

CAIO FABRICIO DEBERALDINI NETTO

**SISTEMA DE PREDIÇÃO DE SÉRIES
TEMPORAIS BASEADO EM TÉCNICAS
NEURO-SIMBÓLICAS**

São Paulo
2020

CAIO FABRICIO DEBERALDINI NETTO

**SISTEMA DE PREDIÇÃO DE SÉRIES
TEMPORAIS BASEADO EM TÉCNICAS
NEURO-SIMBÓLICAS**

Trabalho apresentado à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Engenheiro Mecatrônico.

São Paulo
2020

CAIO FABRICIO DEBERALDINI NETTO

**SISTEMA DE PREDIÇÃO DE SÉRIES
TEMPORAIS BASEADO EM TÉCNICAS
NEURO-SIMBÓLICAS**

Trabalho apresentado à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Engenheiro Mecatrônico.

Área de Concentração:
Engenharia Mecatrônica

Orientador:
Prof. Dr. Fábio Gagliardi Cozman

São Paulo
2020

AGRADECIMENTOS

A realização deste trabalho não seria possível sem a ajuda de um número expressivo de pessoas, as quais sempre estiveram ao meu lado nos bons e maus momentos de minha vida. Aqui, dedico a eles esta singela homenagem:

Aos meus pais, Hilda e Nivaldo, meu imenso agradecimento por possibilitarem a realização deste sonho. Vocês ensinaram-me o real valor do trabalho bem feito, da perseverança, da honestidade e ética. Espero, no futuro, retribuir-lhes tudo aquilo o que vocês, a partir de muito esforço, proporcionaram a mim.

Ao meu irmão, Markus, e minha namorada, Kissya, agradeço pela incansável paciência e incentivos irrestritos ao longo não apenas deste último ano, como de minha graduação.

Ao meu orientador, Prof. Dr. Fábio Gagliardi Cozman, pela excepcional orientação ao longo deste ano atribulado. A amizade formada a partir do respeito mútuo tanto dentro quanto fora do escopo acadêmico foram-me vitais para a realização deste trabalho, assim como a finalização de minha graduação em situação tão adversa. Agradeço, também, pela compreensão e paciência ao longo de cada etapa deste processo.

Aos colegas do curso de Engenharia Mecatrônica, em especial Andrés, André e Paulo. Assim como um agradecimento especial aos amigos de longa data.

A todos os professores e funcionários do departamento, os quais ajudaram direta ou indiretamente para a minha formação profissional, ética e social, bem como para a conclusão deste trabalho.

“Excellence is never an accident. It is always the result of high intention, sincere effort, and intelligent execution; it represents the wise choice of many alternatives - choice, not chance, determines your destiny.”

-- Aristóteles

RESUMO

Sistemas de predição de séries temporais têm apresentado crescente desempenho graças ao uso de modelos estatísticos e modelos neurais, os quais são aprendidos a partir de grandes bases de dados. Combinações desses modelos com regras levam ao aprendizado dito “neuro-simbólico”. Este trabalho desenvolverá um sistema de predição de séries temporais baseado em dados e regras, ou seja, em técnicas neuro-simbólicas.

Palavras-Chave – sistemas de predição, séries temporais, aprendizado de máquina, aprendizado estatístico, modelos neurais, raciocínio simbólico, neuro-simbólico.

ABSTRACT

Time series forecast systems have shown increasing performance thanks to the use of statistical models and neural models, which are learned from huge databases. Joining both models with rules takes it to the so called “neuro-symbolic” learning. This work will develop a time series prediction system based on data and rules, i.e. over neuro-symbolic techniques.

Keywords – forecast systems, time series, machine learning, statistical learning, neural models, symbolic reasoning, neuro-symbolic.

LISTA DE FIGURAS

1	Arquitetura de uma Rede Neural Recorrente conhecida como LSTM (<i>Long Short-Term Memory</i>)	12
2	Diagrama de fluxo representando um exemplo simples de um método simbólico - também conhecido como <i>rule-based AI</i>	13
3	Rede de conexões entre autores e os principais expoentes do tema, através das <i>keywords</i> “ <i>neural networks</i> ” AND “ <i>symbolic reasoning</i> ”.	15
4	Mapa de densidade de publicações dos autores, através das <i>keywords</i> “ <i>neural networks</i> ” AND “ <i>symbolic reasoning</i> ”. Cores mais quentes indicam maior quantidade de publicações.	16
5	Rede de conexões entre autores e os principais expoentes do tema, através das <i>keywords</i> “ <i>neuro symbolic</i> ” OR “ <i>neural-symbolic</i> ”.	17
6	Mapa de densidade de publicações dos autores, através das <i>keywords</i> “ <i>neuro symbolic</i> ” OR “ <i>neural-symbolic</i> ”. Cores mais quentes indicam maior quantidade de publicações.	18
7	Documentos publicados por ano.	18
8	Documentos publicados por país.	19
9	Documentos publicados por autor.	20
10	<i>Framework</i> geral de um sistema baseado em técnica neuro-simbólica.	21
11	Baía de Sepetiba/Ilha Grande e a localização dos principais terminais portuários: 1 - Terminal Angra (Petróleo) TEBIG; 2 - Terminal da Ilha de Guaíba (TIG); 3 - Porto de Sepetiba; 4 - Terminal CSN.	28
12	Localização das bóias que medem a velocidade da corrente e do vento, elevação da maré, temperatura e visibilidade. Nome das bóias: 1) B18; 2) BEV; 3) TIG; 4) TIG 1; 5) Bifurcação; 6) Evolução; 7) Pier; 8) BPA; 9) B22	29
13	Grafo: nós, arestas e atributos global.	30

14	Distribuição para um dos fenômenos da bóia Bifurcação. Devido às restrições para divulgação dos dados, o fenômeno foi anonimizado tanto pela omissão do que ele representa quanto pela transformação dos valores. . . .	34
15	Distribuição para um dos fenômenos da bóia B22. Devido às restrições para divulgação dos dados, o fenômeno foi anonimizado tanto pela omissão do que ele representa quanto pela transformação dos valores.	35
16	Exemplo do processo de predição.	38
17	Resultados para uma janela de tempo selecionada, comparando os modelos e os <i>baselines</i>	39
18	Resultados para outra janela de tempo.	40

LISTA DE TABELAS

1	<i>Dataset</i> dos dados. Os valores dos fenômenos aqui presentes são meramente ilustrativos.	33
2	Hiperparâmetros e seus valores. Seus ajustes foram realizados através de <i>grid-search</i> durante a fase de validação.	37
3	Comparação entre os modelos a partir do erro quadrático médio (MSE, na sigla em inglês). Tanto GNN Não Local e LSTM obtiveram os melhores resultados, apesar do primeiro modelo ter maior eficiência de dados — necessita de menos dados para obter o mesmo resultado.	38

SUMÁRIO

1	Introdução	10
1.1	Objetivos	10
1.2	Estrutura do Trabalho	10
2	Revisão Bibliográfica	11
2.1	Considerações Iniciais	11
2.2	Revisão Bibliométrica	14
2.3	Análise da Literatura	20
3	Metodologia	25
4	Requisitos	26
5	Desenvolvimento	27
5.1	O Problema da Região da Baía de Sepetiba/Ilha Grande	27
5.2	Redes Neurais de Grafos	30
5.2.1	Explorando relações espaço-temporal	30
5.2.2	Modelando o problema em GNN	31
5.3	Dados do Problema	33
5.3.1	Análise Exploratória	33
5.3.2	Imputação dos Dados	35
6	Resultados e Conclusões	37
	Referências	42

1 INTRODUÇÃO

1.1 Objetivos

Os últimos dez anos testemunharam grandes mudanças na área de aprendizado de máquina. Houve significativo avanço em técnicas que empregam otimização local para ajustar modelos de grande porte, como redes neurais profundas, a partir de enormes massas de dados. Por outro lado, avanços em capacidade de processamento e em coleta de dados permitiram que tais modelos e tais massas de dados fossem efetivamente processados. Tópicos como por exemplo tradução automática foram muito afetados: hoje os melhores tradutores automáticos usam modelos de linguagem baseados em redes neurais [1]. Além disso, uma gama variada de sistemas (e.g., sistemas de classificação de objetos, de perguntas/respostas, tradução automática, predição de fenômenos, etc) passaram a depender de modelos latentes (aprendizado de representação) que capturam padrões essencialmente estatísticos em grandes bases de dados [2].

Este trabalho desenvolverá um sistema de predição de séries temporais, baseado em dados e regras, e aprendido por meio de técnicas neuro-simbólicas.

1.2 Estrutura do Trabalho

O presente trabalho está estruturado da seguinte forma: o Capítulo 2 faz uma revisão da literatura sobre o tema, entendendo quais são seus principais autores, o que são as técnicas neuro-simbólicas e como é possível integrar modelos latentes com simbolismo. Ademais, quais os modelos de predição de séries temporais usuais em *Deep Learning* e possíveis caminhos para abordar o problema; depois, têm-se os Capítulos 3 e 4, os quais versam sobre a metodologia empregada e os requisitos para cumprir com os objetivos deste projeto; na sequência, o Capítulo 5 apresenta o problema, as soluções escolhidas, sua modelagem e as dificuldades do problema; por fim, o Capítulo 6 mostra os resultados obtidos, os principais desafios encontrados e a conclusão do trabalho.

2 REVISÃO BIBLIOGRÁFICA

2.1 Considerações Iniciais

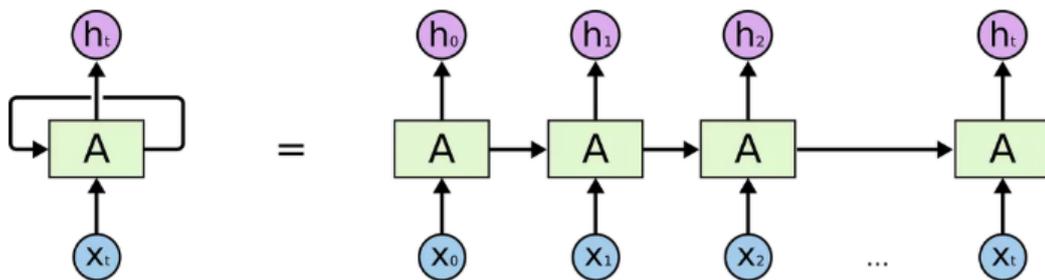
O progresso meteórico nos últimos anos da área de Inteligência Artificial reflete, para algumas pessoas, uma ilusão a respeito dos avanços e pesquisas feitas na área. Tendo seu início datado por volta da metade do século XX, a área, que hoje é considerada o que a eletricidade foi durante a 3ª Revolução Industrial [3], passou por um inverno sofrível até as décadas finais daquele século. Muito em parte pois os computadores não possuíam as bases computacionais e de processamento existentes hoje em dia, assim como pelos resultados irrisórios pela falta de técnicas e modelos que, devido a falta de bases massivas de dados, utilizavam-se de regras para obterem resultados fracos - técnicas conhecidas como simbólicas -, comparando-se com os obtidos pelos sistemas atuais. Desde então houve um aumento significativo na aquisição de novos dados, no seu tratamento e no seu compartilhamento (o surgimento da internet foi, sem dúvidas, um ponto de inflexão) e, estudos antes congelados, como as primeiras tentativas de mimetizar o que, porventura, venha a ser a estrutura neural de um ser humano - as primeiras aplicações de redes neurais datam do fim do século XX -, aqueceram-se e passaram a dominar tanto a área acadêmica, quanto os noticiários e cotidiano das pessoas.

