دانشگاه صنعتی شریف
دانشکده‌ی مهندسی کامپیوتر
طراحان: علی رحیمی، ارشیا ایزدیاری، امیرعلی رستمی، فرید محمود زاده، علیرضا میرزکنی، علیرضا ملک حسینی

هوش مصنوعی

بهار ۱۴۰۴

استاد: احسان تن قطاری

تمرین سوم | **شبکه های بیزی و مدل های مارکوف** | مهلت ارسال: ۱۷ اسفند

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین سقف ۴ روز و در مجموع ۱۰ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده پذیرفته نخواهند بود. همچنین، به ازای هر ساعت تأخیر غیر مجاز نیم درصد از نمره ی تمرین کم خواهد شد.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

سوالات (۱۰۰ نمره)

۱. (۱۸ نمره) درستی یا نادرستی عبارت‌های زیر را با ذکر دلیل مشخص کنید.

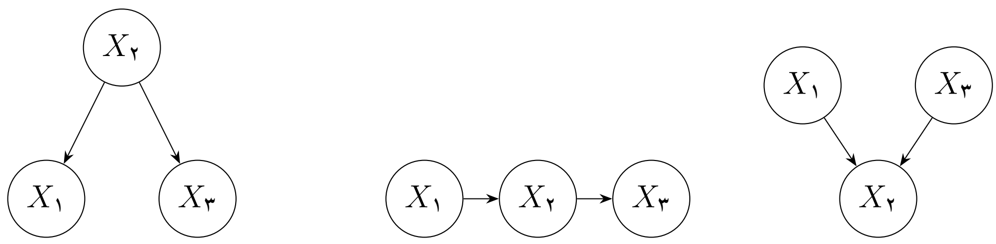
(آ) شبکه بیزی مربوط به توزیع توأم زیر دارای یال‌هایی از گره A به C، از B به C، از B به E، از C به D و از C به E است و هیچ یال دیگری ندارد.

$$P(A)\ P(B)\ P(C \mid A, B)\ P(D \mid C)\ P(E \mid B, C)$$

(ب) حاصل‌ضرب احتمالات شرطی زیر بیانگر یک شبکه بیزی معتبر بر روی متغیرهای A، B، C و D است.

$$P(A \mid B) \quad P(B \mid C) \quad P(C \mid D) \quad P(D \mid A)$$

با توجه به شبکه‌های بیزی زیر، به دو سؤال بعدی پاسخ دهید.



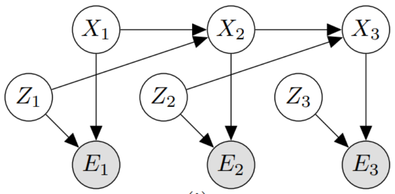
A B C

(ج) هر توزیع توأم $P(X_1, X_2, X_3)$ که با استفاده از توزیع‌های احتمال شرطی (CPD)[1] در شبکه بیزی A قابل بیان باشد، با انتخاب مناسب CPD ها در شبکه بیزی B نیز قابل تولید است.

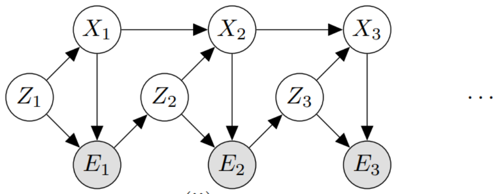
[1] Conditional Probability Distributions

(د) هر توزیع توأم $P(X_1, X_2, X_3)$ که با استفاده از CPD در شبکه بیزی A قابل بیان باشد، با انتخاب مناسب CPD ها در شبکه بیزی C نیز قابل تولید است.

با توجه به مدل‌های مارکوف تغییریافته‌ی زیر، به دو سؤال بعدی پاسخ دهید.



(i)



(ii)

(ه) در مدل (i) معادله‌ی تغییریافته‌ی Elapse Time برابر است با:

$$P(X_t, Z_t|e_{1:t-1}) = \sum_{x_{t-1}, z_{t-1}} P(x_{t-1}, z_{t-1}|e_{1:t-1}) P(X_t|x_{t-1}, z_{t-1})$$

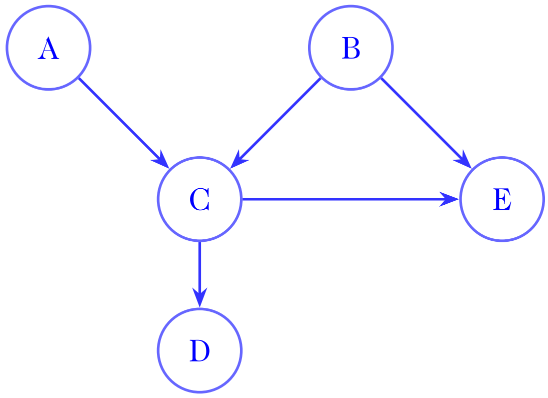
(و) در مدل (ii) معادله‌ی تغییریافته‌ی Observe برابر است با:

$$P(X_t, Z_t|e_{1:t}) \propto P(X_t, Z_t|e_{1:t-1}) P(e_t|X_t, Z_t)$$

(ز) در یک مدل HMM استاندارد مقدار X_∞ مستقل از مقدار X_i به ازای هر i است.

حل.

(آ) درست. این شبکه به صورت زیر است.



