

CLASSIFICADOR DE EMBARCAÇÕES MARÍTIMAS UTILIZANDO LÓGICA FUZZY

Manuel Ramon Vargas Avila

Coppe/UFRJ. Centro de Tecnologia - Bloco B - Ilha do Fundão, Rio de Janeiro - RJ
manuelvargas760@gmail.com

Paulo Collares Moreira Neto

IPqM/Marinha do Brasil. Rua Ipiru, 02, Cacúia, Ilha do Governador, Rio de Janeiro - RJ
pcollares@gmail.com

Rubens L. de Oliveira

IPqM/Marinha do Brasil. Rua Ipiru, 02, Cacúia, Ilha do Governador, Rio de Janeiro - RJ
rubensliv@yahoo.com

Resumo

Este projeto mostra o desenvolvimento de um classificador fuzzy utilizando a ferramenta SimpleFuzzyGrid do software Weka com o objetivo de classificar embarcações que navegam pela Baía de Guanabara, no Rio de Janeiro, quanto ao tipo: barcas, catamarãs, cargueiros e petroleiros. Para isso foram utilizadas informações estáticas e dinâmicas obtidas pelo AIS como a velocidade, comprimento, tipo, calado, entre outras.

Palavras-Chaves: Conjuntos Fuzzy; classificação; AIS, navios.

Abstract

This paper shows the development of a fuzzy classifier, using the SimpleFuzzyGrid tool from Weka software, in order to classify vessels that navigate along of Baía de Guanabara in Rio de Janeiro, in three distinct types: passenger, freighters and tankers. Were used static and dynamic information obtained by an AIS equipment, like speed, length, type, draft, etc.

Keywords: Fuzzy sets; classification; AIS; vessel; ships.

1. INTRODUÇÃO

Uma classificação automática de alvos marítimos vem se tornando a cada dia mais importante para uma navegação mais segura e um melhor desempenho em operações de guerra, nisto se propõe este trabalho, desenvolver um classificador fuzzy de embarcações marítimas quanto ao seu tipo, levando em conta o seu histórico de navegação e informações estáticas e utilizando a ferramenta SimpleFuzzyGrid do software Weka.

Para este trabalho foram usados dados coletados de um AIS (Automatic Identification System) instalado nas dependências do IPqM (Instituto de Pesquisas da Marinha) localizado na Ilha do Governador, RJ, a fim de coletar as informações dos navios que trafegam na Baía de Guanabara. O AIS é um sistema obrigatório, instalado em todas as embarcação mercantes e militares, e tem a função de transmitir constantemente informações do navio portador, como a localização geográfica, rumo, velocidade, tipo de navio, tamanho, entre outras informações.

Para este trabalho foram considerados alguns pontos no interior da Baía de Guanabara onde navios do mesmo tipo geralmente se agrupam, ou trafegam, estes pontos foram nomeados de sítios, e estão descritos abaixo:

Áreas de fundeio: Existem áreas específicas de ancoragem no interior da Baía de Guanabara onde embarcações do mesmo tipo ficam aguardando uma liberação no porto, por exemplo. Foram determinadas duas áreas de fundeio, a área 1 foi definida para embarcações do tipo petroleiro e a área 2 para embarcações do tipo cargueiros, para o caso prático estas áreas foram chamadas de sítios petroleiros e sítios cargueiros, respectivamente. As coordenadas geográficas destes pontos são as mostradas na Tabela 1.

Tabela 1- Coordenadas geográficas Áreas de fundeio.

Sítio	Latitude	Longitude
1. Fundeio de Petroleiro	-22.830	-43.146
2. Fundeio de Carga	-22.84665	-43.143

Porto do Rio de Janeiro: Com mais de 6 km de extensão, um dos maiores portos do Brasil, local de carga e descarga de mercadoria e desembarque de passageiros pelos cruzeiros nacionais e internacionais, para o caso prático foi chamado de sítio Praça Mauá neste trabalho, devido ao ponto geográfico central escolhido para definir este sítio (Tabela 2).

