

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA
RECONCILIAÇÃO ÓTIMA DE SÉRIES
TEMPORAIS HIERÁRQUICAS E AGRUPADAS

VITÓRIA

2022

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA
RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS
HIERÁRQUICAS E AGRUPADAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Guilherme A. A. Pereira

VITÓRIA

2022

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA DE
SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS/ ALBERSON DA SILVA MI-
RANDA. – VITÓRIA, 2022-

26p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Guilherme A. A. Pereira

Dissertação (Mestrado) – UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA, 2022.

1. Palavra-chave1. 2. Palavra-chave2. 2. Palavra-chave3. I. Orientador. II. Universidade
xxx. III. Faculdade de xxx. IV. Título

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA
RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS
HIERÁRQUICAS E AGRUPADAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Economia.

Aprovada em xx de xx de 20xx.

COMISSÃO EXAMINADORA

Prof. Dr. Guilherme A. A. Pereira
Universidade Federal do Espírito Santo
Orientador

Professor
Instituição

Professor
Instituição

RESUMO

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Palavras-chave: palavra-chave1. palavra-chave2. palavra-chave3.

ABSTRACT

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Keywords: keyword1. keyword2. keyword3.

LISTA DE FIGURAS

Figura 1 – Séries Hierárquicas	12
Figura 2 – Séries Agrupadas	13
Figura 3 – Séries Hierárquicas Agrupadas (a)	13
Figura 4 – Séries Hierárquicas Agrupadas (b)	14

LISTA DE QUADROS

Quadro 1 – Artigos de referência em Hyndman e Athanasopoulos (2021)	20
Quadro 2 – Trabalhos encontrados na busca estendida	20

LISTA DE TABELAS

Tabela 1 – Trabalhos mais citados com os termos “hierarquical forecast reconciliation”	19
----------------------------------------------------------------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

MinT	<i>Minimum Trace</i>
MCRL	Modelo Clássico de Regressão Linear
MQO	Mínimos Quadrados Ordinários
MQP	Mínimos Quadrados Ponderados

LISTA DE SÍMBOLOS

t	Tempo dentro-da-amostra
T	Último tempo dentro-da-amostra
h	Horizonte de previsão, tempo fora-da-amostra
Ω	Conjunto de dados dentro-da-amostra
y	Série temporal dentro-da-amostra
\hat{y}	Série temporal estimada
\tilde{y}	Série temporal reconciliada
\mathbf{S}	Matriz de soma
\mathbf{G}	Matriz de reconciliação
$\{\dots\}$	Conjunto
$ \{\dots\} $	Número de elementos em um conjunto

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Séries hierárquicas e séries agrupadas	12
1.2	Abordagens top-down e bottom-up	14
1.3	Coerência e reconciliação	18
1.4	Motivação	18
1.5	Objetivos	19
1.6	Revisão da literatura	19
2	MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO	
	ÓTIMA	21
2.1	O processo de ajuste e sobreajuste	21
2.2	Reamostragem	22
	Referências	25
	Índice	26

1 INTRODUÇÃO

Neste trabalho,

1.1 Séries hierárquicas e séries agrupadas

Séries temporais hierárquicas são aquelas que podem ser agregadas ou desagregadas naturalmente em uma estrutura aninhada (HYNDMAN; ATHANASOPOULOS, 2021). Para ilustrar, tome a série do PIB brasileiro. Ela pode ser desagregada por estado que, por sua vez, pode ser desagregada por município.

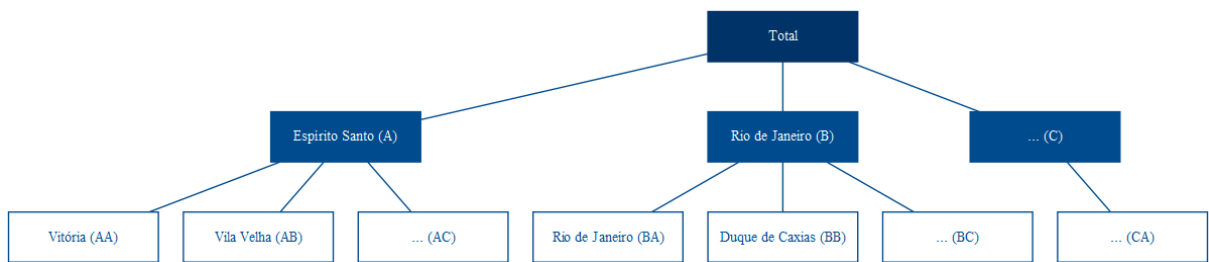


Figura 1 – Séries Hierárquicas

Essa estrutura pode ser representada por equações para qualquer nível de agregação.

