Proceedings of the
47th Annual Iranian Mathematics Conference,
Kharazmi University, Karaj, Iran, 28-31 August 2016

مجموعهٔ مقالات چهل و هفتمین کنفرانس ریاضی ایران
دانشگاه خوارزمی، کرج، ۱۰-۷ شهریور ۱۳۹۵

فرمول بندی تجرید حالت در یادگیری تقویتی چند و ظیفه‌ای

ناهید طاهریان* و سمیه عربی
دانشگاه خوارزمی، taherian@khu.ac.ir
دانشگاه خوارزمی، s.arabi@khu.ac.ir

چکیده

تجرید حالت روشی قدرتمند برای مدیریت زمان و حافظه در یادگیری تقویتی هست. در رویکردهای قدیمی این روش برای افزایش سرعت یادگیری وظیفه فعلی مورد استفاده قرار می‌گیرد. در بیشتر این موارد، هزینه یافتن تجرید مناسب و سپس بهره‌گیری از آن در برابر سود حاصله مقرون به صرفه نیست. با این وجود، وقتی قرار است چندین محیط مشابه از یک حوزه یاد گرفته شوند، از این روش برای بهبود یادگیری در وظایفِ دیگرِ حوزه و در نتیجه افزایش سود در برابر هزینه می‌توان بهره برد. در این مقاله یک فرمول بندی برای شناساییِ متغیرهای حالت که در بیشتر وظایف حوزه تأثیری در یادگیری ندارند ارائه می‌شود تا با حذف آن‌ها در وظایف بعدی، سرعت یادگیری با پذیرش خطایی کنترل شده افزایش یابد.

واژه‌های کلیدی: یادگیری تقویتی چند وظیفه ای، تجرید حالت، فرایند مارکوف فاکتور شده.

رده‌بندی موضوعی ریاضی (2010): 68T05، 68T05.

۱ مقدمه

یادگیری انتقال دانش به معنای بهره‌گیری از دانش یاد گرفته شده در وظایف مبدأ جهت بهبود یادگیری در وظایف هدف مرتبط، اما متفاوت است. در برخی از تحقیقات تلاش شده است تا با انتقال تجرید حالت به سایر وظایف بر مشکلات تجرید حالت سنتی غلبه شود و بدین ترتیب هزینه کشف وظیفه مبدأ با سود حاصله بیشتر جبران گردد. والش و همکاران [۴] تلاش کرده اند تا از دانش حاصل از تعداد زیادی از وظایف از قبل یاد گرفته شده از حوزه وظایف بهره برده و با استفاده از این دانش اولیه فرایند تجرید در وظایف بعدی را سرعت ببخشند. نویسندگان یک مدل بسیار مفید و انعطاف‌پذیر به نام الگوریتم عمومی انتقال تجرید GATA ارائه داده اند و آن را با مقداری بحث‌های تئوری پشتیبانی کرده اند. با این وجود، مبانی تئوری و آزمایش‌های تجربی ارائه شده ناکافی و بسیار ابتدایی است. در این مقاله مدل GATA توسعه داده می‌شود و مبانی تئوری لازم نیز ارائه می‌گردد.

یک حوزه چندوظیفه‌ای را در نظر بگیرید به‌گونه‌ای که تمام وظایف متعلق به یک توزیع D بوده و تمام آن‌ها فرایندهای مارکوف فاکتور شده با مجموعه متغیرهای حالت و مجموعه اعمال کاملا یکسان اما توابع پاداش و گذر دلخواه باشند. اگر به تعداد m فرایند مارکوف نمونه از D قبلا حل شده باشند و سیاست‌های بهینه یا مقادیر ارزش بهینه آن‌ها به‌دست آمده باشند، می‌توان از این دانشِ حاصل از مجموعه نمونه از وظایف برای انجام نوعی تجرید در وظیفه هدف بهره گرفت و بدین ترتیب سرعت یادگیری آن را افزایش داد.

