

Construção de Tábuas de Mortalidade com o uso de Redes Neurais LSTM Bidirecional para Predição das Probabilidades de Morte

Construction of Mortality Tables using Bidirectional LSTM Neural Networks to Predict Death Probabilities

José Douglas Nascimento, Tatiana Escovedo, Marcos Kalinowski

Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, RJ – Brasil

domdoug2012@gmail.com, {tatiana, kalinowski}@inf.puc-rio.br

Abstract. *Mortality Tables are tables structured with mortality data, especially mortality rates observed at all ages, used in pension funds and life insurance markets. This article concerns the application of the neural network model to the construction of future mortality tables, using the Lee-Carter model for comparison. The proposed model was a LSTM (Long-Short Term Memory) Neural Network model, including a bidirectional variation. This network is characterized by the sequential processing of data over time. The data for the prediction came from historical mortality table data prepared by the IBGE (Brazilian Institute of Geography and Statistics) and the Human Mortality Database. The results point to a reasonable use as an auxiliary tool for predicting death probabilities.*

Keywords. *Mortality Table, Pension Funds, Life Expectancy, Forecast, Lee-Carter Model, Neural Network, LSTM.*

Resumo. *As Tábuas de Mortalidade são tabelas estruturadas contendo dados epidemiológicos traduzidos em probabilidades de morte associada a cada idade de vida, utilizadas no mercado de previdência e seguros. Este artigo discorre sobre a aplicação do modelo de redes neurais para a construção de tábuas de mortalidade futuras, tendo como comparação o modelo Lee-Carter. O modelo proposto foi a Rede LSTM (Long-Short Term Memory), sendo implementada também uma variação bidirecional. Esta rede tem como característica o processamento sequencial de dados ao longo do tempo. Os dados para a predição são oriundos de dados históricos de tábuas de mortalidade elaboradas pelo IBGE (Instituto Brasileiro de Geografia e Estatística) e do The Human Mortality Database. Os resultados apontam para uma utilização razoável como ferramenta auxiliar de predição das probabilidades de morte.*

Palavras-Chave. *Tábua de Mortalidade, Previdência, Expectativa de Vida, Previsão, Modelo Lee-Carter, Redes Neurais, LSTM.*

1. Introdução

As tábuas completas de mortalidade do IBGE (Instituto Brasileiro de Geografia e Estatística) são divulgadas desde 1999, por força do Decreto nº 3.266, de 29/11/1999¹. Esta divulgação ocorre até o dia 1º de dezembro de cada ano, com base no total da população brasileira estimada para o dia 1º de julho do ano anterior. Estas projeções têm como base os dados do último Censo Demográfico. Para este estudo, serão utilizados os dados para construção de tábuas *unisex*.

A utilização de tábuas de mortalidade é uma das principais ferramentas para a estimativa das reservas de fundo de entidades de previdência privada, abertas ou fechadas. A área de seguros do ramo vida também utiliza esta ferramenta, agregada nos cálculos de renda. Antes do Decreto 3.266, o Brasil utilizava tábuas de mortalidade de outros países, notadamente dos Estados Unidos. Como exemplos, as tábuas AT49, AT2000, AT83 (AT: *Annuity Table*), CSO58 (*Commissioner's Standard Ordinary*), cuja numeração corresponde, aproximadamente, ao ano quando foram apurados os dados. Estas e outras tábuas de mortalidade, em sua grande maioria, originadas nos EUA, podem ser obtidas por meio de um software denominado *Table Manager*² ou podem ser obtidas em consulta ao site <https://mort.soa.org>, certificado pelo SOA (*Society of Actuaries*).

Embora as tábuas de outros países, notadamente dos EUA, ainda sejam utilizada no mercado brasileiro de previdência privada e principalmente do mercado segurador, o fator motivador para a construção da tábua de mortalidade brasileira foi pela força da legislação para calcular o fator previdenciário, criado em 1999, a ser aplicado como um redutor no cálculo dos benefícios de aposentadoria, com o intuito de equilibrar o custo de despesas do sistema de Previdência Social, além da questão de que não tínhamos tábuas que medissem o perfil biométrico da população brasileira. A Figura 1 apresenta gráficos da pirâmide etária de 2010 e 2060.

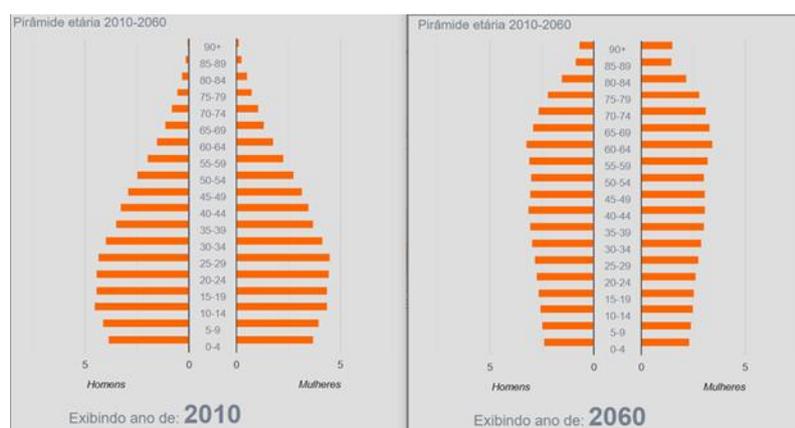


Figura 1. Pirâmide etária de 2010 e 2060. Fonte: <https://www.ibge.gov.br>.

¹ “Art. 2º. Compete ao IBGE publicar, anualmente, até o dia primeiro de dezembro, no Diário Oficial da União, a tábua completa de mortalidade para o total da população brasileira referente ao ano anterior.”

² Projeto inicial da Society of Actuaries. Mantido em: <https://www.blufftop.com/products/TableMgr.htm>

Construir tábuas de mortalidade mais próximas da tendência do perfil biométrico da população permite uma melhor estimativa de benefícios de renda ou expectativa de vida para o cálculo de seguros e planos de saúde. O ideal seria ter a pesquisa censitária sempre atualizada anualmente, porém isto é impossível, dado o alto custo desta operação.

Com base nas exposições iniciais, em [Nascimento e Escovedo 2021] fizemos uma estimativa de novas tábuas de mortalidade ou sobrevivência a partir da aplicação de uma rede neural treinada para fazer uma predição de novos valores com base nas tábuas do IBGE de 1998 a 2018. Nele, foi investigada a aplicação do modelo de redes neurais LSTM para estimar tábuas futuras com base no histórico de mortalidade no Brasil, elaboradas pelo IBGE e foi concluído ser um modelo razoável como ferramenta auxiliar na mensuração de tábuas futuras. O presente artigo estende o trabalho anterior com a aplicação do modelo de redes neurais LSTM bidirecionais que é uma variação do modelo LSTM. Também foi realizada implementação no algoritmo para novas saídas estatísticas. Por fim, foi realizada uma análise comparativa adicional com a utilização de uma base mais robusta, com o intuito de verificar se o volume de dados da base do IBGE utilizada no artigo anterior foi satisfatória para o *framework* de estimativa do modelo.

O restante deste artigo está organizado da seguinte forma. A Seção 2 aborda conceitos relacionados à modelagem de tábuas atuariais. Na Seção 3, apresentamos um modelo consolidado e utilizado nas estimativas dos elementos das tábuas atuariais e que servirá como um baseline para comparar como o modelo proposto neste artigo. A Seção 4 apresenta os conceitos das redes neurais LSTM e traz alguns trabalhos relacionados. Na Seção 5 é apresentado o modelo proposto e são detalhados os resultados dos experimentos realizados. Finalmente, a Seção 6, conclui este artigo.

2. Modelagem de Tábuas Atuariais

Esta Seção apresenta uma revisão da literatura para a modelagem de tábuas atuariais, com as definições e cálculos dos elementos que são fundamentais na construção das respectivas tábuas, além de uma técnica de extrapolação dos valores desses elementos para as idades acima de 80 anos, dado que é a última idade constante nos arquivos de dados.

2.1. Conceito de Tábua de Mortalidade

De acordo com Cordeiro Filho [Cordeiro Filho 2017], as tábuas de mortalidade são instrumentos estatísticos destinados a medir as probabilidades de vida e de morte e pessoas em uma coorte para cada idade e para estas idades são apresentadas as quantidades estimadas de falecimentos e, a taxa de mortalidade específica, a probabilidade de falecimentos, a probabilidade de sobrevivência e a esperança de vida.

Segundo Bowers *et al.* [Bowers *et al.* 1986], o conceito teórico de construção de tábuas de mortalidade leva em consideração os seguintes fatores: (a) a massa populacional observada, no início, consiste de l_x vidas na idade 0; (b) a massa populacional é exposta, para cada idade, a taxa anual de mortalidade expressa pelos valores de q_x , na tábua de mortalidade; e (c) a massa populacional é fechada, sem novos entrados. O único decremento corresponde ao resultado da aplicação da taxa anual de mortalidade.

Em outras palavras, a massa populacional inicializa com um radix arbitrário, $l_0 = 10^6$ diminuindo ano a ano, considerando somente o efeito da taxa de mortalidade, excluindo-se efeitos de imigração e emigração na massa populacional observada até a extinção da massa, na idade final, denominada ω (ômega).

De forma prática, segundo Souza [Souza 2007], a construção de tábuas atuariais é feita pelo “método indireto” utilizando um “grupo aberto”, ou seja, um grupo com novos entrados e saídas de pessoas e que por meio de técnicas estatísticas são feitas as correções nas contagens apuradas.

Segundo Rodrigues [Rodrigues 2008], os elementos l_x e q_x , definem o que conhecemos como função de sobrevivência em uma tábua de mortalidade. Algumas funções foram elaboradas para satisfazer a modelagem da curva de sobrevivência. Ainda segundo Rodrigues [Rodrigues 2008], sob a curva determinada pela função de sobrevivência, determinamos “todas as inferências para a obtenção de cálculos biométricos”. Em outras palavras, a partir da curva, obtemos todas os elementos básicos constante na tábua de mortalidade.

A notação matemática e seus respectivos símbolos dos elementos de uma tábua foram definidos em [London e Layton 1899] no II Congresso Internacional de Atuária, em Londres³.

