

I SIMPÓSIO DE APLICAÇÕES DA INFORMÁTICA EM BIOLOGIA

Campinas, 4 a 6 de agosto de 1993
Instituto de Matemática, Estatística e Computação

ANAIS DE RESUMOS

Universidade Estadual de Campinas



EDITOR

Renato M.E. Sabbatini
Universidade Estadual de Campinas

PROMOÇÃO

Núcleo de Informática Biomédica
Disciplina de Informática Médica da Faculdade de Ciências Médicas Universidade Estadual de
Campinas

APOIO

Instituto de Biologia Empresa Junior de Biologia Universidade Estadual de Campinas
Sociedade Brasileira de Informática em Saúde

COMISSÃO ORGANIZADORA

Renato M.E. Sabbatini (NIB e FCM/UNICAMP)
Marcelo Okamoto Tanaka (IB/UNICAMP e EJB)
Carlos Eduardo Nascimento de Lima (NIB/UNICAMP e EJB)
João Meidanis (IMECC/UNICAMP)
Raul Neder Porrelli (NIB/UNICAMP)
Hilton Silveira Pinto (CEPEAGRI e CCUEC/UNICAMP)
Jacques Vieilliard (IB/UNICAMP)
Alexandre Gianini Martins (NIB/UNICAMP)

SUMÁRIO

Conteúdo

EDITOR.....	2
PROMOÇÃO.....	2
APOIO	2
COMISSÃO ORGANIZADORA	2
INFORMÁTICA E BIOLOGIA: A NOVA ALIANÇA.....	6
Processamento de Sinais e Imagens	7
DIGITAL IMAGE PROCESSING APPLICATIONS TO BIOLOGY.....	7
INTELIGÊNCIA ARTIFICIAL E REDES NEURAIAS EM PROSPECÇÃO AMBIENTAL E SENSORIAMENTO REMOTO	8
INTERPRETAÇÃO DE IMAGENS DE SATÉLITE DA AMAZÔNIA USANDO REDES NEURAIAS	9
ANÁLISE DIGITAL DE GLOMÉRULOS RENAIAS	9
PROCESSAMENTO DIGITAL DE VOCALIZAÇÕES.....	11
FRAGMENTAÇÃO AUTOMÁTICA DO SONO UTILIZANDO REDES NEURONAIAS	12
TÉCNICAS PARA ANÁLISE DE RITMOS BIOLÓGICOS	12
TESTE PARA DETECTAR PERIODICIDADES COMPOSTAS EM SÉRIES TEMPORAIAS.....	13
Aplicações em Biologia Molecular e Engenharia Genética	13
ENGENHARIA GENÉTICA E INFORMÁTICA.....	13
TRÊS MÉTODOS PARA PREVISÃO DE ESTRUTURA SECUNDÁRIA DE PROTEÍNAS, VERSÕES PARA MICROS.....	14
PREVISÃO DE ESTRUTURA SECUNDÁRIA DE PROTEÍNAS USANDO MODELOS LOGÍSTICOS	15
SISTEMAS DE AUXÍLIO AO SEQÜENCIAMENTO DE DNA EM LARGA ESCALA	15
Bases de Dados e Telemática	16
THE BRAZILIAN MOLECULAR BIOLOGY AND BIOTECHNOLOGY NETWORK (BMBBNet)	16
ICGEBnet: AN INTERNATIONAL BIOLOGICAL COMPUTER RESOURCE.....	17
BASES DE DADOS EM BIOLOGIA.....	18
BASE DE DADOS TROPICAL: DISSEMINAÇÃO DE INFORMAÇÕES PARA BIODIVERSIDADE E BIOTECHNOLOGIA.....	19
BIN21 - THE BIODIVERSITY INFORMATION NETWORK.....	20
ECOLOG: UM SISTEMA GERENCIADOR DE BANCOS DE DADOS PARA LEVANTAMENTOS ECOLÓGICOS DE CAMPO	21
Modelagem e Simulação	22

MODELAGEM MATEMÁTICA E SIMULAÇÃO EM BIOLOGIA	22
A SIMPLE MODEL FOR MICELLIZATION: SIMULATION EXPERIMENT	23
ENACT: AN ARTIFICIAL LIFE WORLD IN A FAMILY OF CELLULAR AUTOMATA.....	23
SIMULAÇÃO DA DINÂMICA DE POPULAÇÕES DE PLANTAS	24
UM MODELO PARA O EFEITO DE ANTAGONISTAS DO ESTRÓGENO NA LIGAÇÃO COOPERATIVA DO ESTRADIOL.....	24
Inteligência Artificial e Reconhecimento de Padrões	25
REDES NEURAIS E ALGORITMOS GENÉTICOS EM BIOLOGIA.....	25
BEHAVIOR-BASED ACTIVE VISION	26
CLASSIFICAÇÃO TAXONÔMICA DE BACTÉRIAS ATRAVÉS DE REDES NEURAIS	26
USO DE REDES NEURAIS EM TAXONOMIA NUMÉRICA BASEADA EM DADOS HETEROGÊNEOS	27
RECONHECIMENTO DE SEQÜÊNCIAS COMPORTAMENTAIS UTILIZANDO REDES NEURAIS ARTIFICIAIS	28
Aplicações Educacionais.....	28
USO DE PROGRAMAS PARA ENSINO DE GENÉTICA	28
SIMULAÇÃO DA DERIVA GENÉTICA EM POPULAÇÕES SUBMETIDAS OU NÃO A PROCESSOS DE SELEÇÃO: UM PROGRAMA DE APOIO AO ENSINO DE GENÉTICA DE POPULAÇÕES.....	29
CASOS CLÍNICOS SIMULADOS PARA COMPLEMENTAÇÃO DO ESTUDO DAS ABERRAÇÕES CROMOSSÔMICAS.....	30
Outras Aplicações.....	30
COMPARAÇÃO DE MÉTODOS DE INFERÊNCIA FILOGENÉTICA.....	30
CONJUNTO DE ROTINAS PARA AVALIAÇÃO DA CAPACIDADE SECRETORA E SENSIBILIDADE PERIFÉRICA À INSULINA.....	31

PREFÁCIO

Este pequeno volume reúne as contribuições apresentadas ao I Simpósio de Aplicações da Informática na Biologia (InfoBio'93), realizado na Universidade Estadual de Campinas, SP, em agosto de 1993. O InfoBio'93 foi uma iniciativa conjunta do Núcleo de Informática Biomédica (NIB), da Disciplina de Informática Médica do Departamento de Genética Médica da Faculdade de Ciências Médicas e da Empresa Junior de Biologia (EJBU), todos da Universidade Estadual de Campinas (UNICAMP). Esta iniciativa recebeu o apoio oficial do Instituto de Biologia (IB) da UNICAMP e da Sociedade Brasileira de Informática em Saúde (SBIS). Embora este não tenha sido o primeiro evento do gênero (vários outros congressos biológicos, como os de Zoologia, e algumas reuniões da Sociedade Brasileira para o Progresso da Ciência e da Federação das Sociedades de Biologia Experimental, tiveram em seu bojo iniciativas similares), queremos crer que este é o primeiro evento isolado de relativo porte realizado no Brasil. Como parte da missão institucional do Núcleo de Informática Biomédica (NIB) da UNICAMP, esta a difusão, pesquisa e ensino de novas tecnologias baseadas na Informática nos campos de Biologia e Saúde. Entretanto, os biólogos demoraram a acordar para o potencial e as novas realidades proporcionadas pela Informática em sua área. Em consequência, a demanda inicial sobre o NIB foi bem maior a partir da área médica, a qual se organizou mais cedo do que as demais (o primeiro Simpósio foi em 1985, e o primeiro Congresso Brasileiro em 1986, juntamente com a primeira revista especializada e a primeira sociedade científica na área). Em julho de 1993 o NIB completa 10 anos de atividade. Muito apropriadamente, frente à importância cada vez maior que a Informática assume perante as atividades de pesquisa e ensino em Biologia, resolvemos comemorar o aniversário promovendo um evento específico com uma programação inteiramente dedicada às aplicações em Biologia, e a computação inspirada biologicamente. Embora o número de trabalhos apresentados tenha sido relativamente pequeno, reflete com fidelidade a variedade das áreas de interação entre Biologia e Informática, e demonstra o rápido crescimento desta área multidisciplinar no Brasil. Uma idéia do valor da Informática para a Biologia pode ser dada pelo fato de que, pela primeira vez em nosso grupo, um congresso científico de certa complexidade foi inteiramente organizado utilizando-se o correio eletrônico (redes acadêmicas Internet, Bitnet e Rede Nacional de Pesquisa), desde a divulgação do evento e das sucessivas versões do seu programa, até o envio de todos os resumos e a correspondência com todos os palestrantes. Outro fato significativo é que uma versão eletrônica destes anais foi tornada disponível para acesso remoto, em um computador da UNICAMP ligado a rede Internet. Esperamos sinceramente que este evento não seja o último da série (como infelizmente acontece tantas vezes no Brasil), e que outros venha a ocorrer, fomentando o desenvolvimento de um setor ainda carente no Brasil, mas que já é big science nos países mais desenvolvidos.

