

Arcabouço de Aprendizado Estatístico para Caracterização Assintótica de Canais sem Fio

Felipe Moraes Laburú, Juliano Siloto Assine, Flavio du Pin Calmon e José Cândido Silveira Santos Filho

Resumo— As ferramentas de inteligência artificial têm atraído grande atenção da comunidade de sistemas sem fio, que segue reformulando problemas nessa área por meio do arcabouço de aprendizado estatístico. Este trabalho propõe o seguinte problema: a partir de um conjunto de amostras do canal, estimar o comportamento estatístico de métricas essenciais de desempenho (p. ex., taxas de erro e de interrupção) dentro da faixa operacional do sistema. Uma solução eficaz é apresentada usando técnicas de aprendizado profundo e avaliada à luz da solução via inferência clássica, conceitualmente ótima mas inviável na prática. A análise é disposta em resultados numéricos representativos que revelam como cada parâmetro afeta a estimação.

Palavras-Chave— Aprendizado profundo em sistemas sem fio, canais de desvanecimento, análise de desempenho assintótica, estimador de desempenho.

Abstract— Artificial intelligence tools have attracted great attention from the wireless community, which keeps reformulating problems in this area through the statistical learning framework. This work proposes the following problem: from a set of channel samples, to estimate the statistical behavior of essential performance metrics (e.g., error rate and outage probability) within the system's operating range. An effective solution is presented using deep learning techniques and evaluated in light of the solution via classical inference, conceptually optimal but impracticable. The analysis is arranged into representative numerical results that reveal how each parameter affects estimation.

Keywords— Deep learning in wireless systems, fading channels, asymptotic performance analysis, performance estimator.

I. INTRODUÇÃO

A indústria de sistemas de comunicação sem fio está em constante evolução para atender às crescentes demandas dos consumidores. Enquanto a quinta geração (5G) ainda está sendo implantada, a comunidade científica já faz previsões para a próxima (6G), de que serão utilizados algoritmos de inteligência artificial para obter redes de comunicação mais heterogêneas, inteligentes e eficientes [1].

Nos últimos anos, houve uma proliferação no uso de técnicas de aprendizado de máquina (ML, do inglês *machine learning*) para sistemas sem fio, em toda a comunidade científica. As técnicas de ML têm sido utilizadas como candidatas

F. M. Laburú, J. C. S. Santos Filho e J. S. Assine fazem parte do Departamento de Comunicações, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, SP 13083-852, Brasil (e-mails: flaburu, candido@decom.fee.unicamp.br e jsiloto@dca.fee.unicamp.br).

F. P. Calmon faz parte da School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, Estados Unidos (e-mail: fcalmon@g.harvard.edu).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

aos problemas tradicionais de comunicação sem fio, como classificação de modulação, estimação de canal, detecção de sinais, e alocação de recursos [2]. Alguns pesquisadores vão além, ao proporem o problema de unificar todos os blocos clássicos do sistema utilizando um *autoencoder* [3], [4]. Nesse caso, em lugar de se otimizar cada bloco individualmente, otimiza-se o sistema inteiro, do receptor ao transmissor, em um único processo. Em contraste, neste trabalho não se usa ML para substituir ou melhorar blocos do sistema de comunicação. Em vez disso, incorporam-se conhecimentos específicos da área de comunicações em um problema de ML cuja solução pode ser usada em sinergia com os blocos convencionais do sistema. O contexto e o problema são detalhados a seguir.

Em sistemas sem fio, uma tarefa central é avaliar métricas de desempenho essenciais, como a probabilidade de interrupção (OP, do inglês *outage probability*) e a probabilidade de erro de símbolo (SEP, do inglês *symbol-error probability*), a fim de embasar a alocação de recursos como potência ou banda tal que se tenha uma comunicação mais eficiente. Em um cenário ideal, em que se sabe o valor do ganho do canal de desvanecimento, a avaliação é simples, reduzida à de um canal AWGN (do inglês *additive white Gaussian noise*) sujeito a um fator de escala conhecido. Em um cenário quase estático, também chamado de desvanecimento lento, em que o requisito de atraso da aplicação é menor que o tempo de coerência do canal, o ganho de canal é aleatório mas aproximadamente constante durante um bloco de transmissão. Nesse caso, tal ganho pode ser estimado com base em tons pilotos.

