Антонова Г.О.
Одеська національна академія зв'язку ім. О.С. Попова

ЗАСТОСУВАННЯ АЛГОРИТМУ ВІТЕРБІ ДЛЯ ЗАВАДОСТІЙКОГО ПРИЙОМУ ЦИФРОВИХ ПОСЛІДОВНОСТЕЙ І ДЕКОДУВАННЯ ЗГОРТКОВИХ КОДІВ

Метою цієї роботи є застосування алгоритму Вітербі для мінімізації ймовірності помилок при проходженні цифрового сигналу амплітудно-модульованих систем по вузькосмуговому каналу зв'язку.

Алгоритм динамічного програмування у формі алгоритму Вітербі знайшов широке застосування в різних галузях зв'язку для вирішення різних проблем, зокрема, у техніці зв'язку при рішенні зв'язних завдань. Зручністю алгоритму Вітербі є те, що рішення виноситься не відразу, а методом послідовного пошуку оптимального рішення поставленого завдання. Алгоритм наближень зручно представити у вигляді діаграми станів. При виконанні декодування в техніці зв'язку це має велике значення, оскільки дозволяє більш точно декодувати отриманий сигнал, тобто реалізує так зване м'яке рішення.

Фактично метод динамічного програмування – це пошук такого оптимального керування, яке забезпечить перехід системи з початкового стану в кінцеве за найменшу кількість кроків. При цьому, з метою надання кількісної оцінки оптимальності керування вводять певний числовий критерій W, який буде характеризувати якість та ефективність виконуємих операцій. Задача динамічного програмування – обрати такий спосіб організації дій системи, який дозволить отримати мінімізацію (максимізацію) певного критерію W [1].

Одна з причин інтересу вчених до положень алгоритму Вітербі – спроможність боротися з міжсимвольною інтерференцією (МСІ). Вона виникає при обмеженні смуги частот [2].

Формування компактного (вузького) спектра для передавання інформації дає можливість провайдерам збільшити дозволену швидкість передачі, а також розмістити в одному стовбурі передачі більше абоненских потоків. Однак, необмежене зменшення смуги частот є неможливим, оскільки при цьому порушується умова відліковості Найквіста і, як наслідок, сприяє зростанню впливу міжсимвольної інтерференції на винесення рішення про прийнятий сигнал.

Розглянемо більш докладно проблему втручання МСІ при передачі високошвидкісних цифрових даних. На рис. 1,а, зображено одиночний імпульс, що був переданий по відносно вузькосмуговому каналу, а також відгук каналу на нього. При передачі послідовності імпульсів, як, наприклад, при амплітудній модуляції, де інформація закладена в зміні амплітуд імпульсу, втручання МСІ особливо помітно. При цьому відбувається накладання одного імпульсу, що розтягнувся за часом, на сусідній імпульс, що викликає зростання ймовірності помилки імпульсу та істотно погіршує якість прийому (рис. 1,б). Тому замість значення корисного сигналу в моменти взяття відліку на схемі взяття відліків присутня зважена сума амплітуд імпульсів у декількох сусідніх інтервалах. Це призводить до помилкових рішень про прийнятий сигнал.

