

MEDIDAS DE EXATIDÃO E PRECISÃO EM CLASSIFICADORES BASEADOS EM RAZÃO DE VEROSSIMILHANÇA

Adelino Pinheiro Silva¹²

Jean Michel Barboza Mendonça¹³



RESUMO: O trabalho de elucidação de crimes reúne um grande conjunto de procedimentos e tarefas a serem executadas pela polícia judiciária. As fontes de informação que são exploradas durante a investigação têm naturezas diversas, englobando desde dados de pessoas e amostras de informação ou objetos. Cada informação tem seu valor, e a rastreabilidade e confiança na origem da informação é fundamental para o êxito da investigação. Para espécimes de informação específicas, como amostras biológica, arquivos digitais (imagem e voz), amostras biológicas ou impressões latentes, a determinação da fonte (ou origem) pode ser necessária para materializar um indício. Neste contexto, situa-se o presente trabalho, que tem como propósito descrever de forma acessível os princípios de exatidão e precisão aplicados em sistemas de comparação de amostras (e determinação de fonte) baseados em razão de verossimilhança. Esta apresentação é realizada através de uma revisão bibliográfica com exemplos computacionais hipotéticos fazendo uso de elementos ilustrativos.

Palavras-chave: Razão de verossimilhança, determinação de fonte, regressão logística, precisão, exatidão, classificadores.

MEASURES OF ACCURACY AND PRECISION IN CLASSIFIERS BASED ON LIKELIHOOD RATIO

ABSTRACT: The work of elucidating crimes brings together a large set of procedures and tasks to be carried out by the judicial police. The sources of information that are explored during the investigation are diverse in nature, ranging from data from people to samples of information or objects. Each piece of information has its value, and traceability and trust in the origin of the information is fundamental to the success of the investigation. For specimens of specific information, such as biological samples, digital files (image and voice), biological samples or latent moisture, determination of the source (or origin) may be necessary to materialize a commitment. In this context, the present work is located, which aims to describe in an accessible way the principles of accuracy and precision applied in sample comparison systems (and source determination) based on the likelihood ratio. This presentation is carried out through a literature review with hypothetical computational examples using illustrative elements.

Key words: Likelihood ratio, source determination, logistic regression, precision, accuracy, classifiers.

1 Introdução

A investigação de uma infração penal consiste em um processo de obtenção e processamento de informações que visa a inferir sobre a ocorrência (ou não) de um fato típico, ilícito e culpável. As questões básicas da investigação orbitam sobre o fato acontecido (ou seu resultado), sua autoria, a maneira como o crime foi executado (*modus operandi*) e suas circunstâncias. A resposta dessas questões

¹² Perito Criminal e doutor em Engenharia Elétrica. E-mails: adelinocpp@gmail.com, adelino.pinheiro@policiacivil.mg.gov.br.

¹³ Perito Criminal e bacharel em Engenharia de Minas, Ciências Contábeis e Direito. E-mails: jmichelbm@gmail.com, jean.mendonca@policiacivil.mg.gov.br.

permite à polícia judiciária a acertada formulação da hipótese investigativa e a consecução efetiva de seus fins institucionais. Sob uma ótica pragmática, o trabalho de polícia judiciária envolve a identificação dos infratores penais e das atividades criminosas por eles desenvolvidas e, a partir dessa elucidação, prover de elementos probatórios o patrono da pretensão acusatória (LEE & CHO, 2021).

No trabalho de elucidação de crimes, diferentes atividades de obtenção de informação são realizadas. Fontes de informação de naturezas diversas são identificadas e acessadas, permitindo-se a coleta de dados e amostras, de objetos ou de indivíduos, envolvidas direta ou indiretamente na infração penal investigada. A depender da espécie de fonte de informação, essas atividades de coleta de dados apresentam desafios específicos. Na seara da coleta de dados de objetos, destaca-se como um de seus principais obstáculos à aplicação de melhores práticas de coleta em ambientes de obtenção de dados mais sensíveis – e.g., gravações (de áudio e vídeo), objetos, aparelhos telefônicos, material biológico, documentos, entre outros (SILVA, 2020a).

O material coletado durante o curso da investigação encontra-se, muitas vezes, maculado por fatores interferentes e contaminantes (e.g., sujeito), ou está imerso em um oceano de ruído, por exemplo, um arquivo digital em um computador ou servidor, ou uma mensagem ou ligação em um aplicativo de conversa. Nesse contexto, a tarefa de determinação de fonte mostra-se necessária ou para automatizar a busca de uma informação ou para fundamentar uma circunstância relacionada ou crime, *i.e.*, para materializar um indício¹⁴.

A determinação de fonte ou origem é uma operação que busca comparar duas amostras¹⁵ e indicar se pertencem (ou não) a mesma classe (ou mesma origem), com base em características comuns entre elas (MEUWLY, 2006). São diversos os exemplos de procedimentos de determinação de fonte realizadas em uma investigação criminal: reconhecimento de pessoas; a comparação de

um vestígio biométrico (como uma impressão digital) colhido em local de crime ou presente em um gravador; a localização de dispositivos comunicação (*smatphones*); a comparação de material químico ou biológico; comparação de marcas de objetos (marcas de banda de rolagem de pneumáticos). A qualidade do indício depende da qualidade da comparação realizada, e a qualidade de uma comparação pode ser avaliada segundo dois critérios: precisão e exatidão (MORRISON, 2011).

Outrora, exames comparativos dessa natureza fundavam-se em dois princípios, o da individualidade e o da individualização. Segundo o princípio da individualidade, dois objetos podem ser indistinguíveis, porém, jamais idênticos. O princípio da individualização, por sua vez, pauta-se na ideia de que, se existem semelhanças suficientes entre dois objetos que possibilitem sua coincidência, esses objetos devem ter vindo da mesma fonte. Apesar de intuitivos, esses conceitos não são potencialmente falseáveis (ROBERTSON et al., 2016).

A aplicação dos princípios da individualidade e da individualização pode gerar uma comparação arbitrária uma vez que não existe um critério que indique o número (ou nível) de semelhanças (características convergentes) necessárias para uma individualização (SAKS; KOEHLER, 2008; ROBERTSON et al., 2016). O atual paradigma divide a determinação de fonte em duas tarefas. A primeira diz respeito à comparação propriamente dita, para avaliar o quão compatíveis são entre si, a fim de que sejam considerados de mesma fonte. A segunda é avaliar o alcance da correspondência, *i.e.*, qual a probabilidade de os materiais serem oriundos da mesma fonte ou de se cometer um erro na avaliação (SAKS & KOEHLER, 2008).

No atual paradigma, expressa-se o resultado do teste de comparação como uma razão de verossimilhança (LR – *likelihood ratio*). A razão de verossimilhança avalia estatisticamente

14 Vide definição de indício no Art. 239 do decreto-lei nº 3.689, de 3 de outubro de 1941.

15 As amostras comparadas na determinação de fonte, dependendo da tarefa, podem ser denominadas como de origem conhecida, ou padrão, e de origem desconhecida, ou questionada.

as diferenças e semelhanças na realização de uma comparação. A comparação é baseada em um banco de dados representativo, com reprodutibilidade. Os sistemas baseados em razão de verossimilhança apresentam uma abordagem transparente, quantitativa, célere; disponível ao contraditório; e menos suscetível às variáveis de confusão e menos influenciável por vieses de quem realiza a comparação (SILVA, 2020a).