O aprendizado de máquina (*Machine Learning* - ML) e, mais recentemente, o aprendizado profundo (*Deep Learning* - DL) são a prova desse cenário retrospectivo. Atualmente, os melhores sistemas artificiais ditos inteligentes fazem parte do escopo dessas áreas [4, 5]. A área de Processamento de Linguagem Natural (ou, NLP, em sua sigla em inglês), por exemplo, passou a adotar modelos baseados em grandes massas de dados, em especial modelos neurais (ou latentes) [6], em suas soluções, em contrapartida à soluções mais clássicas de NLP, como o próprio *Watson* da IBM o faz. Essa mudança de paradigma é bastante compreensível, dado que modelos latentes conseguem levar os sistemas a obterem acurácias, precisões e outras métricas nunca antes obtidas - ainda que haja um preço no que tange a interpretabilidade desses modelos, muitas vezes tidos como “caixas-pretas”.

Modelos neurais baseiam-se em formulações matemáticas que buscam aprender representações a partir da distribuição dos dados que os alimentam. Diferentes domínios e tarefas são explorados a partir de arquiteturas diversas. O *Multi-Layer Perceptron* (MLP) foi um dos primeiros modelos neurais que obtiveram resultados significativos em áreas como Visão Computacional [7]. Hoje em dia, as arquiteturas que conseguem os melhores resultados, inclusive melhores do que seres humanos, pautam-se no cerne do aprendizado profundo - utilização de grandes massas de dados aliado a modelos complexos. Arquiteturas como Redes Neurais Convolucionais, Redes Neurais Recorrentes e Redes Generativas Adversárias (CNN, RNN e GAN, respectivamente, nas siglas em inglês) são alguns exemplos de redes neurais que encontram-se neste ápice.

No contexto deste trabalho, assim como será discutido na seção seguinte, os problemas de predição de séries temporais são atacados com grande êxito fazendo-se uso de modelos estatísticos relacionais, e.g. modelos auto-regressivos (AR) [8, 9], médias móveis exponencialmente ponderadas [10] e modelos *bayesianos* (baseados nas evidências de distribuições *a priori*) [11]. Todavia, atualmente, com a popularização do aprendizado profundo e a capacidade de coleta de dados, abordagens *data-driven* passaram a ser adotadas e a fornecerem resultados sobre-humanos [12] — em algum senso. Neste sentido, modelos sequenciais passaram a ser utilizados para tarefas de predição de séries temporais, tais como RNNs [13] (exemplo da arquitetura de RNN conhecida como LSTM na Figura 1).

Figura 1: Arquitetura de uma Rede Neural Recorrente conhecida como LSTM (*Long Short-Term Memory*)

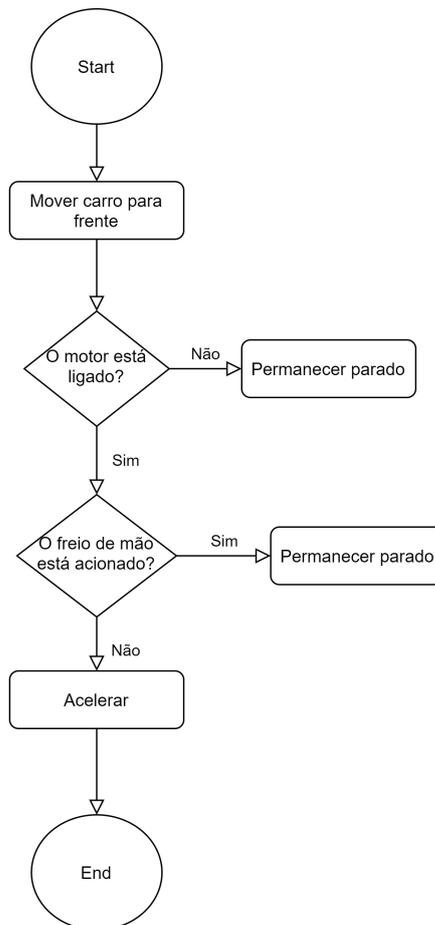


Fonte: Extraído de [14].

No espectro oposto, muitos pesquisadores defendem o uso de modelos simbólicos, com a justificativa de que há uma certa regularidade nas relações estatísticas existentes nos dados e fatos previamente conhecidos, podendo-se extrapolar esse conhecimento *a priori* para inferir novas informações. Trabalhos como [15, 16, 17] abordam o tema sob perspectivas diferentes. Nesse caso, extrair informações contextuais e/ou semânticas a partir

de dados simbólicos, por exemplo de palavras e relações em Grafos de Conhecimento (*Knowledge Graphs* ou KG, na sigla em inglês) é uma tarefa mais próxima de nós, seres humanos, porém para uma máquina não é natural. Os resultados obtidos através de abordagens e técnicas simbólicas são, geralmente, menos precisas, todavia mais interpretáveis. Nesse contexto, a Figura 2 permite visualizar essas características presentes em técnicas e modelos ditos simbólicos.

Figura 2: Diagrama de fluxo representando um exemplo simples de um método simbólico - também conhecido como *rule-based AI*.



Fonte: Autor.

No caso acima, um exemplo simplista do que, porventura, poderiam ser as regras necessárias para um carro autônomo (um modelo simbólico baseado puramente em regras), torna-se bastante claro os motivos das técnicas simbólicas serem menos precisas do que as técnicas e modelos neurais. Em cenários menos abrangentes, muito provavelmente um raciocínio baseado em regras seja valioso, como, por exemplo, na classificação de uma fruta baseada na sua cor, formato, tamanho, etc. Porém, no contexto de carros autônomos, torna-se praticamente inviável estabelecer regras que consigam englobar toda a gama

de situações que um carro autônomo pode enfrentar (tendo em vista que esses veículos baseiam-se em Visão Computacional, uma leve mudança na iluminação da imagem e/ou vídeo do cenário capturado já acarretaria na necessidade de formulação de novas regras). O mesmo é válido para problemas na área de NLP: em problemas complexos, sistemas puramente simbólicos não fornecem os mesmos resultados extraordinários de modelos neurais, apesar de seus benefícios.

Desta forma, de acordo com pesquisas recentes e especialistas da área [18, 19], a combinação dessas duas abordagens pode trazer o melhor de ambas as técnicas e, sob uma ótica otimista, alavancar a Inteligência Artificial a níveis até então inatingidos. A justificativa seria na completude obtida na união das duas técnicas: modelos simbólicos possuem seus pontos fortes nas fraquezas existentes em modelos neurais — e vice-versa. Assim, modelos neuro-simbólicos seriam interpretáveis, robustos, generalizáveis e precisariam de menos dados para serem treinados.

Desse modo, a fim de compreender como este tema (técnicas e metodologias neuro-simbólicas) vem sendo tratado pelos pesquisadores, assim como a possível aplicação em problemas de predição de séries temporais, propõe-se analisar a literatura através de duas abordagens. A primeira por meio de uma análise bibliométrica. A segunda através de uma análise de trabalhos específicos resultantes da análise bibliométrica.

2.2 Revisão Bibliométrica

A revisão bibliométrica é uma abordagem que tem como objetivo entender como um tema qualquer vem sendo pesquisado ao longo dos anos, assim como quem são os principais autores, instituições e países responsáveis por essas pesquisas. Ou seja, é uma abordagem quantitativa que se baseia na captura e análise das informações obtidas nas bases de dados que reúnem produção científica (e.g. *Scopus*, *Web of Science*).

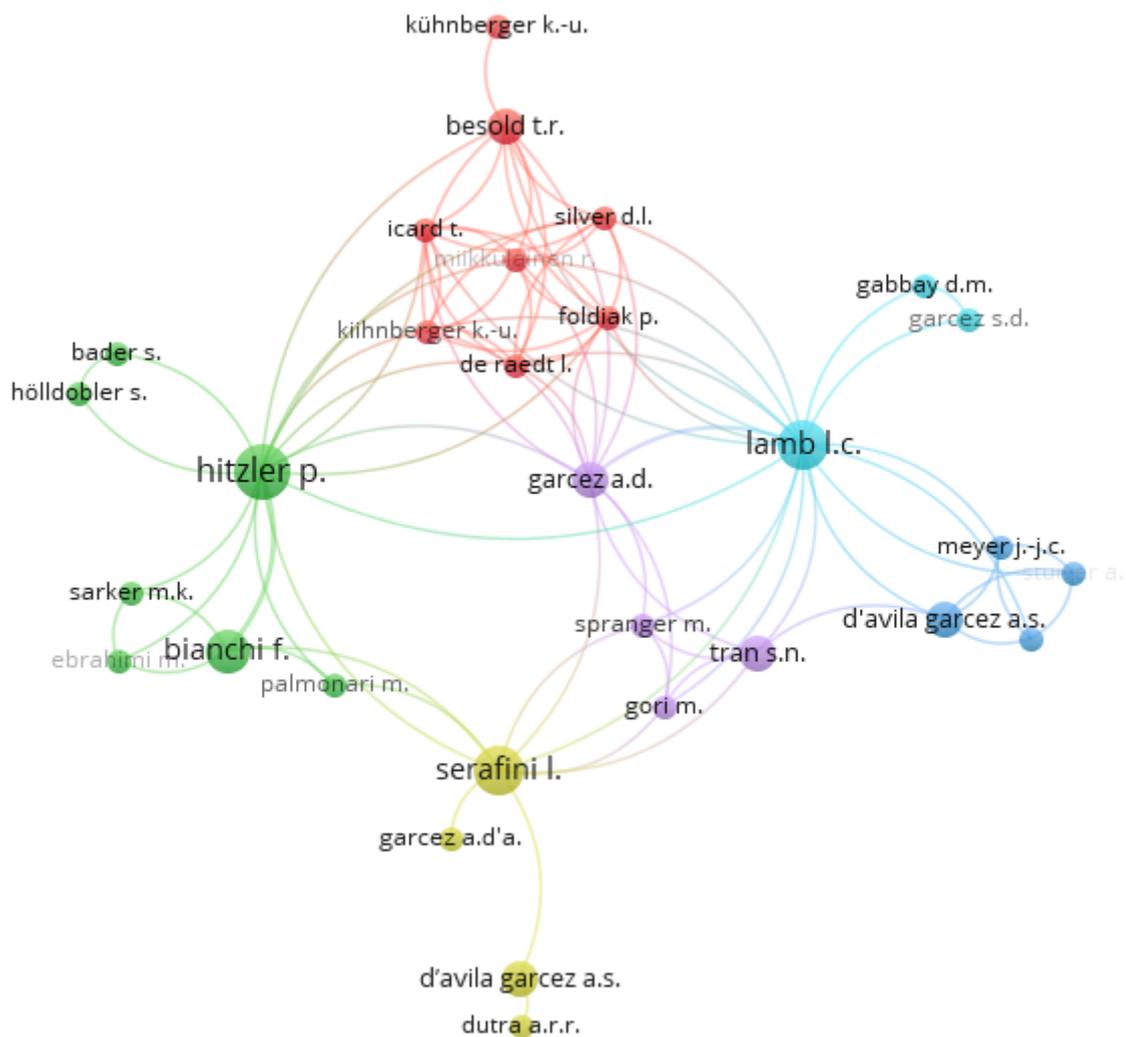
Para a realização da análise bibliométrica do tema em questão, foram utilizadas a própria ferramenta de análise de uma dessas bases de dados (*Scopus*) e o *software* — *VOSviewer* — capaz de criar redes de conexão entre os trabalhos e os autores, formando *clusters* e outras informações visuais, por meio dos metadados extraídos das bases de dados.

Os dados foram obtidos a partir de dois grupos de palavras-chave. O primeiro grupo foi a união das palavras “*neural networks*” e “*symbolic reasoning*” com o conector *AND*. O resultado obtido foram 73 documentos, passando de *surveys* e *overviews* até estudos

de caso.

