



SUAVIZAÇÃO HIPERBÓLICA: UMA NOVA METODOLOGIA DE CLASSIFICAÇÃO EM TAXONOMIA BIOLÓGICA

Maria Gardênia Sousa Batista – Universidade Estadual do Piauí, Laboratório de Ficologia e Limnologia – Teresina, PI – batistamariagardenia@gmail.com;

Adilson Elias Xavier – Universidade Federal do Rio de Janeiro, PESC/COPPE – Rio de Janeiro, RJ Francisca Lúcia de Lima - Universidade Estadual do Piauí, GERATEC – Teresina, PI André Macedo Santana – Universidade Federal do Piauí – Departamento de Informática e Estatística – Teresina, PI

INTRODUÇÃO

No que se referem à classificação dos seres vivos, os critérios são modificados de acordo com o tipo de relação que o homem estabelece com a natureza (Nagem *et al*, 2001). A necessidade da interação entre as diferentes áreas da Biologia e da Matemática Computacional traz um profundo impacto sobre as práticas em sistemática, tornando rápido e sofisticado e acima de tudo eficiente os dados taxonômicos, bem como facilita a disseminação e intercambio de informações. Atualmente, a sistemática ocupa uma posição central na biologia evolutiva e está desempenhando uma função cuja importância tem aumentado gradativamente para outras disciplinas como a informática. (Judd *et al.*, 2009; Amorim, 2002; Hickman *et al.*, 2004). A necessidade de buscar por uma forma sistemática para encontrar grupos em dados levou ao desenvolvimento de técnicas para resolver este problema e são conhecidas como “agrupamento de dados” (Kaufman & Rousseeuw 1990, Everitt *et al.* 2001, Gordon 1999), “taxonomia numérica” ou, ainda, “classificação automática de dados”. O problema de cluster é de interesse em diversas áreas que necessitam agrupar dados, assim, nos últimos anos, tem-se intensificado o número de trabalhos publicados utilizando-se da análise de agrupamentos. A importância, do estudo dos métodos de agrupamento, na comunidade científica, pode ser observada pela publicação da temática em diversos jornais de extrema importância tais como, são apresentados por Xu & Wunsch (2009). A aplicação dos métodos de análise de cluster tem produzido uma dinâmica de melhoramento e produção de novos métodos Bagirov & Yearwood (2006), Xavier (2009), a fim de viabilizar sua utilização e desempenho em diversas áreas.

OBJETIVOS

No propósito de contribuir modestamente para o avanço na sistemática é que ora se apresenta uma nova metodologia para agrupamento de dados biológicos, através do uso de um algoritmo novo e mais eficiente – Metodologia de Classificação via Suavização Hiperbólica (HSCM).

MATERIAL E MÉTODOS

A nova metodologia proposta para agrupamento de dados biológicos utiliza o algoritmo (HSCM). Nesse algoritmo a solução é obtida através da resolução de uma sequência de subproblemas irrestritos. As minimizações irrestritas foram realizadas por meio de um algoritmo Quase-Newton, com fórmula de atualização BFGS, da Harwell Library. Esse algoritmo foi codificado na linguagem Fortran 77, utilizando-se o compilador Digital Visual Fortran versão 6.0A. Para demonstração do desempenho desse algoritmo de agrupamento, foram utilizados os dados de biometria extraídos da morfologia de três estruturas de macroalgas do gênero *Caulerpa* (Barata, 2008), tendo como vetores de características, obtidos da medida de altura mínima e máxima dos ramos assimiladores; do diâmetro mínimo e

máximo do estolão e dos rizoides das algas.

RESULTADOS

São apresentados agrupamentos baseados no uso do algoritmo do tipo partição, HSCM, que adota um enfoque de suavização hiperbólica. Mostrar-se-á um resultado diferenciado quanto à representação taxonômica, em contraposição as metodologias hierárquicas com uma visão multidimensional, através da formação de “cluster” nesse propósito, suscitaram-se a produção de desenhos não lineares, de acordo com os designios específicos da aplicação do especialista. Sendo aqui utilizados para análise das algas que puderam ser agrupadas com base nas características morfológicas e comparadas com análises filogenéticas.