2.2. Definição e Formulação dos Elementos da Tábua de Mortalidade

Os principais elementos constantes nas tábuas de mortalidade são os seguintes:

q_x : probabilidade de uma pessoa de idade x da população geral falecer antes de alcançar a idade $x+I$. A métrica é obtida por meio da modelagem dos óbitos no país. A partir da probabilidade de morte, os outros elementos são calculados:

l_x : número de indivíduos vivos para a idade x na tábua de sobrevivência, onde a raiz inicial é $l_0 = 10^6$. Sua fórmula é dada por: $l_x = l_{x-1} - d_{x-1}$.

d_x : corresponde ao número de indivíduos mortos entre as idades x e $x+I$ na tábua de mortalidade. Sua fórmula é dada por: $d_x = q_x * l_x$.

L_x : número de vivos no meio do ano. Para centralizarmos o número de vivos em uma determinada idade, basta utilizarmos a média aritmética simples. Sua fórmula é dada por:

$$L_x = \frac{l_x + l_{x+1}}{2}.$$

T_x : corresponde aos anos vividos pelos indivíduos da massa populacional desde a idade x até a total extinção do grupo. Segundo Rodrigues [Rodrigues 2008], este elemento corresponde a área sob a curva $s(x)$, curva de sobrevivência, ilustrada na Figura 2. Sua fórmula é dada por: $T_x = L_x + L_{x+1} + L_{x+2} + \dots + L_\omega$.

onde ω é a última idade da tábua.

e_x^0 : corresponde a expectativa de vida (ou esperança de vida, em média, que se espera que vivam os componentes da massa populacional. Sua fórmula é dada por: $e_x^0 = \frac{T_x}{l_x}$.

³ Em <https://archive.org/details/transactionsofse00interich>, p. 618-624, trecho em inglês (acessado em 20/09/2020). O documento consta de 931 páginas escrito em alemão, inglês e francês.

Graficamente, os elementos l_x , d_x e T_x são identificados na curva de sobrevivência, ilustrada pela Figura 2. A camada hachurada representa o elemento L_x , como parte da área sob a curva. Este elemento também pode ser interpretado como a área do trapézio sob a curva ou a integral sob a curva [Bowers *et al.* 1986].

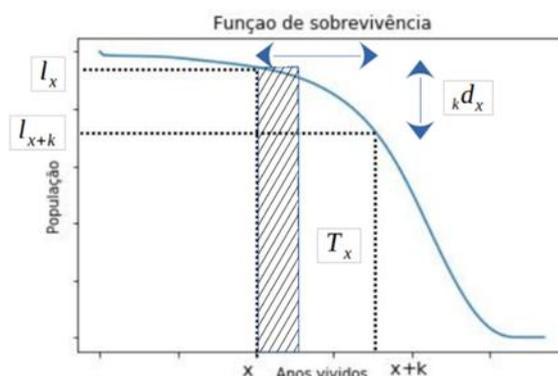


Figura 2. Função de sobrevivência com L_x . Adaptada de [Bowers *et al.* 1986].

Estes elementos básicos da tábua de mortalidade são os insumos dos elementos da tábua completa de comutação para sobrevivência (D_x , N_x e S_x) e morte (C_x , M_x e R_x). Estes elementos são a base para o cálculo da Matemática Atuarial para Seguros e Previdência, que não serão aqui apresentados por estarem fora do escopo do artigo.

2.3. Tábua de Mortalidade do IBGE

A tábua de mortalidade do IBGE, anualmente divulgada, apresenta a expectativa de vida às idades exatas até os 80 anos. É sabido que a população já vive bem mais do que esta idade. A tábua de mortalidade, ou tábua de vida, elaborada pelo IBGE constitui um modelo demográfico que descreve a incidência da mortalidade ao longo do ciclo vital das pessoas, entretanto, faremos uma extrapolação destas idades até o limite ω anos (fim da vida do último sobrevivente) para todas as tabelas divulgadas pelo IBGE, desde 1999.

As tábuas de mortalidade ou sobrevivência do IBGE para o cálculo da estimativa de vida são representativas da mortalidade do país. São discriminadas por sexo ou com a agregação de ambos os sexos. Os dados utilizados para a modelagem dessas tábuas são os óbitos por causas naturais, violentas ou por mortalidade infantil.

O grande problema que temos ao utilizar uma tábua de um determinado ano e assim continuar com ela é que teremos problemas para os cálculos de longo prazo, característica principal dos modelos atuariais, seja para estimativa de benefícios de renda ou expectativa de vida para o cálculo de seguros e planos de saúde. Dado que as tábuas são estáticas, ela não captará o aumento da probabilidade de sobrevivência com o passar dos anos, posto que é fato que a população está vivendo mais, devido a evolução dos tratamentos de saúde e qualidade de vida de uma forma geral.

Como se constatou na evolução do escopo deste artigo, as tábuas de mortalidade ou sobrevivência do IBGE, para o cálculo da expectativa de vida, são apresentadas com as probabilidades de morte da idade 0 até 80 anos, porém, sabe-se que já é realidade que as pessoas já sobrevivem além da idade de 80 anos. Assim, neste artigo, estimamos o

prolongamento das probabilidades de morte além dos 80 anos de idade. Consideramos as tábuas de 1998 a 2018, pois a tábua anterior a 1998 (no caso, a de 1991) ainda era muito experimental e descolada da exigência do Decreto nº 3.266, de 29/11/1999.

2.4. Extrapolação da Tábua de Mortalidade do IBGE para idades acima de 80 anos

Para a extrapolação dessas tábuas, usamos o método descrito em Castro [Castro 2014] na “Nota Técnica sobre a Metodologia adotada pelo Ministério da Previdência Social na Extrapolação das Tábuas de Mortalidade IBGE para as idades acima de 80 anos”, que utiliza um estimador para probabilidades de morte acima de 80 anos de idade. Utilizamos, inicialmente, todas as tabelas por sexo e por ambos os sexos para apresentar se existe diferenças entre as expectativas de vida ao nascer entre homens e mulheres. A base consta em 21 planilhas disponíveis no site do IBGE abrangendo o período de 1998 a 2018⁴ e as 3 principais características (*features*) são número de indivíduos vivos (l_x), probabilidade de morte (q_x) e expectativa de vida ao nascer por idade (e_x^0).

Analisando os dados importados, observou-se que a expectativa de vida ao nascer do sexo feminino é substancialmente maior que a expectativa de vida ao nascer do sexo masculino. Não obstante, fizemos todos os testes com base nos arquivos unissex, denominado “ambos”.

A base para a construção da interpolação para as idades até o fim dos últimos indivíduos vivos foi encontrar os fatores de ajuste (FA) aplicado na *feature* l_x (número de indivíduos vivos), com a exigência que a expectativa de vida na idade de 80 anos seja a mesma que na tábua antes da interpolação. A partir deste fator estimado, as demais *features* foram calculadas com base na *feature* l_x .

Os resultados das principais *features*, após a interpolação podem ser vistos nas Figuras 3 e 4, somente para as tábuas unissex.



Figura 3. Probabilidade de morte (q_x).

⁴ <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=73097>



Figura 4. Expectativa de vida ao nascer por idade (e_x^0).

As tabelas interpoladas foram os insumos para a construção do modelo de rede neural como proposta de estimativa de tábuas futuras. Entretanto, para efeitos comparativos, apresentamos na próxima seção o modelo Lee-Carter, utilizado para projeção de tábuas atuariais.

3. Modelo de Projeção Lee-Carter

O modelo de projeção Lee-Carter foi detalhado no artigo “*Modelling and Forecasting U.S. Mortality*”, publicado no “*Journal of the American Statistical Association*”, em 1992. No citado artigo, Lee e Carter [Lee e Carter 1992] apresenta uma técnica de projeção de mortalidade com base nos dados históricos demográficos dos Estados Unidos, mais precisamente as taxas de mortalidade desde 1933 até 1987. Este modelo pode ser representado matematicamente pela seguinte fórmula: $\ln(m_{x,t}) = a_x + b_x * k_t + \varepsilon_{x,t}$, que tem como restrições: $\sum b_x = 1$; e $\sum k_x = 0$;

Onde:

- $m_{x,t}$: Corresponde a Taxa central de mortalidade para cada idade x , definida pela fórmula: $m_x = \frac{d_x}{L_x}$.
- a_x : Nível médio das taxas de mortalidade no tempo para cada idade x ;
- b_x : Padrão médio da mortalidade à idade x quando o nível geral da mortalidade em função de k_x ;
- k_t : Índice de tendência temporal do nível de mortalidade geral;
- $\varepsilon_{x,t}$: Erro aleatório do tipo ruído branco, inserido no modelo para cada idade x no tempo t .

3.1. Estimadores do Modelo

São dois os estimadores que podem ser utilizados com o modelo de projeção de Lee-Carter. Os estimadores são dados pelas fórmulas definidas a seguir.

Fórmula 1: $\hat{a}_x = \frac{1}{h} \sum_{t=t_1}^{t_n} \ln(m_{x,t})$, sendo: $h = t_n - t_1$.

Fórmula 2: $\beta_x * k_t = \ln(m_{x,t}) - \alpha_x + \epsilon_{x,t}$

Utilizamos a SVD (*Singular Value Decomposition*) para a estimação dos parâmetros b_x e k_t e ARIMA, para previsão das taxas de mortalidade $m_{x,t}$.

A Fórmula 2, na forma matricial Z, onde foi aplicada a SVD, correspondeu, para as idades de x a 116 e anos de 1998 a 2018:

$$Z = \begin{bmatrix} \ln(m_{0,1998}) - \hat{\alpha}_0 & \cdots & \ln(m_{0,2018}) - \hat{\alpha}_0 \\ \vdots & \ddots & \vdots \\ \ln(m_{116,1998}) - \hat{\alpha}_{116} & \cdots & \ln(m_{116,2018}) - \hat{\alpha}_{116} \end{bmatrix}$$

Conforme se observa na Fórmula 1, Lee e Carter [Lee e Carter 1992] utilizou como variável de medição a taxa central de mortalidade m_x entretanto foi utilizada a fórmula do q_x , que está intrinsecamente relacionada ao m_x .

Segundo Bowers *et al.* [Bowers *et al.* 1986], a *relação* entre q_x e m_x , sob a distribuição uniforme, pode ser definida por: $q_{x,t} = \frac{m_{x,t}}{(1-0,5*m_{x,t})} \approx m_{x,t}$, ou ainda

$q_{x,t} = 1 - e^{-m_{x,t}} \approx m_{x,t}$. Esta relação pode ser ilustrada graficamente pela Figura 5.

