

Một số vấn đề về đồ họa máy tính

Dai hoc Khoa học Tự nhiên

Khoa Toán - Cơ - Tin học

Khoa học dữ liệu K4

Tháng 12 năm 2022

Chương 7 - Kỹ thuật trực quan hóa cho dữ liệu hướng thời gian

Nguyễn Mạnh Linh, Nguyễn Đức Thịnh

Tài liệu này được dịch từ cuốn "Ward, M.O. and Grinstein, G. and Keim, D., *Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition, CRC Press, 2015*", chương số 7 "Visualization Techniques for Time-Oriented Data".

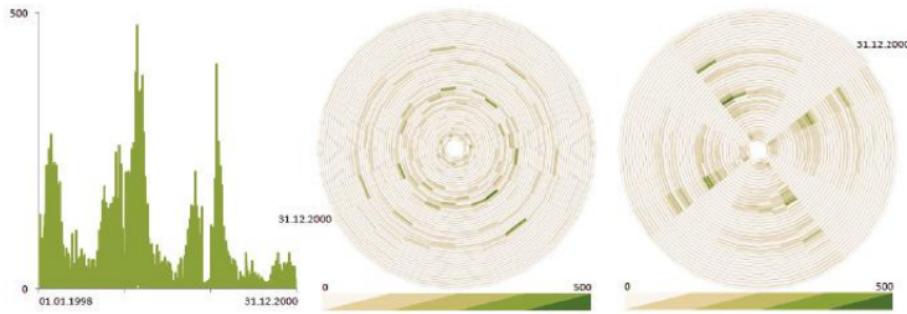
Trong chương này chúng tôi giải thích chi tiết các kỹ thuật trực quan hóa cho dữ liệu hướng thời gian. Chúng tôi tổ chức và cấu trúc chương theo cuốn sách này [5]. Do đó, trước tiên chúng tôi nhấn mạnh tầm quan trọng của việc xử lý các yếu tố thời gian qua các ví dụ. Thứ hai, chúng tôi đưa ra các khái niệm cần thiết và các khía cạnh của thời gian và dữ liệu hướng thời gian. Các tập dữ liệu cơ bản đã có trong chương 2. Ở chương này, chúng tôi tập trung vào các đặc điểm của thời gian. Thứ ba, chúng tôi cung cấp một cái nhìn tổng quan về các kỹ thuật trực quan hóa khác nhau. Thứ tư, chúng tôi giải thích ngắn gọn về *TimeBench*, là mô hình và một thư viện phần mềm phân tích dữ liệu hướng thời gian. Chúng tôi kết luận với đánh giá phân loại theo các kỹ thuật trực quan hóa được trình bày và giới thiệu *TimeViz Browser*, là một kho lưu trữ để hỗ trợ các nhà nghiên cứu và học viên trong việc tìm kiếm các kỹ thuật trực quan hóa tích hợp, tiếp theo là các bài đọc, bài tập và đồ án thực hành.

1 Giới thiệu

Bản thân thời gian vốn là dữ liệu một chiều, là trung tâm của việc phát hiện xu hướng, xác định các mẫu lặp lại và mối các mối quan hệ trong dữ liệu. Thời gian và dữ liệu hướng thời gian có những đặc điểm riêng biệt đáng để chúng ta nghiên cứu và xử lý như một kiểu dữ liệu riêng. Do tầm quan trọng của dữ liệu hướng thời gian, cấu trúc của nó đã được nghiên cứu trong nhiều ấn phẩm khoa học.

Các ví dụ sau minh họa tầm quan trọng của thời gian. Hình (1) thể hiện ba biểu diễn trực quan khác nhau của cùng một tập dữ liệu hướng thời gian, chứa số ca mắc cúm hàng ngày xảy ra ở miền bắc nước Đức trong khoảng thời gian ba năm. Dữ liệu thể hiện một mô hình tuần hoàn mạnh. Hình ngoài cùng bên trái sử dụng một biểu đồ đường đơn giản để trực quan hóa dữ liệu. Mặc dù thời gian cao điểm có thể được nhận ra dễ dàng, tuy nhiên khi kiểm tra biểu diễn này, hành vi theo chu kỳ của dữ liệu chỉ có thể được đoán và thật khó để phân biệt các chu kỳ mà trên thực tế có tồn tại. Ngược lại, hình ở giữa và hình bên phải biểu diễn một vòng tròn nhấn mạnh các đặc tính tuần hoàn của dữ liệu bằng cách sử dụng trục thời gian hình xoắn ốc. Với hình xoắn ốc bên trái, mô hình tuần

hoàn không được phát hiện ra. Điều này là do độ dài chu kì được đặt là 24 ngày, không khớp với sự tuần hoàn của dữ liệu. Biểu diễn xoắn ốc bên phải được cài đặt với chu kì 28 ngày ngay lập tức tiết lộ sự tuần hoàn trong dữ liệu. Sự khác biệt có ý nghĩa thống kê về số trường hợp bệnh cúm được báo cáo lần lượt vào Chủ Nhật và Thứ Hai là khá rõ ràng. Chúng ta cũng sẽ thấy sự tuần hoàn này nếu đặt độ dài chu kỳ là 7 hoặc 14 ngày hoặc một bội (nhỏ) của 7 ngày.



Hình 1: Đặc điểm của thời gian: tuyến tính so với biểu diễn theo chu kỳ của thời gian: những hiểu biết khác nhau có thể thu được từ các biểu diễn trực quan tùy thuộc vào việc tuyến tính hay tuần hoàn

Ví dụ này đã chứng minh một cách dễ hiểu rằng việc quan sát đặc tính thời gian có thể cải thiện đáng kể ý nghĩa của việc biểu diễn trực quan bằng hình ảnh. Do đó, điều quan trọng là phải chọn một biểu diễn trực quan phù hợp với đặc điểm của dữ liệu (theo chu kỳ trong trường hợp này) và tham số hóa phù hợp để có thể phát hiện các mẫu lặp lại ẩn trong dữ liệu.

Trong phần tiếp theo, chúng ta sẽ nghiên cứu các đặc điểm của thời gian và dữ liệu hướng thời gian, thứ cần được xem xét và xử lý để lựa chọn kỹ thuật phù hợp.

2 Định nghĩa: đặc tính của dữ liệu hướng thời gian

Phần này bao gồm các khía cạnh chính để mô tả thời gian và dữ liệu hướng thời gian. Điều quan trọng là phải phân biệt rõ giữa thời gian vật lý và mô hình thời gian trong các hệ thống thông tin. Khi mô hình hóa thời gian trong các hệ thống thông tin, mục tiêu không phải là bắt chước hoàn toàn thời gian vật lý, tuy nhiên để cung cấp mô hình phù hợp nhất để phản ánh các hiện tượng đang được xem xét và hỗ trợ các bài toán phân tích (bằng tay). Hơn nữa, theo Frank [132], không có mô hình đúng duy nhất, có nhiều cách để mô hình hóa thời gian trong các hệ thống thông tin và thời gian cũng được mô hình hóa trong các ứng dụng khác nhau phụ thuộc vào từng bài toán. Các nghiên cứu sâu rộng đã được tiến hành để hình thành khái niệm về thời gian trong nhiều lĩnh vực của khoa học máy tính bao gồm cả trí tuệ nhân tạo, khai phá dữ liệu, mô phỏng, mô hình, cơ sở dữ liệu ... Chúng tôi phỏng theo các nghiên cứu của Frank [132] và Goralwalla cùng cộng sự [154], trong đó các khía cạnh trực giao chính được trình bày để mô tả các khía cạnh khác nhau của các loại thời gian. Dữ liệu cơ bản được trình bày trong chương 2, chúng tôi tập trung vào các đặc tính của thời gian và dữ liệu hướng thời gian nói riêng. Những khía cạnh này sẽ được mô tả chi tiết sau đây.

2.1 Các đặc tính của thời gian

Các đặc tính của thời gian có thể được chia thành các khía cạnh chung yêu cầu mô hình thời gian đầy đủ cũng như tổ chức phân cấp của thời gian và định nghĩa các yếu tố cụ thể. Các khía cạnh tổng quát bao gồm *thang đo*, *phạm vi*, *sắp xếp* và *góc nhìn*.

- *Thang đo*: *thứ tự*, *rời rạc*, *liên tục*. Ở góc độ thứ nhất, chúng tôi xem xét thời gian theo thang đo dọc mà các thành phần của mô hình đã được cho trước. Trong miền thời gian theo *thứ tự*, chỉ có quan hệ thứ tự tương đối được biểu diễn (ví dụ như: trước, sau, trong). Trong miền *rời rạc*, khoảng thời gian cũng được xem xét. Các giá trị thời gian có thể được ánh xạ tới một tập các số nguyên, cho phép lập mô hình định lượng các giá trị thời gian. Miền thời gian rời rạc dựa trên đơn vị (ví dụ: giây, phút). Mô hình thời gian *liên tục* đặc trưng bởi ánh xạ đến các số thực. Nghĩa là giữa hai mốc thời gian, tồn tại một mốc thời gian khác (cũng có thể hiểu là mô hình mật độ thời gian).
- *Phạm vi*: dựa trên *thời điểm*, dựa trên *khoảng*. Chúng tôi xem xét phạm vi của các yếu tố cơ bản cấu thành nên miền thời gian. Thời gian dựa trên *thời điểm* có thể được biểu diễn như các điểm Euclide rời rạc trong không gian, tức là có mốc thời gian bằng 0. Như vậy không có thông tin được đưa về khoảng giữa 2 thời điểm. Trái ngược với đó, miền dựa theo *khoảng* có độ lớn lớn hơn 0. Khái niệm này cũng có liên quan chặt chẽ đến khái niệm về độ chi tiết, sẽ được bàn luận sau. Ví dụ, giá trị thời gian ngày 1/5/2014 có thể liên quan đến một thời điểm riêng lẻ là 2014-05-01 00:00:00 là một điểm theo miền dựa trên *thời điểm* cũng có thể là đoạn [2014-05-01 00:00:00, 2014-05-01 23:59:99] trong miền dựa trên *khoảng*
- *Sắp xếp*: *tuyến tính*, *tuần hoàn*. Tương tự với nhận thức tự nhiên, chúng ta coi thời gian là một quá trình *tuyến tính* từ quá khứ đến tương lai, tức là mỗi giá trị thời gian có một giá trị duy nhất ở hiện tại cũng như tương lai. Trong cách sắp xếp *tuần hoàn* thời gian được xem xét theo các giá trị định kì (ví dụ mùa trong năm).
- *Góc nhìn*: *thứ tự*, *nhánh*, *đa góc nhìn*. Miền thời gian theo *thứ tự* xem xét sự kiện xảy ra sau sự kiện khác. Chi tiết hơn, chúng ta cũng có thể phân biệt giữa thứ tự toàn phần và thứ tự một phần. Trong miền thứ tự toàn phần, chỉ một sự kiện có thể xảy ra trong một thời điểm. Trái với đó, các sự kiện đồng thời hoặc chồng chéo được cho phép trong miền thứ tự một phần. Một dạng phức tạp hơn của mô hình thời gian được gọi là *nhánh* thời gian. Trong mô hình này, nhiều nhánh của thời gian có thể được mô tả với các kịch bản khác nhau (ví dụ mô hình lập kế hoạch). Trái ngược với thời gian phân nhánh, chỉ có một đường thực sự diễn ra theo thời gian, mô hình *đa góc nhìn* cho phép nhiều phương án xảy ra đồng thời theo thời gian. Ví dụ nhiều nhân chứng mô tả cùng một tình huống, mỗi người có một góc nhìn khác nhau và kể một câu chuyện khác nhau.

Phân loại thứ bậc thời gian và các yếu tố thời gian cụ thể được xác định dựa trên độ chi tiết, đơn vị thời gian và tính xác định.

- *Độ chi tiết và lịch*: *rỗng*, *đơn* và *nhiều*. Để giải quyết độ phức tạp về thời gian và đưa ra các độ chi tiết khác nhau, chúng ta có thể sử dụng vài giải thích ngắn gọn hữu ích. Về cơ bản *độ chi tiết* là một khái niệm trừu tượng về thời gian (do chúng ta tự định nghĩa) để hình dung thời gian một cách dễ dàng hơn trong cuộc sống (chẳng hạn như giờ, phút, giây). Tổng quan hơn, độ chi tiết mô tả ánh xạ từ thời gian đến

các đơn vị nhỏ hơn hoặc lớn hơn. Nếu độ chi tiết và hệ thống lịch được hỗ trợ bởi mô hình thời gian, chúng tôi mô tả nó như *nhiều* chi tiết. Bên cạnh biến thể phức tạp này, có thể chỉ có một *đơn* chi tiết hoặc là *không* có khái niệm trừu tượng nào trong này được hỗ trợ.

