

Base de Dados para Avaliação de Algoritmos de Classificação de Modulação

Andrey Silva, Joary Paulo e Aldebaro Klautau
 Universidade Federal do Pará
 {andreysilva, joary, aldebaro}@ufpa.br

Resumo—A classificação automática de modulação (CAM) é um importante processo nos sistemas militares e de rádio cognitivo, a qual consiste em identificar a modulação adotada sem haver cooperação do transmissor. No entanto a avaliação desses sistemas é tipicamente feita adotando-se sinais simulados, os quais podem não representar bem cenários realísticos. A primeira contribuição deste trabalho é disponibilizar uma base de dados pública de sinais digitalizados e com diferentes modulações. Essa base de dados foi obtida utilizando-se *Universal Software Radio Peripheral* (USRP) e o *software GNURadio*. A segunda contribuição é uma comparação de algoritmos de CAM utilizando a base de dados gerada, avaliando-se eventuais discrepâncias entre os resultados dos algoritmos de CAM simulados e os com a citada base de dados. Os resultados indicam que entre os dois algoritmos avaliados, o *Concatenated Sorted Symbols* (CSS) superou o algoritmo HISTO e HISTO2D em performance. Além disso, o Erro Médio Quadrático (EQM), entre os resultados simulados e os obtidos com um canal real, esteve na ordem de 10^{-4} para *Signal-to-Noise Ratio* (SNR) abaixo de 5 dB. Para valores de SNR acima de 5 dB o EQM esteve próximo de zero.

Keywords—Classificação automática de modulação, base de dados de sinais, classificadores discriminativos

I. INTRODUÇÃO

CLASSIFICAÇÃO automática de modulação (CAM) consiste em identificar o esquema de modulação em uso por um dado sistema de comunicação, com um alto grau de confiança baseando-se em um sinal ruidoso com parâmetros desconhecidos [1]. Apesar de muito estudado em áreas militares [2], o interesse por CAM se renovou devido ao advento dos rádios cognitivos, onde CAM atua principalmente no sensoriamento espectral [3].

Dada a miríade de algoritmos de CAM existentes, é importante definir *benchmarks*. Na literatura, vários algoritmos de CAM são testados assumindo canais ideais e parâmetros perfeitamente estimados, o que pode ser irrealístico em vários cenários práticos. A ausência de *benchmarks* tem sido criticada em outros domínios, como genômica [4], por exemplo.

Nesse contexto, uma contribuição deste trabalho é construir e publicar uma base de dados para *benchmark* de algoritmos de CAM. A base foi obtida com o hardware *Universal Software Radio Peripheral* (USRP) utilizando *GNURadio* e permite a exploração de vários aspectos práticos de CAM. Além disso, este trabalho propõe uma comparação de performance de alguns algoritmos de CAM utilizando a base de dados proposta, utilizando o EQM para comparação dos resultados

obtidos com sinais simulados e sinais reais, validando a base proposta.

Este trabalho é organizado como segue: A Seção II apresenta detalhes dos algoritmos utilizados. A Seção III apresenta detalhes da base de dados proposta. A Seção IV apresenta os resultados dos algoritmos utilizando os sinais simulados e faz uma comparação com os obtidos com a base de dados. Por último, encontram-se as conclusões na Seção V.

II. ALGORITMOS UTILIZADOS DE CAM

O *Concatenated Sorted Symbols* (CSS) é um *front-end* utilizado para classificação de modulação proposto em [5], que permite o uso de aprendizado discriminativo, utilizando símbolos de modulações digitais como parâmetro de entrada para os classificadores. O CSS se caracteriza por armazenar a magnitude e a fase dos símbolos recebidos normalizados e em seguida ordená-los separadamente. Os dois vetores ordenados (magnitude e fase) são então concatenados, gerando um novo vetor X com comprimento $D = 2N$, que reflete uma assinatura individual da modulação.

