

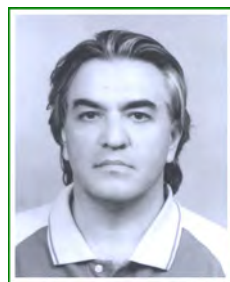
نمونه‌هایی از تعامل ریاضیات و زیست‌شناسی

عبدالحسین عباسیان* و روزبه فرهودی*

نمونه اول



روزبه فرهودی



عبدالحسین عباسیان

شرایطی روی دنباله‌ها، مسئله در رده ساده‌تر الگوریتم‌ها قرار می‌گیرد. این مسئله اهمیت بسیاری در پروژه ژنوم انسان داشت اما به علت ماهیت ساده زیستی آن در مسائل دیگری نیز ظاهر می‌شود. از این رو تحلیل دقیق آن هنوز مسئله مهمی است.

نمونه دوم

گاهی اوقات مسائلی از زیست‌شناسی به پیدایش قضایای جالبی در ریاضیات کمک می‌کند [2]. یکی از مسائل قدیمی زیست‌شناسی، تقسیم‌بندی موجودات زنده است. تلاش بسیاری شده است تا این کار را از روی دی‌ان‌ای انجام دهند. برای این کار نیاز به متریکی روی دنباله‌ها داریم تا محکی برای دوری و نزدیکی دو دنباله تعریف کنیم. به خاطر اشتراک‌های زیاد دی‌ان‌ای موجودات مختلف و انواع جهش‌های ژنتیکی، تعریف متریکی که از نظر زیستی معنادار باشد کمی سخت است. با این حال، متریک‌های زیادی وجود دارد و یکی از معروف‌ترین آنها متریک جوکس-کانتور (Jukes-Cantor distance) است. با داشتن متریک d و n دنباله، به ماتریس $D = (d_{ij})_{i,j=1}^n$ که درایه (i, j) آن فاصله دنباله i ام تا دنباله j ام است می‌رسیم. اما چگونه می‌توانیم از این ماتریس بزرگ به درخت تکاملی این موجودات پی ببریم؟

تمام اطلاعات زیستی هر موجود زنده در دی‌ان‌ای (DNA) آن نوشته شده است. شناسایی این دنباله با کمک آنزیم‌هایی که توانایی تقسیم آن را به قطعات کوچکتر دارند، صورت می‌گیرد. اگر ترتیب کنار هم قرار گرفتن این قطعات مشخص باشد، بازسازی دنباله اولیه کار ساده‌ای است. اما بدون اطلاع از این ترتیب، حالات بسیاری برای بازسازی دنباله نهایی وجود دارد و بدون اطلاعات دیگری نمی‌توانیم دنباله نهایی را بازسازی کنیم. حال تصور کنید این دنباله با دو آنزیم متفاوت تقسیم شده باشد. مثلاً در مجموعه اول تکه‌های A_1, A_2, \dots, A_r و در مجموعه دوم تکه‌های B_1, B_2, \dots, B_s قرار داشته باشد. از جایگاه A_i ها و B_j ها در دنباله اصلی اطلاعی نداریم اما می‌توانیم مشخص کنیم کدام A_i با کدام B_j اشتراک دارد. به بیان دیگر، با بررسی اجزای دو قطعه می‌توان با تقریب خوبی مشخص کرد کدام یک از $A_i \cap B_j$ ها ناتهی است. آیا با کنار هم گذاشتن این اطلاعات می‌توانیم دنباله اصلی را به دست آوریم؟ این سؤال به double digest problem معروف است و جواب آن گاهی مثبت و گاهی منفی است [5]. در بعضی از داده‌های آزمایشگاهی تنها فراوانی طول قطعاتی که آنزیم‌ها جدا کرده‌اند مشخص است که در این حالت مسئله سخت‌تر است. اگر جواب مسئله مثبت باشد با الهام از تکامل، الگوریتم‌های ژنتیکی زیادی برای بازسازی دنباله اولیه پیشنهاد شده است. در این الگوریتم‌ها در ابتدا تعداد زیادی نمونه تصادفی قرار می‌دهیم و در هر مرحله اجازه می‌دهیم تعدادی از آنها با قواعدی با هم ترکیب شوند و نمونه‌های جدید به وجود آورند. سپس آنها را که از محکی برای بقا پیروز نشوند خارج می‌کنیم. این الگوریتم‌ها بسیار کارآمدند، حالات زیادی را بررسی و بهترین نمونه را پیدا می‌کنند. اما گاهی تعداد حالات به قدری زیاد است که مسئله در رده NP خواهد بود. تحقیقات بسیاری در تحلیل این الگوریتم‌ها انجام شده تا مشخص شود تحت چه

* پژوهشگاه

برای این کار با چند تعریف ریاضی شروع می‌کنیم.

