II SIMPÓSIO DE PRODUÇÃO CIENTÍFICA DA UNIFESSPA
A Pesquisa Científica e as Interações com a Realidade Amazônica
1 E 2 DE FEVEREIRO DE 2018

Análise da Prática Pesqueira na Bacia Araguaia-Tocantins utilizando Técnicas de Inteligência Computacional

Renato S. da Silva[1,2]; Cristiane Vieira da Cunha[3]; Hugo P. Kuribayashi[1,2]
[1] *Laboratório de Computação Científica, Unifesspa, Marabá-PA,Brasil*
[2] *Faculdade de Computação e Engenharia Elétrica, IGE-Unifesspa, Marabá-PA,Brasil*
[3] *Faculdade de Educação no Campo, ICH-Unifesspa, Marabá-PA,Brasil*

Palavras-Chave: Prática Pesqueira, Inteligência Computacional, Clusterização.

1. INTRODUÇÃO

O Brasil está entre os dezesseis maiores produtores de pescado no mercado internacional, no entanto esta produção é deficitária de informações básicas, em função de reduzido volume de dados que representem a dimensão do setor na produção nacional de alimentos [1], sobretudo na pesca artesanal, o que dificulta entre outras coisas, o monitoramento para evitar a pesca ilegal e o estabelecimento de políticas eficientes que garantam a subsistência das famílias e o defeso das espécies.

Ao se considerar o rio Araguaia, observa-se que este é uma importante zona de transição entre o cerrado e a floresta amazônica [2] e compõe junto com o rio Tocantins, a segunda maior bacia hidrográfica e mais extensa em termos de área de drenagem integralmente situada no território brasileiro [3]. O Araguaia está numa região que agrega cerca de 40% dos pescadores artesanais do Brasil [4], porém como há escassez de programas de monitoramento de desembarques na bacia Araguaia-Tocantins, a falta de informações contínuas sobre captura e esforço, influencia negativamente na avaliação dos estoques e produtividade da pesca, assim como aumenta o grau de incerteza sobre a aplicação de políticas de proteção dos recursos pesqueiros [5].

Dessa forma, obter dados confiáveis sobre a estatística pesqueira na Amazônia ainda é uma tarefa difícil, demorada e onerosa. O déficit de dados estatísticos oficiais no Brasil, referentes à sazonalidade da produção por município, estado e/ou espécie capturada, dificulta um planejamento adequado para o ordenamento da atividade e dos estoques continentais a longo prazo. Esse déficit de dados prejudica significativamente a construção de políticas públicas mais ajustadas ao setor, e que são prioritárias para o desenvolvimento da região Norte e demais regiões brasileiras [6].

Assim, a partir de dados coletados em comunidades dos municípios de Itupiranga (PA) e São João do Araguaia (PA), em especial dados de desembarque, durante o anos de 2016 e 2017, neste trabalho são realizadas análises de conhecimento aplicando técnicas de Inteligência Computacional que permitam identificação de padrões e uma análise inicial da dinâmica pesqueira na bacia Araguaia-Tocantins.

2. MATERIAL E MÉTODOS

A partir das atividades de monitoramento de desembarque, foram utilizadas variáveis como: a data da pesca especificando quando ocorreu a pesca; o rio em que a pesca foi realizada; o período lunar (Nova, Crescente, Cheia e Minguante), utilizados para relacionar com os dados da pesca; os locais de pesqueiros que especifica a comunidade ou localidade em

que a pesca foi realizada; os ambientes de pesca que descreve o lugar em que foi feita a pescaria como beira do rio, meio do rio, pedral entre outros; as horas totais de pesca que envolvem informações sobre o tempo de trajeto até o local pesqueiro, e o tempo de pesca propriamente dito; kg totais onde é especificada a quantidade total em quilos, somando respectivamente a quantidade peixe em kg para consumo próprio com a quantidade em kg para venda; o número de pescadores que participaram da pesca; a renda líquida, que representa os ganhos com a pesca, excluídos os gastos para realizar a mesma.

Também foram utilizadas variáveis como os instrumentos utilizados para pescar que foram: Malhadeira, Linha de mão, Ceva, Caniço, Espinhel, Zagaia, Flecha, Arpão, Tarrafa, entre outros instrumentos. Assim como as espécies de peixes pescadas, que contemplam espécies típicas da amazônia central, bem como espécies endêmicas da região e espécies não-amazônicas.

Para este estudo inicial, foi utilizada a abordagem de clusterização de dados, método utilizado na mineração de dados, que parte da idéia que dados pertencentes a um mesmo grupo tem mais afinidades entre si, que dados pertencentes a outros grupos. A técnica consiste na utilização unicamente dos dados contidos na base de dados (que podem ser multivariados), que de forma não supervisionada, esses mesmos possam ser automaticamente agrupados em *k* grupos, que são denominados *clusters*. Neste trabalho são consideradas 02 (duas) técnicas de clusterização distintas: A técnica *Expectation Maximization Algorithm* (EM), e a técnica *K-Means*.

