Обчислення складових функції евристичної оцінки станів у просторі станів задачі сегментно-складового розпізнавання мовлення

Ольга Савенкова

Дніпропетровський національний університет
2sol@ukr.net

Abstract

New model of evaluation of heuristic function for the state space of the segment-syllabic speech recognition problem is considered.

1. Вступ

Створення систем розпізнавання мовлення (СРМ) з великим словником потребує: визначення структури об'єктів розпізнавання та операцій, необхідних для розв'язання проблеми; розроблення швидкозбіжних стратегій для ефективного пошуку потенційних розв'язків, які можуть бути згенеровані цими структурами і операціями з урахуванням додаткової інформації (евристик) про досліджувану проблемну область [1-11].

Модель сегментно-складового розпізнавання формулюється таким чином [10, 12]. Нехай заданий словник $\{\mathrm{SL}_k\}$, що містить N мовленнєвих одиниць (МО). Для формування словника МО виконано автоматичний поділ сегментованої послідовності параметрів відповідних МС на акустичні склади-еталони (дво-, три-, чотирисегментні склади, яким відповідають дво- та трисимвольні МО $\mathrm{SL}_{2,k}$, $\mathrm{SL}_{3,k} \subset \mathrm{SL}_k$, $k = 1 \div NSL$) з маркуванням лінгвістичної інформації (назва МО, транскрипція). Для кожної МО обчислені послідовності параметрів або траєкторії параметрів (ТП) $\{Y_k\}$. ТП для МО SL_k сегментована на n_k ($k = 1 \div N$) сегментів-фонем. Нехай є невідомий мовленнєвий сигнал (МС), для якого обчислені послідовності параметрів X та визначені m_p сегментів-фонем, які можна об'єднати в M сполучень SL_p^x з двох, трьох або чотирьох сегментів ($p = 1 \div M$). Необхідно знайти композицію з ТП Y_k елементів словника МО, яка найкраще відповідає послідовності параметрів невідомої реалізації X.

В [10] для розв'язання задачі сегментно-складового розпізнавання було запропоновано застосування повного перебору наявних МО. Аналіз елементарних кроків зіставлення N —сегментних ($N = 2; 3; 4$) МО з ТП поданого МС дозволив визначити математичну модель побудови розв'язків-кандидатів задачі сегментно-складового розпізнавання без урахування порядку слідування МО всередині:

$$f(\eta, \lambda, \mu) = 2 \cdot \eta + 3 \cdot \lambda + 4 \cdot \mu, \qquad (1)$$

де η, λ, μ – кількість МО, що мають дво-, три- та чотирисегментні ТП відповідно.

Очевидно, що слід розглянути $N\,\mathrm{var}$ можливих розв'язків-кандидатів, знайшовши всі набори цілочисельних значень $\{\eta, \lambda, \mu\}$, при яких виконується наступна рівність $2 \cdot \eta + 3 \cdot \lambda + 4 \cdot \mu = NSGX$ за умови обмежень $\eta, \lambda, \mu < NSGX$, $\eta, \lambda, \mu \geq 0$, тобто

$$N\,\mathrm{var} = \sum_{\eta, \lambda, \mu} f(\eta, \lambda, \mu) \cdot \frac{(\eta + \lambda + \mu)!}{\eta! \cdot \lambda! \cdot \mu!}, \qquad (2)$$

де

$$f(\eta, \lambda, \mu) = \begin{cases} 1, & 2 \cdot \eta + 3 \cdot \lambda + 4 \cdot \mu = NSGX, \ \eta, \lambda, \mu \geq 0, \\ 0, & \text{в іншому випадку.} \end{cases}$$

2. Функція евристичної оцінки станів

В загальному випадку алгоритм евристичного пошуку у просторі станів для зменшення обсягу перебору в процесі знаходження розв'язку досліджуваної задачі використовує оцінну функцію (ОФ) $f(n)$ вигляду [9]

$$f(n) = g(n) + h(n), \qquad (3)$$

де в якості $g(n)$ обирають дійсну довжину шляху від початкового вузла до деякого вузла n, знайденого алгоритмом перебору до даного моменту, а оцінка $h(n)$ – евристична оцінка відстані з вузла n до цільового вузла.