$$y_t = y_{A,t} + y_{B,t} + y_{C,t} \quad (1)$$

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BC,t} + y_{CA,t} \quad (2)$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \quad (3)$$

Assim, o agregado nacional pode ser representado apenas pelos agregados dos estados, através de (1), ou como o agregado dos municípios (2). Já o agregado para o estado do Espírito Santo é representado por (3).

Alternativamente, podemos descrever a estrutura completa de forma matricial:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{BC,t} \\ y_{CA,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{BC,t} \\ y_{CA,t} \end{bmatrix} \quad (4)$$

Por outro lado, o PIB pode ser também desagregado de forma cruzada de acordo com a atividade econômica — lavoura, rebanho, indústria de transformação, extrativa, bens de capital, bens intermediários, comércio de vestuário, automotivos, serviços etc. Essa estrutura não pode ser desagregada naturalmente de uma única forma, como é a hierarquia de estados e municípios. Não pode ser aninhada por um atributo como a própria geografia. A esse tipo de estrutura dá-se o nome de séries agrupadas.

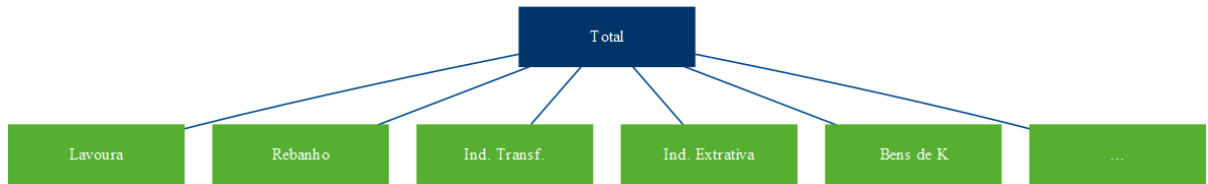


Figura 2 – Séries Agrupadas

Combinando as duas, temos a estrutura de séries hierárquicas agrupadas. Ao contrário da estrutura hierárquica, que só pode ser agregada de uma forma — como com os municípios abaixo dos estados —, a adição da estrutura agrupada pode ocorrer tanto acima (Figura 3) quanto abaixo (Figura 4) da hierárquica.

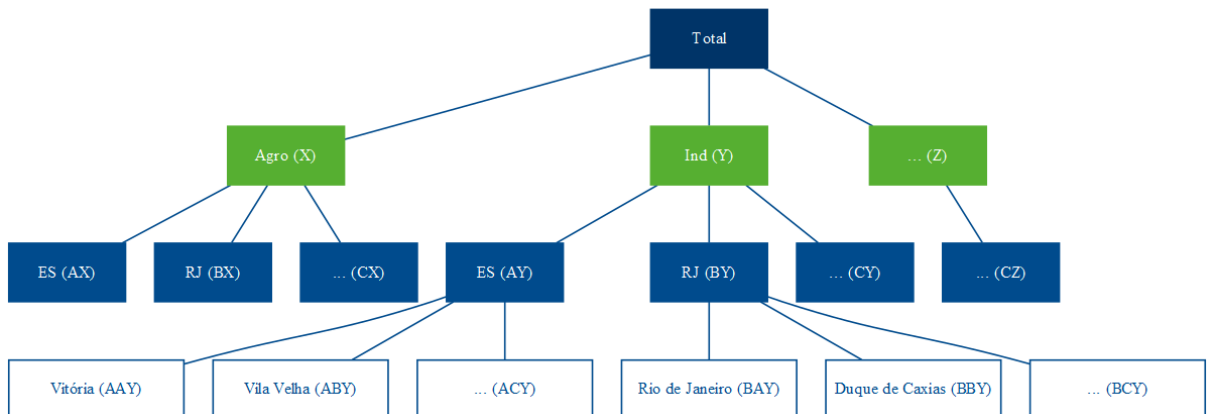


Figura 3 – Séries Hierárquicas Agrupadas (a)