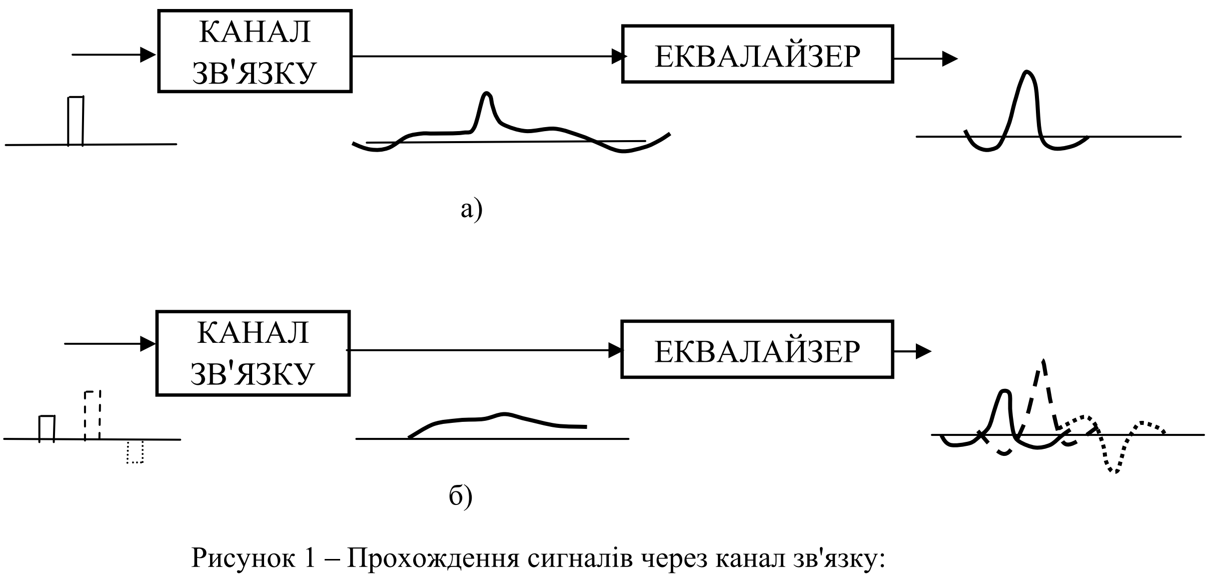

Рисунок 1 – Проходження сигналів через канал зв'язку:
а) проходження одиночного імпульсу; б) проходження пачки імпульсів

Простий перебір прийнятих комбінацій з метою оптимального виявлення цілої послідовності імпульсів не є доцільним, оскільки, обчислювальна складність зростає експоненціально з довжиною передаваної послідовності. До того ж обчислення не може починатися поки повна послідовність не отримана.

Важливість підходу динамічного програмування в тім, що число обчислень, необхідних для оптимального виявлення, росте лише лінійно з довжиною переданої послідовності, і обчислення можуть здійснюватися під час одержання послідовності.

Розглянемо математичну модель сигналу основної смуги частот, тобто, сигнал, що не модулює несучу.

Нехай $s(t)$ – послідовність чисел, які називають «інформаційні символи», передаються по каналу, позначимо їх $a_i,...a_N$. Кількість інформаційних символів N є значною, проте скінченною. Передбачається, що ці символи незалежні й можуть приймати L однаково ймовірних значень. Інформаційна послідовність являє собою сигнали певної амплітуди, що мають місце в певні моменти часу, – тактові інтервали T.

$$s(t) = \sum_{i=1}^{N} a_i p(t - iT), \tag{1}$$

де a_i – певні коефіцієнти, що відповідають інформаційним символам;
$p(t)$ – переданий імпульс із символьною швидкістю $1/T$ Бод.
T – тривалість тактового інтервалу.

Вихід каналу основної смуги частот обчислюється

$$y(t) = \sum_{i=1}^{N} a_i h(t - iT) + n(t), \tag{2}$$

де $n(t)$ – білий гауссівський шум із двосторонньою спектральною густиною потужності $N_o/2$, Вт;
$h(t)$ – відгук каналу на вплив $p(t)$;

Далі будемо посилатися на $h(t)$ як на імпульсну реакцію каналу основної смуги. Будемо вважати, що $h(t)$ має обмежений набір значень m (як пропонується на рис. 1). Це твердження має два наслідки. По-перше, прийом всіх N символів здійснюється на інтервалі $0 < t < \tau$, $(N + m)\, T < t < \tau$. Другий наслідок стосується терміну виконання. Пам'ять каналу зв’язку, тобто його інерційність, може бути записана у вигляді:

$$r_{i-j} = \int_0^{\tau} h(t - iT)\,h(t - jT)\,dt, \tag{3а}$$

Тому що окремий випадок

$$r_{i-j} = 0, \text{ для } |i - j| > m. \tag{3б}$$

Таким чином, ми повинні посилатися на m як на пам'ять каналу в одиницях T.