O HISTO e o HISTO2D [6], diferentemente do CSS, evitam esse procedimento de ordenação que implica em um certo custo computacional. O HISTO calcula dois histogramas com B bins, um para magnitude e um para fase. Os dois histogramas, derivados da sequência de símbolos recebidos, são concatenados para formar um único vetor que é utilizado como parâmetro de entrada para o classificador. O HISTO2D cria um *grid* bidimensional de bins, que correlaciona a magnitude e fase dos símbolos.

III. BASE DE DADOS UFPA

A base de dados em [7] contém sinais digitalizados provenientes tanto de simulações em *software* quanto de transmissões feitas através de canais reais. O nome do arquivo contém a modulação utilizada e o *software/equipamento* com o qual foi gerado. Os valores de SNR presentes nos nomes dos arquivos são representados em dB.

Para a aquisição e correção dos sinais optou-se pelo uso de dois sinais multiplexados em frequência e separados por 500 kHz. O primeiro sinal é previamente conhecido e possui uma potência mais alta (canal de controle) sendo utilizado para calcular os erros de *timing*, frequência e fase. O segundo sinal possui a modulação de interesse (canal de dados) e será decodificado com as correções calculadas pelo sinal anterior. Isto permitiu a obtenção de sinais com baixa SNR no canal

de dados, mantendo a robustez dos algoritmos de sincronismo com o auxílio do canal de controle. Os principais parâmetros utilizados podem ser observados na Tabela I.

Tabela I
PARÂMETROS DA APLICAÇÃO

Parâmetros Gerais	Canal de Controle	Canal de Dados
Frequência de Amostragem	1e6	1e6
Nº de Símbolos	1e6	1e6
Alocação de Potência (%)	80	20
Frequência Central (MHz)	904.75	905.25
Parâmetros do Filtro		
Fator de <i>Oversampling</i>	4	4
Fator de <i>Roll-off</i>	0.35	0.35
Spam do Filtro	1e3	1e3
Tipo de Filtro	SRRC ¹	SRRC ¹
Largura de Banda (kHz)	375	375

¹Square-Root Raised Cosine.

IV. RESULTADOS

Para cada par Avaliação e SNR foram criados dois conjuntos de sinais disjuntos, um para testes outro para treino. Cada um desses conjuntos é formado por 1000 trechos de sinais onde cada trecho possui 250 símbolos. Em todos os sinais as constelações possuem energia unitária.

O algoritmo escolhido para a tarefa de classificação foi: *Support Vector Machine* (SVM) com função de ativação Linear e *Radial Basis Function* (RBF). O HISTO e o HISTO2D tiveram $Q = 32$ bins em suas configurações, o que resultou em $D = 64$ e $D = 1024$ características de entrada para o SVM e o CSS fornece $D = 500$ características.

A Fig. 1 mostra a Probabilidade de Classificação Correta (PCC) em um canal *Additive White Gaussian Noise* (AWGN) obtida pelo grupo de modulações (BPSK, 8PSK, 4PAM, 16QAM). Note que o *front-end* CSS possui uma performance levemente superior ao HISTO e HISTO2D, com ambas funções de ativação (Linear e RBF). A maior diferença de performance acontece com SNR igual a 1 dB, que é de cerca de 1% do CSS_RBF em relação ao HISTO_RBF e de cerca de 5% do CSS_RBF em relação ao HISTO2D_RBF.

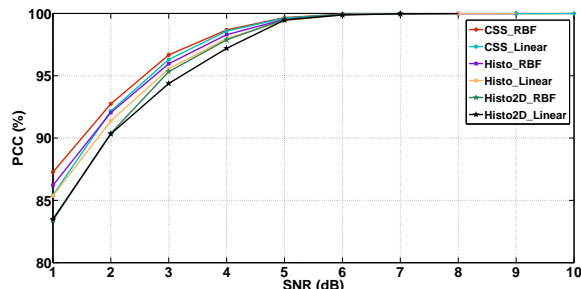


Figura 1. Probabilidade de classificação correta dos algoritmos em um canal AWGN com $N = 250$ símbolos.