Entretanto, em um cenário de desvanecimento rápido, como em comunicações veiculares e em redes integradas espaciais e terrestres [5], [6], o canal pode variar dentro do intervalo de transmissão, ou de pacote, ou mesmo símbolo a símbolo. Essa rápida variação causa erros significativos em estimativas de canal através de tons pilotos, especialmente no final do pacote transmitido. Para diminuir esse efeito, é necessário usar tons pilotos mais frequentes, o que deteriora a eficiência de transmissão e aumenta a energia requerida por bit [7]. Alternativamente, pode-se adotar uma abordagem probabilística, em que se supõe um modelo estatístico para o canal e se estimam os parâmetros desse modelo por meio de métodos como máxima verossimilhança ou casamento de momentos. Todavia, essa abordagem está sujeita a erros na seleção do modelo ou ao aumento de complexidade ao utilizar modelos com maior número de parâmetros a fim de reduzir o viés.

Por outro lado, sabe-se que, para alta relação sinal-ruído (SNR, do inglês *signal-to-noise ratio*), regime operacional de grande interesse prático, a SEP e a OP podem ser aproximadas por uma reta em escala log-dB [8]. A SEP, \mathcal{P}_e , pode ser

representada em termos dos ganhos de codificação G_c e de diversidade G_d como

$$\mathcal{P}_e(1/\bar{\gamma}) \sim (G_c \bar{\gamma})^{-G_d}, \quad (1)$$

e a OP, \mathcal{P}_o , em termos dos ganhos de codificação O_c e de diversidade O_d , também chamado de diversidade de interrupção, como

$$\mathcal{P}_o(1/\bar{\gamma}) \sim (O_c \bar{\gamma})^{-O_d}, \quad (2)$$

em que $\bar{\gamma}$ é a SNR média e “ \sim ” denota equivalência assintótica, de modo que $a(x) \sim b(x) \Leftrightarrow \lim_{x \rightarrow 0^+} g(x)/h(x) = 1$. Dessa forma, basta conhecer os parâmetros G_c , G_d , O_c e O_d para se estimar o comportamento da SEP e da OP na região (de interesse) de alta SNR. Além disso, como demonstrado em [8], esses parâmetros dependem da função densidade de probabilidade (PDF, do inglês *probability density function*) do canal apenas em termos de sua assíntota na origem (*vide* Seção III). Ou seja, a PDF plena do canal é desnecessária. Surge assim a proposta deste trabalho: construir um arcabouço de aprendizado estatístico para estimar, a partir de um conjunto de amostras do canal sem fio, sua PDF assintótica na origem e, por meio dessa, estimar os ganhos de codificação e de diversidade de métricas como SEP e OP. O estimador proposto via aprendizado profundo será comparado com o estimador ótimo via inferência clássica apenas para fins conceituais, pois na prática a inferência clássica é inviável, já que a distribuição dos dados é desconhecida. No texto a seguir, $f_{(\cdot)}(\cdot)$ denota PDF, e $\Gamma(\cdot)$, a função gama.

II. MODELOS PROBABILÍSTICOS DE DESVANECIMENTO: PERSPECTIVA CLÁSSICA REVISITADA

Há um grande número de modelos estatísticos que descrevem o desvanecimento de pequena escala, associado aos multipercursos do sinal. Esse fenômeno é modelado por distribuições simples de um parâmetro, como a de Rayleigh; de dois parâmetros, como a de Nakagami- m , Hoyt, Weibull, e Rice; de três parâmetros, como a κ - μ , η - μ , e α - μ ; a distribuições sofisticadas com mais parâmetros [9]. Cada uma dessas distribuições foi concebida a partir de um subconjunto de aspectos físicos essenciais que governam a propagação de ondas de rádio: desbalanço de potência entre componentes em fase e em quadratura, composição gaussiana, linha de visada entre transmissor e receptor, clusterização de ondas, e não-linearidade do meio de transmissão. Cada novo aspecto físico considerado incorpora mais complexidade analítica ao modelo, evidenciada pelo seu número de parâmetros.

