Оцінка ефективності кодування спектральної обвідної мовленнєвих сигналів в просторах лінійних спектральних параметрів найвищої регресії методом кластерного аналізу

Павлов О.І., Стасевич П.А., Тертичний Г.М.

Національний технічний університет України «Київський політехнічний інститут», Міжнародний науково-навчальний центр інформаційних технологій та систем
opmail@bigmir.net

Анотація

This paper describes proposed and experimental verified analytic expressions for approximation of a "distortion-rate" characteristic of parameters vector quantization processes that encode speech signals spectrum envelope form in different spaces. Estimate of lower Shannon limit has been found. It has been shown that migration to spaces of the highest regression changes the lower Shannon limit and improves vector quantization potential characteristics.

1. Вступ

Під час розробки та перевірки методів ефективного кодування будь-якого джерела аналогових повідомлень в більшості випадків намагаються отримати такий алгоритм стиснення і відновлення даних, який би мав переваги перед існуючими в трикутнику критеріїв «*складність алгоритму* (кількість операцій, швидкодія) — відносне стиснення даних (*необхідна швидкість передачі* кодованих повідомлень) — *спотворення*, що виникають при кодуванні». При цьому оперують різними наборами параметрів, які певним чином кодують аналогові повідомлення джерела і утворюють відповідні параметричні простори. Процес кодування аналогового джерела супроводжується операціями попередньої обробки даних (що надходять від нього) з утворенням первинного образу повідомлення, перетворенням його в певний еквівалентний простір параметрів, формуванням ознак (набору остаточних параметрів) і їх квантуванням (представленням за допомогою певної скінченої кількості розрядів).

Кожен з вказаних вище трьох критеріїв оцінки ефективності кодування джерела аналогових повідомлень має власні обмеження в кожному конкретному просторі параметрів.

Спотворення, що виникають під час кодування даних (мається на увазі середньоквадратична похибка) обмежені нульовим значенням знизу і дисперсією аналогового джерела зверху.

Нульові спотворення відповідають неквантованому аналоговому повідомленню і вимагають нескінченої розрядності для представлення таких повідомлень (а також нескінченої швидкості для їх передачі).

Якщо спотворення перевищують дисперсію аналогового джерела, то немає жодного сенсу передавати (зберігати) таке спотворене повідомлення, замість якого з успіхом можна використовувати «нульове» повідомлення (що відповідає нульовим параметрам) зі спотвореннями рівними дисперсії.

Необхідна швидкість передачі лежить в межах від нескінченості (для некодованих аналогових повідомлень) до *імовірнісної абсолютної ентропії* К.Е. Шеннона для кодованого (дискретного) джерела в конкретному просторі параметрів — мінімально можливої середньої швидкості потоку його параметрів (відповідно рівню спотворень, обраного схемою кодування).

Складність кодування лежить в межах від умовно нульових витрат (для некодованих аналогових повідомлень) до *алгоритмічної ентропії* А.М. Колмогорова (мінімально можливої складності алгоритму, який відтворює повідомлення кодованого джерела).

Нижня межа швидкості і верхня межа складності співвідносяться між собою через алгоритмічний підхід до визначення ентропії, який враховує не тільки всі імовірнісні закономірності джерела повідомлень, але й інші (якщо вони є); якщо інших закономірностей немає, то алгоритмічний підхід приводить до таких самих результатів, що й імовірнісний.

Так саме пов’язані між собою обраний рівень спотворень і відповідна найменша можлива середня швидкість передачі кодованих повідомлень. Цей зв’язок описується функцією $R(D)$ («rate — distortion») або зворотною функцією $D(R)$, котрі повністю визначаються *імовірнісною диференційною ентропією* аналогового джерела.

Функція «спотворення — швидкість» $D(R)$ визначає найменші можливі (середньоквадратичні) спотворення D, які можуть бути забезпечені при ефективному кодуванні джерела повідомлень в разі витрачання в середньому R біт коду на одну координату обраного вектору (набору) параметрів.