منظور از جدول، ماتریسی متقارن $n \times n$ مانند $D = (d_{ij})_{i,j=1}^n$ است که درایه‌های روی قطرش صفرند و در نامساوی مثلثی $(d_{jk} \leq d_{ij} + d_{ik})$ صدق می‌کند. مجموعه تمام جدول‌ها، زیرمجموعه‌ای $(\binom{n}{2})$ بعدی از $\mathbb{R}^{n \times n}$ معروف به کنج متریک است. به هر جدول D می‌توانیم چندوجهی زیر را نسبت دهیم:

$$P_D = \{x \in \mathbb{R}^n \mid \forall i, j : d_{ij} \leq x_i + x_j\}$$

و به راحتی می‌توان نشان داد که اگر D_1, \dots, D_k جدول‌های مختلفی باشند، $D_1 + \dots + D_k$ نیز یک جدول است و

$$P_{D_1 + \dots + D_k} \supseteq P_{D_1} + \dots + P_{D_k}.$$

اگر زیرمجموعه بودن به تساوی تبدیل شود، می‌گوییم جمع D_1 تا D_n همساز است. جدول را جدایی‌پذیر می‌گوییم اگر برای آن افزایش (α, β) از مجموعه $\{1, 2, \dots, n\}$ یافت شود که درایه (i, j) ماتریس متناظرش هرگاه $\alpha \leq i, j \leq \beta$ یا $\{i, j\} \subseteq \beta$ ، برابر صفر و در غیر این صورت یک باشد. فهم جدول‌های جدایی‌پذیر ساده است و علاقه‌مندیم هر جدول را به صورت ترکیبی خطی از آنها بنویسیم. منتها این کار همواره قابل انجام نیست و به همین جهت جدول‌های اول تعریف می‌شوند. یک جدول، اول است اگر نتوان آن را به صورت جمع همساز ضریب مثبتی از یک جدول جدایی‌پذیر و یک جدول دلخواه نوشت. کوچکترین مثال جدول اول اندازه پنج دارد و ماتریس آن برابر فاصله رأس‌ها در گراف کامل دو بخشی $K_{2,3}$ است. قضیه زیر در راستای این مسئله بیان شده است [3]:

قضیه. (تجزیه به جدول‌های جدایی‌پذیر) هر جدول D تجزیه یکتایی به شکل مجموع همساز $D = D_1 + \dots + D_k + D'$ دارد که D_i ‌ها جدول‌هایی جدایی‌پذیر و مستقل خطی هستند و D' جدولی اول است.

نکته قضیه بالا این است که با صرف نظر کردن از D' در تجزیه یک جدول می‌توان متریک اولیه را روی یک درخت نشان داد و به این شکل راهی طبیعی برای نمایش درخت تکاملی پیدا کرد. الگوریتم‌های سریعی برای این تجزیه وجود دارند و در [4] می‌توانید با نرم‌افزارهایی که این محاسبات را انجام می‌دهند آشنا شوید.

نمونه سوم

یکی از مسائل هیجان‌انگیز در ریاضیات زیستی، مسئله حافظه و یادگیری در مغز انسان و سایر موجوداتی است که سیستم عصبی دارند. برای سادگی، فعالیت یک سلول عصبی را می‌توان به صورت صفر و یک یا خاموش و روشن نمایش داد. سلول‌های عصبی معمولاً در شبکه‌ای از نورون‌ها مورد مطالعه قرار می‌گیرند به طوری که هر سلول عصبی با تعداد دیگری از اعضای شبکه ارتباط دارد. این ارتباطات را با ماتریس W یا اصطلاحاً ماتریس وزن‌ها