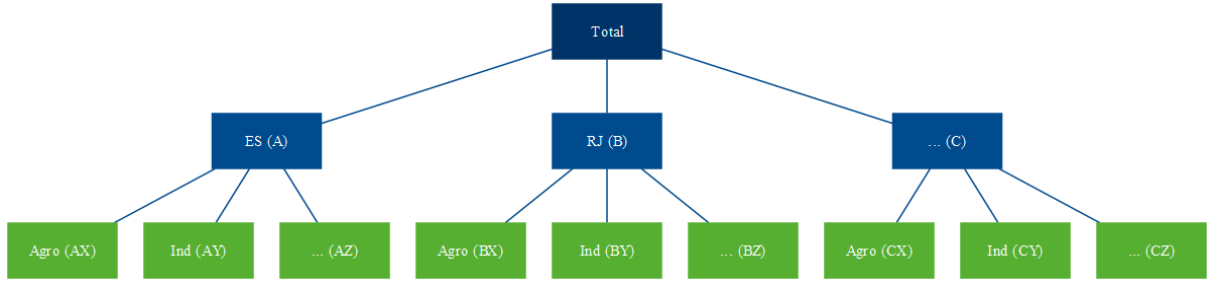


Figura 4 – Séries Hierárquicas Agrupadas (b)

Na notação matricial, a estrutura da Figura 4 é representada como abaixo. Formalmente, o primeiro membro da igualdade é composto pelo vetor \mathbf{y}_t n -dimensional com todas as observações no tempo t para todos os níveis da hierarquia. O segundo membro é composto pela matriz de soma \mathbf{S} de dimensão $n \times m$ que define as equações para todo nível de agregação, e pela matriz \mathbf{b}_t composta pelas séries no nível mais desagregado.

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{Z,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{bmatrix} \quad (5)$$

1.2 Abordagens top-down e bottom-up

Talvez as formas mais intuitivas de se pensar em previsões para esses tipos de estrutura sejam as abordagens top-down e bottom-up. Tome a estrutura descrita na Figura 1, por exemplo. Podemos realizar a previsão para o horizonte de tempo h do agregado do PIB brasileiro,

representado no topo da hierarquia por *Total* (6), e então distribuir os valores previstos proporcionalmente entre os estados e municípios.

$$\hat{y}_{T+h|T} = E[y_{T+h}|\Omega_T] \quad (6)$$

Essa é a abordagem top-down. Nela, a previsão para os níveis mais desagregados da hierarquia são determinadas por uma proporção p_i do nível agregado. Por exemplo, as previsões para Vitória são dadas pela equação 7.

$$\tilde{y}_{AA,T+h|T} = p_1 \hat{y}_{T+h|T} \quad (7)$$

Para isso, temos de definir uma matriz com todos esses pesos, que, seguindo a formulação de Hyndman e Athanasopoulos (2021), vamos chamar de G :

$$G = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (8)$$

G é uma matriz $m \times n$ que multiplica a matriz $\hat{y}_{T+h|T}$ que, por sua vez, é composta pelas previsões base — as previsões individuais para todos os níveis de agregação. A equação para a abordagem *top-down* será, então:

$$\tilde{y}_{T+h|T} = SG\hat{y}_{T+h|T} \quad (9)$$

Na notação matricial para a estrutura da Figura 1, temos:

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \mathbf{S} \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{y}_{T+h|T} \\ \hat{y}_{A,T+h|T} \\ \hat{y}_{B,T+h|T} \\ \hat{y}_{C,T+h|T} \\ \hat{y}_{AA,T+h|T} \\ \hat{y}_{AB,T+h|T} \\ \hat{y}_{AC,T+h|T} \\ \hat{y}_{BA,T+h|T} \\ \hat{y}_{BB,T+h|T} \\ \hat{y}_{BC,T+h|T} \\ \hat{y}_{CA,T+h|T} \end{bmatrix} \quad (10)$$

