

Tổng kết

Xin chào mọi người, cảm ơn Thắng & Nam vì sự trình bày của hai bạn. Sau đây tôi xin tóm tắt lại ý tưởng chính của chương sách chúng tôi trình bày. Ngắn gọn mà nói, chương sách đề cập hai nội dung chính là quy trình và các lỗi hay gặp.

Quy trình bao gồm từ việc: - Xác định một ánh xạ dữ liệu phù hợp để chuyển đổi từ dữ liệu thô trở thành hình ảnh trực quan. - Điều chỉnh chế độ xem, mật độ thông tin, chú thích, màu sắc và cuối cùng là tính thẩm mỹ tổng thể của hình ảnh trực quan.

Mặc dù việc tuân thủ quy trình này cơ bản sẽ giải quyết hầu hết vấn đề, tuy nhiên vẫn tồn tại những sai lầm như là: - Hình ảnh trực quan vô tình hoặc cố ý làm sai lệch về mặt tỷ lệ hoặc các chiều trò hình ảnh khác. - Phép trực quan hiển thị ra là vô nghĩa. - Lỗi khác liên quan nhiều hơn đến dữ liệu, như là lấy lại mẫu, nội suy, loại bỏ dữ liệu khỏi phép trực quan mà không cho phép truy cập vào dữ liệu thô gốc ban đầu.

Về cơ bản, đó là tóm lược những điều hay ho. Vậy là hết rồi. Cảm ơn các bạn đã lắng nghe.

À khoan, kết thúc như này thực sự không hay chút nào. Tôi xin nói thêm một chút nữa. Sự thực, chương sách có hai phần chính và Thắng & Nam đã tranh nhau nói mất hai cái hay ho đó mất rồi. Như vậy nếu không còn cái để nói thì như nhìn hình Steve Jobs chúng ta thấy rằng: Tôi cần tạo ra cái để nói. Giống như không có vấn đề thì mình sẽ tạo ra thêm vấn đề cho nó phức tạp hóa.

Hai vấn đề tôi sẽ đề cập chính là: 1. Thử áp dụng kiến thức về quy trình tạo trực quan hóa và các lỗi sai hay gặp vào bài toán thực tế một cách liền mạch. Rõ ràng nếu chúng tôi đi xuống ngay lúc này thì sẽ có người nhận xét là chúng tôi nói về quy trình nhưng chẳng thấy quy trình đó được dùng vào một bài toán cụ thể như thế nào. Sự thực, đây là một chương tổng hợp và tác giả muốn đề cập vào rất nhiều khía cạnh của trực quan hóa. Điều mà không một bài toán cụ thể nào có thể bao trùm hết ý nghĩa cho những kỹ thuật được áp dụng. Bởi vậy, các bạn có thể thấy các ví dụ trước đó khá rời rạc và không liên quan nhiều đến nhau. Trong vấn đề đầu tiên tôi đề cập đến ở đây chính là ứng dụng lý thuyết đó vào hai bài toán: - Một cái đơn giản (đánh giá tiềm năng mua cổ phiếu Apple với dữ liệu cổ phiếu 5 chiều theo thời gian) để cho mọi người thấy quy trình từng bước kia. - Bài toán còn lại sẽ phức tạp hơn (dữ liệu 39 chiều về các yếu tố kinh tế vĩ mô của nền kinh tế, mục đích là tìm ra các mối liên hệ và sự giải thích phù hợp).