Campinas, agosto de 1993

Renato M.E. Sabbatini

Presidente da Comissão Organizadora

RESUMOS

INFORMÁTICA E BIOLOGIA: A NOVA ALIANÇA

Sabbatini, R.M.E.

Depto. Genética Médica, Faculdade de Ciências Médicas e Núcleo de Informática Biomédica da UNICAMP, Campinas, SP Email: sabbatini@ccvax.unicamp.br

Gradualmente, nos últimos cinco anos, a Biologia tem se apropriado com avidez das ferramentas proporcionadas pela Informática. Esta progressão tem sido tão rápida e avassaladora, que já existem não apenas uma, mas varias subespecialidades da Biologia e da Informática que formam um mosaico multidisciplinar riquíssimo, onde aos poucos a antiga distinção entre as duas áreas se dissolve. Biologia computacional, genética computacional, ecologia matemática, algoritmos genéticos ou evolutivos, redes neurais artificiais, são hoje varias dessas denominações que tipificam a interação entre Informática e Biologia; nos dois sentidos: desde conceitos de Biologia, como evolução natural e auto- organização de tecidos vivos, inspirando a criação de novas áreas das ciências exatas; até novas e poderosas técnicas sendo criadas com a finalidade especifica de resolver complexos problemas de processamento de dados e conhecimentos biológicos, tais como a comparação de seqüências de polímeros biológicos, os bancos de dados genéticos, a predição de estruturas moleculares, etc. De forma mais fascinante ainda, constata-se que o circulo de influencia mutua entre as duas disciplinas se fecha, quando se constata a cada vez mais intensa utilização de tecnologias biologicamente inspiradas (por exemplo, algoritmos genéticos), para resolver problemas da própria Biologia ! Um dos grandes estímulos para o desenvolvimento acelerado da Informática Biológica tem sido, sem duvida, o esforço conjugado de mapeamento do genoma de diversas espécies. Sem a ajuda da Informática, qualquer um desses esforços, mesmo os dos organismos mais simples, como a levedura, seria impossível. Tanto é assim que o famoso projeto HUGO (mapeamento do genoma humano) tem investido grandes somas de dinheiro na infraestrutura computacional. Outro estímulo importante é o crescimento de diversas linhas de pesquisa, reunidas sob o nome de Sistemas Complexos, que procura caracterizar, modelar e simular (novamente com a ajuda de computadores) diversos sistemas biológicos termodinamicamente abertos, como sistemas auto-organizados, aprendizado, sociedades animais, sistemas caóticos, transmissão e armazenamento de informação genética, etc. Um grande estímulo para a Biologia Computacional e para a Informática Biológica tem sido as bases de dados biológicos informatizadas e as redes acadêmicas internacionais. Atualmente, e impensável realizar pesquisa, desenvolvimento e produção em certas áreas da Biologia, como Engenharia Genética, Taxonomia, etc., sem o acesso as bases de dados contendo seqüências de DNA, genomas, proteínas, organismos, etc. O acesso internacional e a contribuição acelerada a essas bases, assim como a comunicação entre os biólogos e demais profissionais trabalhando na área é enormemente facilitada pelas redes Internet, Bitnet e outras, que por sua vez servem de esqueleto para implementar numerosas redes conceptuais, tais como a do Centro Internacional de Engenharia Genética, de Trieste; da Rede Brasileira de Biologia Molecular, da EMBRAPA, da Base de Dados Tropical, da Fundação "André Tosello", etc. Algumas

organizações de pesquisa biológica, como o European Molecular Biology Laboratory (EMBL), em Heidelberg, e outras, assumiram o papel de gigantescas provedoras de informação em forma eletrônica em Biologia. O crescimento da Informática Biológica e da Biologia Computacional também se reflete no grande número de novos títulos de revistas especificamente dedicadas ao tema, congressos, livros, etc.; e até os primeiros cursos de pós-graduação. A realidade do dia-a-dia do biólogo foi irreversivelmente mudada com o uso do microcomputador e da estação de trabalho como ferramentas de acesso, armazenamento e manipulação de dados. É possível que a Biologia seja a primeira ciência natural complexa a transpor os limiares estabelecidos pelas ciências eminentemente quantitativas, como a Física. Se isso acontecer, certamente será por mérito quase exclusivo da Informática, que se constitui como a única metodologia capaz de enfrentar a complexidade matemática dos sistemas biológicos.

Processamento de Sinais e Imagens

DIGITAL IMAGE PROCESSING APPLICATIONS TO BIOLOGY

Amman, J.-J.

Instituto de Biologia da UNICAMP, Campinas, SP Email: jjacques@ibi.unicamp.br

Images have always been one of the most useful tools in biological sciences. Classification of living beings and determination of their structure is principally based on images obtained from naked eye observations, optical microscopes and then electron microscopes, both scanning and transmission, and now from the new generation of tunnelling and atomic force microscopes and from high resolution light microscopes. If the image acquisition is a primordial step in any of these techniques, the analysis of images is not, by far, a simple task. Basically due to the very high amount of information caught in an image, extraction the information that is really significant and useful for a specific research has long had the only support of the extraordinary power of the human brain. However, modern research requires higher and higher precision along with a satisfactory reproducibility in the measurements, two properties in which the human brain is not so comfortable. This is especially critical in the biomedical domain where objects are often subject to large variation, and where the effects of a treatment are often small and require a large scale of time to develop. Thus the development of more powerful computers along with a large flexibility and good users interfaces is giving to image analysis a new scope which will soon become as important as the step of image acquisition itself for the researcher. We will try to present here some basic ideas on digital image processing and analysis and show how these powerful tools can be used in biology. As illustrations, several examples will be shown.

INTELIGÊNCIA ARTIFICIAL E REDES NEURAIS EM PROSPECÇÃO AMBIENTAL E SENSORIAMENTO REMOTO

Engel, P.M.

Instituto de Informática da Universidade Federal do Rio Grande do Sul, Porto Alegre, RS Email: engel@inf.ufrgs.br

A classificação de imagens baseada em computador é de grande importância em muitas aplicações. Especialmente imagens multiespectrais de superfícies, tomadas por satélites de sensoriamento remoto, podem contribuir de forma destacada na área de proteção ambiental, por exemplo, para supervisionar e interpretar danos em florestas. Para esta finalidade, os dados referentes às imagens obtidas devem ser classificados por métodos apropriados de reconhecimento de padrões óticos, após serem preprocessados adequadamente. Métodos convencionais usualmente aplicam uma classificação de Bayes combinada com uma estimativa do tipo máxima probabilidade, ou ainda com métodos de regressão, parametrizados anteriormente a classificação, por áreas de treinamento selecionadas. Ambos estes métodos apresentam restrições em relação à morfologia dos clusters individuais a serem separados no espaço de característica. Por este motivo, os resultados obtidos com estes métodos são somente em parte satisfatórios, principalmente para vetores de característica que não sejam distribuídos de forma gaussiana. Além disso, para os vários canais espectrais o vetor de característica inclui não só valores de pixel puro, mas também informação de textura. A análise desta situação no espaço de característica e a avaliação de áreas de treinamento podem ser realizadas por métodos de análise de clusters, não-supervisionados. Uma alternativa bastante favorável para os métodos convencionais de classificação se constitui no emprego de redes neurais artificiais (RNA). Hoje em dia se dispõe de arquiteturas neurais que apresentam bons desempenhos, tanto para os métodos de reconhecimento supervisionado, como para o reconhecimento não-supervisionado. Dentre as arquiteturas neurais que implementam algoritmos de classificação supervisionada se destaca a chamada rede de Perceptrons multicamada, também conhecida como rede Backpropagation, devido ao nome do seu algoritmo de treinamento. Estas redes apresentam resultados de classificação muito bons, mas exigem a especificação de um arquivo de treinamento, com entradas e saídas conhecidas, que cubra uniformemente o espaço de característica. Devido à complexidade das imagens multiespectrais, uma boa estratégia para a análise destas imagens se constitui na combinação de algoritmos não-supervisionados com métodos supervisionados. Neste sentido, as redes neurais permitem a realização dos chamados mapas de característica auto-organizados (SOM), que implementam a clusterização dos vetores de característica, segundo métodos de auto-organização topológica. Uma rede deste tipo realiza a projeção do espaço multidimensional dos vetores de característica, sobre um espaço bidimensional de neurônios. A partir desta clusterização, pode-se empregar técnicas de classificação supervisionada do tipo quantização vetorial adaptativa (LVQ), atingindo assim resultados de classificação bastante bons. Além das redes neurais, outras técnicas de inteligência artificial podem ser empregadas na análise de imagens multiespectrais. Em especial, a lógica fuzzy tem encontrado aplicação na tomada de decisão dos classificadores. Aqui, a principal aplicação está relacionada com a utilização de funções de pertinência de conjuntos fuzzy, para a classificação dos chamados pixel mistura, onde mais de uma classe ocorre simultaneamente dentro de um

