

Trang chủ/Diễn đàn khoa học/Nghiên cứu - Trao đổi

Nhìn lại 5 năm phát triển chương trình phân tích thống kê Bayes của người Việt

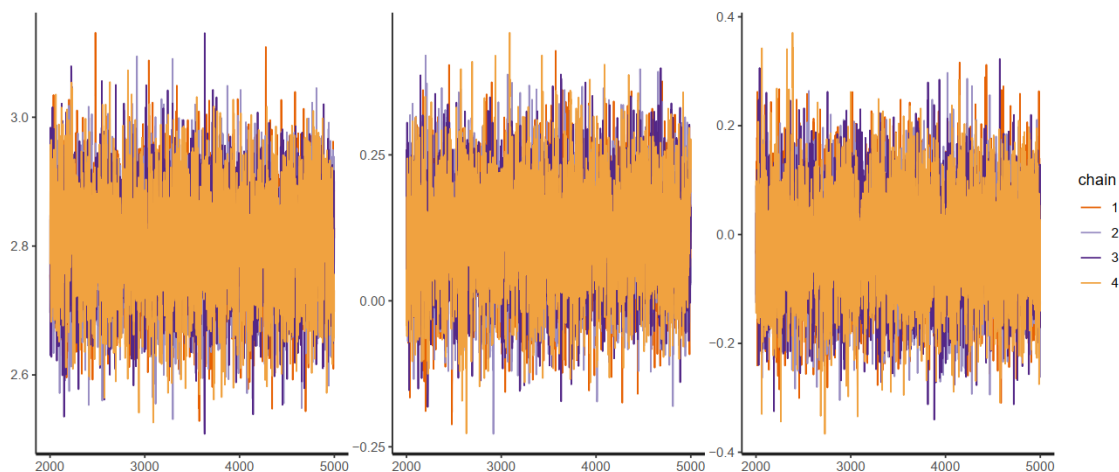
15:27 | 28/04/2024

EFR Chương trình máy tính 'bayesvl', một trong số ít những phần mềm tự phát triển của các nhà khoa học Việt Nam nhằm tăng năng suất lao động...

Chỉ còn một vài ngày nữa cả nước ta sẽ đón chào ngày lễ Quốc tế Lao động 1/5. Sẽ thật ý nghĩa khi ta kiểm đếm và nhìn lại các tác động và giá trị của thành quả lao động trước đây, để từ đó tạo động lực cùng hướng tới tương lai hăng say lao động. Một trong những sản phẩm lao động đáng chú ý của giới khoa học trong những năm gần đây là chương trình máy tính 'bayesvl', một trong số ít những phần mềm tự phát triển của các nhà khoa học Việt Nam nhằm tăng năng suất lao động [1].

Cách đây 5 năm, vào ngày 24/05/2019, chương trình máy tính 'bayesvl' chính thức được xuất bản trên R với tên gọi "bayesvl: Visually Learning the Graphical Structure of Bayesian Networks and Performing MCMC with 'Stan'". Đây là chương trình được phát triển bởi TS. Vương Quân Hoàng và kỹ sư Lê Việt Phương (Trung tâm Nghiên cứu Xã hội Liên ngành ISR, Đại học Phenikaa) với định hướng sư phạm, hỗ trợ người sử dụng làm quen với phương pháp thống kê Bayes dựa trên mô hình lưới, mô phỏng MCMC, hình ảnh hóa các thông số kỹ thuật và kết quả [2].

Tính đến hiện nay, chương trình máy tính 'bayesvl' đã nhận được gần 22 nghìn lượt tải từ khắp nơi trên thế giới, xếp hạng 11,293 trên tổng số hơn 100,000 chương trình máy tính có mặt trên R. Trong bài đánh giá các chương trình máy tính hỗ trợ phân tích Bayes trên R, nhà thống kê sinh học và kỹ sư khoa học dữ liệu Darko Medin cho rằng, 'bayesvl' là một trong những chương trình máy tính tốt nhất hiện có để phân tích lưới và trực quan hóa khung phân tích Bayes [3].