Для побудови функції евристичної оцінки станів $h(n)$ розглянуто наступні характеристики сегментно-складового подання інформації про МС: вкладеність складів $\mathrm{SL}_{2,k} \in \mathrm{SL}_{3,k}$ на рівні МО, на рівні ТП; групові ознаки сегментів ТП МО (сегмент ТП МО може бути віднесений до групи Т, Ш, П); стикування МО усередині поточного розв'язку; значення середньої відстані між ТП складів, що містять різні сполучення групових ознак сегментів. Перераховані характеристики можна використовувати незалежно для обчислення $h(n)$. З урахуванням цього, пропонується обрати вигляд оцінки $h(n)$ в $f(n)$ (3) таким, що складається з декількох доданків, а саме:

1) для вузла n обчислюється значення оцінки $h_v(n)$, на основі аналізу вкладеності складів $\mathrm{SL}_{2,k} \in \mathrm{SL}_{3,k}$;

2) оцінка $h_g(n)$ обчислюється на основі аналізу варіантів сполучення групових ознак сегментів ТП МО (Т-Т, Ш-Т тощо);

3) оцінка відстані $h_d(n)$ для всіх можливих переходів з вузла n для різних сполучень групових ознак сегментів ТП МО;

4) оцінка кількості нерозкритих вузлів $h_o(n)$, що залишаються нерозглянутими в процесі пошуку у напрямку цільового вузла.

Для кожного доданку 1)-4) передбачається відповідний коефіцієнт (v, g, d, o), що має сенс ваги відповідної евристичної оцінки, яка задається експертом, та правило нормування вагових коефіцієнтів.

Таким чином, вираз для обчислення значення евристичної оцінки $h(n)$ з вузла n до цільового вузла набуває наступного вигляду

$$h(n) = \hat{h}_v(n) + \hat{h}_g(n) + \hat{h}_d(n) + \hat{h}_o(n),$$

або

$$h(n) = \frac{1}{v} h_v(n) + \frac{1}{g} h_g(n) + \frac{1}{d} h_d(n) + \frac{1}{o} h_o(n). \quad (4)$$

2.1. Алгоритм обчислення оцінки вкладеності

Нехай на n-му кроці алгоритму пошуку знайдено наступну відповідність символів-фонем МО сегментам-фонемам ТП поданої реалізації МС: $Sol_{n \to n+2} = W^*_{n,n+2}$; $Sol_{n \to n+3} = W^*_{n,n+3}$; $Sol_{n \to n+4} = W^*_{n,n+4}$. Для знайдених розв'язків виконується аналіз вкладеності: а) перевірка повної відповідності $W^*_{n,n+2} \in W^*_{n,n+3}$ (входження символів $W^*_{n,n+2}$ в $W^*_{n,n+3}$); б) перевірка повної відповідності $W^*_{n,n+2} \in W^*_{n,n+4}$ (входження символів $W^*_{n,n+2}$ в $W^*_{n,n+4}$); в) перевірка $W^*_{n,n+3} \cap W^*_{n,n+4}$ (співпадання символів на відповідних позиціях рядків). В залежності від результату аналізу вкладеності знаходится одне або більше двосимвольних сполучень $\mathrm{SL}_{2,k}$, на основі аналізу оцінок h_v, що відповідають знайденим сполученням, обирається значення кроку переходу на графі синтезу.

Відомо [8], що внаслідок характеру сполучень звуків, алофонічної варіативності та інших факторів, перед деякими приголосними виникає сегмент-пауза, можливе також утворення особливо роду подвоєння на стиках приголосних, тому на основі аналізу правил фонемного транскрибування алфавіт символів-фонем розіб'ємо на дві групи, які позначимо Γ_1 та Γ_2: до групи Γ_1 віднесемо символи-фонеми, які відповідають сегменту ТП МС; до групи Γ_2 – двом сегментам ТП МС.