Tabela 2 - Coordenada geográfica do porto do RJ

Sítio	Latitude	Longitude
1. Praça Mauá	-22.892283	-43.1897

Estações das barcas: Na Baía de Guanabara existem cinco estações de embarque e desembarque de passageiros por meio do transporte coletivo marítimo, nestas estações trafegam, exclusivamente, as barcas e catamarãs. Para o caso prático estas estações foram chamadas de sítios, e suas coordenadas estão definidas na Tabela 3.

Tabela 3 - Coordenadas geográficas das estações das barcas

Sítio	Latitude	Longitude
1.Praça XV	-22.9028	-43.1711
2.Araribóia	-22.8946	-43.12435
3.Charitas	-22.9322	-43.0989
4.Cocotá	-22.805	-43.178
5.Paquetá	-22.7622	-43.1063

2. SIMPLEFUZZYGRID NO WEKA

Nesta seção é detalhado o processo utilizado pela ferramenta SimpleFuzzyGrid do software Weka, este possui um sistema de auto geração de regras do tipo SE-ENTÃO. Na equação 1 é observada a estrutura da regra utilizada pela ferramenta, neste caso para um problema com duas dimensões para uma fácil compreensão,

Equação 1: Regra R_{ij}^K : SE x_1 é A_i^K e x_2 é A_j^K então (x_1, x_2) pertencem a C_{ij}^K [2]
com $CF = CF_{ij}^K$

onde K é o número de subconjuntos fuzzy, R_{ij}^K é o nome da regra, x_1 e x_2 são os pontos de um padrão a classificar, A_i^K e A_j^K são cada um dos subconjuntos fuzzy, C_{ij}^K é a consequência da regra, CF_{ij}^K é o grau de certeza da regra, e i e $j = 1, \dots, K$ representam cada uma das duas dimensões do problema, isto com o fim de testar todas as possíveis combinações dos subespaços.

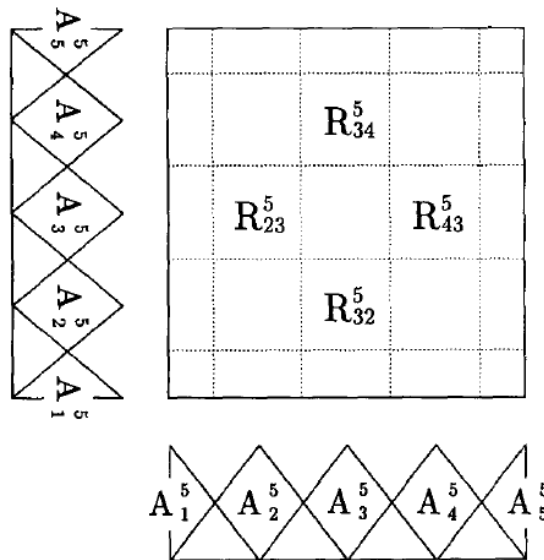


Figura 1 - Partição do espaço em subconjuntos fuzzy [2]

Na Figura 1 temos um exemplo de como é feita a partição do espaço nos subconjuntos fuzzy para o caso de $K=5$, nela observamos a maneira como são etiquetados cada um dos subconjuntos e regras fuzzy e como são geradas ditas regras a partir da interseção dos subconjuntos.

Nesta ferramenta os subconjuntos Fuzzy (A_i^K) são definidos por uma função triangular simétrica descrita pelas equações 2 a 4, onde o valor central da função é a_i^K e o

suporte é o intervalo aberto $(a_i^K - b^K, a_i^K + b^K)$. [2]

$$\text{Equação 2: } \mu_i^K(x) = \max\{1 - |x - a_i^K|/b^K, 0\}, i = 1, 2, \dots, K$$

$$\text{Equação 3: } a_i^K = (i - 1)/(K - 1), i = 1, 2, \dots, K$$

$$\text{Equação 4: } b^K = 1/(K - 1)$$

A seguir são mostrados os passos para a geração das regras fuzzy [2]:

1. Calcular β_{Ct} para $t = 1, 2, \dots, M$ onde M é o numero de classes do problema.

$$\text{Equação 5: } \beta_{Ct} = \sum_{p \in Ct} \mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2})$$

Onde Ct faz referência à classe, $p = 1, 2, \dots, m$ e m é igual ao número de padrões do problema.