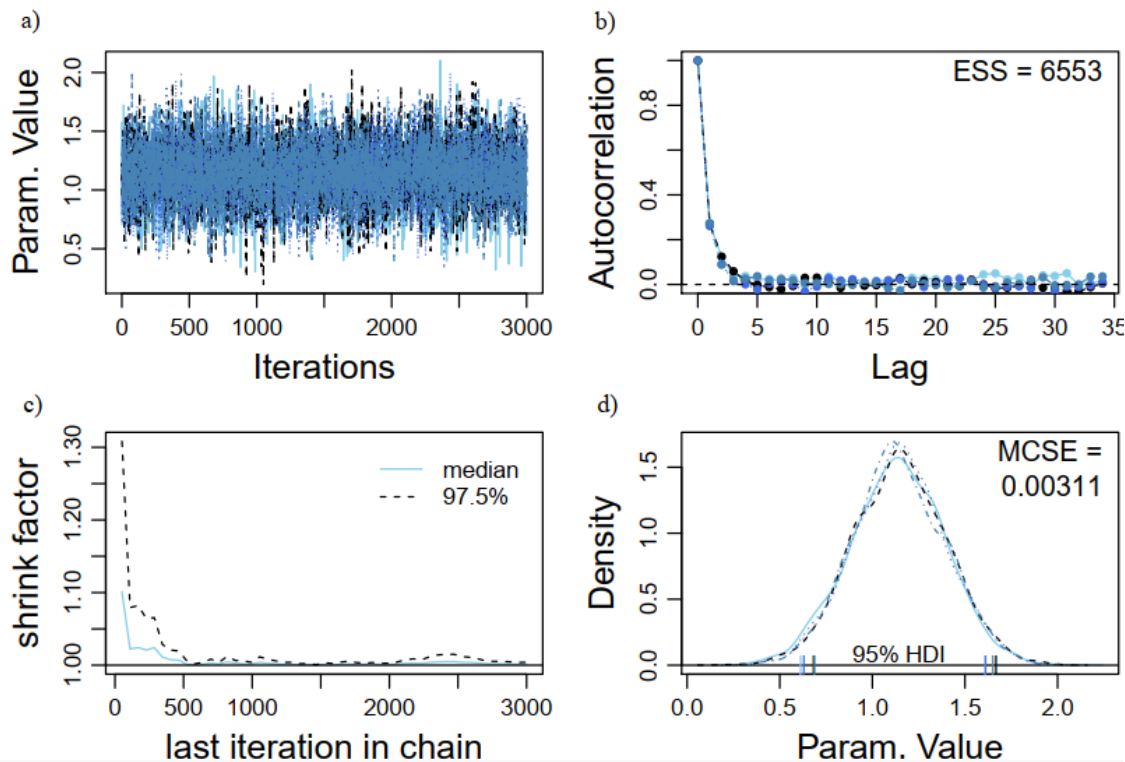


Biểu đồ truy vết trong phân tích Bayes

Có lẽ chính vì khả năng phân tích và trực quan hóa kết quả tốt của 'bayesvl', nên số lượng nghiên cứu khoa học xã hội đã xuất bản sử dụng phần mềm cũng rất đáng kể. Tính tới hôm nay, đã có khoảng gần 80 sản phẩm nghiên cứu khoa học áp dụng chương trình máy tính 'bayesvl' đã được xuất bản. Trong đó, có nhiều sản phẩm nghiên cứu được xuất bản bởi các tạp chí hàn lâm uy tín như *Humanities and Social Sciences Communications* (Nature Portfolio), *Marine Policy* (NXB Elsevier), *Research Evaluation* (NXB Đại học Oxford), *Pacific Conservation Biology* (CSIRO và Viện Hàn lâm Khoa học Úc), *npj Climate Actions* (Nature Portfolio)... Các nhà nghiên cứu sử dụng phần mềm cũng có bối cảnh đa dạng, đến từ các quốc gia phát triển và đang phát triển, như: Đức, Hà Lan, Hàn Quốc, Indonesia, Mỹ, Nam Phi, New Zealand, Nhật Bản, Thái Lan, Trung Quốc, Úc...

