

تشخیص گفتار با استفاده از
مدل مخفی مارکوف
دانیال خشابی

دانشگاه صنعتی امیرکبیر
دی ماه ۱۳۹۰

آنچه خواهیم دید:

◎ تعریف و بررسی مدل مخفی ماکوف (HMM)

➤ تعاریف اولیه و مثال

➤ دو مساله اساسی و راه حل آنها

◎ بررسی عملکرد مدل در سیستم تشخیص گفتار (بلی/خیر)

➤ مدل سازی و حل مساله برای استفاده از HMM

تعاریف اولیه

◎ خاصیت مارکوفی:

$$P[x_t = S_j \mid x_{t-1} = S_i, x_{t-2} = S_k, \dots] = P[x_t = S_j \mid x_{t-1} = S_i]$$

➤ احتمال جابجایی از هر state تنها به state در لحظه ی قبل بستگی دارد.

◎ احتمال و ماتریس گذار حالت ها:

$$a_{ij} = P[x_t = S_j \mid x_{t-1} = S_i], 1 \leq i, j \leq N$$

$$A = [a_{ij}]$$

◎ با شرط اینکه:

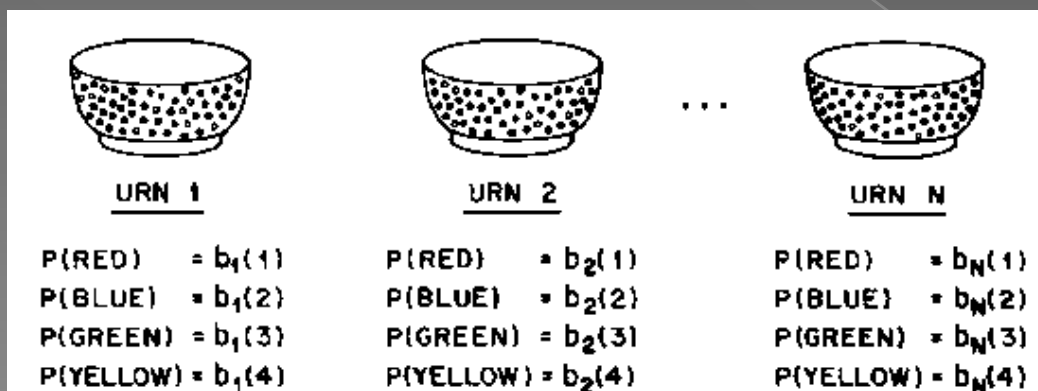
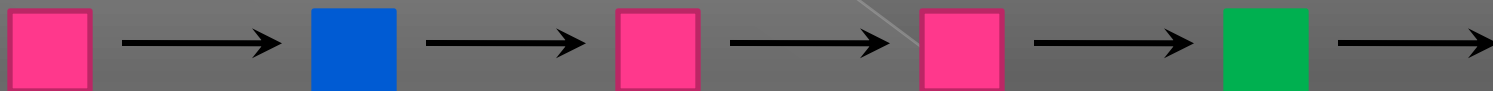
$$\sum_j a_{ij} = 1, a_{ij} \geq 0$$

مدل مخفی مارکوف (Hidden Markov Model)

◎ یک مثال: کیسه ها و گوی های رنگی

➤ فرض کنیم یک زنجیره ی مارکوفی سه حالت از ظرف ها داریم.