Характеристики, які визначають функції $D(R)$ та $R(D)$, можна застосовувати до будь-яких методів кодування джерела (в тому числі з необмеженою складністю алгоритмів, які можуть вносити необмежені в часі затримки, пов’язані з накопиченням даних).

Це дозволяє порівнювати між собою потенційні можливості різних за складністю схем кодування аналогових повідомлень в різних за розмірами просторах параметрів. Кількість біт R, що витрачається на одну координату вектору параметрів може бути з легкістю перерахована в кількість біт на одне кодоване повідомлення (при відомій кількості параметрів, що кодують одне повідомлення), а також у швидкість передачі потоку (при відомій швидкості створення повідомлень джерелом).

Таким чином, задача полягає в тому, щоб визначити наскільки обраний алгоритм (схема) кодування джерела аналогових повідомлень наближається за своїми показниками до оптимальних характеристик, які визначаються функціями $D(R)$ та $R(D)$, а також оцінити потенційні переваги того чи іншого простору параметрів.

Далі розглядається джерело аналогових повідомлень, яким є алгоритм (схема) кодування форми спектральної обвідної мовленнєвих сигналів. Його повідомлення формуються в дискретні моменти часу (що пов'язано з кадровою обробкою мовлення) і являють собою вектор параметрів, що кодують форму «зрізу» спектральної обвідної, яка є неперервною в часі і за значеннями.

Простори, які порівнюються між собою щодо потенційної ефективності їх використання для кодування вказаного вище джерела є еквівалентними (в тому сенсі, що існують взаємно однозначні перетворювання параметрів з одного простору в інший без будь-яких втрат інформації про повідомлення джерела) і належать до класу просторів так званих лінійних спектральних параметрів.

Враховуючи все сказане можна перейти до розгляду методики досліджень.

2. Оцінка ефективності кодування джерела повідомлень за допомогою кластерного аналізу

2.1. Загальні зауваження

Для оцінки ентропії вказаного вище джерела аналогових повідомлень (що генерує послідовність $\{\boldsymbol{x}_j\}$ реалізацій векторної випадкової змінної $\boldsymbol{x}$ з *неперервним* розподілом значень амплітуди $p(\boldsymbol{x})$) його можна розглядати як крайній випадок *дискретного* джерела (що генерує послідовність $\{\boldsymbol{y}_i\}$ реалізацій векторної випадкової змінної $\boldsymbol{y}_i$ з *дискретним* розподілом значень амплітуди $P(\boldsymbol{y}_i)$). При цьому весь простір повідомлень для *дискретного* джерела є поділеним на L комірок $\{C_i, 1 \le i \le L\}$, кожній з яких приписується кодовий вектор $\boldsymbol{y}_i$, за допомогою якого апроксимується вектор $\boldsymbol{x}_j$, якщо останній потрапляє в комірку C_i .

Перехід від *дискретного* джерела до *неперервного* відбувається при зменшенні розмірів комірок C_i до нескінченно малих значень з одночасним збільшенням їх кількості до нескінченності для заповнення всього простору повідомлень.

Такий підхід по своїй суті є кластерним аналізом, а точніше проектуванням кодової книги алгоритму (схеми) оптимального векторного квантування з послідовним збільшенням кількості рівнів квантування до достатньо великих значень. При цьому шуми квантування *дискретного* джерела (в якості якого виступає схема квантування) наближаються до надзвичайно малих, майже нульових значень, його вихідні вектори $\boldsymbol{y}_i$ наближаються до неквантованих векторів $\boldsymbol{x}_j$ (в плоть до того, коли кожен кластер утворюється з одного вектору $\boldsymbol{x}_i$), а *абсолютна* ентропія *дискретного* джерела $H(\boldsymbol{y}) = -\sum_{i=1}^{L} P(\boldsymbol{y}_i)\log_2 P(\boldsymbol{y}_i)$ наближається до **оцінки** *диференційної* ентропії *неперервного* джерела $h(\boldsymbol{x}) = -\int_x p(\boldsymbol{x})\log_2 p(\boldsymbol{x})d\boldsymbol{x}$ в цьому просторі параметрів.