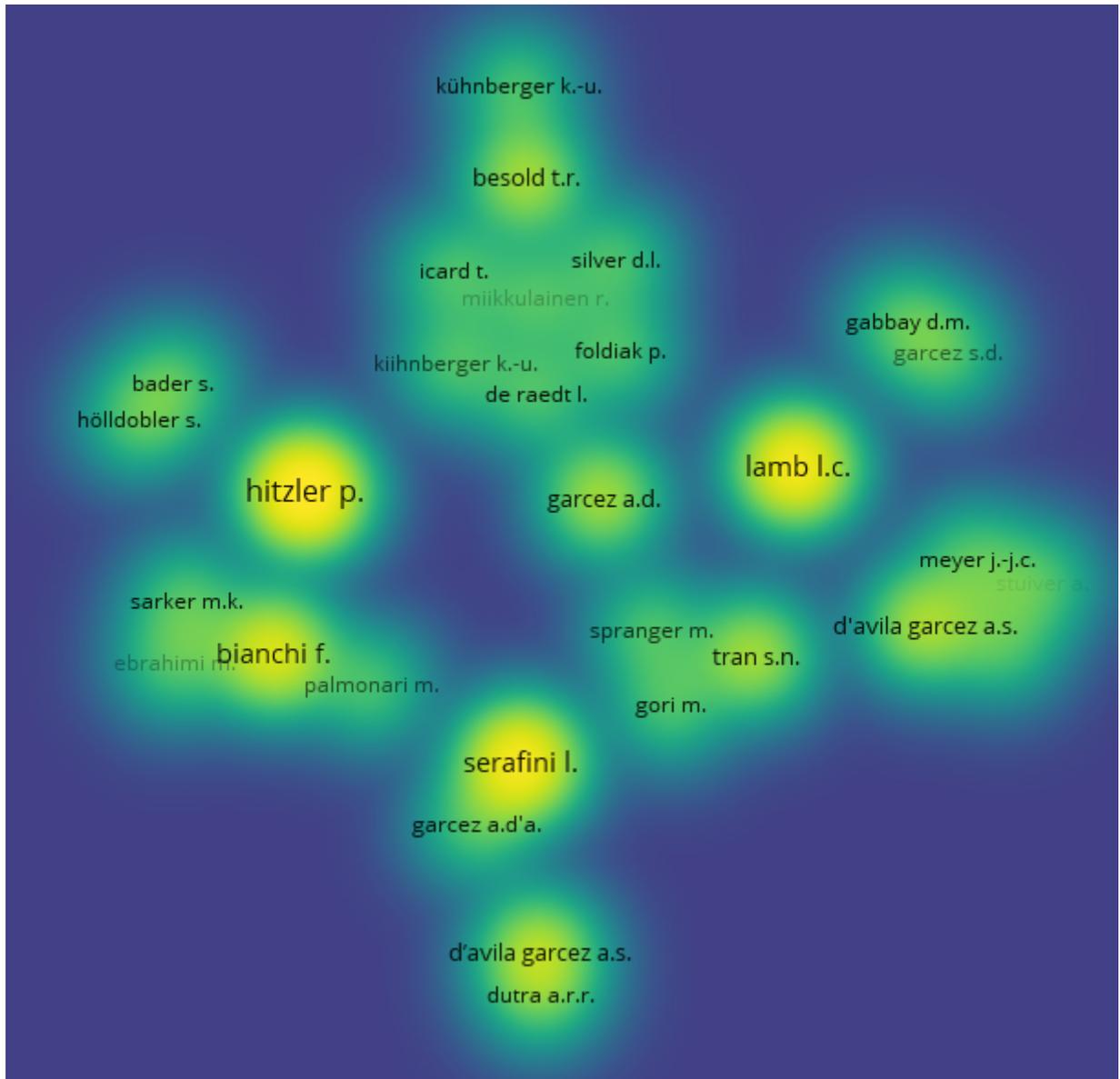
A Figura 3 apresenta a rede de conexões entre os autores, sendo possível notar a formação de *clusters* de autores os quais tiveram trabalhos conjuntos e em co-autorias (mais próximos no *cluster*), quem são os autores que mais publicaram a respeito do tema (vértices maiores) e como eles se relacionam (citações) a partir das arestas entre os nós. A Figura 4 apresenta o mapa de densidade para este mesmo conjunto de palavras-chave, representando quem são os autores que mais publicam sobre o tema (cores mais quentes no mapa).

Figura 3: Rede de conexões entre autores e os principais expoentes do tema, através das *keywords* “neural networks” AND “symbolic reasoning”.



Fonte: Autor com auxílio do *software VOSviewer*.

Figura 4: Mapa de densidade de publicações dos autores, através das *keywords* “*neural networks*” AND “*symbolic reasoning*”. Cores mais quentes indicam maior quantidade de publicações.

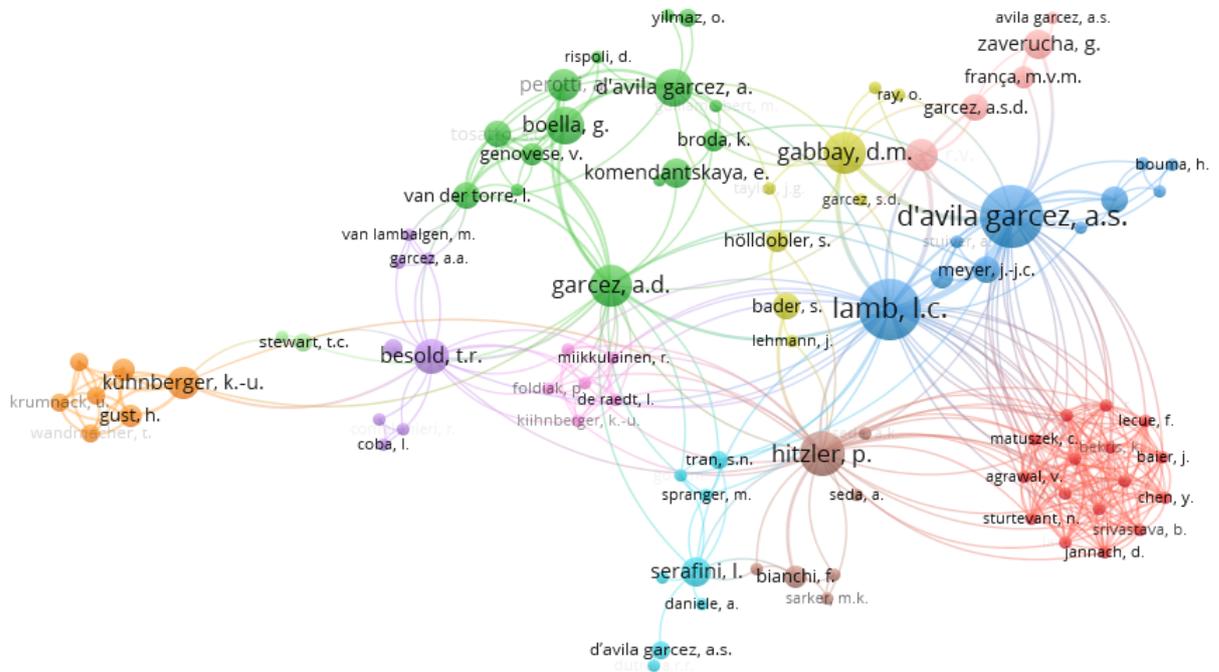


Fonte: Autor com auxílio do *software VOSviewer*.

Ainda que com este primeiro conjunto de palavras-chave tenhamos obtido um número expressivo de trabalhos e conseguimos visualizar quais os autores mais relevantes, buscamos expandir a análise. Assim, o segundo grupo de palavras-chave foi a união das palavras “*neuro symbolic*” e “*neural-symbolic*” com o conector *OR*. O resultado foi quase 5 vezes maior do que o anterior (316 documentos) e pôde fornecer mais informações a respeito da produção científica do tema.

A Figura 5 representa a rede de conexões entre os autores, sendo possível notar a

Figura 5: Rede de conexões entre autores e os principais expoentes do tema, através das *keywords* “*neuro symbolic*” OR “*neural-symbolic*”.



Fonte: Autor com auxílio do *software VOSviewer*.

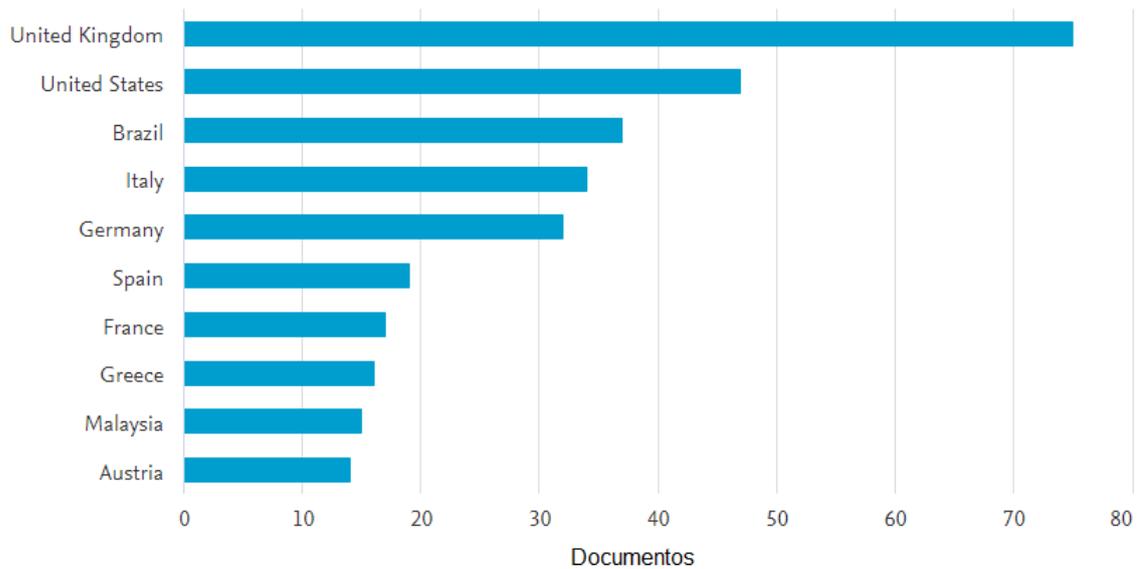
maior diversidade de pesquisadores, apesar de apresentar os mesmos autores com maior expressividade na área — inclusive com relações e quantidade de publicações semelhantes, representadas pelas arestas e o tamanho dos nós. O mapa de densidade para este conjunto de palavras-chave é representado pela Figura 6. É possível notar os autores mais produtores, em consonância com o exposto nas Figuras 3 e 4.

Como o segundo conjunto de palavras-chave foi mais proveitoso, analisamos, a partir da ferramenta disponível na base de dados *Scopus*, como esses 316 trabalhos se distribuíam nas diferentes dimensões que se seguem.

A Figura 7 apresenta como o tema passou a ser explorado ao longo dos anos. É possível notar uma tendência de crescimento de pesquisas sobre o tema, com um pico de trabalhos publicados em 2020 (33 documentos) — vale ressaltar que a grande queda pós-2020 indica os trabalhos com publicação prevista para 2021.

Já a Figura 8 fornece a informação a respeito de quais países estão produzindo pesquisas a respeito do tema. O destaque do Brasil, figurando entre os 3 países com mais documentos sobre o tema (quase 40), poderá ser entendido na sequência. Ademais, países com excelentes reputações no que tange a pesquisas de ponta na área de IA figuram entre aqueles com mais publicações sobre o tema (Reino Unido e EUA).