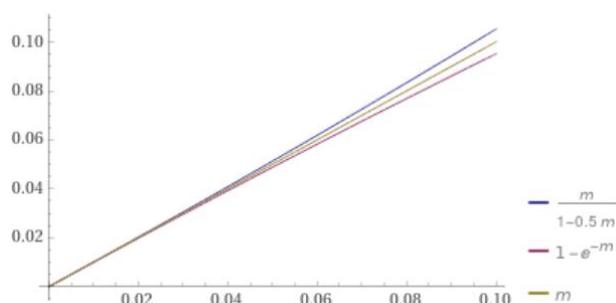


Figura 5. Gráfico q_x e m_x . Fonte: Utilizado recurso gráfico do site <https://www.wolframalpha.com>⁵

Com a aplicação deste método, encontramos os valores dos estimadores a_x , b_x e k_t , conforme Figura 6. Em atenção aos gráficos, os estimadores a_x , b_x estão delimitados pela idade da coorte, de 0 a ω (fim de cada tábua), enquanto o parâmetro k_t é o componente temporal. Neste componente temporal, k_t , foi aplicada a técnica ARIMA, para projeção.

Antes da aplicação do ARIMA nas tábuas de mortalidade do IBGE, foi utilizado um método para verificar a estacionariedade do componente temporal. Os resultados demonstraram que o componente temporal não é estacionário.

Também foi verificado que não existe sazonalidade ruído branco (resíduo) no componente a ser medido. Após estas verificações, foram feitas simulações para encontrar o modelo ARIMA que apresenta o menor MSE (*Mean Squared Error*). O melhor ARIMA foi (0,1,0) com o MSE = 0.524.

⁵ Sintaxe do código: `plot m/(1-0.5*m), 1-exp(1)**(-m), m, (m, 0, 0.1)`

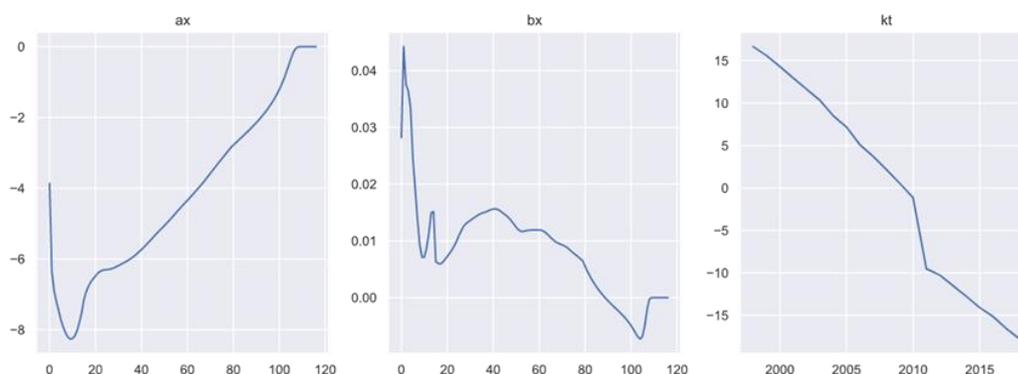


Figura 6. Estimadores calculados a_x , b_x e k_t .

Também foi utilizada outra medida, chamada AIC (*Akaike Information Criterion*) [Pascariu *et al.* 1992], e novamente, o modelo ARIMA que apresentou o melhor resultado, ou menor AIC, calculado no valor de 78.451, foi o ARIMA (0,1,0). O valor de k_t foi modelado e estimado por este ARIMA, resultando no gráfico contido na Figura 7.

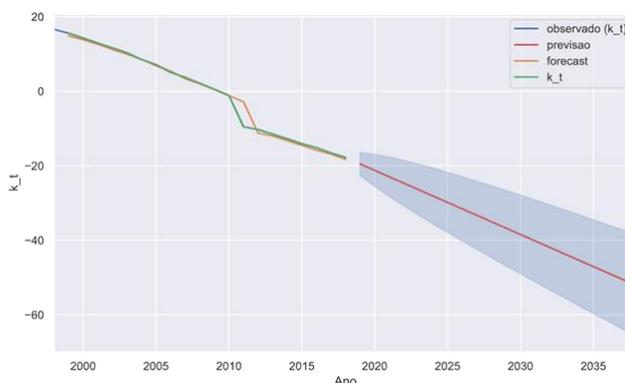


Figura 7. Valores previstos e estimados de k_t .

4. Redes Long-Short Term Memory (LSTM)

Segundo Escovedo e Koshiyama [Escovedo e Koshiyama 2020], a inteligência artificial tende a simular o comportamento da mente humana, com a utilização de máquinas. Machine Learning pode ser considerada uma subárea da inteligência artificial e consiste na descoberta de padrões ou fórmulas matemáticas e estatísticas que expliquem o relacionamento entre os dados e tendem a resolver problemas na forma do pensamento humano. O principal objetivo do modelo construído é generalizar a partir da sua experiência [Bishop 2006].

Segundo Goodfellow *et al.* [Goodfellow *et al.* 2016], no escopo das técnicas de *Machine Learning*, um tipo específico conhecido são as redes neurais, tratadas também, segundo Zaccone *et al.* [Zaccone *et al.* 2018], como *Deep Learning*. As redes neurais

consistem em camadas interconectadas por nós. Cada camada possui um modelo matemático com funções e calibrações entre o modelo observado e o estimado.

Nesta seção, apresentamos uma arquitetura de rede neural que permite modelar o comportamento desta probabilidade de morte ao longo do tempo. Conforme Chollet [Chollet 2018], a rede neural denominada *Recurrent Neural Network* (RNN) é um tipo de rede neural que é desenhada para processamento temporal e reconhecer as sequências dos dados utilizados como input da mesma para reconhecimento, reprodução ou predição. Entretanto, conforme o mesmo autor, esta arquitetura de rede apresenta um problema denominado *vanishing gradient*. Nas redes neurais que utilizam métodos de aprendizado com base em gradiente, durante cada iteração de treinamento, os pesos da rede neural são atualizados de forma proporcional à derivada parcial da função de erro em relação ao peso. Em alguns casos, o gradiente será muito pequeno, fazendo com que o peso não seja efetivamente alterado e, conseqüentemente, podendo impedir a rede neural de aprender. A principal ação deste efeito atua na mitigação dos sinais de calibragem dos neurônios das camadas ocultas, tornando os sinais fracos, dificultando a manutenção da memória no processo de aprendizado da rede.

Um caso particular da rede RNN que atenua o efeito do *vanishing gradient* é a rede *Long-Short Term Memory* (LSTM), que traz subestruturas, denominadas células de memória, que ajudam a gerenciar a memória da RNN e, conseqüentemente, melhorar o aprendizado. Exatamente por conta deste requisito de manter a memória do processo é que esta rede pode ser viável para processos que tenham como objetivo, estimar o próximo valor nos tempos subsequentes, $t+1$, $t+2$, etc. Portanto, a arquitetura de rede neural que foi utilizada para estimar as probabilidades de morte foi a LSTM.

Conforme Goodfellow *et al.* [Goodfellow *et al.* 2016], como exemplo prático de aplicação desta arquitetura de rede neural temos a Modelagem de Linguagem, Tradução de Idiomas, Legendas em Imagens, Geração de Texto. Além dessas aplicações, conforme Zaccane *et al.* [Zaccane *et al.* 2018] esta arquitetura é ideal para predições de sequencias temporais.

A estrutura básica de uma célula LSTM contém os seguintes portões (*gates*) com as suas definições de forma resumida:

- *Input gate*: informação de entrada, por exemplo, a idade x_0 , no tempo t , $t+1$, $t+2$;
- *Forget gate*: portão de reset do estado corrente na célula;
- *Output gate*: informação que deverá ou não passar para a próxima camada da arquitetura LSTM.

A Figura 8 ilustra uma representação gráfica simplificada de uma célula LSTM. A lógica da estrutura de redes neurais LSTM, com esses *gates*, é regular o fluxo da informação, o armazenamento (retenção), o *forget* (descarte) de informações irrelevantes no histórico, atualizando a informação para a saída final.

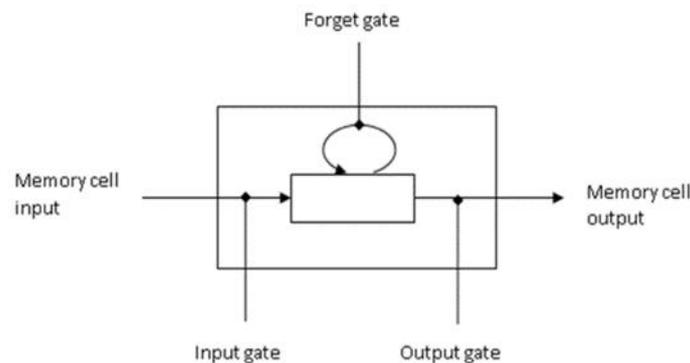


Figura 8. Diagrama simplificado de uma célula LSTM [Zaccone *et al.* 2018].

4.1. Trabalhos Relacionados

Devido à sua capacidade de modelar e prever a dinâmica de sistemas não lineares variantes no tempo, as LSTMs têm sido muito usadas nos últimos anos para previsão de séries temporais, com diversos graus de sucesso [Lindemann *et al.* 2021]. Especificamente para o problema de construção de tabelas de mortalidade, foram encontradas na literatura algumas abordagens recentes baseadas em *deep learning*.

Por exemplo, Petneházi e Gáll [Petneházi e Gáll 2019] treinaram modelos LSTM simples com dados de diferentes países, sexos e idades e compararam o seu desempenho de previsão da taxa de mortalidade com o modelo de Lee-Carter, e verificaram que o LSTM produziu previsões mais precisas. Entretanto, os autores utilizaram apenas um único horizonte de previsão de 10 anos e indicaram a necessidade de obter mais dados para melhorar a abordagem utilizada e permitir a previsão décadas a frente.

Nigri *et al.* [Nigri *et al.* 2019], por sua vez, desenvolveram uma solução baseada em LSTM para integrar o modelo de Lee-Carter tradicional, com a finalidade de construir tábuas de mortalidade para seis países, objetivando construir um modelo com maior capacidade preditiva. A abordagem proposta apresentou desempenho preditivo, considerando a influência significativa da tendência de mortalidade passada.