(weight matrix) نشان می‌دهند. در این ماتریس، درایه w_{ij} یک عدد حقیقی منفی یا مثبت است که میزان اتصال یا وزن معمول i ام و j ام را تعیین می‌کند. معمولاً برای همه سلول‌های عصبی یا نورون‌ها یک دینامیک ساده ولی غیرخطی در نظر گرفته می‌شود اما قبل از اینکه وارد جزئیات مسئله شویم می‌توانیم به دو سؤال که بلافاصله مطرح می‌شود و توجه ریاضیدانان و فیزیکدانان را در چند دهه اخیر به خود جلب کرده است اشاره کنیم. اول اینکه با توجه به دینامیک داده شده تعداد نقاط جاذب یا اصطلاحاً رابینده‌های (attractors) سیستم چه اندازه است و این رابیندها چگونه دسته‌بندی می‌شوند. دوم اینکه چگونه می‌توان ماتریس W را تغییر داد به طوری که حاصل دینامیک داده شده دسته معینی از نقاط دلخواه باشد. همان گونه که می‌توان حدس زد سؤال اول به مسئله حافظه و سؤال دوم به مسئله یادگیری مربوط می‌شود. به عبارت دیگر، نقاط جاذب سیستم دینامیکی را حافظه سیستم می‌نامیم و تغییرات W و یا، بهتر بگوییم، قواعدی را که باعث تغییر W می‌شود قواعد یادگیری می‌نامیم. تعداد سلول‌های عصبی مغز تقریباً از مرتبه 10^{15} و تعداد اتصالات یا اصطلاحاً سیناپس‌ها (synapses) از مرتبه 10^{19} است و مسئله تغییرپذیری یا اصطلاحاً پلاستیسیته (plasticity) مغز از یافته‌های مهم علوم عصب‌شناسی در قرن اخیر است. واضح است که توصیف ناحیه کوچکی از مغز که از هزارها و بلکه میلیون‌ها سلول عصبی تشکیل شده با توجه به رفتارهای غیرخطی هر یک از نورون‌ها کار دشواری است. این را هم اضافه کنیم که ما هنوز تعریف دقیق ریاضی از مفهوم یادگیری و یا رابیندهای سیستم‌های دینامیکی پیچیده نداریم [6,7,8] ولی با وجود این‌ها نتایج اولیه به دست آمده نشان می‌دهد که چالش‌های بزرگی در راه است. در ادامه این مطلب به توصیف نمونه ساده‌ای از مدل‌سازی ریاضی شبکه‌های نورونی می‌پردازیم که سیستم‌های دینامیکی شبکه را به طرز جالبی با خواص گراف‌هایی متناظر با الگوهای صفر و یک حافظه و حوزه‌ای از ریاضیات تحت عنوان «هندسه فاصله» (distance geometry) پیوند می‌دهد.

دینامیک یک شبکه n تایی از نورون‌ها به صورت زیر قابل توصیف است

$$\dot{\mathbf{x}} = -D\mathbf{x} + [W\mathbf{x} + \mathbf{b}]_+, \quad (1)$$

که در آن $D = \text{diag}(1/\tau_1, \dots, 1/\tau_n)$ (τ_i ها ثابت‌های زمانی اند)، W ماتریس وزن‌هاست، $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}^n$ و $[Z]_+ = \max(z, 0)$. البته نقطه ثابت یا پایدار این سیستم لزوماً آرایه‌ای از صفر و یک‌ها نیست ولی به راحتی می‌توان نقطه پایدار $\mathbf{x}^* \in \mathbb{R}_{\geq 0}^n$ را به چنین آرایه‌ای تبدیل کرد:

$$\sigma = \text{supp}(\mathbf{x}^*) = \{i \in [n] \mid x_i^* > 0\}.$$

به عبارت دیگر به ازای هر i چنانچه x_i^* بزرگتر از صفر باشد مقدار σ_i را یک و در غیر این صورت صفر در نظر می‌گیریم. به عنوان مثال، $\sigma = \{1, 7, 11\}$ به این معنی است که در آرایه یا بردار n تایی σ مقادیر $\sigma_1, \sigma_7, \sigma_{11}$ با هم شلیک می‌کنند یا روشن هستند و بقیه خاموش‌اند. به این ترتیب برای یک n تایی با ماتریس اتصال W دسته‌ای از الگوهای

منظور از $X(G)$ مجموعه خوشه (clique) های گراف G است که قبلاً اشاره شد. آنچه اهمیت دارد رابطه $\text{geom}(\mathcal{S})$ است که ابزارهای ریاضی جالبی را مطرح می‌کند:

$\text{geom}(\mathcal{S}) \stackrel{\text{تعریف}}{=} \{\sigma \subset \{1, \dots, n\} \mid \sigma \text{ یک ماتریس فاصله، مربعی و وارون پذیر باشد}\}$

نکته ۱. ماتریس فاصله مربعی (square distance matrix)

ماتریسی $n \times n$ است با این خصوصیت که نقاط $P_1, \dots, P_n \in \mathbb{R}^{n-1}$ وجود داشته باشند به طوری که $S_{ij} = \|P_i - P_j\|^2$.