2. Encontrar as classes X (CX) tal que:

$$\text{Equação 6: } \beta_{CX} = \max\{\beta_{C1}, \beta_{C2}, \dots, \beta_{CM}\}$$

Se dois ou mais classes tem o máximo valor na equação 6, então a regra fuzzy se-então correspondente ao subespaço fuzzy $A_i^K \times A_j^K$ não é gerada, do contrario C_{ij}^K é determinada como CX na equação 6.

3. Se apenas uma classe tem o máximo valor na equação 6, CF_{ij}^K é determinada como:

$$\text{Equação 7: } CF_{ij}^K = |\beta_{CX} - \beta| / \sum_{t=1}^M \beta_{Ct}$$

Onde:

$$\text{Equação 8: } \beta = \sum_{Ct \neq CX} \beta_{Ct} / (M - 1)$$

Neste procedimento, a consequência de C_{ij}^K é determinada como a classe que tem a maior soma de $\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2})$. CF oscila no intervalo de 0 a 1. Se todos os padrões em $A_i^K \times A_j^K$ são da classe X (CX), $CF_{ij}^K = 1$, do contrario $CF_{ij}^K < 1$.

Agora se é necessário classificar um padrão novo deve-se seguir o seguinte procedimento [2]:

Assume-se que $x_p = (x_{p1}, x_{p2})$ é um padrão novo.

1. Calcular α_{Ct} para $t=1, 2, \dots, M$ como:

$$\text{Equação 9: } \alpha_{Ct} = \max\{\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K | C_{ij}^K = Ct; R_{ij}^K \in S_R\}$$

Onde S_R é o conjunto de todas as regras geradas

2. Encontrar a classe X (CX) tal que

$$\text{Equação 10: } \alpha_{CX} = \max\{\alpha_{C1}, \alpha_{C2}, \dots, \alpha_{CM}\}$$

Se dois ou mais classes tomam o máximo valor na equação 10 então x_p não pode ser classificado, do contrário associar x_p à classe X (CX) determinada pela equação 10.

Neste procedimento, o resultado da inferência fuzzy é a consequência da regra fuzzy se-então que tem o máximo produto de $\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K$. Se não existem regras fuzzy se-então tal que $\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) > 0$ e $CF_{ij}^K > 0$ para x_p , o novo padrão x_p não pode ser classificado.

3. DESENVOLVIMENTO DO CLASSIFICADOR

3.1. GERAÇÃO DE DADOS

Neste projeto foram analisamos nove atributos obtidos pelo AIS referentes a navegação (velocidade, rumo atual, etc), e características estáticas do navio (calado, comprimento, tipo, etc) para que, através da logica Fuzzy, seja classificado entre seis classes, Cargueiro, Petroleiros, Barcas, Catamarãs e Cruzeiros.

Os dados foram coletados durante cinco dias, onde mais de 3 milhões de registros

tiveram que ser filtrados e analisados, selecionando apenas aqueles registros de navios que trafegam pela Baía de Guanabara e aqueles que pertençam aos tipos de navios de interesse neste trabalho. Devido a que alguns de estes dados não tinham informação suficientemente relevante para compor uma boa classificação, como é observado na Figura 2 onde é evidente que o comprimento diferencia dois grupos de grande porte como são cargueiros, cruzeiros e petroleiros dos demais tipos de navios, na Figura 3 vemos que a velocidade também não é uma variável discriminativa só ajuda a destacar os catamarãs das demais classes e por último, na Figura 4, o calado tem um comportamento similar ao comprimento na hora de separar as classes, por isso, foi necessário inferir outras informações, com o estabelecimento das áreas de fundeio, portos e sítios de barcas e o uso do rumo e posição informadas pelo AIS, pode-se inferir que quanto mais próximos destes sítios um navio se encontra, e o seu rumo condiz com a direção deste mesmo sítio, há maior certeza do navio pertencer a uma classe relacionada a estes pontos.