Já Pirovano [Pirovano 2020] propôs duas abordagens baseadas em redes neurais para a extensão do modelo de Lee-Carter, também com o objetivo de construir tábuas de mortalidade. Os dois modelos propostos - um utilizando redes neurais do tipo *Gated Recurrent Unit* (GRU) e o outro utilizando LSTM - obtiveram melhores resultados do que o modelo de Lee-Carter tradicional.

Ainda utilizando um modelo LSTM para estender o modelo de Lee-Carter, Lindholm e Palmborg [Lindholm e Palmborg 2021] focaram em diferentes procedimentos para utilizar o conjunto de treinamento de forma mais eficiente para estabilizar o desempenho preditivo e melhorar a estabilidade das previsões de longo prazo do modelo.

Também há na literatura abordagens que não utilizam LSTM, especificamente. Por exemplo, o trabalho de Perla *et al.* [Perla *et al.* 2021] propôs a generalização do modelo de Lee-Carter através de uma rede neural convolucional simples. Este modelo

obteve bons resultados no *The Human Mortality Database* (HMD) e no Banco de Dados de Mortalidade dos Estados Unidos.

5. Solução Proposta e Experimentos Realizados

A solução proposta neste artigo para o problema de construção de tábuas de mortalidade futuras é a utilização de uma rede neural LSTM, cujos conceitos teóricos já foram apresentados na Seção 4. Esta arquitetura de rede foi escolhida para este problema porque ela tem como característica o processamento sequencial de dados ao longo do tempo, sendo especialmente adequada para a previsão de séries temporais, como mostram diversos trabalhos correlatos na literatura, alguns apresentados na Subseção 4.1. Adicionalmente ao modelo baseado na rede neural LSTM, foi implementada neste artigo também uma variação bidirecional.

A arquitetura da rede foi submetida às *features* de entrada como dados sequenciais. No caso, os dados sequenciais foram as idades x em cada tempo t . Em outros termos, a entrada foi por idade em cada ano, iniciando com a idade 0, nos períodos de 1998 a 2018, gerando a série de idade = 0. Para a idade 1, foi gerada a série de idades =1, entre 1998 e 2018 e assim por diante, até a última idade. Cabe salientar que foi feita uma transformação de q_x para a escala logarítmica para uma a melhor visualização. O objetivo é simular um aprendizado da Rede LSTM com base no q_x , estruturado em uma série temporal. A plataforma para a execução desta rede foi o *framework Tensor Flow e Keras*.

Nas subseções a seguir, serão apresentados os resultados de dois experimentos realizados, ambos com o objetivo de comparar o tradicional método Lee-Carter com os modelos propostos neste artigo, a Rede Neural LSTM e a Rede Neural LSTM Bidirecional. O primeiro experimento (Subseção 5.1) utiliza como referencial as tábuas de mortalidade projetadas pelo IBGE. Já o segundo experimento (Subseção 5.2) utiliza os dados do *The Human Mortality Database*. Em ambos os experimentos, foram considerados os seguintes parâmetros:

- $t_{projecao}$: número de anos de projeção (*forecast*);
- n_{epochs} : número de vezes que o vetor q_x passará pela rede, relacionado com o aprendizado da rede;
- $n_{neurons}$: número de neurônios na camada de oculta;
- *dropout*: método de regularização;
- n_{test} : tamanho da base de teste.

Resumidamente, os resultados têm como saída:

- As probabilidades de morte por idade de 0 até 120 anos, no espaço temporal de 2019 a 2028;
- A expectativa de vida ao nascer no espaço temporal de 2019 a 2028.

Para mostrar o efeito das variações dos hiperparâmetros nos resultados de saída, selecionamos os gráficos no ano de 2028, que é o ano mais distante do ano de 2018, no qual temos a última informação disponível nos dados tratados e em que podemos observar as possíveis distorções que possam ocorrer nas estimativas. Foram realizadas

diversas simulações variando os hiperparâmetros, mas considerando sempre o valor do parâmetro $t_{projecao}$ igual a 10 anos, pois o *framework Keras* apresenta erro com valores maiores, dado a quantidade limitada dos dados disponíveis (total de 20 anos de informação), e considerando n_{test} igual a 3, representando os últimos 3 anos. Para o hiperparâmetro n_{epochs} , foram utilizados os valores 500 e 1000, uma vez que valores mais baixos não resultaram em uma boa addropouterência do modelo em estudos prévios realizados. Para o hiperparâmetro $n_{neurons}$, foram utilizados os valores 50 e 100 e para o hiperparâmetro $dropout$, os valores 0% e 20%.

O código-fonte escrito pelos autores na linguagem Python está disponível no *Github*, em: https://github.com/Domdoug/LSTM_Mortality_Table_Actuarial.

5.1. Experimento 1: Dados do IBGE

O objetivo deste experimento é comparar os valores projetados pelo Método Lee-Carter e pelas redes Neurais LSTM e LSTM Bidirecional com os dados projetados pelo IBGE. Para tal, nos modelos LSTM, foram variados os hiperparâmetros de número de neurônios ($n_{neurons}$), número de camadas ocultas (*hidden layers*), épocas (n_{epochs}) e *dropout*, resultando em cinco simulações distintas, em comparação com os valores iniciais utilizados na LSTM de referência: $n_{epochs} = 500$, $n_{neurons} = 50$ e $dropout = 20\%$. Nas figuras desta subseção que ilustrarão os resultados obtidos, serão sempre apresentadas as curvas referentes aos resultados do Método Lee-Carter (identificados como “Lee-Carter” na legenda), dos dados projetados pelo IBGE (identificados como “IBGE” na legenda) e da LSTM de referência (identificados como “LSTM-500” na legenda). A Tabela 1 apresenta o resumo das simulações realizadas, que serão detalhadas a seguir.

Tabela 1. Tabela resumo das simulações

Sim.	Tipo de Rede Neural	Num. neurônios	Camadas ocultas (<i>hidden layers</i>)	<i>epochs</i>	<i>Dropout</i>	Tempo de processamento ⁶
1	LSTM	50	2	1.000	20%	04h:30m:36.02s
2	LSTM	50	1	500	20%	01h:35m:43.05s
3	LSTM	50	2	500	0%	01h:28m:03.50s
4	LSTM	100	2	500	20%	02h:27m:13.41s
5	LSTM Bidirecional	50	2	1.000	20%	05h:36m:44.18s

Cabe observar que no site do IBGE consta um arquivo com a projeção das Tábuas de Mortalidade 2010/2060, com as *features* q_x as resultantes atuariais, explicitadas na Seção 2.2, projetadas para o período de 2010 até 2060⁷. As projeções estimadas das populações do Brasil e das Unidades da Federação foram calculadas pelo IBGE⁸, conforme método das componentes demográficas. Cabe ressaltar que essas estimativas foram utilizadas com base nos dados do último censo oficial, de 2010.

⁶ Computador i7, com 4 núcleos e 8 processadores e 8 Gb de memória RAM

⁷<https://www.ibge.gov.br/novo-portal-destaques/27470-projecao-da-populacao-das-unidades-da-federacao-por-sexo-e-idade-simples.html>

<https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?=&t=resultados>

⁸<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101597.pdf>

<https://biblioteca.ibge.gov.br/visualizacao/livros/liv86444.pdf>

<https://biblioteca.ibge.gov.br/visualizacao/livros/liv41229.pdf>

Observa-se, tomando como exemplo a tábua estimada e a tábua divulgada no ano de 2018 (conforme Tabela 2) que a diferença percentual entre a probabilidade de morte (q_x) estimada e a efetivamente divulgada por força de Lei possuem diferenças significativas (em torno de 82%). Entretanto, a diferença percentual entre a expectativa de vida (E_x) estimada e a efetivamente divulgada possui variação de no máximo 0,1423%. Esta diferença é refletida nos gráficos que envolvam a probabilidade de morte, no caso, $\log q_x$, estas foram apresentadas para manter uma uniformidade comparativa quando se apresenta os gráficos da Expectativa de vida.

Por fim, a expectativa de vida é o resultado divulgado pela mídia e é objeto de análises quanto a um cenário de longevidade da população, porém a probabilidade de morte é a informação primária que é utilizada pelos atuários para estimar custos de modelos previdenciários e de seguros de vida. Cabe salientar que a base utilizada para as projeções foram os dados efetivamente divulgados no período de 1998 a 2018.

Tabela 2. Tábua estimada de 2018 com base no censo de 2010

Tábua Estimada de 2018 com base no censo 2010									Tábua divulgada - Expectativa de Vida em 2018	Diferença % (Ex Divulgada - Ex Estimada)	Tábua divulgada - Prob de Morte (qx) em 2018	Diferença % (qx Divulgada - qx Estimada)
Idade	nM_x	nq_x	nD_x	nPx	l_x	nL_x	T_x	e_x				
0	0,01250	0,01235	1.235	0,98765	100.000	98.872	7.625.151	76,25	76,30515582	0,0704%	0,012358261	0,0291%
1	0,00053	0,00211	208	0,99789	98.765	394.539	7.526.279	76,20	76,25892261	0,0717%	0,000841266	-60,0485%
5	0,00024	0,00121	119	0,99879	98.557	492.484	7.131.740	72,36	72,41731217	0,0766%	0,000285130	-76,4616%
10	0,00029	0,00146	144	0,99854	98.437	491.826	6.639.255	67,45	67,50144266	0,0813%	0,000217382	-85,1505%
15	0,00099	0,00492	484	0,99508	98.293	490.256	6.147.430	62,54	62,59528759	0,0855%	0,000695017	-85,8722%
20	0,00144	0,00717	701	0,99283	97.810	487.295	5.657.173	57,84	57,89216198	0,0925%	0,001320089	-81,5872%
25	0,00150	0,00749	728	0,99251	97.108	483.722	5.169.879	53,24	53,29200475	0,1009%	0,001484503	-80,1919%
30	0,00170	0,00846	815	0,99154	96.381	479.865	4.686.157	48,62	48,67424705	0,1087%	0,001594431	-81,1471%
35	0,00208	0,01034	988	0,98966	95.565	475.356	4.206.292	44,01	44,06788602	0,1206%	0,001884038	-81,7854%
40	0,00280	0,01390	1.315	0,98610	94.577	469.597	3.730.936	39,45	39,50145488	0,1337%	0,002435133	-82,4846%
45	0,00406	0,02008	1.873	0,97992	93.262	461.628	3.261.339	34,97	35,02125224	0,1476%	0,003464418	-82,7498%
50	0,00588	0,02896	2.647	0,97104	91.389	450.328	2.799.711	30,64	30,68529107	0,1639%	0,005037931	-82,6049%
55	0,00844	0,04133	3.667	0,95867	88.742	434.543	2.349.383	26,47	26,52230736	0,1816%	0,007247187	-82,4631%
60	0,01215	0,05895	5.015	0,94105	85.075	412.836	1.914.840	22,51	22,55248779	0,1990%	0,010337476	-82,4649%
65	0,01820	0,08703	6.968	0,91297	80.060	382.879	1.502.004	18,76	18,80076141	0,2114%	0,015176346	-82,5620%
70	0,02817	0,13158	9.617	0,86842	73.092	341.417	1.119.125	15,31	15,34206219	0,2015%	0,023208512	-82,3612%
75	0,04368	0,19690	12.498	0,80310	63.475	286.128	777.709	12,25	12,26969485	0,1423%	0,035858095	-81,7889%

Simulação 1

Nesta simulação, foi utilizado $epochs = 1.000$. Observamos, na Figura 9, quanto maior o número de iterações ou ciclos para treinar a rede, melhor a estimativa, comparativamente como o modelo Lee-Carter. Entretanto, conforme Figura 10, que representa a expectativa de vida ao nascer, a simulação com 1.000 $epochs$ diverge dos outros resultados, porém aponta para uma estabilidade que, em tese, é lógica, na manutenção da expectativa de vida. Os outros modelos apontam um crescimento indefinido.