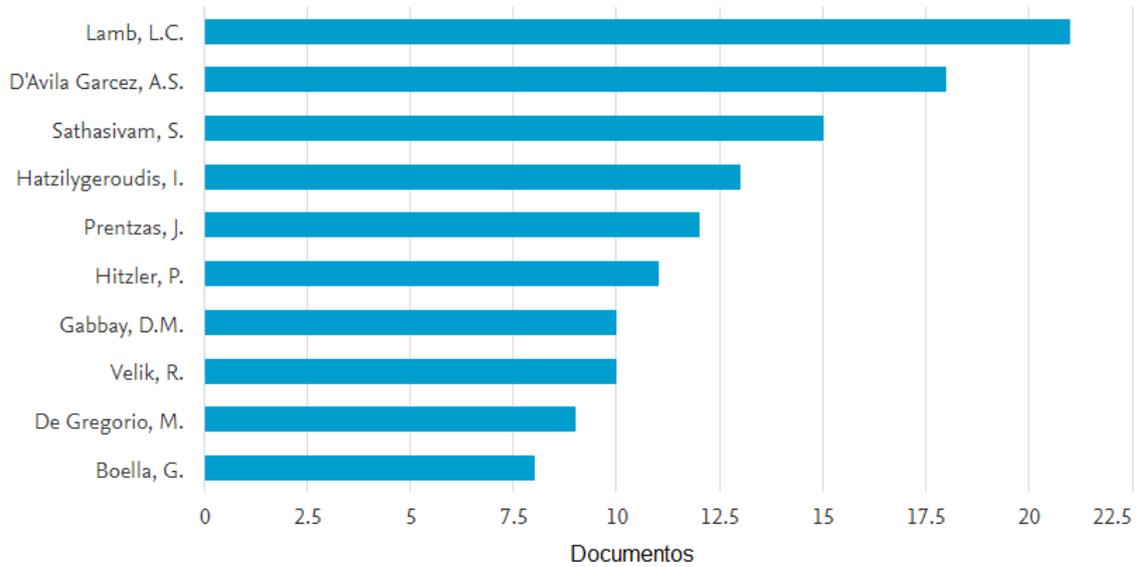
Figura 8: Documentos publicados por país.



Fonte: Autor adaptado de *Scopus*.

Por fim, como pôde ser observado pelas Figuras 3, 4, 5 e 6, além da posição expressiva do Brasil, representada pela Figura 8, o autor com maior número de publicações sobre o tema é Lamb, pesquisador brasileiro da Universidade Federal do Rio Grande do Sul (UFRGS) — apresentado na Figura 9. Vale destacar as co-autorias entre este autor e outros com grande destaque, como D’Avila Garcez, Hitzler e Besold, pesquisadores estes principalmente de universidades americanas e/ou britânicas — *University of London*, *Wright State University* e *University of London*, respectivamente.

Figura 9: Documentos publicados por autor.



Fonte: Autor adaptado de *Scopus*.

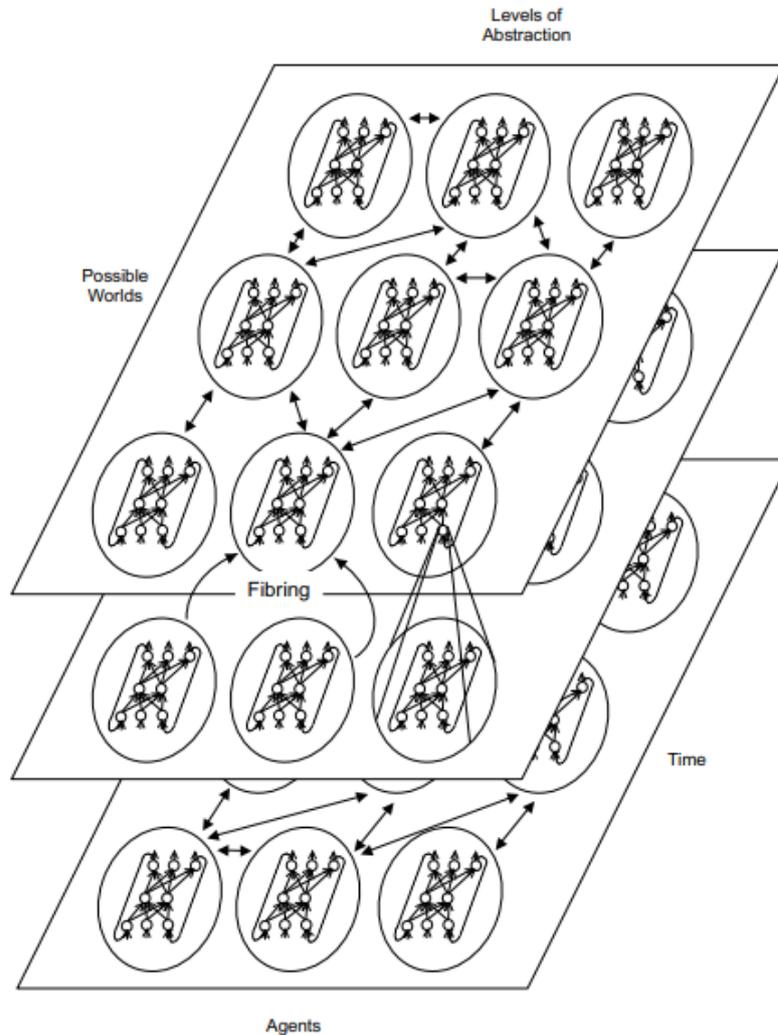
Com este panorama, seguiu-se para uma análise dos trabalhos, em especial *surveys*, destes autores que mais se destacaram no tema, além de como sistemas de predição de séries temporais enquadram-se neste cenário — apresentados na seção 2.3.

2.3 Análise da Literatura

Besold et al. (2017) [18] propõem uma visão holística sobre as principais motivações para a união entre os paradigmas conexionista e de raciocínio lógico, em termos computacionais, além de como os estudos de técnicas neuro-simbólicas permeiam diferentes áreas e aplicações, seus fundamentos teóricos e desafios. Os autores argumentam a respeito da fragilidade e dependência de modelos neurais a conjuntos de dados representativos das tarefas a que se propõe a solucionar, e como a integração de técnicas simbólicas podem auxiliar na resposta destes modelos ao acoplar um módulo de raciocínio (lógico ou probabilístico). Ou seja, o objetivo principal é, de fato, proporcionar a unificação deste dois paradigmas.

Uma forma de se chegar a esta unificação, de acordo com os autores, é pensando no problema a ser atacado de maneira modular. O problema precisa ser subdividido em pequenas partes, as quais serão responsáveis por tratar de um aspecto específico, por exemplo raciocínio sobre conceitos genéricos e conhecimento obtido a partir do aprendizado de representação. Assim, cada módulo seria responsável por auxiliar na construção

Figura 10: *Framework* geral de um sistema baseado em técnica neuro-simbólica.



Fonte: Extraído de [18].

de novos conhecimentos, possibilitando o modelo a raciocinar sobre novos conceitos. Vale destacar que o aprendizado e o raciocínio simbólico são traduzidos a partir de arquiteturas neurais, as quais são representadas, inicialmente, antes mesmo do aprendizado, ou seja, as particularidades do problema são integradas a arquitetura do modelo. Esta *framework* discutida pelos autores está representada na Figura 10. Diferentes redes neurais, cada uma responsável por raciocinar sobre um aspecto do problema em diferentes níveis de abstração, intra e interconectadas. Esta abstração aponta diretamente para a utilização de modelos em que exista um forte viés relacional entre suas entidades, não apenas do ponto de vista conexionista das redes neurais, mas inclusive do ponto de vista lógico, pois as diferentes redes neurais artificiais precisam ser concatenadas a fim de representarem uma abstração lógica.

Consequentemente, como apontado pelos autores, os principais desafios, tanto práticos

quanto teóricos, passam por (1) conseguir traduzir o raciocínio simbólico, que está relacionado com a lógica (proposicional, *booleana*, *fuzzy*, etc), (2) criar modelos de aprendizado e de raciocínio — híbridos e não “puro-sangue” — e (3) traduzir as redes em lógica (extração de conhecimento).

Sob esta perspectiva, buscamos entender como e quais os modelos neuro-simbólicos posicionam-se atualmente. *Besold et al.* (2017) citam alguns exemplos como BN (*Bayesian Networks*), MLN (*Markov Logic Networks*), RNM (*Relational Neural Machines*, como NTMs — *Neural Turing Machines*), dentre outros. E, seguindo linha parecida, *De Raedt et al.* (2020) [20], ao traçar paralelos entre os modelos ditos estatísticos relacionais e os neuro-simbólicos, comparando-os em 7 diferentes dimensões, expande esse leque de modelos que unem aprendizado e regras, mas possuindo muitas intersecções com o argumento expresso em [18].

Tanto *Besold et al.* (2017) quanto *De Raedt et al.* (2020) definem os modelos neuro-simbólicos em termos de modelagem em grafos (sejam diretos ou indiretos). Neste sentido, grande destaque é dado para as chamadas Redes Neurais de Grafos (GNNs, na sua sigla em inglês). *Lamb et al.* (2020) [21] apresentam, a partir deste *survey*, como GNNs ligam-se (ou podem ser relacionadas) com computação neuro-simbólica. O viés indutivo relacional das Redes Neurais de Grafos, assim como sua propriedade de invariância a permutação — as quais serão melhor explicitadas na seção 5.2.1 — são características comuns entre os modelos neuro-simbólicos supracitados e as GNNs.

Todos estes trabalhos [18, 20, 21] discutem sobre a utilização de modelos neuro-simbólicos em diferentes áreas: biologia computacional, avaliação de riscos, processamento de linguagem natural, visão computacional, etc. Todavia, apesar de algumas destas áreas compartilharem conhecimentos e técnicas aplicáveis ao estudo de séries temporais (e.g., mecanismos de memória e atenção, emprestados dos estudos feitos na área de NLP), existe uma lacuna no que tange a técnicas neuro-simbólicas em conjunto com modelos de aprendizado profundo aplicados a séries temporais.

Lim e Zohren (2020) [22] apresentam um panorama sobre como aprendizado profundo vem sendo utilizado em problemas de séries temporais, assim como quais são as técnicas clássicas utilizadas em tarefas de predição de conjuntos de dados de séries temporais. Apesar dos autores deixarem claro que este *survey* não tem nenhuma pretensão de ser um trabalho exaustivo, a classificação seguida pelos autores permite entender quais os prós e contras de cada categoria de modelos e técnicas aplicadas a séries temporais. Neste sentido, os métodos, de acordo com os autores, podem ser classificados em três categorias:

1. **Abordagem Clássica:** Os métodos tradicionais de modelagem de séries temporais têm como foco modelos puramente estatísticos e paramétricos, os quais dependem do ajuste de especialistas no domínio [22]. São representantes desta classe:
 - 1.1. **ARIMA** — *Autoregressive Integrated Moving Average* [8]
 - 1.2. **Exponential Smoothing** [10, 23]

2. **Abordagem por ML e DL:** Métodos baseados em aprendizado de máquina e aprendizado profundo buscam aprender a dinâmica temporal do conjunto de dados guiados estritamente nos padrões presentes na própria distribuição (i.e., são *data-driven*). Porém, o segundo conjunto de modelos, de DL, conseguem incorporar em suas arquiteturas os vieses relacionais que podem existir no conjunto de dados, permitindo o aprendizado de representações mais complexas. Exemplos desta classe de métodos:
 - 2.1. **Redes Neurais Convolucionais (CNN)** [24, 25]
 - 2.2. **Redes Neurais Recorrentes (RNN)** [13, 26, 27]
 - LSTM — Mecanismos de memória
 - *Transformers* — Mecanismos de atenção

3. **Abordagem Híbrida:** Apesar dos benefícios da utilização de modelos de aprendizado de máquina ou aprendizado profundo em problemas de séries temporais (e.g., reduzir ou extinguir a etapa de *feature engineering*), por serem *data-driven* esses modelos estão suscetíveis a sobreajustarem à distribuição dos dados, especialmente quando trabalha-se com modelos complexos como aqueles pertencentes ao aprendizado profundo [22]. Neste sentido, os métodos híbridos mostraram-se ideais para a solução de muitos problemas envolvendo séries temporais, conseguindo unir o melhor de ambos os paradigmas: incorporação de conhecimento sobre o domínio na estrutura do modelo em conjunto com o aprendizado de padrões. Assim, parte do modelo é responsável por capturar as informações temporais, enquanto a outra tem como função direcionar a solução do sistema, a partir das particularidades do domínio. Destacam-se os seguintes modelos:
 - 3.1. **ES-RNN** [28]
 - 3.2. **Gaussian Processes** [29]
 - 3.3. **Deep State Space** [30]

Os autores finalizam o trabalho explicitando algumas limitações que muitos modelos de aprendizado profundo, desenvolvidos para problemas de séries temporais, possuem. Primeiro, a área ainda carece de abordagens que consigam lidar com conjuntos de dados faltantes ou que possuam aquisição aleatória, tendo em vista que grande parte destes modelos tratam as relações temporais de forma regular. Por último, os modelos atuais carecem de arquiteturas que consigam capturar o caráter hierárquico dos conjuntos de dados de séries temporais. Como será visto no problema abordado neste trabalho, séries temporais podem possuir grupos logicamente conectados que influenciam no resultado do fenômeno observado. Estes agrupamentos podem advir tanto da estrutura do problema quanto de relações de causalidade latentes.