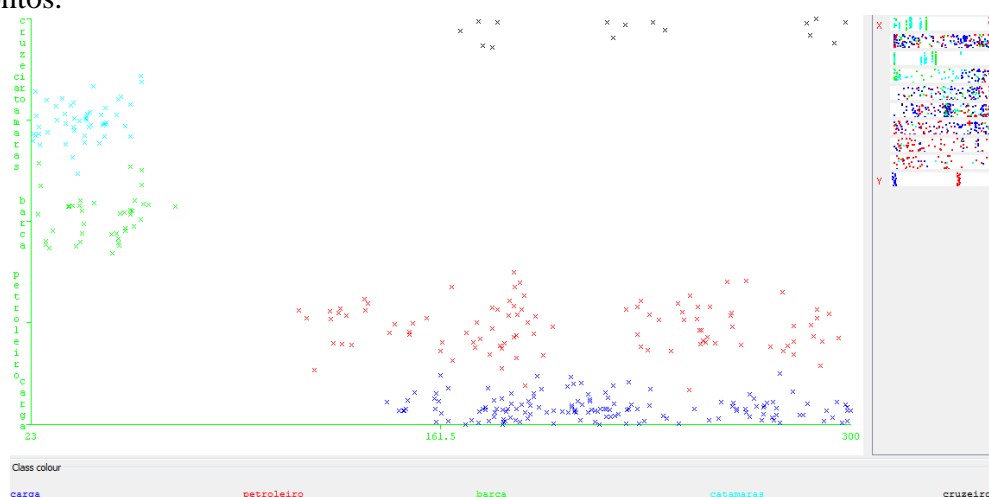


Figura 2 - Comprimento vs Classes

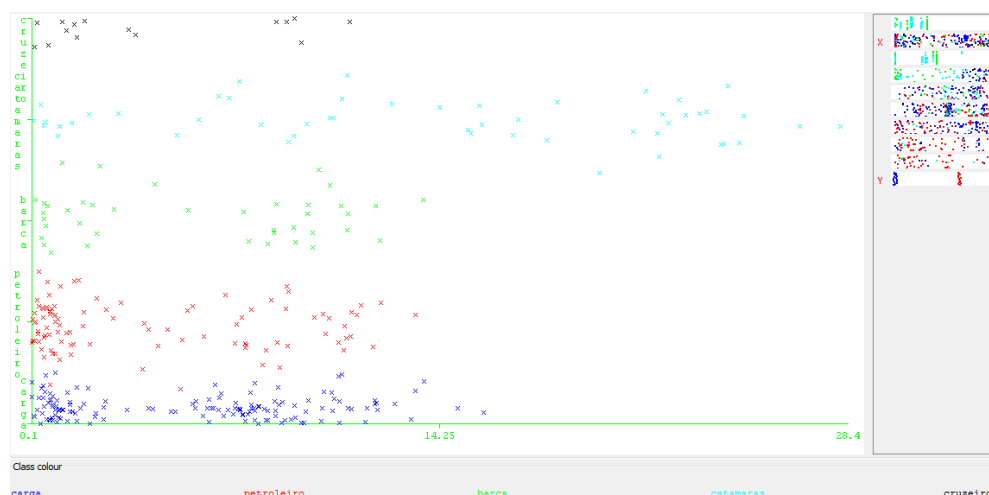


Figura 3 - Velocidade vs Classes

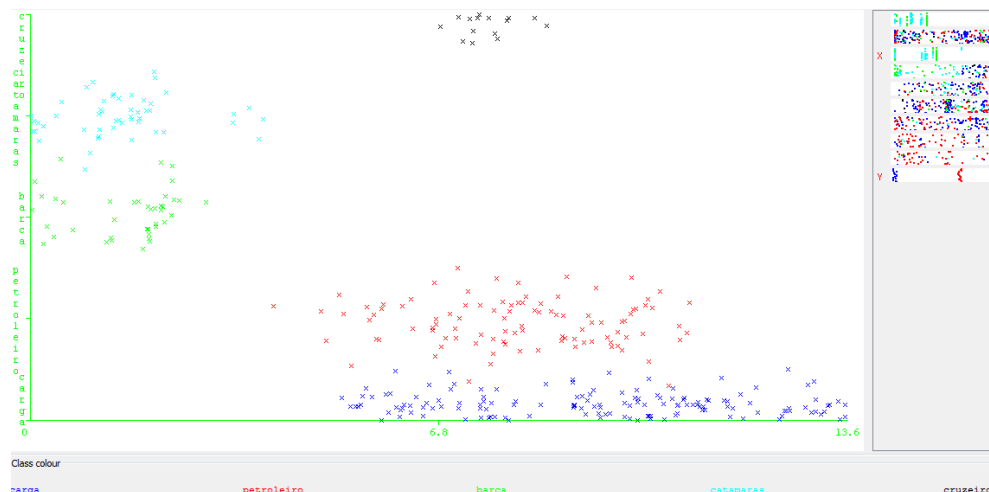


Figura 4 - Calado vs Classes

Além do AIS proporcionar dados das características dos navios, também permite conhecer a classe da embarcação que é essencial para o treinamento e validação da classificação, porém, através do AIS só foi possível identificar três classes utilizadas neste projeto (embarcações de passageiros, cargueiros e petroleiros), uma vez conhecendo as embarcações de passageiro, identificamos manualmente, por meio do MMSI no site do Marinetraffic as demais classes (Barcas, catamarãs e cruzeiros).

Para cada navio selecionado foram usados os seguintes dados:

Comprimento: Tamanho do navio em metros da popa a proa;

Velocidade: Velocidade em Nós, no instante da aquisição;

Calado: profundidade em metros a que se encontra o ponto mais baixo da quilha do navio;

Distância do sítio das barcas: Distância em milhas náuticas de uma das cinco estações de barcas mais próxima no momento da aquisição;

Diferença de rumo do sítio das barcas: Valor em graus, resultante da diferença do rumo empregado pelo navio no momento da aquisição e da marcação da estação das barcas mais próxima;

Distância do sítio do petroleiro: Distância em milhas náuticas do ponto de fundeio do petroleiro no momento da aquisição;

Diferença de rumo do fundeio do petroleiro: Valor em graus, resultante da diferença do rumo empregado pelo navio no momento da aquisição e da marcação do ponto de fundeio do petroleiro;

Distância do sítio de carga: Distância em milhas náuticas da zona de fundeio de carga mais próxima no momento da aquisição;

Diferença de rumo do fundeio de carga: Valor em graus, resultante da diferença do rumo empregado pelo navio no momento da aquisição e da marcação do ponto de interesse de carga;

Tipo do navio: Classe do navio para treinar e validar a classificação.