Для кожної двосимвольної МО $\mathrm{SL}_{2,k}$ з навчальної множини МО обчислюється значення оцінки h_v за наступним алгоритмом.

1. Для всіх M-символьних МО $\mathrm{SL}_{M,k}$ обчислюються значення частоти появи $P_{M,k}$ у словнику предметної області W ($M = 2; 3$).

2. Обирається k—та двосимвольна МО $\mathrm{SL}_{2,k}$, яку позначимо $\mathrm{R} = \mathrm{R}_1\mathrm{R}_2$, де R_i – символи-фонеми, що відповідають складовим $\mathrm{R}_{i,k}$ МО $\mathrm{SL}_{2,k} = \mathrm{R}_{1,k}\mathrm{R}_{2,k}$.

3. Виконується пошук трисимвольних МО $\mathrm{SL}_{3,k}$ згідно шаблону $\mathrm{S} = \mathrm{R}_2\mathrm{L}_1\mathrm{L}_2$, де L_1, L_2 – будь-який голосний або приголосний. Нехай знайдено множину SL^*_3 таких МО, тоді для знайдених МО виконується упорядкування і групування за першими двома символами $\mathrm{R}_2\mathrm{L}_1$ (відповідна множина МО SL^*_2).

4. Для знайдених груп МО виконується наступне упорядкування і групування таким чином: до першої групи включаються МО з множини SL^*_3, всі символи-фонеми яких належать до групи Γ_1 у випадку, коли $\mathrm{R}_2 \in \Gamma_1$, або МО з два символи-фонеми яких належать до групи Γ_1 у випадку, коли $\mathrm{R}_2 \in \Gamma_2$; до другої групи – МО множини SL^*_3, в яких один чи більше символів-фонем належать до групи Γ_2 (K_1, K_2 – відповідна кількість МО з множини SL^*_3 у першій, другій групах; $K_{1,j}$, $K_{2,j}$ – відповідна кількість МО з множини SL^*_2, що відповідають шаблону $\mathrm{R}_2\mathrm{L}_1$ у першій, другій групах).

5. Значення евристичної оцінки вкладеності h_v для k—ї двосимвольної МО $\mathrm{SL}_{2,k}$ обчислюється як відношення зважених сум значень частот появи МО так

$$h_v = \frac{h_{1,v}}{h_{2,v}}, \quad (5)$$

$$h_{1,v} = \sum_{j=1}^{K_1} P_{2,j} \cdot \left(\sum_{m=1}^{K_{1,j}} P_{3,m} \right),$$

$$h_{2,v} = \sum_{j=1}^{K_2} P_{2,j} \cdot \left(\sum_{m=1}^{K_{2,j}} P_{3,m} \right).$$

Правило вибору кроку переходу на графі синтезу: якщо (h_{vk} >1) ∧ ($R_2 \in \Gamma_1$), то $l_n \to l_{n+2}$, інакше – $l_n \to l_{n+3}$; якщо (h_{vk} >1) ∧ ($R_2 \in \Gamma_2$), то $l_n \to l_{n+3}$, інакше – $l_n \to l_{n+4}$; якщо $(K_{1,j} = 0)$ $\wedge$ ($R_2 \in \Gamma_2$), то $l_n \to l_{n+4}$.

2.2. Алгоритм обчислення оцінки сполучень групових ознак

Для кожної МО обчислюється значення оцінки h_g , яка дозволяє обрати шлях на графі синтезу з поточного стану у напрямку цільового стану таким чином.

1. Обираються МО SL_k , що мають M — сегментні ТП ($M = 2; 3; 4$). Для обраних МО здійснюється групування за варіантом сполучення ознак ТШП-транскрипції сегментів-фонем GT . Нехай визначено NG_M таких груп. Наприклад, нехай для $M = 2$ у множині МО знайдено $NG_2 = 4$ груп: Т-Т, Ш-Ш, Т-Ш, Ш-Т, тоді кожна група позначиметься відповідним індексом сполучення: G_{T-T} , $G_{Ш-Ш}$, $G_{T-Ш}$, $G_{Ш-T}$. Теоретична кількість варіантів сполучення ознак ТШП-транскрипції у групі МО, що мають M — сегментні ТП, $NGteor_M = 3^M$.