(ب) نادرست. زیرا این شبکه $P(A \mid B) \cdot P(B \mid C) \cdot P(C \mid D) \cdot P(D \mid A)$ نشان‌دهنده‌ی یک گراف جهت‌دار شامل دور است: $A \to D \to C \to B \to A$. از آن‌جا که شبکه‌های بیزی باید به صورت DAG (گراف‌های جهت‌دار بدون دور) باشند، این ساختار معتبر نیست.

(ج) درست. اگر یک توزیع در شبکه A قابل نمایش باشد، می‌توان آن را به صورت زیر نوشت:

$$P(X_2) \cdot P(X_1 \mid X_2) \cdot P(X_3 \mid X_2)$$

با استفاده از قاعده‌ی بیز، می‌توان این عبارت را به فرم زیر بازنویسی کرد:

$$P(X_2) \cdot P(X_3 \mid X_2) \cdot \frac{P(X_2 \mid X_1) \cdot P(X_1)}{P(X_2)} = P(X_3 \mid X_2) \cdot P(X_2 \mid X_1) \cdot P(X_1)$$

که دقیقاً با ساختار شبکه B مطابقت دارد:

$$P(X_1) \cdot P(X_2 \mid X_1) \cdot P(X_3 \mid X_2)$$

(د) نادرست. شبکه A می‌تواند توزیع‌هایی را نمایش دهد که در آن $X_۱$ می‌تواند به $X_۳$ وابسته باشد به شرط نداشتن $X_۲$، در حالی که شبکه C نمی‌تواند.

(ه) نادرست. به مراحل زیر توجه کنید:

$$P(X_t, Z_t \mid e_{۱:t-۱}) = \sum_{x_{t-۱}, z_{t-۱}} P(X_t, Z_t, x_{t-۱}, z_{t-۱} \mid e_{۱:t-۱})$$

$$= \sum_{x_{t-۱}, z_{t-۱}} P(x_{t-۱}, z_{t-۱} \mid e_{۱:t-۱}) P(X_t \mid x_{t-۱}, z_{t-۱}, e_{۱:t-۱}) P(Z_t \mid X_t, x_{t-۱}, z_{t-۱}, e_{۱:t-۱})$$

$$= \sum_{x_{t-۱}, z_{t-۱}} P(x_{t-۱}, z_{t-۱} \mid e_{۱:t-۱}) P(X_t \mid x_{t-۱}, z_{t-۱}) P(Z_t)$$

(و) درست. به مراحل زیر توجه کنید:

$$P(X_t, Z_t \mid e_{۱:t}) \propto P(X_t, Z_t, e_t \mid e_{۱:t-۱})$$

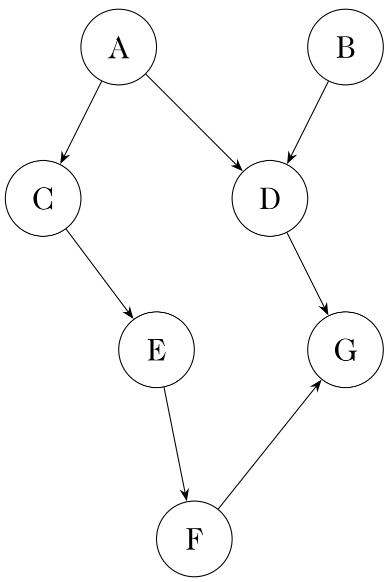
$$\propto P(X_t, Z_t \mid e_{۱:t-۱}) P(e_t \mid X_t, Z_t, e_{۱:t-۱})$$

$$\propto P(X_t, Z_t \mid e_{۱:t-۱}) P(e_t \mid X_t, Z_t)$$

(ز) درست. در یک مدل HMM استاندارد، توزیع مقدار X_∞ فقط با توجه به توزیع انتقال حالت به دست می‌آید. این توزیع با داشتن یا نداشتن X_i به طور ثابت به دست می‌آید و این دو متغیر مستقل‌اند.

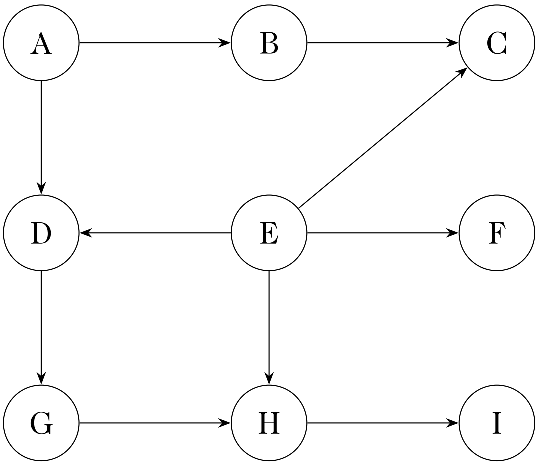
۲. (۱۶ نمره) با توجه به شبکه‌های بیز هر بخش، به سوالات آن پاسخ دهید.

(آ) مطابق شبکه بیز زیر، مشخص کنید کدام یک از تساوی‌های زیر حتماً برقرار هستند. (برای پاسخ خود استدلال بیاورید.)



