

KÍNH GỬI PHẢN BIỆN BÀI BÁO KHOA HỌC

“Dự báo tỉ giá ngoại tệ với mô hình học cộng đồng kết hợp giải thuật tiến hóa đa mục tiêu”

Trước hết nhóm tác giả xin tiếp thu và gửi lời cảm ơn tới nhà khoa học đã phản biện có những nhận xét hết sức xác đáng để chúng tôi thấy được những vấn đề còn tồn tại và cần bổ sung, làm rõ trong bài báo.

Chúng tôi đã chỉnh sửa, bổ sung cho phù hợp và giải thích một số vấn đề như dưới đây:

Ý kiến nhận xét của Phản biện 1

Ý kiến 1: Về nội dung

1. Bài báo nên đánh số thứ tự của từng mục nhỏ theo đề mục lớn để người đọc dễ theo dõi. Ví dụ 2. Lý thuyết nền tảng thì 2.1 Dự báo chuỗi thời gian, 2.1.1 Khái niệm và tương tự cho các phần khác:

Nhóm tác giả tiếp thu ý kiến và đã sửa

2. Bài báo vẫn còn chứa nhiều lỗi chính tả và cách hành văn. Cần phải sửa cách hành văn, câu cú nhiều chỗ không rõ ràng

Nhóm tác giả tiếp thu ý kiến và đã sửa

3. Hình 7. Sơ đồ khối thuật toán mạng Neural kết hợp với thuật toán NSGA-II cần được thể hiện rõ ràng hơn.

Nhóm tác giả tiếp thu ý kiến và đã sửa (phần 3.4 trang 7)

4. Chưa nhất quán cách trình bày hình 9, 10, 11: Đơn vị của các trục?

Do đơn vị tiền tệ của mỗi nước khác nhau nên tỉ giá cũng như đơn vị biểu diễn của các đồng tiền này cũng khác nhau. Do đó ở đây chúng tôi biểu diễn mỗi loại tiền trên từng đồ thị riêng biệt.

5. Hình 12, 13 của bài báo không rõ ràng:

✓ Không có đơn vị đo của các trục đồ thị.

Nhóm tác giả tiếp thu ý kiến và đã sửa (phần 4.1 trang 8)

✓ Không rõ tập dữ liệu cụ thể nào áp dụng cho biểu đồ kết quả (Chưa thể hiện rõ kết quả bảng 3).

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 9)

- ✓ Hình 12 chưa cho thấy rõ tính hiệu quả của khử xu hướng.

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 9)

- ✓ Ngoài ra, bảng 3 và hình 12, 13 cần được đánh giá về thời gian so với MSE.

Việc đánh giá về thời gian phụ thuộc vào nhiều yếu tố như cấu hình máy và các tiến trình khác đang diễn ra, ... Do đó thay vì đánh giá yếu tố thời gian ở đây chúng tôi đã bổ sung đánh giá độ phức tạp của thuật toán.

Cụ thể:

Với N là số cá thể trong quần thể; M là số hàm mục tiêu; C là số cá thể con trong mỗi cá thể, T là số thế hệ, B là số vòng lặp trong quá trình tinh chỉnh. (với $M=2$; $C=1$ hoặc 3 ; $N < B \leq T$)

Ta xem xét độ phức tạp của một vòng lặp của thuật toán. Các bước tính toán chính bao gồm:

(*) *Quá trình tiến hóa qua các thế hệ*

+ Tính toán hàm mục tiêu f1 (MSE): $O(NC)$.

+ Tính toán hàm tính mục tiêu 2 (DIV): Trong đó có 3 kiểu khoảng cách:

- Khoảng cách tập trung: $O(N)$
- Khoảng cách tới quần thể: $O(N^2)$
- Khoảng cách đường tròn: $O(N^2)$

⇒ Độ phức tạp chung của hàm tính DIV là: $O(N^2)$.

+ Thực hiện xếp hạng (Ranking): $O(MN^2)$

➔ Độ phức tạp chung của vòng lặp này là $O(TMN^2)$

(*) *Bước tinh chỉnh: $O(NB)$*

Do đó độ phức tạp chung của cả thuật toán là $O(TMN^2)$

Nhóm tác giả đã bổ sung phần đánh giá độ phức tạp của thuật toán này trong phần III.4 (trang 8) của bài báo. (Tuy nhiên nếu phản biện vẫn yêu cầu đánh giá về thời gian chúng tôi xin phép bổ sung sau).