*سخنران و مسئول مکاتبات

یکی از ساده‌ترین شکل‌های تجرید این است که فضای حالت را به یک زیرفضا با بُعد کمتر تصویر کنیم. این کار معادل با این است که از برخی متغیرهای حالت صرف‌نظر کنیم. این دقیقا همان کاری است که راویندران [۳] به‌عنوان هomومورفیسم تصویر ساده برای یافتن فرایند مارکوف مجرد از یک فرایند مارکوف فاکتور شده با مدل کاملا معلوم ارائه داده است. هرچند در الگوریتم ارائه شده توسط راویندران، دو مرحله تجرید حالت و مرحله انتخاب متغیرهای مرتبط با هم ترکیب شده اند و بطور هم‌زمان انجام می‌شوند، اما روش مذکور تنها برای فرایندهای مارکوف با مدل بیزی کاملا معلوم قابل استفاده است. در این مقاله، این دو مرحله از یکدیگر مجزا شده‌اند و تلاش شده است تا الگوریتمی ارائه گردد که برای مواردی که مدل فرایند مارکوف هدف شناخته شده نیست هم قابل استفاده باشد. در روش پیشنهادی، زیرمجموعه‌ای از متغیرهای مرتبط از مجموعه نمونه از وظایف استخراج می‌شوند و به‌عنوان دانش اولیه در اختیار وظیفه هدف قرار می‌گیرند؛ سپس با استفاده از این دانش، فضای حالت در وظیفه هدف به زیرفضا با بعد کمتر تصویر می‌شود و بدین ترتیب تجرید بدون هیچ هزینه برخطی صورت می‌گیرد. برای وظایف مبدأ نمونه با مدل معلوم، ابتدا تجرید حالت بر مبنای یک معیار مثل مقادیر ارزش Q بهینه یا سیاست‌های بهینه یا معیارهای دیگر صورت می‌گیرد؛ سپس از هریک از این تجریدها، زیرمجموعه‌ای از متغیرهای حالت به‌عنوان متغیرهای مرتبط استخراج می‌شوند؛ و در نهایت یک زیرمجموعه نهایی از متغیرها، به عنوان مثال اجتماع تمام متغیرهای استخراج شده از وظایف مبدأ نمونه، به‌عنوان دانش اولیه در اختیار وظیفه هدف قرار می‌گیرند.

۲ فرمول بندی مسئله و نتایج اصلی

یک فرایند مارکوف فاکتور شده، یک فرایند مارکوف است که در آن فضای حالت بصورت حاصلضرب کارتزین حوزه‌های یک مجموعه متناهی از متغیرهای تصادفی $Z = \{S_1, ..., S_n\}$ تعریف شده است.

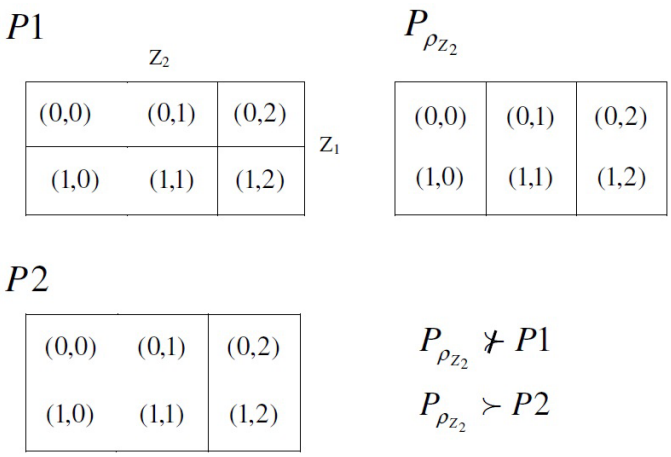
تعریف ۱.۲. یک مجموعه فاکتور شده S در یک فضای N بُعدی را با مجموعه متغیرهای Z در نظر بگیرید. اگر S^j حوزه مربوط به متغیر Z_j باشد، Z_j -تصویر[۱] S یک نگاشت $\rho_{Z_j} : S \to S^j$ است که بصورت $\rho_{Z_j}(\vec{z}) = z_j$ تعریف شده و $\vec{z} = \langle z_1, \cdots, z_N \rangle$.