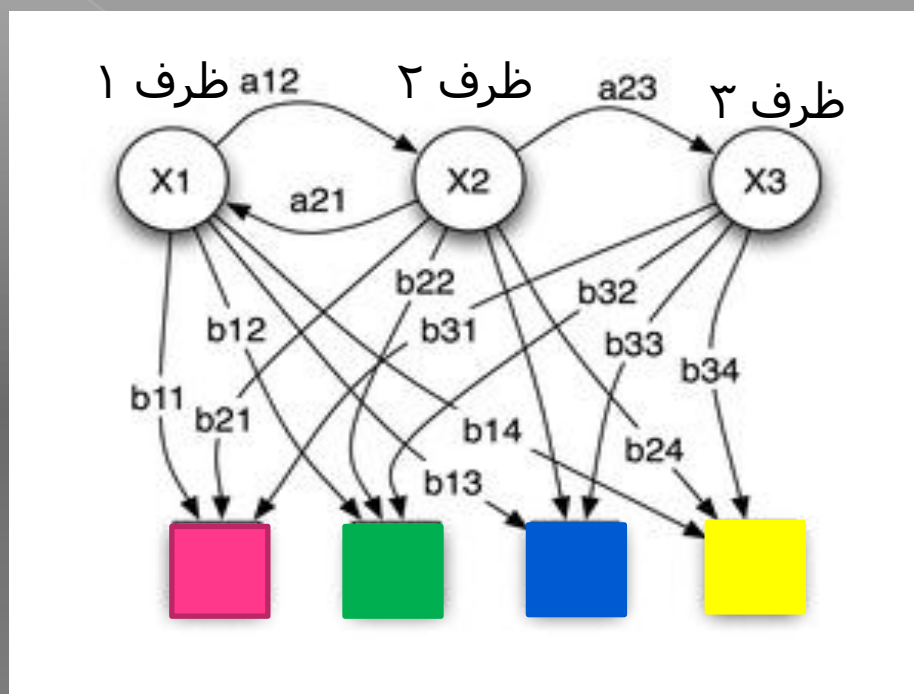
➤ نفر دومی در هر لحظه یک گوی از به صورت تصادفی از یکی از کیسه ها برداشته و رنگ مورد نظر را اعلام می کند بدون اینکه بدانیم از کدام کیسه گوی را برداشته است!



مدل مخفی مارکوف (Hidden Markov Model)

◎ ادامه ی مثال: کیسه ها و گوی های رنگی

➤ در واقع می توان سیستم مورد نظر را به صورت زیر مدل کرد:



➤ مخفی بودن مدل به خاطر عدم مشاهده ی مستقیم حرکت روی حالت ها است!!

مدل مخفی مارکوف (Hidden Markov Model)

تعریف عمومی:

۱. N : تعداد state ها (تعداد ظرف ها)

۲. M : تعداد مشاهده های ممکن در هر حالت (تعداد رنگ های گوی ها)

۳. احتمال گذار: $a_{ij} = P[x_t = S_j | x_{t-1} = S_i]$

۴. احتمال مشاهده ی هر کدام از observation ها در یک state مشخص

$$b_j(k) = P[y_k @ t | x_t = S_j]$$

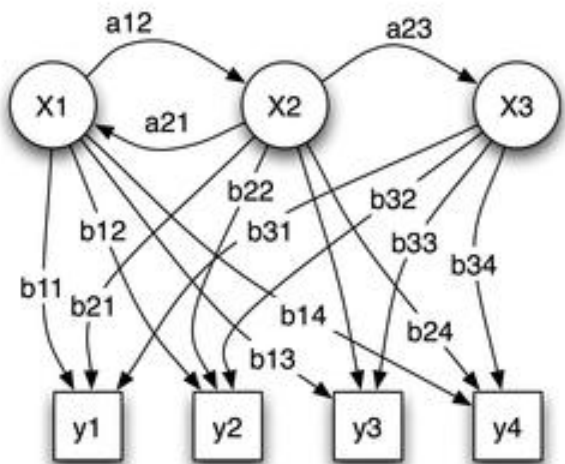
۵. وضعیت اولیه ی زنجیر:

$$\pi_i = P[x_1 = S_i]$$

✓ پس یک زنجیره ی مارکوف پنهان به صورت

زیر توصیف می گردد:

$$\lambda = (A, B, \pi)$$



بررسی ویژگی های HMM

◎ سوال های اساسی که می توان مطرح کرد:

۱. احتمال مشاهده ی یک دنباله ی مشخص از رنگ ها به شرط داشتن یک زنجیره ی مارکوف پنهان معلوم؟ (ارزیابی)

$$\text{given : } Y = y_1, y_2, \dots, y_T \Rightarrow P[Y | \lambda = (A, B, \pi)] = ?$$

۳. یک زنجیره با state های ثابت، به ازای کدام مشخصات زنجیره، احتمال مشاهده ی یک دنباله ی مشخص از رنگ ها را حداکثر می کند؟ (آموزش)

$$\text{given : } Y = y_1, y_2, y_3, \dots, y_T \Rightarrow \arg \max_{\lambda} P[Y | \lambda = (A, B, \pi)]$$