O que nos dá uma proporção do total para cada elemento no nível mais desagregado.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \mathbf{S} \begin{bmatrix} p_1 \hat{y}_{T+h|T} \\ p_2 \hat{y}_{T+h|T} \\ p_3 \hat{y}_{T+h|T} \\ p_4 \hat{y}_{T+h|T} \\ p_5 \hat{y}_{T+h|T} \\ p_6 \hat{y}_{T+h|T} \\ p_7 \hat{y}_{T+h|T} \end{bmatrix} \quad (11)$$

Substituindo a matriz \mathbf{S} , temos as equações que definem cada previsão da estrutura em função de proporções da previsão do agregado.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1 \hat{y}_{T+h|T} \\ p_2 \hat{y}_{T+h|T} \\ p_3 \hat{y}_{T+h|T} \\ p_4 \hat{y}_{T+h|T} \\ p_5 \hat{y}_{T+h|T} \\ p_6 \hat{y}_{T+h|T} \\ p_7 \hat{y}_{T+h|T} \end{bmatrix} \quad (12)$$

Já a abordagem bottom-up parte do raciocínio inverso e define as previsões de cada elemento da estrutura a partir das previsões dos elementos mais desagregados. Para tanto, basta modificar a matriz \mathbf{G} .

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

O que resulta nas equações desejadas. Portanto, \mathbf{G} define a abordagem — se *top-down* ou *bottom-up* —, e \mathbf{S} define a maneira da qual as previsões são somadas para formar as equações de previsão para cada elemento da estrutura. Portanto, chamo \mathbf{G} de matriz de reconciliação.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{y}_{AA,T+h|T} \\ \hat{y}_{AB,T+h|T} \\ \hat{y}_{AC,T+h|T} \\ \hat{y}_{BA,T+h|T} \\ \hat{y}_{BB,T+h|T} \\ \hat{y}_{BC,T+h|T} \\ \hat{y}_{CA,T+h|T} \end{bmatrix} \quad (14)$$

1.3 Coerência e reconciliação

Seja somando as previsões do nível mais desagregado para formar os níveis superiores da hierarquia (*bottom-up*) ou distribuindo proporcionalmente as previsões do nível mais agregado (*top-down*), o vetor $\tilde{\mathbf{y}}_t$ representa as previsões *coerentes*. Isso significa que as previsões são totalizadas corretamente — as previsões de cada elemento agregado corresponde ao somatório das previsões dos níveis inferiores da hierarquia. Isso é garantido pela multiplicação das matrizes $\mathbf{S}\mathbf{G}$.

Não fosse essa pré multiplicação, nada garantiria a coerência das previsões. Tomando a estrutura da Figura 1 como exemplo, seria um acaso improvável que as previsões do agregado para o estado do Espírito Santo sejam exatamente a soma das previsões individuais de seus municípios. Isso porque não há qualquer razão para que cada série siga o mesmo processo (e.g., arima) com coeficientes idênticos.

Os métodos de gerar previsões coerentes a partir de previsões base são chamados de métodos de *reconciliação*. Os métodos de reconciliação tradicionais apresentados, *top-down* e *bottom-up*, utilizam informação limitada. No método *top-down*, utiliza-se apenas informações do nível mais agregado — por isso, apenas a primeira coluna em (8) é diferente de zero. Já na abordagem *bottom-up*, utiliza-se apenas as informações dos níveis mais desagregados, o que resulta na submatriz identidade $m \times m$ em (13), enquanto as colunas que representam os níveis mais agregados são nulas.

Alternativamente, podemos pensar numa matriz \mathbf{G} qualquer que utilize toda a informação disponível e tenha algumas propriedades que garantam que as previsões coerentes tenham o menor erro o possível. Esse é o problema de pesquisa trabalhado na *reconciliação ótima*.

1.4 Motivação

Atualmente, os métodos analíticos, especificamente o MinT, são os mais populares na literatura da reconciliação ótima. Entretanto, tais métodos são sujeitos a uma série de restrições, como as do MCLR, e têm sua capacidade preditiva reduzida quando suas hipóteses são violadas.