Ступінь останнього наближення для кожного простору параметрів різна і залежить також і від розмірів блоку параметрів (об'єднання параметрів декількох векторів в один вектор більшого розміру).

Основний постулат теорії кодування говорить про те, що *при використанні векторного квантування можна скільки завгодно близько наблизитися до функції «спотворення — швидкість» шляхом збільшення розмірності векторів*.

Скористатися наведеним постулатом і дослідити вплив збільшення розмірів блоку параметрів на оцінку диференційної ентропії доцільно лише при розгляді задачі порівняння потенційної ефективності кодування джерела (з метою отримання оптимального коду для збереження кодованих повідомлень) в різних просторах параметрів.

При цьому слід очікувати, що навіть в еквівалентних просторах *диференціальна* ентропія буде різною.

В разі розгляду задачі кодування повідомлень з метою їх передачі, специфіка джерела повідомлень (алгоритм (схема) кодування спектральної обвідної мовленнєвих сигналів) є такою, що надмірне збільшення розмірів блоку параметрів і внесення затримки в передачу мовлення більше 10 — 30 мс є неприпустимим.

При цьому оцінка *диференційної* ентропії *неперервного* джерела (як результат наближення *абсолютної* ентропії *дискретного* джерела) буде відображати не тільки перетворення статистичних характеристик параметрів кодування повідомлень в різних просторах, але й вплив обраного фіксованого розміру блоку параметрів.

Слід відзначити, що самі по собі значення диференційної ентропії не мають певного смислу, але набувають його лише відносно інших значень диференційної ентропії. Так, різниця двох значень диференційної ентропії характеризує відмінність інформативності (яка вимірюється в бітах) відповідних джерел.

2.2. Верхня та нижня межа Шеннона

В разі оцінки спотворень як середньоквадратичної похибки (відхилення) при апроксимації неквантованих векторів $\boldsymbol{x}_j$ за допомогою содових векторів $\boldsymbol{y}_i$ для характеристики «спотворення — швидкість» $D(R)$ існує нижня межа, яка має назву *нижньої межі Шеннона*, і визначається так: $D^*(R) = \frac{1}{2\pi e} 2^{2h(\boldsymbol{x})} 2^{-2R}$, де $h(\boldsymbol{x})$ — диференційна ентропія джерела без пам'яті. Для багатьох джерел нижня межа Шеннона може бути досягнута лише при $R \to \infty$.

Використовуючи визначення *нормованого спотворення*, $D(R) = 10\lg\frac{D(R)}{\sigma^2}$, дБ, де σ^2 — дисперсія параметрів, що кодують повідомлення джерела в розрахунку на координату вектору, нижня межа Шеннона у вигляді нормованого спотворення визначається формулою $D^*(R) = -6.02R - 6.02[h_G(\boldsymbol{x}) - h(\boldsymbol{x})]$, дБ, де $h_G(\boldsymbol{x}) = \frac{1}{2}\log_2(2\pi e \sigma^2)$ — *диференційна* ентропія джерела з гауссовою ФЩІ і дисперсією σ^2, яка визначає *верхню межу Шеннона* $D_G(R) = 10\lg\frac{D_G(R)}{\sigma^2} = -6.02R$; $h(\boldsymbol{x})$ — *диференційна* ентропія джерела аналогових повідомлень, $h(\boldsymbol{x}) \le h_G(\boldsymbol{x}) = \frac{1}{2}\log_2(2\pi e \sigma^2)$.

Таким чином, для будь-якого простору параметрів що кодують повідомлення джерела експериментальна крива $D(R)$ в цьому просторі завжди лежить між незмінною верхньою і відповідною для цього простору нижньою межею Шеннона: $D^*(R) \le D(R) \le D_G(R)$.