mesmo pixel. Com a utilização desta metodologia consegue-se aumentar sensivelmente o índice de classificação correta do sistema.

INTERPRETAÇÃO DE IMAGENS DE SATÉLITE DA AMAZÔNIA USANDO REDES NEURAIIS

Machado, R.J.

Centro Científico Rio, IBM Brasil Email: machado@riosc.bitnet

O Centro Científico Rio da IBM Brasil participa conjuntamente com o INPE de um projeto que prevê o desenvolvimento de um sistema de extração automática de informação de imagens da Amazônia brasileira, obtidas por sensoriamento remoto, com o objetivo de identificar os diversos temas de interesse para avaliação do processo de desflorestamento. Atualmente este tipo de atividade é realizado manualmente por especialistas humanos, cuja produtividade não é suficiente para dar conta do grande volume de dados novos continuamente produzidos, bem como para analisar imagens muito complexas, tais como as imagens da colonização do estado de Rondônia. A abordagem usada para interpretação das imagens processa-se em duas fases. A primeira fase, ao contrário da abordagem usual pixel-a-pixel, consiste na segmentação da imagem em regiões espectralmente homogêneas (chamadas de segmentos) pela aplicação de uma técnica de crescimento de regiões. Na segunda fase, cada segmento é então classificado em uma ou mais das categorias temáticas Floresta, Cerrado, Água, Área Desflorestada, Nuvem, Sombra, e Pluma. As quatro primeiras categorias englobam toda informação relevante a ser monitorada sendo chamadas de categorias básicas. As três últimas categorias refletem as interferências ocasionadas ao processo de classificação pela presença de nuvens, sombras e fumaça nas imagens. A classificação dos segmentos nessas categorias segue uma abordagem de lógica nebulosa, ou seja, um segmento pode pertencer a múltiplas categorias com graus parciais de pertinência. Dessa forma é possível modelar fenômenos de transição, por exemplo, uma área desmatada que se regenera através de um processo de rebrota. A arquitetura do classificador é baseada numa rede neural treinada usando o modelo de Retropropagação de Erros (backpropagation). A arquitetura implementa também um procedimento de relaxação no topo da rede neural treinada com backpropagation, com o intuito de se beneficiar das relações de vizinhança entre segmentos adjacentes. O resultado obtido durante testes dessa arquitetura, usando um conjunto de cenas típicas e de cenas consideradas difíceis, revela um grande potencial para a automação da tarefa de automatização da tarefa de monitoramento ambiental da Amazônia.

ANÁLISE DIGITAL DE GLOMÉRULOS RENAIIS

Andrade, M.C.^{1,2}; Araujo, A.A.¹; Santos, A.M.M.²; Lameiras, F.S.²; Bambirra, E.A.³

¹ Departamento de Ciência da Computação, Universidade Federal de Minas Gerais Belo Horizonte; ² Centro de Desenvolvimento da Tecnologia Nuclear, Comissão Nacional de Energia Nuclear, Belo Horizonte, ³ Departamento de Anatomia Patológica, Universidade Federal de Minas Gerais, Belo Horizonte Email: andrade@dcc.ufmg.br, arnaldo@dcc.ufmg.br, cbtn@dcc.ufmg.br

A análise digital da estrutura dos glomérulos renais, através de imagens de seções planas de fragmentos cirúrgicos ou biopsias do córtex renal, fornece ao nefrologista um padrão sistemático de avaliação da morfologia destas microestruturas. Nos casos de transplantes renais, o fármaco imunossupressor utilizado para evitar a rejeição do órgão transplantado, por exemplo, a Ciclosporina (CsA), induz alterações morfológicas significativas na estrutura glomerular [Kumar-89], que podem ser quantificadas através da análise digital de imagens. A estrutura microscópica dos rins revela a presença de elementos denominados néfrons que são essencialmente idênticos em estrutura e função. Cada néfron consiste de um glomérulo e um túbulo. A parte interna do glomérulo é composta de um tufo de capilares, de forma quase esférica, envolto por uma cápsula de parede dupla, denominada cápsula de Bowman. Esta parede constitui a membrana filtrante do glomérulo. O espaço entre o tufo de capilares e a parede interior da cápsula está em contato direto com a luz do túbulo. Neste trabalho, a experiência inicial dos autores com a análise digital de glomérulos renais é descrita e ilustrada. Duas situações foram estudadas: a evolução das estruturas glomerulares em biopsias percutâneas, feitas em rim transplantado em paciente tratado com CsA [Araujo-93], para três instantes de tempo após o transplante e a determinação da distribuição volumétrica destas estruturas, em um fragmento cirúrgico do córtex renal [Andrade-93], empregando-se o método estereométrico de Saltykov [Saltykov-74], [Elias-71]. Os fragmentos foram retirados da superfície do córtex renal, fixados em parafina e seccionados através de micrótomo produzindo seções de poucos micrometros de espessura. As seções foram coletadas em laminais histológicas de vidro, fotografadas através de microscópio ótico e, então, digitalizadas através de um scanner de mesa com resolução mantida em 200 dpi e 256 níveis de cinza. Nas imagens aqui utilizadas, frequentemente, se fez necessário retificar, de forma manual, as bordas mal definidas das cápsulas de Bowman. Após a etapa de edição as imagens foram segmentadas e a partir delas determinaram-se as áreas e os diâmetros equivalentes de cada uma das seguintes regiões de interesse: a área total da seção de corte do glomérulo, a área da luz glomerular entre a cápsula de Bowman e o tufo de capilares e, também, a área do tufo de capilares. Cada região de interesse foi previamente preenchida, com colorações convenientemente escolhidas, de forma a permitir o processamento automático das imagens. Os seguintes resultados foram obtidos: para o caso das biopsias observou-se que o efeito da CsA sobre os glomérulos significou um aumento global do volume glomerular durante o tratamento. Determinou-se, também, que a fração volumétrica glomerular representou 7% do volume do córtex renal para os três instantes de tempo analisados. Para os fragmentos cirúrgicos mediu-se a distribuição de tamanhos de cápsulas e tufos dos glomérulos do córtex renal com auxílio de método estereométrico de Saltykov. Observou-se que os volumes mais frequentes aparecem na classe de maior diâmetro ($\sim 212 \mu\text{m}$) e que as freqüências de ocorrência nas demais classes são significativamente menores. Deduz-se que os glomérulos possuem volumes concentrados numa faixa de valores relativamente pequena. Tanto as cápsulas quanto os tufos analisados apresentaram alta esfericidade e volumes numa pequena faixa de valores (relação diâmetro máximo / diâmetro mínimo de 2,1 para os tufos e 2,7 para as cápsulas). Neste trabalho, foi utilizado um analisador de imagens [Andrade-93] desenvolvido pelo Núcleo de Processamento Digital de Imagens - NPDI (DCC-UFGM), originalmente concebido para análise de materiais cerâmicos [Chermant-86]. A versão protótipo do analisador foi implementada sobre a interface PIXELWARE [Davis-92] em microcomputador 486, dotado de placa gráfica SuperVGA. Parte dos algoritmos de