Inobstante as vantagens dos sistemas baseados em LR, muitos atores envolvidos no sistema de investigação criminal têm dificuldade de interpretar seus resultados estatísticos. Desta forma o presente trabalho tem como objetivo descrever os fundamentos de uma comparação baseada em LR, bem como suas métricas de precisão e exatidão. Mais especificamente o trabalho busca, a partir de exemplos hipotéticos: ilustrar o funcionamento de classificadores baseados em LR; explicar o processo de calibração de um classificador; e a interpretação de resultados e métricas de exatidão e precisão.

A relevância do presente trabalho repousa no fato de que os atores da investigação criminal são os produtores de indícios por excelência e, em um primeiro momento, seus destinatários, uma vez que esses atores serão os intérpretes dos indícios, e suas conclusões pautarão a tomada de decisão no decorrer da investigação. Já, em um segundo momento, o destinatário principal é o Poder Judiciário cujas decisões precisam ser legitimadas a partir de conjunto probatório que engloba indícios de diferentes naturezas (MONTEIRO, 2007). Desta forma, divulgar os fundamentos metodológicos da LR permitirá que os diversos intérpretes dos elementos probatórios tenham uma visão crítica acerca do resultado dos exames de determinação de fonte.

Metodologicamente, os fundamentos dos sistemas de determinação de fonte baseados em LR serão apresentados a partir da revisão da literatura, com exemplos computacionais hipotéticos. Busca-se detalhar as diferentes terminologias e métricas utilizando rotinas desenvolvidas na linguagem *python*.

2 Desenvolvimento

2.1 Precisão e Exatidão

Uma das motivações que leva a explorar medidas de exatidão e precisão é o relatório do Conselho de Pesquisa Estadunidense (NRC; 2009, p. 23) que recomenda o desenvolvimento e o estabelecimento de “medidas quantitativas de confiabilidade e de precisão”. Esta recomendação alinha-se com o atual paradigma sobre a determinação de fontes, em que medidas de confiabilidade e precisão “devem refletir a prática do estado da arte em cenários de casos realistas, calculados, em média, a partir de uma amostra representativa” (NRC; 2009, p. 23; SAKS; KOEHLER, 2008).

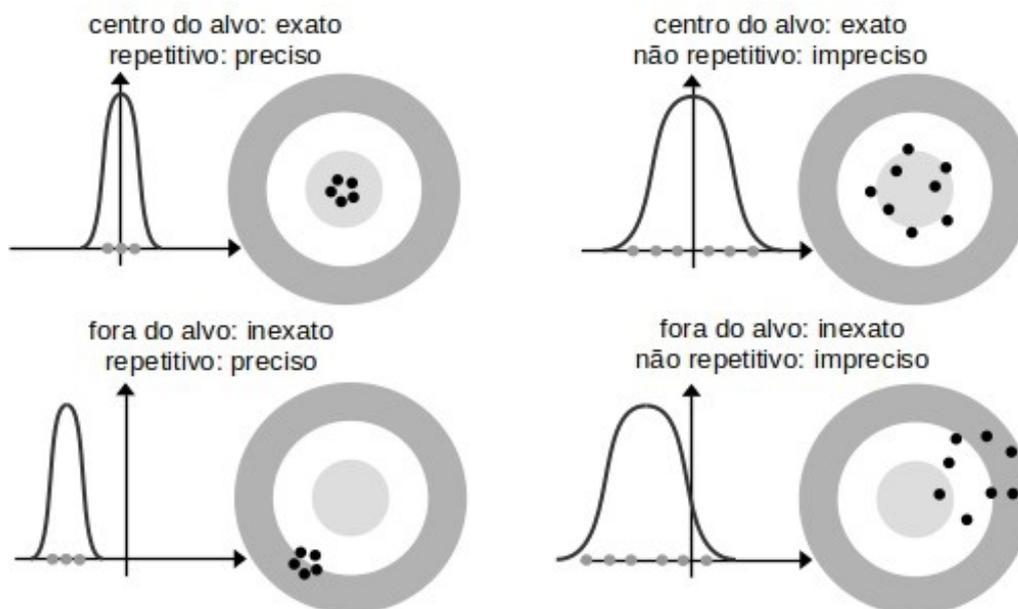
Segundo o vocabulário internacional de medição (IVO; 2012) a exatidão¹⁶ é uma medida de proximidade e concordância entre uma grandeza medida e o valor da verdadeira grandeza. Por outro lado, a precisão (*precision*) indica a proximidade da concordância (repetibilidade) dos valores de uma grandeza obtidos por repetidas medidas que foram replicadas no mesmo objeto/amostra (ou em semelhantes) sob condições específicas.

Uma forma adicional de compreender os conceitos, a exatidão é a capacidade de um sistema de aproximar-se do valor esperado, i.e., o valor do desvio que uma medida apresenta do seu valor real. Por outro lado, a precisão indica o quanto uma medida se dispersa em torno do valor medido. Associando com medidas da estatística descritiva, a exatidão está associada ao valor da tendência central amostra (e.g., média) e a precisão ao valor da dispersão (e.g., desvio padrão).

A Figura 1 apresenta de forma ilustrada os conceitos de precisão e exatidão através de uma curva de distribuição e um alvo bidimensional. Na imagem tem-se na porção superior esquerda um resultado preciso e exato, onde as ocorrências acontecem no centro do alvo (próximo ao valor médio) e concentrados (baixa dispersão). As demais apresentam variações, na porção superior direita os valores aparecem centrados, mas

dispersos (exato, mas não preciso). Na porção inferior, valores exatos, mas não precisos (à esquerda) e valores não precisos e não exatos (à direita).

Figura 1 – Representação gráfica dos conceitos de precisão e exatidão. Na porção superior esquerda um resultado preciso e exato, onde as ocorrências acontecem no centro do alvo (próximo ao valor médio) e concentrados (baixa dispersão). As demais apresentam variações, na porção superior direita os valores aparecem centrados, mas dispersos (exato, mas não preciso). Na porção inferior, à esquerda uma ocorrência precisa (pouca dispersão) mas não exata (longe do valor real) e à direita valores dispersos e tendência central distante do valor real.



Fonte: Adaptado de Melo (2014).

2.2 Determinação de Fonte por Razão de Verossimilhança

A determinação de fonte é tarefa de reconhecimento de padrões. O reconhecimento de padrões é o ato de medir os dados disponíveis e realizar uma ação baseada na classe (ou categoria) que é definida pela diferença entre os dados (Duda, Hart, Stock; 2001). Uma classe pode ser definida como um agrupamento de unidades que possuem características¹⁷ comuns entre si. Um ponto importante é que o uso de classificadores na determinação permite quantificar um agrupamento, medindo um grau de semelhança momentâneo. A classificação por

grau de semelhança não é definitiva e não significa identificar ou individualizar.

Um exemplo de classificação está na determinação de classe¹⁸. Suponha que uma planta industrial deseja separar, em um lote de pescado, as unidades (peixes) que são salmão das demais unidades. Na planta industrial as unidades são dispostas em uma esteira e fotografadas. Se existirem características que podem distinguir estas classes (salmão e demais peixes), logo é possível desenvolver um classificador (modelo matemático) para realizar a distinção. O modelo do classificador baseia-se no valor das características de cada classe.

