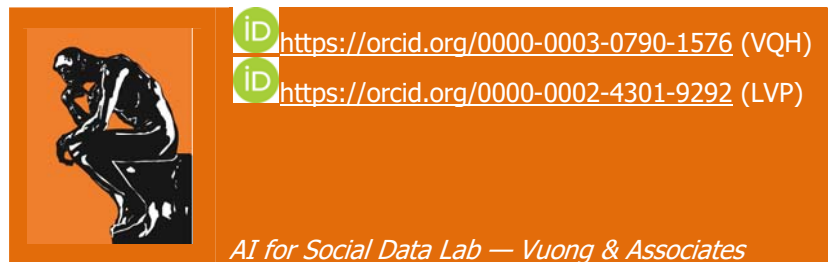


Tiến triển của *bayesvl* kể từ sau khi xuất bản trên CRAN

Vương Quân Hoàng
Lã Việt Phương

AISDL



Hà Nội, Việt Nam
Ngày 29 tháng 2 năm 2020

Vài nét sơ lược

Chương trình máy tính *bayesvl* chạy trên môi trường ngôn ngữ tính toán thống kê R được chính thức xuất bản trên CRAN ngày 24-5-2019, phiên bản v.0.8.5, tại địa chỉ: <https://cran.r-project.org/web/packages/bayesvl/index.html> [1]. Các tiến triển về chương trình cũng như ứng dụng được cập nhật mở trên kho mã nguồn mở Github: <https://github.com/sshpa/bayesvl>, hiện tại đang được đánh số phiên bản v.0.9 [2].

Sau khi vượt qua kiểm tra kỹ thuật của R Core Team, *bayesvl* được sản xuất thành các phiên bản mã máy, kèm theo tài liệu của R Documentation, cung cấp chính thức tại địa chỉ: <https://www.rdocumentation.org/packages/bayesvl/versions/0.8.5/topics/bayesvl%20stan%20utilities>. Ngoài Github [2] như đã nói, chương trình có thể dễ dàng tải trực tiếp từ nhiều máy chủ khu vực, đặt tại các trường đại học và trung tâm nghiên cứu liên kết với mạng lưới tài liệu kỹ thuật của R. Vài ví dụ về các nơi có thể tải tài liệu chính thức của *bayesvl* dưới đây (đã kiểm tra truy cập ngày 29/2/2020):

- <https://mran.microsoft.com/package/bayesvl>
- <https://cran.csiro.au/web/packages/bayesvl/bayesvl.pdf>
- <https://rdr.io/cran/bayesvl/>
- <https://www.r-pkg.org/pkg/bayesvl>
- <https://www.crantastic.org/packages/bayesvl>
- <http://packages.renjin.org/package/org.renjin.cran/bayesvl>
- <ftp://stat.ethz.ch/CRAN/web/packages/bayesvl/index.html>
- https://ftp.uni-sofia.bg/CRAN/web/checks/check_results_bayesvl.html

- http://espejos.ucr.ac.cr/CRAN/web/checks/check_results_bayesvl.html
- http://repo.miserver.it.umich.edu/cran/web/checks/check_results_bayesvl.html
- https://espejito.fder.edu.uy/cran/web/checks/check_results_bayesvl.html

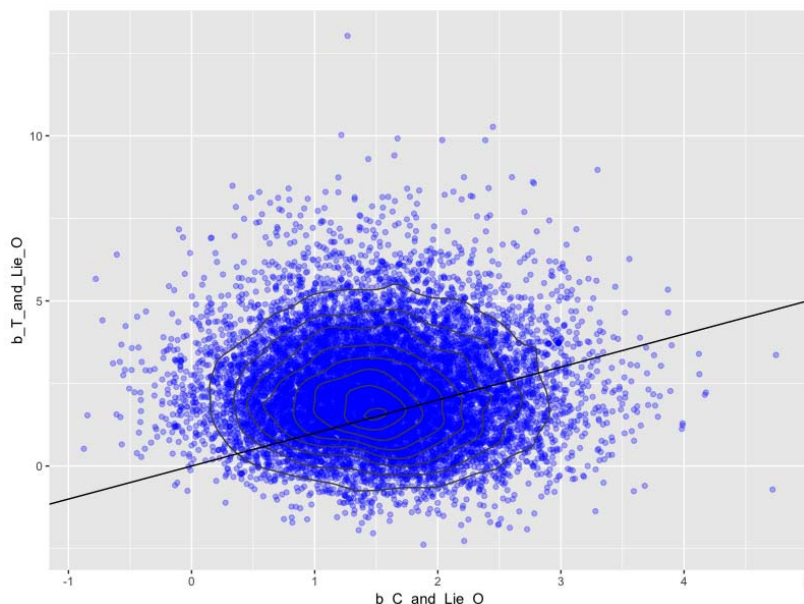
Bên cạnh đó, chương trình cũng được giới thiệu một số nơi như báo *Khoa học & Phát triển* [3], trang web các cơ quan KH-CN, và trường đại học (trường ĐH Phenikaa, ĐH Ngoại thương,...).