- *Nguyên thủy thời gian*: *tức thời*, *khoảng thời gian*, *nhiệt*. Những nguyên thủy thời gian này có thể được xem như một lớp trung gian giữa các phần tử dữ liệu và miền thời gian. Về cơ bản, thời gian nguyên thủy có thể được chia thành neo nguyên thủy (tuyệt đối) và không neo (tương đối). *tức thời* và *khoảng thời gian* thuộc nhóm đầu tiên, tức là chúng nằm trên một ví trí cố định trên miền thời gian. Ngược lại, một *nhiệt* là một tính chất tương đối, tức là nó không có vị trí tuyệt đối trên miền thời gian. Tức thời là mô hình dành cho các điểm riêng lẻ trên trực thời gian (đôi khi còn gọi là thời điểm, ví dụ 10/5/2014), khoảng thời gian là khoảng giữa hai thời điểm (ví dụ từ 10/5/2014 đến 16/5/2014), nhiệt là thời lượng (của khoảng) mà không có mốc cố định (ví dụ 6 ngày).
- *Tính xác định*: *xác định*, *không xác định*. Sự không chắc chắn là một khía cạnh quan trọng khác cần được xem xét của dữ liệu hướng thời gian. Nếu không có thông tin đầy đủ hay chính xác về thời gian hoặc nếu thời gian gốc được chuyển đổi từ độ chi tiết này sang độ chi tiết khác thì sự không chắc chắn sẽ xuất hiện và cần được xử lý. Ví dụ cho điều này là sự không chính xác đến từ kiến thức thực tế (ví dụ thời điểm trái đất hình thành), các dữ liệu được lập kế hoạch trong tương lai (ví dụ nhiệm vụ cần khoảng 2 đến 3 tháng để hoàn thành) hoặc là các sự kiện không chắc chắn (ví dụ khoảng 2 đến 3 ngày trước). Lưu ý rằng tính không xác định thời gian cũng như tính tương đối của các tham chiếu đến thời gian chủ yếu là các tiêu chuẩn của các phát biểu hơn là các sự kiện mà chúng biểu thị. Tính không xác định có thể được hiểu bởi các thông số rõ ràng (ví dụ bắt đầu sớm nhất và bắt đầu gần nhất của một khoảng thời gian) hoặc ngầm hiểu trong trường hợp đa độ chi tiết nhỏ. Ví dụ, xem xét một câu khẳng định "Hoạt động A bắt đầu vào ngày 14 tháng 5 năm 2014 và kết thúc vào ngày 17 tháng 5 năm 2014" - khẳng định này có thể được mô hình hóa bằng điểm bắt đầu "14/5/2014" và điểm kết thúc "17/5/2014" cả hai đều có độ chi tiết tính đến ngày. Nếu ta nhìn vào khoảng này với độ chi tiết theo giờ, khoảng có thể bắt đầu và kết thúc vào các điểm giữa 0 AM và 12 PM của một ngày nhất định. Chính vì vậy, *tính xác định* của tham số thời gian cho trước cần phân được phân tích. Một thông số được biểu diễn khi có kiến thức đầy đủ về tất cả các khía cạnh của thời gian.

Trong phần tiếp theo, chúng tôi khám phá và xác định dữ liệu hướng thời gian một cách chi tiết hơn.

2.2 Các đặc tính của dữ liệu hướng thời gian

Cũng như các tính chất của thời gian, dữ liệu có tác động lớn đến việc thiết kế phương pháp trực quan hóa. Chúng ta cùng lược qua những yếu tố chính của dữ liệu liên quan đến thời gian.

- *Quy mô*: *định tính*, *định lượng*. Dữ liệu định lượng được dựa trên một thang đo (rời rạc hoặc liên tục). Dữ liệu định tính mô tả tập hợp các phần tử dữ liệu có thứ tự hoặc không có thứ tự.

- *Hệ quy chiếu: trừu tượng, không gian.* Dữ liệu trừu tượng (ví dụ tài khoản ngân hàng) được thu thập trong bối cảnh phi không gian. Dữ liệu không gian (ví dụ dữ liệu điều tra dân số) chứa thông tin về không gian ví dụ như vị trí địa lý.
- *Loại dữ liệu: sự kiện, trạng thái.* Sự kiện có thể được hiểu là các điểm đánh dấu sự thay đổi trạng thái (ví dụ sự khởi hành của đoàn tàu), trong khi đó trạng thái đặc trưng cho các giai đoạn liên tục giữa các sự kiện (ví dụ đoàn tàu đang chạy trên đường sắt).
- *Số biến: đơn biến, đa biến.* Dữ liệu đơn biến chỉ chứa một giá trị dữ liệu tại một thời điểm, trong trường hợp đa biến, dữ liệu tại mỗi thời điểm thể hiện bằng nhiều giá trị dữ liệu.

Những phạm trù cơ bản này tạo thành cơ sở cho việc lựa chọn, thiết lập kĩ thuật trực quan hóa phù hợp cho dữ liệu hướng thời gian.

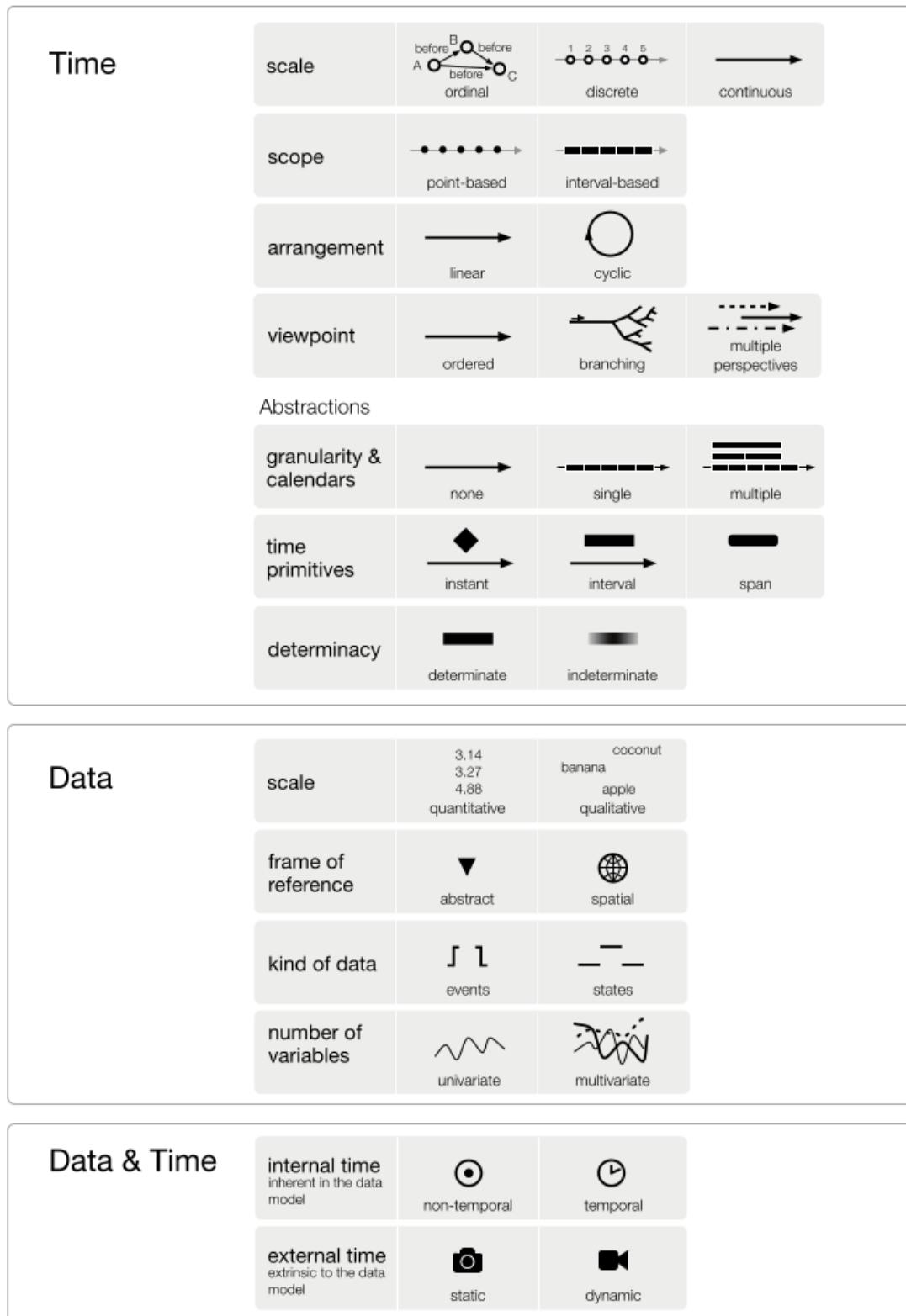
2.3 Dữ liệu liên quan và thời gian

Các khía cạnh liên quan của dữ liệu phụ thuộc vào thời gian đã được kiểm tra rộng rãi trong lĩnh vực cơ sở dữ liệu về thời gian [274,397]. Ở đây, chúng tôi điều chỉnh và phát triển các định nghĩa trong lĩnh vực này. Theo đó, bất kì tập dữ liệu nào cũng có liên quan đến 2 miền thời gian: (1) thời gian bên trong và (2) thời gian bên ngoài.

Thời gian bên trong được coi là chiều thời gian vốn có của mô hình dữ liệu. Thời gian bên trong mô tả thời điểm thông tin tồn tại trong tập dữ liệu là hợp lệ. Ngược lại *thời gian bên ngoài* là thời gian nằm ngoài mô hình dữ liệu. Thời gian bên ngoài là cần thiết để mô tả cách một tập dữ liệu phát triển theo thời gian. Dựa vào số biến thời gian nguyên thủy của thời gian bên trong và thời gian bên ngoài, dữ liệu liên quan đến thời gian có thể được phân loại như sau:

- *Dữ liệu tĩnh phi thời gian.* Nếu cả thời gian bên trong lẫn bên ngoài bao gồm một yếu tố thời gian thì dữ liệu hoàn toàn không phụ thuộc vào thời gian. Trong chương này chúng ta sẽ không nghiên cứu kiểu dữ liệu đó.
- *Dữ liệu thời gian tĩnh.* Nếu thời gian bên trong chứa nhiều hơn một yếu tố thời gian nguyên thủy, trong khi thời gian bên ngoài chỉ chứa một thì dữ liệu có thể được xem là phụ thuộc vào thời gian. Vì các giá trị được lưu trong dữ liệu phụ thuộc vào thời gian bên trong, dữ liệu thời gian tĩnh có thể được hiểu là một cái nhìn lịch sử về cách một số mô hình xem xét các yếu tố khác nhau của thời gian trong. Chuỗi thời gian chung là một ví dụ tiêu biểu về dữ liệu thời gian tĩnh.
- *Dữ liệu động phi thời gian.* Nếu thời gian bên trong chỉ chứa một mà thời gian bên ngoài chứa nhiều yếu tố thời gian nguyên thủy thì dữ liệu phụ thuộc vào thời gian bên ngoài. Để cho dễ hiểu, dữ liệu thay đổi qua thời gian thì chúng được gọi là động. Vì thời gian bên trong không được xem xét nên chỉ có trạng thái hiện tại của dữ liệu là bảo toàn. Cái nhìn quá khứ không được duy trì ở đây nữa. Có vài kĩ thuật trực quan hóa sẵn có tập trung vào dữ liệu động phi thời gian, tuy nhiên do thời gian bên trong và thời gian bên ngoài có thể ánh xạ lẫn nhau, một số kĩ thuật trực quan hóa cho dữ liệu tĩnh cũng có thể áp dụng được.