¹⁷ Uma característica pode ser definida, no conceito de reconhecimento de padrões, como uma propriedade mensurável (e quantificável) de um objeto ou fenômeno (Bishop; 2006).

¹⁸ Exemplo adaptado do capítulo 1 por Duda, Hart e Stock (2001). O exemplo tem caráter explicativo com foco nos conceitos de classificação, características, erro, decisão e razão de verossimilhança. Isto posto, alguns passos e técnicas que poderiam melhorar a classificação são ignorados visando tornar o exemplo mais compreensível para o público.

Ainda no exemplo, suponha-se um lote de peixes já classificados (classe conhecida) que serão utilizados para o treinamento (calibração) do classificador. A classe alvo (o salmão) apresenta características, por exemplo, o comprimento e o brilho, que podem distinguir a classe. A Figura 2 apresenta dois histogramas com as ocorrências de comprimento e brilho de salmão e de outros peixes. Nas imagens cada coluna indica o número de ocorrências obtidas das classes para o valor da característica no eixo horizontal. Na imagem da esquerda (Figura 2a) a característica é o comprimento do peixe e na direita o brilho (Figura 2b). A linha vertical indica o limiar de decisão. Na Figura 2 nota-se que as classes apresentam diferentes distribuições, entretanto, existe uma região de sobreposição.

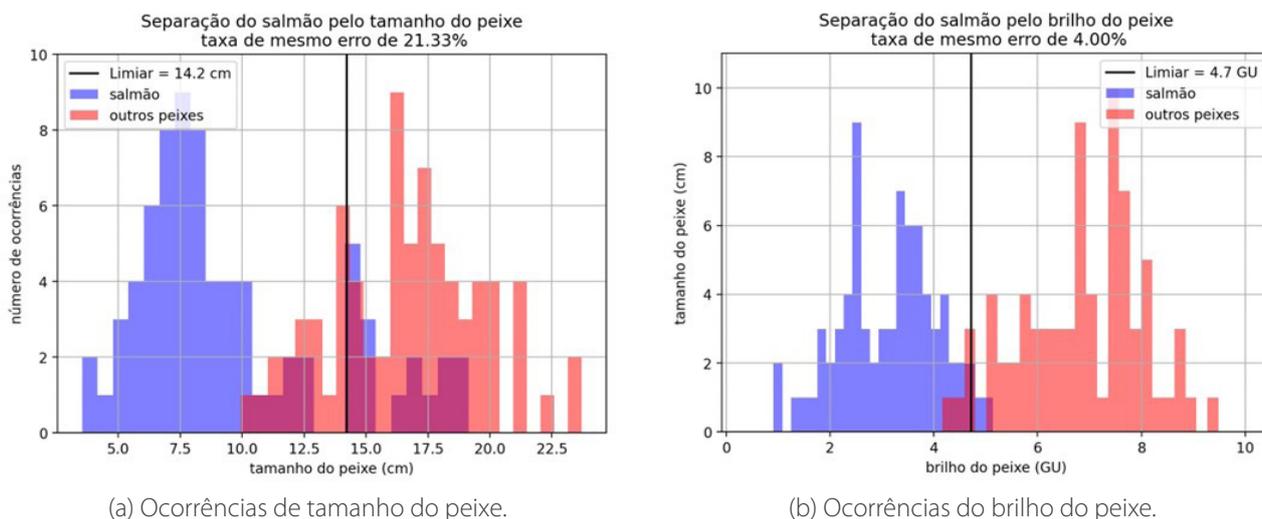
A partir de um valor de limiar é possível separar os peixes pelo tamanho e pelo brilho. O valor do limiar de decisão é obtido a partir de amostras de treinamento de salmão e outros peixes. O limiar de decisão é o valor, de comprimento e de brilho, que divide as classes

e a partir do qual decide-se de qual classe são as características. Nota-se, nos dois casos, que o limiar de separação não distingue perfeitamente as unidades de salmão (em azul) dos demais peixes (em vermelho). Como a distinção não é perfeita, existem erros associados.

Existem dois tipos de erro que podem ocorrer: uma unidade ser classificada como salmão quando se trata de outro peixe (falso positivo ou erro do tipo I); ou não ser classificada como salmão quando se trata de um salmão (falso negativo ou erro do tipo II), como indicado na Tabela 1.

Ao alterar o valor do limiar de decisão (deslocar para valores maiores ou menores) muda-se as taxas de erros (do tipo I e do tipo II). Ao se realizar uma classificação pode-se optar por reduzir um tipo de erro em relação ao outro. Por exemplo, deseja-se escolher o limiar de decisão de forma que as unidades classificadas como salmão sejam (em grande parte) salmão, não importando se unidades de salmão forem classificadas como outros peixes. Neste caso, deseja-se reduzir o falso positivo.

Figura 2 – Histograma apresentando a distribuição de valores medidos dos pescados. Em azul tem-se as ocorrências da classe “salmão” e em vermelho os “demais peixes”. A linha vertical preta indica o limiar de separação com menor taxa de mesmo erro.



Fonte: Os autores.

Nos exemplos das figuras 2 e 3 escolheu-se como limiar de decisão o valor que minimiza a taxa de mesmo erro (EER - *equal error rate*), i.e., o limiar admite taxas iguais de falso positivo e falso negativo. Porém, o limiar de decisão pode ser escolhido associando-se custos aos erros. Por exemplo, se trocar um salmão por outro peixe gerar mais custos, pode-se ponderar os erros e minimizar os custos.

Tabela 1 – Denominação dos tipos de erro que podem ocorrer em uma classificação.

		Condição real	
		Outro peixe	Salmão
Decisão ou classificação	Outro peixe	Verdadeiro Negativo	Falso negativo Erro do tipo II
	Salmão	Falso positivo Erro tipo I	Verdadeiro positivo

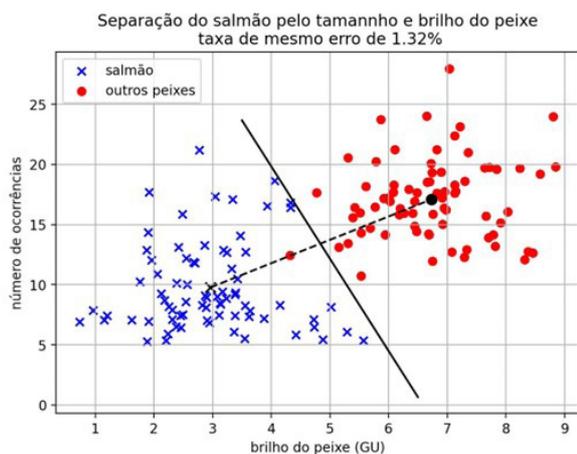
Fonte: Os autores.

