

13ª Mostra da Produção Universitária

Rio Grande/RS, Brasil, 14 a 17 de outubro de 2014.

APRENDIZADO ATIVO E SEMI-SUPERVISIONADO NA CLASSIFICAÇÃO DE MICROALGAS

FARIA, Matheus Frantz
DREWS, Paulo Lilles
(autor/es)

TAVANO, Virginia (orientador)

mffaria@furg.br

Evento: Congresso de Iniciação Científica

Área do conhecimento: Ciências exatas e da terra, ciência da computação

Palavras-chave: aprendizado, semi-supervisionado, microalgas

1 INTRODUÇÃO

O trabalho consiste na classificação automatizada de microalgas obtidas a partir de uma amostra de água. As microalgas são organismos unicelulares que apresentam diversas formas, tamanhos e estruturas. Utilizar uma classificação de forma manual dessas microalgas acaba sendo muito custoso, já que em uma pequena amostra de água analisada podem ser encontradas milhares de microalgas.

2 REFERENCIAL TEÓRICO

A metodologia é baseada no Modelo de Misturas Gaussianas e em técnicas de aprendizado de máquina. O modelo de misturas Gaussianas (GMM) é uma função densidade de probabilidade (PDF) dada por uma combinação linear de distribuições de densidade de probabilidade Gaussianas [McLachlan and Peel, 2000].

3 PROCEDIMENTO METODOLÓGICO

Os dados sobre as microalgas são obtidos através do equipamento chamado FlowCAM, um equipamento para análise óptica que obtêm imagens das microalgas. Porém, este equipamento não possui capacidade de classificar automaticamente as microalgas. A partir dessas imagens diversos parâmetros para cada microalga são estimados: largura, altura, etc.

Neste trabalho é proposto o uso de GMM com aprendizado semi-supervisionado e aprendizado ativo. O aprendizado semi-supervisionado é aquele no qual o especialista humano classifica uma pequena porcentagem dos dados (<5%) e o algoritmo generaliza para os outros dados. Enquanto o aprendizado ativo, o especialista humano é convidado a classificar uma pequena quantidade de dados que pode maximizar o ganho de informação.

Foi utilizado o aprendizado semi-supervisionado para inicializar a classificação. Com essa classificação inicial estima-se a GMM por meio do algoritmo

13ª Mostra da Produção Universitária

Rio Grande/RS, Brasil, 14 a 17 de outubro de 2014.

Expectation-Maximization (EM) [Dempster et al. 1977].

Após a execução do aprendizado semi-supervisionado, utiliza-se o aprendizado ativo, que divide os dados em dois grupos: classificados e não-classificados. Dentre as não-classificadas, o algoritmo escolhe aqueles dados que maximizam a quantidade de informação.

4 RESULTADOS e DISCUSSÃO

Resultados foram obtidos em um conjunto de dados reais de testes, obtendo uma classificação entre 80% e 90% de acertos na classificação das microalgas, com uma supervisão total entre 5% a 10% dos dados. Resultado expressivo visto que especialista conseguem manualmente classificar corretamente entre 67% e 83%, segundo a literatura da área [Culverhouse et al. 2003].

5 CONSIDERAÇÕES FINAIS

O trabalho em questão propôs o uso de aprendizado ativo e semi-supervisionado para a classificação dos dados obtidos, tendo excelentes resultados em seus testes e boa aceitação na sua implementação.

REFERÊNCIAS

Culverhouse, P., Williams, R., Reguera, B., Herry, V., and González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology-progress Series*, 247:17–25.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38.

Drews-Jr, P. , Colares, R. G., Machado, P., Faria, M., Detoni, A., Tavano, V. (2012) Aprendizado Ativo e Semi-Supervisionado na Classificação de Microalgas. Encontro Nacional de Inteligência Artificial (ENIA).

McLachlan, G. and Peel, D. (2000) *Finite Mixture Models*. Wiley, New-York,