ЛЕКЦИЯ 6. ПРИЗНАКИ. КЕПСТРАЛЬНЫЕ КОЭФФИЦИЕНТЫ. MFCC

Сергей Николенко

1. Выделение признаков

Пусть есть какой-то входной сигнал. Для того, чтобы его оцифровать, то есть привести в вид, который используется алгоритмом распознавания, исходный сигнал делится на *окна*, на каждом из которых отдельно рассчитываются свойства. Деление происходит таким образом, чтобы каждая точка перекрывалась два раза. Например, если мы разделим сигнал на части шириной δh, то каждое окно будет иметь ширину $2\delta h$, тогда указанное выше условие будет выполнено.

Рассмотрим более подробно процесс выделения признаков.

1.1. **Банки фильтров.** Берется сигнал, над ним производится преобразование Фурье, далее это преобразование делится на фильтр по частотам. Считаются свойства сигналов на каждой из этих частот.

$$x(t) \longrightarrow X(iw)$$

Фильтры по частотам складываются в вектор, а затем над ними производятся дополнительные действия. Минусы способа: слишком большой результирующий вектор, не достаточно хорошее выделение признаков, так как признаки, полученные на разных отрезках частот будут довольно сильно коррелировать между собой, что приведет к потере информации.

1.2. **Линейное предсказательное кодирование (LPC).** Сигнал делится на источник и фильтр. Делается предположение, что фильтр можно представить в виде следующего многочлена:

$$H[n] = \frac{A}{1 - a_k z^{-k}}$$

Коэффициенты a_k и являются признаками сигнала. Их надо подвергать дополнительному преобразованию в зависимости от того, как мы воспринимаем сигнал (PLP признаки).

1.3. **Mel Frequency Cepstral Coefficients (MFCC).** Доказано, что чувствительность человека к звуковому сигналу зависит от его частоты: чем ниже частота, тем чувствительность выше. В 1937 году была выведена формула, по которой можно перевести частоту в *герцах* в частоту в *мелах.*

Мел - специальная единица измерения частоты звука, основанная на статистической обработке большого числа данных о субъективном восприятии высоты звуковых тонов.

Законспектировали Селифонов Евгений и Тихомиров Андрей.

$$x \longrightarrow \boxed{FT} \longrightarrow \boxed{ln} \longrightarrow \boxed{IFT}$$

$$x * h \quad XH \quad \hat{X} + \hat{H} \quad \hat{x} + \hat{h}$$

Рис. 1. MFCC.

Формулы перевода частоты в герцах (f) и в мелах (m) представлены ниже:

$$m = 1127.01048 \ln(1 + f/700)$$

$$f = 700(e^{m/1127.01048} - 1)$$

Сигнал представляется как свёртка двух функций: исходного сигнала и фильтра, параметры которого мы хотим оценивать. Мы хотим выделить эти отдельные компоненты. Для этого нам необходимо преобразование, которое бы превращало свёртку в сумму:

$$x * h \to \hat{x} + \hat{h}$$

Для этого вводится преобразование *кепстр*:

$C[n] = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \ln|X(e^{i\omega})| e^{i\omega n}\, d\omega$ - ***вещественный кепстр***

$C[n] = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \ln(X(e^{i\omega})) e^{i\omega n}\, d\omega$ - ***комплексный кепстр***

Итак, схема преобразований сигнала будет выглядеть следующим образом:

- Преобразование Фурье: $x * h \to XH$
- Преобразование кепстр: $XH \to \hat{X} + \hat{H}$
- Обратное преобразование Фурье: $\hat{X} + \hat{H} \to \hat{x} + \hat{h}$

Таким образом, разделение на источник и фильтр происходит само собой.

На рисунке представлена схема преобразования сигнала.

При этом в преобразовании $x * h \to \hat{x} + \hat{h} x[n] * h[n] \to \hat{x}[n] + \hat{h}[n]$ получаем, что $\hat{x}[n]$ - исток, а $\hat{h}[n]$ - соответственно фильтр.

Причем $\hat{x}[n] \approx 0$ при больших n, $\hat{h}[n] \approx 0$ при малых n.

Пример 1

Пример расчета кепстральных коэффициентов.

$$H(z) = \frac{Az^r \Pi(1 - a_k z^{-1}) \Pi(1 - u_k z)}{\Pi(1 - b_k z^{-1}) \Pi(1 - v_k z)}$$

$\Rightarrow$ Здесь преобразование Фурье уже выполнено.

$\hat{H}(z) = \ln A + \ln|z^r| + \Sigma \ln(1 - a_k z^{-1}) + \Sigma \ln|1 - u_k z| - \Sigma \ln(1 - b_k z^{-1}) - \Sigma \ln|1 - v_k z|$

Здесь $\ln A + \ln|z^r|$ - сдвиг фазы.

Нам необходимо найти $\hat{h}[n]$ - коэффициенты при z^{-n}.

