

Toán học, khoa học tính toán và khoa học về sự sống

Một cách tương đối có thể xem có ba loại toán có đặc điểm và động lực phát triển khác nhau. Một là toán học thuần túy được nghiên cứu với động lực phát triển nội tại của toán học. Toán học thuần túy phát triển nhanh và sâu theo các chuyên ngành đến mức nếu như các nhà toán học kiệt xuất David Hilbert và Henri Poincaré ở thế kỷ 19 có thể hiểu sâu được nhiều lĩnh vực toán học thời đó, thì ngày nay dường như không ai có thể nhìn được bức tranh tổng thể về toán học sâu sắc như vậy. Hai là toán học đại chúng, gồm những kiến thức toán mọi người cần cho cuộc sống và công việc hàng ngày. Điều đáng chú ý ở đây là giáo dục của ta sẽ đưa thứ toán đại chúng nào và đưa thế nào đến cho mọi người ở thời máy tính và internet này? Ba là toán cho các khoa học khác, có nội dung và động lực phát triển từ các vấn đề cụ thể của các khoa học này. Chính ở đây, toán học thể hiện vai trò cơ bản của mình trong khoa học và ảnh hưởng sâu sắc đến sự phát triển của xã hội.

Toán học, khoa học máy tính và khoa học tính toán

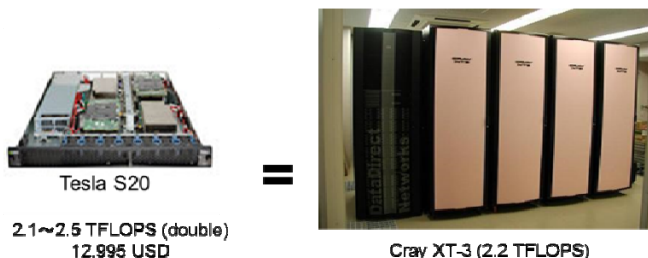
Khoa học máy tính (computer science) hay *tin học* (informatics) gần đây được gọi ở ta với tên *công nghệ thông tin* (information technology) theo nghĩa rộng của từ này. Khoa học máy tính được xem là “khoa học về xử lý thông tin tự động bằng máy tính điện tử”. Về bản chất, *xử lý thông tin là quá trình biến đổi dữ liệu từ dạng này sang dạng khác để thu được thông tin và tri thức mới*.

Cốt lõi của quá trình “biến đổi dữ liệu” chính là các biến đổi toán học, và do vậy cốt lõi của khoa học máy tính cũng chính là toán học. Một người bạn ở Pháp kể khi là sinh viên tin học, anh đã bàng hoàng lúc thầy giáo cho xem bức ảnh một rừng cây do không quân Mỹ chụp ở Trường Sơn trong chiến tranh Việt Nam, và qua các phép biến đổi toán học lớp cây xanh đã bị bóc đi để lộ rõ bên dưới hình các xe tăng của bộ đội miền Bắc. Đây là một thí dụ về biến đổi dữ liệu trong xử lý ảnh.

Trong khoa học máy tính ta luôn gặp các vấn đề toán học, từ bài toán nền tảng $NP \neq P$ (một trong bảy bài toán do Viện Toán học Clay công bố vào năm 2000, thách thức các nhà toán học giải trong thế kỷ 21) đến các bài toán thiết yếu như lập lịch biểu thời gian thực, chống bùng nổ tổ hợp khi tìm kiếm lời giải trong trí tuệ nhân tạo, các hệ hình thức để kiểm tra tính đúng đắn của chương trình trong công nghệ phần mềm, ...

Nếu như khoa học máy tính được hiểu như trên với đối tượng nghiên cứu là máy tính và các phương pháp tính toán trên chúng, thì *khoa học tính toán* (computational science) lại là khoa học về việc dùng *toán học* và *khoa học máy tính* để làm việc trong các ngành khoa học khác. Dưới tên gọi này, có rất nhiều lĩnh vực cụ thể như ngôn ngữ tính toán (computational linguistics), sinh học tính toán (computational biology) hoặc tin sinh học (bioinformatics), tài chính tính toán (computational finance), khoa học vật liệu tính toán (computational materials science), vật lý tính toán, hóa học tính toán, ...