برای یک زیرمجموعه از متغیرها $X \subseteq Z$ ، می‌توان تعریف بالا را توسعه داد. X-تصویرِ S، تصویر S بروی زیرفضای حاصل از متغیرهای X است. این را می‌توان با نگاشت $\rho_X : S \to \times_{(Z_j \in X)} S^j$ که بصورت $\rho_X = \times_{(Z_j \in X)} \rho_{Z_j}$ تعریف شده، نشان داد.

تعریف ۲.۲. هر تابع ϕ با دامنه S یک رابطه هم‌ارزی روی S تعریف می‌کند بطوری‌که $s_1 \equiv_\phi s_2$ اگر و تنها اگر $\phi(s_1) = \phi(s_2)$. می‌گوییم s_1 و s_2، ϕ-هم‌ارز هستند اگر $s_1 \equiv_\phi s_2$. افراز حاصل از این رابطه هم‌ارزی روی S را با P_ϕ نمایش می‌دهیم [۳].

تعریف ۳.۲. دو افراز P و P' روی مجموعه S را در نظر بگیرید. به یاد آورید که یک افراز S عبارتست از تقسیم S به "کلاس‌ها" یا "بلوک‌ها"ی ناتهی که با هم همپوشانی نداشته باشند. می‌گوییم افراز P یک تظریفِ[۲] افراز P' است و آن را با $P \succeq P'$ نمایش می‌دهیم اگر و تنها اگر هر بلوک از P زیرمجموعه‌ای از بلوکی از P' باشد. در این حالت می‌گوییم P ظریف‌تر از P' است [۱]. اگر علاوه بر این داشته باشیم $P \neq P'$، آنگاه P اکیدا ظریف‌تر از P' است و با $P \succ P'$ نمایش داده می‌شود. هم‌چنین گفته می‌شود که P' (اکیدا) درشت‌تر از P است و با $P' \preceq P$ ($P' \prec P$) نشان داده می‌شود[۲].

[۱] $Z_j - projection$
[۲] Refinement



شکل ۱: یک مثال از متغیر نامرتبط؛ Z_1 نسبت به افراز $P1$ نامرتبط نیست اما نسبت به افراز $P2$ نامرتبط است.

تعریف ۴.۲. یک افراز P روی مجموعه فاکتور شده S در نظر بگیرید. فرض کنید Z مجموعه تمام متغیرهای S و Y زیرمجموعه‌ای از متغیرها و $X = Z - Y$ باشند. می‌گوییم Y نسبت به افراز P نامرتبط است (P-نامرتبط[۳]) اگر P_{ρ_X}، افراز معرفی شده بوسیله X-تصویر S، ظریف‌تر از P باشد. در این حالت می‌گوییم X، P-سازگار[۴] است.
یک مجموعه P-سازگار ماکسیمال است اگر افزودن هر متغیر دلخواهی به آن منجر به یک مجموعه غیرِ P-نامرتبط شود. یک مجموعه P-سازگار مینیمال است اگر حذف هر یک از متغیرهای آن منجر به یک مجموعه غیرِ P-سازگار شود.

این بدان معناست که افراز حاصل از صرف نظر کردن از متغیرهای مجموعه Y در افراز P می‌نشیند و در نتیجه محدودیت‌های شدیدتری (اطلاعات جزئی‌تری) نسبت به P دارد، اما از محدودیت‌های (بلوک‌های) P تجاوز نمی‌کند. برای روشن شدن بیشتر مطلب، یک مجموعه دو بعدی S با متغیرهای Z_1 و Z_2 را در نظر بگیرید. فرض کنید $Z_1 \in \{0,1\}$ و $Z_2 \in \{0,1,2\}$. دو افراز $P1 = \{\{(0,0),(0,1)\},\{(0,2)\},\{(1,0),(1,1)\},\{(1,2)\}\}$ و $P2 = \{\{(0,0),(0,1),(1,0),(1,1)\},\{(0,2),(1,2)\}\}$ روی S را در نظر بگیرید.