Em previsões de séries temporais, o objetivo na maioria dos casos é prever valores futuros com a maior acurácia possível. Em vista disso, métodos de *machine learning* são mais gerais, no sentido de permitir parâmetros não lineares e poderem aproximar virtualmente qualquer função. Além disso, são focados na capacidade preditiva, muitas vezes em detrimento da explicativa. Espera-se, portanto, que esses métodos alcancem melhor performance no problema da reconciliação ótima, devendo receber mais atenção

1.5 Objetivos

O objetivo geral da dissertação é estudar o problema da reconciliação ótima de previsões pontuais a partir de métodos de *machine learning*.

Como objetivos específicos, tenho:

1. Estudar métodos para estimação da matriz de reconciliação aplicando algoritmos e fluxos de trabalho de *machine learning*, como *tuning* e *resampling*;
2. Identificar possíveis vantagens e limitações da abordagem por *machine learning* na reconciliação de previsões pontuais a partir de aplicação do método estudado na previsão de saldos de crédito do Banestes.

1.6 Revisão da literatura

A primeira etapa da revisão de literatura consistiu na pesquisa bibliográfica relacionada à reconciliação ótima de previsões de séries temporais hierárquicas e agrupadas e sua interseção com o tema *machine learning*.

A pesquisa bibliométrica na base de dados do Google Acadêmico, pesquisando pelas palavras-chave “*hierarchical forecast reconciliation*” para qualquer lugar no corpo do texto, encontrando 27.600 resultados. Utilizando um script para ordenar os resultados pelo número de citações¹, verifiquei que o trabalho mais citado é Hyndman e Athanasopoulos (2021).

Tabela 1 – Trabalhos mais citados com os termos “*hierarchical forecast reconciliation*”

Autor	Citações	Ano
Hyndman, G Athanasopoulos	5222	2018
Dellarocas, X Zhang	2082	2007
Hyndman, A Lee, E Wang, S Wickramasuriya	1023	2013
Badre, M D’esposito	974	2009
Hong, S Fan	912	2016
Tashman	798	2000

Utilizando essa obra como texto base, obtive os textos referenciados no capítulo 11 “*Forecasting Hierarchical and Grouped Time-Series*”, subcapítulo 3 “*Forecast Reconciliation*”, além de Hyndman, Lee e Wang (2016), onde o método por MQP foi desenvolvido porém não está citado nas referências do capítulo.

Adicionando o termo “*machine learning*” e refinando a pesquisa para encontrar as palavras chave no título dos trabalhos, 10 resultados foram encontrados.

Previsões pontuais de séries temporais hierárquicas não é um assunto novo. Ao menos desde a década de 70, pesquisas foram publicadas acerca de abordagens *bottom-up* e *top-down*, suas vantagens e desvantagens, e tentativas de se definir qual é o melhor método². Entretanto,

¹ <https://github.com/WittmannF/sort-google-scholar>

² Uma revisão dessa literatura pode ser encontrada em Athanasopoulos, Ahmed e Hyndman (2009).

Quadro 1 – Artigos de referência em Hyndman e Athanasopoulos (2021)

Autor	Ano
Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., Shang, H. L.	2011
Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., Hyndman, R. J.	2021
Wickramasuriya, S. L., Athanasopoulos, G., Hyndman, R. J.	2019
Rob J. Hyndman and Alan J. Lee and Earo Wang	2016

Quadro 2 – Trabalhos encontrados na busca estendida

Autor	Ano	Citacoes
Li, SMM Rahman, R Vega, B Dong	2016	114
Mancuso, V Piccialli, AM Sudoso	2021	17
Abolghasemi, RJ Hyndman, G Tarr	2019	13
Afolabi, SU Guan, KL Man, PWH Wong, X Zhao	2017	20
Saatloo, A Moradzadeh, H Moayyed	2021	6
Abolghasemi, G Tarr, C Bergmeir	2022	0
C Neto, BL Fernando...	2022	0
Moon	2012	0
YAN, C SHENG	2018	1
Varone, C Ieracitano, A Özyüksel, T Hussain...	2022	0

é apenas em [Hyndman, Ahmed et al. \(2011\)](#) que é formalizada uma abordagem prática que utiliza toda a informação disponível, (i.e. as previsões de todos elementos de todos os níveis da hierarquia) a partir da estimação da matriz G via regressão linear por mínimos quadrados generalizados (MQG).