Наше завдання-– діяти на отриманий сигнал $y(t)$, $0 \le t \le \tau$ таким чином, щоб робити оцінку послідовності $a_1^*, a_2^*,...a_N^*$, що відповідає переданим символам $a_1, a_2,... a_N$.Даний сигнал зіпсований адитивним шумом, і ми не можемо впевнено відтворювати передану послідовність. Скоріше ми прагнемо мінімізувати ймовірність послідовності помилок, тобто, імовірність того, що $a_1^*, a_2^*,...a_N^*$ буде відмінна від $a_1, a_2,... a_N$. Відповідно до наших припущень для якісної обробки прийнятої послідовності необхідно забезпечити мінімальну ймовірність помилки кожного з символів.

Як критерій, що потребує, (в нашому випадку мінімізації), для забезпечення оптимального керування діями системи, оберемо функцію густини ймовірності. Будемо визначати її як імовірність сигналу $y(t)$ на інтервалі $0 \le t \le \tau$, який прийнято за умови, що передавалися символи $\hat{a}_1, \hat{a}_2, ..., \hat{a}_N$.

$$P\left[y(t), 0 \le t \le \tau \mid a_1 = \hat{a}_1, a_2 = \hat{a}_2, ... a_N = \hat{a}_N\right], \tag{4}$$

Необхідно звернути увагу на те, що для прийнятого сигналу існує L^N можливих значень, оскільки кількість послідовностей становить $\hat{a}_1, \hat{a}_2, ..., \hat{a}_N$.. Отже, для мінімізації ймовірності послідовності помилок потрібно L^N обчислень. Перевагою алгоритму Вітербі є те, що число обчислень, необхідних для визначення максимуму оцінки ймовірності послідовності, зросте лінійно зі зростанням N і вимагає меншого об'єму обчислень, ніж експонентне.

Ми зараз одержали вираз для ймовірності, що наочно показує необхідні обчислення. З формули (2) видно, що при передачі по каналу зв'язку послідовності a_1, $a_2, \ldots$ a_N і оцінок цього сигналу $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N$, сумарний сигнал $y(t)$ є гауссівським процесом із середнім значенням

$$s(t) = \sum_{i=1}^{N} \hat{a}_i p(t - iT),$$

що дозволяє нам написати

$$p\left[y(t), 0 \le t \le \tau \mid a_1 = \hat{a}_1, a_2 = \hat{a}_2, \ldots a_N = \hat{a}_N\right] = K \exp\left[-\frac{1}{N_0}\int_0^{\tau}\left[y(t) - \sum_{i=1}^{N} \hat{a}_i p(t - iT)\right]^2 dt\right], \qquad (5)$$

де K – константа, що не залежить від $y(t)$ і послідовності $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N$. При пошуку оптимальної послідовності, його метою є не абсолютна величина ймовірності, а її відносна величина для різних наборів $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N$. Мінімальна ймовірність помилки досягається шляхом вибору набору $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N$, що мінімізує інтеграл

$$P(y(t)|a_N = \int_0^{\tau}\left[y(t) - \sum_{i=1}^{N} \hat{a}_i p(t - iT)\right]^2 dt.$$

У результаті розкриття квадратичного члена під знаком інтеграла маємо три члени, один з яких має вигляд:

$$\int_0^{\tau} y(t)^2 dt.$$

Цей член не залежить від послідовності $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N$ і тому може не братись до уваги при подальших розрахунках. Запишемо вираз для об'єктивної функції ймовірності, яка потребує мінімізації

$$D = -2\sum_{i=1}^{N} \hat{a}_i Z_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \hat{a}_i \hat{a}_j r_{i-j}, \qquad (6)$$

де

$$Z_i = \int_0^{\tau} y(t) h(t - iT) dt.$$

Два члени у формулі (6) відображають два типи інформації в наших теоретичних викладках. Вираз

$$Z_i = \int_0^{\tau} y(t) h(t - iT) dt,\ i = 1, 2, \ldots, N$$

може розглядатися як вихід фільтра з відповідним імпульсним відгуком $h(t)$. Доданок $\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{a}_i \hat{a}_j r_{i-j}$ відображає наші знання про пам'ять каналу (див. (3,б)).Хоча формула (5) показує істотне спрощення, рішення завдання мінімізації прямим шляхом (переборний метод вирішення задачі) вимагає L^N обчислень.