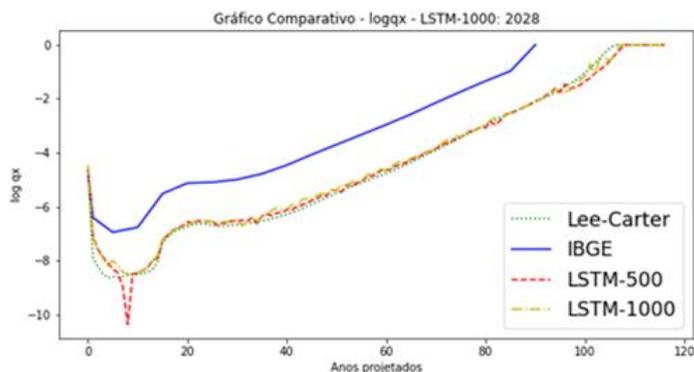


Figura 9. Gráfico do $\log qx$ projetado para epochs = 1000.

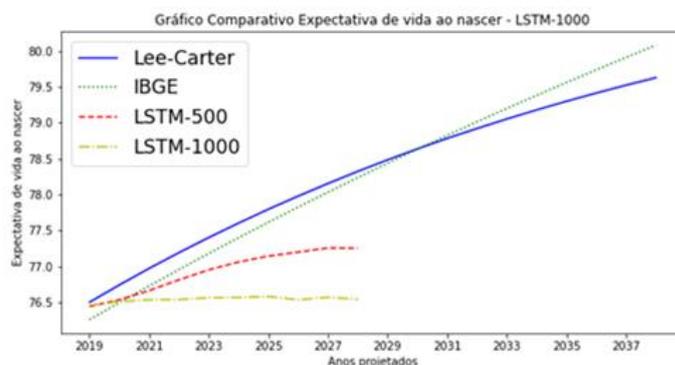


Figura 10. Gráfico da expectativa de vida para epochs = 1000.

Simulação 2

Nesta simulação, foi utilizada uma camada oculta, com a função de ativação ReLU (*Rectified Linear Unit*).

A Figura 11 ilustrou um ajuste em boa parte da curva, comparado com a LSTM-500, entretanto, no trecho crítico de mínimo, na região onde tem-se as idades iniciais, existe uma diferença acentuada. Esta região, em outras palavras, é a faixa onde a mortalidade infantil é relevante. Já a Figura 12 apontou para uma expectativa de vida crescente, o que não foi observado em nenhuma das curvas.

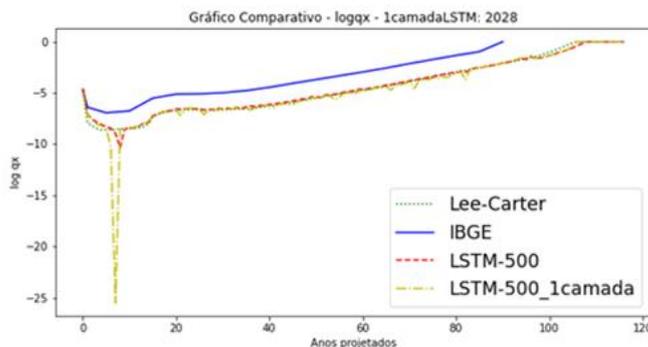


Figura 11. Gráfico do $\log qx$ projetado com ReLU e uma camada.



Figura 12. Gráfico da expectativa de vida com ReLU e uma camada.

Simulação 3

Nesta simulação, verificou-se o que ocorre com a exclusão da camada de *dropout*. Conforme Zaccone *et al.* [Zaccone *et al.* 2018] o *dropout*, é uma técnica de otimização utilizada para reduzir o *overfitting* dentro de uma rede neural.

Conforme ilustra a Figura 13, com a aplicação do *dropout* de 20% (LSTM-500), a curva ficou mais próxima da estimada pelo método Lee-Carter. Sem *dropout*, ocorreram alguns distúrbios ao longo da curva, apontando que a rede neural ficou muito sensível sem a aplicação da camada de *dropout*. A Figura 14, também, mostra este descolamento, comparado com a LSTM-500.

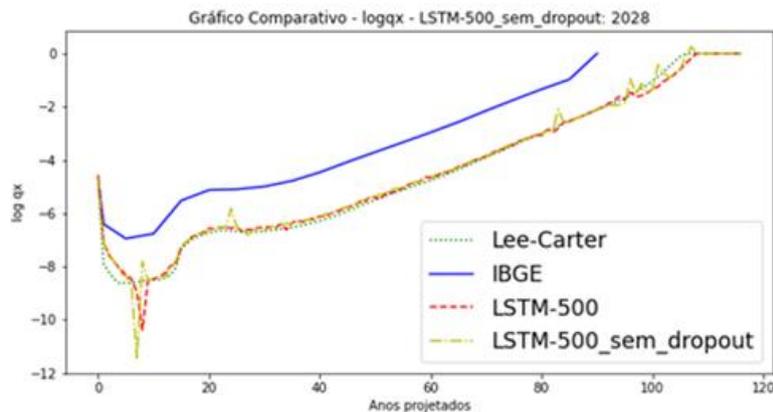


Figura 13. Gráfico do $\log qx$ projetado sem *dropout*.

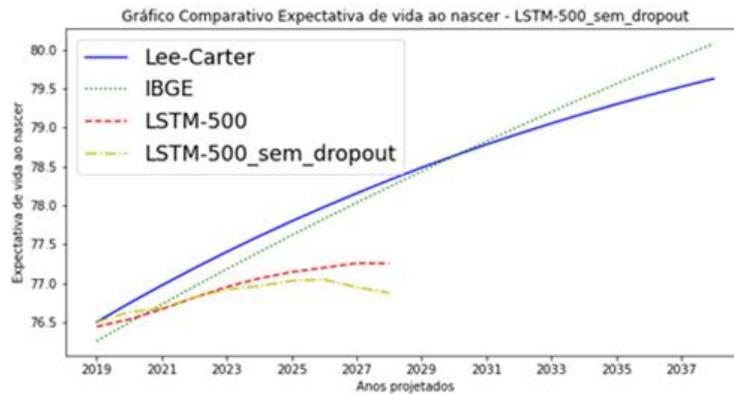


Figura 14. Gráfico da expectativa de vida sem *dropout*.

Simulação 4

Nesta simulação, foram acrescentados mais neurônios nas camadas ocultas. A curva LSTM com 100 neurônios ficou bem mais próxima da modelagem pelo método Lee-Carter, porém esta implementação, juntamente com o número de iterações (*epochs*) maior, aumentou o recurso de processamento da máquina. Os resultados foram ilustrados pelas Figuras 15 e 16.

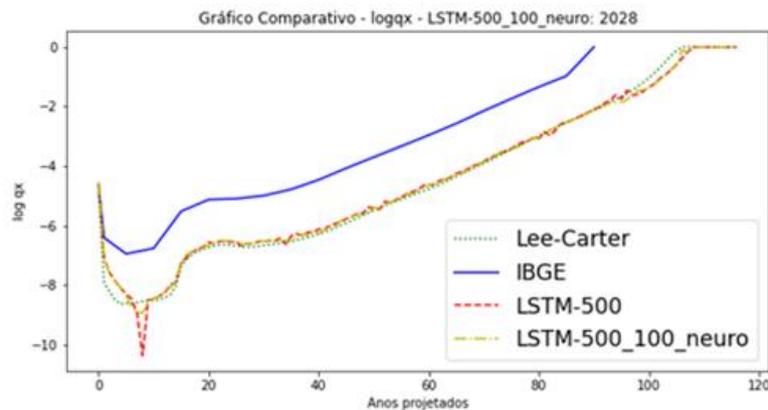


Figura 15. Gráfico do $\log qx$ projetado com 100 neurônios.

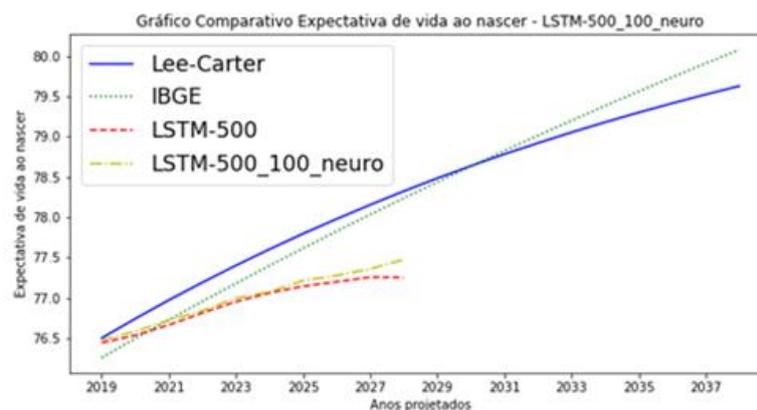


Figura 16. Gráfico da expectativa de vida com 100 neurônios.