Uma das formas de melhorar a classificação é combinar as características, gerando um espaço latente em que seja possível separar as classes¹⁹. A Figura 3a apresenta um plano de separação por tamanho e brilho. Na imagem, cada “x” (em azul) indica uma unidade de salmão e cada “o” (em vermelho) um peixe que não é salmão. Este plano é o espaço latente²⁰ das classes com o limiar de decisão na linha preta contínua. Os “centros” das classes são marcados em preto e foram calculados pela média dos valores das características. A linha contínua preta indica o limiar (vetor) de separação entre as classes²¹. As unidades de um lado da linha são classificadas como “salmão” e do outro lado

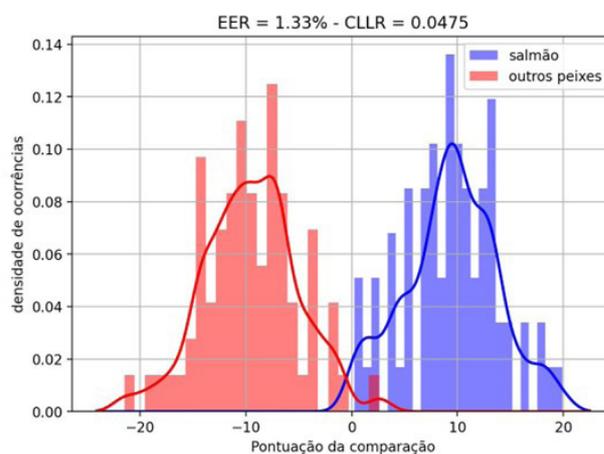
como “outros peixes”.

A linha pontilhada preta indica a direção normal à linha de separação. A partir das características é possível estabelecer uma pontuação/estatística de classificação (*score*). No exemplo, pode-se considerar uma pontuação à distância (paralela à linha pontilhada) entre uma unidade e a linha de separação (linha contínua preta). O processo obtenção dos valores de “tamanho” e “brilho” é denominado extração de características. Enquanto o cálculo da distância no plano é a tarefa de análise, tratamento e transformação.

Figura 3 – Separação dos peixes no plano “tamanho” vs. “brilho”. Em azul as unidades da classe “salmão” e em vermelho a classe “demais peixes”. A linha preta a direita indica uma linha de separação e a esquerda a distribuição de distâncias (paralelas ao eixo pontilhado) das unidades até linha.



(a) Espaço latente das variáveis apresentado como o plano de separação das classes.



(b) Aproximação por densidade das ocorrências de distância.

Fonte: Os autores.

19 Naturalmente existem várias formas de combinar características para gerar um espaço latente. Diferentes técnicas de análise tanto lineares – e.g. como análise fatorial conjunta –, ou não lineares – e.g. por análise de *kernel* ou características de gargalo. Entretanto, este trabalho ate-se aos fundamentos da classificação visando a acessibilidade dos principais conceitos.

20 O espaço latente é formado pela medição das características, de forma que cada característica é uma dimensão desse espaço. Por exemplo, se utilizarmos apenas o tamanho do peixe o espaço possui uma dimensão. No exemplo o espaço é um plano “tamanho” vs. “brilho” de duas dimensões dividido por um elemento geométrico de uma dimensão abaixo, ou uma reta. Um classificador que utiliza N características possui um espaço de N dimensões que é separado por um hiperplano de N-1 dimensões.

21 A linha de separação foi calculada utilizando máquina de vetor de suporte (SVM – *support vector machine*). Este limiar poderia ser traçado de diferentes maneiras no espaço latente inclusive para minimizar o erro – e.g. redes neurais, lógica fuzzy –, entretanto é importante ressaltar que o presente trabalho avisa a acessibilidade dos principais conceitos.

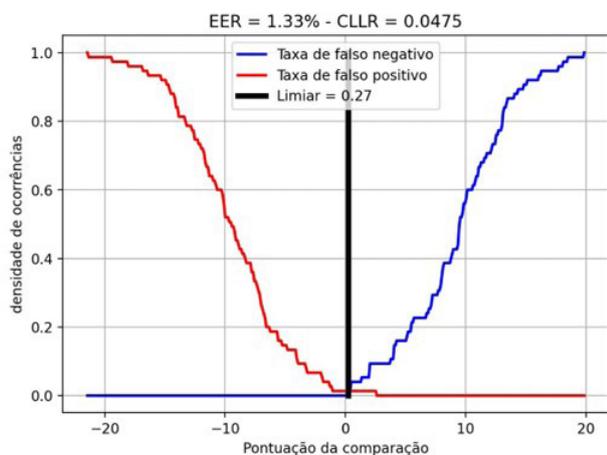
A partir dos dados de treinamento pode-se assumir que o quanto mais distante da linha contínua preta mais “segura” será a decisão, se a unidade pertence ou não aquela classe. Na Figura 3b é apresentada a distribuição de ocorrência (número de ocorrências dividido pelo número de unidades) das distâncias de cada unidade, em azul a classe salmão e em vermelho a classe demais peixes. A linha contínua indica a aproximação da densidade de probabilidade²². No eixo horizontal tem-se a distância medida positiva esquerda da linha de divisão e negativa à direita.

É no espaço latente, definido pelo conjunto de características que são utilizadas para classificação, que a decisão sobre a classe é tomada. Quanto mais características são utilizadas, maior o número de dimensões²³. Utilizando

a distância da linha de separação como uma estatística d_p de decisão para duas classes (“salmão” e “outros peixes”) pode-se variar o limiar de decisão e verificar quais os erros associados.

O valor do limiar λ_0 indicado na Figura 4a (linha vertical preta) marca o EER (taxa de mesmo erro) que minimiza as taxas de falso positivo e falso negativo. Por exemplo, observando a Figura 2a, se um peixe apresentar um tamanho superior a 14,2 cm é mais provável que não seja um salmão. Neste caso o limiar, valor a partir do qual eu tomo a decisão é $\lambda_0 = 14,2$. Para uma característica (e.g. tamanho) limiar é um ponto que divide a dimensão em duas semirretas. Na Figura 3a o limiar é uma reta, que separa o plano (“tamanho” vs. “brilho”) em dois semiplanos.

Figura 4 – Avaliação da distinção entre as classes. A curva *tippet* indica as taxas de falso positivo (em azul) e falso negativo (em vermelho) de acordo com a seleção do limiar no eixo horizontal. A regressão logística (em verde) apresenta a probabilidade da amostra pertencer a classe “salmão” de acordo com a medida de distância.



(a) Curva tippet apresentando limiar de decisão no eixo horizontal. A linha preta apresenta o limiar de mínimo EER.



(b) Regressão logística indicando a probabilidade de pertencer a classe “salmão” a partir da distância medida e a amostras de calibração na vertical.

Fonte: Os autores.

No exemplo, considera-se como falso positivo o resultado da classificação de uma unidade de outro peixe como salmão e como falso negativo a classificação de outro peixe como salmão (vide Tabela 1). Ao variar o valor do limiar de decisão (distância da linha de divisão) pode-se computar, para cada limiar, a taxa (percentual) de ocorrências de falso positivo e de falso negativo. A Figura 4a apresenta na curva azul a taxa de falso negativo e a vermelha a taxa de falso positivo como o valor do

22 A curva foi aproximada usando aproximação de densidade por *kernel* (núcleo) gaussiano.

23 Quando dados atingem uma alta dimensionalidade o tratamento e a análise podem ser dificultados por diferentes fatores, levando a fenômenos conhecidos como a maldição da dimensionalidade (*curse of dimensionality*). Existem diferentes técnicas de redução de dimensionalidade que podem ser utilizadas – e.g. análise de componentes principais –, entretanto descrever tais técnicas não faz parte do escopo deste trabalho.

limiar no eixo horizontal. Nota-se que a medida que desloca-se o limiar de decisão entre as classes para a direita a taxa de falso positivo (curva azul) aumenta enquanto a taxa do falso negativo (curva vermelha) tende a zero. Por outro lado, ao deslocar o limiar de decisão para esquerda a taxa do falso negativo (curva vermelha) aumenta enquanto a taxa de falso positivo (curva azul) tende a zero.