A técnica EM segue uma abordagem interativa, subótima, tentando buscar parâmetros da distribuição de probabilidade que tenham a máxima probabilidade de seus atributos mesmo com a presença de dados ausentes/latentes [7], se aproximando da função máxima da verossimilhança, sendo o objetivo do algoritmo maximizar a probabilidade global ou a probabilidade dos dados, dado os clusters.

Com relação ao algoritmo K-means, este algoritmo assim como anterior trabalha de forma não supervisionada, classificando um determinado conjunto de dados através de um número *k* de clusters previamente especificados. São definidos *k* centróides para cada um dos clusters, os dados do conjuntos são associados ao centróide mais próximo com base na distância euclidiana. Por fim, cabe destacar que a ferramenta Weka[1] foi utilizada para a execução do algoritmo EM, enquanto que a ferramenta R[2] utilizada para o algoritmo K-means.

Com base nos dados utilizados, a estratégia inicial esteve baseada na determinação ideal do número de clusters ou agrupamentos a serem utilizados na referida análise, de modo que o grau de similaridade entre duas instâncias de dados quaisquer refletir o padrão de agrupamento dos dados analisados. Ao se considerar o Algoritmo de *Expectation Maximization* (EM), o mesmo utiliza técnicas de validação cruzada (cross-validation) que podem ser aplicadas para estimar o número de agrupamentos de uma amostra ou distribuição [8, 9]. A partir da utilização da técnica em questão, a quantidade de clusters estimada é de 03 (três) unidades.

Com relação à utilização do algoritmo k-means, o mesmo precisa ser parametrizado com o número de clusters desejados previamente, e assim, uma questão fundamental é como determinar o valor deste parâmetro. Assim, foi realizado uma análise inicial com base na porcentagem de variação obtida em função do número de clusters considerados. Assim, conforme a Figura 01, o grau de variabilidade tende a estabilizar a partir de 03 a 04 clusters,

[1] Weka: Data Mining Software in Java - https://www.cs.waikato.ac.nz/ml/weka/
[2] R: Project for Statistical Computing - https://www.r-project.org/

uma vez que soma das áreas dos clusters tende a reduzir, conforme comportamento “joelho” observado em tal gráfico.

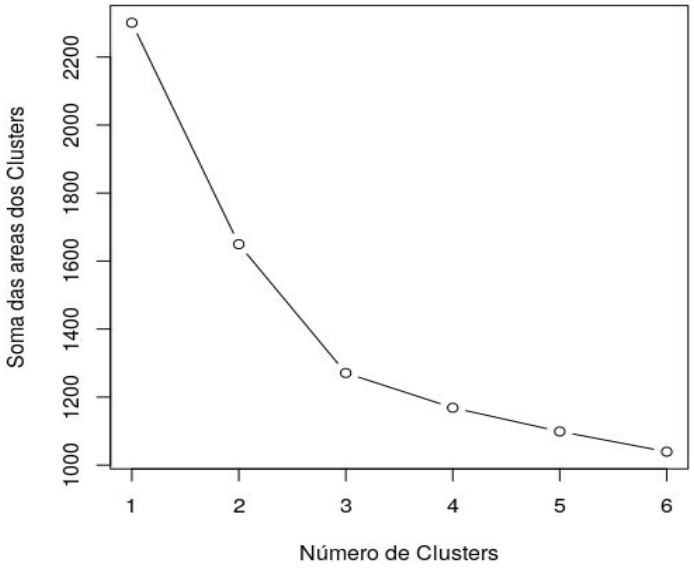

Figura 01: Soma das áreas dos clusters em função da quantidade de clusters.

3. RESULTADOS E DISCUSSÃO

Assim, ao se considerar o parâmetro k (número de clusters) como sendo 03 unidades, observou-se que os três clusters possuem assimetrias significativas em relação ao número de amostras classificadas, para ambos os algoritmos utilizados. Para o algoritmo K-means, os clusters 01, 02 e 03 se apresentaram respectivamente com as seguintes populações: 84 (34,56%); 152 (62.55%); e 07 (2,88%). O algoritmo EM se apresentou as seguintes populações para os clusters 01, 02 e 03, respectivamente: 97 (40%); 117 (48%); 32 (12%). Conforme pode ser observado, os clusters 01 e 02 possuem tamanhos razoavelmente proporcionais em ambos os algoritmos, variação esta que pode ser justificada pelas especificidades de cada algoritmo, bem como pelo grau de interseção presente, ao se observar a Figura 02.

Adicionalmente, é importante observar que também em ambas as técnicas, o cluster 03 apresenta uma menor população em comparação com os demais clusters, que pode indicar a presença de um fenômeno específico na amostra de dados, ou ainda ruídos que devem ser melhor tratados em análise futuras e mais complexas, porém o mesmo é marcado por comportamento diferenciado no que tange aspectos de renda líquida e quantidade de pescado (kg), bem acima das médias dos clusters 01 e 02, característica esta que deve ser analisada em trabalhos futuros.