Sob uma perspectiva de engenharia, uma tarefa central é estimar, com base em amostras do canal, os parâmetros de um modelo de desvanecimento pré-selecionado. Essa pré-seleção é feita considerando observações da comunidade para aplicações e ambientes específicos, como a adequação do modelo Nakagami- m para comunicações veiculares em ambientes suburbanos [10]. A maioria dos estimadores propostos se baseia em escolher o parâmetro que torna os dados mais prováveis (método da máxima verossimilhança) ou em combinar certos momentos estatísticos (método do casamento de momentos). Ambos os métodos supõem que o modelo pré-selecionado representa a distribuição verdadeira do canal.

Esse arcabouço — escolher um modelo de desvanecimento e estimar seus parâmetros por meio de métodos estatísticos clássicos — é como dados de canais sem fio têm sido usados para aprender modelos de desvanecimento. Entretanto, a premissa de que o modelo escolhido representa de fato o modelo probabilístico dos dados de desvanecimento pode não ser verdadeira. Nesse caso, há uma perda da otimalidade da solução e de exatidão. Em particular, quando há um erro de especificação do modelo, sob certas condições de regularidade, a solução que maximiza a verossimilhança do falso modelo converge, na realidade, para o parâmetro que minimiza a entropia relativa entre a distribuição verdadeira e o modelo pré-definido [11].

Uma maneira de se contornar o problema da seleção de modelo é utilizar um modelo mais complexo, com maior número de parâmetros e, portanto, com maior poder de ajuste aos dados. Em [12], os autores mostram a superioridade do poder ajuste da distribuição α - η - κ - μ a medições de campo quando comparada a distribuições com menos parâmetros. Uma outra abordagem explorada na literatura consiste em pré-selecionar um subconjunto de modelos candidatos e selecionar o modelo mais apropriado a partir dos dados. No entanto, essas abordagens aumentam a complexidade e a dificuldade de projeto do sistema de comunicação. A próxima seção introduz uma alternativa promissora.

III. MODELOS PROBABILÍSTICOS DE DESVANECIMENTO: CARACTERIZAÇÃO ASSINTÓTICA EM ALTA SNR

O sinal recebido em um sistema sem fio sujeito a desvanecimento pode ser modelado por [8]

$$R_i = H_i S_i + Z_i, \quad (3)$$

em que i é um índice temporal, R_i representa o sinal recebido, H_i indica o ganho de canal do transmissor à antena do receptor, S_i é a mensagem transmitida, representada por um sinal independente e identicamente distribuído com energia E_s , e Z_i é um ruído aditivo gaussiano com partes real e imaginária independentes e de variância igual a $N_0/2$.

A SNR instantânea do sinal recebido é $\Gamma_i = |H_i|^2 \bar{\gamma} \triangleq \theta_i \bar{\gamma}$, em que define-se $\bar{\gamma} \triangleq E_s/N_0$. Assim, a SNR média é dada por $\bar{\gamma}$ quando $\mathbb{E}[\theta_i] = 1$. Como atestado em [8], sob condições suaves, a PDF do canal pode ser aproximada por $f_{\Theta}(\theta) \sim \alpha \theta^{\beta}$ quando $\theta \rightarrow 0^+$, e seus coeficientes assintóticos α e β governam o comportamento da SEP e da OP em regime de alta SNR ($\bar{\gamma} \gg 1$):

$$\mathcal{P}_e(1/\bar{\gamma}) \sim \frac{2^{\beta} \alpha \Gamma(\beta + \frac{3}{2})}{\sqrt{\pi} \beta + 1} (k \bar{\gamma})^{-(\beta+1)} \quad (4)$$

$$\mathcal{P}_o(1/\bar{\gamma}) \sim \frac{\alpha}{\beta + 1} \left(\frac{\gamma_{\text{th}}}{\bar{\gamma}} \right)^{\beta+1}, \quad (5)$$

em que γ_{th} é o limiar de interrupção escolhido e k é uma constante positiva associada ao esquema de sinalização, p. ex. $k = 2$ para BPSK. Conforme mencionado na Introdução, os coeficientes assintóticos do canal α e β parametrizam as métricas de desempenho. Portanto, estimar esses coeficientes é o objeto central do arcabouço de aprendizado proposto a seguir.