3.2. PRÉ-PROCESSAMENTO

Para uma melhor classificação, foram usados cinco dados de cinemática por cada navio, estes dados foram selecionados aleatoriamente da base de dados feita para o presente trabalho. Com isso e já filtrada a base de dados, são obtidos 70 navios dos quais 29 são cargueiros, 20 petroleiros, 8 barcas, 10 catamarãs e 3 cruzeiros, com um total de 350 instancias para a classificação.

Para facilitar o treinamento e funcionamento da classificação fuzzy é preciso fazer um pré-processamento dos dados, para isso foi realizada uma análise de correlação mostrado na Tabela 4, na qual observamos que a variável “distância do sitio de carga” não tem correlação forte com a saída pelo que é possível que seja descartada no treinamento e validação, diminuindo o custo computacional.

Tabela 4 - Correlação entre os dados

Valor mínimo de correlação 0,1054	Comprimento	Velocidade	Calado	Dist. Barca	Dif.rumo barca	Dist. Carga	Dif.rumo carga	Dist. Petro.	Dif.rumo petro.	Saída
Comprimento	1									-0.5115
Velocidade	-0,3308	1								0.2914
Calado	0,9181	-0,3077	1							-0.5342
Dist. Barca	0,4543	-0,1576	0,4203	1						-0.4807
Dif. rumo barca	0,133	0,0198	0,1028	0,3704	1					-0.1361
Dist. Carga	-0,086	0,3029	-0,0674	0,0326	0,4208	1				-0.0323
Dif. rumo carga	-0,2755	0,1709	-0,2349	-0,1991	-0,1054	0,1684	1			0.2242
Dist. Petro.	-0,0528	0,3245	-0,0295	-0,2611	0,1014	0,5029	0,4138	1		0.2826
Dif.rumo petro.	-0,098	0,274	-0,0549	-0,3175	-0,3288	-0,1696	0,6198	0,5714	1	0.3022

4. RESULTADOS

Para realizar a classificação foram usados 4 subconjuntos fuzzy (4⁴) e para validar o resultado foi utilizada a validação cruzada com 10 subgrupo de dados da amostra, com isso foram feitas 4 provas que serão descritas abaixo:

1. São utilizadas apenas as variáveis comprimento, calado e velocidade para classificar os 5 tipos de navios, os resultados obtidos são mostrados na tabela 5, onde obtemos 229 (65,4286%) instancias classificadas corretamente e 121 (34.5714%) classificadas de maneira incorreta.