Na Fig. 2 é apresentado o EQM da PCC normalizada para diferentes valores de SNR, comparando o resultado dos algoritmos de CAM que utilizaram sinais simulados com os que utilizaram sinais reais. O erro para SNR com valores abaixo de 5 dB mantiveram-se na ordem de 10^{-4} , o qual

é causado pela dificuldade de se obter sinais reais precisos com valores baixos de SNR. Entretanto, pode-se notar que os resultados dos classificadores é bastante similar aos obtidos com os dados simulados para valores de 5 dB em diante.

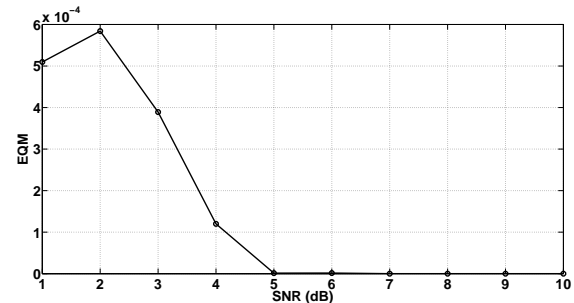


Figura 2. Error Quadrático Médio da PCC Normalizada por SNR.

V. CONCLUSÃO

Os resultados experimentais obtidos mostraram que o *front-end* CSS possui uma performance levemente superior ao HISTO e ao HISTO2D. Entretanto, o HISTO possui menos características que o CSS e o HISTO2D, o que facilita sua implementação em *hardware*. Em relação ao HISTO2D, apesar de possuir o maior número de características dos classificadores utilizados, perde em performance para o CSS e HISTO para valores de SNR abaixo de 5 dB.

É importante notar que o classificador utilizando funções de ativação Linear possui resultados tão bons quanto o que utiliza RBF. A base de dados foi criada e seus dados foram utilizados para comparar a eficiência dos classificadores utilizando dados simulados e dados reais, o erro para valores abaixo de 5 dB mantiveram-se na faixa de 10^{-4} e de 5 dB em diante os resultados são praticamente iguais. Dessa forma, os resultados comprovam a eficiência da base proposta em auxiliar pesquisadores de CAM que desejam utilizar dados reais e não somente simulados.

Este trabalho está em desenvolvimento e pretende evoluir a base proposta para melhor avaliar a robustez dos algoritmos de classificação frente a sinais reais, bem como os efeitos causados pelas imperfeições de canal e sincronismo nos algoritmos de classificação de modulação.

REFERÊNCIAS

- [1] Z. Zhu and A. Nandi, *Automatic modulation classification: principles, algorithms, and applications*. Chichester, West Sussex, United Kingdom: Wiley, 2015.
- [2] M. Aslam, Z. Zhu, and A. Nandi, "Automatic digital modulation classification using genetic programming with K-nearest neighbor," in *Military Communications Conference*, Nov. 2010, pp. 1731–1736.
- [3] S. Haykin, D. Thomson, and J. Reed, "Spectrum sensing for cognitive radio," *Proceedings of the IEEE*, vol. 97, pp. 849–877, May 2009.
- [4] U. Braga-Neto, "Fads and fallacies in the name of small-sample microarray classification - a highlight of misunderstanding and erroneous usage in the applications of genomic signal processing," *IEEE Signal Processing Magazine*, vol. 24, pp. 91 – 99, Jan. 2007.
- [5] F. Muller, C. Cardoso, and A. Klautau, "A front end for discriminative learning in automatic modulation classification," *IEEE Commun. Lett.*, vol. 15, no. 4, pp. 443–445, Apr. 2011.
- [6] C. C. Junior, "Classificação automática de modulação baseada em aprendizagem discriminativa." Ph.D. dissertation, UFPA, 2015.
- [7] LaPS, "Ufpa AMC dataset," 2016. [Online]. Available: https://legacy.redmine.lasseufpa.org/projects/knowledgebase/wiki/Laps_Modulation