Розпишемо об'єктивну функцію як

$$D = -2\sum_{i=1}^{N} \hat{a}_i Z_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \hat{a}_i \hat{a}_j r_{i-j} = -2\sum_{i=1}^{N-1} \hat{a}_i Z_i + \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} \hat{a}_i \hat{a}_j r_{i-j} - 2\hat{a}_N Z_N + 2\hat{a}_N \sum_{i=N-m}^{N-1} \hat{a}_i r_{N-i} + \hat{a}_N^2 r_0. \qquad (7)$$

Очевидним є той факт, що кожна зі складових залежить тільки від одного виходу відповідного фільтра Z_N.

Використовуючи уведені нами раніше позначення, об'єктивна функція може бути представлена в більш зручному вигляді. Визначимо систему поля векторів

$$\sigma_k = \{\hat{a}_{k-m+1}, \hat{a}_{k-m+2}, \ldots, \hat{a}_k\},\ k = m, m+1, \ldots, N. \qquad (8)$$

Відзначимо, що до складу σ_k входять всі інформаційні символи, за винятком $\hat{a}_{k+1}$, що визначається y_{k+1}. При цьому існує взаємозв'язок між послідовністю в полі векторів $\sigma_m, \sigma_{m+1}, \ldots, \sigma_N$

та оцінкою послідовності переданих символів $\hat{a}_1, \hat{a}_2, ..., \hat{a}_N$, а отже набір поля векторів має надмірність. Проблема вибору оптимальної послідовності з набору $a_1,...,a_N$ може розглядатись як проблема вибору оптимальних значень $\sigma_m, \sigma_{m+1}, ..., \sigma_N$. Оцінка оптимальної послідовності станів $\sigma_m, \sigma_{m+1}, ..., \sigma_N$ може розглядатися як пошук оптимального шляху по решітчастій діаграмі станів, що відображають стан кодера в різні моменти часу. На рис. 3 така решітчаста діаграма станів показана для випадку, коли $L = 2$ й $m = 3$. Пунктирні лінії вказують переміщення, які можуть бути зроблені від одного стану до іншого.

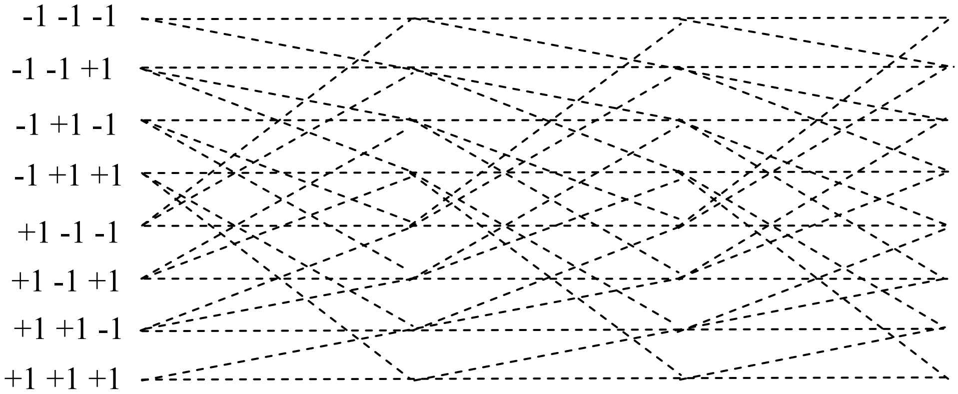

Рисунок 2 – Решітчаста діаграма для $L = 2$, $m = 3$.