- $P(A, E \mid G) = P(A \mid G) * P(E \mid G)$
- $P(A \mid B = b) = P(A)$
- $P(E, G \mid D) = P(E \mid D) * P(G \mid D)$
- $P(A, B \mid F) = P(A \mid F) * P(B \mid F)$

(ب) مطابق شبکه بیز زیر، مشخص کنید کدام‌یک از عبارت‌های زیر صحیح هستند. (برای پاسخ خود استدلال بیاورید.)



- $A \perp\!\!\!\perp F \mid \emptyset$ قطعا برقرار است.
- $A \perp\!\!\!\perp D \mid \emptyset$ قطعا برقرار است.
- $A \perp\!\!\!\perp I \mid E$ قطعا برقرار نیست.
- $C \perp\!\!\!\perp G \mid A, I$ قطعا برقرار است.

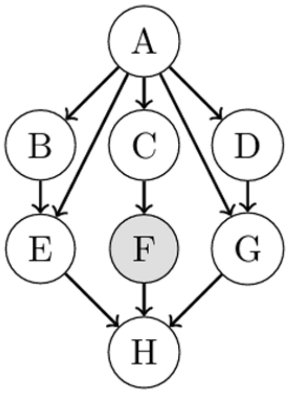
حل.

(آ)
- نادرست. لزوماً استقلال شرطی بین A و E با شرط داشتن G وجود ندارد، چون مسیر $A - C - E$ فعال است.
- درست. A و B مستقل هستند، زیرا مسیرهای $A - D - B$ و $F - G - D$ غیرفعال هستند. پس هیچ مسیر فعالی بین A و B وجود ندارد.
- نادرست. لزوماً استقلال شرطی بین E و G وجود ندارد، چون مسیر $E - F - G$ فعال است.
- درست. A و B به شرط F مستقل هستند، زیرا هر دو مسیر $A - D - B$ و $E - F - G$ غیرفعال هستند. پس هیچ مسیر فعالی بین این دو وجود ندارد.

(ب)
- درست. زیرا مسیرهای $B - C - E$، $A - D - E$ و $G - H - E$ غیرفعال هستند.
- نادرست. زیرا D فرزند مستقیم A است.
- نادرست. مسیر $A - D - G - H - I$ فعال است. بنابراین A و I به شرط E لزوماً مستقل نیستند. اما برای این‌که به طور قطعی بگوییم مستقل نیستند، به CPT های آن‌ها نیاز داریم.
- نادرست. زیرا دو مسیر $C - E - D - G$ و $C - E - H - G$ فعال هستند.

۳. (۲۴ نمره) به پرسش های زیر در مورد الگوریتم Elimination Variable پاسخ دهید.

(آ) در این قسمت باید طبق ترتیب مشخص‌شده در شبکه‌ی بیزی در شکل زیر متغیرها را حذف کنید. تمام متغیرهای این پرسش baniary هستند.



متغیرها را به ترتیبی که در ادامه آمده است حذف کنید. نوشتن به فرمت

$$f_۱(X, +y) = \sum_z p(z \mid +y)\, p(X \mid z).$$

برای بخش های ۱ تا ۷ کافی است.

(۱) پس از در نظر گرفتن شواهد چه فاکتورهایی در ابتدا داریم؟

(۲) متغیر A را حذف کنید تا فاکتور جدید $f_۱$ ساخته شود.

(۳) متغیر B را حذف کنید تا فاکتور جدید $f_۲$ ساخته شود.

(۴) متغیر C را حذف کنید تا فاکتور جدید $f_۳$ ساخته شود.

(۵) متغیر D را حذف کنید تا فاکتور جدید $f_۴$ ساخته شود.

(۶) متغیر E را حذف کنید تا فاکتور جدید $f_۵$ ساخته شود.

(۷) متغیر G را حذف کنید تا فاکتور جدید $f_۶$ ساخته شود.

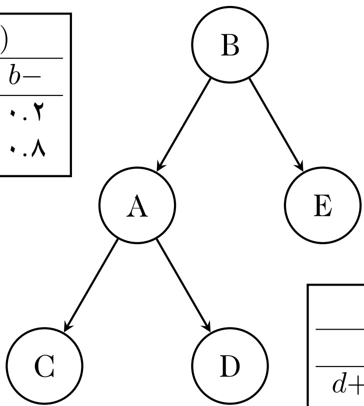
(۸) توضیح دهید به وسیله فاکتور های محاسبه شده چگونه می‌توان $P(+h| + f)$ را محاسبه کرد.

(ب) شبکه‌ی بیزی زیر و جداول احتمالات آن را در نظر بگیرید. جدول احتمال $P(A \mid b+, c-)$ را به روش Elimination به دست آورید.

$P(B)$	
$b+$	۰.۸
$b-$	۰.۲

$P(A \mid B)$	$b+$	$b-$
$a+$	۰.۶	۰.۲
$a-$	۰.۴	۰.۸

$P(E \mid B)$	$b+$	$b-$
$e+$	۰.۲۵	۰.۳
$e-$	۰.۷۵	۰.۷



$P(C \mid A)$	$a+$	$a-$
$c+$	۰.۵	۰.۵
$c-$	۰.۵	۰.۵

$P(D \mid A)$	$a+$	$a-$
$d+$	۰.۷	۰.۹
$d-$	۰.۳	۰.۱

حل.