3 METODOLOGIA

Conforme indicado nas seções anteriores, pretende-se nesse trabalho focar na implementação de um sistema de predição de séries temporais baseado em técnicas neuro-simbólicas. O primeiro passo será a escolha do domínio; pretende-se atuar em duas etapas. Durante o primeiro semestre, construir o sistema de predição de séries temporais que consiga abordar um problema a partir da união do aprendizado em cima de grandes massas de dados e técnica simbólica. O problema em questão diz respeito a predição de fenômenos marítimos em uma região litorânea do Brasil (especificamente, a região da Baía de Sepetiba/Ilha Grande), de interesse da indústria de minérios, óleo e gás. Nesta primeira etapa, será possível compreender os principais desafios, assim como fazer o maior número de ajustes possíveis para que o sistema possa fornecer respostas robustas. Em seguida, implementar modelos *baseline*, com o intuito de comparação com o(s) modelo(s) implementados na primeira etapa. Assim, poderemos concluir se: (1) o uso de técnicas neuro-simbólicas faz sentido para a predição de séries temporais, onde há uma forte relação entre os atributos espaciais e temporais do problema; (2) validar a hipótese de que este tipo de sistema (baseados em técnicas neuro-simbólicas) de fato é mais robusto do que outras soluções e possui ganhos em termos de eficiência no aprendizado (necessidade de menos dados para obter melhores resultados).

4 REQUISITOS

Desse modo, apresentados a área da pesquisa, seus termos e trabalhos correlatos, com suas vantagens e desvantagens, seguem-se os requisitos para o projeto:

- **Escolha de um domínio:** Tendo em vista que o cerne do trabalho consiste na junção de duas técnicas de inteligência artificial (à título de esclarecimento, técnicas e modelos neurais unidos com técnicas simbólicas de raciocínio) e que, segundo apresentado na revisão bibliográfica e na revisão do estado-da-arte, técnicas simbólicas são precedidas pelo conhecimento das regras e procedimentos do domínio, estabelecer um domínio que seja amplamente discutido e conhecido.
- **Construção de um modelo neuro-simbólico de um sistema de predição de séries temporais:** Implementar um modelo neuro-simbólico, ou seja, que seja capaz de extrair informações estatístico-relacional dos dados, através de arquiteturas neurais, unido com as regras e conhecimento do domínio, por meio de técnicas simbólicas. A revisão do estado-da-arte aponta para a direção de modelos sequenciais que capturam relações espaço-temporais (e.g. GNNs).
- **Testes do sistema:** Após treinamento do modelo, testá-lo na base de dados do problema do domínio escolhido.
- **Validação da metodologia:** A fim de que outros modelos e técnicas sejam comparadas com a metodologia adotada, validar o modelo proposto com modelos *baseline* e modelos estado-da-arte na base de dados do problema escolhido.

5 DESENVOLVIMENTO

5.1 O Problema da Região da Baía de Sepetiba/Ilha Grande

Diferentes situações práticas requerem, com frequência, previsões onde tanto as informações presentes nos dados quanto os padrões estruturais dos problemas referentes a estas situações devem ser considerados conjuntamente. Por exemplo, se atributos são relacionados por meio de predicados, de modo que os dados cheguem na forma de grafos, é possível explorar o aprendizado relacional para construir um modelo de predição por meio de GNNs [31, 32]. Outro cenário diferente que parece pedir também alguma estrutura teórica em grafos acontece caso tenhamos conhecimento prévio sobre as relações entre atributos. Seja devido à relações espaciais, temporais ou espaço-temporais. Redes Neurais de Grafos parecem oferecer o veículo ideal para explorar essa estrutura de tempo/espaço subjacente.

Examinamos um problema real de importância tecnológica e econômica, onde é preciso lidar com um fluxo substancial de dados temporais e também com uma estrutura espacial conhecida. O problema é determinar a velocidade da corrente de água em um local específico dentro de uma importante área portuária da América do Sul, usando medições coletadas por uma rede de bóias marítimas próximas umas as outras. A predição das condições da água na área portuária é uma grande preocupação para as autoridades portuárias que precisam coordenar um grande número de navios pesados que transportam, por exemplo, petróleo ou minério. Os modelos físicos atuais são caros para desenvolver e manter, pois exigem medições de alta qualidade do ambiente, condições de contorno e uma representação espacial 3D precisa da localização.

Abstratamente, esse é um problema em que dados sequenciais são coletados através de uma rede de sensores com uma estrutura espacial fixa, onde os sensores geralmente estão com defeito. Assim, nosso problema específico é representativo de uma grande classe de desafios de predições enfrentados por várias indústrias e que poderiam se beneficiar

Figura 11: Baía de Sepetiba/Ilha Grande e a localização dos principais terminais portuários: 1 - Terminal Angra (Petróleo) TEBIG; 2 - Terminal da Ilha de Guaíba (TIG); 3 - Porto de Sepetiba; 4 - Terminal CSN.



Fonte: Autor.

amplamente da alavancagem de GNNs.

Neste sentido, foi desenvolvido, no presente trabalho, um modelo baseado em GNN cuja estrutura captura relações espaciais no domínio e cujos parâmetros são aprendidos a partir de dados históricos, para solucionar o problema de previsão de séries temporais da região da Baía de Sepetiba/Ilha Grande.

A Baía de Sepetiba/Ilha Grande está localizada no Estado do Rio de Janeiro, região sudeste do Brasil. É uma área protegida perto da cidade do Rio de Janeiro, com diferentes instalações portuárias e tráfego intenso de navios. A Figura 11 mostra a região de interesse na América do Sul (à esquerda) e os quatro principais terminais portuários: TEBIG, TIG, Porto de Sepetiba e CSN (à direita).

Condições marítimas e climáticas possuem um impacto direto nas operações exercidas nos terminais portuários devido às possíveis consequências para as embarcações comerciais que ali se locomovem, afetando frontalmente a segurança da navegação. Como exemplo, em condições de eventos extremos (e.g., correntes marítimas fortes ou baixa visibilidade), o tráfego de embarcações na região deve ser interrompido. Portanto, a previsão a curto prazo destes fenômenos ambientais (dentro de 24 a 48 horas) é uma tarefa essencial das autoridades portuárias. No caso de canais compartilhados com vários tamanhos de navios, como no caso da região em análise, a previsão torna-se ainda mais crítica. Desta forma, o interesse neste trabalho é realizar previsões dentro de um período de 24 horas, pois essa é a janela de previsão mais típica nesse contexto.

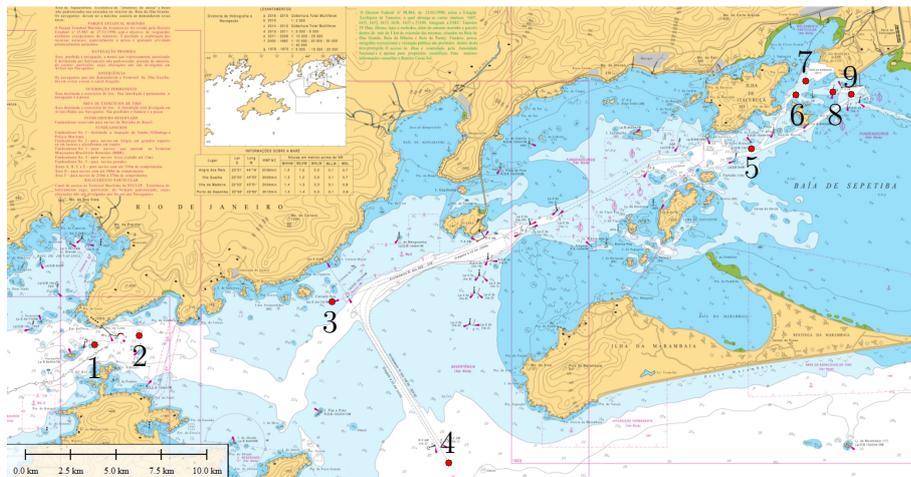
A previsão de parâmetros ambientais geralmente é feita com uma série de modelos de circulação atmosférica e hidrodinâmica. As entradas do método são as condições de limite, como variação das marés, informações de satélite de baixa resolução ou modelo global e medições locais do vento [33]. Eles também dependem de um modelo de grade 3D preciso

da área (litoral e batimetria). Os últimos anos testemunharam um número crescente de dados medidos, devido a novos sensores, tecnologias de transmissão e armazenamento de dados, e uma conseqüente chamada para técnicas mais sofisticadas baseadas em dados para previsão de séries temporais [34]. O modelo baseado em dados é independente dos modelos físicos e das entradas externas, sendo atualizado automaticamente assim que um novo conjunto de medições estiver disponível. É dentro dessa tendência de modelagem baseada em dados que busca-se operar neste trabalho.

Adotamos uma nova abordagem de modelagem que usa uma rede de sensores distribuída espacialmente ao longo da Baía de Sepetiba/Ilha Grande para prever a velocidade da corrente de água em uma única bóia conhecida como Bifurcação. A rede consiste em um sistema de nove bóias localizadas conforme indicado na Figura 12 (observe que a bóia Bifurcação aparece sobre o número 5). Cada bóia coleta medições de variáveis relacionadas à elevação da maré, velocidade da corrente e do vento, temperatura e visibilidade.

Cada bóia coleta uma rodada de medições a cada dez minutos. Devido à dinâmica relativamente lenta do sistema, convencionou-se fazer uma subamostragem e analisar intervalos de vinte minutos entre medidas. Notavelmente, a coleta de dados é bastante falha (problema abordado na seção 5.3): vários atributos estão ausentes em muitas rodadas de medição e, em alguns casos, bóias não relatam um ou mais atributos por meses. Isso ressalta o benefício da abordagem GNN que explora as informações espaciais.

Figura 12: Localização das bóias que medem a velocidade da corrente e do vento, elevação da maré, temperatura e visibilidade. Nome das bóias: 1) B18; 2) BEV; 3) TIG; 4) TIG 1; 5) Bifurcação; 6) Evolução; 7) Pier; 8) BPA; 9) B22



Fonte: Autor.

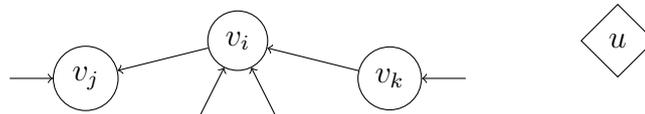
5.2 Redes Neurais de Grafos

5.2.1 Explorando relações espaço-temporal

Redes neurais profundas agora oferecem excelente, e muitas vezes surpreendente, desempenho em uma variedade de configurações onde os padrões devem ser identificados e postos em prática. No entanto, as redes neurais totalmente conectadas enfrentam dificuldades em capturar relações entre entidades. Em alguns casos, as redes neurais adotam pesos que são compartilhados por muitas unidades, por exemplo em Redes Neurais Convolucionais (CNNs) ou Redes Neurais Recorrentes (RNN); contudo, mesmo nesses casos, as conexões são principalmente locais dentro do modelo (espacial ou temporalmente).