1. Vấn đề thứ hai là cái tôi quyết định thêm vào sau khi nghe qua đầy đủ 19 nhóm trình bày trước đó. Ngay buổi đầu nhận sách ra đọc, tôi đã vô cùng ấn tượng với chương sách 13 này. Nó có tính tổng quát hóa cao cho cả môn học - có chữ quy trình đã show rõ ra rồi. Giống như câu nói tôi tâm đắc: “Nhìn cây thì phải thấy rừng, học cái cụ thể phải biết khái quát hóa thành quy luật lớn hơn.” Và phần tổng kết chương sách này sẽ còn điều gì tuyệt vời hơn nếu như tôi đưa ra lời bình luận, ý kiến cảm quan tôi đúc rút ra được về chương sách này nói riêng và cả quyển sách hay những nội dung các nhóm đã nói trước đó nói chung. Sau mỗi lần học hỏi được điều gì mới chúng ta đều cần quá trình gọi là review, xem lại để đúc rút ra. Tôi rất vui khi đứng đây để nói ra ba điều quan trọng nhất tôi đúc rút ra được và tôi xem là tinh hoa cho học phần Trực quan hóa dữ liệu này.

Hãy bắt đầu đến với phần đầu tiên, cách ứng dụng lý thuyết chương sách này vào bài toán cụ thể. Bài toán đầu tiên, data có là 5 chiều thông tin về một cổ phiếu theo thời gian. Mục đích tôi trực quan hóa ở đây là thuyết phục các bạn mua cổ phiếu Apple. Hy vọng sau phần trình bày của tôi, bạn nào ghét Apple thì sẽ đi mua cổ phiếu Apple :) Dữ liệu 5 chiều bao gồm: mở cửa, đóng cửa, thấp nhất, cao nhất và khối lượng giao dịch.

- Bước 1: Chúng ta cần lựa chọn một ánh xạ dữ liệu phù hợp. Do chứng khoán là cái nhiều người quan tâm nên dữ liệu dạng này được nghiên cứu kỹ và ánh xạ dữ liệu phù hợp là biểu đồ nến. Biểu đồ nến có thể hiện được 4 chiều thông tin: mở cửa, đóng cửa, thấp nhất, cao nhất của một cổ phiếu. Kết hợp thêm một chiều nữa là khối lượng giao dịch, như vậy toàn bộ thông tin về 5 chiều được hiển thị trên một hình ảnh trực quan. Quả thực là một ý tưởng thông minh và hiệu quả. Người nghĩ ra cái này xứng đáng được tràng pháo tay. Mọi người vỗ tay đi ạ. Vâng, hy vọng tiếng vỗ tay giúp các bạn tỉnh ngủ và chuẩn bị xuống tiền mua cổ phiếu Apple.
- Bước 2: Điều chỉnh chế độ xem và các bước còn lại. Chúng ta cần cung cấp nhiều chế độ xem cho người dùng. Có thể chọn xem theo ngày, theo tuần, theo tháng hoặc theo năm; cho phép zoom in, zoom out để xem rõ hơn. Một số người có thể gặp khó khăn trong việc nhìn nhận một số màu sắc, hoặc do sở thích cá nhân. Bởi vậy, ở đây hình ảnh trực quan cho phép lựa chọn thay đổi nhãn màu sắc. Một khía cạnh khác là cung cấp một góc nhìn nhận thông tin khác đi. Ví dụ nhìn mốc tăng gần như hàm mũ thế kia thì ai mà chẳng muốn mua phải không ạ? Tuy nhiên, nếu muốn thấy rõ sự tăng giảm trong một khoảng thời gian cụ thể, việc lấy sai phân theo hàm log có thể giúp chúng ta nhìn nhận rõ hơn. Mọi người có thấy chuỗi thời gian thu được này bớt tăng nóng quá không? Điều này giúp chúng ta nhìn nhận rõ hơn về sự tăng giảm của cổ phiếu. Đây là ví dụ về việc cung cấp nhiều chế độ xem để đa dạng góc nhìn của người xem.
- Các bước tiếp theo chính là: lựa chọn mật độ thông tin phù hợp. Do dạng thuyết trình nên tôi chỉ cần làm nổi bật sự tăng trưởng của mã cổ phiếu theo thời gian. Không cần thêm nhiều ký hiệu màu mè hơn, làm lệch khỏi mục đích ban đầu. Với tùy chọn màu sắc, chú giải, thẩm mỹ... của từng người sẽ giúp tạo ra nhiều phiên bản hình ảnh trực quan khác nhau tùy vào sở thích của người xem. Giờ đây lựa chọn mua cổ phiếu Apple là của các bạn, tôi chỉ show cái hình ra thế thôi. Ai có nhu cầu có thể liên hệ tôi. Công ty Harito chúng tôi cung cấp dịch vụ phân tích các danh mục đầu tư đa dạng. Cuối kỳ này chúng tôi sẽ public nghiên cứu giúp các bạn xây dựng danh mục tiết kiệm, đầu tư theo từng giai đoạn cuộc đời cho phù hợp. Rất mong cuối kỳ các bạn đến nghe và bắt đầu tiết kiệm đi khi theo như thống kê của Đại học Thương Mại thì 50% sinh viên hiện tại chẳng biết tiết kiệm là cái gì.