Для того, щоб сформулювати проблему як пошук оптимального шляху, ми повинні сформулювати зручні критерії для визначення відстаней. Продовжимо визначення величин

$$U(Z_1,...,Z_k;\sigma_m,...,\sigma_k) \stackrel{\Delta}{=} -2\sum_{i=1}^{N} \hat{a}_i Z_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \hat{a}_i \hat{a}_j r_{i-j}, \qquad (9а)$$

а також

$$V(Z_k;\sigma_{k-1},\sigma_k) \stackrel{\Delta}{=} -2\hat{a}_k Z_k + 2\hat{a}_k \sum_{i=1}^{N} \hat{a}_i r_{k-i} + a_k^2\ r_0,\ \ k = m+1,...,N. \qquad (9б)$$

Використовуючи введені раніше позначення, ми можемо написати

$$U(Z_1,...,Z_k;\sigma_m,...,\sigma_k) = U(Z_1,...,Z_{k-1};\sigma_m,...,\sigma_{k-1}) + V(Z_k;\sigma_{k-1},\sigma_k). \qquad (10)$$

Проблема пошуку набору станів, які мінімізують об'єктивну функцію (6), вимагає мінімізації виразу (10) за умови мінімальних значень $\sigma_M,...,\sigma_N$

$$I(Z_1,...,Z_N) \stackrel{\Delta}{=} \min_{\sigma_m,...,\sigma_N} U(Z_1,...,Z_N;\sigma_m,...,\sigma_N) = \min_{\sigma_N} \min_{\sigma_m,...,\sigma_{N-1}|\sigma_N} U(Z_1,...,Z_N;\sigma_m,...,\sigma_N). \qquad (11)$$

Вираз (11) наочно ілюструє, що мінімізація здійснюється у два етапи.

1) Для значень σ_N, що приймають значення з набору можливих значень L^m, виконують мінімізацію над значеннями $\sigma_m,...,\sigma_{N-1}$. При цьому, кількість обчислень дорівнює $L^{(N-1)}$.

2) Використовуючи значення, отримані на першому кроці, мінімізують σ_N (кількість порівнянь дорівнює L).

Перепишемо вираз (10) з урахуванням складових $U(Z_1,...,Z_k;\sigma_m,...,\sigma_k)$ та $V(Z_k;\sigma_{k-1},\sigma_k)$.

$$I(Z_1,...Z_N) = \min_{\sigma_N} \min_{\sigma_{N-1}|\sigma_N} \min_{.\sigma_m,...,\sigma_{N-2}|\sigma_{N-1},\sigma_N} \left[U(Z_1,...Z_{N-1};\sigma_1,...,\sigma_{N-1}) + V(Z_N;\sigma_{N-1},\sigma_N)\right]. \qquad (12)$$

Є очевидним, що поставлена раніше задача пошуку набору $\hat{a}_1, \hat{a}_2, ..., \hat{a}_N$, яка максимізує $P\left[y(t), 0 \le t \le \tau \mid a_1 = \hat{a}_1, a_2 = \hat{a}_2, ... a_N = \hat{a}_N\right]$ та вираз (12), є еквівалентними. Як було раніше зазначено, ми маємо однозначний зв'язок між $\sigma_m, \sigma_{m+1}, ..., \sigma_N$ і набором символів $\hat{a}_1, \hat{a}_2, ..., \hat{a}_N$. Відповідно до величин з виразу (12) можна зробити два спостереження, які є критичними при побудові алгоритму.

$$\min_{\sigma_m,...,\sigma_{N-2}|\sigma_{N-1},\sigma_N} V(Z_N;\sigma_{N-1},\sigma_N) = V(Z_N;\sigma_{N-1},\sigma_N). \qquad (13а)$$

Це говорить про те, що якщо σ_{N-1} і σ_N є фіксованими, то зміна станів $\sigma_m,...,\sigma_{N-2}$ може характеризуватися значенням $V(Z_N;\sigma_{N-1},\sigma_N)$:

$$\min_{\sigma_m,...,\sigma_{N-2}|\sigma_{N-1},\sigma_N} U(Z_1,...,Z_{N-1};\sigma_m,...,\sigma_{N-1}) = \min_{\sigma_m,...,\sigma_{N-2}|\sigma_{N-1}} U(Z_1,...,Z_{N-1};\sigma_m,...,\sigma_{N-1}). \quad (13б)$$

