

1 Introdução

1.1. Contexto

Análise e previsão de séries temporais é notadamente uma disciplina com muitas aplicações práticas, capaz de oferecer suporte à decisão em diversas áreas:

1. Energia;
 2. Cadeia de suprimentos;
 3. Finanças;
 4. Administração pública;
 5. Gestão de pessoas;
- etc.

Há na literatura uma lista extensa de aplicações usando métodos de séries temporais para previsão de grandezas tais como demanda, preço, volume e vazão (Harvey, 1991; Pankratz, 1991; Pole et al. 1994; Makridakis et al. 1998; Souza & Camargo, 2004; Chase, 2009). Logicamente a atividade de previsão pode ser mais ou menos precisa, de acordo com a natureza da variável sendo estudada. Por exemplo, a previsão do **preço** do petróleo é tipicamente uma atividade mais complexa do que a previsão da **demand**a por óleo diesel ou gasolina, assim como os modelos de previsão meteorológica têm ainda hoje considerável grau de incerteza. Estes e outros exemplos da complexidade de modelagem e previsão em sistemas econômicos ou físicos podem ser vistos em Orrell (2010).

No processo de tomada de decisão, normalmente múltiplas previsões para uma mesma variável estão disponíveis para o gestor. Neste contexto, como observado em Timmerman (2006), uma questão surge naturalmente: qual a melhor forma de se aproveitar as informações de cada previsor? Como será visto ao longo deste trabalho, a multiplicidade de previsões não deve ser considerada indesejada; ao contrário, ela permite a construção dos chamados sistemas **multi-**

previsores, que integram, de alguma maneira, as previsões disponíveis. Se bem projetados, estes sistemas levam a previsões consensuais superiores às previsões individuais; esta questão é abordada de maneira bastante prática em Chase (2009).

1.2.

Sistemas multi-previsores: seleção versus combinação

Considere-se um cenário com uma série temporal de interesse e um conjunto de previsores disponíveis para ela. Ao decidir por usar todas as informações ao seu alcance, um analista institui um sistema multi-previsor, ora **selecionando** o previsor de maior potencial, ora estabelecendo um esquema de **combinação** ponderada. Esta seção procura ilustrar este processo.

Com o objetivo de prever a capacidade de geração energética em parques eólicos, Sanchez (2008) exemplifica, de maneira didática, o funcionamento de um sistema multi-previsor. Considerando 2 parques eólicos e 4 previsores (*experts*¹), a Figura 1 exibe a evolução do erro quadrático médio (MSE – *Mean Squared Error*) observado em previsões até 18 passos a frente.

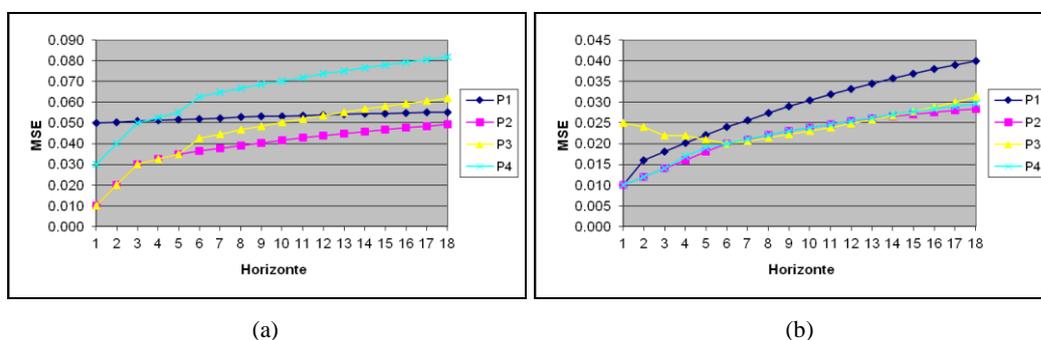


Figura 1 – Evolução do erro quadrático médio (MSE) na previsão da capacidade de geração energética em 2 parques eólicos. Para cada um dos parques, (a) e (b), há 4 previsores disponíveis: P₁, P₂, P₃ e P₄. O horizonte de previsão é de até 18h a frente. Observa-se que o parque (a) apresenta erros maiores; isso é explicado por fatores meteorológicos.

A análise dos gráficos acima leva a conclusões distintas para cada parque gerador. Por exemplo, para o parque (a) parece haver uma certa dúvida sobre qual é o melhor previsor, P₂ ou P₃, para horizontes $h < 6$. Para horizontes $h > 6$, P₂ tem o melhor desempenho. Assim, alguém poderia concluir que uma boa estratégia

¹ Neste trabalho a palavra *expert* é usada como sinônimo de previsor.

² Os termos **combinação de previsores** e **combinação de previsões** são usados intercaladamente.

para esse parque poderia ser (i) usar a média simples entre as previsões de P_2 e P_3 para horizontes $h < 6$ e (ii) usar P_2 sozinho para horizontes maiores. Já para o parque (b), não está claro qual é o melhor previsor, especialmente para horizontes $6 \leq h \leq 15$. Uma estratégia imediata seria (i) usar a média simples entre as previsões de P_2 e P_4 para horizontes $h < 6$ e (ii) adicionar as previsões de P_3 a esta média, quando os horizontes forem maiores.