A partir da estatística d_p (distância de uma medida no plano “tamanho” vs. “brilho” até a linha de separação)²⁴ pode-se realizar uma regressão logística. Esta regressão define uma equação que aproxima a diferença entre as distribuições da Figura 3b por uma curva de probabilidade a partir do valor de d_p .

A Figura 4b apresenta a regressão logística dos dados da Figura 3b. Na imagem, os pontos indicam as ocorrências da distância d_p das classes, sendo que os pontos azuis no eixo horizontal indicam as unidades de salmão (a maioria a direita do limiar $\lambda_0 = 1,2$) e os pontos vermelhos indicam as unidades de outros peixes. A curva verde indica no eixo vertical a probabilidade estimada de uma unidade ser da classe salmão para uma distância d_p . Nota-se que à esquerda do limiar (valor no eixo horizontal) a probabilidade (curva verde) é próxima de zero, enquanto à direita do limiar a curva tende a um.

Ao avaliar a regressão logística, obtêm-se a uma probabilidade condicionada de uma unidade do lote de pescado ser da classe salmão a partir da informação d_p . Desta forma, define-se a verossimilhança da classe w_0 para a medida (informação) d_p , pela equação da regressão logística como

$$\mathcal{L}(w_0|d_p) = \frac{1}{1+e^{\beta_0+\beta_1 d_p}} \quad (1)$$

Na Equação 1 w_0 representa a classe “salmão” e d_p é a distância da linha divisória no plano latente formado pelos eixos de tamanho e brilho do peixe

(vide Figura 3a). Os valores β_0 e β_1 definem a forma da curva²⁵ da Figura 4b. A razão de verossimilhança (*likelihood ratio* - LR) é definida como

$$LR = \frac{\mathcal{L}(w_0|d_p)}{\mathcal{L}(w_1|d_p)} \quad (2)$$

Em que w_1 representa a classe “outros peixes”. Por se tratar de uma razão de probabilidades o valor de LR é limitado entre zero e infinito. Como o problema é dicotômico, a verossimilhança da classe w_1 para a medida (informação) d_p é complementar.

Na definição do problema as classes são dicotômicas. Esse fato implica que suas verossimilhanças são complementares. O valor de LR indica, quantas vezes mais provável uma unidade pertencer a classe “salmão” frente às informações disponíveis de medida e calibração. Para um valor de $LR > 1$ a informação colabora para a hipótese de a unidade pertencer à classe w_0 , enquanto valores de $LR < 1$ colaboram para a hipótese contrária.

O raciocínio da distinção entre duas classes pode ser extrapolado para a detecção de fonte. Ao comparar as características de duas medidas, uma extraída de uma fonte conhecida (denominada como padrão) e outra de origem desconhecida (denominada questionada), pode-se levantar a hipótese destas características serem oriundas da mesma fonte. Neste caso, o treinamento, teste, validação e calibração tentariam separar duas classes, as de mesma origem w_0 e as de origem diferente w_1 . Com um raciocínio parecido com o do exemplo, mas com uma modelagem mais complexa, desenvolve-se uma pontuação (*score*) para distinguir comparação de características e calcular um valor de LR.

A partir destes conceitos pode-se definir um classificador como um método algorítmico, de etapas sequenciais, que pode distinguir entre duas ou mais classes a partir de suas características. Um classificador, por exemplo, pode ser utilizado

24 No plano definido pelo tamanho (eixo y) vs. brilho (eixo z), a distância d_p entre um ponto (y_0, z_0) e a reta $ay + bz + c = 0$ – que separa as classes –, é $d_p = |ay_0 + bz_0 + c| / \sqrt{a^2 + b^2}$.

25 Gostaria de colocar para o leitor que a apresentação da Equação 1 é ilustrativa e que os parâmetros da regressão logística (β_0 e β_1) são obtidos a partir de métodos numéricos de otimização.

para classificar, a partir de características extraídas de duas amostras, fornecer um rótulo como de mesma origem ou de origem diferente (Suárez & García, 2009).

Nesta seção foi apresentado um exemplo de como características (medidas quantitativas) extraídas de uma população de interesse podem ser utilizadas para obter um modelo de classificador e calcular a LR que provém informação para a tomada de decisão. Nas próximas seções serão apresentadas formas de como estimar a exatidão e a precisão de resultados de classificadores baseados em LR .

2.3 Exatidão na Razão de Verossimilhança

Na tarefa de reconhecimento de padrões, mais particularmente em duas classes dicotômicas, a exatidão pode ser entendida como a propriedade do modelo de classificação em indicar corretamente as classes. Ao realizar uma classificação sobre a determinação de fonte de suas amostras, o valor mensurável desta classificação é o LR . Esse valor indica o quanto aqueles dados analisados (no exemplo, brilho e tamanho do peixe) oferece de informação para decidir entre as amostras serem da mesma fonte ou de fontes diferentes. Um modelo de classificação²⁶ será mais exato quanto maior a concordância entre uma classe medida (estimada a partir do modelo) e o valor verdadeiro da classe. Em outras palavras, uma classificação mais exata apresenta menores valores de erro, do tipo I e do tipo II (vide Tabela 1).

Uma métrica desenvolvida para avaliar a exatidão de um modelo de classificação foi o custo

$$C_{LLR} = \frac{1}{2} \left(\frac{1}{N_{MO}} \sum \log_2 \left(1 + \frac{1}{LR_{MO}} \right) + \frac{1}{N_{OD}} \sum \log_2 (1 + LR_{OD}) \right). (3)$$

Um classificador com boa exatidão o valor de C_{LLR} se aproxima de zero. Um classificado para determinação de fonte de alto desempenho fornecerá valores de $LR_{MO} \gg 1$ e $LR_{OD} \ll 1$. A Equação 3, que pode ser um pouco complexa de analisar, indica que quanto maiores os valores