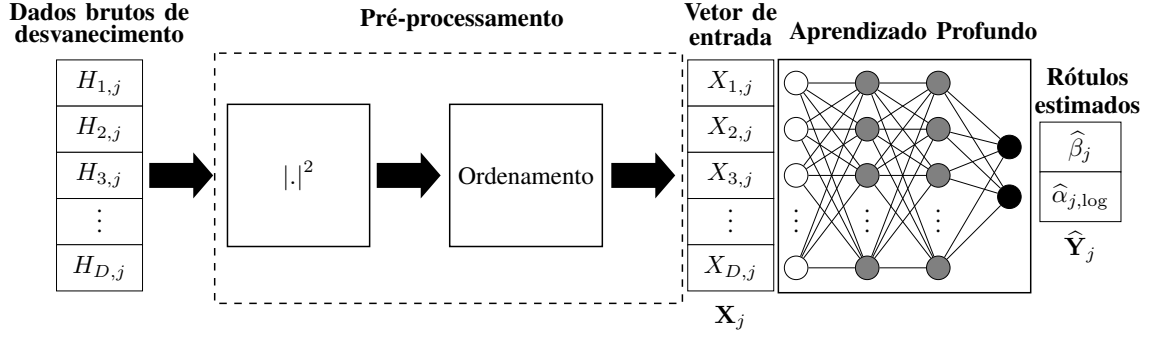


Fig. 1. Arcabouço proposto para inferência do modelo assintótico do canal sem fio por meio de aprendizado profundo.

IV. PROBLEMA PROPOSTO E SOLUÇÃO VIA APRENDIZADO PROFUNDO

Em vista de (4) e (5), propõe-se aqui aprender um modelo de inferência que observe D amostras do canal e forneça uma estimativa dos coeficientes α e β da PDF subjacente (desconhecida). Segue uma descrição do esquema proposto, ilustrado na Figura 1.

Construção do conjunto de dados: Neste trabalho, optou-se por trabalhar com dados sintéticos, para controlar aspectos primordiais do conjunto de dados, como o tamanho do vetor de entrada, o número de amostras e a distribuição estatística dos dados. Nesse cenário, é possível obter a solução ótima por inferência clássica para compará-la com a inferência do modelo de aprendizado, dado que os dados são gerados a partir de uma distribuição pré-definida. Outras vantagens são o baixo custo envolvido na geração desses dados quando comparado a dados de campo e a possibilidade de gerar um conjunto de dados rotulados de maneira controlada e precisa.

O conjunto de dados pode ser sintetizado a partir de modelos de distribuição de desvanecimento bem estabelecidos na literatura. Apesar de o arcabouço proposto ser geral e poder ser utilizado em uma ampla gama de cenários de propagação sem fio, a análise apresentada neste trabalho é ilustrada utilizando a distribuição Nakagami- m [8],

$$f_{|H|}(h) = \frac{2m^m h^{2m-1}}{\Gamma(m)\Omega^m} \exp\left(-\frac{mh^2}{\Omega}\right), \quad (6)$$

em que Ω é a potência média recebida e m é o parâmetro de desvanecimento. Essa distribuição foi escolhida por ser o modelo mais simples da literatura que produz ganhos de diversidade e de codificação arbitrários, permitindo criar rótulos α e β mais diversificados. Os coeficientes assintóticos se relacionam com o parâmetro m , supondo-se $\Omega = 1$, por

$$[\alpha, \beta] = \left[\frac{m^m}{\Gamma(m)}, m - 1 \right]. \quad (7)$$

A hipótese $\Omega = 1$ não traz perda de generalidade, visto que os dados de desvanecimento podem ser normalizados tal que $\Omega \approx 1$. Essa normalização simplifica a etapa de aprendizado e melhora os erros de estimação na etapa de teste.

O conjunto de dados sintetizados é composto por N amostras rotuladas, conforme $\mathcal{D} = \{\mathbf{X}_j, \mathbf{Y}_j\}_{j=1}^N$. Cada vetor de covariáveis \mathbf{X}_j é formado por D realizações $X_{ij} = |H_{ij}|^2$

do quadrado de uma variável aleatória Nakagami- m com parâmetro de desvanecimento M_j , denotado aqui em letra maiúscula por ser também tratado como variável aleatória — do contrário, não haveria incerteza sobre α e β . Na geração de cada amostra \mathbf{X}_j , o parâmetro M_j é sorteado a partir de uma distribuição uniforme entre 0.5 e 3.0, faixa típica para representar casos de desvanecimento severo a suave [8]. Note que um único sorteio de M_j é usado para gerar as D componentes de cada amostra \mathbf{X}_j , ou seja, supõe-se que, para cada amostra, as estatísticas do canal são aproximadamente constantes. O j -ésimo vetor de rótulo é composto por $\mathbf{Y}_j = [\beta_j, \log_{10} \alpha_j]$, em que α_j e β_j se relacionam com M_j como em (7). Escolheu-se utilizar o logaritmo de α pois a OP e a SEP são normalmente representadas em escala log-dB.