Nhìn chung với dữ liệu đơn giản 5 chiều như kia, công việc trực quan tương đối dễ dàng và ít khả năng mắc phải các lỗi nghiêm trọng. Tuy nhiên, với dữ liệu 39 chiều về các yếu tố kinh tế vĩ mô của nền kinh tế, mục đích là tìm ra các mối liên hệ và sự giải thích phù hợp thì sẽ phức tạp hơn. Đầu tiên để tôi giới thiệu qua về cấu trúc dữ liệu. Ở đây tôi show ra cái heatmap đánh giá khả năng tương quan tuyến tính giữa các yếu tố kinh tế vĩ mô. Mục đích của tôi ở đây là show ra cho các bạn thấy được các mối liên hệ giữa các yếu tố kinh tế vĩ mô, để giúp các bạn hiểu hơn về nền kinh tế.

Nói về cái các tương quan trong heatmap, ...

Được rồi, các bạn làm rõ cấu trúc dữ liệu rồi chứ? Mục đích của tôi (xin nhấn mạnh lại) là phân tích sự tương quan của các yếu tố kinh tế vĩ mô. Tôi sẽ show ngẫu nhiên vài cái hình vẽ ra và gọi ngẫu nhiên vài bạn điểm danh nói xem hình tôi vẽ mắc lỗi gì nhé.

1. Hình 1: ... Tránh tạo các trực quan vô nghĩa. <Đây là hình ảnh histogram chứng minh rằng hầu hết thời gian trong lịch sử, VN có GDP thấp và chứng tỏ sự yếu kém của nước nhà.>
2. Hình 2: ... Ẩn giấu, lược bớt đi thông tin xu hướng thời gian trong tỷ trọng các ngành trong nền kinh tế. Hình ảnh bị lợi dụng để làm người xem hiểu sai (thủ thuật tinh vi để lừa những người không biết phân tích và phản biện hình trực quan mình nhìn được). Chi tiết hơn tôi sẽ đề cập ở phần phía sau (phần lời bình nhận xét với tư cách người tổng kết, đưa ra kết luận về quá trình chúng ta đọc các quyển sách qua). <Đây là hình ảnh tỷ lệ các ngành nghề, các thành phần khác kia tôi không rõ nó là gì đâu, tại lúc crawl dữ liệu về thì số phần trăm cộng lại không được là 100%. Có thể do sai lệch trong thu thập. Tuy nhiên, do chưa xác định được lý do nên theo sách này tôi đọc, tôi không nên loại bỏ khoảng trống đó đi, ví dụ áp dụng hàm

softmax ở đây, kiểu vậy.>

3. Hình 3: ... Áo tưởng sự tương quan. Rõ ràng là do dữ liệu thu thập chưa đủ lâu dài để đưa ra kết luận. Có nhiều yếu tố khác ảnh hưởng đến GDP và dân số như chính sách, môi trường, nguồn lực, ... Những điều này rõ ràng cần phải được xem xét kỹ lưỡng hơn trước khi tôi đưa ra kết luận khẳng định như vậy, có đúng không?
4. Hình 4: ... Điều chỉnh lượng thông tin phù hợp cho dữ liệu lớn. Tuy nhiên, tôi cũng cần nhấn mạnh một ý mà chương sách này có nói và tôi thấy rất hay, mặc dù tôi không thấy Thắng đề cập. Đó là phần bước: "Lựa chọn ánh xạ dữ liệu phù hợp".
5. Rõ ràng rằng việc chọn những ánh xạ dữ liệu nhìn trực quan thường thấy rất tốt. Tuy nhiên, đôi khi chọn các phép ánh xạ mà số đông người khác coi là không tốt hóa ra lại tốt.