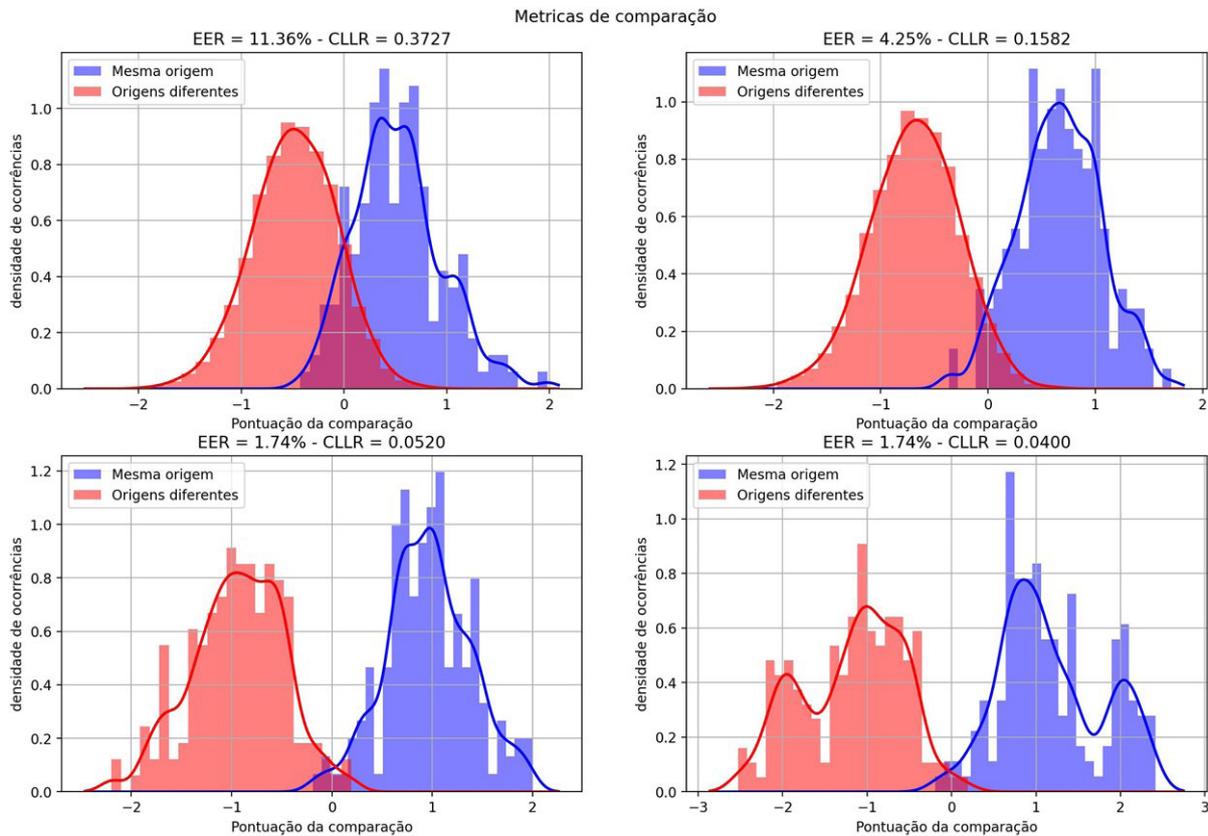


do logaritmo da razão de verossimilhança (*log likelihood ratio cost* – C_{LLR}). Essa métrica, proposta por Gonzalez-Rodriguez (2006), é obtida a partir de um evento de calibração com N_{MO} pares de classificação da classe “mesma origem” e N_{OD} amostras da classe “origem diferente”. Para cada amostra do total ($N_{MO} + N_{OD}$), o classificador gera os valores de LR_{MO} e LR_{OD} , respectivamente, para comparação entre “mesma origem” e “origem diferente”. O custo obtido na calibração é definido como

de LR_{MO} e menores os valores de LR_{OD} menor será o valor de C_{LLR} . Em outras palavras, um sistema baseado em LR que apresenta um custo (C_{LLR}) baixo. Se o custo for baixo maior será o poder de distinção entre as classes (“mesma origem” e “origem diferente”) para os valores de LR obtidos.

²⁶ Neste contexto, modelo de classificação refere-se a uma forma matemática de obter a avaliação de uma classe a partir de medidas de características.

Figura 5 – Comparação entre métricas de avaliação entre diferentes classificadores. Na imagem superior esquerda um classificador com desempenho inferior até a imagem inferior esquerda onde tem-se um classificador de alto desempenho.



Fonte: Os autores.

A Figura 5 apresenta comparação entre classificadores hipotéticos com diferentes desempenhos. No gráfico da porção superior direita tem-se o classificador de pior desempenho²⁷. Nota-se na imagem que mesmo estabelecendo um limiar de decisão (no eixo horizontal), amostras de mesma origem serão classificadas como de origem diferente e vice e versa. Este resultado apresenta EER de 11,36% e $C_{LLR} = 0,3727$. Na imagem adjacente as classes se distanciam e tem-se EER de 4,25% e $C_{LLR} = 0,1582$. Nos gráficos da segunda linha da Figura 5 nota-se a diferença entre a medida de EER e de C_{LLR} . Como o valor de ERR contabiliza apenas a ocorrência do erro, sem considerar o valor da razão de verossimilhança, na segunda linha da Figura 5 os dois classificadores apresentam a mesma EER = 1,74%, porém, o valor da C_{LLR} diferentes. A diferença ocorre porque na imagem da direita (com menor C_{LLR}) as ocorrências

das classes se distanciam mais do limiar de decisão.

Desta forma, ao calibrar um classificador, i.e., escolher o valor do limiar de decisão λ_0 que distingue entre as classes, é importante considerar o valor do custo C_{LLR} , pois quanto menos esse valor mais exato será a classificação.

2.4 Precisão na Razão de Verossimilhança

A precisão de um sistema de classificação está associada à dispersão do valor da LR em relação a mudanças no valor da característica de entrada. Conforme o NRC (2009, p. 121) a precisão apresenta uma medida do intervalo com alta probabilidade de conter o valor verdadeiro (parâmetro). Uma forma de realizar uma inferência no espaço paramétrico (do valor verdadeiro), onde é definido o valor probabilístico do problema, é

²⁷ Para classificadores, o desempenho é a capacidade de obter valores verdadeiros positivos e verdadeiros negativos com o menor risco e custo de cometer erros.

utilizar o teste de significância genuinamente bayesiano (*Full Bayesian Significant Test – FBST*) proposto por Pereira e Stern (1999).

A aplicação do FBST para realização de inferência intervalar – denominado intervalo de evidência –, foi proposto por Silva (2020a) e baseia-se no cálculo do intervalo no espaço paramétrico (Madruga, Pereira, Stern; 2003)²⁸. Suponha-se um classificador para determinação de fonte para distinguir pares de amostras padrão e questionadas entre as classes de mesma origem (MO) e de origem diferentes (OD). Caso disponha-se de n_p amostras do padrão e n_q amostras questionadas, o total de comparações, e de valores de LR, será $n = n_q \cdot n_p$, cada associação de padrão e questionado disponíveis²⁹.

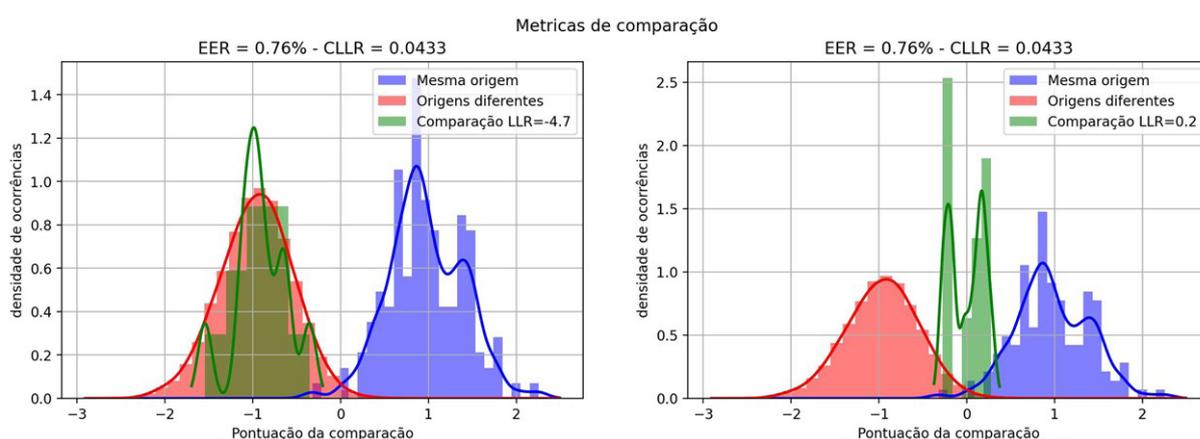
A Figura 6 apresenta exemplos da distribuição de valores de LR em um sistema calibrado com um limiar de decisão definido. O classificador da figura 6 apresenta $EER = 0,76\%$ e $C_{LLR} = 0,0433$. As ocorrências de calibração são apresentadas em azul, para mesma origem, e vermelho, para origens diferentes. Nas imagens as n ocorrências da comparação são representadas em verde. Como os valores de LR não ocorrem

sobre o mesmo ponto sobre o eixo horizontal, existe uma dispersão associada ao resultado da comparação.