- *Dữ liệu động theo thời gian.* Nếu cả thời gian bên trong và bên ngoài đều có nhiều yếu tố thời gian nguyên thủy, thì dữ liệu được xem xét là phụ thuộc vào thời gian. Nói cách khác, dữ liệu chứa các biến phụ thuộc vào thời gian (bên trong) và trạng thái dữ liệu cũng thay đổi theo thời gian (bên ngoài). Thường thì trong trường hợp này, thời gian bên trong và bên ngoài phụ thuộc lẫn nhau và có thể ánh xạ vào nhau. Sự phân biệt rõ ràng giữa thời gian bên trong và bên ngoài thường không được thực hiện bởi các phương pháp trực quan hóa hiện tại bởi xem xét cả 2 chiều thời gian để trực quan hóa là một thách thức.



Hình 2: Các khía cạnh của dữ liệu hướng thời gian

3 Trực quan hóa dữ liệu hướng thời gian

Dữ liệu và thời gian đã được trình bày trong phần trước, tuy nhiên chúng ta cũng phải xem xét các vấn đề về thiết kế ở cấp độ biểu diễn trực quan. Để biểu diễn sự phụ thuộc

vào thời gian của dữ liệu, ta cần đưa vào trực thời gian. Sự đa dạng của các kĩ thuật trực quan hóa bao gồm các cách tiếp cận rất khác nhau. Để tóm lược sự đa dạng này, chúng tôi tập trung vào 2 tiêu chí cơ bản:

- *Ánh xạ của thời gian.* Có hai lựa chọn cho ánh xạ thời gian: ánh xạ thời gian vào không gian và ánh xạ từ thời gian vào thời gian. Khi nói đến ánh xạ từ thời gian vào không gian, điều đó có nghĩa là thời gian và dữ liệu được thể hiện trong một hình ảnh duy nhất. Biểu diễn này không tự thay đổi theo thời gian, đó là lý do tại sao chúng tôi gọi nó là trực quan hóa dữ liệu hướng thời gian *tĩnh*. Ngược lại biểu diễn *động* sử dụng thời gian vật lý để truyền đạt sự phụ thuộc vào thời gian của dữ liệu, nghĩa là thời gian ánh xạ vào thời gian. Các kết quả này trong biểu diễn tự thay đổi theo thời gian (ví dụ bản trình chiếu hay hiệu ứng hoạt họa). Lưu ý rằng, có hay không sự tương tác để điều hướng thời gian không ảnh hưởng đến việc cách tiếp cận trực quan hóa là tĩnh hay động.
- *Kích thước của không gian biểu diễn.* Chúng ta có thể phân biệt giữa biểu diễn 2D và 3D của dữ liệu hướng thời gian. Cách trực quan hóa sử dụng không gian 2D phải đảm bảo nhấn mạnh trực thời gian vì chiều thời gian và dữ liệu thường có chung biểu diễn sẵn có. Trong trường hợp biểu diễn 3D, chiều hiển thị thứ 3 được đưa vào. Thực tế, nhiều kĩ thuật sử dụng nó như một chiều dành riêng cho trực thời gian, tách thời gian ra khỏi các chiều dữ liệu khác một cách rõ ràng.

3.1 Phân loại

Để thuận lời cho việc phân loại một cách dễ dàng và giữ cho phân loại các kĩ thuật trực quan hóa đơn giản, chúng tôi tập trung vào những khía cạnh cốt lõi của dữ liệu, thời gian và trực quan hóa. Qua nhiều ví dụ trực quan hóa, chúng tôi sẽ minh họa những khả năng ứng dụng của chúng và những tính năng của các kĩ thuật trực quan hóa này.

- **Dữ liệu**
 - *Khung tham chiếu* - trừu tượng, không gian
 - *Các biến* - đơn biến, đa biến
- **Thời gian**
 - *Sắp xếp* - tuyến tính, tuần hoàn
 - *Thời gian nguyên thủy* - thời điểm, khoảng
- **Trực quan hóa**
 - *Ánh xạ* - tĩnh, động
 - *Số chiều* - 2D, 3D

Để thiết lập những kĩ thuật trực quan hóa khác nhau cho dữ liệu hướng thời gian, chúng ta cần trả lời những câu hỏi sau:

- *Cái gì* được biểu diễn ? *Thời gian và dữ liệu*
- *Tại sao* nó được biểu diễn ? *Tác vụ của người dùng*

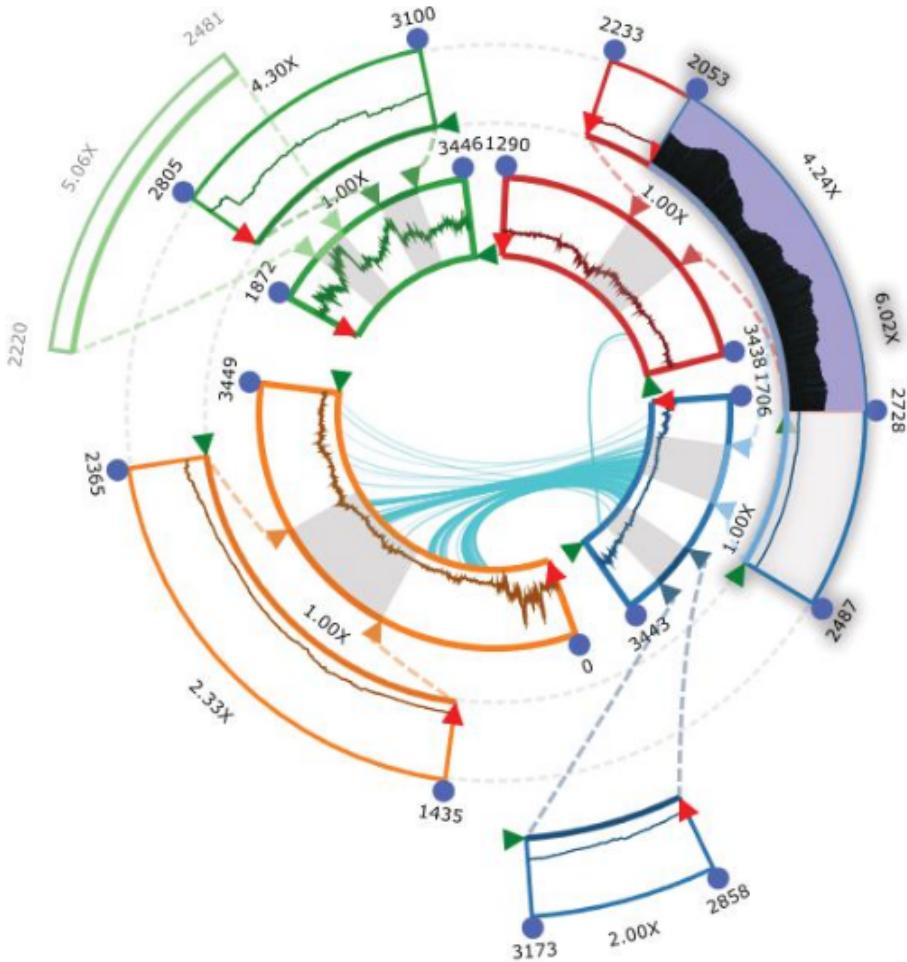
- Nó được biểu diễn thế nào ? *Biểu diễn trực quan*

Trong chương này, chúng tôi chủ yếu nói về thời gian, dữ liệu và biểu diễn trực quan, bỏ qua những tác vụ của người dùng để mọi thứ trở nên dễ hiểu nhất có thể. Các kĩ thuật trực quan hóa đòi hỏi phải hiểu rõ những khâu cụ thể được thực hiện trong quá trình khai phá dữ liệu và trực quan hóa. MacEachen [281] đã đề xuất một phương pháp ở cấp độ thấp giả quyết một phần cụ thể trong miền thời gian. Các khâu này được định nghĩa bởi một tập các câu hỏi quan trọng mà người dùng có thể tìm kiếm câu trả lời bằng biểu diễn trực quan hóa, ví dụ như *sự tồn tại của thành phần dữ liệu*: liệu thành phần này có tồn tại trong một thời điểm (khoảng) nào đó, hoặc là *tốc độ thay đổi*: thành phần dữ liệu này thay đổi nhanh như thế nào theo thời gian.

Trong phần tiếp theo, chúng tôi sẽ đưa ra các ví dụ cho mỗi loại được nêu trên.

3.2 Dữ liệu: Khung tham chiếu - trừu tượng

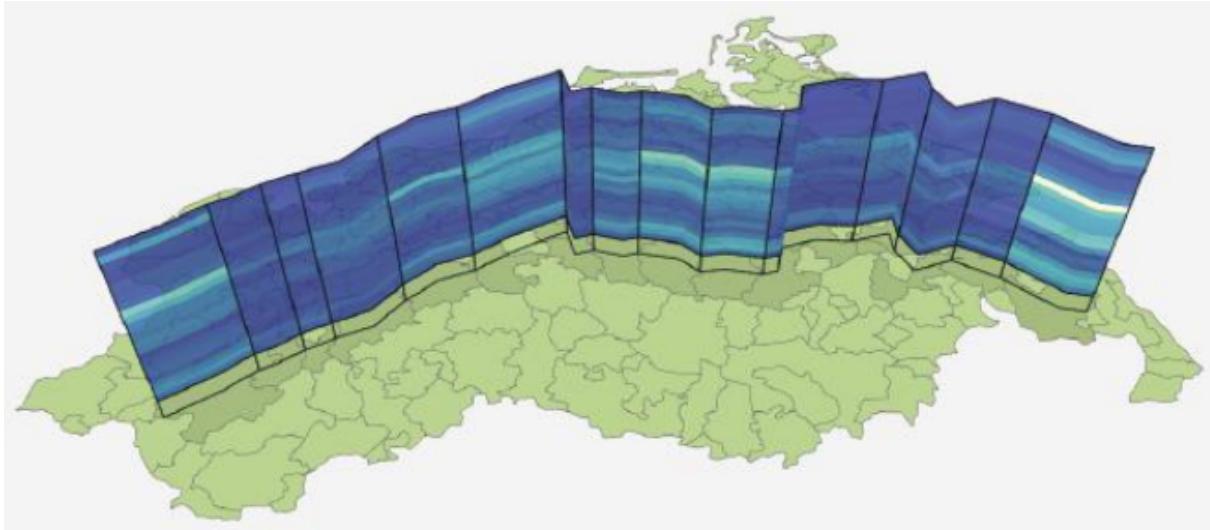
KronoMiner [488] là một công cụ thăm dò chuỗi thời gian đa năng cung cấp khả năng điều hướng phong phú và hỗ trợ phân tích (hình (3)). Biểu diễn của nó dựa trên bối cảnh tâm phân cấp, cho phép người dùng có thể quan sát chi tiết bằng cách tập trung vào các mảnh khác nhau. Các mảnh dữ liệu có thể được xoay, kéo, co dãn hoặc thu nhỏ một cách dễ dàng, hỗ trợ nhiều kiểu phân tích dữ liệu thời gian. KronoMiner cũng giới thiệu 2 kỹ thuật phân tích: (1) MagicAnalytics Lens, thể hiện mối tương quan của 2 thành phần dữ liệu xếp chồng lên nhau và (2) chế độ Best Match, trong đó một hình cung được hiển thị thể hiện sự tương tự của hai thành phần dữ liệu theo một phép đo cụ thể.



Hình 3: KronoMiner [488]. KronoMiner là một công cụ khám phá chuỗi thời gian đa mục đích cung cấp các tính năng điều hướng và phân tích phong phú

3.3 Dữ liệu: Khung tham chiếu - Không gian

Tominski và Schulz [415] giới thiệu kĩ thuật trực quan hóa cho dữ liệu thời gian có yếu tố không gian địa lý 2D và thời gian tuyến tính 1D. Ý tưởng là xây dựng một lát cắt không phẳng được gọi là *Great Wall of Space-Time* (hình (4)) thông qua không-thời gian liên tục 3D (2D+1D). Kiến trúc bức tường này được xây dựng dựa trên các khía cạnh tô tô và hình học của không gian địa lý. Dựa trên một đồ thị lân cận, một đường tô pô được thiết lập tự động hoặc có tương tác. Đường tô pô này được biến đổi thành đường địa lý thông qua những tính chất địa lý của khu vực trên bản đồ. Đường này được dựng thành một bức tường 3D mà chiều thứ 3 có thể được dùng để ánh xạ đến miền thời gian. Những biểu diễn trực quan khác nhau có thể được chiếu lên bức tường này để biểu thị dữ liệu. Bức tường có lợi thế là nó thể hiện một đường khép kín qua không gian mà không có khoảng trống giữa các pixel mang thông tin trên màn hình.

