Phương pháp làm trơn dữ liệu (Smoothing Methods)

Các phương pháp làm trơn là tập hợp các phương pháp, trong đó các giá trị dự báo mới ở một thời đoạn sẽ được ước lượng bằng việc kết hợp giá trị dự báo và giá trị quan sát ở thời đoạn ngay trước đó. Đây là một trong những kĩ thuật đơn giản nhưng hiệu quả để sử dụng dự báo đối với dữ liệu chuỗi thời gian có tính chất tuyến tính. Gồm có các hiện thực sau:

1.Kĩ thuật làm trơn lũy thừa giản đơn (Simple exponential methods)

(1)

Với α được gọi là hệ số làm trơn (0 < α < 1).

Đây là hiện thực giản đơn nhất trong các phương pháp làm trơn dữ liệu. Từ công thức (1) ta có thể biến đổi thành

Vì thế kĩ thuật này mới được gọi là kĩ thuật làm trơn lũy thừa. Đối với giá trị α ta hoàn toàn có thể ước lượng bằng cách sử dụng các phương pháp ước lượng tối thiểu SSE, MSE, MAE, …

2. Kĩ thuật làm trơn lũy thừa Holt

Đối với dự liệu chuỗi thời gian có tính xu hướng rõ rệt, việc sử dụng kĩ thuật làm trơn giản đơn không còn hiệu quả. Vì thế, một kĩ thuật cải tiến đã được sử dụng đó là kĩ thuật làm trơn Holt. Trong kĩ thuật này, dữ liệu chuỗi thời gian được tách thành hai thành phần là thành phần bậc (level) và thành phần xu hướng (trend), khi đó giá trị Y tại thời điểm (t+p) sẽ được ước lượng bởi

Trong đó, Lt và Tt là hai thành phần bấc và xu hướng được ước lượng như sau:

Đây chính là một sự mở rộng từ kĩ thuật làm trơn giản đơn nếu cho thành phần xu phướng trong dữ liệu bằng 0.

3.Kĩ thuật làm trơn lũy thừa Winter

Đặc biệt, đối với dữ liệu chuỗi thời gian có tính mùa (seasonal), một phương pháp được sử dụng là kĩ thuật làm trơn Winter. Kĩ thuật này có hai mô hình hiện thực: mô hình nhân (multiplicative) và mô hình cộng (addictive). Công thức cho từng mô hình này như sau:

Mô hình nhân: thành phần mùa được xem là một thành phần nhân trong dữ liệu

Mô hình cộng: thành phần mùa được xem là một thành phần cộng trong dữ liệu

Trong đó:

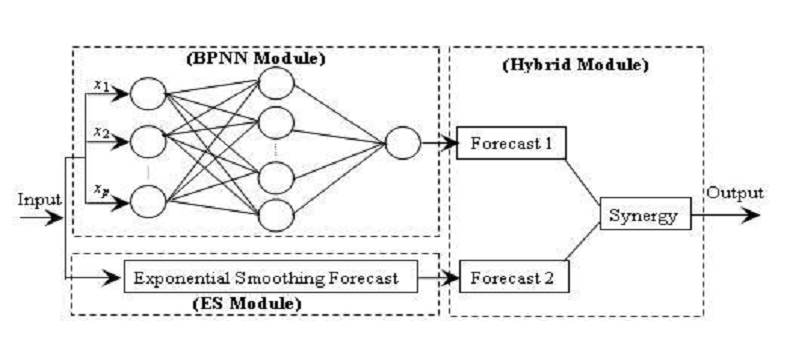
L,T, S lần lượt là các thành phần bậc, xu hướng, mùa của chuỗi thời gian.

α,β,γ là các hệ số làm trơn (0 ≤ α,β,γ ≤1) tương ứng với các thành phần L, T, S

Nếu hệ số γ trên bằng 0, nghĩa là thành phần mùa không tồn tại trong dữ liệu, kĩ thuật này trở thành kĩ thuật làm trơn Holt. Nếu hệ số β cũng bằng 0, mô hình này trở thành mô hình làm trơn lũy thừa giản đơn. Như vậy, đây chính là kĩ thuật tổng quát nhất trong 3 kĩ thuật (giải quyết trọn vẹn cho ba kiểu dữ liệu).

4.Kết hợp kĩ thuật làm trơn lũy thừa với mô hình Neuron Network

Dữ liệu chuỗi thời gian, đặc biệt là dữ liệu chuỗi thời gian có tính mùa và xu hướng rõ rệt thì kĩ thuật làm trơn lũy thừa xử lí khá tốt và cho kết quả với độ sai số nhỏ. Tuy nhiên trong trường hợp thành phần ngoại lai của chuỗi dữ liệu chiếm nhiều thì mô hình neuron network lại xử lí tốt hơn và ngược lại. Vì thế, để cải tiến hiệu suất của dự báo, một mô hình mới, mô hình lai (hybrid) giữa mô hình Neuron Network và kĩ thuật làm trơn lũy thừa được đưa ra. Cấu trúc mô hình được mô tả ở hình bên dưới:



Mô hình lai gồm ba phần tách biệt: Neuron Network, Exponential Smoothing và Hybrid. Trong đó, hai thành phần Neuron Network và Exponential Smoothing là hai thành hoàn toàn độc lập nhau. Mỗi thành phần sẽ được huấn luyện trên cùng một tập dữ liệu (tìm các trọng số giữa các perceptrons và bias tương ứng của chúng đối với mô hình Neuron Network và tìm ba hệ số α,β,γ đối với mô hình Exponential Smoothing). Sau đó, cả hai mô hình này sẽ truyền toàn bộ các kết quả dự báo của mình cho mô hình Hybrid, tại đây sẽ cho ra kết quả cuối cùng bằng cách kết hợp hai kết quả từ Neuron Network và Exponential Smoothing theo công thức:

Trong đó, α được gọi là hệ số lai (0 ≤ α ≤ 1). Nếu α bằng 0, nghĩa là kết quả dự đoán bằng kĩ thuật làm trơn lũy thừa là tốt nhất, nếu α bằng 1, nghĩa là kết quả dự đoán bằng mô hình Neuron Network là tốt nhất. Việc ước lượng giá trị α có thể sử dụng phương pháp tối thiểu đại lượng MSE của chuỗi dự báo và chuỗi giá trị thực.