Redes Neurais de Grafos (GNNs, na sigla em inglês) foram desenvolvidas de forma a levar em consideração a estrutura de um domínio, expresso através de relações entre entidades [35, 31]. O objetivo é trabalhar na intersecção entre técnicas neurais e modelagem simbólica, onde as relações e conexões subjacentes capturam os aspectos simbólicos do domínio. Para codificar o domínio, uma GNN usa nós (que pertencem a uma ou mais classes), arestas (que correspondem a predicados binários) e atributos globais. Figura 13 representa um fragmento de um grafo com tais objetos.

Figura 13: Grafo: nós, arestas e atributos global.



Fonte: Autor.

Para ilustrar melhor os componentes de uma GNN, considere uma configuração simples que não está relacionada ao aplicativo que descrevemos a seguir. Considere o problema físico de prever a posição da massa em um sistema massa-mola. Pode-se modelar o sistema físico usando um grafo com nós representando as massas do sistema e arestas representando relações entre nós - isto é, interações entre massas devido às molas. Os atributos de massas v_i são suas posições, velocidades e valores de massa. Os atributos de arestas e_k são a rigidez e o comprimento natural da mola que conecta cada massa (vértice). Por último, o atributo global u , compartilhado com todas as entidades, é a força da gravidade. Esses objetos definem o grafo subjacente. Para prever os atributos do modelo, uma função de predição deve ser aplicada às entidades; por exemplo, pode-se querer prever a posição de uma massa (nó) após 5 etapas de tempo no problema hipotético do sistema massa-mola.

Para determinar o comportamento de uma GNN, várias funções devem ser especificadas. *Battaglia et al.* (2017) [36] definem, em seu esquema geral, as funções associadas aos vértices, arestas e atributos globais:

$$\begin{aligned} e'_k &= \phi^e(e_k, v_{r_k}, v_{s_k}, u), & \bar{e}'_i &= \rho^{e \rightarrow v}(E'_i) \\ v'_i &= \phi^v(\bar{e}'_i, v_i, u), & \bar{e}' &= \rho^{e \rightarrow u}(E') \\ u' &= \phi^u(\bar{e}', \bar{v}', u), & \bar{v}'_i &= \rho^{v \rightarrow u}(V') \end{aligned}$$

onde os atributos das entidades arestas e nós, e atributos globais, respectivamente e_k , v_i e u , são atualizados pelas funções ϕ . Além disso, os atributos das arestas são agregados por funções ρ , que podem ser a média, mediana, soma (qualquer função de agregação) de forma a serem obtidos pelos vértices apontados pelas arestas ou pelos atributos globais. O mesmo é válido para agregar atributos dos nós com o propósito de atualizar os atributos globais. Para uma GNN, cada função ϕ é codificada por uma rede neural. Todas essas redes devem ter sua estrutura pré-definida e seus pesos serão aprendidos — como será discutido posteriormente.

5.2.2 Modelando o problema em GNN

Seguindo *Battaglia et al.* (2017), tomamos um grafo atribuído (direcionado) como uma tupla $G = (V, E)$, em que $V = \{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ é um conjunto dos atributos do nó (vetores com valor real) e $E = \{(\mathbf{e}_1, r_1, s_1), \dots, (\mathbf{e}_M, r_M, s_M)\}$ é um conjunto de triplas contendo um atributo de aresta \mathbf{e}_k (um vetor de valor real) sobre a aresta (r_k, s_k) ¹.

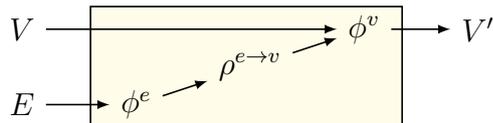
Nosso grafo é especificado da seguinte forma: cada vértice representa uma bóia cujos atributos são componente de velocidade da corrente no eixo x, componente de velocidade da corrente no eixo y, componente da velocidade do vento no eixo x, componente da velocidade do vento no eixo y, nível do mar, temperatura e visibilidade locais. Quanto às arestas, um especialista forneceu conhecimento de domínio selecionando quais atributos dos nós afetam mais os demais atributos dos vértices adjacentes. Os atributos da aresta são uma projeção dos atributos componente de velocidade da corrente no eixo x, componente de velocidade da corrente no eixo y e nível local do mar do vértice adjacente.

Consideramos dois tipos de topologia de grafos: uma versão totalmente conectada (denominado modelo “Não Local”) e uma versão totalmente desconectada (denominado modelo “Local”). A primeira versão captura efeitos entre bóias (como todas compartilham

¹[36] também define um atributo global \mathbf{u} que não usamos.

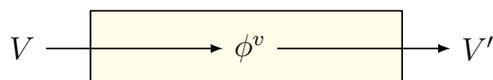
uma localização geográfica, estão relacionadas espacialmente); já a segunda versão não aproveita o viés indutivo relacional e foi construído apenas para fins de comparação e obtenção de evidências sobre a hipótese de melhoria do modelo ao compartilhar informações entre as entidades. Ou seja, têm-se dois esquemas:

- A GNN Não Local atualiza os atributos dos vértices levando em consideração todas as informações relacionais [37].



A função $\mathbf{e}'_k = \phi^e(\mathbf{e}_k)$ atualiza cada atributo de aresta \mathbf{e}_k . Esses atributos atualizados são agregados através de uma soma ponderada como $\bar{\mathbf{e}}_i = \rho^{e \rightarrow v}(\{\mathbf{e}_k\}_{k:r_k=i}) = \sum_{k:r_k=i} \mathbf{w}_k \mathbf{e}'_k$, em que \mathbf{w}_k é uma função da distância entre o nó emissor e o nó receptor. Para um nó receptor, calculamos o inverso da raiz quadrada da distância física (fornecida por especialista na área) entre esse nó e o nó emissor adjacente. Em seguida, normalizamos esse fator com a soma de todas essas distâncias invertidas de cada nó emissor para aquele receptor, obtendo nosso peso \mathbf{w}_k . Em resumo, queremos que, para todas as bóias, os atributos dos seus parentes mais próximos tenham maior relevância. Finalmente, a função $\mathbf{v}'_i = \phi^v(\bar{\mathbf{e}}_i, \mathbf{v}_i)$ atualiza cada atributo de nó \mathbf{v}_i levando em consideração o atributo agregado $\bar{\mathbf{e}}_i$. Conforme especificado na Seção 5.2.1, em uma abordagem GNN, as funções aplicadas às entidades do grafo são redes neurais. Nesse sentido, as funções de atualização descritas acima como ϕ^e e ϕ^v são redes neurais (NN) alimentadas por atributos da entidade. Em outras palavras, dados os atributos das bóias em uma etapa de tempo, que estão nos nós e arestas do grafo, seus valores na próxima etapa são previstos atualizando os atributos das arestas, $\mathbf{e}'_k = NN^e(\mathbf{e}_k)$, que é uma variável independente para atualizar os atributos dos nós: $\mathbf{v}'_i = NN^v(\bar{\mathbf{e}}_i, \mathbf{v}_i)$.

- O bloco de GNN Local, por sua vez, analisa apenas os atributos dos vértices, isto é, temos apenas $\mathbf{v}'_i = \phi^v(\mathbf{v}_i)$.



Observe que o segundo esquema (Local) ainda se beneficia da estrutura em grafo, pois a função ϕ^v é compartilhada entre todos os nós.

Para permitir que as GNNs capturem a evolução temporal dos sinais, concatenamos, para cada nó e aresta, medições das últimas 48h, coletadas a cada 20 minutos. Portanto,

gera-se atributos de nós $\mathbf{v}_i \in \mathbb{R}^{145 \times 7}$ e atributos de arestas $\mathbf{e}_k \in \mathbb{R}^{145 \times 3}$ (ou seja, 144 pontos de dados observados mais o ponto do próximo passo a ser previsto).

5.3 Dados do Problema

5.3.1 Análise Exploratória

O conjunto de observações consiste de uma grande base de dados de 01-01-2018 00h00 a 31-12-2019 23h50 (2 anos de medições), amostrados de 10 em 10 minutos, dos quatro locais da Baía de Sepetiba/Ilha Grande (TEBIG, TIG, CSN e Porto de Sepetiba), o que fornece, aproximadamente, 105 mil pontos².

Os dados consistem de medições de fenômenos marítimos e climáticos através de uma rede de sensores, instalados em bóias, totalizando 9 pontos de medição. Sete são os fenômenos observados: componentes de velocidade da corrente de água nos eixos x e y, componentes de velocidade do vento nos eixos x e y, nível do mar e temperatura local, e visibilidade. A Tabela 1 ilustra a disposição dos dados obtidos.

Tabela 1: *Dataset* dos dados. Os valores dos fenômenos aqui presentes são meramente ilustrativos.

t	bcx_media	bcy_media	...	B18v
01 – 01 – 2018 00h00	0.789	0.567	...	<i>NaN</i>
01 – 01 – 2018 00h10	0.743	0.598	...	1.0
...
31 – 12 – 2019 23h50	0.722	<i>NaN</i>	...	1.0

Fonte: Autor.

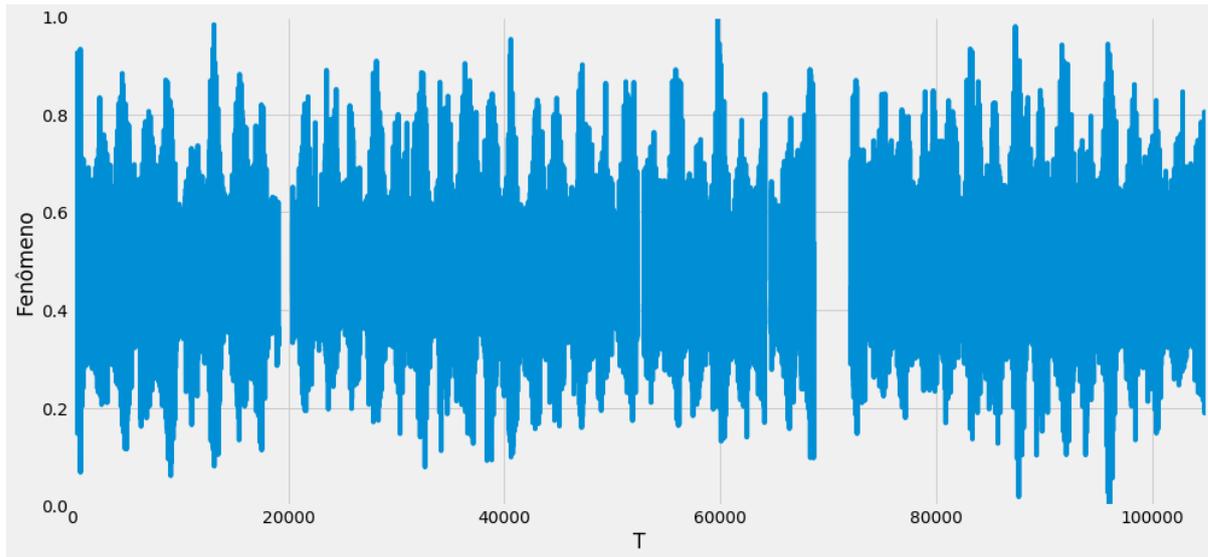
Apesar do grande avanço nos últimos anos em termos de captura, processamento e distribuição de dados em diferentes áreas, e tendo em vista que a referida rede de sensores encontra-se “*into the wild*”, ou seja, estão vulneráveis a toda sorte de intempéries, seria ingênuo pensar que os dados não apresentariam faltas. Deste modo, existia uma deficiência considerável, a ponto de ser necessário implementar algoritmos capazes de completar nossa base com certa fidelidade: cerca de 3,2 milhões de células estavam ausentes neste conjunto de dados, o que corresponde a aproximadamente 43% das células.

A título de exemplificação, a Figura 14 apresenta a distribuição para um fenômeno observado na bóia Bifurcação. É possível notar faixas de dados faltantes, tanto no ano de

²Vale destacar que o pesquisador responsável por fornecer os dados exigiu a sua confidencialidade. Assim, apesar do modelo ser exposto e aberto, os dados aqui representados servirão para ilustrar a estrutura da base a qual tínhamos em mãos e foi utilizada para o treinamento e teste do nosso modelo.