A questão sobre se é melhor selecionar um único previsor, potencialmente melhor que qualquer outro, ou estabelecer um método de combinação ponderada dos previsores disponíveis será aprofundada a seguir.

1.3. Combinação de previsores²

Do ponto de vista teórico, selecionar um previsor é o mesmo que fazer a média ponderada dos previsores disponíveis, com todo peso concentrado no previsor escolhido (Sanchez, 2008). Contudo, do ponto de vista prático, estes procedimentos levam a diferentes implementações computacionais, e o analista precisa definir se é melhor **selecionar** um previsor dominante (o que normalmente é feito pela análise histórica dos desempenhos), ou se haverá, a cada instante, uma **combinação**, sendo a contribuição de cada previsor devidamente ponderada (o que normalmente também é feito pela análise histórica dos desempenhos).

O resultado de uma combinação de previsores não será sempre melhor do que a melhor previsão individual disponível, mas na prática é menos arriscado combinar previsores do que selecionar um método individual (Hibon & Evgeniou, 2005). Um dos fatores que contribui para o risco de seleção é a necessidade de técnicos que tenham, ao mesmo tempo, especialização em séries temporais e conhecimento na área de negócio sendo estudada: em muitos casos isso não é factível por falta de tempo, dinheiro ou mão de obra (Lemke & Gabrys, 2010). Além disso, conclusões empíricas demonstram que mesmo que o melhor previsor possa ser identificado a cada instante, a combinação pode ainda ser uma estratégia atraente, por oferecer ganho potencial em **diversificação** (ou complementaridade entre previsores). Contudo, o sucesso desta abordagem vai depender do quão bem

² Os termos **combinação de previsores** e **combinação de previsões** são usados intercaladamente.

os pesos da combinação são determinados (Timmerman, 2006). Ainda nesta linha, Zou & Yang (2004) afirmam que quando há muita incerteza na determinação do melhor modelo de previsão, a combinação reduz a **instabilidade** dos resultados e melhora o desempenho. Os mesmos autores indicam que não é uma boa ideia combinar cegamente todos os previsores disponíveis; uma análise preliminar, e.g. análise das autocorrelações dos resíduos, deve ser realizada para obtenção de uma lista de previsores razoáveis. Conclusão similar também aparece em Timmerman (2006): corte dos piores previsores (*trimming*) e agrupamento dos previsores com desempenhos similares podem trazer melhoria considerável na previsão combinada.

A combinação de previsores foi usada com sucesso em diversas áreas (Clemen, 1989). Ainda hoje, esta linha de pesquisa é citada como um tópico atraente do ponto de vista prático (Chase, 2009), e servirá de base para os estudos realizados neste trabalho.

1.4. Motivação

Ponderar um previsor é associar a ele um valor (peso) representando o quanto ele irá contribuir em uma futura combinação com outros previsores. A forma mais simples de ponderação é a **estática**: estima-se um único vetor de pesos com base no histórico e repete-se este vetor ao longo de todo o horizonte de previsão. Esta alternativa pode ser incrementada ao considerar-se métodos para geração **dinâmica** de pesos, i.e., métodos onde os vetores de pesos gerados variam ao longo do horizonte de previsão. Na prática, não há garantia de que a geração dinâmica será a melhor opção sempre, mas variar os pesos da combinação no tempo pode muitas vezes melhorar o desempenho final (Timmerman, 2006).

Como será visto na seção 2.4, quase todos os métodos consagrados de ponderação estão fortemente baseados nos desempenhos dentro da amostra³ dos previsores envolvidos (uma exceção importante é a combinação por média simples, onde independente dos desempenhos, os pesos são sempre iguais).

³ A expressão **dentro da amostra** se refere aos dados (históricos) usados para ajustar um modelo estatístico. A expressão **fora da amostra** se refere aos dados separados para teste.

A alternativa para esta abordagem **tradicional** seria tentar modelar **explicitamente** a variação temporal (evolução) dos pesos de combinação, uma tarefa de complexidade elevada e sem garantia de superioridade. A ponderação baseada **exclusivamente** nos erros históricos de previsão (ou em medidas derivadas destes erros) é a mais citada na literatura e apresenta resultados difíceis de serem superados (*hard to beat*); mas a falta de um modelo auxiliar que **atualize** as informações fora da amostra ainda pode ser vista como uma **limitação**.

A dependência **exclusiva** dos desempenhos medidos dentro da amostra, observada como uma característica potencialmente limitante na geração tradicional de pesos em horizontes de previsão maiores do que 1, motivou, em grande parte, a realização deste trabalho. Desta motivação inicial, chegou-se a seguinte proposta: desenvolver uma metodologia para geração dinâmica de pesos múltiplos passos a frente, que possa **relaxar** a dependência dos desempenhos históricos e assim agregar valor aos procedimentos tradicionais. O termo **agregar valor** significa trazer alguma vantagem, seja em **desempenho** ou em **abstração** da complexidade envolvida na modelagem explícita – abstrair (**encapsular**) a complexidade facilita a automação do processo de combinar previsores.