Функції $D_G(R)$ і $D^*(R)$ в логарифмічному масштабі завжди подаються паралельними прямими з нахилом $-6{,}02$ дБ/біт. Функція $D(R)$ наближається до нахилу $-6{,}02$ дБ/біт при високих швидкостях (зазвичай $R > 3$ біт).

Знаючи нижню межу Шеннона $D^*(R)$ у вигляді нормованого спотворення $D^*(R)$, можна порівнювати з нею характеристики векторного квантування параметрів в різних просторах враховуючи можливу зміну дисперсії σ^2 при еквівалентних перетвореннях кодованих повідомлень.

Незважаючи на те, що функції $R(D)$ и $D(R)$ визначаються просто, аналітичне їх обчислення (за виключенням декількох окремих випадків) є складним.

В [1] наведені асимптотичні формули для характеристик схем векторного квантування. При великій кількості рівнів квантування L оптимальний (в сенсі мінімуму спотворень ступені 2) L-кодова схема квантування в N-вимірному векторному просторі має спотворення $D_2 = -6.02R + F_{VQ}(p, N)$, де R — кількість біт на координату, а член $F_{VQ}(p, N)$ залежить від ФЩІ $p(\boldsymbol{x})$ і розмірності векторного простору. Складнощі використання наведеного виразу полягають в тому, що точний вигляд для члену $F_{VQ}(p, N)$ є відомим лише в декількох окремих випадках. Саме тому отримання аналітичних виразів функцій, що апроксимують експериментальні дані є також актуальною задачею.

В однаковій мірі це відноситься і до диференційної ентропії джерел, що розглядаються.

По аналогії з відомими мало чисельними окремими випадками можна очікувати, що для диференційної ентропії $h(\boldsymbol{x}) = h_G(\boldsymbol{x}) - \left(D_G(R) - D^*(R)\right)/6.02$, яка також має вигляд $h(\boldsymbol{x}) = \frac{1}{2}\log_2(2\pi e \sigma^2) - \left(D_G(R) - D^*(R)\right)/6.02$, аналітичний вираз слід шукати у вигляді $h(\boldsymbol{x}) = \frac{1}{2}\log_2(C\sigma^2)$, де C — деяка константа.

2.3. Результати експериментальних досліджень

Далі розглядаються характеристики «спотворення — швидкість» векторного квантування кодованих повідомлень про форму спектральної обвідної джерела мовленнєвих сигналів, поданих в 4 еквівалентних просторах параметрів (по 10 параметрів на кожний спектральний «зріз»), які дозволяють взаємно однозначний перехід від одного простору до іншого без внесення спотворень в повідомлення джерела, а саме в просторах [2 — 9]:

- лінійних спектральних частот найвищої регресії (ЛСЧ-НР),
- лінійних спектральних проекцій найвищої регресії (ЛСП-НР),
- класичних лінійних спектральних частот (ЛСЧ),
- класичних лінійних спектральних проекцій (ЛСП).

Таблиця 1: Результати векторного квантування

Швидкість, біт на координату	Нормовані спотворення, $10\log\left(D_{ave}/\sigma^2\right)$, дБ,			
	ЛСЧ-НР	ЛСП-НР	ЛСЧ	ЛСП
0,0	0,000	0,000	0,000	0,000
0,1	-3,984	-4,251	-1,710	-2,283
0,2	-6,766	-7,437	-2,747	-3,429
0,3	-8,504	-9,669	-3,712	-4,610
0,4	-9,931	-11,608	-4,591	-5,586
0,5	-11,203	-13,159	-5,368	-6,519
0,6	-12,350	-14,619	-6,095	-7,364
0,7	-13,432	-15,993	-6,798	-8,187
0,8	-14,441	-17,244	-7,493	-8,972
0,9	-15,397	-18,439	-8,186	-9,748
1,0	-16,321	-19,587	-8,882	-10,527
1,1	-17,225	-20,703	-9,586	-11,301
1,2	-18,117	-21,800	-10,314	-12,095
1,3	-19,021	-22,897	-11,082	-12,926