Entretanto, para ser capaz de estimar o modelo por MQG, é necessária a matriz de variância-covariância dos erros. [Hyndman, Ahmed et al. \(2011\)](#) usam a matriz de erros de coerência, ou seja, a diferença entre as previsões reconciliadas e as previsões base, que tem posto incompleto e não identificada e, portanto, não pode ser estimada. Os autores contornam esse problema adotando no lugar da matriz de variância-covariância dos erros uma matriz diagonal constante, ou seja, assumem variância constante dos erros de reconciliação, e estimam a matriz G por mínimos quadrados ordinários (MQO).

A estimação por esse método resulta numa reconciliação ótima que depende apenas da matriz S , ou seja, da estrutura hierárquica, e independe da variância e covariância das previsões base \hat{y}_{T+h} — o que não é uma conclusão satisfatória.

[Hyndman, Lee e Wang \(2016\)](#) tentam aperfeiçoar o método usando as variâncias das previsões base estimadas (dentro da amostra) como estimativa para a matriz de variância-covariância dos erros de reconciliação, de forma a as utilizar como pesos e realizar a reconciliação ótima por mínimos quadrados ponderados (MQP). Assim, previsões base mais acuradas têm peso maior do que as mais ruidosas. Entretanto, não fornecem justificativa teórica para usar a diagonal da matriz de variância-covariância de \hat{e}_t .

[Wickramasuriya, Athanasopoulos e Hyndman \(2019\)](#) argumentam que o que de fato interessa é que as previsões reconciliadas tenham o menor erro. Então, corrigem a abordagem de

reconciliação ótima para o objetivo de minimização dos erros das previsões reconciliadas \tilde{y}_{t+h} , ao invés dos erros das previsões base \hat{y}_{t+h} . Dado que isso implica na minimização da variância de \tilde{e}_{t+h} , ou seja, na minimização do somatório da diagonal, o traço, da matriz de variância-covariância de \tilde{e}_{t+h} , eles chamaram esse método de Traço Mínimo (MinT, na sigla em inglês). Paralelamente, usam desigualdade triangular para demonstrar que as previsões reconciliadas obtidas por esse método são ao menos tão boas quanto as previsões base.

[Panagiotelis et al. \(2021\)](#) reinterpreta a literatura de coerência e reconciliação de previsões pontuais a partir de uma abordagem geométrica, trazendo provas alternativas para conclusões anteriores ao mesmo tempo em que fornece novos teoremas. Além disso, [Panagiotelis et al. \(2021\)](#) estende essa interpretação geométrica para o contexto probabilístico, fornecendo métodos paramétricos e não paramétricos (via *bootstrapping*) para reconciliação de previsões probabilísticas, ou seja, para reconciliar previsões \hat{y}_t obtidas a partir de toda a distribuição, e não apenas a média.

2 MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA

Neste capítulo...

2.1 O processo de ajuste e sobreajuste

Considere uma função de ajuste f , um conjunto de pontos $D = d_1, \dots, d_n$ com $d_i = (x_i y_i)'$, variáveis de decisão ou parâmetros $x_i \in \mathbb{R}^m$ e imagem $y_i = f(x_i) \in \mathbb{R}$. Diferentemente da abordagem clássica, em que, no caso do modelo clássico de regressão linear, há um modelo teórico de coeficientes estimados por mínimos quadrados ordinários (MQO) que é garantido pelo teorema de Gauss-Markov ser o melhor estimador linear não viesado (BLUE), em *machine learning* o objetivo é encontrar, de forma iterativa, um meta-modelo que melhor aproxima a função f usando a informação contida em D , ou seja, queremos ajustar uma função de regressão \hat{f}_D aos nossos dados D de forma que $\hat{y} = \hat{f}_D(x, \varepsilon)$ tenha o menor erro de aproximação ε .

Para verificar o quão bem o modelo \hat{f}_D se aproxima da função real f , é necessário uma função de perda $L(y, \hat{f}(x))$ que, no caso de regressão, será a perda quadrática $(y - \hat{f}(y))^2$ ou a perda absoluta $|y - \hat{f}(y)|$. Esses valores são agregados pela média para formar as funções de custo erro médio quadrático (MSE) e erro médio absoluto (MAE).