Khoa học tính toán gần đây được nhìn nhận rõ hơn và đang có một sự dịch chuyển trong nhiều ngành khoa học từ thực nghiệm tới kết hợp thực nghiệm với tính toán. Có được điều này do nhiều yếu tố, một trong đó là tiến bộ rất nhanh của khoa học máy tính. Thí dụ như hiện tượng đặc sắc về công nghệ GPGPU (các bộ xử lý đồ họa với mục tiêu tổng quát). Gần đây, các bộ GPGPU của Nvidia Tesla S2070 với giá hơn 10 nghìn đôla, khi lắp vào các máy tính thông thường có thể cho phép ta đạt đến tốc độ hàng nghìn tỷ phép tính giây (Teraflops) của các siêu máy tính, chẳng hạn máy tính Cray XT-3 ra đời năm 2004.



Những thành phần cơ bản của khoa học tính toán gồm *mô hình hóa* (modeling) nhằm tìm các mô hình, tức các mô tả khái quát của một hiện thực khi tách phần không cốt yếu ra khỏi phần cốt yếu, *mô phỏng* (simulation) nhằm tạo ra các vật thể và hiện tượng như thật trên máy tính, và *phân tích dữ liệu* (data analysis) nhằm rút ra những kết luận cần thiết từ các tập dữ liệu.

Toán học với khoa học về sự sống và sinh-y học

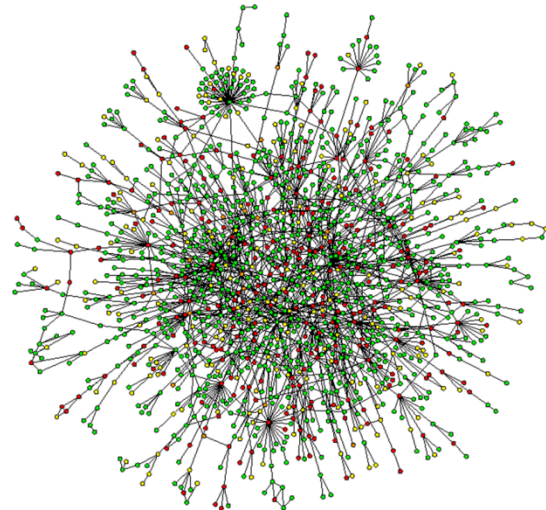
Sinh học phân tử cho ta biết mỗi tế bào trong hàng nghìn tỷ tế bào của mỗi người chứa toàn bộ *hệ gen* người, khoảng 20-25 nghìn *gen*. Các *gen* này – tức các đoạn đặc biệt của các dãy DNA nằm trong các cặp nhiễm sắc thể của tế bào – chứa mật mã di truyền, điều khiển sự phát triển và kế thừa của tế bào. Các *protein* – các dãy amino acid tổng hợp từ DNA và RNA – là các thành phần thiết yếu của các tổ chức sống và hoạt động của chúng. *Sinh học tính toán* (computational biology) hay *tin-sinh học* (bioinformatics) là ngành khoa học liên quan đến việc dùng các phương pháp toán học và khoa học máy tính để nghiên cứu các bài toán của sinh học.

Khoa học về sự sống bao gồm tất cả các lĩnh vực khoa học liên quan đến việc nghiên cứu các sinh vật, trong đó sinh học và y học là lĩnh vực trung tâm. Nhìn tổng thể, khoa học sự sống bao gồm nghiên cứu ở các cấp độ: phân tử, tế bào, sinh vật sống, quần thể, và các hệ sinh thái. Đặc trưng cốt lõi của khoa học về sự sống trong thế kỷ 21 là việc nghiên cứu đang trở nên có tính định lượng và dựa trên dữ liệu. Những điều này có được nhờ vào:

- Sự phát triển và sử dụng rộng rãi các thiết bị tự động đã tạo ra rất nhiều dữ liệu ở tất cả các cấp độ của khoa học về sự sống. Lượng dữ liệu này tăng nhanh theo hàm mũ, gần đây thường được tạo ra bởi các tổ chức và đề án lớn, lưu trữ trong các cơ sở dữ liệu cho phép mọi người cùng sử dụng.
- Thành công của đề án quốc tế về hệ gen người (human genome project) vào năm 2003 đã tạo ra những nguồn dữ liệu trung tâm của sinh học.
- Các hệ máy tính hiệu năng cao, được nối mạng, có trên bàn của các nhà sinh học và toán học.