Vale ressaltar que, apesar de a abordagem em regime de alta SNR (média) se basear na condição em que o ganho (instantâneo) de canal se aproxima assintoticamente de zero [8], as amostras são aqui coletadas em todo o suporte da distribuição de probabilidade. A utilização de um limiar para coletar amostras apenas em uma região perto da origem reduziria o número de amostras no conjunto de dados e poderia implicar em uma deterioração do desempenho final de inferência. Além disso, seria necessário um maior tempo de observação para tentar se coletar as D amostras do vetor de entrada, o que iria contradizer a premissa de que as estatísticas do canal permanecem constantes durante este período.

Modelo de aprendizado: Nos últimos anos, as redes neurais artificiais profundas têm evoluído e atingido um sucesso sem precedentes em inúmeras áreas da ciência. Esse sucesso pode ser explicado pela sua capacidade de representar uma ampla variedade de funções, quando os parâmetros da rede são ajustados adequadamente ao conjunto de dados. Inspirado nesse sucesso e grande poder de ajuste, o modelo de aprendizado adotado neste trabalho é o de aprendizado profundo.

No modelo de rede profunda proposto, utiliza-se, por ser computacionalmente eficiente, como função de ativação a ReLU (do inglês *Rectified Linear Unit*) para as camadas escondidas, e a norma L_2 como função-custo. Além disso, utilizam-se duas camadas escondidas e 800 neurônios em uma rede densa. Esses dois hiperparâmetros foram escolhidos a partir de uma otimização por busca exaustiva. O tamanho do vetor de entrada é o mesmo do vetor de covariáveis \mathbf{X}_j . A saída não possui função de ativação e corresponde a um vetor de tamanho dois.

A inicialização dos pesos da rede é a Inicialização de He. Para se encontrar o peso ótimo da rede a partir dos dados sintéticos, utiliza-se o otimizador ADAM (do inglês *adaptive moment estimation*). Os hiperparâmetros do otimizador são ajustados conforme recomendado em [13]: taxa de aprendizado de 0.001, decaimento exponencial das estimativas do primeiro e segundo momentos com $\beta_1 = 0.9$ e $\beta_2 = 0.999$, e tamanho de *batch* de 32. É utilizada a estratégia de parada antecipada (*early stopping*) para prevenir sobreajuste (*overfitting*) no problema de regressão. Na etapa de treinamento, sintetizam-se N_{tr} amostras para o conjunto de treinamento e $N_{tr}/5$ amostras para o conjunto de validação. O conjunto de teste, com N_t amostras, é criado apenas na etapa de inferência.

Pré-processamento é a chave: Para melhorar o desempenho do aprendizado da rede, é proposta neste trabalho uma etapa crucial de pré-processamento. Nessa etapa, aplica-se o quadrado do módulo no vetor de dados de desvanecimento bruto contendo as D realizações de canal $\{H_{ij}\}_{i=1}^D$, para se obter o vetor de covariáveis \mathbf{X}_j apresentado anteriormente. Em seguida, as amostras são rearranjadas em ordem crescente, formando um novo vetor que contém as D estatísticas de ordem do vetor original \mathbf{X}_j . A motivação dessa última etapa é fornecer à rede neural a função de distribuição acumulada empírica. Essa etapa diminui de maneira significativa o erro de inferência, sendo uma etapa fundamental da solução de aprendizado estatístico. O esquema de inferência, após a rede ter sido apropriadamente treinada, é mostrado na Figura 1.