imageamento microscópico, aqui utilizados, baseou-se no conjunto de subrotinas SPIDER [Tamura-82]. M. C. Andrade, Imageamento microscópico. Belo Horizonte: UFMG, 1993. Dissertação (mestrado em computação), UFMG, 1993. A. de A. Araujo, et al. Digital processing of histopathological aspects in renal transplantation. Proceedings of the IS&T/SPIE. Conf. 1905, Biomedical Image Processing and Biomedical Visualization, 1993. J. L. Chermant, Characterization of the ceramics by image analysis. Ceramics International. v. 12, p. 67-80, 1986. C. Davis Jr., PIXELWARE: Um sistema de processamento digital de imagens. Belo Horizonte: UFMG, 1992. Dissertação (mestrado em computação), UFMG, 1992. S.A. Saltykov, Stereometrische Metallographie, Leipzig, VEB, 1974. H. Tamura et al., Design and implementation of SPIDER - a transportable image processing software package. Computer Vision, Graphics and Image Processing, v. 23, p. 273-294, 1982. H. Elias et al. Stereology: Application to biomedical research. Physiological Reviews, v. 51, n. 1, p.158-200, 1971. M. S. A. Kumar et al., Chronic cyclosporine nephrotoxicity in renal transplantation: Is it effect or preservation?, Transplantation Proceedings, v. 21, p. 1552-1553, 1989. Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (400190/90-7, 500908/91-5) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG (TEC 1113-90) pelo apoio financeiro a este trabalho.

PROCESSAMENTO DIGITAL DE VOCALIZAÇÕES

Viellard, J.M.E.

Departamento de Zoologia, Instituto de Biologia UNICAMP e CNPq)

Vocalizações são os sons produzidos pelos órgãos vocais de certos animais (principalmente as siringes das aves e laringe dos mamíferos e anfíbios) através da compressão periódica de um fluxo de ar. São usadas como sinais portadores de informação para a comunicação. Essas variações de pressão do meio de transmissão (ar ou água) são sinusoidais, simples (som puro) ou complexas. Apesar das sinusoidais serem curvas contínuas, elas podem ser facilmente digitalizadas por serem, no caso das vocalizações, periódicas, ou seja, repetitivas (o que não se verifica nos sons instrumentais estalidos). A teoria indica que uma sinusóide pode ser reconstituída com somente 3,4 valores por ciclo; daí a taxa de amostragem precisa ser somente 3,4 vezes a frequência mais alta do som a ser digitalizado. No caso dos sons complexos ou sobrepostos, a resultante é facilmente decomposta em sinusóides simples por tratamento matemático (Transformação de Fourier): hoje o computador realiza esta operação quase instantaneamente (real time) com o algoritmo FFT (Fast Fourier Transform). O som digitalizado pode ser arquivado sob a forma binária (disco CD, fita DAT, computador) para ser reproduzido posteriormente. Ele também pode ser tratado através de programas de edição. Finalmente, pode ser representado por seus parâmetros físicos (frequência, amplitude e duração calculado diretamente a partir dos valores das componentes sinusoidais digitalizados. Isto facilita muito a análise fonográfica (medição e representação gráfica dos parâmetros sonoros) das vocalizações animais.

FRAGMENTAÇÃO AUTOMÁTICA DO SONO UTILIZANDO REDES NEURONAIS

Coimbra, A.J.F.

Grupo de Engenharia Biomédica, Centro de Tecnologia da Universidade Federal de Santa Catarina, Florianópolis, SC Email: alex@gpeb.ufsc.br

Está sendo implementado um sistema de análise automatizada de EEG para fragmentação do sono utilizando alguns algoritmos de extração de características do sinal e redes neuronais para classificação das fases do sono. Análises visuais de registros prolongados de EEG são trabalhosas e consomem muito tempo do especialista. Registros prolongados são necessários, por exemplo, na monitoração contínua de pacientes epiléticos em vigília e/ou durante o sono, em polissonografia noturna para caracterizar distúrbios do sono e/ou para realizar fragmentação do sono. O estudo da fragmentação do sono pode ser interessante para avaliar o efeito de drogas ou stress ambiental sobre as fases do sono. Variações nas características das fases podem ser utilizadas como marcadores biológicos em distúrbios psiquiátricos tais como a depressão endógena. Segmentadores automáticos de sono vêm sendo desenvolvidos já há alguns anos e, com o advento dos microcomputadores (e com o aprimoramento de suas características de velocidade e poder de processamento), vem se tornando mais e mais acessíveis. Atualmente estão sendo pesquisados os algoritmos de extração de característica do sinal. Estes algoritmos extraem do sinal informações como as contribuições energéticas de cada faixa de frequência de interesse (teta, delta, alfa, beta), os coeficientes de modelos auto-regressivos (AR) do sinal, e as características de amplitude-frequência das ondas (evento contido entre dois extremos locais) componentes do sinal. Estas informações são fornecidas as entradas das redes neuronais que realizam a classificação das fases. Pretende-se utilizar o sistema para auxiliar em pesquisas de efeitos de drogas em pombos (COLUMBA LIVIA) realizadas no Depto. de Ciências Fisiológicas - CCB da Universidade Federal de Santa Catarina. Futuramente pretende-se implementar um sistema similar para humanos, que encontrara aplicação no Grupo de Eletroencefalografia do Hospital Universitário da UFSC.

TÉCNICAS PARA ANÁLISE DE RITMOS BIOLÓGICOS

Benedito-Silva, A.A.; Menna-Barreto, L.

Grupo Multidisciplinar de Desenvolvimento e Ritmos Biológicos. Depto. Fisiologia e Biofísica, ICB, Universidade de São Paulo, Brasil. Email: mennaicb@brusp.bitnet

Serão apresentados alguns sistemas computacionais que implementam técnicas de análise de ritmos biológicos. Estes sistemas foram confeccionados em nosso laboratório de acordo com as necessidades atuais das nossas pesquisas. Foram desenvolvidos em Fortran-77 o programa COSANA, baseado no procedimento do cosinor individual, e o programa denominado COSPOP que estima os parâmetros rítmicos de uma população. Foi também desenvolvido um programa em FORTRAN-77 para investigar os componentes espectrais de uma serie temporal de dados de ciclo vigília/sono, de acordo com um algoritmo que correlaciona uma serie de dados com ondas quadradas, no caso as funções de Walsh. Foram também desenvolvidos programas em dBase 3 Plus para detectar a periodicidade dominante numa serie temporal, baseado no teste de Fisher (1929), bem como periodicidades compostas, baseado no teste de Siegel (1980). Um

programa em Turbo- Pascal foi também confeccionado para analisar dados através das técnicas de estatística circular.

TESTE PARA DETECTAR PERIODICIDADES COMPOSTAS EM SÉRIES TEMPORAIS

Benedito-Silva, A.A.; Menna-Barreto, L.

Grupo Multidisciplinar de Desenvolvimento e Ritmos Biológicos. Depto. Fisiologia e Biofísica, ICB, Universidade de São Paulo, Brasil. Email: mennaicb@brusp.bitnet

Testes para detectar periodicidades compostas em series temporais biológicas são úteis em pesquisas relacionadas com intermodulação de frequências em ritmos biológicos. Uma aplicação possível destes testes é a investigação de componentes espectrais vigília/sono nos primeiros meses de vida. O objetivo deste trabalho é apresentar a adaptação a Cronobiologia de um teste para detecção de periodicidades compostas (J.Amer.Stat.Assoc. 75: 345, 1980). Este teste consiste numa modificação do teste de Fischer (Proc. Royal Soc. Ser.A, 125: 54, 1929) uma vez que ele se baseia nas ordenadas do periodograma, e não apenas na maior delas. Os espectros de potencia Fourier do ciclo vigília/sono de cinco crianças, com idades variando de 3 a 14 meses foram coletados e submetidos a este teste. A partir dos resultados obtidos, foi possível verificar que o conjunto de oscilações rítmicas simultaneamente significativas varia de acordo com, contribuindo para uma investigação ontogenética do ciclo vigília/sono.

Aplicações em Biologia Molecular e Engenharia Genética

ENGENHARIA GENÉTICA E INFORMÁTICA

Leite, A.

Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas, SP Email: cbmeg@ccvax.unicamp.br

Devido à característica modular dos genes e proteínas, suas seqüência e propriedades armazenadas nos grandes bancos de dados constituem importante fonte de informações para projetos nas áreas de Engenharia Genética e Engenharia de Proteínas. O desenvolvimento de técnicas mais rápidas, simples e confiáveis de seqüenciamento de DNA determinou um grande aumento no número de seqüência publicada e armazenada nestes bancos nos últimos 20 anos. Softwares específicos permitem o acesso às seqüências sejam através da determinação de similaridades entre seqüência, seja a partir de palavras-chave do documentário que acompanha cada uma das seqüências publicadas. Bancos de dados podem ser acessados diretamente (on line) via centrais de coleta e distribuição de dados, em estações de trabalhos individuais ou ainda, através de servidoras locais. A manipulação dos dados dispõe de um grande número de softwares voltados para a análise de seqüência. Tais programas permitem a edição, alinhamento e tradução de seqüência, construção de mapas de restrição, predição de estruturas de RNAs e de conformação estrutural de proteínas e a localização de regiões codificadoras, tRNAs e sinais regulatórios. A alta repetitividade de tarefas realizadas num

laboratório de Biologia Molecular, assim como a demanda pela reprodutibilidade dos experimentos, tem levado a automação de processos laboratoriais. Desde a pipetagem de reagentes, passando pela realização dos experimentos (como é o caso de colunas cromatográficas, géis de seqüenciamento e aparelhos para amplificação de ácidos nucléicos), até a leitura e interpretação dos resultados, podem hoje ser feitas por sistemas completamente automatizados. A velocidade exponencial na qual os bancos de dados têm crescido e a automação das tarefas laboratoriais devera ser ainda incrementada com os grandes projetos internacionais visando o seqüenciamento de genomas completos. Tal velocidade de crescimento constitui um grande desafio para a Ciência da Computação, a qual será responsável pelo fornecimento de ferramentas destinadas a obtenção e análise do incrível número de seqüência com certeza disponíveis nas próximas décadas.

TRÊS MÉTODOS PARA PREVISÃO DE ESTRUTURA SECUNDÁRIA DE PROTEÍNAS, VERSÕES PARA MICROS

Panepucci, E.H.¹, Barbosa, J.A.R.G.¹, Ventura, M.M.²

¹ Instituto de Física e Química de São Carlos, USP, São Carlos, S.P.; ² Depto. de Biologia Celular, Laboratório de Biofísica, UnB, Brasília.

Já faz algum tempo desde que Chou e Fasman publicaram seu método para previsão de estrutura secundaria de proteínas. Existem hoje em dia diversos métodos com esse propósito, podemos agrupá-los de acordo com a metodologia utilizada para previsão (redes neurais, teoria da informação, aspectos estereoquímico, simples estatística). A maioria desses métodos se utiliza de tabelas de valores para cada um dos 20 resíduos de aminoácidos. Os algoritmos para obtenção da predição a partir dessas tabelas estão publicados em artigos e são transformados em programas, facilitando o seu uso. Apesar de existirem versões desses programas para computadores de grande porte, os mesmos podem ser aplicados a microcomputadores com uma eficiência muito boa. Nosso objetivo é fornecer, a pessoas que não tenham acesso a computadores de grande porte, uma saída para obter informações estruturais com suas seqüências. Esse trabalho apresenta três programas de predição em duas versões: TurboPascal e C, que podem ser aplicados a qualquer seqüência protéica. Os algoritmos utilizados foram extraídos de Holley e Karplus [1], McGregor, Flores e Sternberg [2] e Garneir, Osguthorpe e Robson [3]. Os dois primeiros métodos são fruto da metodologia de redes neurais, enquanto que o ultimo é uma aplicação da teoria da informação. É um fato inquestionável que a predição de estruturas ainda tem um erro muito grande. Para minimizar o mesmo são utilizados vários programas de predição, sendo seus resultados compilados de maneira tal que regiões onde a maioria dos métodos prevê uma determinada estrutura, esta é aceita. Hoje existem formas mais confiáveis de se prever estruturas, utilizando-se do fato das proteínas poderem ser agrupadas em famílias, e que dentro dessas famílias a estrutura é conservada, pode-se prever, através de alinhamentos, o folding de uma proteína. Em um projeto futuro, pretendemos fornecer um pequeno pacote com capacidade de satisfazer as pessoas que não tenham acesso aos computadores de grande porte, afim de que elas possam trabalhar com sua seqüência. [1] Holley, L.H. and Karplus, M. Proc. Natl. Acad. Sci., 86:152-156, 1989. [2] McGregor, M.J., Flores, T.P. and Sternberg, M.J.E. Protein Engineering, 2(7):521-526,

1989. [3] Garneir, J., Osguthorpe, D.J. and Robson, B. J. Mol. Biol., 120:97-120, 1978. Agradecemos pelo apoio financeiro fornecido pelo CNPq.

PREVISÃO DE ESTRUTURA SECUNDÁRIA DE PROTEÍNAS USANDO MODELOS LOGÍSTICOS

Munson, P.J.; di Francesco, V.; Porrelli, R.N.

Laboratory of Structural Biology, Division of Computer Research and Technology, National Institutes of Health, USA e Nucleo de Informática Biomédica, UNICAMP Email: porrelli@ccsun.unicamp.br

Utilizamos técnicas estatísticas para tentar melhorar a capacidade de prever estrutura secundária de proteínas baseada na seqüência de aminoácidos. Funções discriminantes logísticas foram estendidas para acomodar termos quadráticos e assim poder competir com algoritmos baseados em redes neurais artificiais e aprendizado de máquina. Nosso modelo limita o número de parâmetros quadráticos explorando a natureza periódica das alfa- hélices e placas pregueadas. Também utilizamos técnicas semi-paramétricas para melhor ajustar o modelo aos dados. Com esta abordagem fomos capazes de elevar a acurácia do modelo GOR original, obtida com validação cruzada, para 62.5% (modelo otimizado de máximo-verossimilhança penalizada). Quando selecionamos os pares de aminoácidos a serem considerados nos parâmetros quadráticos, a acurácia com validação cruzada subiu para 65.9%. A seleção do melhor modelo foi baseada no cálculo do número efetivo de parâmetros de vários modelos diferentes. O resultado ótimo correspondeu a um modelo com cerca de 800 parâmetros efetivos. Uma das vantagens da utilização de técnicas estatísticas é a possibilidade de interpretação dos parâmetros do modelo. Ainda que não tenha sido incluída qualquer informação sobre a natureza química dos diferentes aminoácidos, os valores obtidos para os parâmetros lineares refletem algumas propriedades conhecidas dos mesmos. Os valores dos parâmetros quadráticos, correspondentes a interação entre pares de aminoácidos, mostram grande concordância com potenciais de contato calculados por outros autores.

SISTEMAS DE AUXÍLIO AO SEQÜENCIAMENTO DE DNA EM LARGA ESCALA

Meidanis, J.

Departamento de Ciência da Computação, Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, Campinas, SP Email: meidanis@dcc.unicamp.br

Com o advento de técnicas laboratoriais rápidas para seqüenciamento de DNA os cientistas ganharam acesso ao nível mais elementar de informação genética, a saber, a seqüência de bases de uma molécula de DNA. A princípio, apenas moléculas relativamente curtas puderam ser assim desvendadas, na faixa de até 5000 bases. Hoje em dia projetos bem mais ambiciosos são objeto de atenção, culminando com o famoso Projeto do Genoma Humano, cuja meta final é obter as seqüências de todos os cromossomos do homem. Apesar de todos os avanços tecnológicos, uma limitação ainda permanece: só é possível ler diretamente de um filme de autoradiografia seqüência de no máximo 300 a 500 bases. Para seqüenciar moléculas maiores deve-se quebrá-las em fragmentos, seqüenciar os fragmentos e a seguir montá-los para obter

a molécula original. Nesta montagem o uso de computadores é indispensável, devido ao grande volume de dados. Nesta palestra faremos uma revisão dos diversos métodos e programas que tem sido desenvolvidos para este fim, mostrando vantagens e desvantagens em relação a vários aspectos. Descreveremos ainda um sistema novo que esta sendo desenvolvido no Departamento de Ciência da Computação da UNICAMP. Este sistema esta voltado especialmente para projetos de seqüenciamento em larga escala (envolvendo milhares de fragmentos), que possuem toda uma gama de problemas e desafios próprios, não encontrados em projetos menores.