(آ) (۱)

$$\Big\{p(A), p(B \mid A), p(C \mid A), p(D \mid A), p(E \mid A, B), p(+f \mid C), p(G \mid A, D), p(H \mid E, +f, G)\Big\}.$$

(۲)

$$f_۱(B,C,D,E,G) = \sum_a p(a)\,p(B \mid a)\,p(C \mid a)\,p(D \mid a)\,p(E \mid a,B)\,p(G \mid a,D)$$

(۳)

$$f_۲(C,D,E,G) = \sum_b f_۱(b,C,D,E,G)$$

(۴)

$$f_۳(D,E,+f,G) = \sum_c f_۲(c,D,E,G)\,p(+f \mid C)$$

(۵)

$$f_۴(+f,E,G) = \sum_d f_۳(d,E,+f,G)$$

(۶)

$$f_۵(+f,G,H) = \sum_e f_۴(e,G,+f)\,p(H \mid e,+f,G)$$

(۷)

$$f_۶(+f,H) = \sum_g f_۵(+f,g,H)$$

(۸) برای محاسبهٔ $P(+h \mid +f)$، ابتدا مقادیر $f_۶(+f,+h)$ و $f_۶(+f,-h)$ را از عامل نهایی استخراج می‌کنیم و سپس با نرمالیزه کردن آن‌ها به صورت زیر احتمال شرطی را به دست می‌آوریم:

$$P(+h \mid +f) = \frac{f_۶(+f,+h)}{f_۶(+f,+h) + f_۶(+f,-h)}$$

(ب)

$$P(A \mid +b,-c) \propto P(A,+b,-c) = \sum_{d,e} P(+b)\,P(e \mid +b)\,P(A \mid +b)\,P(-c \mid A)\,P(d \mid A)$$

$$= P(+b)\,P(A \mid +b)\,P(-c \mid A)\left(\sum_d P(d \mid A)\right)\left(\sum_e P(e \mid +b)\right)$$

$$= P(+b)\,P(A \mid +b)\,P(-c \mid A)$$

محاسبهٔ مقادیر عددی با استفاده از جداول:

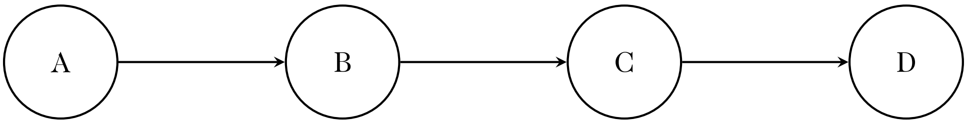
$$P(+a,+b,-c) = P(+b)\;P(+a \mid +b)\;P(-c \mid +a) = ۰/۸ \times ۰/۶ \times ۰/۵ = ۰/۲۴$$

$$P(-a,+b,-c) = ۰/۸ \times ۰/۴ \times ۰/۵ = ۰/۱۶$$

$$P(+a \mid +b,-c) = \frac{۰/۲۴}{۰/۲۴ + ۰/۱۶} = \frac{۰/۲۴}{۰/۴۰} = ۰/۶۰$$

$$P(-a \mid +b,-c) = ۱ - ۰/۶۰ = ۰/۴۰$$

۴. (۲۰ نمره) شبکه‌ی بیزی زیر داده شده است و توزیع‌های مربوط به متغیرهای موجود در شبکه‌ی بیزی نیز مشخص شده‌اند.



$P(A)$	
$-a$	۳/۴
$+a$	۱/۴

$P(B \mid A)$	$+a$	$-a$
$-b$	۴/۵	۲/۳
$+b$	۱/۵	۱/۳

$P(C \mid B)$	$+b$	$-b$
$-c$	۱/۲	۱/۴
$+c$	۱/۲	۳/۴

$P(D \mid C)$	$+c$	$-c$
$-d$	۵/۶	۱/۸
$+d$	۱/۶	۷/۸

(آ) شما نمونه‌های زیر را در اختیار دارید:

$$\begin{array}{cccc} -d & -c & +b & +a \\ -d & +c & -b & +a \\ -d & +c & +b & -a \\ -d & +c & -b & -a \end{array} \qquad \begin{array}{cccc} +d & -c & -b & +a \\ -d & +c & +b & +a \\ +d & -c & +b & -a \\ -d & +c & -b & -a \end{array}$$

(i) فرض کنید این نمونه‌ها با استفاده از روش نمونه‌گیری پیشین (Prior Sampling) به دست آمده‌اند. برآورد نمونه‌ای $\mathrm{P}(+c)$ را محاسبه کنید.

(ii) اکنون می‌خواهیم $\mathrm{P}(+c \mid +a, -d)$ را برآورد کنیم. تمامی نمونه‌هایی را که در روش نمونه‌گیری حذفی (Rejection Sampling) برای این کار استفاده نخواهند شد، مشخص کرده و با علامت‌گذاری مناسب نشان دهید که کدام نمونه‌ها کنار گذاشته می‌شوند. سپس در نهایت $\mathrm{P}(+c \mid +a, -d)$ را برآورد کنید.

(ب) در این بخش، از روش نمونه‌گیری با وزن‌دهی احتمالی (Likelihood Weighting Sampling) برای برآورد $\mathrm{P}(-a \mid +b, -d)$ استفاده کرده‌ایم. نمونه‌های زیر حاصل شده‌اند. وزن مربوطه برای هر نمونه بنویسید.

$$\begin{array}{cccc} -d & +c & +b & -a \\ -d & +c & +b & +a \\ -d & -c & +b & +a \\ -d & -c & +b & -a \end{array}$$

(ج) با استفاده از نمونه‌های وزنی در سؤال قبل، $\mathrm{P}(-a \mid +b, -d)$ را برآورد کنید.