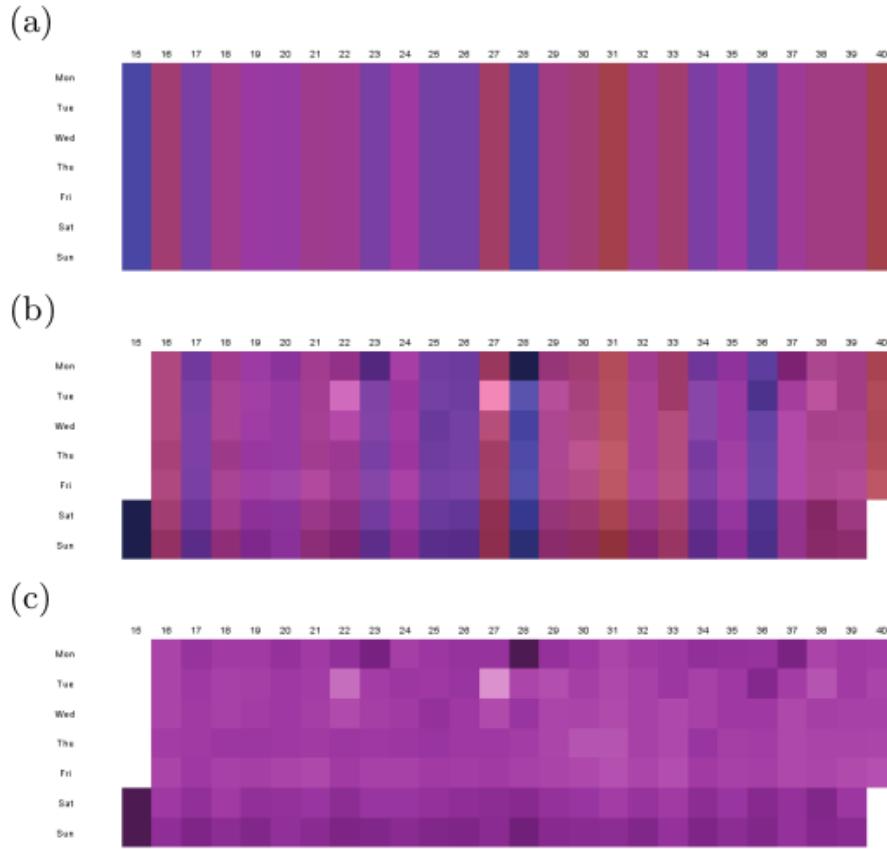


Hình 4: Biểu diễn trực quan hóa của dữ liệu sức khỏe con người sử dụng Great Wall of Space-Time. Một đường trong không gian được thiết lập. Dọc theo đường này, bức tường thể hiện 24 tháng của dữ liệu tại mỗi vùng nó đi qua. Các màu tối thể hiện giá trị dữ liệu thấp, màu sáng thể hiện giá trị cao. Hình trên mô tả số ca mắc cúm

3.4 Dữ liệu: Số biến - Đơn biến

Trực quan hóa GROOVE (Granularity Overview OVErlay) [262] mở rộng trực quan hóa dựa trên pixel bằng cách phủ lên lớp tổng hợp một hoặc nhiều trong 3 phương pháp: (1) lớp phủ theo màu, (2) lớp phủ theo độ mờ hoặc (3) lớp phủ không gian. Các mức tổng hợp là kết quả của các mức độ chi tiết thời gian khác nhau. Các lớp phủ cho phép đọc vi mô và vĩ mô và trán chuyển động của mắt giữa việc quan sát tổng quan và chi tiết. Với mục đích minh họa, chúng tôi giữ sự hiển thị đơn giản và lượng dữ liệu trực quan không nhiều. Không gian trống xuất hiện do sự bất thường của việc kết hợp các chi tiết tuần và tháng.

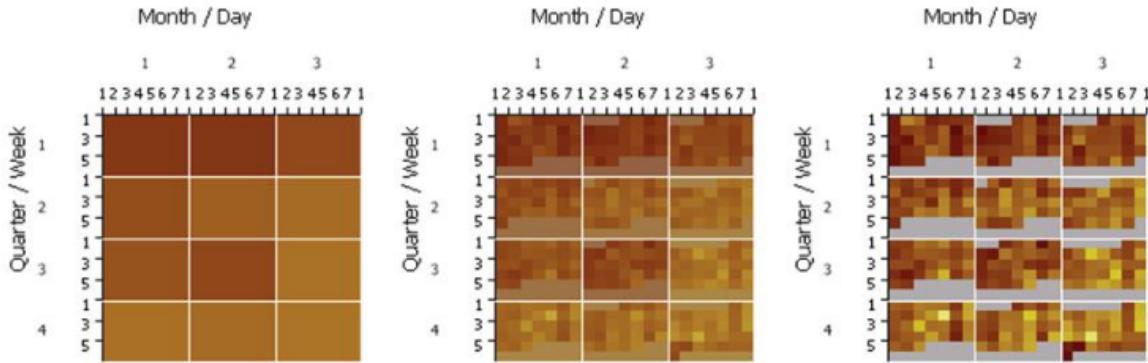
Đầu tiên, các thành phần màu có thể được sử dụng với lớp phủ màu (hình (5)). Hình này vẽ sơ đồ dữ liệu lưu lượng trong vài tuần. Màu sắc thay đổi từ xanh sang đỏ cho mức thấp hơn hoặc cao hơn trung bình hàng tuần. Hình 5 (a) chỉ thể hiện các giá trị theo tuần. Độ sáng của một ngày cụ thể cao hơn trong trường hợp lưu lượng trong ngày đó cao hơn. Hình 5 (c) chỉ thể hiện giá trị theo ngày. Hình 5 (b) thể hiện lớp phủ màu.



Hình 5: Trực quan hóa GROOVE kết hợp đọc chi tiết và tổng quan sử dụng bố cục thông thường dựa trên mức độ chi tiết của thời gian. Ví dụ biểu diễn dữ liệu lấy từ tập Dodgers Loop Sensor [195]. Ở đây lớp phủ màu được sử dụng: (a) Lưu lượng trung bình trên một đoạn đường cao tốc (tuần trong năm từ 15 đến 40) được ánh xạ màu từ xanh qua tím qua đỏ. (b) Lưu lượng trung bình theo ngày được ánh xạ bằng độ sáng. (c) Lưu lượng trung bình theo ngày được ánh xạ bằng độ sáng. (Nguồn: GROOVE modules for TimeBench [339])

Thứ hai, lớp phủ mờ (xem Hình (6)) áp dụng hiệu ứng xen kẽ tương tác giữa hiển thị tổng quan và hiển thị chi tiết. Hình thứ 2 của chúng tôi biểu diễn dữ liệu doanh thu của một cửa hàng trong một năm và sử dụng bố cục đê quy [223]. Các hàng và cột của các khối lớn hơn được kết hợp với sự sắp xếp các pixel trong các khối cho cấu trúc chi tiết. Các sắp xếp khác nhau có thể được chọn lựa (ví dụ hàng-hàng, sau-trước). Mỗi khối đại diện cho một tháng và các khối được sắp xếp với mỗi hàng cho mỗi quý. Trong các khối, các pixels được sắp xếp với mỗi hàng cho mỗi tuần và mỗi pixel cho một ngày giống như tờ lịch.

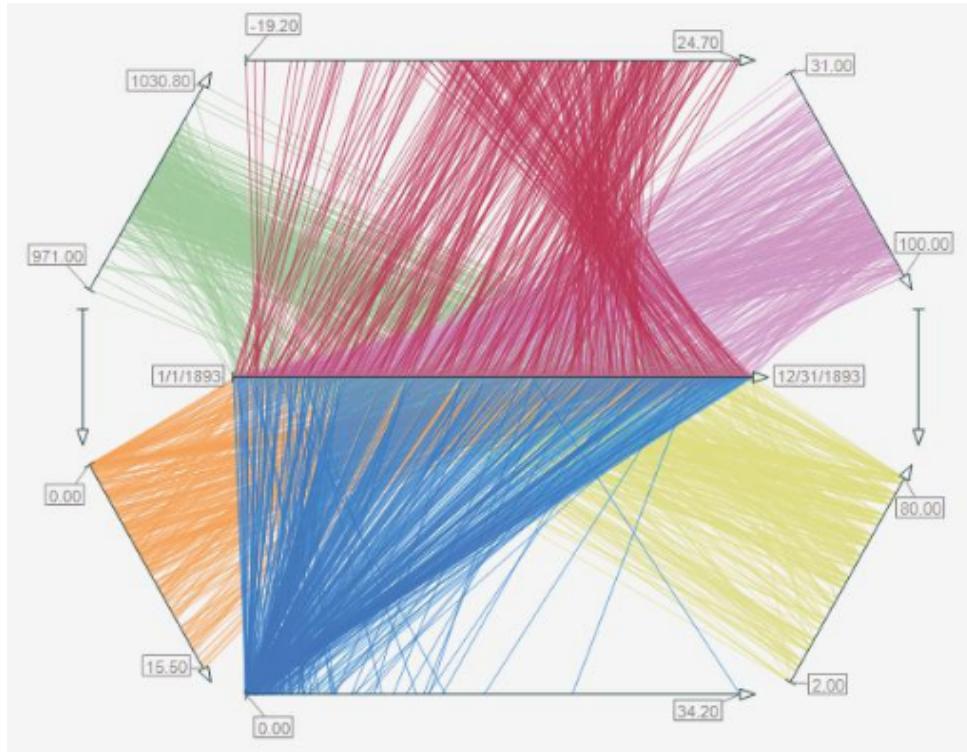
Thứ ba, lớp phủ không gian có thể được sử dụng bằng cách hiển thị các giá trị trung bình như viền bao xung quanh các giá trị chi tiết. Lớp phủ không gian có thể được kết hợp với hai phương pháp khác. Nó cũng có thể được áp dụng sử dụng tương tác bằng cách mở rộng hoặc thu hẹp diện tích theo nhu cầu.



Hình 6: Trực quan hóa GROOVE sử dụng lớp phủ mờ. Dữ liệu được vẽ cho doanh thu hàng ngày của một cửa hàng trong 1 năm; mỗi khối mô tả một tháng: màu tối thể hiện giá trị thấp và vàng sáng thể hiện giá trị cao. Trong ví dụ này, tổng quan theo tháng được kết hợp với chi tiết ngày sử dụng nhiều độ mờ khác nhau. (Nguồn: GROOVE)

3.5 Dữ liệu: Số biến - Đa biến

TimeWheel [417] là một kỹ thuật để trực quan hóa nhiều biến phụ thuộc thời gian. TimeWheel bao gồm một trực thời gian và nhiều trực cho các biến dữ liệu khác. Trực thời gian được đặt ở trung tâm của hình vẽ để nhấn mạnh đặc tính thời gian của dữ liệu. Các trực của biến dữ liệu khác được ánh xạ bởi các màu khác nhau và được sắp xếp theo hình tròn quanh trực thời gian. Để trực quan hóa dữ liệu, các đường tỏa ra từ trực thời gian đến mỗi trực khác thiết lập một kết nối giữa thời điểm và các giá trị dữ liệu liên quan. Những đường này tạo thành các nhịp điệu để người dùng xác định mối tương quan với trực thời gian, xu hướng hoặc ngoại lai. Mỗi mẫu như vậy có thể được phân biệt một cách rõ nhất với các trực song song với trực thời gian. Để tập trung vào trực mong muốn, người dùng có thể xoay TimeWheel. Các trực được tập trung có thể được nhấn mạnh bằng cách kéo dài ra. Các trực vuông góc với trực thời gian thì khó giải thích hơn, vì thế chúng được làm nhạt bớt màu đi và co lại. Khám phá tương tác, bao gồm điều hướng trong thời gian được hỗ trợ thông qua các trực tương tác khác nhau.