Ta điem qua ở đây những thách thức chính của toán học trong khoa học về sự sống ở các cấp độ khác nhau kể trên:

(1) *Hiểu về các phân tử*: Ba đại phân tử đóng vai trò quan trọng nhất là DNA, RNA và protein. Mỗi phân tử DNA hay RNA là một dãy của 4 đơn phân tử (ký hiệu bởi các chữ các A, G, C, T hoặc U) và mỗi protein là một dãy của 20 amino acid. Các lớp bài toán chủ yếu cần giải quyết để có thể hiểu các phân tử gồm: (a) phân tích dãy, (b) phân tích cấu trúc phân tử, (c) động học của phân tử, và (d) tương tác phân tử. Một số bài toán ở đây là các thách thức lớn cần sự tham gia của toán học, như việc dự đoán sự xoắn cuộn của protein từ dãy amino acid và thông tin về môi trường (protein folding) được xem là một trong những vấn đề hóc búa nhất của sinh học và vẫn chưa có lời giải, hay sắp thẳng hàng n dãy ($n > 2$) là bài toán tổ hợp tối ưu NP-đầy đủ. Bài toán xác định và xây dựng các mạng tương tác protein-protein (liên quan đến lý thuyết đồ thị, tô pô, và thống kê) là nền tảng để nghiên cứu nhiều cơ chế của bệnh tật và chế thuốc.



Mạng tương tác proteins, Jeong et al., Nature 441 (2001)

(2) *Hiểu về tế bào*: Việc này đòi hỏi ta phải hiểu cấu trúc các đại phân tử trong tế bào, các cơ chế và mẫu dạng không gian, thời gian của động học tế bào, liên quan giữa động học và chức năng của tế bào, liên quan giữa tế bào và các tổ chức sống khác ở mức cao hơn, như mô và các bộ phận khác. Hiểu về tế bào thậm chí còn khó hơn việc hiểu các phân tử do ta không có các nguồn thông tin như các dãy đơn phân tử hay amino acid của DNA, RNA hay protein. Với sự tiến bộ rất nhanh của công suất máy tính và các thiết bị đo đạc y sinh, ta có thể xây dựng được hợp lý và chính xác các mô hình về động học của DNA và protein, trong khi đó các mô tả định lượng của tế bào mới chỉ thực hiện được một cách xấp xỉ. Thách thức chủ yếu về phân tích toán học của tế bào sẽ không nằm ở việc tính toán mà ở việc mô hình được những thứ con người quan tâm. Những thách thức này ở một vài thập kỷ tới chính là việc thiết lập một cách hệ thống các mô hình của cấu trúc và động học tế bào, xuất phát từ các dữ liệu phức tạp thu được và các thực nghiệm về các mô hình này. Đã có một lịch sử về nghiên cứu toán học cho các mô hình tế bào. Những bài toán chủ yếu được theo đuổi gồm: (a) nghiên cứu cấu trúc tế bào, (b) phát hiện mạng tế bào và chức năng của chúng, (c) từ mạng tế bào đến chức năng tế bào, và (d) từ tế bào đến mô. Trọng tâm của sinh học tế bào tới đây sẽ là sự chuyển dịch từ các mô tả về hiện tượng sang các mô hình dự đoán tương thích với những lượng dữ liệu rất lớn. Các nghiên cứu tới đây sẽ nhấn mạnh sự kết hợp chặt hơn giữa thực nghiệm, xây dựng và đánh giá mô hình cũng như tích hợp dữ liệu.