В якості функції загального вигляду, що використовувалася для апроксимації експериментальних даных, була обрана функція виду $f(x) = p_0\left(3 - e^{p_1 x} - e^{p_2 x} - e^{p_3 x}\right) - 6{,}02x$, невідомим вектором-параметром якої був вектор $\mathbf{p} = \left[p_0 \quad p_1 \quad p_2 \quad p_3\right]^{\mathrm{T}}$, для пошуку якого використовувався градієнтний метод, реалізований в пакеті *Mathcad* 2001*i Professional*.

Для просторів ЛСЧ-НР ЛСП-НР, ЛСЧ, ЛСП після декількох ітерацій обчислень були отримані такі результати:

- ЛСЧ-НР: $\mathbf{P}_{fhi}^{(5)} = \left[-5{,}626 \quad -6{,}945 \quad -0{,}915 \quad -0{,}266\right]^{\mathrm{T}}$,
- ЛСП-НР: $\mathbf{P}_{phi}^{(2)} = \left[-7{,}178 \quad -4{,}896 \quad -0{,}597 \quad -0{,}597\right]^{\mathrm{T}}$,
- ЛСЧ: $\mathbf{P}_{lsf}^{(4)} = \left[-1{,}648 \quad -7{,}876 \quad -0{,}479 \quad -0{,}48\right]^{\mathrm{T}}$,

- ЛСП: $\mathbf{P}_{lsp}^{(11)} = \left[-2{,}220 \quad -7{,}577 \quad -0{,}734 \quad -0{,}733\right]^{\mathrm{T}}$.

Оцінка нижньої межі Шеннона для характеристик векторного квантування що досліджувалися шукалася у вигляді $\mathrm{L}(x) = \lim_{x \to \infty} f(x) = 3p_0 - 6{,}02x$.

Для кожного з вказаних просторів таке наближення дало:

- ЛСЧ-НР: $\mathrm{L}(x)_{fhi} = -16{,}878 - 6{,}02x$,
- ЛСП-НР: $\mathrm{L}(x)_{phi} = -21{,}534 - 6{,}02x$,
- ЛСЧ: $\mathrm{L}(x)_{lsf} = -4{,}944 - 6{,}02x$,
- ЛСП: $\mathrm{L}(x)_{lsp} = -6{,}660 - 6{,}02x$.

Експериментальні результати дослідження процесів векторного квантування в 4 вказаних вище просторах подані на рис.1 разом з результатами їх апроксимації, незмінною верхньою межею Шеннона та визначеними оцінками нижньої межі Шеннона для кожного простору.

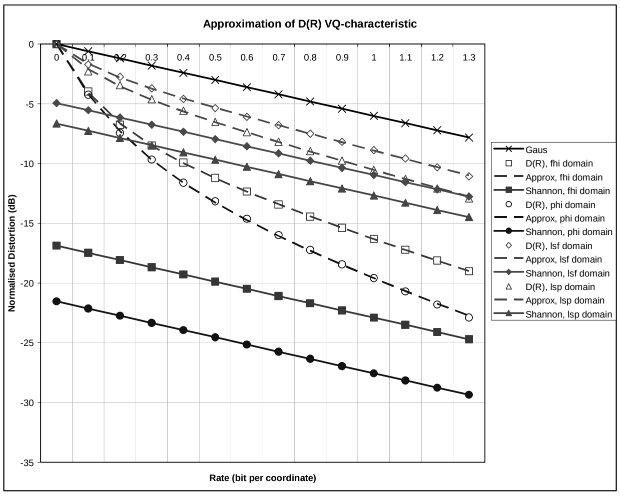

Рисунок 1: Апроксимація характеристик векторного квантування.