(د) استفاده از روش نمونه‌گیری با وزن‌دهی احتمالی (Likelihood Weighting Sampling) برای محاسبه‌ی $P(A|D)$ مناسب تر است یا $P(D|A)$؟ دلیل خود را توضیح دهید.

(ه) فرض کنید تنها شواهد موجود $A = +a$ باشد. مشخص کنید کدام یک از توالی‌های زیر می‌تواند توسط نمونه‌گیری گیبس تولید شود، دلایل خود را توضیح دهید.(به خاطر داشته باشید که در طی نمونه‌گیری گیبس ، نمونه‌ها از طریق یک پروسه‌ی iterative تولید می‌شوند.)

توالی ۱

۱: $+a \quad -b \quad -c \quad +d$

۲: $+a \quad -b \quad -c \quad +d$

۳: $+a \quad -b \quad +c \quad +d$

توالی ۲

۱: $+a \quad -b \quad -c \quad +d$

۲: $+a \quad -b \quad -c \quad -d$

۳: $-a \quad -b \quad -c \quad +d$

توالی ۳

۱: $+a \quad -b \quad -c \quad +d$
۲: $+a \quad -b \quad -c \quad -d$
۳: $+a \quad +b \quad -c \quad -d$

توالی ۴

۱: $+a \quad -b \quad -c \quad +d$
۲: $+a \quad -b \quad -c \quad -d$
۳: $+a \quad +b \quad -c \quad +d$

حل.

(آ) (i) از میان ۸ نمونهٔ داده‌شده، ۵ نمونه دارای $+c$ هستند، بنابراین

$$P(+c) = \frac{5}{8}$$

(ii) ابتدا تمام نمونه‌هایی را که با شواهد $\{A = +a, D = -d\}$ سازگار نیستند، حذف می‌کنیم از میان ۳ نمونهٔ باقیمانده، ۲ نمونه دارای $+c$ هستند، پس

$$P(+c \mid +a, -d) = \frac{2}{3}$$

(ب) نمونه‌های تولیدشده و وزن آن‌ها به شرح زیر است:

شماره نمونه	A	B	C	D	وزن
۱	$-a$	$+b$	$+c$	$-d$	$P(+b \mid -a)\, P(-d \mid +c) = \frac{1}{3} \times \frac{5}{6} = \frac{5}{18} \approx 0{,}277$
۲	$+a$	$+b$	$+c$	$-d$	$P(+b \mid +a)\, P(-d \mid +c) = \frac{1}{5} \times \frac{5}{6} = \frac{1}{6} \approx 0{,}167$
۳	$+a$	$+b$	$-c$	$-d$	$P(+b \mid +a)\, P(-d \mid -c) = \frac{1}{5} \times \frac{1}{8} = \frac{1}{40} = 0{,}025$
۴	$-a$	$+b$	$-c$	$-d$	$P(+b \mid -a)\, P(-d \mid -c) = \frac{1}{3} \times \frac{1}{8} = \frac{1}{24} \approx 0{,}042$

(ج) با توجه به نمونه های بخش ب داریم:

$$P(-a \mid +b, -d) = \frac{\frac{5}{18} + \frac{1}{24}}{\frac{5}{18} + \frac{1}{6} + \frac{1}{40} + \frac{1}{24}} \approx 0{,}625.$$

(د)

$$\boxed{P(D \mid A)}$$

چون در وزن‌دهی احتمال (probability weighting) تنها می‌توانیم بر شواهد «بالادستی» (evidence) شرط بگیریم. در این حالت، متغیر A در بالادستی D قرار دارد و بنابراین فقط می‌توانیم احتمال شرطی $P(D \mid A)$ را محاسبه کنیم. دقت کنید، برای محاسبه احتمال شرطی $P(D \mid A)$ هیچ‌گونه شواهدی از متغیرهای پایین‌دستی D که ممکن است بر آن تأثیر بگذارند، در محاسبات در نظر گرفته نمی‌شود.

(ه) در Gibbs Sampling تنها یک متغیر غیرویژه در هر گام تغییر می‌کند و متغیرهای شاهد ثابت می‌مانند. بنابراین دنباله‌های ۱ و ۳ مجاز اند.

$+d \quad -c \quad -b \quad +a$
دنباله ۱: $+d \quad -c \quad -b \quad +a$
$+d \quad +c \quad -b \quad +a$

در این دنباله تنها یک متغیر (C) بین ردیف‌های دوم و سوم تغییر کرده است، که طبق قاعده Gibbs Sampling تغییر تنها یک متغیر در هر گام مجاز است.

$+d \quad -c \quad -b \quad +a$
دنباله ۲: $-d \quad -c \quad -b \quad +a$
$+d \quad -c \quad -b \quad -a$

در این دنباله دو متغیر (A و D) بین ردیف‌های دوم و سوم تغییر کرده‌اند که خلاف قاعده Gibbs Sampling است که فقط یک متغیر باید در هر گام تغییر کند.