Hình 7: Trung tâm của TimeWheel [417] là trục thể hiện thời gian. Các trục khác thể hiện các biến phụ thuộc vào thời gian, hình này thể hiện 8 biến khác nhau tương ứng với 8 trục. Ở giữa, các đường đỏ thể hiện nhiệt độ trung bình, nó tăng khi bắt đầu và giảm về cuối trục thời gian. Các đường xanh dương là lượng mưa. Có một vài ngoại lệ với lượng mưa cao nhưng lượng mưa cả năm thì không bất thường (Nguồn: VisAxes)

3.6 Thời gian: Sắp xếp theo trình tự thời gian tuyến tính

Một trong những cách đơn giản nhất để biểu diễn dữ liệu chuỗi thời gian là sử dụng hệ tọa độ Descartes với trực hoành thể hiện mốc thời gian và trực tung thể hiện giá trị biến định lượng tương ứng. Mỗi cặp gồm mốc thời gian và giá trị tương ứng được thể hiện bằng một điểm trên hệ tọa độ. Cách biểu diễn này thường được gọi là *đồ thị điểm*, *biểu đồ điểm*, hoặc *biểu đồ phân tán* (scatterplot). Harris [170] mô tả cách biểu diễn này là một dạng đồ thị hai chiều trong đó độ lớn của các biến định lượng được thể hiện thông qua khoảng cách của mỗi điểm đến các trục tọa độ. Ngoài dạng cơ bản này còn có nhiều biến thể mở rộng khác, chẳng hạn như kỹ thuật 3D (biểu đồ lớp) hoặc kỹ thuật sử dụng nhiều ký hiệu khác nhau thay vì chỉ đơn thuần là các điểm giống nhau. Kỹ thuật này đặc biệt phù hợp trong trường hợp muốn nhấn mạnh từng giá trị riêng lẻ. Bên cạnh đó, việc biểu diễn dữ liệu bằng các thang đo chung thống nhất sẽ giúp cho người đọc có thể dễ dàng nắm bắt được thông tin một cách trực quan và chính xác nhất.

Phát triển trên nền tảng của biểu đồ điểm, *biểu đồ đường* cho phép ta thể hiện được mối liên hệ về mặt thời gian giữa các điểm dữ liệu thông qua việc gắn các điểm dữ liệu đó trên một đường nối. Khác với biểu đồ điểm khi tập trung nhấn mạnh vào các điểm dữ liệu riêng lẻ, rời rạc, biểu đồ đường sẽ giúp ta tập trung vào xu hướng tổng quan của dữ liệu theo thời gian. Tùy thuộc vào đối tượng đang được xem xét đến ta có thể sử dụng các kiểu kết nối khác nhau giữa các điểm dữ liệu như: đường thẳng, đường bậc thang (liên

tục thay đổi giá trị) hoặc đường cong Bezier. Tuy nhiên, dù sử dụng bất kỳ kiểu kết nối nào nêu trên cũng cần lưu ý rằng việc phản ánh giá trị của các điểm nằm giữa hai điểm dữ liệu đã biết đều chỉ mang tính tương đối và ta sẽ không thể đảm bảo chắc chắn về độ chính xác cho mọi điểm dữ liệu được thể hiện trên đường nối. Một vấn đề khác cần lưu ý là việc thiếu dữ liệu. Trong các trường hợp thiếu dữ liệu của một khoảng thời gian nhất định mà người phân tích bỏ qua yếu tố này và chỉ đơn thuần kết nối các điểm dữ liệu liền trước và liền sau khoảng thời gian đó thì nhiều khả năng sẽ dẫn đến những kết luận không chính xác. Do đó, cần có cách thức phù hợp để biểu diễn cho người xem ý thức được vấn đề này, chẳng hạn bằng cách sử dụng các đường nối bằng nét đứt.

Bên cạnh các dạng biểu đồ nêu trên, ta có thể sử dụng một số dạng mở rộng và biến thể khác như: biểu đồ nhiệt, biểu đồ dải, biểu đồ đường lốp, biểu đồ miền, biểu đồ chỉ số hoặc biểu đồ kiểm soát (còn gọi là biểu đồ hành vi – quy trình) (xem [170]).



Hình 8: TiMoVA. Giao diện thể hiện biểu đồ chuỗi thời gian dưới dạng biểu đồ đường (dữ liệu đầu vào; trên cùng bên trái), công cụ lựa chọn mô hình và nhiều chế độ xem khác để định hướng, hỗ trợ quá trình lựa chọn mô hình. (Nguồn: Thực hiện bằng phần mềm nguyên mẫu TiMoVA.)

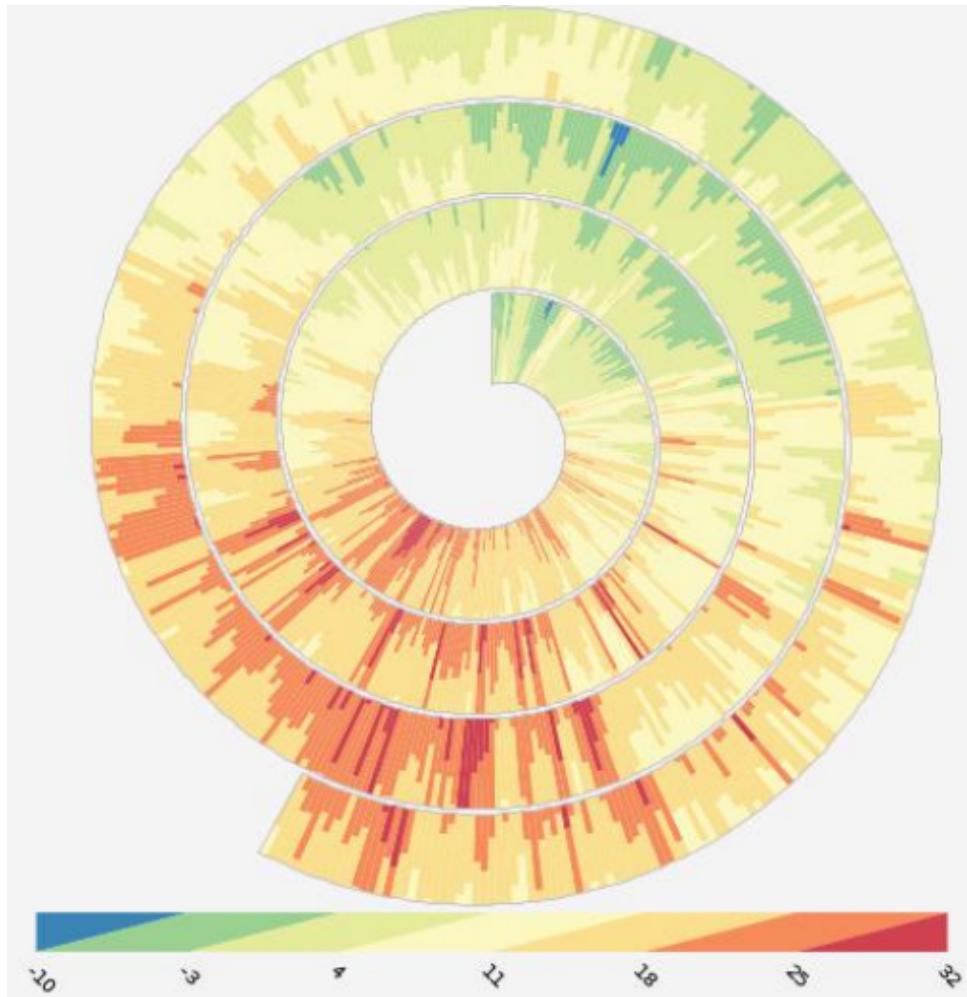
Các loại biểu đồ điểm, biểu đồ đường cũng như các biểu đồ trực quan hóa khác (ví dụ: biểu đồ cột) được sử dụng một cách thường xuyên trong quá trình *TiMoVA* (“*Ti*”: *Time series analysis - Phân tích chuỗi thời gian*, “*Mo*”: *Model selection - Lựa chọn mô hình*, “*VA*”: *applies Visual Analytics methods - áp dụng các phương pháp Phân tích hình ảnh trực quan*) [42]. Phân tích thống kê với dữ liệu chuỗi thời gian là một nhiệm vụ đầy thách thức và đòi hỏi phải được thực hiện bởi các chuyên gia trong nhiều lĩnh vực khác nhau, ví dụ như khi một nhân viên y tế công muôn dự đoán số người cần được điều trị tim mạch trong năm tới. Thông thường, việc lựa chọn mô hình là một nhiệm vụ phức tạp. Chính vì vậy, TiMoVA là một công cụ hữu hiệu, cung cấp giao diện tương tác để định hướng cho người dùng khai phá dữ liệu và lựa chọn mô hình chuỗi thời gian phù hợp. TiMoVA hỗ trợ chuyên gia trong các lĩnh vực ngành nghề trong việc: (1) lựa chọn thử tự mô hình tương tác thông qua một giao diện trực quan, (2) cung cấp phản hồi về kết quả mô hình trong quá trình chọn thử tự mô hình một cách trực quan và nhanh chóng tức thì, (3)

quyết định về việc có nên cải thiện, thay đổi mô hình hay không (Hình 7.8). Một đánh giá thực nghiệm sử dụng bộ dữ liệu dịch tễ và sự tham gia đánh giá của 02 (hai) chuyên gia thống kê cho thấy TiMoVA hỗ trợ rất hiệu quả trong việc tạo ra một môi trường tương tác khám phá dữ liệu từ đó giúp lựa chọn được các mô hình phù hợp.

3.7 Thời gian: Sắp xếp theo trình tự thời gian tuần hoàn

Tominski và Schumann [416] sử dụng mã màu hai tông (two-tone) nâng cao của Saito và cộng sự [354] để trực quan hóa dữ liệu phụ chuỗi thời gian dọc theo hình xoắn ốc. Mỗi đơn vị khoảng thời gian được thể hiện bằng một đoạn duy nhất trên xoắn ốc. Mỗi phân đoạn được chia thành hai phần tô màu theo phương pháp tô màu hai tông. Ưu điểm của việc sử dụng phương pháp tô màu hai tông là nó giúp hiện thực hóa khái niệm tổng quan + chi tiết bằng một thiết kế trực quan. Hai màu được sử dụng trên mỗi đoạn xoắn ốc cho phép người dùng nhanh chóng nhận ra khoảng giá trị của đoạn đó (tổng quan). Nếu ta quan tâm tới một khoảng giá trị nhất định, tỷ lệ của hai màu sẽ giúp biểu thị giá trị dữ liệu cụ thể một cách chính xác (chi tiết). *Xoắn ốc tương tác nâng cao* có thể được điều chỉnh tương tác theo nhiều cách khác nhau. Số lượng đơn vị giai đoạn thời gian, số chu kỳ và các tham số hình học bổ sung sẽ quyết định hình dạng của đường xoắn ốc. Việc biểu diễn dữ liệu chủ yếu được kiểm soát bởi các thang màu sắc đã sử dụng và các tham số như số lượng màu, hướng ánh xạ và chức năng ánh xạ (tuyến tính hay logarit). Việc điều hướng có thể được thực hiện tức thì thông qua việc điều chỉnh trực tiếp xoắn ốc.

Nội dung mục này và Hình (9) minh họa một hình xoắn ốc nâng cao với mã màu hai tông. Mã giả dưới đây là một ví dụ minh họa cách vẽ một hình xoắn ốc đơn giản trong đó màu sắc và độ rộng của các thanh thể hiện độ lớn giá trị của dữ liệu tương ứng (so sánh hình ảnh ở giữa và bên phải trong Hình (1)). Các tham số đầu vào bao gồm dữ liệu (*data[]*) - một mảng gồm n giá trị dữ liệu không âm với *max* là giá trị lớn nhất. Tọa độ tâm của hình xoắn ốc được ký hiệu là *cx* và *cy*. Bán kính trong và bán kính ngoài của hình xoắn ốc được cho bởi các tham số *inRad* và *outRad*. Số lượng giá trị dữ liệu hiển thị trong một chu kỳ xoắn ốc đầy đủ, hay nói cách khác là số lượng đoạn thẳng trên mỗi chu kỳ được cho bởi tham số *cyclLen*.