Dada a função de perda, pode-se definir o risco associado ao modelo de função de ajuste

$$R(f, p) = \int_{\mathbb{R}} \int_{\mathbb{R}^m} L(y, f(x)) p(x, y) dx dy$$

em que $p(x, y)$ é a função densidade de probabilidade conjunta. Como não temos a função real mas procuramos uma função estimada que se aproxime dela, temos

$$GE(\hat{f}_D, p) = \int_{\mathbb{R}} \int_{\mathbb{R}^m} L(y, \hat{f}_D(x)) p(x, y) dx dy \quad (1)$$

que é o erro de generalização ou risco condicional associado ao preditor.

Então, podemos estimar o erro de generalização do modelo. Como não conhecemos a distribuição P , a substituímos pela amostra de teste D^* e ficamos com

$$\widehat{GE}(\hat{f}_D, D^*) = \sum_{(xy)' \in D^*} \frac{L(y, \hat{f}_D(x))}{|D^*|}$$

Se substituirmos a amostra teste pela amostra treino D usada para ajustar o modelo, teremos o chamado erro de resubstituição

$$\widehat{GE}_{\text{resub}} = \widehat{GE}(\hat{f}_D, D)$$

Naturalmente, nesse caso estaríamos usando os dados de treino tanto para treinar o preditor quanto para estimar o erro de generalização, o que nos levaria a uma estimativa enviesada do erro de generalização. Caso usássemos essa estimativa para seleção de modelos, esse viés favoreceria modelos mais adaptados à amostra.

O problema é que nesses modelos, dadas suficientes iterações, o erro de resubstituição tende a zero. Isso acontece porque conforme o preditor se adapta cada vez mais aos dados de treinamento ele irá memorizar a relação entre o conjunto de pontos D e a imagem $f(x_i)$, ou seja, irá se ajustar perfeitamente ao formato da função a ser modelada. E não necessariamente um modelo perfeitamente ajustado se traduz na capacidade de predição de dados futuros (fora da amostra).

De forma geral, espera-se que o preditor reduza seu viés durante o treino apenas o suficiente para que seja capaz de generalizar sua predição para fora da amostra em um nível ótimo de acurácia. A partir desse ponto, a redução no viés é penalizada com o aumento da variância, ou seja, com a redução de sua capacidade de prever dados futuros. A esse processo se dá o nome de *overfitting* ou **sobreajuste**. Isso quer dizer que não podemos considerar a performance do preditor em D se desejamos estimar honestamente a performance real do modelo.

2.2 Reamostragem

Uma forma de se corrigir esse problema é dividindo a amostra em um conjunto para treino D_{treino} e outro conjunto para teste D_{teste} de forma que $D_{\text{treino}} \cup D_{\text{teste}} = D$ e $D_{\text{treino}} \cap D_{\text{teste}} = \emptyset$. Assim, pode-se treinar o modelo em D_{treino} para se obter $\hat{f}_{D_{\text{treino}}}$ e calcular seu erro

de generalização usando os dados de D_{teste} . Essa abordagem é chamada de *hold-out* e ela é de simples implementação e utilização, uma vez que as observações do conjunto teste são completamente independentes das observações com as quais o modelo foi treinado. A estimativa do erro de generalização então se torna

$$\widehat{GE}_{\text{hold-out}} = \widehat{GE}(\hat{f}_{D_{\text{treino}}}, D_{\text{teste}})$$

Dois problemas permanecem:

1. É necessária uma amostra grande, uma vez que deve-se ter dados suficientes tanto na amostra treino para ajustar um modelo adequado, quanto na amostra teste para realizar uma avaliação de performance estatisticamente válida.
2. Esse método não é suficiente para detectar variância e instabilidades na amostra treino. Modelos mais complexos, especialmente não lineares, podem produzir resultados muito diferentes com mudanças pequenas nos dados de treino.