دنباله ۳:

$$\begin{array}{cccc} +d & -c & -b & +a \\ -d & -c & -b & +a \\ -d & -c & +b & +a \end{array}$$

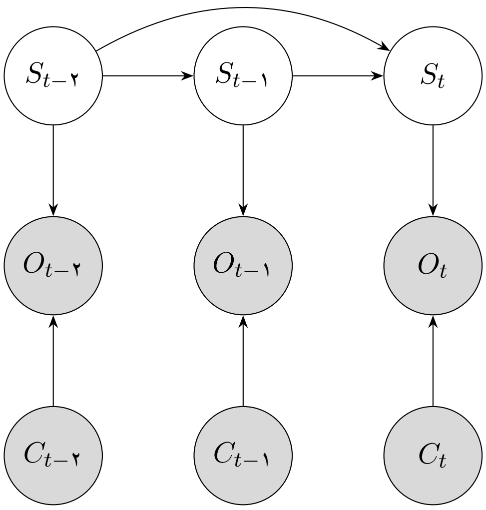
این دنباله مجاز است زیرا تنها یک متغیر (B) بین ردیف‌های دوم و سوم تغییر کرده است و سایر متغیرها ثابت باقی مانده‌اند.

دنباله ۴:

$$\begin{array}{cccc} +d & -c & -b & +a \\ -d & -c & -b & +a \\ +d & -c & +b & +a \end{array}$$

در این دنباله دو متغیر (D و B) بین ردیف‌های دوم و سوم تغییر کرده‌اند که بر خلاف قاعده Gibbs Sampling است.

۵. (۲۰ نمره) پکمن در تلاش است تا یک روح را در راهرویی بی‌نهایت با موقعیت‌هایی مانند تصویر زیر شکار کند. او از حسگرهایی که نصب کرده استفاده می‌کند تا موقعیت واقعی روح یعنی S_t را پیدا کند. در هر گام زمانی، حسگرها مقدار نویزی از موقعیت روح به نام O_t می‌دهند. اما روح هم پیشرفت کرده و می‌تواند در هر زمان خودش را نامرئی کند که با C_t نمایش داده می‌شود و نویز بیشتری به داده‌های حسگر اضافه می‌کند.



مدل خطای حسگر در جدول زیر آمده است که بسته به اینکه روح نامرئی باشد یا نه تغییر می‌کند. همچنین پکمن یک مدل پویایی برای حرکت روح طراحی کرده که موقعیت‌های روح در دو گام زمانی گذشته را در نظر می‌گیرد.

Dynamics model:

$$P(S_t \mid S_{t-۱}, S_{t-۲}) = F(D_۱, D_۲)$$

$$D_۱ = |S_t - S_{t-۱}| \quad D_۲ = |S_t - S_{t-۲}|$$

$D_۲$	$D_۱$	$F(D_۱, D_۲)$
0	0	0.7
1	0	0.2
2	0	0.0
0	1	0.3
1	1	0.3
2	1	0.5

Observation model:

$$P(O_t \mid S_t, C_t) = E(C, D)$$

$$D = |O_t - S_t|$$

D	C	$E(C, D)$
0	+	0.4
1	+	0.2
2	+	0.1
0	−	0.6
1	−	0.2
2	−	0.0

(آ) فرض کنید در حال حاضر دو ذره داریم:

- $S_6 = 7, S_7 = 8$
- $S_6 = 6, S_7 = 6$

با فرض مشاهدات $C_6 = +$، $C_7 = -$، $O_6 = 5$، $O_7 = 8$، وزن هر ذره را محاسبه کنید.

(ب) فرض کنید پکمن دیگر نمی‌تواند ببیند که آیا روح نامرئی شده است یا نه، اما فرض می‌کند که در هر گام زمانی با احتمال ۰.۵ نامرئی است. وزن هر ذره را با فرض مشاهدات بالا محاسبه کنید.

(ج) برای جلوگیری از انتشار خطا، فرض کنید پس از وزن‌دهی و نمونه‌گیری مجدد، یک ذره جدید داریم: $(S_6 = 6, S_7 = 7)$

(i) احتمال اینکه پس از اعمال مدل پویایی به این ذره به $(S_7 = 6, S_8 = 6)$ برسیم چقدر است؟ (۲ نمره)

(ii) احتمال اینکه به $(S_7 = 7, S_8 = 8)$ برسیم چقدر است؟ (۲ نمره)

(د) فرض کنید سه ذره با وزن‌های مشخص در اختیار داریم:

Particle	weight
$(S_۷ = ۵, S_۸ = ۶)$	0.1
$(S_۷ = ۷, S_۸ = ۶)$	0.25
$(S_۷ = ۷, S_۸ = ۷)$	0.3

با توجه به این ذرات، باور پکمن نسبت به موقعیت روح در زمان $t = 8$ را محاسبه کنید.

حل.