نکته ۲. شرط وارون پذیری ماتریس فاصله وقتی برقرار است که دترمینان

ماتریس زیر که به دترمینان کیلی-منگر (Cayley-Menger) معروف است غیرصفر باشد:

$$\text{cm}(\mathcal{S}) = \det \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathcal{S} \end{bmatrix} \neq 0.$$

($\mathbf{1} \in \mathbb{R}^n$ برداری است که همه مؤلفه‌های آن یک است).

رابطه (۲) از این جهت حائز اهمیت است که همه حافظه‌های مجاز شبکه عصبی و سیستم دینامیکی داده شده را برحسب خوشه‌های گراف G ، $X(G)$ ، و مفهوم ماتریس فاصله دسته‌بندی می‌کند.

نتیجه

در اینجا سه نمونه از مسائل ریاضیات زیستی را بیان کردیم. تعداد مسائل مشابه بسیار زیاد است. در دادوستد بین ریاضیات و زیست‌شناسی، به خصوص نقش اولی در دومی در چند دهه گذشته بسیار پررنگ شده است. گاهی وضعیت فعلی تعامل ریاضیات و زیست‌شناسی با مشابه تاریخی آن، ریاضیات و فیزیک، مقایسه می‌شود. این مقایسه از بسیاری جهات قابل تأمل است. باید به یاد داشته باشیم که از نظر یک ریاضیدان، راه و روش فیزیک بسیار متفاوت با زیست‌شناسی است [1, 4]. احتمالاً هیچ‌گاه قوانینی مشابه قوانین نیوتن در زیست پیدا نخواهد شد. اگرچه اصول کلی مانند وراثت وجود دارد ولی اکثراً نه با معادلاتی ریاضی قابل بیان‌اند و نه توصیف کاملی از پیچیدگی پدیده‌های زیستی ارائه می‌دهند. علاوه بر این، پراکندگی مسائل در زیست بسیار زیاد است و احتمالاً ریاضیدان در مواجهه با آنها در وهله اول سردرگم می‌شود! وجود یک ماده شیمیایی جدید در محیط چه اثری بر یک گونه جانوری دارد؟ نورون‌های مغز ما چگونه در ارتباط با یکدیگر باعث فکر کردن، احساس کردن و ... می‌شوند؟ چگونه ارگان‌های مختلف بدن انسان در کنار یکدیگر کار می‌کنند؟ سرطان و بیماری‌های مشابه چگونه رشد می‌یابند؟ در کنار این‌ها آزمایش کردن در زیست‌شناسی کار راحتی نیست: اولاً به علت ارتباطات پیچیده سیستم‌های زیستی جدا کردن یک قسمت و بررسی آن اغلب به نتایج اشتباهی می‌انجامد. ثانیاً هیچ دو سلول، گیاهی و یا انسانی

صفر و یک متناظر با نقاط پایدار سیستم دینامیکی به دست می‌آید. کار جالب دیگری که می‌شود انجام داد این است که به این شبکه یک گراف G با n رأس نسبت می‌دهیم به طوری که به ازای الگوهای داده شده، σ ، رئوسی که با هم شلیک می‌کنند یال‌های مشترک دارند. با این مقدمات می‌توان به دو قضیه نسبتاً جالب اشاره کرد (برای دنبال کردن جزئیات بیشتر مسئله می‌توان به اصل منبع یادشده مراجعه کرد).

قضیه با توجه به رابطه (۱) و ماتریس‌های داده شده D و W مجموعه الگوها یا حافظه مجاز سیستم، $\mathcal{P}(W)$ ، از رابطه زیر به دست می‌آید

$$\mathcal{P}(W) = \text{stab}(-D + W)$$

$$\text{stab}(A) \stackrel{\text{def}}{=} \{\sigma \subset \{1, \dots, n\} \mid A_\sigma \text{ یک ماتریس پایدار باشد}\}$$

نکته به ماتریسی پایدار می‌گویند که جزء حقیقی مقادیر ویژه آن منفی باشد و منظور از A_σ زیرماتریسی از A است که متناظر با اندیس‌های غیرصفر σ است. مثلاً در حالت قبلی $\sigma = \{1, 2, 11\}$ ، ماتریس A_σ شامل زیرماتریسی با سطر و ستون ۱، ۲ و ۱۱ است.