A partir do valor médio amostral desta comparação pode ser calculado o intervalo de evidência, que são os limites que contêm o intervalo mais provável para conter o valor paramétrico da média dos resultados de LR para uma confiabilidade e desvio padrão amostral definidos. A Figura 7 apresenta a curva *tippet* (análoga a Figura 4a) do classificador, indicando o limiar de decisão em preto e em verde a média dos valores de LR, associados a Figura 6.

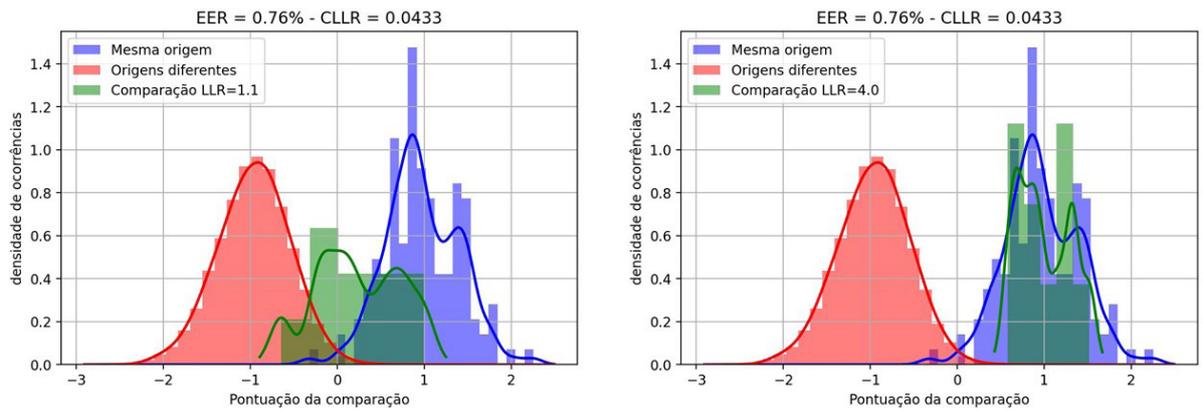
Sem avaliar a precisão do resultado (i.e., usando apenas a média pontual), o classificador indica como de mesma origem caso a média seja superior ao limiar e de origem diferente caso contrário. Entretanto, quando é realizada a inferência indicando os limites prováveis do valor paramétrico, um resultado intermediário pode ocorrer. Neste resultado o valor pontual aparece em uma das classes, mas o intervalo ultrapassa o limiar, indicando que o valor paramétrico pode estar contido na outra classe (do outro lado do limiar).

Figura 6 – Exemplos de dados de amostras comparadas em um sistema calibrado com $EER = 0,76\%$ e $C_{LLR} = 0,0433$. As ocorrências de calibração aparecem em azul, para mesma origem, e vermelho, para origens diferentes. Em verde as amostras comparadas, sendo uma tendendo a esquerda do eixo horizontal, duas centralizadas e uma tendendo a direita do eixo.



28 A formulação do cálculo do intervalo de evidência a partir do FBST envolve uma formulação matemática que não faz parte do escopo e da proposta deste artigo. Detalhes da formulação e aplicação podem ser encontrados em Silva (2020a) e (2020b).

29 Imagine que as amostras podem ser um conjunto de imagens, onde tem-se várias fotos padrões de um indivíduo, por exemplo, da carteira de motorista, das redes sociais etc., e as questionadas são diferentes momentos arquivos de vídeos, como de câmeras de segurança e de *smartphones*, associados a um fato típico. Caso possua-se duas fotos do padrão e três questionadas ter-se-ia um total de 6 comparações.



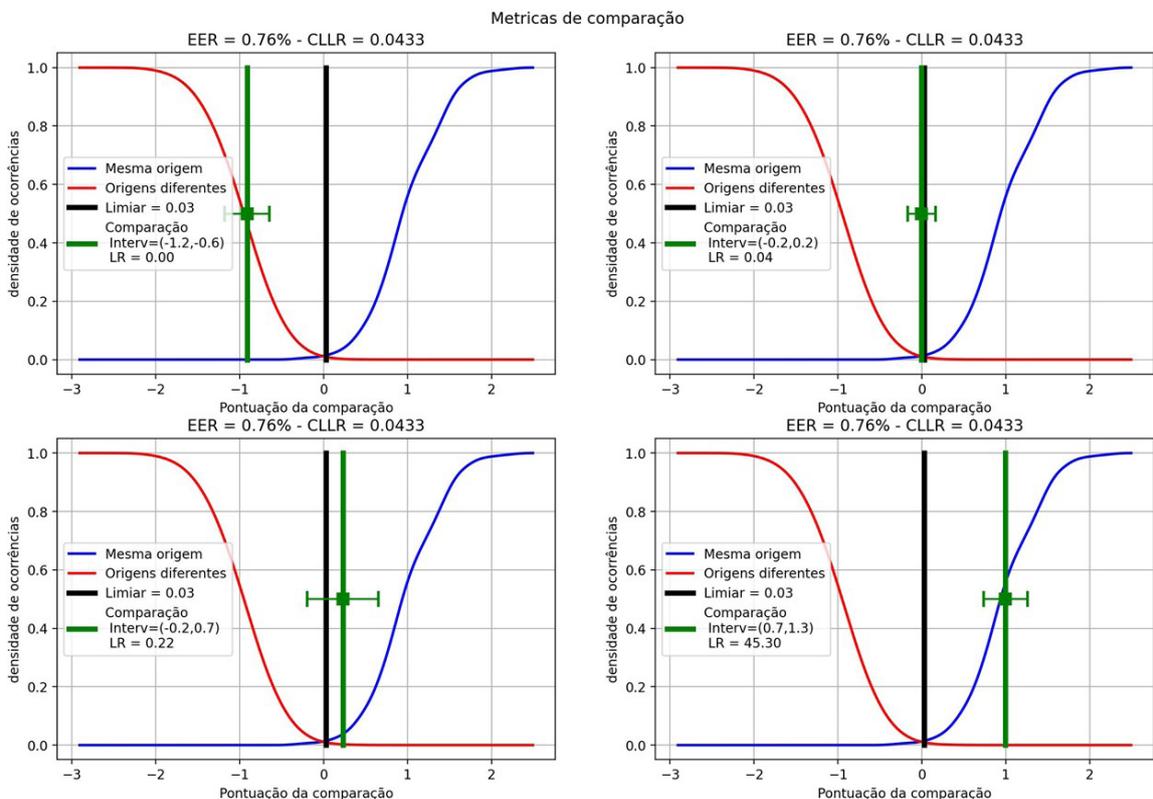
Fonte: Os autores.