(آ) برای هر ذره، احتمال مشاهده‌ها را با استفاده از مدل مشاهده محاسبه می‌کنیم:

- $(S_6 = 7, S_7 = 8)$:

$$P(O_۶ = ۵, C_۶ = + | S_۶ = ۷) = P(C_۶ = +) \times P(O_۶ = ۵ | S_۶ = ۷, C_۶ = +)$$
$$= ۱ \times ۰/۱ = ۰/۱$$
$$P(O_۷ = ۸, C_۷ = - | S_۷ = ۸) = ۱ \times ۰/۶ = ۰/۶$$
$$\Rightarrow \text{وزن} = ۰/۱ \times ۰/۶ = ۰/۰۶$$

• $(S_6 = 6, S_7 = 6)$:

$$P(O_۶ = ۵, C_۶ = +|S_۶ = ۶) = ۱ \times ۰٫۲ = ۰٫۲$$
$$P(O_۷ = ۸, C_۷ = -|S_۷ = ۶) = ۱ \times ۰ = ۰$$
$$\Rightarrow \text{وزن} = ۰٫۲ \times ۰ = ۰$$

(ب) اکنون باید احتمال مشاهده را با فرض عدم مشاهده C_t و در نظر گرفتن احتمال $P(C = +) = P(C = -) = 0.5$ محاسبه کنیم:

• $(S_6 = 7, S_7 = 8)$:

$$P(O_۶ = ۵|S_۶ = ۷) = \sum_{c_۶} P(C_۶ = c_۶) \times P(O_۶ = ۵|S_۶ = ۷, C_۶ = c_۶)$$
$$= ۰٫۵ \times ۰٫۱ + ۰٫۵ \times ۰٫۰ = ۰٫۰۵$$
$$P(O_۷ = ۸|S_۷ = ۸) = ۰٫۵ \times ۰٫۶ + ۰٫۵ \times ۰٫۰ = ۰٫۳$$
$$\Rightarrow \text{وزن} = ۰٫۰۵ \times ۰٫۳ = ۰٫۰۱۵$$

• $(S_6 = 6, S_7 = 6)$:

$$P(O_۶ = ۵|S_۶ = ۶) = ۰٫۵ \times ۰٫۲ + ۰٫۵ \times ۰٫۰ = ۰٫۱$$
$$P(O_۷ = ۸|S_۷ = ۶) = ۰٫۵ \times ۰٫۰ + ۰٫۵ \times ۰٫۰ = ۰$$
$$\Rightarrow \text{وزن} = ۰٫۱ \times ۰ = ۰$$

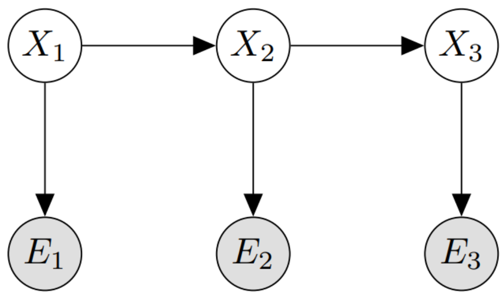
(ج) (i) $P(S_7 = 7, S_8 = 6|S_6 = 6, S_7 = 7) = 0$ چون با مدل پویایی مشخص است که برای $D_1 = 1, D_2 = 1$ احتمال رفتن به $S_8 = 6$ صفر است.

(ii) $P(S_8 = 8|S_6 = 6, S_7 = 7) = F(D_1 = 1, D_2 = 2) = 0.5$

(د) با استفاده از وزن‌های ذرات موجود:

موقعیت	احتمال $P(S_۸)$
۵	۰
۶	$\frac{۰٫۱+۰٫۲۵}{۰٫۱+۰٫۲۵+۰٫۳} = \frac{۰٫۳۵}{۰٫۶۵} = \frac{۷}{۱۳}$
۷	$\frac{۰٫۳}{۰٫۶۵} = \frac{۶}{۱۳}$
۸	۰

۶. (۱۶ نمره) ساختار HMM زیر را در نظر بگیرید.



می‌دانیم الگوریتم Forward یک الگوریتم دو مرحله‌ای بازگشتی است که برای تقریب توزیع احتمال $P(X_t|e_1, \dots, e_t)$ استفاده می‌شود. دو مرحله این الگوریتم به صورت زیر هستند:

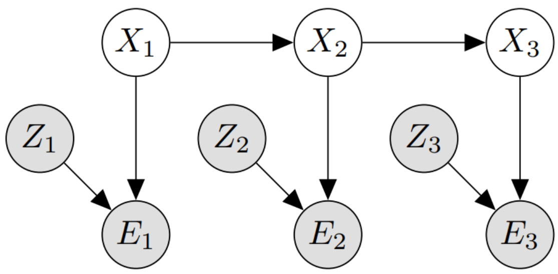
- مرحله Elapse Time

$$P(X_t|e_1, \dots, e_{t-1}) = \sum_{x_{t-1}} P(X_t|x_{t-1}) P(x_{t-1}|e_1, \dots, e_{t-1})$$

- مرحله Observe

$$P(X_t|e_1, \dots, e_t) = \frac{P(e_t|X_t) P(X_t|e_1, \dots, e_{t-1})}{\sum_{x_t} P(e_t|x_t) P(x_t|e_1, \dots, e_{t-1})}$$

می‌خواهیم با اعمال تغییراتی روی ساختار گراف HMM الگوریتم Forward را تغییر دهیم. هدف ما همچنان طراحی الگوریتمی بازگشتی است که بتواند توزیع متغیرهای پنهان X_t را با توجه به تمام حالت‌های evidence موجود از زمان ۰ تا t محاسبه کند.
گراف زیر را در نظر بگیرید که در آن متغیرهای مشاهده‌شده جدیدی با نام Z_i معرفی شده‌اند و روی حالت‌های evidence تأثیر می‌گذارند.