V. INFERÊNCIA CLÁSSICA COMO BASE DE COMPARAÇÃO

No problema de aprendizado supervisionado, o principal objetivo é obter uma função que faça uma inferência com erro de generalização tão pequeno quanto o erro (mínimo) da inferência clássica (ótima). Ao contrário do paradigma tradicional de aprendizado de máquina, em que a distribuição de probabilidade conjunta dos dados $f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ é desconhecida, neste trabalho ela é conhecida, pois os dados são sintetizados. No caso em questão (Nakagami- m), pode-se demonstrar que as funções de inferência que minimizam o erro médio quadrático para os rótulos são dadas por (demonstração omitida por limitação de espaço)

$$\hat{\alpha}_{\log}^*(\mathbf{x}) = \frac{\int_{0.5}^3 \log_{10} \left(\frac{m^m}{\Gamma(m)} \right) \frac{m^{Dm}}{\Gamma(m)^D} \exp[(m-1) \sum_{i=1}^D \ln(x_i) - m \sum_{i=1}^D x_i] dm}{\int_{0.5}^3 \frac{m^{Dm}}{\Gamma(m)^D} \exp[(m-1) \sum_{i=1}^D \ln(x_i) - m \sum_{i=1}^D x_i] dm} \quad (8)$$

$$\hat{\beta}^*(\mathbf{x}) = \frac{\int_{0.5}^3 (m-1) \frac{m^{Dm}}{\Gamma(m)^D} \exp[(m-1) \sum_{i=1}^D \ln(x_i) - m \sum_{i=1}^D x_i] dm}{\int_{0.5}^3 \frac{m^{Dm}}{\Gamma(m)^D} \exp[(m-1) \sum_{i=1}^D \ln(x_i) - m \sum_{i=1}^D x_i] dm}, \quad (9)$$

em que \mathbf{x} corresponde ao vetor de covariáveis, $\hat{\alpha}_{\log}^*(\cdot)$ é a função de inferência ótima para $\log_{10} \alpha$, e $\hat{\beta}^*(\cdot)$ é a função de inferência ótima para β . As integrais em (8) e (9) não apresentam solução em forma fechada, devendo ser resolvidas numericamente para se obterem os rótulos correspondentes.

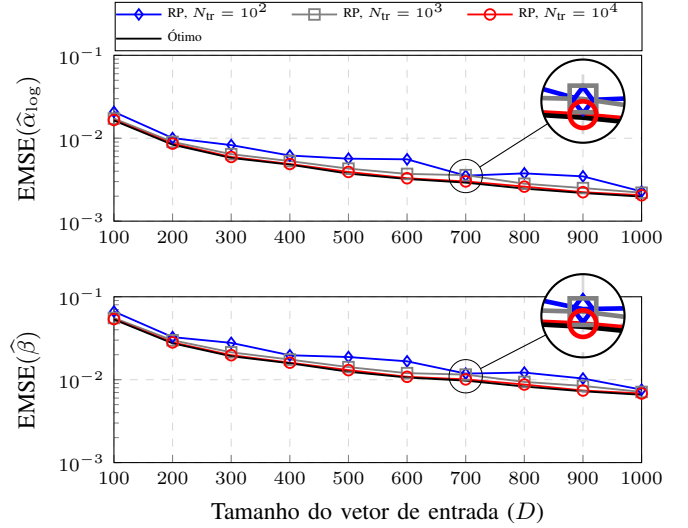


Fig. 2. Erro empírico: rede profunda (RP) versus inferência clássica (ótimo), em função do tamanho do vetor de entrada (D).

Vale ressaltar que, em uma situação real com dados de campo, a inferência clássica não é aplicável, pois a distribuição dos dados é desconhecida. Essa solução ótima é introduzida aqui apenas como base de comparação para o erro empírico da inferência da rede profunda.

VI. RESULTADOS NUMÉRICOS E DISCUSSÕES

Nesta seção, são investigados os resultados e o desempenho da inferência dos coeficientes assintóticos por rede profunda, comparada ao valor ótimo por inferência clássica. Analisa-se a influência do tamanho do vetor de entrada (D) e do número de amostras de treinamento (N_{tr}). A métrica de desempenho utilizada é o erro médio quadrático empírico (EMSE, do inglês *empirical mean squared error*), tipicamente usada em regressão, dos rótulos, avaliado sobre o conjunto de teste:

$$\text{EMSE}(\hat{\delta}) = \frac{1}{N_t} \sum_{k=1}^{N_t} (\delta_k - \hat{\delta}_k)^2, \quad (10)$$

em que δ denota α_{\log} ou β , conforme o caso, δ_k é o rótulo da k -ésima amostra, e $\hat{\delta}_k$ é o rótulo correspondente estimado pela rede ou pela função de inferência ótima. O conjunto de teste empregado nos experimentos contém $N_t = 10^3$ amostras. Por fim, utilizam-se os coeficientes estimados pela rede profunda para avaliar a curva de OP em um cenário de comunicações sem fio. É apresentado apenas OP, pois a SEP segue um padrão análogo conforme, Eqs. 1 e 2.