Na Figura 7 os resultados do gráfico superior esquerdo e inferior direito são de classificações dentro de uma classe, tanto pelo valor da média, quanto pelo intervalo. Entretanto, no gráfico da porção superior direita o valor pontual cai sobre o limiar de decisão e no inferior esquerdo cai na classe de mesma origem, com intervalo abrangendo o limiar. Nestes dois últimos casos o classificador cairia em uma região intermediária, que no caso de um procedimento policial, daria

um *In dubio pro reo*.

A proposta da utilização do intervalo de evidência é devido à redução das ocorrências de erro Tipo I (falso positivo), em detrimento do aumento de resultados intermediários, em cenários com contaminação de ruído (SILVA, 2020b). Esta redução é superior a outras metodologias de inferência intervalar, como a de Gosset (STUDENT, 1908; ZABELL, 2008) e mostra-se um indicador de precisão em sistemas baseados em *LR*.

Figura 7 - A Figura 7 apresenta a curva *tippet* do classificador indicando o limiar de decisão na linha vertical preta. A linha vertical preta indica a média dos valores de *LR* da comparação. A linha verde vertical indica o intervalo de evidência. No gráfico superior esquerdo o resultado indica a classe de origem diferente, no superior direito a média cai sobre o limiar de decisão. No inferior esquerdo a média está a direita do limiar mas o intervalo de evidência se sobrepõe ao limiar de decisão.



Fonte: Os autores.

3 Considerações Finais

O presente trabalho descreveu, a partir de exemplos hipotéticos, os fundamentos de uma comparação baseada em LR e suas métricas de precisão e exatidão. Foram apresentados, de forma ilustrativa, os conceitos de exatidão e precisão, bem como foi descrito e ilustrado os fundamentos da classificação baseada em razão de verossimilhança (LR – *likelihood ratio*). Por fim, detalhou-se a medida de custo do logaritmo da razão de verossimilhança (*log likelihood ratio cost* – C_{LLR}) como parâmetro de exatidão de um classificador, quanto menor o CLLR mais exata a classificação; e o intervalo de evidência como medida de precisão. A consequência da medida de precisão é o aparecimento de um resultado intermediário dentro da classificação dicotômica.

A aplicação de métricas de exatidão e precisão contribuem com o atual paradigma de determinação de fonte. O C_{LLR} na exatidão, permite avaliar quão compatíveis duas amostras são para serem consideradas de mesma fonte. A precisão – através do intervalo de evidência –, permite avaliar o alcance da correspondência ou de se cometer um erro na avaliação.

Em áreas de maior difusão de acesso a coleta de dados – e.g., assinatura de documentos, fotos de indivíduos, gravações de áudio, fotos de impressões digitais, imagem de ranhuras em projéteis ou objetos, arquivos digitais – a construção de banco de dados juntamente com o desenvolvimento e evolução de protocolos, permite difundir a tecnologia e dar celeridade à investigação criminal. Esta celeridade contribui para aprimorar a eficiência, eficácia e transparência da atividade de polícia judiciária na integração com o sistema de justiça.

As rotinas exemplificativas utilizadas para gerar os dados e as visualizações deste trabalho estão disponíveis no repositório do *github* de um dos autores. Como continuidade do presente trabalho, propõe-se estudos específicos para estabelecimento de protocolos robustos baseados em LR, protocolos estes que podem ser aplicados em diferentes níveis de determinação de fonte e

em diferentes etapas do processo de investigação criminal. ■

Referências

BISHOP, Christopher; NASRABADI, Nasser. **Pattern recognition and machine learning**. New York: springer, 2006.

BRASIL. Decreto-Lei Nº 3.689, de 3 de Outubro de 1941. **Código de Processo Penal**. Diário Oficial da União, 3 de outubro de 1941.

DUDA, Richard; HART, Peter; STORK, David. **Pattern classification** 2nd edition. New York, USA: John Wiley&Sons, v. 35, 2001.

GONZALEZ-RODRIGUEZ, Joaquin et al. Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. **Computer Speech & Language**, v. 20, n. 2-3, p. 331-355, 2006.

INTERNATIONAL VOCABULARY OF METROLOGY (IVM). **Basic and general concepts and associated terms**. 2012.

LEE, Yung; CHO, Sujung. The Different Crime Solving Styles of Police Agencies across the United States: A Latent Class Analysis of Criminal Investigation Goals. **American Journal of Criminal Justice**, v. 46, p. 496-527, 2021.

MELO, Danilo; Volpi, Edmilson; Silva, Hailton; Menezes, Leonardo. Decifrando o georreferenciamento. **Geografia Ensino & Pesquisa**, p. 85-102, 2014.

MADRUGA, Maria; PEREIRA, Carlos; STERN, Júlio. Bayesian evidence test for precise hypotheses. **Journal of Statistical Planning and Inference**, Elsevier, v. 117, n. 2, p. 185–198, 2003.

Meuwly, Didier. (2006). Forensic Individualization from Biometric Data. **Science & Justice**, v. 46, n. 4, p. 205-213.

MONTEIRO, Cláudia. Fundamentos para uma teoria da decisão judicial. In: **Anais do XVI Congresso Nacional do CONPEDI**. 2007. p. 6104-6125.

MORRISON, Geoffrey. Measuring the validity and reliability of forensic likelihood-ratio systems. **Science & Justice**, v. 51, n. 3, p. 91-98, 2011.

NATIONAL RESEARCH COUNCIL (NRC) et al. **Strengthening forensic science in the United States: a path forward**. National Academies Press, 2009.

ROBERTSON, Bernard; VIGNAUX, Tony.; BERGER, Charles. **Interpreting evidence: evaluating forensic science in the courtroom**. [S.l.]: John Wiley & Sons, 2016.

SAKS, Michael; KOEHLER, Jonathan. The individualization fallacy in forensic science evidence. **Vand. L. Rev.**, v. 61, p. 199, 2008.

SILVA, Adelino. **Intervalo de evidência e pareamento fuzzy utilizando relação sinal ruído aplicados à comparação forense de locutores**. Tese (doutorado) – Universidade Federal de Minas Gerais. Belo Horizonte. 138p. 2020a.

SILVA, Adelino; VIEIRA, Maurílio; BARBOSA, Adriano. Forensic speaker comparison using evidence interval in full Bayesian significance test. **Mathematical Problems in Engineering**, v. 2020, p. 1-9, 2020b.

STUDENT. The probable error of a mean. **Biometrika, Journal Storage (JSTOR)**, p. 1–25, 1908.

SUÁREZ, Oscar; GARCÍA, Gloria. Component Analysis in Artificial Vision. In: **Encyclopedia of Artificial Intelligence**. IGI Global, 2009. p. 367-371.

ZABELL, Sandy. On student's 1908 article "the probable error of a mean". **Journal of the American Statistical Association**, Taylor & Francis, v. 103, n. 481, p. 1–7, 2008.