Ý kiến 2: Về trình bày

1. Phần lý thuyết nền tảng khá nhiều không cần thiết, chỉ nên đề tham khảo đến các sách hay bài báo gốc về các lý thuyết nền tảng này (gần 50% của bài báo là tổng quan và lý thuyết nền tảng):

Nhóm tác giả đã rà soát và chỉnh sửa cho ngắn gọn và xúc tích hơn.

2. Đối với bài toán liên quan đến xử lý dữ liệu chuỗi thời gian, tập dữ liệu cần phải lớn, bài báo chỉ áp dụng tập dữ liệu khoảng 900, tập dữ liệu này là quá nhỏ để chứng minh hiệu quả thực hiện của bài báo. Bài báo cần áp dụng với tập dữ liệu lớn hơn từ 5000 – 10000 và mục tiêu cần đề cập đến thời gian thực hiện bài toán.

Trong lĩnh vực dự báo, thông thường tất cả các dữ liệu này được thu thập từ trên Internet và là những bộ dữ liệu khá phổ biến được dùng để thử nghiệm trong các nghiên cứu. Đối với chuỗi thời gian thì những dữ liệu quá lâu thường không có ảnh hưởng nhiều với dữ liệu hiện tại đặc biệt với những dữ liệu phổ biến hiện nay như tiền tệ, tài chính. Ở đây nhóm tác giả cho rằng 900 điểm dữ liệu (được thu thập từ 19/10/2012 đến 06/04/2015) là đủ lớn để thực hiện thực nghiệm.

3. Thuật toán NSGA-II (Nondominated Sorting Genetic Algorithm II) cần ghi rõ nguồn gốc tham khảo, các tác giả chỉ đề cập đến tác giả của NSGA. Thuật toán này tuy ghi rõ trong phần Lý thuyết nền tảng phía sau, nhưng trong phần Giới thiệu ban đầu đã có đề cập tới mà không ghi rõ nguồn gốc sẽ làm người đọc nhầm lẫn đây là thuật toán của các tác giả đề xuất.

Nhóm tác giả tiếp thu ý kiến và đã bổ sung tài liệu tham khảo và trích dẫn [10], [11] (trang 4)

4. Mục 4. Mô tả chi tiết giải thuật của phần III, ở Bước 1: Khởi tạo quần thể (kích thước quần thể là N , kích thước nhóm cá thể là K), việc thực hiện tách rời 2 lần duyệt cho mỗi lần tính hàm mục tiêu MSE và Diversity sẽ làm cho thời gian thực hiện thuật toán tăng đối với tập dữ liệu lớn. Cần tối ưu bước này.

Mặc dù việc tính toán tách rời 2 lần duyệt cho mỗi lần tính hàm mục tiêu sẽ làm cho thời gian thực hiện thuật toán tăng. Tuy nhiên do xuất phát từ ý tưởng lựa chọn hàm mục tiêu f_2 (DIV) của chúng tôi là để việc lựa chọn các cá thể diễn ra trên toàn không gian tìm kiếm, tránh việc hội tụ sớm (nếu chỉ đơn thuần sử dụng 1 hàm mục tiêu MSE) vào các cực trị địa phương. Cụ thể trong bài báo này chúng tôi tính toán hàm DIV dựa trên khoảng cách giữa các giá trị MSE của từng cá thể. Điều này dẫn đến tất cả các cá thể sẽ phải được tính toán giá trị MSE trước sau đó mới có thể tính toán được giá trị DIV.

Trong quá trình thiết kế thuật toán khi lựa chọn hàm DIV chúng tôi đã xác định sẽ đánh đổi việc tiêu tốn về thời gian thực thi với việc đảm bảo quần thể sẽ đa dạng hơn và không bị hội tụ về cực trị cục bộ.

Độ phức tạp của thuật toán chung vẫn không thay đổi như chúng tôi đã giải trình trong Ý kiến 1.

Nhóm tác giả đã bổ sung giải thích cho ý này ở phần phân tích thuật toán trong trang 7.

