛЯЦІЙНОГО ПІДХОДУ

The elimination indistinct algorithmes in knowledge bases, build on relation schema, are described. The efficient gathering of sufficient statistics for classification from knowledge bases are described.

Запропоновано алгоритми усунення невизначеностей у базах знань, побудованих на основі реляційного підходу. Процес усунення невизначеностей інтерпретовано як класифікування об’єктів. Проаналізовано методи аналізу об’єктів з ієрархічною структурою.

1. ВСТУП

Як відомо, до бази даних (БД) часто формуються запити з нечітко заданими параметрами, поданими у вигляді інтервалів, лінгвістичних змінних, ступенів довіри тощо.