Em adição ao exposto, a partir das análises realizadas sobre os dados pesqueiros coletados na bacia Araguaia-Tocantins, observa-se que ainda que a técnica de clusterização indique claramente de modo evidente que a amostra de dados pode ser agrupada em pelo menos 02 (dois) agrupamento expressivos. O primeiro agrupamento se configura por atividades de pesca realizadas principalmente no rio Tocantins, tendo como base de maior volume o mês de Julho, em especial no período lunar Cheio, com uma renda líquida média de R$ 148,00 por desembarque, com em média tendo aproximadamente 02 (dois) pescadores, utilizando malhadeira e tarrafa como instrumentos de pesca. No segundo agrupamento se configura por atividades localizadas principalmente Araguaia, tendo como base de maior

volume o mês de Maio, em especial na lua crescente, com uma renda líquida média de R$ 112,00 por desembarque, com uma média de 02 pescadores, utilizando principalmente a malhadeira como instrumento de pesca. Tais características de ambos os agrupamento eventualmente podem considerar conhecimento tácito já de conhecimento empíricos da população de pescadores, porém compatível com as expectativas deste trabalho.

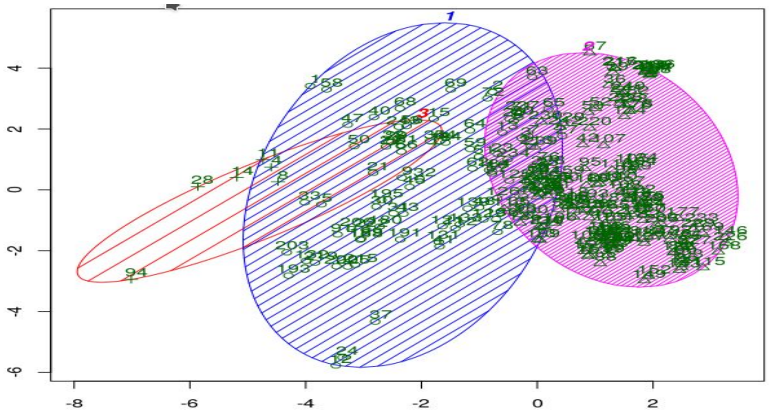

Figura 02: Representação Gráfica 2D da solução de clusterização.

4. CONCLUSÃO

A partir das informações geradas com base na utilização de técnicas de Inteligência Computacional, pode-se considerar que a rede de emalhe (malhadeira) é uma das principais ferramentas empregadas na bacia Araguaia-Tocantins para pesca, sendo complementada por demais tipos de instrumentos, porém como não há um programa institucional de monitoramento de desembarques, há ainda falta de informações que possam dar melhor embasamento estatístico. Como trabalhos futuros pretende-se aplicar refinar o estudo atual com técnicas adicionais, bem como realizar a adoção de técnicas de IC complementares para explorar melhor as características dos grupos clusterizados.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Beard, Jr. et al. The World report on ageing and health: a policy framework for healthy ageing, The Lancet,The Lancet, Volume 387, Issue 10033,2016, Pages 2145-2154, ISSN 0140-6736, https://doi.org/10.1016/S0140-6736(15)00516-4.

[2] Ferreira et al., The fish fauna of the Parque Estadual do Cantão, Araguaia River, State of Tocantins, Brazil. 2011. Biota Neotropica.

[3] ANA (Brasil), Plano Estratégico de recursos hídricos da bacia dos rios Tocantins e Araguaia: relatório síntese / Agência Nacional de Águas. Brasília: ANA; SPR, 2009. 256 p.: II. ISBN 987-85-89629-55-3.

[4] Santos, M.A.S. The production chain of artisanal fisheries in the state of Pará: a case study from Northeast Pará. Amazônia: Ciência e Desenvolvimento, Belém, 2005.

[5] Dias, C. Silva, A. A rede de emalhe na pesca artesanal do rio Araguaia, Tocantins, Brasil. Anais of the XX CONBEP (Congresso Brasileiro de Engenharia de Pesca) 2017 Out 08-11; Florianopólis.

[6] Silva, A.P. da, Dinâmica da pesca, produtividade e composição de captura da frota motorizada de Valença-BA, Brasil. Embrapa Pe., Palmas, 2013.

[7] Neto, J. Expectation-Maximization [Internet], 2016. Available from: http://www.di.fc.ul.pt/~jpn/r/EM/EM.html.

[8] Wei Fu, Patrick O. Perry. Estimating the number of clusters using cross-validation, 2017. Available from: https://arxiv.org/pdf/1702.02658.pdf.

[9] Kawamoto, T.; Kabashima, Y.; Cross-validation estimate of the number of clusters in a network. Nature: *Scientific Reports* **7**, Article number: 3327. 2017. doi:10.1038/s41598-017-03623-x.