Trên thực tế, *bayesvl* phần nào đã trả lời câu hỏi trước đây nhóm nghiên cứu nêu trên *European Science Editing* năm 2019 [4], tuy nhiên cần thêm nhiều ứng dụng để có thể kiểm tra tốt hơn khả năng hỗ trợ công việc nghiên cứu, giảng dạy, cũng như tìm kiếm hướng để cải thiện tính năng kỹ thuật và thực hành cho người dùng.

Một số ứng dụng trong công việc

Trước tiên, *bayesvl* hỗ trợ được quá trình đào tạo của AISDL cho các cán bộ nghiên cứu trong việc tiếp cận triết lý, logic và cấu trúc dữ liệu, và cách đánh giá các kết quả thống kê Bayesian. Nói như thế không có nghĩa là chỉ có *bayesvl* làm được việc này, mà về tính năng sư phạm thì những chương trình như *Rethinking* của Richard McElreath đã và đang làm rất tốt cả với phần sách và mã máy tính [5]. Tuy nhiên, tiếp cận qua lưới Bayesian, và kiểm tra bằng MCMC thì thực sự là mục tiêu của *bayesvl*, và hiện tại vẫn là ý tưởng riêng của AISDL. Trên thực tế, *bnlearn* [6] của Scutari & Denis không phát triển phần kỹ thuật MCMC cho các kiểm tra thống kê mô hình lưới.

Trong một vài nghiên cứu gần đây, đã công bố [7,8] hoặc đã hoàn thành bản thảo ở trạng thái preprint (chưa qua bình duyệt) [9,10], chương trình tỏ ra hữu dụng cả về xử lý thống kê lẫn cung cấp đồ họa chất lượng (ví dụ, Hình 1).



Hình 1. Một đồ họa sử dụng trong [9]

Một số ví dụ vừa điểm ở trên cho thấy giá trị sử dụng khá linh hoạt của *bayesvl*. Đồng thời, cũng còn nhiều tiềm năng có thể tiếp tục khai thác, và những điểm cần hoàn thiện cho việc sử dụng được thuận tiện hơn.

Những vướng mắc

Đầu tiên là hạn chế của *Bản hướng dẫn sử dụng*. Do ban đầu được chuẩn bị để đảm bảo yêu cầu kỹ thuật của R Core Team đối với công tác kiểm tra, đánh giá quy chuẩn chương trình, bản hướng dẫn chỉ tập trung vào các lệnh và khai báo. Các phần hỗ trợ mô hình và lý thuyết tương đối mỏng.

Thiếu các ví dụ tiêu biểu của các mô hình ứng dụng phân tích thống kê bằng *bayesvl*. Điều đáng nói là các ứng dụng này phải đảm bảo tiêu chuẩn: đi từ đơn giản đến phức tạp. Nếu tốt hơn nữa, để đảm bảo tin cậy, thì các ứng dụng nên được bình duyệt và đã có hiệu chỉnh dựa trên đánh giá của chuyên gia.

Cần được bổ sung những phần lý giải hoặc so sánh với các mô hình *frequentist*. Mỗi liên hệ này tỏ ra có ích với những người đã quen với thống kê *frequentist*, nhưng hiện còn thiếu những ví dụ toàn diện, đặc biệt là qua dữ liệu và bài toán thực tế.

Dường như cần có một tài liệu có tính sư phạm tốt hơn (kỹ lưỡng, toàn vẹn và đầy đủ nội dung dẫn chiếu tại chỗ).