Bases de Dados e Telemática

THE BRAZILIAN MOLECULAR BIOLOGY AND BIOTECHNOLOGY NETWORK (BMBBNet)

Neshich, G.

EMBRAPA, Brasília, DF Email: bbrc@cenargen.embrapa.br

BMBBnet is an initiative to develop the Brazilian infrastructure for academic and commercial information services in molecular biology and biotechnology. With the growing importance of biotechnology for agriculture and of genome projects for biotechnology and medicine, the long-range impact of a Bioinformatics Resource in Brazil can be considerable, providing essential information services to the entire national research community in biotechnology in a very cost-effective way. The network consists of a national central node: Brazilian Bioinformatics Resource Center (BBRC) and already existing network (Internet) of the RNP. The latter functions as the backbone network connecting various state capitals. It is supported and maintained by the Genetic Resource and Biotechnology Center (CENARGEN) of the Brazilian Agriculture Research Corporation (EMBRAPA), an enterprise affiliated to the federal Ministry of Agriculture. The Network activities are: Data Distribution: In collaboration with the EMBL-Heidelberg the BBRC will receive the daily updates of the EMBL data bases. The BBRC should be therefore supplied with the most comprehensive collection of sequence data and it is furthermore licensed by the EMBL to distribute its data-bases to local state nodes within Brazil and also, to all interested Latin American countries. Similar mechanisms could be useful for distribution of other databases. A computational assistance to molecular biologists in Brazil will be also provided. Computer Conferencing System: The development of a bulletin-board and conferencing system, as well as a biotechnology discussion list (bbrc-l) on the network will provide the vehicle for information exchange between all network partners. To be effective, this will involve relevant research topics and will be easily accessible to laboratory scientists on the computers they commonly use for their work. Access to Remote Facilities: Systems for interactive access to remote facilities should be developed, for example: access to specialized hardware and software, database hosts, software collections or other resources. Training: Where necessary, training for variety of potential users should be arranged. Possible topics include the installation and support of particular software packages and technical networking issues. Full participation of the EMBL technical staff is guaranteed. All information on BMBBNet and the BBRC is available through the on-line information server gopher. Interested

laboratories could contact the BBRC using the E-mail address, or write to: Dr. Goran Neshich, MBNet coordinator, EMBRAPA/CENARGEN, P.O.Box 02.372, 70879-970, Brasília - DF, Brazil, phone: +55 61 272-3212, fax: +55 61 274-3212

ICGEBnet: AN INTERNATIONAL BIOLOGICAL COMPUTER RESOURCE

Bevilacqua, V.; Pongor, S.

ICGEB - International Centre For Genetic Engineering and Biotechnology, Area Science Park, 34012 Trieste, Italy Email: valeria@icgeb.trieste.it

The International Centre for Genetic Engineering and Biotechnology (ICGEB) is an international organization presently operating as a special programme of the United Nations Industrial Development Organization (UNIDO). ICGEB's aim is to help its member countries (Afghanistan, Algeria, Argentina, Bhutan, Bolivia, Brazil, Bulgaria, Chile, China, Colombia, Congo, Costa Rica, Croatia, Cuba, Ecuador, Egypt, Greece, Hungary, India, Indonesia, Iran, Iraq, Italy, Kuwait, Mauritania, Mauritius, Mexico, Morocco, Nigeria, Pakistan, Panama, Peru, Poland, Russia, Senegal, Sri Lanka, Sudan, Syria, Thailand, Trinidad and Tobago, Tunisia, Turkey, Venezuela, Viet Nam, Yugoslavia, Zaire) in molecular biology and biotechnology research. ICGEBnet is a biocomputing resource of ICGEB that currently provides login facilities for 550 users in Latin America, Central and Eastern Europe, Asia and Africa via X.25 and Internet connections. Biological sequence databases of genes and proteins play a fundamental role in molecular biology. Maintenance of such databases and the necessary softwares packages require a complex know-how and is best done via centralized computer resources such as the EMBnet Informatics Network of the European Molecular Biology Organization. In 1990 ICGEB started the ICGEB computer resource project, (ICGEBnet), in order to provide similar services for developing countries. The primary purpose of the ICGEBnet is to disseminate the best of currently available computational technology to the molecular biologists of the ICGEB research community. ICGEBnet provides on-line access to up-to-date databases, softwares for sequence analysis and communication tools to promote a rapid sharing of information among the ICGEB member countries scientists. Access to the ICGEBnet resource is available free of charge to all ICGEB member country scientists; however, preference is given to those scientists whose research is directly related to the research goals of ICGEB. The ICGEBnet can be accessed through the Internet or via the X25 Public Data Network. Hardware comprises one SUN4 server/390 and one SPARC server10 with 8 Gbytes disk space, accessible through X.25 and Internet lines. Molecular modeling facility is provided by a Silicon Graphics IRIS Indigo XS24. Software components are: 1) Biological databases: EMBL, GenBank, TFD, HIV-NA, PIR, Swiss-Prot, SEQDB, PROSITE, SBASE, HIV-AA, LIMB, OMIM, Brookhaven PDB, Restriction Enzymes, Biocomputing Bibliography, PBASE. Sequence analysis software available are GCG Wisconsin Sequence Analysis package, Clustal V, Phylip, Plsearch, blast, Signal Scan and IRX. Gopher based information containing many databases, help for system facilities and links to other information servers world-wide. USENET news including bionet, embnet, news and gnu groups. Menu systems have been developed for ICGEBnet facilities, GCG programs, Phylip and Waissearch; a file manager for novice user on UNIX was also implemented in C using curses library. Automated database update protocols and the SBASE protein domain database were also developed in-house. The programs and databases developed in-house are available

through anonymous ftp from (ftp.icgeb.trieste.it). The ICGBnet resource became an EMBnet node in 1992 and now provides services that are beyond the European average, using relatively inexpensive hardware components. ICGBnet is ready to help the member countries in establishing similar biocomputing resources. Pongor, S., Simon, G., Falaschi, A. (1992) The UNIDO computer resource for molecular biology. The Internet Society News, 2, 23-24 Pongor, S., Skerl, V., Cserzo, M. and Hatsagi, Z., Simon, G. and Bevilacqua, V. (1993): The SBASE domain library: A collection of annotated protein sequence segments. Protein Engineering, 6(4), 391-395 Simon, G. and Pongor, S. (1992): ICGBnet: The UNIDO computer resource for molecular biology. Bioinformatics, 1, 12

BASES DE DADOS EM BIOLOGIA

Canhos, V.P.

Fundação André Tosello, Campinas, SP Email: vcanhos@bdt.ftpt.br

Em função dos avanços em instrumentação e informática, os especialistas em biologia contam agora com inúmeras ferramentas e métodos avançados para a sistemática, estudos fisiológicos e aplicações biotecnológicas. Um dos problemas fundamentais é de como se manter atualizado com a oferta dinâmica de novas tecnologias disponíveis para o estudo das relações entre organismos e a sua aplicação no desenvolvimento de novas tecnologias em agricultura, indústria, medicina e meio ambiente. O estabelecimento de bases de dados biológicos compreensíveis e abrangentes, assim como a disseminação via redes eletrônicas é cada vez mais importante para desenvolvimento de programas de ecologia, biodiversidade e desenvolvimento socioeconômico sustentável. Como resultado do investimento de países industrializados na implantação de infraestruturas de redes de transmissão eletrônica de dados a altas velocidades, estamos vivendo uma verdadeira revolução que esta levando a uma era de pesquisa cooperativa nunca vista. Estes desenvolvimentos podem ser medidos pela rápida evolução da conectividade da rede Internet (no Brasil, a RNP - Rede Nacional de Pesquisa) que esta mudando a forma pela qual nos percebemos a interligação de centros provedores de dados e usuários. Muito desta evolução esta ocorrendo em software que e de domínio publico, e é acessível a qualquer individuo que tenha uma conexão Internet. A democratização do acesso a Internet torna possível o compartilhamento de informações sem a imposição de estruturas rígidas. Bases de dados de domínio publico estão proliferando em todas as áreas da biologia. Através da Internet podemos fazer buscas em varias centenas de bases de dados disponíveis em servidores distribuídos pelo mundo afora. De fato e possível, em certo grau, qualquer usuário se tornar um provedor de dados. Como exemplo podemos citar as bases de dados de domínio publico do Genbank e EMBL. De forma coordenada e através de métodos constantemente aprimorados de acessos a bases de dados de biologia molecular, milhares de pesquisadores espalhados pelo mundo estão usando e contribuindo para compilações de dados sobre proteínas e DNA. Os desenvolvimentos em biologia molecular fornecem um bom modelo para o desenho e implementação de outras bases de dados e redes temáticas. A consolidação da Internet permite o fortalecimento de bases de dados de domínio publico, o que possibilita a maximização do uso das informações disponíveis, alem de promover a coordenação de atividades de pesquisa e fomento. No Brasil, a Base de Dados Tropical (BDT) vem atuando desde 1985 na área de bancos de dados biológicos e comunicação eletrônica.