اکنون اگر از محدودیت‌های ایجاد شده توسط متغیر Z_1 صرف‌نظر شود، افراز حاصله روی مجموعه S، افراز $P_{\rho_{Z_2}} = \{\{(0,0),(1,0)\},\{(0,1)(1,1)\},\{(0,2),(1,2)\}\}$ خواهد بود. این افراز برخی از محدودیت‌های افراز $P1$ را نقض می کند؛ بنابراین Z_1 نسبت به $P1$ نامرتبط نیست. اما مشاهده می‌شود که با صرف نظر کردن از Z_1 تعدادی از محدودیت‌های اضافی که در $P2$ موجود نبودند، از بین می‌روند و هر چند افراز حاصله هم‌چنان محدودیت‌های اضافه‌تر و جزئی‌تری نسبت به $P2$ دارد اما از $P2$ تجاوز نمی‌کند و محدودیت‌های آن را نقض نمی‌کند؛ بنابراین Z_1 نسبت به $P2$ نامرتبط است (شکل ۱ را ببینید.)

هم‌چنان که از تعداد بیشتر و بیشتری از متغیرها صرف نظر شود، افراز حاصله درشت‌تر و درشت‌تر می‌شود. علت این است که بُعد فضای تصویر کاهش می‌یابد. این فرایند را باید تا جایی ادامه داد که افراز حاصله از افراز P تجاوز نکند. با این کار محدودیت‌های (اطلاعات) اضافی نسبت به محدودیت‌های P که متغیرها ایجاد می‌کنند از بین می‌روند و کنار گذاشته می‌شوند. برای یافتن مجموعه متغیرهای نامرتبط، والش و همکاران [۴] یک الگوریتم ابتکاری افزایشی ارائه می‌دهند که در اینجا با استفاده از تعریف های دقیق صورت گرفته دوباره ساختاردهی و ارائه شده است (شکل ۲).

به‌وضوح، ممکن است بیشتر از یک مجموعه از متغیرهای نامرتبط نسبت به یک افراز وجود داشته باشد که برخی از آن‌ها ماکسیمال هستند. این الگوریتم یک ترتیب روی مجموعه متغیرها در نظر می‌گیرد و همیشه یک مجموعه ماکسیمال یکسان از

[۳] P-irrelevant
[۴] P-consistent

1. Set the set of irrelevant features as empty, $Y = \emptyset$.
2. Set $Z1 = Z$, the set of all features.
3. Set $stop = false$.
4. While $Z1 \neq \emptyset$ && $stop == false$,
   (a) Choose a feature F from the set of features $Z1$ according to the ordering defined on features.
   (b) $Y1 = Y \cup \{F\}$.
   (c) $X = Z - Y1$.
   (d) if P_{ρ_X} is finer than P then $stop = true$ and $Y = Y1$.
   (e) $Z1 = Z1 - \{F\}$.
5. If $stop == true$ then $Z1 = Z - Y$. and go to step 3.
6. Print Y as the set of irrelevant features.

شکل ۲: الگوریتم افزایشی والش برای یافتن متغیرهای نامرتبط نسبت به افراز P روی مجموعه S

متغیرهای نامرتبط برای هر وظیفه پیدا می‌کند. تضمینی وجود ندارد که الگوریتم بهترین مجموعه از متغیرهای نامرتبط را بیابد. بهترین مجموعه، آن مجموعه‌ای است که صرف‌نظر کردن از آن منجر به درشت‌ترین افراز روی S و در نتیجه نزدیک‌ترین آن‌ها به افراز P شود.

تعریف ۵.۲. دو فرایند مارکوف $M = \langle S, A, \Psi, T, R\rangle$ و $\bar{M} = \langle \bar{S}, A, \bar{\Psi}, \bar{T}, \bar{R}\rangle$ را در نظر بگیرید. می‌گوییم یک تابع پوشای $\phi : S \to \bar{S}$ یک تابع تجرید[۵] از فرایند مارکوف زمینه M به فرایند مارکوف مجرد $\bar{M}$ است اگر یک تابع وزن $w : S \to [0, 1]$ وجود داشته باشد بطوری که برای هر $\bar{s} \in \bar{S}$ داشته باشیم $\sum_{s\in\phi^{-1}(\bar{s})} w(s) = 1$ و

$$\bar{R}(\bar{s}, a) = \sum_{s\in\phi^{-1}(\bar{s})} w(s)R(s, a); \quad \bar{T}(\bar{s}, \bar{s'}, a) = \sum_{s\in\phi^{-1}(\bar{s})} \sum_{s'\in\phi^{-1}(\bar{s'})} w(s)T(s, a, s').[۲]$$