2018 (início das observações) quanto em 2019 (faixa mais larga, onde, provavelmente, o sensor sofreu algum dano). Apesar disso, é possível notar que na maior parte do período observável houve a aquisição dos dados.

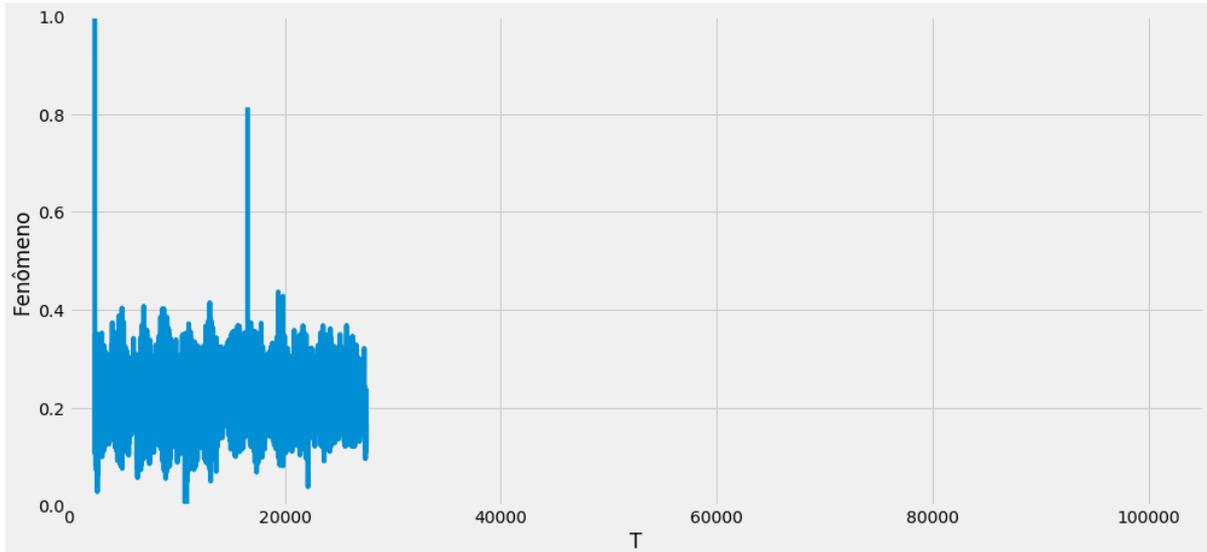
Figura 14: Distribuição para um dos fenômenos da bóia Bifurcação. Devido às restrições para divulgação dos dados, o fenômeno foi anonimizado tanto pela omissão do que ele representa quanto pela transformação dos valores.



Fonte: Autor.

De maneira totalmente oposta, a Figura 15, que apresenta a distribuição para o mesmo fenômeno, porém observado pela bóia B22, mostra que alguns desses sensores deixaram de medir estes fenômenos em grande parte do período (se não, em alguns casos, como o ilustrado na Figura 15, na maioria dele).

Figura 15: Distribuição para um dos fenômenos da bóia B22. Devido às restrições para divulgação dos dados, o fenômeno foi anonimizado tanto pela omissão do que ele representa quanto pela transformação dos valores.



Fonte: Autor.

5.3.2 Imputação dos Dados

Tendo sido explicitado o problema relativo aos dados faltantes da base, dois métodos distintos foram aplicados a fim de realizar a imputação dos dados.

1. **MICE (*Multiple Imputation by Chained Equations*)** [38, 39, 40]: este método funciona a partir do esquema de maximização da probabilidade *a posteriori* (ou MAP, na sigla em inglês). A ideia central consiste em utilizar as informações das distribuições *a priori* das variáveis faltantes junto com a relação entre elas. Inicialmente, completa-se a base a partir das distribuições das próprias variáveis. Na sequência, define-se uma destas como sendo a variável dependente e faz-se uma regressão tendo como variáveis independentes as demais variáveis. Repete-se este processo para todas as demais variáveis, atualizando os seus valores em cada rodada de regressão, até que os valores das variáveis deixem de alterar, ou seja, quando houver a otimização dos valores.
2. **Imputação via distribuição espaço-temporal:** suponha que consideremos uma bóia e um atributo esteja ausente em algum momento. Caso este atributo estiver presente para a maioria das bóias, imputamos a média delas quando a variância for menor do que um *threshold*; caso contrário, imputamos a mediana. Todavia,

se a maioria das bóias também tiverem este atributo ausente, tomamos a média deste atributo em relação ao tempo para as medições da bóia. Ou seja, no primeiro esquema é levado em conta a distribuição espacial do atributo, enquanto que no segundo, a distribuição temporal. O valor de *threshold* foi obtido através de várias tentativas, seguidas de amostragens de intervalos para definir, qualitativamente, se a imputação teve bom resultado ou não — o valor encontrado foi de 0.25.

Nossas experiências indicaram que a imputação via distribuição espaço-temporal levou a um desempenho um pouco melhor. Deste modo, relatamos, na sequência, os resultados com um conjunto de dados completo por este método.

6 RESULTADOS E CONCLUSÕES

Implementamos nossa GNN usando a biblioteca de redes de grafos do DeepMind¹, adaptando-o conforme necessário ao nosso modelo. Dadas as premissas de modelagem descritas na seção anterior, temos 2.415.367 parâmetros para treinar o esquema Não Local e 2.020.359 parâmetros no esquema Local. Os modelos foram treinados em um lote de 5k pontos de dados com *sliding window* de 6 pontos de dados para treinamento e 3 pontos de dados para validação (uma divisão de 67%/33% do conjunto de treinamento/teste). O otimizador Adam foi utilizado com *learning rate* igual a 1e-4. Os modelos foram treinados para 5 iterações. As arquiteturas de redes neurais usadas para as entidades tinham 5 camadas e 256 unidades/camada. A Tabela 2 resume esta definição.

Tabela 2: Hiperparâmetros e seus valores. Seus ajustes foram realizados através de *grid-search* durante a fase de validação.

Hiperparâmetro	Valor
<i>Learning Rate</i> (α)	1e-4
# camadas	5
# unidades/camada	256

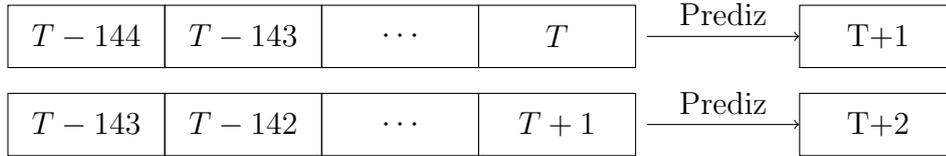
Fonte: Autor.

Em seguida, usamos o modelo para prever a corrente de água na bóia Bifurcação pelas próximas 24 horas de forma sequencial. Ou seja, usamos dados das últimas 48h para prever as medições para os próximos 20 minutos. Em seguida, alteramos nossa janela de tempo para incorporar essa previsão como se fosse uma medida e predizemos as medidas para o segundo intervalo de 20 minutos do dia. Continuamos dessa maneira até que todas as $24[h] \times 3[\text{medidas/h}] = 72$ medidas tenham sido previstas. Um exemplo de previsão de 2 etapas é mostrado na Figura 16. Como afirmado antes, dado que o objetivo é prever a velocidade da corrente de água para as 24h seguintes, são necessários 72 passos semelhantes aos mostrados na Figura 16.

Os testes indicaram que, para ambas as topologias, Local e Não Local, os modelos

¹https://github.com/deepmind/graph_nets.

Figura 16: Exemplo do processo de predição.



Fonte: Autor.

aprendidos são bem-sucedidos no rastreamento do comportamento dinâmico dos sinais. Em particular, as previsões foram capazes de ajustar o número de variações atuais no período de previsão — representadas pelas cristas e vales do sinal —, um comportamento não trivial, pois esse número varia em função da maré e das condições meteorológicas (vide Figuras 17 e 18). No entanto, ambos os modelos tiveram dificuldade em se ajustar aos valores de pico dos sinais. O erro quadrático médio é de 0.02 nós^2 para o modelo não local e de 0.10 nós^2 para o modelo local (o erro quadrático é medido subtraindo os valores observados e previstos ao longo do tempo).

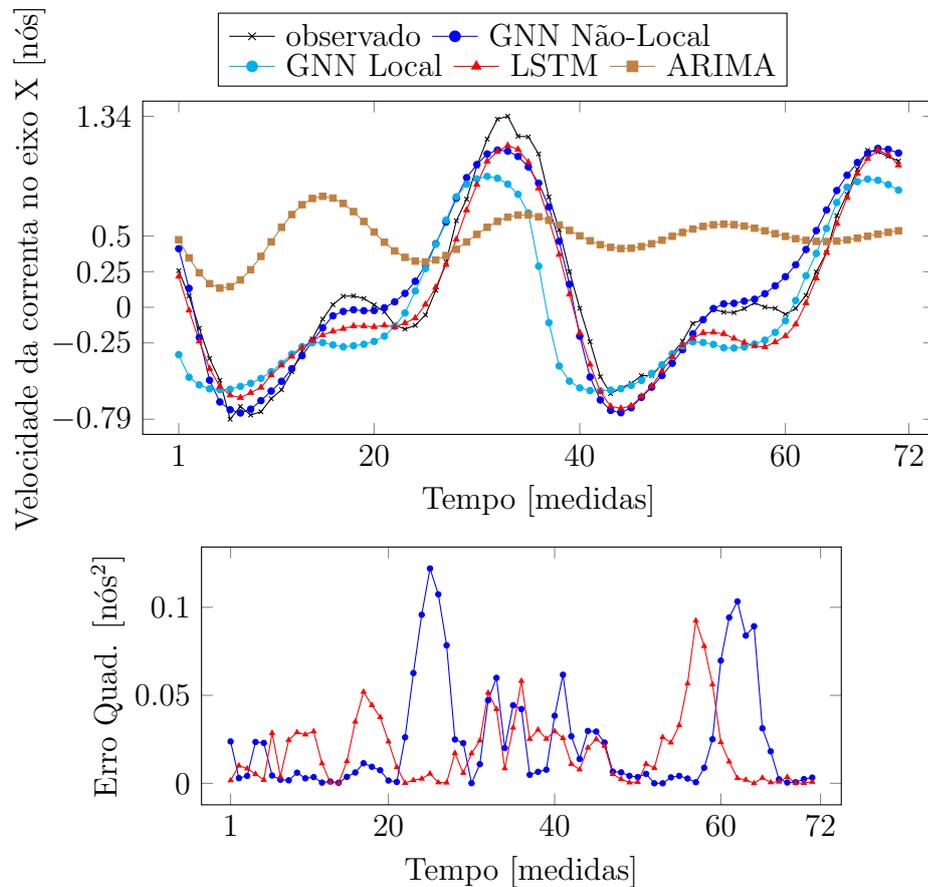
A Figura 17 mostra uma execução de predição, comparando nosso modelo com modelos *baseline* amplamente usados em predição de série temporal — especificamente, ARIMA [9] e LSTM [13]. Ambos os *baselines* foram treinadas em uma base de 20k dados e 50 iterações. Nosso modelo tem um desempenho significativamente melhor do que ARIMA e um pouco melhor do que o LSTM, mas com notável eficiência de dados — menos pontos de dados e iterações para treinar (cerca de um quarto dos pontos de dados). Comparando os dois modelos GNN, Não local e Local, os melhores resultados são obtidos com o primeiro, evidência que sustenta nossa hipótese de que um modelo poderia se beneficiar do compartilhamento e combinação de informações entre suas entidades. Dado que o *baseline* LSTM e o modelo GNN Não Local foram os melhores modelos, também apresentamos na Figura 17 o erro quadrático de ambos os modelos ao longo da janela de tempo. O resultado agregado dos modelos é apresentado na Tabela 3.