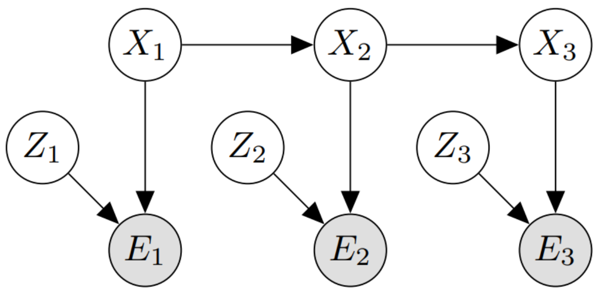
(آ) معادله‌ی تغییر یافته برای Elapse Time را بیان کنید:

$$P(X_t|e_1, \dots, e_{t-1}, z_1, \dots, z_{t-1}) = \dots$$

(ب) معادله‌ی تغییر یافته برای Observe را بیان کنید:

$$P(X_t|e_1, \dots, e_t, z_1, \dots, z_t) = \dots$$

حالا گراف زیر را در نظر بگیرید که متغیرهای Z_i مشاهده نشده‌اند.



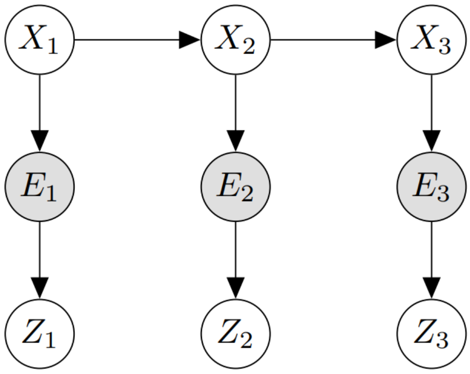
(ج) معادله‌ی تغییر یافته برای Elapse Time را بیان کنید:

$P(X_t|e_1, \ldots, e_{t-1}) = \ldots$

(د) معادله‌ی تغییر یافته برای Observe را بیان کنید:

$P(X_t|e_1, \ldots, e_t) = \ldots$

حالا گراف زیر را در نظر بگیرید که متغیرهای معرفی‌شده مشاهده نشده‌اند و حالت‌های evidence روی آن‌ها تاثیر می‌گذارند.



(ه) معادله‌ی تغییر یافته برای Elapse Time را بیان کنید:

$P(X_t|e_1, \ldots, e_{t-1}) = \ldots$

(و) معادله‌ی تغییر یافته برای Observe را بیان کنید:

$P(X_t|e_1, \ldots, e_t) = \ldots$

حل.

(آ)

$$P(X_t|e_۱, \ldots, e_{t-۱}, z_۱, \ldots, z_{t-۱}) = \sum_{x_{t-۱}} P(X_t \mid x_{t-۱}) P(x_{t-۱} \mid e_{۱:t-۱}, z_{۱:t-۱})$$

(ب)

$$P(X_t \mid e_۱, \ldots, e_t, z_۱, \ldots, z_t) = \frac{P(X_t \mid e_{۱:t-۱}, z_{۱:t-۱}) P(z_t) P(e_t \mid z_t, X_t)}{\sum_{x_t} P(x_t \mid e_{۱:t-۱}, z_{۱:t-۱}) P(z_t) P(e_t \mid z_t, x_t)}$$
$$= \frac{P(X_t \mid e_{۱:t-۱}, z_{۱:t-۱}) P(e_t \mid z_t, X_t)}{\sum_{x_t} P(x_t \mid e_{۱:t-۱}, z_{۱:t-۱}) P(e_t \mid z_t, x_t)}$$

در اینجا باید متغیرهای جدید Z_i را در نظر بگیریم. از آنجایی که این متغیرها مشاهده شده‌اند، فرض می‌کنیم که متغیر Z_t دارای مقدار z_t است.

(ج)

$$P(X_t \mid e_۱, \dots, e_{t-۱}) = \sum_{x_{t-۱}} P(X_t \mid x_{t-۱}) P(x_{t-۱} \mid e_{۱:t-۱})$$

متغیرهای Z_i تاثیری بر مرحله‌ی Elapse Time ندارند.

(د)

$$P(X_t \mid e_۱, \dots, e_t) = \frac{P(X_t \mid e_{۱:t-۱}) \sum_{z_t} P(z_t) P(e_t \mid z_t, X_t)}{\sum_{x_t} P(x_t \mid e_{۱:t-۱}) \sum_{z_t} P(z_t) P(e_t \mid z_t, x_t)}$$

در اینجا نیاز داریم که متغیرهای جدید Z_i را در نظر بگیریم.

(ه)

$$P(X_t \mid e_۱, \dots, e_{t-۱}) = \sum_{x_{t-۱}} P(X_t \mid x_{t-۱}) P(x_{t-۱} \mid e_{۱:t-۱})$$

هیچ تغییری رخ نمی‌دهد زیرا متغیرهای Z_i مستقل از X_i اگر E_i را داشته باشیم هستند.

(و)

$$P(X_t \mid e_۱, \dots, e_t) = \frac{P(X_t \mid e_{۱:t-۱}) P(e_t \mid X_t)}{\sum_{x_t} P(x_t \mid e_{۱:t-۱}) P(e_t \mid x_t)}$$

هیچ تغییری رخ نمی‌دهد زیرا متغیرهای Z_i مستقل از X_i اگر E_i را داشته باشیم هستند.