Em resumo, para efeito comparativo, os gráficos foram construídos considerando a probabilidade de morte nas idades de 0 até 116 anos, para o ano de 2028, último possível para estimar. Os resultados apresentaram uma boa aderência. Além da probabilidade de morte, foi utilizada a expectativa de vida ao nascer. Para esta medida, de uma forma geral, para o ano de 2028, em todas as simulações, a diferença na expectativa de vida entre os resultados dos modelos LSTM e Lee-Carter ficou entre o mínimo de 0,7, na simulação 4 e o máximo de 1,7 anos, na simulação 1.

Com base nas simulações 1 a 4, os melhores parâmetros para o modelo LSTM que mais se aproximaram da modelagem pelo método Lee-Carter são *epochs* igual a 1000, pela tendência da expectativa de vida para uma estabilidade lógica, além de ser um ponto positivo na quantidade de vezes que se treinou a rede, porém exigindo maior poder de processamento computacional. A utilização da camada de *dropout* resultou em uma melhor estabilidade na curva, e a aplicação de 100 neurônios nas camadas ocultas tendeu a aproximar com a modelagem pelo método Lee-Carter.

Simulação 5

Nesta simulação foi utilizada uma modelagem similar a rede neural LSTM, porém com a passagem do vetor de mortalidade também em sentido contrário ao fluxo normal da rede LSTM. Esta rede é conhecida como rede LSTM Bidirecional [Brownlee 2020]. A Figura 17 ilustra uma arquitetura simplificada desta rede.

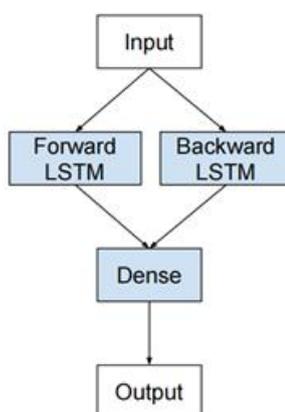


Figura 17. Arquitetura simplificada da Rede LSTM Bidirecional. Fonte: [Brownlee 2020]

O trabalho para implementá-lo foi simplificado, pois foi aproveitada a modelagem LSTM. A biblioteca utilizada foi a *Tf.keras.layers.Bidirectional*. Conforme Figura 18, não houve melhora significativa neste gráfico, que representa a mortalidade de $\log q_x$ projetado, em relação às demais simulações. Entretanto, conforme Figura 19, verificou-se uma melhoria na estimativa da expectativa de vida ao nascer, comparada aos modelos LSTM com 500 ou 1000 *epochs*.

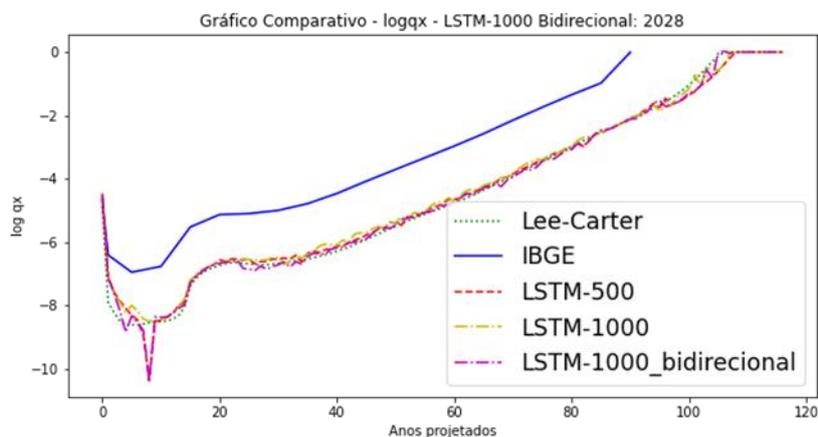


Figura 18. Gráfico do $\log qx$ projetado com 50 neurônios.

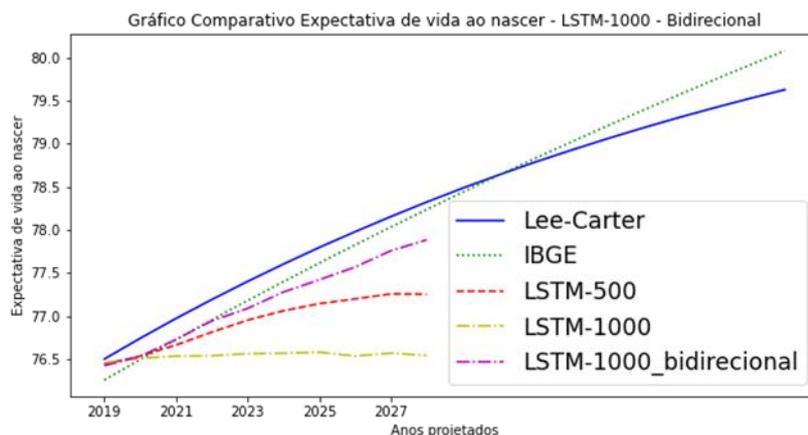


Figura 19. Gráfico da expectativa de vida com 50 neurônios.

Comparação das Simulações do Experimento 1

Durante a rotina de treinamento da base, foram obtidas medidas através do método *CSVLogger*, do pacote *keras.callbacks* da biblioteca do Tensor Flow. Do arquivo de *log* gerado para cada simulação, foi obtido o MAE (*Mean Absolute Error*) de cada *epoch* (500~1000) para cada idade (0 – 116) e calculada sua mediana e construído os respectivos gráficos. A métrica MAE é de fácil interpretação e é amplamente utilizada para a avaliação de modelos de *deep learning* e, por este motivo, foi utilizada como uma medida de erro de estimação do algoritmo.

Conforme Tabela 3, a mediana desses pontos foi menor na simulação 3, embora a variância dos dados seja menor nas simulações 1 e 4. Como são valores muito pequenos, não existe diferença fundamental entre as simulações. A Figura 20 ilustra o MAE obtido em cada uma das cinco simulações.

Tabela 3. Tabela resumo das simulações

Simulação: Base IBGE	Modelo: epoch - camada - dropout - neurônios	MAE mediana	MAE variância
1	LSTM 1000 - 2 - 20% - 50	0,037244	0,002484
2	LSTM 500 - 1 - 20% - 50	0,053395	0,004120
3	LSTM 500 - 2 - 0% - 50	0,030877	0,003135
4	LSTM 500 - 2 - 20% - 100	0,042178	0,002962
5	LSTM Bidirecional 1000 - 2 - 20% -50	0,044114	0,003294

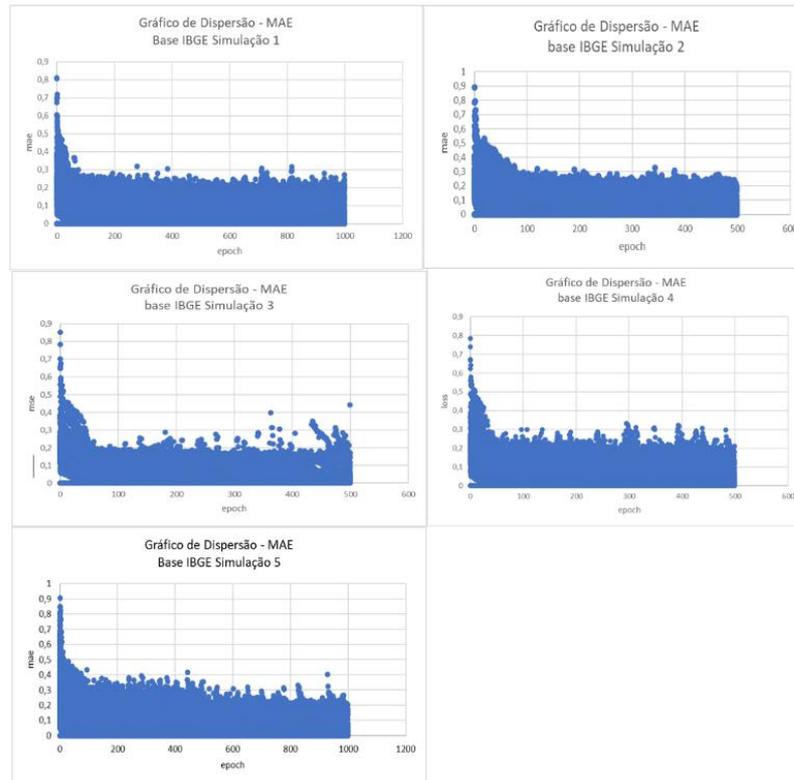


Figura 20. Gráficos de Dispersão das MAE das simulações.

Para comparar os resultados das simulações após a realização da etapa de treinamento, na fase de teste optou-se por usar a métrica RMSE (*Root Mean Squared Error*), métrica que ilustra a diferença entre os valores observados e previstos, mas elevados ao quadrado. Esta métrica também é amplamente utilizada para a avaliação de modelos de *deep learning* e possibilita uma melhor visualização comparativa. Os valores de RMSE foram calculados para cada idade e podem ser vistos na Figura 21. Conforme é possível observar, em todas as simulações os valores de RMSE ficaram em sua grande maioria próximos de zero, sem tantos pontos discrepantes.

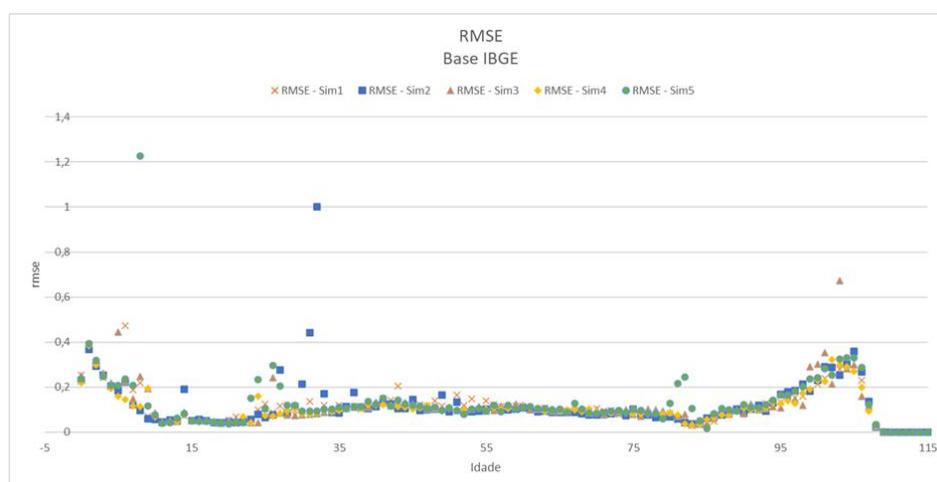


Figura 21. Gráficos de Dispersão das RSME das simulações.