Bên cạnh đó, nhờ vào định hướng hỗ trợ sư phạm, nên phần mềm đã tiếp tục được phát triển và để tích hợp với Lý thuyết

Mindsponge, từ đó tạo thành phương pháp phân tích Bayesian Mindsponge Framework (BMF) [4,5]. Phương pháp này đã được thuyết giảng tại Hội thảo Thống kê Ứng dụng năm 2021 (tổ chức bởi VIASM – Viện Nghiên cứu Cao cấp về Toán) và Hội thảo về “Đổi mới phương pháp giảng dạy môn toán trong lĩnh vực khoa học xã hội” năm 2023 (tổ chức bởi VIASM và Trường Đại học Hà Nội). Sản phẩm khoa học với sự đóng góp của 34 giảng viên và nhà nghiên cứu đến từ 13 tổ chức trong buổi thuyết giảng đã được xuất bản chính thức sau 3 vòng phản biện khắc khe trên tạp chí *The VMOST Journal of Social Sciences and Humanities* [6,7]. Đây là tạp chí hàn lâm được thành lập và đầu tư trọng điểm bởi Bộ Khoa học và Công nghệ, để cung cấp nền tảng xuất bản khoa học mở uy tín, giúp lan tỏa sản phẩm tri thức đến cộng đồng quốc tế.



Kết quả phân tích sử dụng phân tích Bayes

Hiện nay, phương pháp phân tích BMF đã được sử dụng để đào tạo cho hơn 90 nhà nghiên cứu đến từ 54 tổ chức ở 15 quốc gia; những cái tên tiêu biểu có thể kể đến, như: Đại học Calcutta (Ấn Độ), Đại học Chính Pháp Trung Quốc (Trung Quốc), Đại học Monash (Úc), Đại học Pretoria (Nam Phi), Đại học Pepperdine (Hoa Kỳ), Viện Nghiên cứu Chính trị Paris (Pháp), Đại học Tây Ontario (Canada), Cao đẳng Saint Louis (Thái Lan), Đại học Công giáo Widya Mandala Surabaya (Indonesia), Trường Đại học Tôn Đức Thắng và Đại học Bách khoa Hà Nội (Việt Nam). Trong đó, 87,5% thành viên là các nhà nghiên cứu đến từ các nước đang phát triển [8].

Bởi vì chương trình máy tính 'bayesvl' được xuất bản trên R, một nền tảng mở, nên nhiều nhà nghiên cứu, đặc biệt là các nhà nghiên cứu trẻ và nhà nghiên cứu từ các nước đang phát triển, có thể dễ dàng tiếp cận và tiết kiệm chi phí cho việc nghiên cứu khoa học. Ngoài ra, để các nhà nghiên cứu có thể triển khai phân tích một cách hiệu quả, đội ngũ phát triển phần mềm đã xuất bản 2 quyển sách (1 tiếng Anh và 1 tiếng Việt) và 5 bản hướng dẫn chi tiết. Các bản hướng dẫn chi tiết đều được xuất bản mở và trải qua quá trình phản biện nghiêm ngặt trên các tạp chí uy tín, như: *SoftwareX*, *MethodsX*, *The VMOST Journal of Social Sciences and Humanities*, và *Software Impacts* [5,9-14]. Trong khi đó, sách về phương pháp phân tích đang được lưu trữ ở nhiều thư viện trên thế giới, như: thư viện của Đại học Harvard, hệ thống trường của Đại học California (tại Berkeley, Irvine và San Diego), Đại học Bonn, Đại học New York, Học viện Pratt, Đại học Ritsumeikan, Đại học bang Campinas... [15].

Ngoài các tài liệu trên, R Basics, một cộng đồng chuyên cung cấp các tài liệu về lập trình cho những người mới tiếp cận với R, đã đưa ra một bản hướng dẫn hoàn toàn độc lập cho chương trình 'bayesvl'. Bản hướng dẫn bao gồm các chi tiết cơ bản, chi tiết kỹ thuật, khả năng phân tích của chương trình [16].

R là một trong những ngôn ngữ lập trình được sử dụng nhiều nhất thế giới để tính toán thống kê và trực quan hóa dữ liệu. Ước tính khoảng 70% những nhà khai thác dữ liệu (data miner) đều sử dụng ngôn ngữ R. Ngôn ngữ này cũng được sử dụng trong các công ty công nghệ hàng đầu thế giới, như: Google, Facebook, Amazon, Netflix, Airbnb, Twitter, Microsoft.../.