Tabela 5 - Resultados só com as variáveis comprimento, velocidade e calado

Cargueiros	Petroleiros	Barcas	Catamarãs	Cruzeiros	
122	23	0	0	0	Cargueiros
58	42	0	0	0	Petroleiros
0	0	38	2	0	Barcas
0	0	23	27	0	Catamarãs
11	4	0	0	0	Cruzeiros

2. Nesta prova são utilizadas todas as variáveis descritas na seção 3.1 (comprimento, velocidade, calado, distância do sitio das barcas, etc) na classificação, os resultados desta prova são mostrados na tabela 6, onde obtemos um total de 277 (79,1429%) instâncias classificadas corretamente, 71 (20,2857%) incorretas e 2 (0,5714%) sem classificação.

Tabela 6 - Resultados com 9 variáveis de entrada

Cargueiros	Petroleiros	Barcas	Catamarãs	Cruzeiros	
129	13	0	0	3	Cargueiros
23	77	0	0	0	Petroleiros
0	1	25	13	0	Barcas
0	0	13	36	0	Catamarãs
5	0	0	0	10	Cruzeiros

3. Agora é aplicado o pré-processamento dos dados descrito na seção 3.2, onde por meio da análise de correlação, a variável Distância Carga é eliminada por ter baixa correlação com a classe. Na tabela 7 são mostrados os resultados, nos quais obtemos 280 (80%) instancias classificadas corretamente, 69 (19,7143%) incorretas e 1 (0,2857%) sem classificação.

Tabela 7 - Resultados utilizando o pré-processamento

Cargueiros	Petroleiros	Barcas	Catamarãs	Cruzeiros	
132	12	0	0	1	Cargueiros
21	76	0	0	3	Petroleiros
0	1	24	14	0	Barcas
0	0	12	38	0	Catamarãs
5	0	0	0	10	Cruzeiros

4. Por último aplicamos a ferramenta attribute selection do Weka, com avaliador do atributo Correlation-based Feature Subset Selection (CfsSubsetEval) [1] e método de busca BestFirst, o objetivo desta ferramenta é selecionar as variáveis que tenham alta correlação com a classe e baixa inter-correlação analisando subconjuntos de características, neste caso ao aplicar esta ferramenta são eliminadas as variáveis que são inferidas pela diferença de rumos, ficando apenas 6 variáveis de entrada. Na tabela 8 observamos os resultados desta prova a qual classificou corretamente 286 (81,7143%) instâncias e 64 (18,2857%) instâncias incorretas.

Tabela 8 - Resultados utilizando attribute selection

Cargueiros	Petroleiros	Barcas	Catamarãs	Cruzeiros	
129	16	0	0	0	Cargueiros
17	83	0	0	3	Petroleiros
0	2	27	11	0	Barcas
0	0	12	38	0	Catamarãs
6	0	0	0	9	Cruzeiros

5. CONHECIMENTO EXTRAÍDO

Baseado nos resultados extraídos do WEKA, pudemos desenvolver algumas regras de classificação nas embarcações estudadas, esses dados serão de suma importância para o desenvolvimento de trabalhos futuros.

Barcas: (comprimento muito pequeno) e (velocidade muito baixa ou baixa) e (calado muito pequeno) e (distância do sítio das barcas muito pequena ou pequena) e (distância do sítio de cargueiros e de cruzeiros pequena ou grande) e (distância do sítio de petroleiro pequena ou grande).

Catamarãs: (comprimento muito pequeno) e (velocidade baixa ou alta ou muito alta) e (calado muito pequeno ou pequeno) e (distância do sítio das barcas muito pequena ou pequena) e (distância do sítio de cargueiros e de cruzeiros grande ou muito grande) e (distância do sítio de petroleiro pequena ou grande).