Hướng giải quyết

AISDL đặt ra hướng giải quyết các hạn chế nói trên như sau:

- Bổ sung các lớp bài toán từ dễ tới khó, sử dụng dữ liệu xã hội có thực, tốt nhất là từ các công bố quốc tế đã qua phản biện kỹ lưỡng. Chuẩn bị các đoạn chương trình cho việc xây dựng các mô hình đó sử dụng *bayesvl*, và các kết quả kèm theo.
- Sản xuất một giáo trình phục vụ việc đào tạo và sử dụng *bayesvl* trong công việc nghiên cứu, sao cho 3 ngày có thể đủ cho việc tiếp cận ban đầu và tự đọc tiếp tài liệu liên quan.
- Sử dụng tài liệu đào tạo thử trong phạm vi nhỏ, trực tiếp vào các nghiên cứu thực tế đang triển khai.

Dự án trên bắt đầu ngay từ cuối quý 1 năm 2020. Khi hoàn thành cơ bản thì cũng là lúc AISDL sẽ nâng số hiệu phiên bản chương trình lên v.1.0.0. Hy vọng công việc này còn góp phần thúc đẩy tích cực theo hướng chủ động tiếp cận khuynh hướng và nhu cầu phân tích dữ liệu của cộng đồng khoa học quốc tế [11].

Tài liệu tham khảo:

- [1] La, V. P., & Vuong, Q. H. (2019). *bayesvl*: Visually learning the graphical structure of Bayesian networks and performing MCMC with 'Stan'. *The Comprehensive R Archive Network (CRAN)*. <https://cran.r-project.org/package=bayesvl> (v.0.8.5). (Truy cập: 28/2/2020)

- [2] Vuong, Q. H., & La, V. P. (2019). BayesVL package for Bayesian statistical analyses in R. *Github*. <https://github.com/sshpa/bayesvl> (v. 0.9). (Truy cập: 28/2/2020).
- [3] Hạnh, H. P. (2019). Chương trình máy tính bayesvl trong môi trường R: Đóng góp Việt cho khoa học thế giới. *Khoa học & Phát triển* (13/6/2019). <http://khoa-hoc-phat-trien.vn/khoa-hoc/chuong-trinh-may-tinh-bayesvl-trong-moi-truong-r-dong-gop-viet-cho-khoa-hoc-the-gioi/20190613093044113p1c160.htm> (Truy cập: 29-2-2020).
- [4] Vuong, Q. H., Ho, T. M., & La, V. P. (2019). 'Stargazing' and p-hacking behaviours in social sciences: some insights from a developing country. *European Science Editing*, 45(2), 56-57.
- [5] Scutari, M., & Denis, J. B. (2014). *Bayesian Networks: with Examples in R*. New York: Chapman and Hall/CRC.
- [6] McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. New York: Chapman and Hall/CRC.
- [7] Vuong, Q. H., La, V. P., Vuong, T. T., ... , & Ho, M. T. (2018). Cultural additivity: Behavioural insights from the interaction of Confucianism, Buddhism, and Taoism in folktales. *Palgrave Communications*, 4, 143. doi:10.1057/s41599-018-0189-2.
- [8] Vuong, Q. H., Bui, Q. K., La, V. P., ..., & Ho, M. T. (2019). Cultural evolution in Vietnam's early 20th century: a Bayesian networks analysis of Hanoi Franco-Chinese house designs. *Social Sciences & Humanities Open*, 1(1), 100001; doi:10.1016/j.ssaho.2019.100001.
- [9] Vuong, Q. H., La, V. P., Ho, T. M., ... , & Ho, T. M. (2020). On how religions could accidentally incite lies and violence: Folktales as a cultural transmitter. *OSF Preprints*, doi:10.31219/osf.io/nb7tg.
- [10] Vuong, Q. H., Pham, T. H., Tran, T., ... , & Ho, T. M. (2020). STEM education and outcomes in Vietnam: Views from the social gap and gender issues. *OSF Preprints*, doi:10.31219/osf.io/unfa2.
- [11] Vuong, Q. H. (2019). Breaking barriers in publishing demands a proactive attitude. *Nature Human Behaviour*, 3(10),1034. doi:10.1038/s41562-019-0667-6.