É exatamente para lidar com essas situações que foram desenvolvidas as técnicas de reamostragem. Todas essas técnicas geram repetidamente i subconjuntos de treino $D_{\text{treino}}^{(i)}$ e teste $D_{\text{teste}}^{(i)}$ com o dataset disponível, ajustam um modelo com cada conjunto de treino e atestam sua qualidade no conjunto de teste correspondente. A estimativa do erro de generalização então se torna

$$\widehat{GE}_{\text{samp}} = \frac{1}{k} \sum_{i=1}^k \widehat{GE}(\hat{f}_{D_{\text{treino}}^{(i)}}, D_{\text{teste}}^{(i)}) \quad (2)$$

O erro de generalização dado na equação (1) depende tanto do tamanho da amostra usada para treinar quanto para testar o modelo ajustado. Portanto, devemos garantir que o tamanho da amostra usado para verificar o erro de generalização de um modelo estimado a partir de n data points seja próximo de n . Se, por exemplo, o conjunto de treino for muito menor que a amostra total o erro será superestimado, uma vez que muito menos informação foi usada para calcular o estimador.

Da mesma forma, a qualidade do estimador do erro de generalização obtido em (2) a partir de uma estratégia de reamostragem também depende muito do tamanho dos conjuntos $D^{(i)}$ em relação à amostra original, da quantidade k de subconjuntos utilizados e da estrutura de dependência entre os subconjuntos $D^{(i)}$ — novamente, modelos mais complexos são mais sensíveis a alterações no dataset e a variância entre os subconjuntos tende a ser maior. O erro do estimador é geralmente medido pelo erro médio quadrático (MSE):

$$\text{MSE}(\widehat{GE}_{\text{samp}}) = \mathbb{E}[(\widehat{GE}_{\text{samp}} - GE(\hat{f}_D, P))^2]$$

Esse estimador também pode ser representado como a soma do quadrado do viés e a variância:

$$\text{MSE}(\widehat{GE}_{\text{samp}}) = \text{Bias}(\widehat{GE}_{\text{samp}})^2 + \text{Variância}(\widehat{GE}_{\text{samp}})$$

Sendo que o viés expressa a diferença média entre um estimador e o valor real, enquanto a variância mede a dispersão média do estimador. Essas quantidades são definidas da seguinte forma:

$$\text{Bias}(\widehat{GE}_{\text{samp}}) = \mathbb{E}[\widehat{GE}_{\text{samp}}] - \mathbb{E}[GE(\hat{f}_D, p)]$$

e

$$\text{Variância}(\widehat{GE}_{\text{samp}}) = \mathbb{E}[(\widehat{GE}_{\text{samp}} - \mathbb{E}[\widehat{GE}_{\text{samp}}])^2]$$

REFERÊNCIAS

- ATHANASOPOULOS, George; AHMED, Roman A.; HYNDMAN, Rob J. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, v. 25, n. 1, p. 146–166, 2009. ISSN 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2008.07.004>. Citado na p. 19.
- HYNDMAN, R.J.; AHMED, R.A. et al. Optimal Combination Forecasts for Hierarchical Time Series. *Computational Statistics and Data Analysis*, v. 55, p. 2579–2589, 2011. Citado na p. 20.
- HYNDMAN, R.J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. 3. ed. Melbourne, Australia, 2021. Acessado em 14/09/20. Disponível em: <<https://otexts.com/fpp3/>>. Citado nas pp. 12, 15, 19.
- HYNDMAN, Rob J.; LEE, Alan J.; WANG, Earo. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and Data Analysis*, v. 97, p. 16–32, 2016. ISSN 0167-9473. DOI: <https://doi.org/10.1016/j.csda.2015.11.007>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016794731500290X>>. Citado nas pp. 19, 20.
- PANAGIOTELIS, Anastasios et al. Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, v. 37, n. 1, p. 343–359, 2021. ISSN 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2020.06.004>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169207020300911>>. Citado na p. 21.
- WICKRAMASURIYA, Shanika L.; ATHANASOPOULOS, George; HYNDMAN, Rob J. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association*, Taylor & Francis, v. 114, n. 526, p. 804–819, 2019. DOI: [10.1080/01621459.2018.1448825](https://doi.org/10.1080/01621459.2018.1448825). eprint: <https://doi.org/10.1080/01621459.2018.1448825>. Disponível em: <<https://doi.org/10.1080/01621459.2018.1448825>>. Citado na p. 20.

ÍNDICE

coerência, 18

matriz S , 14

matriz de reconciliação, 17

reconciliação, 18