Cargueiros: (comprimento pequeno ou grande ou muito grande) e (velocidade muito baixa ou baixa) e (calado grande ou muito grande) e (distância do sítio das barcas pequena ou grande) e (distância do sítio de cargueiros e de cruzeiros muito pequena ou pequena ou grande) e (distância do sítio de petroleiro pequena ou grande ou muito grande).

Petroleiros: (comprimento pequeno ou grande) e (velocidade muito baixa ou baixa) e (calado pequeno ou grande) e (distância do sítio das barcas grande ou muito grande) e (distância do sítio de cargueiros e de cruzeiros pequena) e (distância do sítio de petroleiro muito pequena ou pequena).

Cruzeiros: (comprimento grande ou muito grande) e (velocidade muito baixa e calado grande) e (distância do sítio das barcas pequena) e (distância do sítio de cargueiros e cruzeiros muito pequena) e (distância do sítio de petroleiro grande).

Sendo que:

Comprimento:

Muito pequeno: de 23 metros à 115,33 metros;

Pequeno: de 23 metros à 207,67 metros;

Grande: de 115,33 metros à 300 metros;

Muito grande: de 207,67 metros à 300 metros.

Velocidade:

Muito baixa: de 0,1 nós à 9,53 nós;

Baixa: de 0,1 nós à 18,97 nós;

Alta: de 9,53 nós à 28,4 nós;

Muito alta: de 18,97 nós à 28,4 nós.

Calado:

Muito pequeno: de 0 metros à 4,53 metros;

Pequeno: de 0 metros à 9,07 metros;

Grande: de 4,53 metros à 13,6 metros;

Muito grande: de 9,07 metros à 13,6 metros.

Distância barcas:

Muito pequena: de 0,02 MN à 1,02 MN;

Pequena: de 0,02 MN à 2,03 MN;

Grande: de 1,02 MN à 3,03 MN;

Muito grande: de 2,03 MN à 3,03 MN.

Distância cargueiro e cruzeiro:

Muito pequena: de 0,12 MN à 1,89 MN;

Pequena: de 0,12 MN à 3,67 MN;
Grande: de 1,89 MN à 5,44 MN;
Muito grande: de 3,67 MN à 5,44 MN.

Distância petroleiro:

Muito pequena: de 0,17 MN à 2,63 MN;
Pequena: de 0,17 MN à 5,08 MN;
Grande: de 2,63 MN à 7,53 MN;
Muito grande: de 5,08 MN à 7,53.

6. CONCLUSÃO

- ☐ O uso das variáveis diferença de rumo e distância aos sítios de interesse (áreas de fundeio, etc), melhoram a classificação em quase 14%, comparado com o caso onde não foram utilizadas. Estas variáveis foram um diferencial na classificação de classes muito similares entre si, como por exemplo, no caso dos cargueiros, petroleiros e cruzeiros.
- ☐ Ao comparar os resultados obtidos pelas provas 2 e 4, é observado que as variáveis inferidas pelos rumos e os sítios de interesse, atrapalha um pouco na classificação, apesar da diferença ser de apenas 2.5%, o qual pode não parecer muito significativo, mas uma tarefa como esta exige a máxima precisão possível. Já na hora de analisar as provas 3 e 4 a diferença mais significativa nelas, é que o número de regras geradas na classificação na prova 3 gera 191 regras fuzzy e na prova 4 apenas 118, que auxilia na redução do custo computacional na hora de implementar o classificador.
- ☐ A eliminação da variável rumo na prova 4 possivelmente ocorre devido que os dados selecionados para a classificação contem navios em movimento, além de que e na Baía de Guanabara o trajeto entre dois pontos não necessariamente será uma linha reta, e o rumo atual pode não coincidir com o rumo do ponto de interesse ou o mesmo navio, estando perto do sítio, pode estar apontando para outra direção, portanto esta variável não forneceria informações relevantes para auxiliar na classificação.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] M. A. Hall (1998). Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New Zealand.
- [2] H. Ishibuchi, K. Nozaki, H. Tanaka (1993). Efficient fuzzy partition of pattern space for classification problems.