بررسی ویژگی های HMM

◎ سوال اول:

given : $Y = y_1, y_2, \dots, y_T \Rightarrow P[Y | \lambda = (A, B, \pi)] = ?$

consider : $X = x_1, x_2, x_3, \dots, x_T$ پاسخ: ◎

$$\begin{cases} P[X | \lambda] = \pi_{x_1} \cdot a_{x_1, x_2} \cdot a_{x_2, x_3} \dots a_{x_{T-1}, x_T} \\ P[Y | X, \lambda] = \prod_{t=1}^T P[y_t | x_t, \lambda] = b_{x_1}(y_1) \cdot b_{x_2}(y_2) \dots b_{x_T}(y_T) \end{cases}$$

$$P(Y | \lambda) = \sum_{all\ X} P[Y | X, \lambda] P[X | \lambda] =$$

$$= \sum_{all\ X} \left(\pi_{x_1} \cdot a_{x_1, x_2} \cdot a_{x_2, x_3} \dots a_{x_{T-1}, x_T} \right) \cdot \left(b_{x_1}(y_1) \cdot b_{x_2}(y_2) \dots b_{x_T}(y_T) \right)$$

◎ محاسبه دارای پیچیدگی محاسباتی $T \cdot N^T$ است.

> یعنی به ازای $T=100$ و $N=5$ به مرتبه 10^{72} محاسبه احتیاج است!

بررسی ویژگی های HMM

◎ روش سریعتر برای حل این مساله: Forward-Backward Procedure
> متغیر forward را به صورت زیر تعریف می کنیم:

$$\alpha_t(i) = P(y_1, y_2, \dots, y_T, x_T = S_i \mid \lambda)$$

> می توان نشان داده الگوریتم زیر به جواب سوال اول منجر می شود:

1. Initialization: $\alpha_1(i) = \pi_i b_i(y_1), 1 \leq i \leq N$
2. Induction: $\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(y_{t+1}), 1 \leq i \leq N, 1 \leq t \leq T-1$
3. Termination: $P(Y \mid \lambda) = \sum_{i=1}^N \alpha_T(i)$

> پیچیدگی محاسباتی این الگوریتم برابر است با N^2T

تشخیص بله / خیر

سوال دوم: $\text{given : } Y = y_1, y_2, y_3, \dots, y_T \Rightarrow \arg \max P[Y | \lambda = (A, B, \pi)]$

➤ پیچیده ترین سوال موجود در HMM ؛ بدون حل آنالیتیک.

پاسخ تقریبی: هدف بدست آوردن بهترین تقریب Maximum Likelihood برای پارامترهای a_{ij} و π_i و $b_i(k)$ است.

➤ می توان نشان داد تخمین های زیر خواسته های فوق را برآورده می کنند:

$\bar{\pi}_i$ = Expected frequency (number of times) in state x_i at time $(t=1) = \gamma_1(i)$

$$\bar{a}_{ij} = \frac{\text{Expected number of transitions from state } x_i \text{ to state } x_j}{\text{Expected number of transitions from state } x_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(j)}$$

$$\bar{b}_i(k) = \frac{\text{Expected number of times in state } j, \text{ Observing Symbol } y_k}{\text{Expected number of times in state } j} = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}$$

$$s.t. : \begin{cases} \gamma_t(i) = P(x_i = S_i | Y, \lambda) \\ \xi_t(i, j) = P(x_i = S_i, x_{i+1} = S_j | Y, \lambda | Y, \lambda) \end{cases}$$

تشخیص بله / خیر

◎ مراحل:

Offline



Online



تشخیص بله / خیر

◎ مرحله ی اول:

➤ دریافت ورودی: نمونه گیری با نرخ 16 KHz

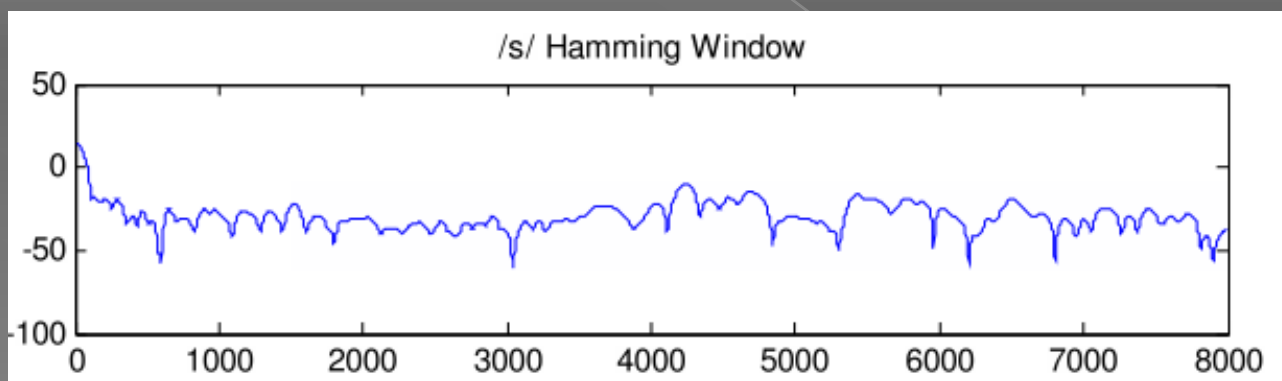
➤ تابع پیش تاکید با پارامتر $\alpha=0.975$

$$H(z) = 1 - \alpha z^{-1}$$

➤ تبدیل فوریه ی زمان-گسسته زمان-کوتاه (ST-DFT)

• طول هر فریم 30 ms و دارای اشتراک 20 ms

• نمونه گیری با پنجره ی Hamming:



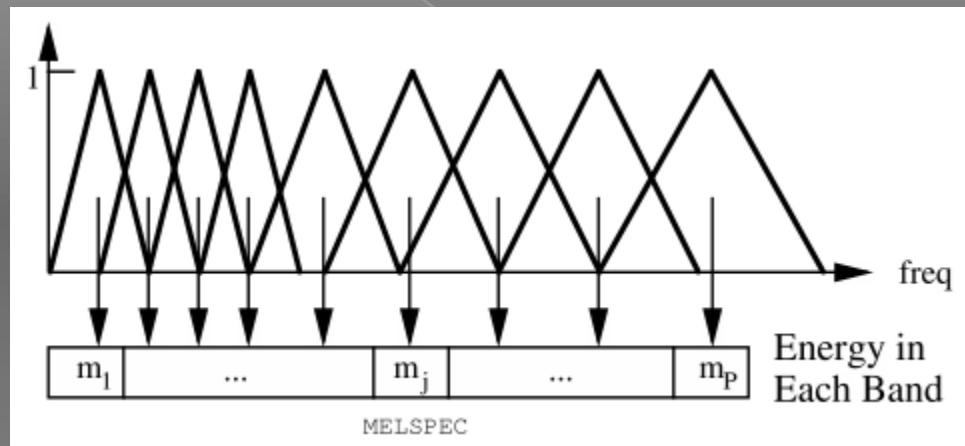
تشخیص بله / خیر

● استخراج ویژگی های طیفی (MFCC)

- > ویژگی های طیفی بهتر از ویژگی های زمانی می توانند نوع صدا را توصیف کنند!
- > قله های هر صدا در شکل طیفی آن مخصوص خود آن صدا است!

● مراحل:

۱. ضرب طیف فرکانسی در مجموعه ای پنجره های مثلثی



۲. محاسبه ی ضرایب تابع تبدیل فوریه ی کسینوسی گسسته (DCT)

✓ به این ترتیب ۳۹ ضریب به عنوان ویژگی های فریم مورد نظر بدست می آید.

تشخیص بله / خیر

◎ آموزش HMM برای دو کلمه:

➤ آنچه در اختیار است تعداد بردار ویژگی در هر لحظه است (Observation)