2. У кожній групі p , що відповідає певному варіанту сполучення ознак ТШП-транскрипції сегментів-фонем, підраховується кількість МО $K_{M,p}$ ($p = \overline{1, NG_M}$).

3. Обчислюється значення вагового коефіцієнта p — ї групи

$$KG_{M,p} = \frac{K_{M,p}}{NSL_M}, \tag{6}$$

де NSL_M – кількість МО, що мають M — сегментні ТП.

Значення $KG_{M,p}$ ($p = \overline{1, NG_M}$) є оцінками h_g для груп МО, що відповідають певному варіанту сполучення ознак ТШП-транскрипції. Вважатимемо, що чим більше МО в групі, тим вагоміше внесок цієї групи у загальну евристичну оцінку h_g . Застосування оцінки h_g : на поточному кроці n виконується аналіз значень h_g , що відповідають сполученням ознак ТШП-транскрипції сегментів-фонем поданого МС, для переходів на графі синтезу по дугам $l_n \to l_{n+N}$ ($N = 2; 3; 4$), і обирається перехід, що відповідає максимальному значенню h_g .

2.3. Алгоритм обчислення оцінки відстані

Для кожної МО обчислюється значення оцінки h_g , яка дозволяє обрати шлях на графі синтезу з поточного стану у напрямку цільового стану таким чином.

1. Обираються МО SL_k , що мають M — сегментні ТП ($M = 2; 3; 4$). МО, групуються за варіантами сполучення ознак ТШП-транскрипції сегментів-фонем GT . Нехай визначено NG_M таких груп.

2. Попарно обчислюються відстані $D_{p,q,r}$ між елементами у кожній групі p методом ДП ($p = \overline{1, NG_M}$, $q, r = \overline{1, NSL_M}$). Визначається середнє значення відстані $\overline{D}_{p,q,r}$ для кожної групи.

3. Для кожної групи p обчислюються оцінка відстані таким чином

$$h_d = \frac{\overline{D}_{p,q,r}}{KG_{M,p}}. \tag{7}$$

де $KG_{M,p}$ – значення вагового коефіцієнта групи p (6).

Застосування оцінки h_d : на поточному кроці n виконується аналіз значень h_d , що відповідають сполученням ознак ТШП-транскрипції сегментів-фонем поданого МС для переходів на графі синтезу по дугам $l_n \to l_{n+N}$ ($N = 2; 3; 4$), і обирається перехід, що відповідає мінімальному значенню оцінки h_d .

2.4. Обчислення оцінки кількості нерозкритих вузлів

Оцінка кількості нерозкритих вузлів, що залишаються нерозглянутими у процесі пошуку в напрямку цільового вузла обчислюється за алгоритмом.

1. Нехай відомі значення параметрів R_N , що відображують надійність розпізнавання N — сегментних МО ($N = 2; 3; 4$).

Для визначення значень R_N пропонується обрати множину реалізацій МС, ТП яких сегментовані на таку кількість сегментів $NSGX$, що $(NSGX \mod 2 = 0) \wedge (NSGX \mod 3 = 0) \wedge (NSGX \mod 4 =$. Це дозволить для однієї реалізації МС побудувати декілька розв'язків-кандидатів і визначити групи МО, розгляд яких підвищує надійність розпізнавання. Визначення значень R_N пропонується виконати згідно наступного критерія

$$F = \frac{n}{N} \to \max, \tag{8}$$

де n – кількість вірно розпізнаних МО у МС тестовій вибірці, N – загальна кількість МО у вибірці.

В загальному випадку вибір припустимих переходів $l_n \to l_{n+N}$ на графі синтезу з поточного стану n (

$N = 2; 3; 4$) пропонується здійснювати у черговості згідно з відсортованими значеннями R_N у порядку спадання.