همان طور که گفته شد، کوشش‌های فراوانی انجام گرفته تا الگوهای حافظه سیستم به روش‌های دقیق دسته‌بندی شود. مثلاً در فیزیک افرادی نظیر هاپفیلد [9]، امیت و سومپولینسکی [11] نتایج جالبی از گذر فاز چنین سیستم‌هایی ارائه کرده‌اند که نحوه پدیداری و از بین رفتن این الگوها را بررسی می‌کند. در این نوشته قصد نداریم به تعبیر جالب فیزیکی مسئله اشاره کنیم و چنانکه خواهیم دید در قضیه بعدی شکل تعمیم یافته قضیه اول که از لحاظ ریاضی و ابزارهای به کار رفته جالب است مطرح شده است و مسئله مجاز بودن الگوهای حافظه را از زاویه دیگری مطرح می‌کند.

نخست به این نکته مهم باید اشاره کنیم که مغز انسان یا موجودات دیگر با هر اتصال دلخواهی بین نورون‌ها سازگار نیست و هر تغییری در اتصال‌ها و به زبان دقیق‌تر تغییرات W در چهارچوب قید و شرط‌های زیستی-تکاملی قابل شکل‌گیری است. به عبارت دیگر، همواره اتصال‌های ثابتی که با این ماتریس S نشان داده می‌شوند وجود خواهند داشت که تغییرات W را محدود می‌کنند. با این توصیف، قضیه بعدی دسته‌بندی جامع‌تری از حافظه‌های مجاز سیستم ارائه می‌دهد:

قضیه اگر S یک ماتریس $n \times n$ متقارن و C مجموعه‌ای از الگوهای داده شده σ باشد، آنگاه قاعده یادگیری خاصی برای W براساس S وجود دارد به طوری که الگوهای حافظه سیستم $\mathcal{P}(W)$ از رابطه زیر به دست می‌آید:

$$\mathcal{P}(W) = \text{geom}(\mathcal{S}) \cap X(G). \quad (2)$$

- 4 **A. Friedman**, *What is mathematical biology and how useful is it?*, Notices Amer. Math. Soc. **57**(7) (2010), 851-857.
- 5 **T. Gowers** (Ed) and et al, *The Princeton Companion to Mathematics*, Princeton University Press, New York, 2009.
- 6 **M. Gromov**, *Ergostructures, ergologic and the universal learning problem*,
[http://www.ihes.fr/gromov/PDF/ergologic3'\(1\).pdf](http://www.ihes.fr/gromov/PDF/ergologic3'(1).pdf), 2013.
- 7 **J. W. Milnor**, *Attractor*,
<http://www.scholarpedia.org/w/index.php>.
- 8 **D. Amit**, *Modeling brain function : The world of attractor neural networks*, Cambridge University press: Cambridge, 1989.
- 9 **J. J. Hopfield**, *Neurons networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci, **79**(8) (1982), 2554-2558.
- 10 **D. Amit, H. Gutfreund, and H. Sompolinsky**, *Statistical mechanics of neural networks near saturation*, Ann. Physics **173** (1987), 30-67.
- 11 **C. Curto, A. Degeratu, and V. Itskov**, *Encoding binary neural codes in networks of threshold-linear neurons*, Neural Comput. **25** (2013), 2858-2903.

دقیقاً مشابه یکدیگر نیستند و تنوع یکی از پایه‌های زیست است. ثالثاً در بسیاری از آزمایش‌ها محدودیت‌های ذاتی و ابزاری وجود دارد. مثلاً نمی‌توان رفتار تمام سلول‌های مغزی را در یک آزمایش مشاهده کرد. با وجود این زمینه‌های کاربردی‌تر ریاضیات مانند معادلات دیفرانسیل و آمار و احتمالات برای فهم بسیاری از مدل‌های زیستی کفایت می‌کنند. تجربه کاربردی ریاضی در علوم دیگر مانند فیزیک نشان داده است که با گذشت زمان رشته‌های متنوع‌تری از ریاضیات درگیر این مباحث می‌شوند و با مجرد شدن بعضی روش‌ها، مفاهیم ریاضی جدیدی ابداع می‌شود. درس تاریخی این است که ریاضیدانان جوان نباید از ورود به موضوعات جدید بهراسند و باید خود را آماده ابداع روش‌های نوین در بررسی پدیده‌های زیستی کنند.

مراجع

- 1 **M. C. Reed**, *Why is mathematical biology so hard?*, Notices Amer. Math. Soc. **51**(3) (2004), 338-342.
- 2 **B. Sturmfels**, *Can biology lead to new theorems?*, Annual Report of the Clay Mathematics Institute, Boston, 2005.
- 3 **H. J. Bandelt and A. Dress**, *A canonical decomposition theory for metrics on a finite set*, Adv. in Math. **92** (1992), 47-105.

شوخی‌های ژنتیکی