5.2. Experimento 2: Dados do The Human Mortality Database

O experimento anterior tomou por base as tábuas de mortalidade do IBGE, no período de 1999 a 2018. Esta base resulta em uma matriz de 120 linhas (idades) por 20 colunas (anos entre 1999 até 2018), que correspondem ao período de 20 anos. Sabemos que os modelos propostos de redes neurais exigem uma grande quantidade de dados e sendo a definição do que seja, em termos absolutos, uma grande quantidade de dados, optou-se por utilizar outra base de dados para efeitos comparativos.

Para uma verificação das modelagens considerando uma base mais robusta, foram utilizados, como alternativa, os dados disponíveis em *The Human Mortality Database* (HMD) (www.mortality.org). Este projeto foi criado para fornecer informações para pesquisadores, estudantes e demais profissionais interessados com o intuito de realizar comparativos de taxas de mortalidade de diversos países. Para este artigo, foi utilizada a base dos Estados Unidos, que é uma das mais robustas em quantidade de informações e abrange o período de 1933 a 2019. Entretanto o processamento, comparativamente ao experimento 1, foi extremamente demorado, conforme ilustra a Tabela 4.

Tabela 4. Tabela resumo das simulações

Sim.	Modelo	Num. neurônios	Camadas ocultas (<i>hidden layers</i>)	<i>epochs</i>	<i>Dropout</i>	Tempo de processamento ⁹
1	LSTM	50	2	500	20%	17h:10m:14.56s
2	LSTM Bidirecional	50	2	500	20%	22h:32m:14.93s

Ainda conforme Tabela 3, foram feitas 2 simulações, mantendo os mesmos hiperparâmetros, porém alternando o modelo entre LSTM e LSTM Bidirecional. A Simulação 1 deste experimento tem os mesmos hiperparâmetros da Simulação 4 do experimento 1.

⁹ Computador i7, com 4 núcleos e 8 processadores e 8 Gb de memória RAM

Simulação 1 e 2

Como estas simulações envolvem mais dados, foi utilizada uma base de teste de 30 anos, diferentemente das simulações anteriores, que utilizaram uma base de teste contendo apenas 3 anos. Os resultados projetados também foram mais abrangentes, até o ano de 2050. Os outros modelos apontam um crescimento indefinido.

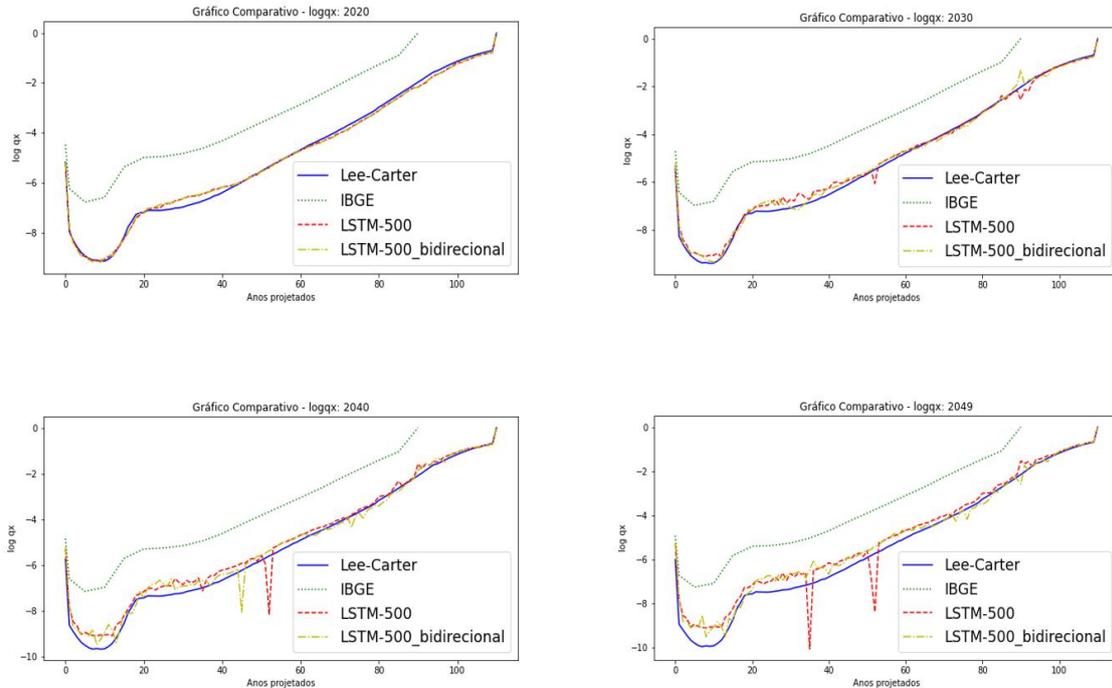


Figura 22. Gráficos do $\log qx$ projetados com 50 neurônios.

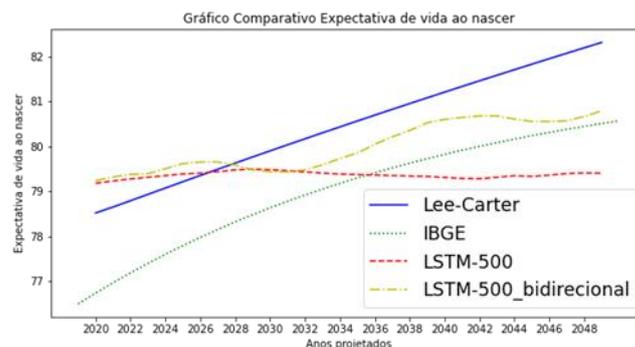


Figura 23. Gráfico da expectativa de vida com 50 neurônios.

Conforme Figura 22, a Rede LSTM 500 bidirecional se mostrou mais estável e comparável com o modelo *baseline*, Modelo Lee-Carter, diferente dos sobressaltos da rede LSTM 500. Na Figura 23, vemos que a rede LSTM sofreu forte distorção para a expectativa de vida ao nascer. A rede LSTM Bidirecional tendeu melhor para o modelo *baseline*. Talvez a utilização de uma rede com 100 neurônios poderia ter melhorado esta relação ainda mais, o que pode ser investigado em trabalhos futuros. A base do IBGE ficou bem deslocada porque é um outro perfil biométrico. O procedimento foi repetido para detectar se houve erro, mas foi obtido o mesmo resultado. Em trabalhos futuros

podem ser realizadas novas simulações alterando os hiperparâmetros da rede neural LSTM com o intuito de buscar uma melhor modelagem para estes dados.

Comparação das Simulações do Experimento 2

Assim como no experimento anterior, para avaliar o desempenho das redes LSTM e LSTM Bidirecional, foi calculado o MAE de ambos. O MAE representa a diferença entre o valor observado e o valor previsto, ou seja, buscamos o menor MAE possível para indicar um erro baixo. Vale observar que um MAE igual a zero poderia evidenciar a ocorrência de um *overfitting*, indicando que o modelo simplesmente “decorou” os padrões apresentados, não havendo a generalização para exemplos não vistos. Conforme Tabela 5, o modelo LSTM 500 Bidirecional tem uma mediana mais próxima de zero.

Tabela 5. Tabela medidas MAE e MSE das bases de treino e teste

Simulação: Base HMD	MAE mediana	MAE variância
LSTM 500	0,030256	0,0005562
LSTM 500 Bidirecional	0,028147	0,0005079

Conforme visto nas figuras 24 e 25, o modelo LSTM 500 Bidirecional convergiu um pouco mais que o modelo LSTM 500. Para uma simulação com mais neurônios e mais iterações(*epochs*), a MAE tenderá a convergir mais, entretanto a demora no processamento para a configuração que foi utilizada será relativamente considerável.

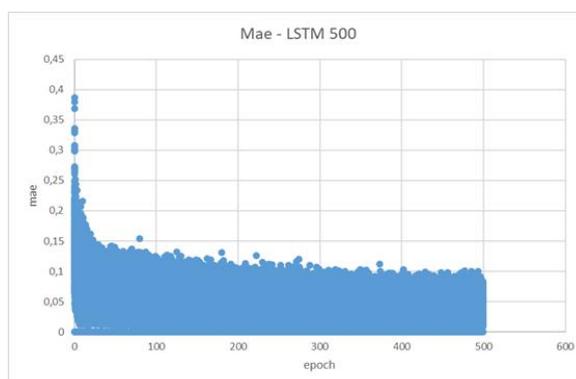


Figura 24. MAE para LSTM 500 com a base HMD.

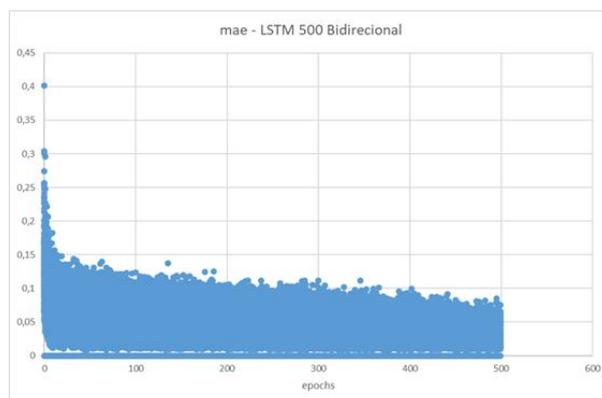


Figura 25. MAE para LSTM 500 Bidirecional com a base HMD.

Assim como no experimento anterior, os valores de RMSE foram calculados para cada idade. A distribuição dos resultados por idade foram as apresentadas na Figura 26. Conforme observou-se, embora a simulação LSTM Bidirecional tenha tido alguns pontos discrepantes, esta simulação foi a que mais flutuou próximo ao RMSE = 0.