3. Обговорення результатів

Зменшення нормованої похибки квантування (потенційний виграш, що очікується) в разі квантування параметрів спектральної обвідної мовленнєвих сигналів в просторах найвищої регресії складає:

- в просторі ЛСЧ-НР по відношенню до простору ЛСЧ: $B_{fhi/lsf} = (3p_0)_{fhi} - (3p_0)_{lsf} = -11{,}934$ [дБ],
- в просторі ЛСП-НР по відношенню до простору ЛСП: $B_{phi/lsp} = (3p_0)_{phi} - (3p_0)_{lsp} = -14{,}874$ [дБ],
- в просторі ЛСП-НР по відношенню до простору ЛСЧ-НР: $B_{phi/fhi} = (3p_0)_{lsp} - (3p_0)_{fhi} = -4{,}656$ [дБ].

4. Висновки

1. Запропоновані та експериментально перевірені аналітичні вирази апроксимуючих функцій для характеристик «спотворення — швидкість» процесів векторного квантування параметрів спектральної обвідної мовленнєвих сигналів в просторах ЛСЧ-НР, ЛСП-НР, ЛСЧ, ЛСП.
2. Знайдені оцінки аналітичних виразів для функції нижньої межи Шеннона при векторному квантуванні параметрів спектральної обвідної мовленнєвих сигналів у вказаних просторах.
3. Показано, що перехід до просторів найвищої регресії зменшує нижню межу Шеннона і покращує потенційні характеристики векторного квантування.
4. Оцінений потенційний виграш що очікується при векторному квантуванні за рахунок зменшення нижньої межі Шеннона: в просторі ЛСЧ-НР по відношенню до класичного простору ЛСЧ — 11,934 дБ, в просторі ЛСП-НР по відношенню до класичного простору ЛСП — 14,874 дБ, в просторі ЛСП-НР по відношенню до простору ЛСЧ-НР — 4,656 дБ.

5. Посилання

[1] Дж. Макхоул, С. Рукос, Г. Гиш. Векторное квантование при кодировании речи // ТИИЭР, 1985 , т.73, №11, с. 19 — 61.

[2] Павлов О.И. Прямое П-преобразование в линейном предсказании речи // Радиоэлектроника. — 2000. — N12. — С. 53 — 66. (Изв. высш. учеб. заведений).

[3] Павлов О.И. Упрощение реализации метода линейных спектральных пар (частот) в линейном предсказании речи // Труды 3-й Международной конференции “Цифровая обработка сигналов и ее применение”, Т.3, стр. 128 — 132, Москва, 2000.

[4] Павлов О.И. Быстрый алгоритм и графическое представление прямого преобразования в методе линейных спектральных частот высшего порядка // Труды 3-й Международной конференции “Цифровая обработка сигналов и ее применение”, Т. 3, стр. 132 — 136, Москва, 2000.

[5] Павлов О.И. Алгоритм быстрого прямого П-преобразования и особенности его математического аппарата // Радиоэлектроника. — 2001. — N2.— С. 61 — 73. (Изв.высш.учеб.заведений).

[6] Павлов О.И. Обратное П-преобразование в линейном предсказании речи // Радиоэлектроника. — 2001. — N1. — С. 61 — 73. (Изв. высш. учеб. заведений).

[7] Павлов О.И. Алгоритм быстрого обратного П-преобразования // Радиоэлектроника. — 2001. — N8.— С. 67 — 77. (Изв.высш.учеб.заведений).

[8] Павлов О.И. Свойства линейных спектральных частот высших порядков // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2001. Вып. 117. С. 62 — 64

[9] Павлов О.И. Межкадровая интерполяция спектральной огибающей речевого сигнала в пространстве линейных спектральных частот наивысшей регрессии // Радиоэлектроника. — 2008. — N4.— С. 56 — 69. (Изв.высш.учеб.заведений).