Заметим, что $\ln(1 - x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - ... = -\Sigma \frac{x^k}{k}$

Отсюда:

$$\hat{H}(z) = \sum_{n=1}^{\infty} \left(-\sum_k \frac{a_k^n z^{-n}}{n} - \sum_k \frac{u_k^n z^n}{n} + \sum_k \frac{b_k^n z^{-n}}{n} + \sum_k \frac{v_k^n z^n}{n} \right)$$

Таким образом:

$$\hat{h}[n] = \begin{cases} \ln A & n = 0 \\ \sum\limits_k \frac{v_k^n - u_k^n}{n} & n < 0 \\ \sum\limits_k \frac{b_k^n - a_k^n}{n} & n > 0 \end{cases}$$

$\Leftarrow$

Пример 2

Пример расчета коэффициентов для LPC.

$$H(z) = \frac{G}{1 - \sum\limits_1^p a_k z^{-k}}$$

$\Rightarrow$

$$\hat{H}(z) = \ln G - \ln(1 - \sum_1^p a_k z^{-k}) = \sum_{n=-\infty}^{\infty} \hat{h}[n] z^{-n}$$

Нам необходимо найти коэффициенты при z^{-k}, т.е. $\hat{h}[n]$.
Для этого возьмем производную от левой и правой частей:

$$\sum_{n=-\infty}^{\infty} -n\hat{h}[n] z^{-n-1} = \frac{\sum\limits_k k a_k z^{-k-1}}{1 - \sum\limits_{k=1}^p a_k z^{-k}}$$

Проведем дальнейшие преобразования:

$$\sum_{n=-\infty}^{\infty} n\hat{h}[n](1 - \sum_{k=1}^p a_k z^{-k}) z^{-n} = \sum_{k=1}^p k a_k z^{-k}$$

Раскроем левую часть:

$$\sum_{n=-\infty}^{\infty} n\hat{h}[n] z^{-n} - \sum_{n=-\infty}^{\infty} \left(\sum_k a_k \hat{h}[n] z^{k-n} \right) = \sum_{k=1}^p k a_k z^{-k}$$

Рассмотрим коэффициенты при z^{-n}.
На множестве $(n < 1) \times (n > p)$ мы получаем следующее выражение:

$$\sum_n \sum_k n a_k \hat{h}[n] z^{-k-n} = \sum_n z^{-n} \left(\sum_{k=1}^p (n+k) a_k \hat{h}[n+k] \right)$$

Получим следующее соотношение:

$$\begin{cases} (n < 1) \times (n > p) & 0 = n\hat{h}[n] - \sum\limits_{k=n-p}^{n-1} k\hat{h}[k] a_{n-k} \\ 0 < n \leqslant p & n a_n = n\hat{h}[n] - \sum\limits_{k=n-p}^{n-1} k\hat{h}[k] a_{n-k} \end{cases}$$

Во втором выражении a_{n-k} имеет смысл рассматривать только в пределах от 1 до p, т.к. остальные будут равны нулю.

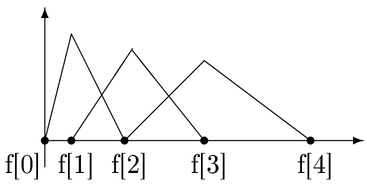

Рис. 2. Треугольные фильтры.

Решением данного соотношения является:

$$\hat{h}[n] = \begin{cases} a_n + \sum\limits_{k=1}^{n-1} \frac{k}{n}\hat{h}[k]a_{n-k} & 0 < n < p \\ \sum\limits_{k=n-p}^{n-1} \frac{k}{n}\hat{h}[k]a_{n-k} & n > p \\ \ln G & n = 0 \end{cases}$$

Таким образом, мы научились выражать кепстральные коэффициенты для LPC. $\Leftarrow$

1.4. **Треугольные фильтры.** Умножая функцию на фильтр, мы усредняем ее на некотором участке. Рассмотрим треугольные фильтры. Будем исходить из того, что человек лучше различает звуки на низких частотах.

Формула для разделения оси частот на треугольные фильтры будут иметь следующий вид:

$$f[k] = \frac{N}{F_s} m^{-1}(m(f_{min}) + k\frac{m(f_{max} - f_{min})}{N+1}$$

Здесь N - количество семплов, F_s - частота дискретизации, $m(f)$ - функция перевода частоты в герцах в частоту в мелах, рассмотренная раньше.

Формула для высоты треугольного фильтра:

$$H_m[n] = \begin{cases} 0 & n < f[m-1] \times n > f[m+1] \\ \frac{2(n-f[m-1])}{(f[m+1]-f[m-1])(f[m]-f[m-1])} & f[m-1] \leqslant n < f[m] \\ \frac{2(f[m+1]-n)}{(f[m+1]-f[m-1])(f[m+1]-f[m])} & f[m] \leqslant n < f[m+1] \end{cases}$$

Важную роль играет понятие *логарифмической энергии* на участке:

$$S[m] = \ln\left(\sum_{k=0}^{N-1} |X[k]|^2 H_m[k]\right)$$

К этой энергии применяется *дискретное косинусное преобразование*:

$$c[k] = \sum_{m=0}^{I-1} S[m]\cos\left(\frac{\pi}{I}(m+\frac{1}{2})k\right), 0 \leqslant k \leqslant I$$

Это преобразование обладает свойством компактности энергии: большей энергии соответствует меньшее количество коэффициентов.

Таким образом, общая схема преобразований выглядит так:

$$x[n] \rightarrow^{FT} X[n] \rightarrow^{TF} \begin{cases} X[n]H_1[n] \rightarrow^{LE} S[1] \\ X[n]H_2[n] \rightarrow^{LE} S[2] \\ ... \\ X[n]H_I[n] \rightarrow^{LE} S[I] \end{cases} \rightarrow^{DCT} c[k]$$

Здесь:

- FT - преобразование Фурье
- TF - применение треугольных фильтров
- LE - логарифмическая энергия
- DCT - дискретное косинусное преобразование

На выходе мы получаем коэффициенты $c[k]$, из которых в качестве признаков обычно берут первые 13-20.