تعریف ۶.۲. فرض کنید ϕ یک تابع تجرید روی فضای حالت یک فرایند مارکوف فاکتور شده با متغیرهای حالت Z باشد. می‌گوییم زیرمجموعه Y از متغیرها نسبت به ϕ نامرتبط است (ϕ - نامرتبط) اگر و تنها اگر نسبت به P_ϕ، افراز حاصله از تجرید، نامرتبط باشد. در این حالت، می‌گوییم مجموعه $X = Z - Y$ ، ϕ - سازگار است.

قضیه ۷.۲. یک ساختار چند-وظیفه‌ای با وظیفه هدف t و m وظیفه مبدأ نمونه را در نظر بگیرید. فرض کنید ϕ_t و ϕ_i به‌ترتیب تجریدهایی روی وظیفه هدف و i اُمین وظیفه مبدأ نمونه باشند. هم‌چنین فرض کنید Y^t مجموعه ϕ_t- نامرتبط ماکسیمالی باشد که با الگوریتم والش بدست آمده است. فرض کنید از بین N متغیر حالت، K تا از آن‌ها X_j؛ $j = 1..K$ در مجموعه مکمل Y^t باشند که با $X^t = Z - Y^t$ نشان داده می‌شوند. به‌وضوح X^t یک مجموعه ϕ_t - سازگار مینیمال است. هم‌چنین، فرض کنید Y^{s_i} و X^{s_i} به‌ترتیب مجموعه ϕ_i- نامرتبط ماکسیمال و مجموعه ϕ_i - سازگار مینیمال باشند که با الگوریتم والش به‌دست آمده‌اند. فرض کنید برای هر متغیر $X_j \in X^t$، چگالی آن متغیر در مجموعه‌های تجرید-سازگار مینیمال ساختار چند-وظیفه‌ای بزرگ‌تر از صفر است. به‌عبارت دیگر $Pr(X_j \in X^{s_i}) = p_j > 0$. فرض کنید E^m_{miss} رخداد نبودن حداقل یکی از متغیرهای X^t در اجتماع مجموعه‌های تجرید-سازگار m وظیفه مبدأ نمونه باشد؛ $E^m_{miss} = \exists X_j \in X^t$ بگونه‌ای که $X_j \notin \bigcup_{i=1}^m X^{s_i}$.

هم‌چنان که m، تعداد وظایف مبدأ نمونه، به بی‌نهایت میل می‌کند، احتمال رخداد E^m_{miss} به صفر همگرا می‌شود. به‌علاوه، بعد

[۵] Abstraction function

از مشاهده m وظیفه نمونه، با احتمال حداقل $1 - \sum_{j=1}^{K}(1-p_j)^m$ می‌توان مطمئن بود که یک مجموعه ϕ_t - سازگار برای وظیفه هدف بدست آید.

برهان. برای هر $X_j \in X^t; j = 1..K$، فرض کنید $\{E_j^m\}$ دنباله رخدادهایی باشد که بصورت $E_j^m = X_j \in \bigcup_{i=1}^{m} X^{s_i}$ تعریف شده‌اند. داریم:

$$Pr(E_j^m) = Pr(X_j \in X^{s_1} \cup (X^{s_2} - X^{s_1}) \cup (X^{s_3} - (X^{s_1} \cup X^{s_2})) \cup \cdots) =$$

$$Pr(X_j \in X^{s_1}) + Pr(X_j \in X^{s_2} - X^{s_1}) + Pr(X_j \in X^{s_3} - (X^{s_1} \cup X^{s_2})) + \cdots =$$

$$Pr(X_j \in X^{s_1}) + Pr(X_j \in X^{s_2} \wedge X_j \notin X^{s_1}) + Pr(X_j \in X^{s_3} \wedge X_j \notin X^{s_1} \wedge X_j \notin X^{s_2}) + \cdots =$$