References

- [1] La VP, Vuong QH. (2019). bayesvl: Visually Learning the Graphical Structure of Bayesian Networks and Performing MCMC with 'Stan'. *The Comprehensive R Archive Network*.
- [2] Hạnh HP. (2019). Chương trình máy tính bayesvl trong môi trường R: Đóng góp Việt cho khoa học thế giới. <https://khoaocphattrien.vn/khoa-hoc/chuong-trinh-may-tinh-bayesvl-trong-moi-truong-r-dong-gop-viet-cho-khoa-hoc-the-gioi/20190613093044113p1c160.htm>
- [3] Medin D. (2022). Perspectives on 20 Bayesian libraries / R implementations. <https://www.linkedin.com/pulse/perspectives-16-bayesian-libraries-r-implementations-darko-medin/>
- [4] Vuong QH. (2023). Vuong QH. (2023). *Mindspunge Theory*. Walter de Gruyter GmbH. <https://www.amazon.com/dp/B0C3WHZ2B3>
- [5] Vuong QH, Nguyen MH, La VP. (Eds.)(2022). *The mindspunge and BMF analytics for innovative thinking in social sciences and humanities*. Walter de Gruyter GmbH. <https://www.amazon.com/dp/B0C4ZK3M74>
- [6] Yến NTQ. (2023). Sinh hoạt khoa học VIASM HANU mang BMF Analytics đến với KHXH&NV. <https://kinhtevadubao.vn/sinh-hoat-khoa-hoc-viasm-hanu-mang-bmf-analytics-den-voi-khxhvn-27529.html>
- [7] Nguyen MH, et al. (2024). Effects of water scarcity awareness and climate change belief on recycled water usage willingness: Evidence from New Mexico, United States. *The VMOST Journal of Social Sciences and Humanities*, 66(1), 62-75. https://d.vjst.vn/index.php/vmost_jossh/article/view/344
- [8] Thường C. (2024). Nhóm nhà khoa học Việt phát triển Cổng thông tin hỗ trợ đào tạo nghiên cứu. <https://vjst.vn/vn/tin-tuc/8976/nhom-nha-khoa-hoc-viet-phat-trien-cong-thong-tin-ho-tro-dao-tao-nghien-cuu-.aspx>
- [9] Hoàng VQ, et al. (2021). *Bản hòa tấu dữ liệu xã hội*. Nxb Khoa học Xã hội.
- [10] Vuong QH, et al. (2020). Bayesian analysis for social data: A step-by-step protocol and interpretation. *MethodsX*, 7, 100924. <https://www.sciencedirect.com/science/article/pii/S2215016120301448>
- [11] Nguyen MH, et al. (2022). Introduction to Bayesian Mindspunge Framework analytics: An innovative method for social and psychological research. *MethodsX*, 9, 101808. <https://www.sciencedirect.com/science/article/pii/S2215016122001881>
- [12] La VP, et al. (2022). The bayesvl package: An R package for implementing and visualizing Bayesian statistics. *SoftwareX*, 20, 101245. <https://www.sciencedirect.com/science/article/pii/S2352711022001637>
- [13] Vuong QH, Nguyen MH, Ho MT. (2022). bayesvl: An R package for user-friendly Bayesian regression modelling. *The VMOST Journal of Social Sciences and Humanities*, 64(1), 85-96. https://d.vjst.vn/index.php/vmost_jossh/article/view/268
- [14] Vuong QH, et al. (2020). Improving Bayesian statistics understanding in the age of Big Data with the bayesvl R package. *Software Impacts*, 4, 100016. [https://www.softwareimpacts.com/article/S2665-9638\(20\)30003-8/fulltext](https://www.softwareimpacts.com/article/S2665-9638(20)30003-8/fulltext)
- [15] Phương M, Thường C. (2024). Sách lý thuyết và phương pháp luận khoa học của người Việt được lưu trữ trong thư viện Đại học Harvard. <https://vjst.vn/vn/tin-tuc/8895/sach-ly-thuyet-va-phuong-phap-luan-khoa-hoc-cua-nguoi-viet-duoc-luu-tru-trong-thu-vien-dai-hoc-harvard.aspx>
- [16] R Basics. (2024). The ultimate guide to the bayesvl package in R. <https://rbasics.org/packages/bayesvl-package-in-r/>

ThS. Nguyễn Thị Hồng Huệ

URL: <https://kinhtevadubao.vn/nhin-lai-5-nam-phat-trien-chuong-trinh-phan-tich-thong-ke-bayes-cua-nguoi-viet-28667.html>

© Kinh tế và Dự báo - Bộ Kế hoạch và Đầu tư