A Figura 2 mostra o EMSE para os rótulos $\hat{\alpha}_{\log}$ e $\hat{\beta}$ inferidos pela rede profunda e pela solução ótima considerando um conjunto de treinamento de tamanho 10^2 , 10^3 ou 10^4 para diferentes tamanhos do vetor de entrada. Observa-se que o EMSE diminui conforme se aumenta o tamanho do vetor de entrada. Essa dependência pode ser observada na própria solução ótima dada por (8) e (9). Além disso, observa-se que o EMSE da rede se aproxima muito estreitamente do valor ótimo à medida que se aumenta o tamanho do conjunto de treinamento, o que atesta a efetividade do esquema de aprendizado proposto.

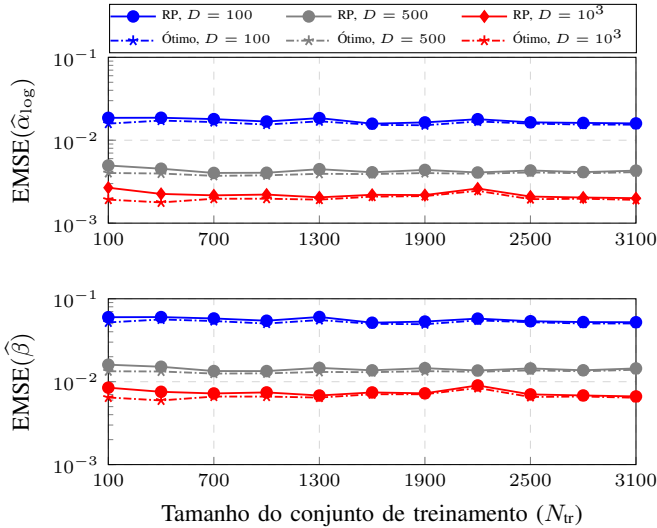


Fig. 3. Erro empírico: rede profunda (RP) versus inferência clássica (ótimo), em função do número de amostras do conjunto de treinamento (N_{tr}).

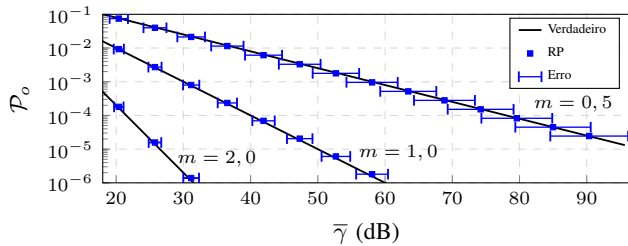


Fig. 4. Probabilidade de interrupção inferida pela rede profunda (RP).

A Figura 3 ilustra o erro médio empírico para os rótulos inferidos considerando um vetor de entrada de tamanho 100, 500 ou 1000, e variando-se o tamanho do conjunto de treinamento. Como no caso anterior, conforme se aumenta o tamanho do vetor de entrada, diminui-se o EMSE. Além disso, apesar de não ocorrer uma clara dependência entre o EMSE e N_{tr} , observa-se novamente que o erro da rede torna-se praticamente indistinguível do erro ótimo conforme N_{tr} aumenta.

Para avaliar a probabilidade de interrupção, parametrizada assintoticamente por (5), adota-se um limiar arbitrário $\gamma_{th} = 1$, e consideram-se os coeficientes $\hat{\alpha}$ e $\hat{\beta}$ estimados pela abordagem da rede profunda. Utilizam-se três valores do parâmetro m da distribuição Nakagami- m , para representar casos de desvanecimento severo ($m = 0,5$) a suave ($m = 2,0$). Os parâmetros da rede são $D = 10^3$, $N_{tr} = 10^4$ e $N_t = 10^3$. As estimativas de \mathcal{P}_o são representadas na Figura 4 em escala log-dB, na qual (5) aparece como uma reta. Mostram-se o valor médio de \mathcal{P}_o e barras de erro (de um desvio padrão) para o nível correspondente de $\bar{\gamma}$. Observa-se que a probabilidade de interrupção calculada pelos coeficientes assintóticos estimados pela rede profunda se aproxima do valor verdadeiro com excelente acurácia. Além disso, o erro aumenta conforme $\bar{\gamma}$ aumenta, por causa do efeito multiplicativo da SNR média nos erros de estimação dos coeficientes.