$$p_j + p_j(1-p_j) + p_j(1-p_j)^2 + \cdots + p_j(1-p_j)^{m-1} =$$

$$p_j \frac{1-(1-p_j)^m}{1-(1-p_j)} = 1-(1-p_j)^m$$

$$So, \lim_{m\to\infty} Pr(E_j^m) = 1.$$

$$0 \leq Pr(E_{miss}^m) = Pr(\exists X_j \in X^t; X_j \notin \bigcup_{i=1}^{m} X^{s_i}) = Pr(\bigcup_{j=1}^{K} \neg E_j^m) \leq \sum_{j=1}^{K} Pr(\neg E_j^m).$$

از آن جا که $\lim_{m\to\infty} \sum_{j=1}^{K} Pr(\neg E_j^m) = 0$، بنابراین بنا بر قضیه فشردگی، $\lim_{m\to\infty} Pr(E_{miss}^m) = 0$.

برای قسمت آخر داریم $Pr(\forall j, E_j^m) = Pr(\bigcap_{j=1}^{K} E_j^m) = 1 - Pr(\bigcup_{j=1}^{K} \neg E_j^m)$.

$Pr(\bigcup_{j=1}^{K} \neg E_j^m) \leq \sum_{j=1}^{K} Pr(\neg E_j^m) = \sum_{j=1}^{K}(1-p_j)^m$ و این بخش پایانی قضیه را ثابت می‌کند. □

قضیه۷.۲ می گوید که هم‌چنان که m، تعداد وظایف مبدأ نمونه، به بی‌نهایت میل می‌کند، احتمال این که تمام متغیرهای یک مجموعه مینیمال خاص از متغیرهای سازگارِ آن وظیفه هدف در اجتماع مجموعه‌های مینیمال از متغیرهای سازگار وظایف نمونه که با الگوریتم والش بدست آمده‌اند موجود باشند، به یک میل می‌کند. به‌وضوح ممکن است مجموعه‌های مینیمال دیگری از متغیرهای سازگار برای آن وظیفه هدف موجود باشند و حتی در حالتی که مجموعه سازگارِ خاصِ مورد نظر از وظیفه هدف در این اجتماع موجود نباشد، آن مجموعه‌های دیگر ممکن است بوسیله مجموعه اجتماع در بر گرفته شوند. در هر یک از این حالت‌ها، تجرید مبتنی بر نتایج وظایف نمونه، بهینگی فرایند مارکوف هدف زمینه را حفظ خواهد کرد. بنابراین، احتمال واقعی بدست نیاوردن جواب بهینه درست برای فرایند مارکوف هدف، حتی از E_{miss}^m هم کمتر است.

۳ نتیجه گیری

در این مقاله به فرایندهای مارکوف فاکتور شده پرداخته شده است و متغیرهایی از فضای حالت که در وظایف قبلی حوزه تأثیری در مقادیر تابع ارزش نداشته‌اند استخراج و شناسایی شده‌اند و با صرف نظر کردن از آن‌ها در وظایف بعدی تجرید و کوچک کردن سایز فضا بدون هزینه برخط صورت گرفته است. ثابت شده است که با حذف متغیرهایی که در تمام وظایف قبلی نسبت به تابع ارزش کاملا نامرتبط بوده‌اند، یادگیری وظیفه جدید با درجه صحت بالا صورت می‌گیرد.

مراجع

[۱] T. Dean and R. Givan، Model Minimization in Markov Decision Processes، in Proceedings of the Fourteenth National Conference on Artificial Intelligence(1997)، 106 – 111.

[۲] L. Li، T. J. Walsh، M. L. Littman، Towards a Unified Theory of State Abstraction for MDPs، in Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics(2006)، 531 – 539.

[۳] B. Ravindran، An Algebraic Approach to Abstraction in Reinforcement Learning، PHD thesis، University of Massachusetts Amherst، 2004.

[۴] T.J. Walsh، L. Li، M.L. Littman، Transferring State Abstractions Between MDPs، in Proceedings of the International Conference on Machine Learning–Workshop on Structural Knowledge Transfer for Machine Learning، 2006.