Ý kiến nhận xét của Phản biện 2

Ý kiến 1: Về nội dung

1. Xây dựng mô hình công đồng là cách làm phổ biến để cải thiện hiệu suất của loại bài toán phân lớp và hồi quy, vì các nhà khoa học đã minh chứng được dự báo dùng phương pháp công đồng thì hiệu suất tốt hơn phương pháp đơn.

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 1)

2. Vấn đề ứng dụng (Evolutionary Multi Objective Optimization – EMOO) được triển khai nhiều trong lĩnh vực dự báo chuỗi thời gian. -> câu này không chính xác.

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 1)

3. Trong phần 3, mô hình toán học của bài toán nên phát biểu đầu đủ Đầu vào, Đầu ra, Các ràng buộc, Các hàm mục tiêu chứ không nên chỉ viết mỗi hàm mục tiêu.

Nhóm tác giả tiếp thu ý kiến và đã sửa (phần 3.2 trang 6)

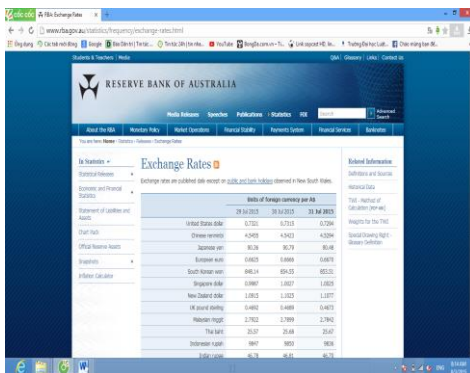
4. Phần phát biểu bài toán hoặc phần giới thiệu cần chỉ rõ bài toán Dự báo chuỗi thời gian có là bài toán NP-khó hay không mà lại sử dụng giải thuật di truyền để giải quyết.

Nhóm tác giả tiếp thu ý kiến và đã sửa (phần 3 trang 5)

5. Phần Lý thuyết nền tảng nên rút ngắn lại: cùng với ý kiến của phản biện 1.

6. Trang web chưa dữ liệu thử nghiệm ở tài liệu tham khảo [24] không truy cập được nữa.

Trong ngày 03/08/2015 tác giả đã kiểm tra lại và thấy đường link vẫn truy cập bình thường (như hình bên dưới).



Rates of foreign currencies per A\$			
	28 Jul 2015	30 Jul 2015	28 Jul 2015
United States dollar	0.7301	0.7315	0.7296
Chinese renminbi	6.5403	6.5413	6.5394
Singapore dollar	81.38	81.73	81.46
Japanese yen	106.71	106.89	106.53
South Korean won	886.14	894.35	885.31
Indonesian Rupiah	1,586.7	1,587.7	1,585.9
New Zealand dollar	1.0473	1.0473	1.0471
UK pound sterling	0.6403	0.6409	0.6403
Malaysian Ringgit	2.7323	2.7389	2.7362
Thai Baht	25.57	25.68	25.67
Indonesian Rupiah	1,586.7	1,587.7	1,585.9
Chinese renminbi	6.5403	6.5413	6.5394

7. Nếu tỷ lệ dữ liệu kiểm tra thay đổi thì kết quả có thay đổi không.

Nếu chọn tỉ lệ dữ liệu kiểm tra thay đổi thì kết quả có thể thay đổi. Tuy nhiên sẽ không thay đổi nhiều nếu tính theo MSE. Đây là tỉ lệ thường chọn trong lĩnh vực học máy (train-test: 70%-30%).

8. Các tham số cài đặt thay đổi thì kết quả có thay đổi không.

Các tham số cài đặt thay đổi thì kết quả có thể thay đổi. Đây là các tham số thường chọn trong giải thuật tiến hóa.

9. Các hình vẽ cần bổ sung giá trị trên các trục toạ độ

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 8-10)

10. Hình 14, 15, 17, giá trị ở trục tung không chính xác.; $0.0001 < 0.000012$???

Trên hình vẽ thể hiện $0.00012 (> 0.0001)$ chứ không phải **0.000012**

11. Tài liệu tham khảo viết chưa chuẩn

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 10)

12. Sửa lại tiêu đề tiếng Anh.

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 1)

13. Phần Cài đặt thử nghiệm, tác giả nên trình bày rõ hơn.

Nhóm tác giả tiếp thu ý kiến và đã sửa (trang 9)