Através de sua atuação perseverante com recursos reduzidos, a BDT tem se mostrado capaz de incorporar os avanços significativos que vem ocorrendo no panorama internacional. Desde o início de suas atividades, utilizando-se do Serviço Cirandão da EMBRATEL, até o momento atual, a BDT progrediu através da adoção de soluções criativas e adequadas para resolver problemas de infraestrutura e recursos limitados. A BDT hoje, oferece cerca de 40 bancos de dados via linha X.25 e/ou Internet, mantém uma lista de discussão internacional (biodiv-l, Biodiversity Information Network List) e oferece ainda um Gopher server, com os bancos de dados nacionais (13), a lista de discussão, relação de cursos, a publicação eletrônica da SBPC (SBPC Hoje) e a Publicação Eletrônica C&T Radiobras. Oferece também o Usenet News Groups relacionados com ciências biológicas.

BASE DE DADOS TROPICAL: DISSEMINAÇÃO DE INFORMAÇÕES PARA BIODIVERSIDADE E BIOTECNOLOGIA

Souza, S.; Brefe, C.A.F.; Canhos, D.A.L.; Oliveira, P.; Canhos, V.P.

Fundação André Tosello, Campinas, SP Email: vcanhos@bdt.ftpt.br

A Base de Dados Tropical, BDT, e o centro de informações da Fundação Tropical de Pesquisas e Tecnologia André Tosello, uma entidade de direito privado, sem fins lucrativos, órgão de utilidade pública federal, estabelecida em marco de 1971. A BDT tem como meta a disseminação de informação eletrônica, como ferramenta na organização da comunidade científica e tecnológica do país. Atua especificamente na área de informação biológica, de interesse industrial e ambiental, e pretende, através de sua atuação, contribuir diretamente para a conservação e utilização racional da biodiversidade do país. Um pacote único de bancos de dados especializados e distribuído através da BDT-Net, incluindo informações de interesse para microbiologia, biotecnologia, genética e biodiversidade. Diferentes fontes de dados incluem informações sobre isolados microbianos caracterizados em coleções de culturas do mundo inteiro, informações taxonômicas de bactérias e vírus, hibridomas, linhagens celulares, sondas moleculares, e outros. Inclui também bancos de dados sobre pesquisadores atuantes nos diferentes campos da biodiversidade e biotecnologia. Todos os bancos de dados do MSDN (Microbial Strain Data Network) e do IRRO (Information Resource on Release of Organisms into the Environment) também estão disponíveis. São 38 bancos de dados nacionais e internacionais gerenciados pela BDT. Catálogos de Coleções de Culturas: - Catalogo Nacional de Linhagens: Bactérias - Catalogo Nacional de Linhagens: Fungos - Catalogo Nacional de Linhagens: Algas - Catalogo Nacional de Linhagens: Protozoários - Catalogo Nacional de Linhagens: Vírus - Catalogo Nacional de Linhagens: Linhagens Celulares - ATCC Animal Viruses and Antisera Catalogue - ATCC Bacteria and Bacteriophages Catalogue - ATCC Filamentous Fungi and Yeasts Catalogue - ATCC Plant Viruses and Antisera Catalogue - ATCC Protozoa and Algae Catalogue - ATCC Clones, Vectors, Libraries, and Hosts Catalogue - ATCC Cell Lines Catalogue - DSM Deutsche Sammlung von Mikroorganismen und Zellkulturen - IMI Catalogue of Fungi - Peterhoff Genetic Collection of Saccharomyces Cerevisiae Yeasts - CCALA Catalogue of Algae and Cyanobacteria - CCF Catalogue of Filamentous Fungi - CCM Catalogue of Bacteria Quem é Quem: - Quem é Quem em Biodiversidade - Quem é Quem em Botânica - Herbários Brasileiros - Censo dos Zoológicos Brasileiros - Cadastro das Coleções de Culturas Brasileira - MSDN Directory Nomenclatura e Informações Taxonômicas: - Nomenclatura Bacteriana - IJSB,

Intern. J. Systematic Bacteriology - AVIS Animal Vírus Information System Biotecnologia: - Agentes Antimicrobianos - BIA BioIndustry Association Databases - BEMET Biotechnology Courses Database - BKS Biotech Knowledge Sources Databases - BIOCAT Database on Insect Control by Release of Natural Enemies Meio Ambiente: - a lista de discussão BIODIV-L - Unido Code of Conduct for Environmental Release of GMOs - OECD BIOTRACK Database on Environmental Release of GMOs Estão também disponíveis aos usuários da BDT-Net a revista eletrônica SBPChoje, da Sociedade Brasileira para o Progresso da Ciência, além de uma relação de cursos de treinamento, simpósios e congressos de interesse para Biotecnologia e/ou Biodiversidade. A BDT-Net faz parte da Internet, uma rede mundial de mais de 2 milhões de computadores interconectados. Existe uma variedade incrível de bases de dados de acesso público e listas de discussão disponível a qualquer usuário BDT interessado. Assim, todo usuário BDT tem a sua disposição serviços de correio eletrônico, buscas em bases de dados nacionais e internacionais, além da possibilidade de participar em listas de discussão dos mais variados assuntos. A BDT não visa lucros. Tem por objetivo estabelecer elos de comunicação eletrônica para promover um maior intercâmbio científico e tecnológico no país e no exterior. Apenas para auxiliar na manutenção de suas atividades, é cobrada a seguinte taxa: Assinatura por um período de 6 (seis) meses: assinante individual: o equivalente a 220 UFIRs; assinante institucional: o equivalente a 2200 UFIRs (com direito de até 20 contas). Para maiores informações, envie uma mensagem para manager@bdt.ftpt.br; ou escreva para: Base de Dados Tropical, Fundação Tropical de Pesquisas e Tecnologia André Tosello, Rua Latino Coelho 1301, Parque Taquaral, 13.087-010 Campinas, São Paulo, Brasil, Telefone: (0192) 42-7022, Fax: (0192) 42-7827

BIN21 - THE BIODIVERSITY INFORMATION NETWORK

*Canhos, D.A.L.: Brefe, C.A.F.: Souza, S.: Oliveira, P. e Canhos, V.P.
Fundação André Tosello, Campinas, SP Email: dora@bdt.ftpt.br*

Em julho de 1992, foi realizado na Base de Dados Tropical (BDT), em Campinas, São Paulo, Brasil um workshop para discutir o Estabelecimento de uma Rede de Informações sobre Biodiversidade patrocinada pela União Internacional de Ciências Biológicas (IUBS), pela União Internacional das Sociedades de Microbiologia (IUMS) e pela Federação Mundial de Coleções de Culturas (WFCC), com o apoio do Governo Brasileiro (IBAMA, CNPq, Finep), da UNEP e do Conselho Britânico. O evento foi organizado pela BDT, e pelo Microbial Strain Data Network (MSDN). Como resultado do workshop foi estabelecida uma iniciativa internacional conhecida por BIN21 - Biodiversity Information Network/Agenda 21. O secretariado interino da BIN21, hoje esta localizado no Brasil, na própria BDT, que e também a mediadora da lista de discussão BIODIV-L (biodiv-l@bdt.ftpt.br), lançada em junho de 1992. A BDT foi convidada, pela WFCC (World Federation for Culture Collections), a organizar o evento (juntamente com o MSDN-Microbial Strain Data Network) por estar localizada no Brasil, país rico em biodiversidade, e pelo seu trabalho na informatização de dados biológicos. Participaram do Workshop 18 representantes de universidades e instituições de pesquisa do Brasil (ALTERNEX/IBASE, BDT/FTPT, EMBRAPA, FAPESP, IBAMA, IBICT, IBGE, SMA/SP, TELEBRAS, UNICAMP), e 21 representantes de organizações internacionais (México, Estados Unidos, Inglaterra, Suécia, Itália, Israel, Japão, Coreia, Rússia, Austrália). A participação não só da BDT, como também dos