Câu chuyện về người thợ rèn: Tóm tắt câu chuyện: - Tìm kiếm: Một thợ rèn đã cố gắng tìm cách làm cho các sản phẩm của mình cứng hơn để đạt được chất lượng tốt hơn. Ông đã thử nghiệm nhiều phương pháp nhưng không thành công. - Nghỉ ngơi: Sau nhiều giờ làm việc vất vả và không đạt được kết quả như mong muốn, ông quyết định nghỉ ngơi và rời khỏi xưởng rèn. - Phát hiện: Khi trở về, ông nhận thấy rằng những món đồ mà ông đã rèn và để nguội một cách tự nhiên đã trở nên cứng hơn rất nhiều so với khi ông rèn chúng ngay lập tức. - Kết luận: Qua trải nghiệm này, ông hiểu rằng việc để thép nguội từ từ trước khi xử lý tiếp có thể làm cho nó cứng hơn. Đây là một trong những cách mà thợ rèn đã phát hiện ra nguyên tắc quan trọng trong quá trình rèn thép.

Ý nghĩa: Câu chuyện này không chỉ nói về việc tìm kiếm sự hoàn thiện trong nghề rèn mà còn mang một thông điệp sâu sắc về việc học hỏi từ những thất bại và sự quan sát. Nó nhấn mạnh tầm quan trọng của việc nghỉ ngơi và cho phép bản thân có thời gian để suy nghĩ, dẫn đến những phát hiện và cải tiến mới. Còn trong trường hợp quyển sách này đề cập thì đôi khi thay đổi ánh xạ, một hướng nhìn, một hướng tiếp cận lạ lùng không thường thấy có thể giúp bạn phát hiện ra điều mới tuyệt vời.

Vậy là phần trình bày của tôi cũng tương đối dài rồi nhỉ? Tôi vẫn còn muốn nói 3 ý tổng kết mà tôi đúc rút được sau khi học học phần này được nửa kỳ. Sau ngày hôm nay, có thể hầu hết các bạn ở đây sẽ quên sách những gì nhóm tôi đã trình bày, và cả các nhóm trước đây nữa. Nhưng tôi tin rằng việc biết 3 ý này của tôi sẽ giúp các bạn trong tương lai.

Hoặc có khi cô Thủy có thể có 1 bài tập thu hoạch cho cả lớp: Hãy viết ra 3 điều tâm đắc nhất đúc rút được sau khi nghe qua các nhóm đã trình bày: từ các công cụ, các kỹ thuật và cả là quy trình. Có thể tỷ có bạn ra bảo sao tạo công ăn việc làm cho anh em mk thế. Tôi biết thêm 1 chút việc cũng có chút mệt. Nhưng thử 1 chút thời gian để suy ngẫm về 3 điều mà bản thân thấy là quan trọng nhất cũng hay mà. Tôi tin rằng 3 điều mà các bạn tự rút ra cho mình thì chắc chắn lúc về hưu các bạn vẫn nhớ đó.

Điều cuối cùng, tôi biết có người thấy chán, có người sắp quên. Nhưng các bạn quên của nhóm khác thì được chứ tuyệt đối đừng quên nhóm của tôi. Nhóm chúng tôi là nhóm nói về quy trình, các bước làm tổng quát. Hay nói cách khác, nhóm tôi đã show ra cả khu rừng từ từng cây nhỏ của các nhóm khác. Xin cảm ơn các bạn đã lắng nghe tôi. Chúc các bạn có một buổi học vui vẻ và hiệu quả.