(3) *Hiểu về các vật sống*: Phân tích toán học ở đây liên quan hai vấn đề chính, một là để hiểu các bộ phận của một sinh vật tham gia vào hệ thống phức tạp của sinh vật này ra sao, và hai là liên hệ sinh học giữa tế bào và các bộ phận của sinh vật. Tầm quan trọng của các mô hình toán được thừa nhận rộng rãi trong nghiên cứu các tổ chức rất phức tạp của sinh vật. Nhiều thí dụ cho thấy khi người làm thực nghiệm và lý thuyết ngồi cùng nhau, họ đã có những phát minh quan trọng vốn không thể làm một mình. Chẳng hạn như sự dao động trong chu trình tế bào dẫn đến việc phân chia tế bào ra sao hay virus HIV được sinh ra và mất đi trong tế bào như thế nào? Các mô hình toán học có trong tất cả các hướng nghiên cứu để hiểu về sự sống: (a) sinh lý học tim mạch, (b) sinh lý học tuần hoàn, (c) sinh lý học hô hấp, (d) sinh lý học nội tiết, (e) hình thái học và sự hình thành các mẫu dạng di truyền, (f) sự vận động, (g) quá trình nhiễm HIV, và (h) ung thư (được tiến hành ở gen, tế bào và u). Ở cấp độ này, thí dụ về những vấn đề cần sự đóng góp của khoa học tính toán là xác định con đường thuốc đi tới tế bào đích, cơ chế hoạt động của thuốc, sự phát triển và phân chia của các quần thể tế bào, tiến triển của sự kháng thuốc. Việc dùng các mô hình toán học để mô tả quá trình hoạt động của các hệ thống chức năng sinh lý sẽ cải thiện hiểu biết của ta về tương tác động giữa các quá trình này và giúp cho sự chuyển dịch từ khoa học cơ bản đến các ứng dụng chữa bệnh, cũng như khả năng xuất hiện các nội dung mới của toán học.

(4) *Hiểu về các quần thể*: Các nội dung ở cấp độ này đều rất cần đến toán. (a) Về di truyền học quần thể: Cho đến cuối những năm 1960 các nghiên cứu toán chủ yếu là phân tích suy diễn với các mô hình di truyền và chọn lọc tự nhiên dựa trên giả thuyết các quần thể sinh học đã tiến hóa từ một hai nơi trên trái đất. Gần đây đã có sự dịch chuyển qua các mô hình quy nạp nhằm xây dựng lịch sử tiến hóa và bản chất của quá trình tiến hóa, với mục tiêu hiểu được các dạng của biến đổi di truyền và các dạng của khác biệt di truyền giữa các loài. (b) Về các khía cạnh sinh thái của các quần thể: Ngay từ 1838, Verhulst đã phát triển một số mô hình về tăng dân số với các mô hình toán học về mật độ, trong đó phương trình logistic vẫn là mô hình chuẩn dùng ngày nay. Rất nhiều khía cạnh khác liên quan đến các vấn đề sinh thái của quần thể đang thách thức và chờ đón đóng góp của toán học, như các yếu tố không gian, thời gian, địa lý, khí hậu, sông ngòi rừng biển. (c) Về tổng hợp sinh thái và tiến hóa: Hiện nay nghiên cứu được tập trung nhiều vào các nguyên nhân liên quan đến sự sống và không có sự sống trong đa dạng sinh học. Rất nhiều thách thức toán học khi các quá trình sinh thái và tiến hóa được xem xét đồng thời.

(5) *Hiểu về các cộng đồng và các hệ sinh thái*: Một cộng đồng sinh thái là một tụ tập các quần thể của nhiều loài (cây cối, động vật, vi khuẩn, ...) tại một nơi và cùng một thời gian. Sự kết hợp của một cộng đồng và một môi trường thường được xem là một hệ sinh thái. Ta vẫn thường nói về cân bằng sinh thái hoặc mất cân bằng sinh thái. Toán học đóng một vai trò thiết yếu trong các khái niệm của sinh thái học cộng đồng. Từ quãng 50 năm trở lại đây, có nhiều mô hình toán học và kết quả lý thú về liên quan giữa tính đa dạng và tính ổn định của một cộng đồng sinh thái, hoặc rộng hơn về mối quan hệ giữa hiệu quả sản xuất, tính ổn định

và tính đa dạng. Thách thức toán học tiêu biểu ở đây là các mô hình cho phép mô tả với nhiều yếu tố như tính phi tuyến, hành vi không cân bằng, kết cấu di truyền, không gian, nhân khẩu học, và tính ngẫu nhiên của môi trường.