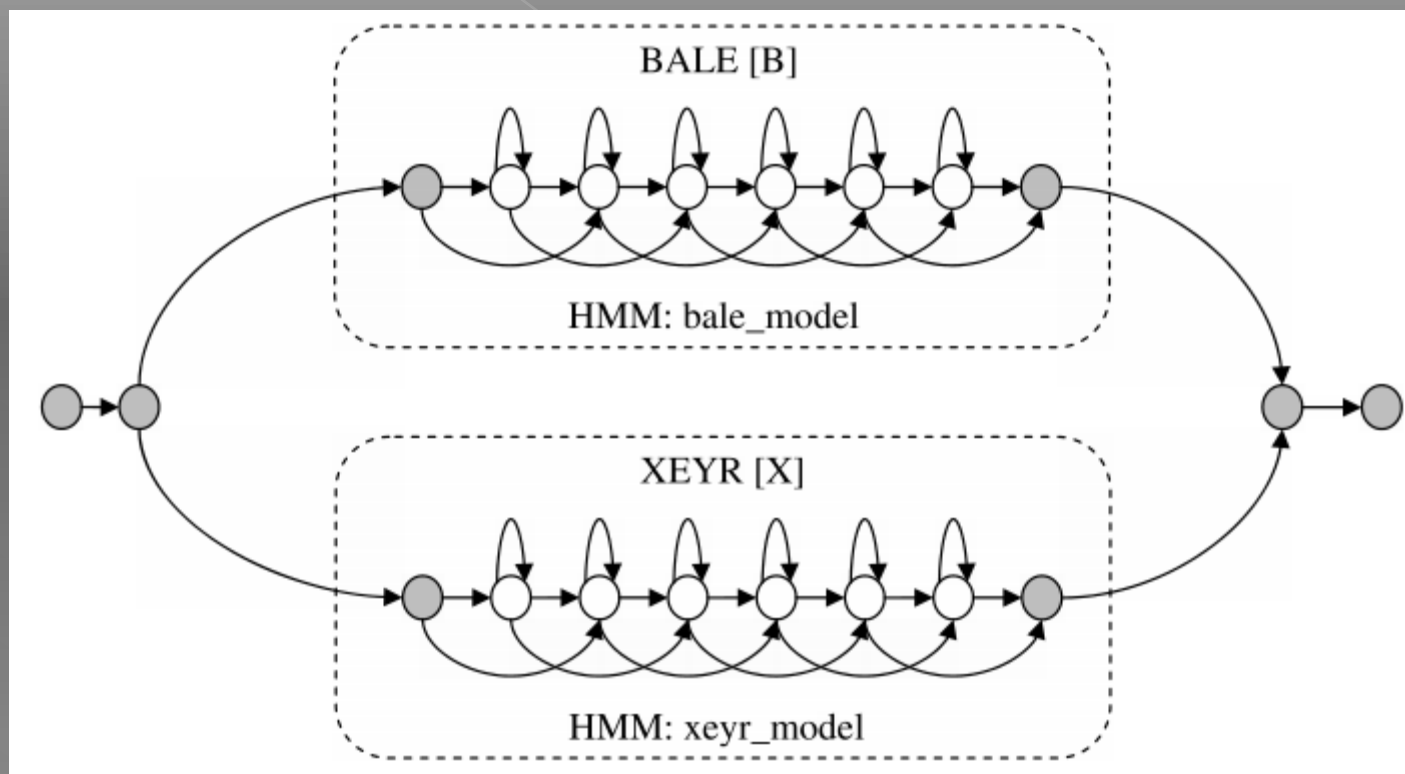
➤ آموزش با استفاده از جواب مساله ی سوم انجام می گیرد.

$$\text{given : } Y = y_1, y_2, y_3, \dots, y_T \Rightarrow \arg \max_{\lambda} P[Y | \lambda = (A, B, \pi)]$$

تشخیص بله / خیر

◎ در مرحله ی تشخیص کافی است داده ی نمونه ی جدید گرفته شده و با استفاده از جواب مساله ی اول HMM درست تشخیص داده می شود:

$$\text{given : } Y = y_1, y_2, \dots, y_T \Rightarrow P[Y | \lambda = (A, B, \pi)] = ?$$



◎ اشکالات مدل معرفی شده تاکنون:

- > حساسیت خیلی زیاد به Noise محیط
- > حساسیت به صدای گوینده (شخص، جنس، لحن، کشیدگی صدا،...)
- > Prototype مربوط به HMM می تواند خیلی تاثیر گذار باشد.
- > تنها برای داده های stationary صادق است.

با تشکر از توجه شما!



1. A tutorial on Hidden Markov Models and selected applications in speech recognition, L. Rabiner, 1989, Proc. IEEE 77(2):257--286.
2. Spoken Language Processing, Huang, Acero, Hon, 2001.
3. The HTK book, 2000.
4. N. Bathaee, Final Project Report, Digital Speech Processing Course.
5. HTK, Hidden Markov Model Toolkit, C Codes.
6. Wikipedia: The Free Encyclopedia