DISCUSSÃO

Como a atividade do sistemata é um processo trabalhoso e, muitas vezes, lento e exaustivo (Senna & Magrin, 1999), nossa perspectiva é de poder ajudar ao sistemata a conhecer e entender o fechamento do ciclo, ou seja, daquele laborioso processo de identificação das espécies que produz uma grande quantidade de informações que servirá para uma interpretação adequada do ecossistema e conseqüentemente de novas abordagens científicas. O método matemático utilizado nesse trabalho teve a capacidade de separar as espécies morfológicamente semelhantes (*C. ashmeadii*, *C. sertularioides*, *C. mexicana*, e *C. taxifolia*) em clusters distintos, apesar de serem próximos, o que demonstra que apesar de apresentarem grande plasticidade as algas puderam ser agrupadas com base nas características morfológicas e comparadas com análises filogenéticas. Olsen e colaboradores (2006) relataram em um universo de 241 amostras do Gênero *Caulerpa*, 12,7% foram classificados morfológicamente de forma errada por Ficologistas experientes, que demonstraram que as espécies de *C. ashmeadii*, *C. sertularioides*, *C. mexicana*, e *C. taxifolia* são morfológicamente próximas e podem ser confundidas. Outro exemplo da eficiência demonstrada com o uso do método foi a obtenção de *C. scalpelliformes* em um grupo separado e isolado dos demais. Dados da literatura de trabalhos por análise filogenética indicam que *C. scalpelliformes* é uma espécie separada das outras, ficando num clado isolado e terminal, em consonância com os nossos resultados (Pillmann, *et al*, 1997 e Barata, 2008).

CONCLUSÃO

Temos a veleidade, que não ostentamos uma proposta que sanou as fragilidades da metodologia hierárquica, mas que se entende ser uma ferramenta que pode oferecer uma alternativa para enriquecimento na interpretação taxonômica, podendo ser utilizada a partir da taxonomia consagrada. É importante destacar que a nova metodologia aqui apresentada, mesmo utilizando somente dados morfológicos foi possível comparar com os resultados obtidos de experimentos moleculares.

REFERÊNCIAS BIBLIOGRÁFICAS

- Amorim, D.S. Fundamentos de sistemática filogenética. Ribeirão Preto: Editora Holos, 2002. Bagirov, A.M. & Yearwood, J. (2006), A new nonsmooth optimization algorithm for minimum sum-of-squares clustering problems, *European Journal of Operational Research*, 170(2), 578-596.
- Barata, D., 2008. Taxonomia e Filogenia do Gênero *Caulerpa* J. V. Lamour. (Bryopsidales, Chlorophyta) no Brasil. D.Sc. Thesis – Instituto de Botânica, São Paulo. Everitt, B.S, Landau S., Leese M. (2001) *Cluster Analysis*, 4th edn. Arnold, London.
- Gordon, A.D. 1999. How many clusters? An Investigation of five procedures for detecting nested cluster structure. In: *Data Science, Classification, and Related Methods*, edited by C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H. Bock, and Y. Baba. Tokyo: Springer-Verlag. Hickman, J.R., C. P.; Roberts, L. S.; Larson, A. *Princípios Integrados*

de Zoologia. 11. ed., Ed. Guanabara Kogan: Rio de Janeiro, 2004.

Judd, W. S.; Campel, C. S., Kellongg, E. A., Steens P. F.; Donogue, M. J. Sistemática Vegetal: um enfoque filogenético. 3.ed. Porto Alegre: Artmed, 2009. 612p.

Kaufman, L., & Rousseeuw, P. J. Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons, Inc. 1990.

Nagem, R.L., Carvalhes, D.O. & Dias, J. A.Y. Uma proposta de metodologia de ensino com analogias. Revista Portuguesa de Educação, (2).14. 2001.

Senna, P.A.C.; Magrin, A.G.E. A importância da “boa” identificação dos organismos fitoplanctônicos para os estudos ecológicos. In: Pompêo, M.L.M. (ed.) Perspectiva da Limnologia no Brasil. São Luís: Gráfica e Editora União, 1999. p. 131-146. 1999.

Xavier, A. E. The Hyperbolic Smoothing Clustering Method; Pattern Recognition Letters. 2009, doi 10.1016/j.patcog.2009.06.018. 2009. Xu, R., Wunsch, D., Clustering, IEEE Press, 2009.

Agradecimento