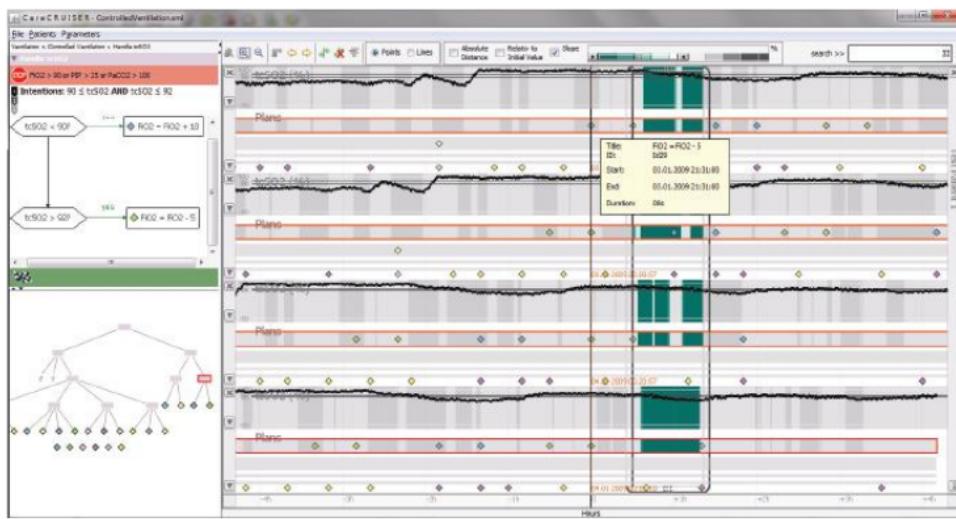


Hình 9: Xoắn ốc tương tác nâng cao [416]. Hình xoắn ốc biểu diễn dữ liệu thời tiết trong 3 năm 6 tháng. Mỗi chu kỳ thể hiện 365 giá trị nhiệt độ trung bình hàng ngày ở thành phố Rostock. Ta có thể thấy mùa đông những năm gần đây (chu kỳ ngoài cùng) có khí hậu ôn hòa hơn so với các mùa đông trước. (Nguồn: Hình vẽ được thực hiện bằng công cụ biểu diễn xoắn ốc tương tác nâng cao.)

3.8 Thời gian: Nguyên thủy thời gian – Tức thời

CareCruiser [163] là một hệ thống trực quan hỗ trợ việc xác định sơ bộ ảnh hưởng của các hoạt động lâm sàng đối với tình trạng của bệnh nhân. Để đạt được mục tiêu này, CareCruiser trực quan hóa các chi tiết trong pháp đồ điều trị kết hợp với các thông số của bệnh nhân, đặc biệt là thông số liên quan đến các quá trình diễn tiến thay đổi theo thời gian. Nó hỗ trợ khai phá thông tin thông qua việc căn chỉnh, đánh dấu màu, lọc và cung cấp thông tin về bối cảnh chung cũng như thông tin chi tiết cần tập trung. Sắp xếp các kế hoạch điều trị lâm sàng theo chiều dọc sẽ giúp đơn giản hóa việc so sánh hiệu quả của các phương pháp điều trị khác nhau hoặc so sánh hiệu quả khác nhau của cùng một phương pháp điều trị nhưng trên nhiều đối tượng bệnh nhân khác nhau. Bên cạnh đó, việc theo dõi hiệu quả của một hoạt động lâm sàng trong tất cả các lần thực hiện (ví dụ: sử dụng lặp lại một loại thuốc nhất định) sẽ giúp tạo điều kiện thuận lợi cho việc so sánh hiệu quả của hoạt động lâm sàng này đối với tình trạng của bệnh nhân. Ba cách phối màu khác nhau được sử dụng để làm nổi bật các phần thú vị trong quá trình phát triển

một tham số: việc làm nổi bật (tô đậm) khoảng cách giữa các giá trị thực tế với giá trị dự kiến giúp xác định các giá trị quan trọng; việc làm nổi bật (tô đậm) tiến trình của các giá trị thực tế so với các giá trị ban đầu cho thấy kế hoạch điều trị được áp dụng có tác dụng như đạt mức độ nào so với mong đợi; và việc làm nổi bật độ dốc (sự thay đổi đột ngột) của một giá trị sẽ giúp ta khám phá những tác động tức thời của các hoạt động lâm sàng được áp dụng. Một thanh trượt được cung cấp để lựa chọn màu sắc làm nổi bật cho các sự kiện được chọn (xem Hình (10), phía trên cùng) và một cửa sổ cung cấp hình ảnh đã làm mờ thông tin màu bên ngoài đường viền được sử dụng để hỗ trợ quan sát tập trung vào một vùng cần quan tâm cụ thể (ví dụ: phần thời gian nhất định sau khi áp dụng một hành động lâm sàng). Các vùng ít được nhấn mạnh hơn một chút bên ngoài cửa sổ đã nêu trên biểu thị đường cong các giá trị liên tiếp của hoạt động, từ đó cung cấp một số thông tin ngữ cảnh xung quanh cho các giá trị nổi bật mà chúng ta đang quan tâm phân tích.

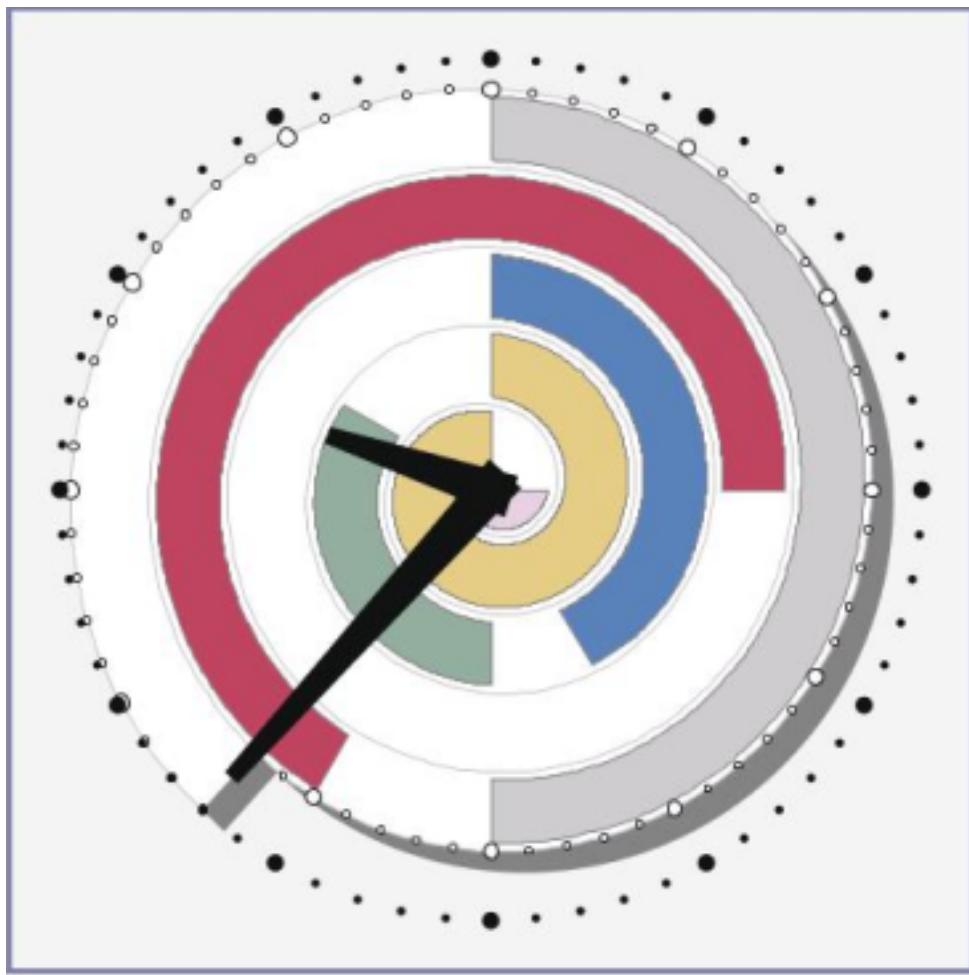


Hình 10: CareCruiser [163]. Chế độ xem theo thời gian (ở phía bên phải) sắp xếp các thông số của bệnh nhân cùng với các hoạt động lâm sàng đã được thực hiện (các chấm nhỏ bên dưới biểu đồ đường) đọc theo trực thời gian nằm ngang. Các chế độ xem theo trật tự móc nối logic và trật tự phân cấp ghi lại các kế hoạch điều trị được hiển thị ở phía bên trái. Khu vực đang được chọn ở cửa sổ bên phải thể hiện giá trị tcSO2 của bệnh nhân giảm chậm sau khi áp dụng một hoạt động lâm sàng cụ thể. (Nguồn: Thực hiện bằng phần mềm nguyên mẫu CareCruiser.)

3.9 Thời gian: Nguyên thủy thời gian – Theo khoảng

Đồng hồ xoắn ốc (SpiraClock) [98] trực quan hóa thời gian bằng cách sử dụng hình ảnh một chiếc đồng hồ. Hình thức thể hiện bao gồm một mặt đồng hồ và hai kim chỉ giờ và chỉ phút. Mặt trong của đồng hồ hiển thị một hình xoắn ốc kéo dài từ viền của đồng hồ về phía tâm của nó. Mỗi chu kỳ của hình xoắn ốc đại diện cho 12 giờ, với thời gian thực được hiển thị ở chu kỳ ngoài cùng và các mốc thời gian tương lai được hiển thị ở gần trung tâm (trong Hình (11) là khoảng 9 tiếng sau trong tương lai). Các khoảng thời gian (ví dụ: các cuộc họp) được thể hiện dưới dạng các phân đoạn được tô màu đọc theo hình xoắn ốc. Các phân đoạn này sẽ thể hiện cho ta biết được thời điểm bắt đầu và kết thúc. Từ đó, người dùng cũng có thể nhanh chóng quan sát được xem liệu các cuộc họp nhất định có bị chồng chéo nhau hay chương trình làm việc có quá dày đặc do nhiều sự kiện

liên tiếp nhau hay không. Khi thời gian trôi qua, vòng xoắn ốc được cập nhật liên tục và các khoảng thời gian trong tương lai dần dần di chuyển ra ngoài cho đến khi chúng trở thành thời gian hiện tại (thời gian thực). Song song với đó, các khoảng thời gian quá khứ sẽ dần dần mờ đi. Với cách thức hoạt động như trên, SpiraClock là một phiên bản nâng cấp của chiếc đồng hồ cổ điển khi nó cho phép người dùng xem trước các khoảng thời gian của tương lai gần và đồng thời cũng có thể nắm bắt khái quát về các khoảng thời gian quá khứ. SpiraClock cũng cho phép người dùng di chuyển kim đồng hồ để truy cập tới các thời điểm khác nhau, bên cạnh đó ta cũng có thể đánh dấu để làm nổi bật các khoảng thời gian mà chúng ta quan tâm và tạo ra các ghi chú dạng văn bản kèm theo tương ứng.



Hình 11: SpiraClock [98] được xây dựng dựa trên ý tưởng của màn hình một chiếc đồng hồ. Trong hình, kim phút hiện đang chỉ vào một cuộc họp đang diễn ra. Các cuộc hẹn trong tương lai được xếp dọc theo hình xoắn ốc trên mặt đồng hồ. (Nguồn: Tác giả tổng hợp)

3.10 Trực quan hóa: Ánh xạ tinh

PostHistory [438] khai phá trực quan các mẫu hình (pattern) hoạt động e-mail khác nhau (ví dụ: các mạng xã hội, tốc độ trao đổi e-mail) và vai trò của thời gian trong các mô hình này (xem Hình (12)). PostHistory lấy người dùng làm trung tâm và tập trung vào các tương tác trực tiếp của một người dùng với những người khác thông qua e-mail. Các

mẫu hình được tạo ra từ việc phân tích thông tin tiêu đề e-mail. Vì vậy, không phải nội dung của thư, mà lưu lượng được lưu vết mới là yếu tố được sử dụng làm cơ sở để phân tích các e-mail trao đổi của mọi người theo thời gian. Về cơ bản, giao diện người dùng trực quan hóa hoạt động trao đổi e-mail trong cả một năm và giao diện này được chia thành hai bảng chính: bảng lịch ở bên trái và bảng danh bạ ở bên phải. Bảng lịch hiển thị cường độ hoạt động e-mail hàng ngày trong đó mỗi ô vuông biểu thị một ngày và mỗi hàng biểu thị một tuần. Kích thước của hình vuông được xác định bởi số lượng e-mail nhận được vào ngày đó và màu sắc của nó thể hiện hướng trung bình tính của thư (tức là thư được nhận qua hình thức TO, CC hay BCC). Màu sắc càng sáng thể hiện các thư nhận được trong ngày hôm đó càng được định hướng rõ ràng. Bảng danh bạ được sử dụng để hiển thị tên của những người đã gửi e-mail cho người dùng.