Підставляючи вирази (11), (12), і (13) один в одний, маємо

$$\min_{\sigma_N}\min_{\sigma_m,...,\sigma_{N-1}|\sigma_N} U(Z_1,...,Z_N;\sigma_m,...,\sigma_N) =$$
$$= \min_{\sigma_N}\min_{\sigma_{N-1}|\sigma_N}\left[V(Z_N;\sigma_{N-1},\sigma_N) + \min_{\sigma_m,...,\sigma_{N-2}|\sigma_{N-1}} U(Z_1,...,Z_{N-1};\sigma_m,...,\sigma_{N-1})\right]. \quad (14а)$$

Використовуючи ті ж кроки, які привели до виразу (14а), ми можемо розкласти другий член виразу (14а). Тоді ми знаходимо

$$\min_{\sigma_m,...,\sigma_{N-2}|\sigma_{N-1}} U(Z_1,...,Z_{N-1};\sigma_m,...,\sigma_{N-1}) =$$
$$= \min_{\sigma_{N-2}|\sigma_{N-1}}\left[V(Z_N;\sigma_{N-2},...,\sigma_{N-1}) + \min_{\sigma_m,...,\sigma_{N-3}|\sigma_{N-2}} U(Z_1,...,Z_{N-2};\sigma_m,...,\sigma_{N-2})\right]. \quad (14б)$$

Продовжуючи наші перетворення, пишемо

$$F(\sigma_{k+1}) \stackrel{\Delta}{=} \min_{\sigma_k|\sigma_{k+1}}\left[V(Z_{k+1};\sigma_k,\sigma_{k+1}) + F(\sigma_k), k = m,...,N-1\right], \quad (15)$$

де

$$F(\sigma_k) \stackrel{\Delta}{=} U(Z_1,...,Z_m;\sigma_m) = \min_{\sigma_m,...,\sigma_{k-1}|\sigma_k} U(Z_1,...,Z_k;\sigma_m,...,\sigma_k), k = m+1,...,N-1.$$

Хоча це не є очевидним, проте $F(\sigma_k)$ усе ще залежить від $Z_1,...,Z_k$. Вирази (14а) і (15) втілюють алгоритм Вітербі. Використовуючи вираз (15), знаходимо значення L^m виразу $F(\sigma_{m+1})$. Позначаючи σ_{m+1} відповідно до рис. 2, ми отримуємо, що σ_{m+1} має L можливих попередніх станів таких, що $\min_{\sigma_m|\sigma_{m+1}}$ вимагає L порівнянь для кожного L^m можливих значень для σ_{m+1}.

Кількість $F(\sigma_{m+1})$ є "мінімальною відстанню" до конкретного значення для σ_{m+1}. Згідно з принципом оптимальності, "оптимальний повний шлях" через діаграму станів повинен проходити по одному зі шляхів до кожного з L^m можливих реалізацій стану σ_{m+1}. Аналогічним чином, як обчислено $F(\sigma_{m+1})$, розраховуються наступні величини $F(\sigma_{m+2}), F(\sigma_{m+3}),...,F(\sigma_N)$.Позначимо кожний з L^m можливих шляхів як σ_N. На заключному етапі вибирається шлях, для якого $F(\sigma_N)$ мінімально. Потрібно підкреслити, що L^m пов'язане із числом рівнів для переданих символів і пам'яттю каналу, але не із числом переданих символів *N*. Таким чином, виходить, що алгоритм Вітербі вимагає фіксованого числа обчислень на символ незалежно від числа отриманих символів.

По роботі можна зробити такі висновки:

– алгоритм Вітербі один з декількох нелінійних підходів до сигнального виявлення. Він дозволяє значно зменшити кількість обчислень, необхідних для його реалізації порівняно із методом простого перебору;

– в загальному випадку алгоритм Вітербі є придатним для проблеми виявлення марковського процесу дискретного часу в присутності адитивного шуму;

– в області передачі даних алгоритм Вітербі набув широкого використання при декодуванні згорткових кодів цифрової ФМ, текстовому розпізнаванні і запису на магнітну стрічку.

Література

1. *Вентцель Е. С.* Элементы динамического программирования. – М.: Наука, 1964. – 176 с.
2. *Скляр Б.* Цифровая связь. Теоретические основы и практическое применение. – 2-е изд., испр.: Пер. с англ. – М.: Вильямс, 2003. –1104 с.