2. Обчислимо кількість сегментів, що залишилося, $\Delta n = NSGX - n$.

3. Враховуючи модель (1), для Δn можна побудувати $N\,\mathrm{var}(\Delta n)$ шляхів $pathg_k$ ($k = 1 : N\,\mathrm{var}(\Delta n)$) до цільового вузла, знайшовши набори цілочисельних значень $\{\eta, \lambda, \mu\}$ при яких виконується наступна рівність $2 \cdot \eta + 3 \cdot \lambda + 4 \cdot \mu = \Delta n$ за умови обмежень $\eta, \lambda, \mu < \Delta n$, $\eta, \lambda, \mu \geq 0$ (η, λ, μ – кількість дво-, три- та чотирисегментних МО відповідно). Для кожного шляху $pathg_k$ обчислюється кількість переходів cnt_k .

4. На поточному кроці n обирається перехід вздовж дуги $l_n \to l_{n+N}$ ($N = 2; 3; 4$), що відповідає найкоротшому шляху $pathg_{k_{\min}}$ ($k_{\min} = \arg\min_k cnt_k$), в якому на першому переході знаходиться N —сегментна МО, що має максимальне значення R_N ($N = 2; 3; 4$).

3. Експериментальні дослідження

Для створення навчальної множини МО і обчислення складових функції евристичної оцінки станів (4) був проаналізований словник з 5000 найуживаніших слів мови, для якого сформовано мовленнєву базу даних (МБД). Записи МБД згруповані у наступні вибірки: SW_1 – 1-а вибірка МС для множини слів W_1 , з яких сформовано навчальну множину МО; SW_2 – 2-а вибірка МС для множини слів W_1 ; SW_3 – вибірка МС для множини слів W_2 . Множини слів W_1 , W_2 такі, що: $W_1 \cap W_2 = \varnothing$. У таблиці 1, на рисунках 1-3 наведено кільісні дані про навчальну множину МО.

Таблиця 1: Кількісні дані про навчальну множину МО

Група МО $SL_{M,k}$	$k = 2$	$k = 3$	$k = 4$
$M = 2$	438	121	10
$M = 3$	0	1499	1579
Кількість $NSL_{M,k}$	438	1620	1589
Загальна кількість МО NSL	3647		

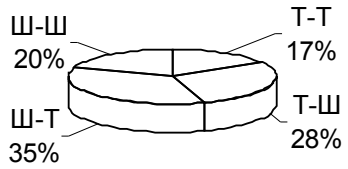

Рисунок 1: Група двосегментних МО

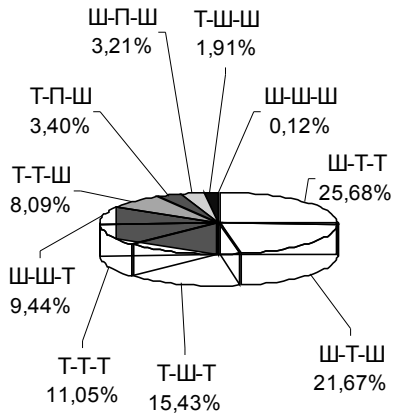

Рисунок 2: Група трисегментних МО

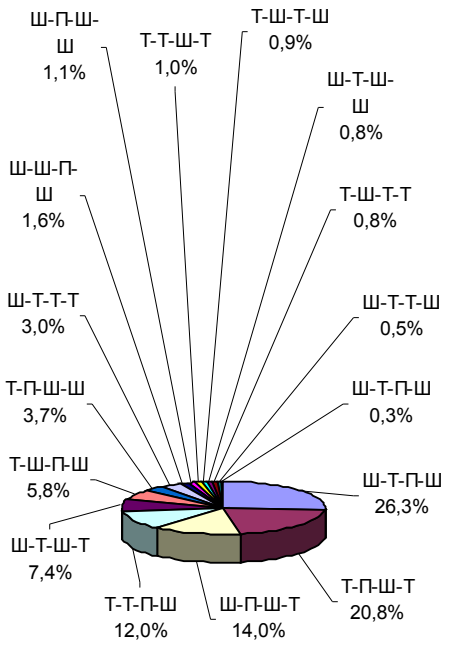

Рисунок 3: Група чотирисегментних МО

У експериментах з обчислення оцінки кількості нерозкритих вузлів отримано наступний результат:

найбільшу надійність розпізнавання мають трисегментні МО. Згідно з критерієм вибору кроку переходу з поточного стану (надійність розпізнавання дво-, три- та чотирисегментних МО) для підвищення надійності та швидкодії розпізнавання обирається наступний порядок здійснення припустимих переходів з деякого поточного стану i ($N = 2; 3; 4$): 3 ($R_N = 0{,}90$); 2 ($R_N = 0{,}85$); 4 ($R_N = 0{,}60$).

Застосування функції евристичної оцінки станів дозволяє прискорити швидкість розпізнавання на 10% у порівнянні з базовим алгоритмом [12] і зменшити похибку розпізнавання до 5%.

4. Висновки

Для підвищення надійності розпізнавання та швидкості евристичного пошуку у просторі станів задачі сегментно-складового синтезу розроблено модель обчислення функції евристичної оцінки станів, яка складається з доданків: оцінки, обчисленої на основі аналізу вкладеності МО; оцінки, обчисленої на основі аналізу варіантів сполучення ознак сегментів ТП МО; оцінки евклідової відстані для всіх можливих переходів з поточного вузла для різних сполучень групових ознак сегментів ТП МО; оцінки кількості нерозкритих вузлів, що залишаються нерозглянутими в процесі пошуку у напрямку цільового вузла. В наступному передбачається подальше удосконалення розробленої моделі.

5. Література

[1] Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К., Наукова думка, 1987. – 264 с.

[2] Вінцюк Т.К. Образний комп'ютер // Сучасні проблеми в комп'ютерних науках. Зб. наук. праць, Вид-во Нац. ун-ту «Львівська політехніка», Львів, 2000, с. 5-14.

[3] Ronzhin A.L., Yusupov R.M, Li I.V., Leontieva A.B. Survey of Russian Recognition Systems. // In Proc. SPECOM'2006, St. Petersburg, 2006, pp. 54-60.

[4] Ронжин А.Л., Карпов А.А., Ли И.В. Система автоматического распознавания русской речи SIRIUS // Искусственный интеллект. – 2005. – № 3. – С. 590-601.

[5] Meisel W.S., Anikst M.T., Pirzadeh S.S., Schumacher J.E., Soares M.C., Trawick D.J. The SSI large-vocabulary speaker-inependent continuous speech recognition system // In Proc. ICASSP'91 Int. Conf. Acoust., Speech and Signal Process., Toronto, 1991, pp. 337-340.

[6] Пилипенко В.В. Распознавание дискретной и слитной речи из сверхбольших словарей на основе выборки информации из баз данных // Искусственный интеллект. – 2006. - №3. – С. 548-558.

[7] Tatarnikova M., Tampel I., Oparin I., Khokholov Yu. Building acoustic models for a large vocabulary continuous speech recognizer for Russian // In Proc. SPECOM' 2006, St. Peterburg, 2006, pp. 83-87.

[8] Кодзасов С.В., Кривнова О.Ф. Общая фонетика. – М.: РГГУ, 2001. – 592 с.

[9] Рассел С., Норвиг П. Искусственный интеллект: современный подход – М., 2006. – 1408 с.

[10] Карпов О.Н. Технология построения устройств распознавания речи. – Д., 2001. – 184 с.

[11] Kopeček I. Speech synthesis based on the composed syllable segments. // Proc. of the First Workshop on Text, Speech and Dialogue – TSD'98, 1998, pp. 259-262.

[12] Карпов О.Н., Савенкова О.А. Эвристический алгоритм поиска оптимальных решений сегментно-слогового синтеза // Искусственный интеллект. – 2007. - №4. – С. 378-385.