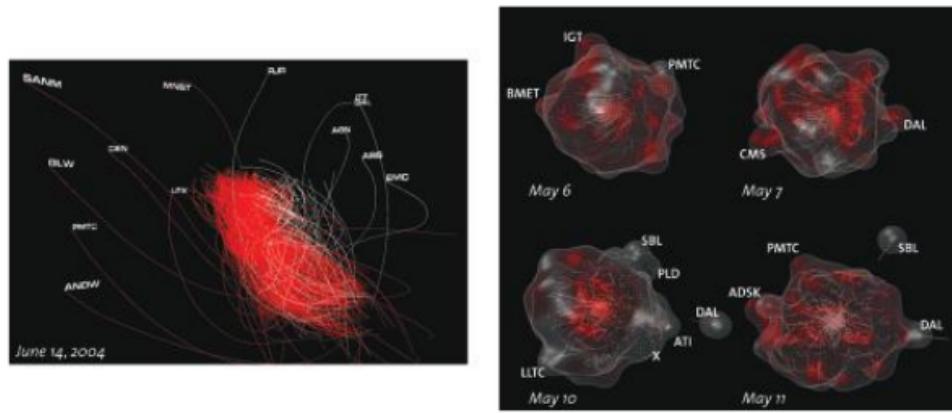


Hình 12: PostHistory [438]. Bảng lịch ở bên trái hiển thị hoạt động e-mail hàng ngày, trong đó số lượng e-mail và hướng trung bình của chúng được lần lượt thể hiện bởi kích thước và màu sắc của các hộp vuông tương ứng. Bảng danh bạ ở bên phải hiển thị tên của những người đã gửi e-mail cho người dùng.

3.11 Trực quan hóa: Ánh xạ động

Dữ liệu thị trường chứng khoán thay đổi linh hoạt trong ngày do giá thị trường liên tục được cập nhật trong phiên giao dịch. Vande Moere [436] đề xuất một cách thức trực quan hóa những dữ liệu như vậy bằng các chùm thông tin (*flocking boids*). Thuật ngữ “boids” mượn từ sự mô phỏng của các loài chim trong đàn (các đối tượng chim = các boid). Để trực quan hóa giá thị trường chứng khoán, mỗi cổ phiếu được coi là một boid với vị trí ban đầu ngẫu nhiên trong không gian ba chiều. Khi có dữ liệu mới, các vị trí boid được cập nhật tự động theo một số quy tắc. Các quy tắc này thường được xây dựng để hướng tới các mục đích như: (i) tránh sự va chạm giữa các đối tượng di chuyển, (ii) đảm bảo

tốc độ di chuyển các đối tượng trong cùng một chùm là giống nhau, (iii) các đối tượng có xu hướng di chuyển về phía trung tâm của đàn, và (iv) đảm bảo các đối tượng có độ tương đồng cao sẽ di chuyển gần nhau trong khi các đối tượng khác nhau sẽ di chuyển xa nhau. Cách thức biểu diễn trực quan hóa này có tính “động” đặc thù giúp thúc đẩy khả năng của người phân tích để nhận biết các xu hướng, mẫu hình (nếu có) khi các chùm di chuyển. Để làm được điều này, các đối tượng (boids) và đường đi của chúng phải được hiển thị dưới dạng các đường cong sinh động, như thể hiện ở Hình (13) (hình bên trái). Ta cũng có thể nâng cấp cách biểu diễn trực quan 3D này bằng cách bao đóng các chùm bên trong các bề mặt ẩn, giúp người dùng có thể dễ dàng hơn trong việc nhận diện cấu trúc không gian của chùm (xem Hình (13), bên phải). Các chùm có thể sẽ rất hữu ích trong việc khai phá, tìm kiếm các xu hướng, mẫu hình khác nhau trong bộ dữ liệu, chẳng hạn như sự tồn tại của các chùm dữ liệu, sự tách biệt của các chùm con khỏi chùm chính hoặc hành vi khác nhau của các đối tượng trong cùng một chùm.

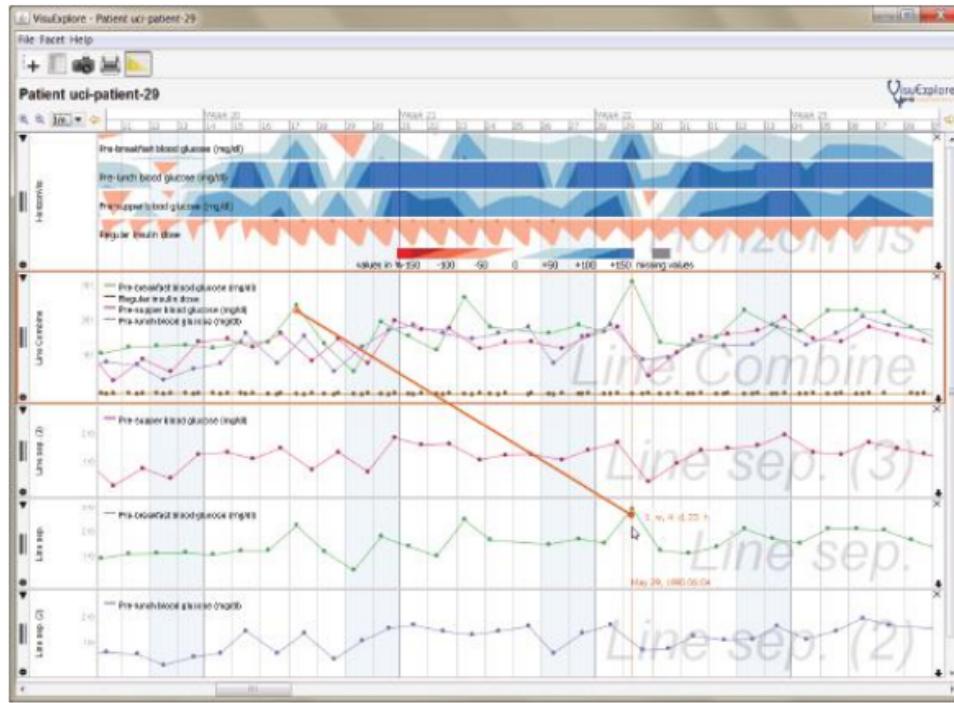


Hình 13: Chùm dữ liệu [436]. Dữ liệu thị trường chứng khoán được thể hiện dưới dạng các chùm di chuyển trong không gian ba chiều. Bên trái: các đối tượng tách khỏi chùm cho thấy giá cổ phiếu tương ứng có xu hướng vận động khác với phần lớn giá của các cổ phiếu khác; bên phải: các bề mặt ẩn bao quanh các khối giúp người dùng nhận biết cấu trúc không gian của chùm.

3.12 Trực quan hóa: Số chiều – 2 chiều (2D)

VisuExplore [338] là một hệ thống tương tác trực quan phục vụ cho việc khai phá chuỗi các dữ liệu, thông số y tế không đồng nhất theo thời gian. Hệ thống này sử dụng nhiều chế độ xem khác nhau nằm song song với nhau theo chiều dọc với chung một trục hoành là trục thời gian, từ đó giúp cho việc tham chiếu các thông số y tế có liên quan với nhau. VisuExplore cung cấp một môi trường mở rộng chứa các công cụ, kỹ thuật hiển thị trực quan với tôn chỉ đơn giản, dễ dàng sử dụng trong thực hành y tế như: biểu đồ đường, biểu đồ dòng thời gian (timeline), biểu đồ cột, biểu đồ miền [331], biểu đồ sự kiện, biểu đồ đường với chú thích có thể thu phóng (xem Hình (14), trên cùng). Ngoài ra, dữ liệu cũng có thể được trình bày dưới dạng các bảng chứa văn bản để đa dạng các cách thức trực quan hóa. Các tính năng tương tác của VisuExplore cho phép các bác sĩ có được cái nhìn tổng quan về nhiều thông số y tế và tập trung vào các phần của dữ liệu. Có thể thêm hình ảnh hiển thị một hoặc nhiều biến bổ sung. Người dùng có thể thêm, xóa, thay đổi kích thước và sắp xếp lại các chế độ xem trực quan hóa vì VisuExplore cũng cung cấp khả năng tương tác phong phú để thêm, xóa, sắp xếp và điều chỉnh chế độ xem. Ngoài ra, một công cụ đo lường cũng được tích hợp để có thể xác định khoảng thời gian giữa

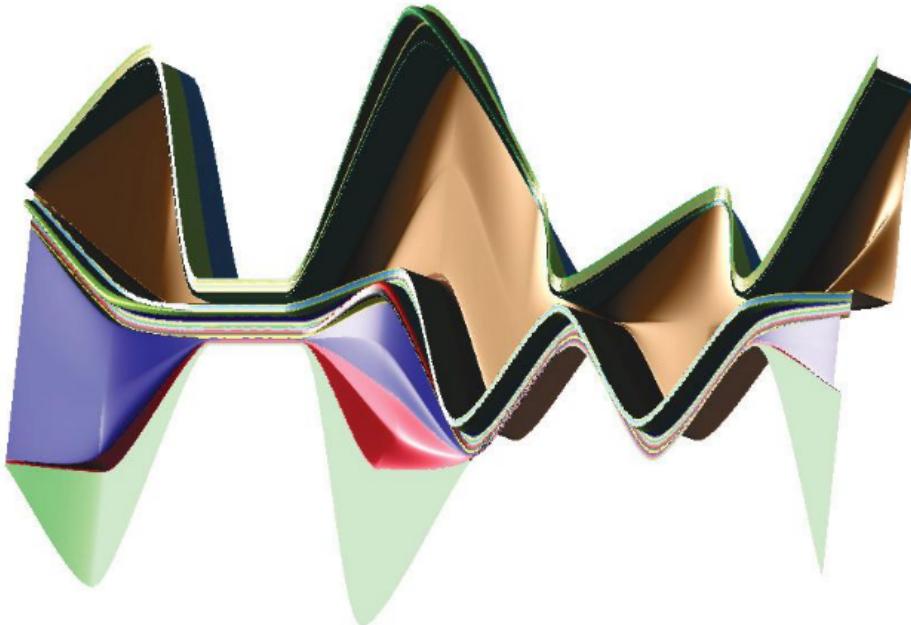
các hai thời điểm mà người dùng quan tâm và lựa chọn; chức năng này không hoạt động trong phạm vi của một chế độ xem cụ thể mà còn có thể hoạt động xuyên suốt trên các chế độ xem khác nhau.



Hình 14: VisuExplore [338]. Các phương pháp biểu diễn trực quan, đơn giản và dễ hiểu bằng cách sử dụng biểu đồ đường, biểu đồ cột, biểu đồ sự kiện, biểu đồ dòng thời gian và biểu đồ đường kèm chú thích. Trong đó, công cụ đo lường được minh họa: Một đường nối minh họa cho khoảng thời gian giữa hai mục được chọn (đường nối này hoạt động trên các khung khác nhau). (Được tạo bằng phần mềm nguyên mẫu VisuExplore.)

3.13 Trực quan hóa: Số chiều – 3 chiều (3D)

3D ThemeRiver [197] (xem Hình (15)) là một biến thể 3D của kỹ thuật ThemeRiver [172] (xem Hình 7.16). Phương pháp 3D kế thừa thiết kế hình ảnh cơ bản từ phiên bản 2D của nó: các dữ liệu chuỗi thời gian được thể hiện bằng độ rộng của các dòng có màu sắc khác nhau tạo thành một dòng chảy xuyên suốt theo trục thời gian nằm ngang. Trong biến thể 2D, chỉ có thể biểu diễn một biến dữ liệu trên mỗi dòng, cụ thể là bằng cách thay đổi độ rộng của dòng đó. Giải pháp cải tiến của Imrich cùng công sự và các công sự sẽ giúp giải quyết hạn chế này. Việc mở rộng thiết kế thêm một chiều thứ ba giúp chúng ta có thể thể hiện thêm được nhiều thông tin trực quan bổ sung: chiều cao (ở dạng 3D) của mỗi dòng có thể được thay đổi để thể hiện thêm những thông tin cần thiết. Thiết kế này đặc biệt phù hợp để hình dung các xu hướng đồng biến bậc ba trong dữ liệu. Imrich và các công sự [197] đã tiến hành khảo sát người dùng để đánh giá tính hữu ích của việc biểu diễn 3D này và thực sự đã thu được kết quả tích cực cho thấy biến thể 3D có nhiều ưu điểm hơn so với biến thể 2D. Cụ thể, sự săn sóc của các công cụ tương tác điều hướng 3D thích hợp được nhấn mạnh là một yếu tố quan trọng góp phần vào sự thành công của 3D ThemeRiver.