Tabela 3: Comparação entre os modelos a partir do erro quadrático médio (MSE, na sigla em inglês). Tanto GNN Não Local e LSTM obtiveram os melhores resultados, apesar do primeiro modelo ter maior eficiência de dados — necessita de menos dados para obter o mesmo resultado.

Modelo	<i>MSE</i>
GNN Não Local	0.02
LSTM	0.02
GNN Local	0.10
ARIMA	0.43

Fonte: Autor.

Figura 17: Resultados para uma janela de tempo selecionada, comparando os modelos e os *baselines*.

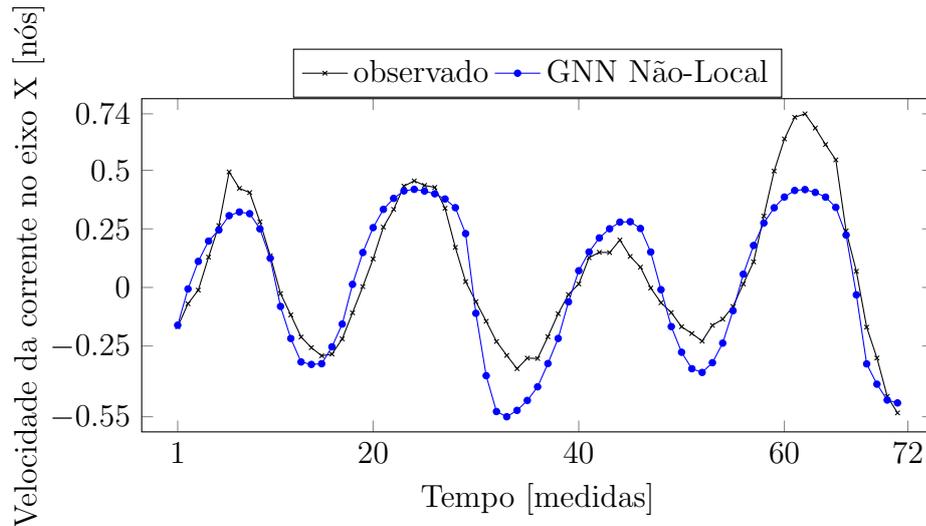


Fonte: Autor.

Os resultados obtidos e apresentados ao longo deste capítulo permitem concluir que a utilização de Redes Neurais de Grafos em problemas de previsão de séries temporais é um veículo adequado, tanto do ponto de vista da modelagem do problema, a qual passa pelo conhecimento holístico deste, quanto dos resultados práticos: há ganhos na acurácia, ao compararmos com modelagens clássicas e aquelas que não levam em conta o compartilhamento de informações, o modelo é mais eficiente em termos de consumo de dados e a forma exigida pela modelagem permite um melhor entendimento de como o aprendizado ocorre.

Vale destacar, ainda, que os resultados deste trabalho foram publicados na forma de um artigo para o evento intitulado *Symposium on Knowledge Discovery, Mining and Learning* (KDMiLe, 2020), evento conjunto com o *Brazilian Conference on Intelligent Systems* (BRACIS, 2020). O trabalho foi muito bem recebido pelos pesquisadores da área e obteve a 2^a colocação dentre os melhores *papers* do evento.

Figura 18: Resultados para outra janela de tempo.



Fonte: Autor.

Além disso, tendo sido vista a posição das GNNs nas técnicas neuro-simbólicas, foi possível alargar o rol de aplicações de modelos que as empregam. A hipótese de que a união entre os paradigmas conexionistas e de raciocínio simbólico, incorporado a partir de regras e restrições na arquitetura do modelo e das redes neurais, alavanca os sistemas de IA obteve mais uma evidência com os resultados colhidos a partir deste trabalho.

Todavia, é evidente que nosso sistema exige melhorias em certos pontos. Ambos os modelos de GNN obtiveram resultados insatisfatórios quando nas proximidades de picos do sinal original. Futuros trabalhos poderiam analisar a influência da arquitetura para este fenômeno, tendo em vista que, provavelmente, as estruturas totalmente conectada e desconectada (Não Local e Local, respectivamente) não representam o teto e o piso da performance desse modelo. Criar um modelo de GNN que consiga aprender, por exemplo, qual seria a melhor disposição de conexões entre os vértices pode ser um caminho a ser seguido. Outro modo, porventura, seria a reunião de diferentes modelos de GNN, treinados cada qual para uma tarefa específica na captura do comportamento da série temporal (e.g., tendência, sazonalidade, evento adverso). Assim, um *ensemble* de GNNs — semelhante à *framework* apresentada durante a revisão bibliográfica da computação neuro-simbólica — alavancaria ainda mais a performance do sistema.

Por fim, em se tratando de um problema de grande interesse econômico (predição de fenômenos críticos à atividade de navegação em regiões portuárias), seria interessante comparar os resultados obtidos pelo modelo GNN com aquele resultante de modelagem hidrodinâmica da região. Seria promissor uma solução totalmente guiada por dados que conseguisse resultados semelhantes, ou até mesmo melhores, do que aqueles obtidos a

partir de *softwares* matemáticos de modelagem hidrodinâmica, muitas vezes caros, rígidos e que exigem manutenções e/ou atualizações constantes, tendo em vista as mudanças físicas do local.

REFERÊNCIAS

- 1 DEVLIN, J. et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019.
- 2 FENG, M. et al. Applying deep learning to answer selection: A study and an open task. *CoRR*, abs/1508.01585, 2015. Disponível em: [⟨http://arxiv.org/abs/1508.01585⟩](http://arxiv.org/abs/1508.01585).
- 3 NG, A. *Why AI Is the New Electricity*. 2017. Disponível em: [⟨https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity⟩](https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity). Acessado em: 14 de mar. 2020.
- 4 WATSON Health: tecnologia a favor da medicina de precisão. 2016. Disponível em: [⟨https://pebmed.com.br/watson-health-tecnologia-favor-da-medicina-de-precisao/⟩](https://pebmed.com.br/watson-health-tecnologia-favor-da-medicina-de-precisao/). Acessado em: 14 de mar. 2020.
- 5 HAWKINS, A. J. *Inside Waymo's strategy to grow the best brains for self-driving cars*. 2018. Disponível em: [⟨https://www.theverge.com/2018/5/9/17307156/google-waymo-driverless-cars-deep-learning-neural-net-interview⟩](https://www.theverge.com/2018/5/9/17307156/google-waymo-driverless-cars-deep-learning-neural-net-interview). Acessado em: 14 de mar. 2020.
- 6 STROH, E.; MATHUR, P. Question answering using deep learning. In: . [S.l.: s.n.], 2016.
- 7 POMERLEAU, D. Alvin: An autonomous land vehicle in a neural network. In: *NIPS*. [S.l.: s.n.], 1988.
- 8 BOX, G. E. P.; JENKINS, G. *Time Series Analysis, Forecasting and Control*. USA: Holden-Day, Inc., 1990. ISBN 0816211043.
- 9 BROCKWELL, P. J.; DAVIS, R. A. *Time Series: Theory and Methods*. [S.l.]: Springer, 1987.
- 10 GARDNER JR, E. S. Exponential smoothing: The state of the art. *Journal of Forecasting*, v. 4, n. 1, p. 1–28, 1985. Disponível em: [⟨https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980040103⟩](https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980040103).
- 11 HARVEY, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. [S.l.]: Cambridge University Press, 1990.
- 12 BROWN, T. B. et al. *Language Models are Few-Shot Learners*. 2020.
- 13 HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, v. 9, p. 1735–80, 1997.
- 14 Data Science Academy. *Deep Learning Book Brasil Capítulo 51 – Arquitetura de Redes Neurais Long Short Term Memory (LSTM)*. Disponível em: [⟨http://deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory/⟩](http://deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory/). Acessado em: 09 de maio 2020.

- 15 NICKEL, M. et al. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, Institute of Electrical and Electronics Engineers (IEEE), v. 104, n. 1, p. 11–33, Jan 2016. ISSN 1558-2256. Disponível em: <http://dx.doi.org/10.1109/JPROC.2015.2483592>.
- 16 POLLETI, G. P.; COZMAN, F. G. Faithfully explaining predictions of knowledge embeddings. In: *Encontro Nacional de Inteligência Artificial*. [S.l.: s.n.], 2019. p. 1–12.
- 17 RUSCHEL, A. et al. Explaining completions produced by embeddings of knowledge graphs. In: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. [S.l.]: Springer International Publishing, 2019. p. 324–335. ISBN 978-3-030-29765-7.
- 18 BESOLD, T. R. et al. *Neural-Symbolic Learning and Reasoning: A Survey and Interpretation*. 2017.
- 19 GOERTZEL, B.; POTAPOV, A. *What are neural-symbolic AI methods and why will they dominate 2020?* 2020. Disponível em: <https://thenextweb.com/podium/2020/01/15/what-are-neural-symbolic-ai-methods-and-why-will-they-dominate-2020/>. Acessado em: 14 de mar. 2020.
- 20 RAEDT, L. D. et al. *From Statistical Relational to Neuro-Symbolic Artificial Intelligence*. 2020.
- 21 LAMB, L. C. et al. *Graph Neural Networks Meet Neural-Symbolic Computing: A Survey and Perspective*. 2020.
- 22 LIM, B.; ZOHREN, S. *Time Series Forecasting With Deep Learning: A Survey*. 2020.
- 23 WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. *Management Science*, INFORMS, v. 6, n. 3, p. 324–342, 1960. ISSN 00251909, 15265501. Disponível em: <http://www.jstor.org/stable/2627346>.
- 24 OORD, A. van den et al. *WaveNet: A Generative Model for Raw Audio*. 2016.
- 25 BOROVYKH, A.; BOHTE, S.; OOSTERLEE, C. W. *Conditional Time Series Forecasting with Convolutional Neural Networks*. 2018.
- 26 SALINAS, D.; FLUNKERT, V.; GASTHAUS, J. *DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks*. 2019.
- 27 VASWANI, A. et al. *Attention Is All You Need*. 2017.
- 28 SMYL, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, v. 36, n. 1, p. 75 – 85, 2020. ISSN 0169-2070. M4 Competition. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0169207019301153>.
- 29 WANG, Y. et al. *Deep Factors for Forecasting*. 2019.
- 30 RANGAPURAM, S. S. et al. Deep state space models for time series forecasting. In: BENGIO, S. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2018. v. 31, p. 7785–7794. Disponível em: <https://proceedings.neurips.cc/paper/2018/file/5cf68969fb67aa6082363a6d4e6468e2-Paper.pdf>.

- 31 SPERDUTI, A.; STARITA, A. Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, v. 8, n. 3, p. 714–735, 1997.
- 32 GORI, M.; MONFARDINI, G.; SCARSELLI, F. "a new model for learning in graph domains". In: "*IEEE International Joint Conference on Neural Networks*". [S.l.: s.n.], "2005". "2", p. "729–734".
- 33 PIANC MarCom. *Use of Hydro/Meteo Information for Port Access and operations*. [S.l.], 2012.
- 34 WU, M. et al. Prediction of short-term wind and wave conditions for marine operations using a multi-step-ahead decomposition-ANFIS model and quantification of its uncertainty. *Ocean Engineering*, v. 188, p. 106300, 2019.
- 35 SCARSELLI, F. et al. The graph neural network model. *IEEE Transactions on Neural Networks*, v. 20, n. 1, p. 61–80, 2009.
- 36 BATTAGLIA, P. W. et al. *Relational inductive biases, deep learning, and graph networks*. [S.l.], 2018.
- 37 WANG, X. et al. Non-local neural networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 7794–7803.
- 38 RAGHUNATHAN, T. et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, v. 27, p. 85–95, 2001.
- 39 BUUREN, S. V. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, v. 16, p. 219–242, 2007.
- 40 AZUR, M. et al. Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, John Wiley and Sons Ltd, v. 20, n. 1, p. 40–49, mar. 2011. ISSN 1049-8931.