1.5. Objetivo

O objetivo principal deste trabalho é, considerando a combinação linear de previsores, desenvolver uma metodologia para geração dinâmica de pesos que possa agregar valor aos procedimentos tradicionais, principalmente, nas previsões múltiplos passos a frente. Para cumprir esta meta, propõe-se o sistema batizado de Ponderação Neural de Experts (NEW – *Neural Expert Weighting*), que deve permitir a criação de modelos robustos para ponderação dinâmica de previsores. A ideia fundamental do sistema é utilizar modelos de **redes neurais** para aprender e generalizar a partir de um ou mais métodos tradicionais de ponderação, de maneira que a dependência dos desempenhos históricos dos previsores envolvidos possa ser relaxada.

1.6. Descrição do trabalho e contribuição

O uso de redes neurais para análise e previsão de séries temporais não é novidade (Adya & Collopy, 1998; Zhang, 1998). Crone & Graffeuille (2004), citados em (NN3, 2011), afirmam haver mais de 2000 publicações nesta linha. Apesar disso, o emprego de redes neurais com o objetivo específico de geração **dinâmica** de pesos (para combinação linear de previsores) não é uma área explorada.

Prudêncio & Ludermir (2006) propõem um sistema de ponderação baseado em redes neurais. A abordagem utilizada por eles sugere a inferência de pesos **estáticos** (fixos) para cada um dos previsores envolvidos, com base em um conjunto de características extraídas da série em questão (comprimento, tendência básica, percentual de pontos de inflexão etc.). A definição de qual o melhor conjunto de características para uma dada série constitui um ponto bastante sensível do sistema, uma vez que as atividades desta natureza (seleção de variáveis) são quase sempre não triviais.

Este trabalho propõe o sistema NEW como uma nova metodologia para a geração dinâmica de pesos múltiplos passos a frente – e posterior combinação linear de previsores – inspirando-se na ideia de que a dependência dos desempenhos históricos dos previsores envolvidos na combinação possa ser relaxada. O sistema emprega redes neurais para simular funções de ponderação, e tem seu funcionamento baseado nas seguintes variáveis (capítulo 3): (i) **previsões** (dentro e fora da amostra), (ii) **série de referência** e (iii) **pesos históricos**. A capacidade das redes neurais de aprender e generalizar funções não lineares garante a robustez necessária para a construção de modelos de ponderação potencialmente eficazes, sempre visando horizontes de previsão maiores do que 1.

1.7. Notação básica

Com o objetivo de uniformizar a notação utilizada, define-se aqui as grandezas mais referenciadas ao longo do texto:

- a) Série de **treinamento** com τ observações (série realizada ou série histórica):

$$y^\tau = [y_1 \quad y_2 \quad \dots \quad y_\tau]' \quad (1)$$

- b) Série de **teste** com horizonte máximo H (série fora da amostra):

$$y^{\tau+H|\tau} = [y_{\tau+1} \quad y_{\tau+2} \quad \dots \quad y_{\tau+H}]' \quad (2)$$

- c) Vetor de previsões no instante $t + h$ ($h \leq H$), estimado com dados até t , para N previsores:

$$\hat{\mathbf{y}}_{t+h|t} = [\hat{y}_{t+h|t,1} \quad \hat{y}_{t+h|t,2} \quad \dots \quad \hat{y}_{t+h|t,N}]' \quad (3)$$

- d) Vetor de pesos de combinação em $t + h$, estimado com dados até t , com M dimensões⁴:

$$\hat{\mathbf{w}}_{t+h|t} = [\hat{w}_{t+h|t,1} \quad \hat{w}_{t+h|t,2} \quad \dots \quad \hat{w}_{t+h|t,M}]' \quad (4)$$

- e) Previsão combinada em $t + h$, estimada com dados até t . A função f representa um modelo genérico de combinação:

$$y_{t+h|t}^c = f(\hat{\mathbf{y}}_{t+h|t}, \hat{\mathbf{w}}_{t+h|t}) \quad (5)$$

⁴ No caso mais geral de combinação (combinação não linear), o número de pesos (M) não é necessariamente igual ao número de previsores (N).

1.8. Organização da tese

O restante do texto está organizado da seguinte forma: o capítulo 2 apresenta conceitos básicos sobre sistemas multi-previsores, tratando especialmente dos modelos de combinação linear. O capítulo 3 introduz o sistema NEW, detalhando sua estrutura matemática e suas possibilidades de uso. O capítulo 4 traz o estudo de casos selecionados e divide-se em duas grandes seções: derivados do petróleo (seção 4.4) e competição NN3 (seção 4.5). O capítulo 5 apresenta as considerações finais, com conclusões e sugestões para estudos futuros.