Hình 15: 3D ThemeRiver [197]. Các dòng có màu riêng biệt tạo thành hình dạng tổng thể của 3D ThemeRiver. Ngoài ra, chiều rộng và chiều cao của các dòng được thay đổi để thể hiện giá trị thay đổi của dữ liệu theo thời gian. Trong hình này, chiều rộng biểu diễn sự phân bố tổng thể của 17 cụm dữ liệu aerosol và chiều cao biểu thị tỷ lệ kẽm. (Được sử dụng với sự cho phép của các tác giả.)

Bảng 16 cung cấp một cách tổng quan về tất cả các kỹ thuật có trong chương này cùng với các thông tin phân loại chúng. Ta cũng có thể sử dụng bảng này để tìm kiếm các kỹ thuật trực quan hóa đáp ứng các tiêu chí nhất định.

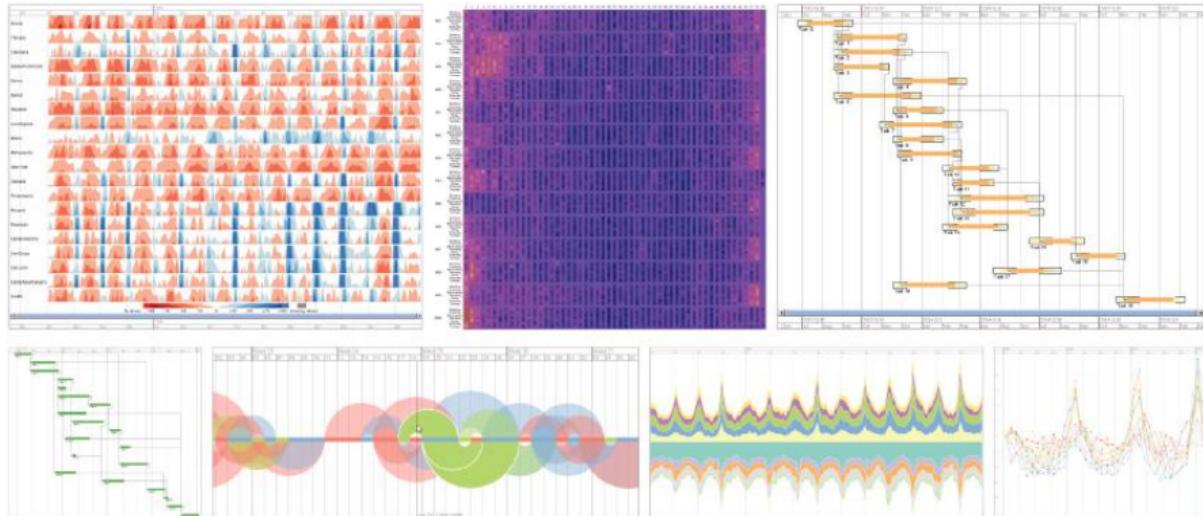
	data	time	vis
	abstract spatial	frame of reference	
	univariate variables	multivariate arrangement	
		linear cyclic	
		instant interval	time primitives
KronoMiner	■	■	■
Great Wall of Space-Time	■	■	■
GROOVE	■	■	■
TimeWheel	■	■	■
Line Plot	■	■	■
Enhanced Interactive Spiral	■	■	■
CareCruiser	■	■	■
SpiraClock	■	■	■
PostHistory	■	■	■
Flocking Boids	■	■	■
VisuExplore	■	■	■
3D ThemeRiver	■	■	■

Hình 16: Tổng quan và phân loại các kỹ thuật trực quan hóa

4 TimeBench: Một mô hình dữ liệu và thư viện hỗ trợ cho việc phân tích trực quan hóa dữ liệu chuỗi thời gian

Dữ liệu chuỗi thời gian đóng một vai trò quan trọng trong nhiều trường hợp cần phân tích hình ảnh và trực quan hóa, chẳng hạn như khi tìm kiếm những thông tin giá trị từ bộ dữ liệu hồ sơ sức khỏe điện tử hoặc xác định các vấn đề mới phát sinh và các kết nối dễ bị hư hại trong mạng truyền thông. Tuy nhiên, nhiều thư viện phần mềm hỗ trợ phân tích trực quan hiện nay chỉ coi thời gian là một loại dữ liệu số phẳng và vì thế đã giải quyết không triệt để sự phức tạp của miền thời gian, chẳng hạn như độ chi tiết của biểu lịch và các khoảng thời gian. Do đó, các nhà phát triển công cụ, kỹ thuật trực quan hóa nâng cao phải triển khai lặp đi lặp lại các nền tảng liên quan đến thời gian trong bộ mã của họ. TimeBench [339] là một phần mềm mã nguồn mở và miễn phí cung cấp các cấu trúc dữ liệu nền tảng và thuật toán phục vụ cho việc phân tích và trực quan hóa dữ liệu chuỗi thời gian (<http://timebench.org>). Khả năng biểu đạt thông tin và khả năng dễ dàng tiếp

cân của nó đã được đánh giá thông qua nhiều ứng dụng cụ thể khi giải quyết các thách thức liên quan đến dữ liệu chuỗi thời gian và các nghiên cứu phát triển dài hạn được ở cả cấp độ các chương trình nghiên cứu lẫn các dự án của sinh viên. Hình (17) thể hiện 7 ứng dụng khác nhau được phát triển với thư viện TimeBench.



Hình 17: TimeBench [339]. Ví dụ về các ứng dụng được xây dựng dựa trên TimeBench: (1) Dữ liệu sức khỏe hàng tháng của 20 thành phố trong 14 năm trong biểu đồ miền [331]; (2) Dữ liệu sức khỏe hàng ngày trong khoảng thời gian 14 năm được thể hiện trong một biểu đồ GROOVE [262]; (3) Bản kế hoạch dự án sử dụng PlanningLines [4]; (4) Bản kế hoạch dự án dưới dạng biểu đồ Gantt [170]; (5) Sơ đồ vòng cung [451] thể hiện mối quan hệ giữa ba loại sự kiện; (6) Trực quan hóa ThemeRiver [172]; (7) Các biểu đồ đường đi kèm chỉ mục [34]. Dữ liệu sức khỏe từ nghiên cứu NMMAPS [316] được sử dụng trong (1), (2), (5), (6) và (7). (<http://timebench.org>).

5 Tổng kết

Trong chương này, chúng ta đã khám phá các khái niệm và phương pháp để trực quan hóa dữ liệu chuỗi thời gian. Chương này cũng đã trình bày cách sử dụng các đặc trưng của dữ liệu chuỗi thời gian trong các kỹ thuật trực quan hóa khác nhau. Các đặc trưng này bao gồm dữ liệu, thời gian và hình ảnh biểu diễn trực quan (xem Phần 3.1). Vì có thể quan sát được những điểm cân xứng và bất cân xứng liên quan đến cách phân loại của chúng tôi, do đó, độc giả nên dành thời gian để xem xét kỹ lưỡng những phần diễn giải cụ thể ở các nội dung đã trình bày.

- *Dữ liệu: hệ quy chiếu.* Trong khảo sát này [5], các tác giả chủ yếu tập trung vào loại dữ liệu trừu tượng. Việc biểu diễn dữ liệu chuỗi thời gian trong khung tham chiếu không gian đòi hỏi nhiều hơn đáng kể trong hoạt động thiết kế vì phải đóng gói nhiều thông tin hơn vào trong cùng một biểu đồ trực quan. Đặc biệt là khi đã có các ngành khoa học liên quan đến bản đồ học, vốn là những lĩnh vực nghiên cứu độc lập có lịch sử hình thành và phát triển lâu dài, đã phát triển các phương pháp tiếp cận để kết hợp trực quan hóa các khía cạnh thời gian và không gian của dữ liệu [16, 17, 256]. Hơn nữa, dữ liệu không gian nói chung và không gian địa lý nói riêng cũng đã được trình bày trong Chương 5 và 6.

- *Thời gian: sắp xếp.* Hầu hết các kỹ thuật trong khảo sát của chúng tôi [5] tập trung vào việc biểu diễn thời gian theo trình tự tuyến tính; các cách tiếp cận với thời gian theo trình tự tuần hoàn được đề cập ít hơn tương đối. Lý do cho điều này có thể là do người dùng thường quan tâm đến các xu hướng phát triển từ quá khứ, đến hiện tại, đến tương lai hơn là tìm kiếm các chu kỳ tuần hoàn trong dữ liệu.
- *Thời gian: nguyên thủy thời gian.* Thời điểm tức thời là nguyên thủy thời gian được sử dụng phổ biến nhất trong khảo sát của chúng tôi [5]. Điều này có vẻ hoàn toàn tự nhiên bởi vì dữ liệu thường được đo lường tại một thời điểm cụ thể. Việc đo lường theo khoảng thời gian thường ít xảy ra hơn và chủ yếu là trong các trường hợp cần đến việc lập kế hoạch.
- *Trực quan hóa: ánh xạ.* Rõ ràng, các kỹ thuật ánh xạ tĩnh thì thường phù hợp hơn để biểu diễn trên một trang sách thay vì ánh xạ động. Vì thế, các kỹ thuật trực quan hóa được chọn ở chương này có chút thiên lệch khi tập trung nhiều vào các phương pháp ánh xạ tĩnh. Tuy nhiên, các kỹ thuật ánh xạ động cũng vô cùng quan trọng và nó thường là giải pháp đầu tiên được đưa ra khi trực quan hóa dữ liệu chuỗi thời gian.
- *Trực quan hóa: số chiều.* Các phương thức biểu diễn trực quan 2D thường được ưa chuộng hơn so với 3D, vì chúng thường khái quát hơn và do đó dễ hiểu hơn. Và vì thế, một lần nữa các kỹ thuật trực quan hóa đã được giới thiệu trong chương này chủ yếu thiên về phương pháp tiếp cận 2D. Đặc biệt là các kỹ thuật sơ khai có xu hướng gắn liền với không gian hai chiều đơn giản do khả năng tính toán hạn chế đó. Tuy nhiên, các công nghệ hiện đại đã giúp các nhà thiết kế trực quan hóa dễ dàng triển khai trực quan hóa 3D hơn và giúp điều hướng góc nhìn của người dùng khi khám phá các không gian thực tế ảo ba chiều. Điều này đặc biệt hữu ích khi trực quan hóa dữ liệu có tham chiếu không gian.
- *Trình duyệt TimeViz.* Chúng tôi đã xây dựng một kho lưu trữ (<http://survey.timeviz.net>) giúp tìm kiếm các phương pháp có sẵn phù hợp với cách thức phân loại đã được giải thích trong Phần 3.1 (xem Hình (18)). Việc tìm kiếm được thực hiện đơn giản thông qua việc lựa chọn các đặc trưng về thời gian, dữ liệu và biểu diễn trực quan, từ đó các kỹ thuật trực quan hóa phù hợp sẽ được đề xuất dựa trên các đặc trưng đã chọn. Việc cấu trúc kho của chúng tôi trên cơ sở từng trang (per-page) cho phép truy cập dễ dàng khi cần tham khảo nhanh một kỹ thuật cụ thể. Mỗi trang mô tả ngắn gọn bối cảnh, giải thích ý tưởng và các khái niệm chính, đồng thời chỉ ra ứng dụng của một kỹ thuật cụ thể. Phần mô tả chi tiết cũng giới thiệu tham chiếu đến các tài liệu gốc hoặc danh sách các tài liệu tham khảo trong trường hợp có nhiều án phẩm cần tham chiếu hoặc nhiều án phẩm khác nhau sử dụng cùng một phương pháp. Có thể nói Trình duyệt TimeViz là một công cụ tuyệt vời để tìm kiếm các kỹ thuật trực quan phù hợp theo ba khía cạnh đã nêu trên.



Hình 18: Trình duyệt TimeViz. Kho lưu trữ các kỹ thuật trực quan hóa cho dữ liệu chuỗi thời gian (<http://survey.timeviz.net>) [5].