Apesar de omitido neste trabalho por restrições de espaço, foi observado que a utilização do bloco de ordenamento,

como na Figura 1, diminuiu o EMSE em uma ordem de grandeza. Além disso, foi feita uma comparação entre um modelo de aprendizado mais simples (regressão de Ridge) e a rede profunda proposta. Observou-se que, em comparação com a regressão de Ridge, a rede profunda apresenta um erro percentual relativo ao EMSE ótimo entre 15% e 30% menor para os coeficientes assintóticos α e β .

VII. CONCLUSÕES

Neste trabalho, é proposto um novo problema de aprendizado estatístico em sistemas de comunicação sem fio e uma solução correspondente utilizando redes profundas. Trata-se de uma investigação preliminar do problema, sujeita a uma análise conceitual em que se tem acesso à distribuição do canal e, portanto, aos estimadores ótimos. Resultados numéricos mostraram que o erro dos coeficientes estimados na abordagem de aprendizado de máquina se aproximou satisfatoriamente do erro mínimo via inferência clássica, e que esses coeficientes levam a uma estimativa da probabilidade de interrupção próxima à verdadeira. Pode-se considerar a abordagem apresentada como um candidato promissor para a avaliação de desempenho de sistemas sem fio em um contexto prático e realista, livre de premissas sobre a distribuição do canal.

REFERÊNCIAS

- [1] S. K. B. Sangeetha and R. Dhaya, "Deep learning era for future 6G wireless communications—theory, applications, and challenges", *Artificial Intelligent Techniques for Wireless Communication and Networking*, pp. 105-119, 2022.
- [2] Z. Guo *et al.*, "Multi-Agent Reinforcement Learning based Distributed Channel Access for Next Generation Wireless Networks", *IEEE Journal on Selected Areas in Communications*, 2022.
- [3] M. E. Morocho-Cayamcela, J. N. Njoku, J. Park and W. Lim, "Learning to Communicate with Autoencoders: Rethinking Wireless Systems with Deep Learning", *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 2020, pp. 308-311.
- [4] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer", *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563-575, Dec. 2017.
- [5] N. Hassan, C. Schneider, R. Thomä and G. Del Galdo, "Characterization of Slow and Fast Fading in V2I Channels for Smart Cities", *2021 15th European Conference on Antennas and Propagation (EuCAP)*, 2021, pp. 1-5.
- [6] H. Zhang, X. Huang and J. A. Zhang, "Adaptive Transmission with Frequency-Domain Precoding and Linear Equalization over Fast Fading Channels", *IEEE Transactions on Wireless Communications*, 2021, pp. 1-1.
- [7] K. Motosugi, K. Yonei, K. Maruta and C. Ahn, "Fast Fading Compensation by Frame Splitting and Decision Directed Channel Extrapolation for OFDM", *2019 25th Asia-Pacific Conference on Communications (APCC)*, 2019, pp. 131-135.
- [8] W. Zhengdao and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels", *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1389-1398, Aug 2003.
- [9] M. D. Yacoub, "The α - η - κ - μ Fading Model", *IEEE Trans. Antennas Propag.*, vol. 64, no. 8, pp. 3597-3610, Aug, 2016.
- [10] W. Viriyasitvat, M. Boba and H. Tsai and A. Vasilakos, "Vehicular Communications: Survey and Challenges of Channel and Propagation Models", *IEEE Veh. Technol. Mag.*, vol. 10, no. 2, pp. 55-66, May 2015.
- [11] C. D. Richmond and L. L. Horowitz, "Parameter Bounds on Estimation Accuracy Under Model Misspecification", *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2263-2278, May. 2015.
- [12] A. A. dos Anjos, T. R. R. Marins, R. A. A. de Souza and M. D. Yacoub, "Higher Order Statistics for the $\alpha - \eta - \kappa - \mu$ Fading Model", *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 6, pp. 3002-3016, June 2018.
- [13] D. P. Kingma and J. L. Ba. "Adam: A method for stochastic optimization". *arXiv preprint arXiv:1412.6980*, 2014.