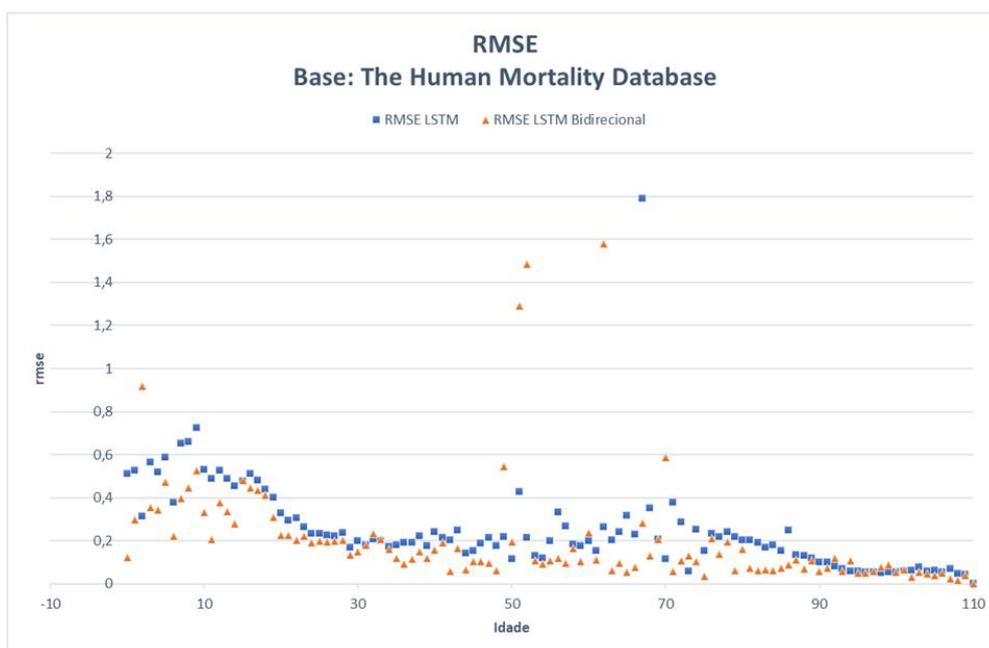


Figura 26. Gráficos de Dispersão das RSME das simulações com a base HMD.

6. Conclusão

O trabalho apresentado neste artigo, como uma extensão do artigo submetido ao SBSI 2021 [Nascimento e Escovedo 2021], que propôs a utilização de redes neurais LSTM, discorreu com a proposta da uma variação desta rede, denominada LSTM Bidirecional. Ambas as redes são um caso específico da rede RNN. Para tanto, foram utilizadas tábuas atuariais do IBGE, que anualmente são disponibilizada pelo órgão do Governo

Federal responsável pela parte previdenciária, com o intuito de auxiliar nos cálculos previdenciários, pelas entidades de previdência. Estas tábuas foram utilizadas como base primária para o modelo de base (o método Lee-Carter), para o modelo LSTM, assim como para o modelo adicional apresentado neste artigo, a rede neural LSTM Bidirecional.

O modelo com base na rede LSTM se apresentou razoável, como uma ferramenta auxiliar, para a predição de valores para a construção de tábuas de mortalidade. Foram utilizadas duas visões para verificar a aderência do modelo aos resultados futuros, a saber, a probabilidade de morte e a expectativa de vida ao nascer. Na ótica da probabilidade de morte, ao longo de todas as idades de uma tábua, o modelo LSTM teve, comparativamente ao modelo determinado pelo método Lee-Carter, boa aderência ao longo do tempo. Já na expectativa de vida ao nascer, a menor diferença observada foi alcançada na simulação que considerou mais neurônios aplicados na camada oculta. Também foi utilizada uma simulação com esta mesma base de dados para a rede LSTM Bidirecional. A diferença das simulações do artigo anterior, em confronto com o modelo LSTM Bidirecional, apresentado na simulação 5 foi que, em termos práticos, verificou-se uma melhoria na estimativa da expectativa de vida ao nascer, comparada aos modelos do simulado anterior. Já nas medidas como resultado comparativo da base de teste com a base submetida ao algoritmo de treino, não houve diferença substancial entre as simulações.

A obtenção de modelos que consigam realizar uma boa predição de valores para as tábuas de mortalidade de um país é um problema relevante para a sociedade porque realizar, de forma mais precisa, uma predição de morte ao longo do tempo é algo necessário em diversos cálculos e planejamentos, tanto do setor público quanto privado. Neste sentido, os resultados obtidos neste trabalho sugerem que os modelos propostos poderiam ser utilizados na prática, por exemplo, em entidades de previdência e de seguros de vida. Em especial, uma das contribuições mais importantes deste trabalho é a utilização de dados reais, que refletem a realidade brasileira.

Uma das principais limitações da aplicação da rede LSTM é o pequeno volume de dados. Para investigar a hipótese se a base de dados do IBGE possui volume de dados insuficiente para uma modelagem mais concisa, foi testada uma série mais robusta, como os dados utilizados no artigo *Modelling and Forecasting U.S. Mortality* na estimativa pelo método Lee-Carter, que considerou dados históricos demográficos de taxas de mortalidade dos Estados Unidos, desde 1933 até 1978, porém o presente artigo considerou a atualização até 2019.

A utilização de uma base mais robusta mostrou-se fundamental na melhoria das respostas das redes neurais quando comparada com o modelo baseline. Em relação aos gráficos de mortalidade geral e expectativa de vida ao nascer, a rede LSTM Bidirecional se mostrou mais precisa e com menos ruídos do que a rede LSTM, comparada com o baseline Lee-Carter. Em relação ao comparativo da base de treino e teste, o modelo LSTM Bidirecional apresentou as medidas estatísticas MAE e RSME bem melhores que o modelo LSTM.

O estudo carece de mais algumas simulações, principalmente nas alterações dos hiperparâmetros. Por exemplo, podem ser variados o número de camadas ocultas, com as respectivas camadas de *dropout*, assim como o aumento no número de neurônios. Ainda assim, este estudo tem grande relevância, pois permite uma alternativa de

mensuração do risco nas estimativas dos recursos previdenciários a serem geridos pelas companhias que atuam no mercado segurador.

Uma das limitações deste trabalho é o intervalo muito grande (10 anos) para o censo demográfico no Brasil. Este fator limita a acurácia dos cálculos, mas esbarra-se nos custos para se promover censo demográfico em um país com grande extensão territorial.

Um possível trabalho futuro interessante seria a aplicação de redes neurais LSTM na estimativa dos custos associados à massa de beneficiários de planos de saúde. Dada a longevidade do grupo objeto da modelagem, seria necessário formular um componente de entrada que representasse, além da probabilidade de morte, o valor pago pelo plano de saúde e o custo de utilização do plano, considerando uma grande massa de participantes.

Referências

- Bishop, C. M. (2006) “Pattern Recognition and Machine Learning”. Berlin: Springer.
- Bowers Jr., N.L., Gerber, H.U., Hickman, J.C., Jones, D., Nesbitt, C.J. (1986) “Actuarial Mathematics”. The Society of Actuaries.
- Brownlee, J. (2020) “Long Short-Term Memory Networks with Python”, Edition: v1.7, E-book.
- Castro, L. G. (2014) “Nota Técnica sobre a Metodologia adotada pelo Ministério da Previdência Social na Extrapolação das Tábuas de Mortalidade IBGE para as idades acima de 80 anos”. Disponível em: <http://sa.previdencia.gov.br/site/2015/06/NOTA-TECNICA-ATUARIAL-EXTRAPOLACAO-DA-TABUA-IBGE-MPS.pdf>. Acessado em 20/09/2020.
- Chollet, F. (2018) “Deep learning with Python”. Shelter Island, NY: Manning Publications Co. Print.
- Cordeiro Filho, A. (2017) “Cálculo atuarial aplicado: teoria e aplicações: exercícios resolvidos e propostos”. 2. ed. -[3.Reimpr.]. São Paulo: Atlas.
- Escovedo, T. e Koshiyama, A. (2020) “Introdução a Data Science - Algoritmos de Machine Learning e métodos de análise”, Casa do Código.
- Goodfellow, I., Yoshua, B., Courville, A. (2016), “Deep Learning”. MIT Press.
- Lee, R. D. e Carter, L. R. (1992) “Modelling and Forecasting U.S. Mortality”. Journal of the American Statistical Association. 87, 659-671.
- Lindemann, B., Müller, T., Vietz, H., Jazdi, N. Weyrich, M. (2021) “A survey on long short-term memory networks for time series prediction”. Procedia CIRP, v. 99, pp. 650-655.
- Lindholm, M., Palmborg, L. (2021) “Efficient Use of Data for LSTM Mortality Forecasting”. SSRN Electronic Journal, January 2021.
- London, C. e Layton, E. (1899); New York, The Spectator company; [etc, etc.], Transactions of the second International Actuarial Congress, held in the hall of the Institute of Actuaries, Staple Inn, Holborn, London, May 16 to 20, 1898, by International Congress of Actuaries (2nd : 1898 : London). Disponível em: <https://archive.org/details/transactionsofse00interich>. Acessado em: 20/09/2020.
- Nascimento, J.D. e Escovedo, T. (2021) “Construction of Mortality Tables using LSTM Neural Networks”. In XVII Brazilian Symposium on Information Systems (SBSI).
- Nigri, A, Levantesi, S., Marino, M., Scognamiglio, S., Perla, F. (2019) “A Deep Learning Integrated Lee–Carter Model”. Risks, v. 7, 33.
- Pascariu, M.D., Canudas–Romo, V., Vaupel, J.W. (1992) “Modelling and forecasting mortality: The Double-Gap Life Expectancy Forecasting Model”. University of Southern Denmark. October.
- Perla, F., Richman, R., Scognamiglio, S., Wuthrich, M. V. (2021) “Time-Series Forecasting of Mortality Rates using Deep Learning”. Scandinavian Actuarial Journal, v. 2021, 7.

- Petneházi, G., Gáll, J. (2019) “Mortality rate forecasting: can recurrent neural networks beat the Lee-Carter model?”. Disponível em: <https://arxiv.org/pdf/1909.05501.pdf>. Acesso em 20/02/2022.
- Pirovano, S. (2020) “Projection of Mortality Tables: A Neural Network Approach”. Master of Science in Statistical and Actuarial Sciences Thesis. Università Cattolica del Sacro Cuore Milano. Disponível em: <https://www.previdai.it/wp-content/uploads/2021/06/Tesi-Pirovano-Simone.pdf>. Acesso em 20/02/2020.
- Rodrigues, J.A. (2008) “Gestão de Risco Atuarial”. 1. ed. São Paulo: Saraiva.
- Souza, S. (2007) “Seguros: contabilidade, atuária e auditoria”. 2. ed.. São Paulo: Saraiva.
- Zaccone, G., Karim, R., e Menshawy, A. (2018) “Deep Learning with TensorFlow”, Birmingham: Packt Publishing. Print.