Nhiều người cho rằng toán học sẽ ảnh hưởng đến sự phát triển của sinh học và y học trong thế kỷ này nhiều như sự ảnh hưởng của toán đến vật lý trong thế kỷ vừa qua. Sinh học hiện đang ở bước chuyển pha từ cách nghiên cứu thực nghiệm tới việc kết hợp cách tiếp thực nghiệm và tính toán để tìm hiểu các hiện tượng phức tạp của sự sống. Y-sinh học, tức việc nghiên cứu y học với tri thức và tiến bộ của sinh học phân tử, đang dần trở thành một hướng đi chủ đạo trong nghiên cứu y học.

Dưới đây là thí dụ về vài bài toán y sinh chúng tôi tham gia giải quyết trong 10 năm qua. Từ một cơ sở dữ liệu lâm sàng rất lớn thu thập trong 20 năm (1980-2000) về bệnh gan ở bệnh viện đại học Chiba (Nhật Bản) và nhiều cơ sở dữ liệu về thông tin di truyền, một số bài toán nghiên cứu về viêm gan được đặt ra. Một bài toán là tìm phương pháp tính toán để đoán nhận cấp độ bệnh viêm gan (có 5 cấp độ từ F0 đến F4) khi chỉ dựa vào dữ liệu xét nghiệm máu nhưng không làm sinh thiết như cách hiện nay. Bài toán khác về việc tìm nguyên nhân ở mức phân tử tại sao những người bệnh viêm gan C lại kháng thuốc hoặc không kháng thuốc (hiện cách điều trị viêm gan C là phối hợp hai loại thuốc interferon và ribavirin, tuy nhiên số người kháng thuốc vẫn trên 50%). Đây là bài toán thời sự khi gần đây giới nghiên cứu cho rằng protein NS5A, một phần của virus viêm gan C, chính là nơi liên quan nhiều nhất đến việc nhận hay kháng các loại thuốc trên. Bài toán thứ ba là việc tìm ra các gen gây bệnh. Hầu hết các bệnh của con người do rối loạn gen, và một số trong 20-25 nghìn gen của người đã được xác định là nguyên nhân gây ra một số bệnh. Bài toán đặt ra cho tính toán là những gen nào trong số còn lại cũng gây ra bệnh?

Trong hợp tác quốc tế về nghiên cứu y học, nhiều khi chúng ta chỉ đóng vai trò cung cấp dữ liệu vì không làm chủ được các phương pháp phân tích toán học. Tuy nhiên, có những tình huống buộc chúng ta phải tự làm, như khi một bệnh nhiệt đới bùng phát hay khi dịch lợn tai xanh xuất hiện khắp nơi,... Trong một chuyến thăm gần đây tới Bệnh Viện Nhi Trung Ương, chúng tôi vô cùng khâm phục khi bệnh viện đã xây dựng các phòng thí nghiệm về sinh học phân tử, về di truyền, đã có thể đo đạc xác định được dữ liệu từ các mẫu của người bệnh, dù công việc khám chữa bệnh ở đây bận rộn vô cùng. Nếu các nhà toán học, tin học phối hợp được với các bác sĩ nghiên cứu của bệnh viện, chúng ta có thể làm được nhiều nghiên cứu thiết thực và giá trị cao như ở bất kỳ nơi nào.

Nước Pháp có nhiều nhà toán học xuất sắc và rất nhiều người trong họ làm nghiên cứu toán cho những khoa học khác. Chính Cedric Villani, nhà toán học Pháp vừa nhận giải Fields, trong một bài phỏng vấn vào tháng 5.2010 cho rằng, “những điều bất ngờ nhất sẽ nằm trong những lĩnh vực có sự tương tác giữa toán học và các ngành khác”.

Hy vọng chúng ta sẽ có nhiều người làm toán dành công sức và đam mê trong những lĩnh vực khoa học đang cần nhiều toán học.

Hồ Tú Bảo

Tài liệu tham khảo chính

1. Bascombe, J., Biology and Mathematics, Arbor, CLXXXIII, 347-361, 2007.
<http://arbor.revistas.csic.es/index.php/arbor/article/view/107/108>
2. National Academy of Sciences, “Mathematics and 21st Century Biology”, The National Academies Press, 2005, http://www.nap.edu/catalog.php?record_id=11315
3. Simon A. L., “Mathematics and biology”, <http://www.bio.vu.nl/nvtb/Contents.html>