demais Brasileiros no workshop foi muito positiva. Fomos nomeados coordenadores de alguns grupos de trabalho e conseguimos fazer com que os interesses dos países em desenvolvimento fossem preservados. A BDT hoje é a secretaria interina da BIN21. A configuração pretendida da rede prevê, não somente a produção de diretórios de informações já disponíveis em redes internacionais, mas a criação e/ou fortalecimento de nos regionais, com o papel de organizar e difundir a informação local, oferecer treinamento regional e auxiliar na busca de informação internacional. O grupo de trabalho técnico da BIN21 está discutindo formas de organizar esta informação em grupos temáticos, sendo que uma das ferramentas mais importantes atualmente disponíveis na Internet para a realização desta tarefa é o Gopher. Está se discutindo também a criação de interfaces para permitir o acesso dessa informação através de correio eletrônico. No Brasil, existem várias instituições gerando informações sobre biodiversidade e há também uma comunidade de usuários em grande expansão. Exemplos de instituições geradoras de informações são: IBAMA, EMBRAPA, IBGE, FIOCRUZ, INPA, GOELDI, Universidades como a USP, UFRJ, UnB, UFMG, UFRS, UFPR, UFPE, UNICAMP, UNESP (existem mais de 100), Jardins Botânicos, como os do Rio de Janeiro e São Paulo (aproximadamente 15), Jardins Zoológicos, como os de São Paulo e Rio de Janeiro (aproximadamente 40), Museus, como o Museu Nacional do Rio de Janeiro, e o Museu de Zoologia de São Paulo, INPE, IBICT, BINAGRI, BIRENE, ONGs, como Funatura, Biodiversitas, SOS Mata Atlântica, SPVS, Alternex/Ibase, BDT, além dos Departamentos Estaduais de Meio-Ambiente, entre outras. A grande maioria destas instituições não possui um sistema de informação de acesso público. É fundamental que este quadro seja revertido o mais rapidamente possível. A RNP em seu servidor Gopher lista 6 servidores Gopher no Brasil, sendo que somente a BDT e o BBRC (Brazilian Bioinformatics Resource Center) trabalham com informação biológica. Com informações relevantes para o tema biodiversidade, existe também a AlterNex/Ibase (não disponível na rede gopher), responsável pela disseminação de informação eletrônica durante a ECO-92. Os anais do evento foram publicados pela UNEP (United Nations Environment Protection), e estão sendo distribuídos gratuitamente para instituições internacionais atuantes em biodiversidade. Para maiores informações favor contatar: Dora Ann Lange Canhos, Gerente de Projetos Base de Dados Tropical, Fundação Tropical de Pesquisas e Tecnologia André Tosello, Rua Latino Coelho, 1301 - Parque Taquaral, 13087-010 Campinas, SP, Brasil, Tel: +55 192 42-7022, Fax: +55 192 42-7827, E-Mail: dora@bdt.ftpt.br.

ECOLOG: UM SISTEMA GERENCIADOR DE BANCOS DE DADOS PARA LEVANTAMENTOS ECOLÓGICOS DE CAMPO

Cavalcanti, M.J.

Departamento de Biologia Geral, Centro de Ciências Biológicas, Universidade Santa Úrsula, Rio de Janeiro, RJ Email: maurobio@ibase.br

A execução de inventários sobre a diversidade de espécies da flora e da fauna de uma região constitui-se num pré-requisito indispensável à conservação e uso sustentável dos recursos naturais por elas representados. Em particular nos países tropicais, aonde estes recursos vem sendo degradados rapidamente e em larga escala como consequência de políticas ambientais inadequadas, os inventários de biodiversidade tem adquirido uma importância cada vez maior. Dada a velocidade com que os ambientes naturais nos trópicos estão sendo degradados ou

inteiramente destruídos, é urgente que o maior número possível de áreas sejam inventariadas quanto as suas espécies de animais e plantas, muitas das quais talvez venham a ser extintas antes mesmo que possam ser descritas e, sobretudo, que tenham seus usos potenciais analisados. Contudo, o volume de dados obtido por inventários de biodiversidade freqüentemente constitui-se em um grande problema, no que se refere ao armazenamento, atualização, recuperação e análise. A dificuldade de acesso rápido as informações oriundas de tais inventários tem sido, em parte, responsável pela lentidão do processo de geração de resultados úteis para apoiar os esforços de conservação e uso sustentável dos recursos naturais. Neste contexto, um Sistema Gerenciador de Bancos de Dados (SGBD) apresenta-se como um eficiente meio de armazenamento e organização dos dados sobre diversidade de espécies, reduzindo seu tempo de acesso e processamento e garantindo sua integridade e consistência. Sistemas de bancos de dados vêm sendo empregados no manejo de dados de inventários de biodiversidade pelo menos desde a década de 70. Não obstante, até recentemente, poucos sistemas específicos estavam disponíveis, sendo todos invariavelmente desenvolvidos para computadores de grande porte ou minicomputadores; a proliferação dos microcomputadores, a partir dos anos 80, levou ao desenvolvimento de varias novas aplicações, de baixos custos operacionais e mais amplamente acessíveis. Neste trabalho, é apresentada uma aplicação de SGBD com capacidade relacional, especificamente projetada para o manejo de dados sobre locais, espécies e indivíduos obtidos em levantamentos de campo e inventários de biodiversidade. Este sistema foi desenvolvido para uso em equipamentos de baixo custo e fácil acesso (microcomputadores da linha IBM-PC, sob o sistema operacional MS-DOS), apresentando como principais características: a facilidade de interação com o usuário; a flexibilidade na recuperação de registros para consulta e geração de relatórios (a partir de qualquer campo ou combinação de campos); a integração com outros sistemas; e, sobretudo, a produção de análises quantitativas de dados ecológicos (índices de diversidade, curvas de rarefação, etc.). Para o desenvolvimento do sistema, foi utilizado o compilador Clipper 5.01, fazendo-se amplo uso de técnicas de Programação Orientada para Objeto (POO). Algumas questões relacionadas ao desenvolvimento de sistemas computacionais para a área biológica através da metodologia de POO são também apresentadas e discutidas. Apoio MacArthur Foundation/Shell do Brasil/CNPq/Jardim Botânico do Rio de Janeiro-IBAMA.

Modelagem e Simulação

MODELAGEM MATEMÁTICA E SIMULAÇÃO EM BIOLOGIA

Meyer, J.F.C.A.

Departamento de Matemática Aplicada, Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, Campinas, SP Email: joni@ime.unicamp.br

Neste trabalho é apresentada uma descrição matemática do processo de diálise extracorpórea. Para tanto é usada a transformação de uma Equação a Derivadas Parciais de Difusão-Advecção. E também proposto um método de aproximação da solução dessa EDP, com um esquema numérico para serem realizadas simulações computacionais. Este Método -

Crank-Nicolson na variável do tempo, e Elementos Finitos nas variáveis espaciais - é modificado usando Petrov-Galerkin devido às características do problema e, com base nas simulações realizadas, uma modificação possível é apresentada como sugestão para os aparelhos de hemodiálise.

A SIMPLE MODEL FOR MICELLIZATION: SIMULATION EXPERIMENT

Bernardes, A.T.¹; Bisch, P.M.²; Henriques, V.B.³

¹ *Universidade Federal de Ouro Preto, MG;* ² *Centro Brasileiro de Pesquisas Físicas e* ³ *Instituto de Química da Universidade de São Paulo Email: atbernar@brufmg.bitnet*

We have studied the characteristic properties of aggregation into micelles of amphiphiles in water through Monte Carlo simulations of a very simple model system. Amphiphiles and water are modeled as trimers and monomers respectively, on a square lattice. The model micellization presents both characteristics of experimental micellization: monomer vs amphiphile concentration with a plateau (a cmc) and a distribution of micelle sizes (polydispersity). Analysis of the model behaviour shows that usual criteria for the definition of the cmc might be contradictory. A normal type distribution of micelle sizes is observed only in the plateau range of concentrations. Sample size and relaxation properties of the simulation have also been analysed.

ENACT: AN ARTIFICIAL LIFE WORLD IN A FAMILY OF CELLULAR AUTOMATA

Balbi de Oliveira, P.P.

Instituto Nacional de Pesquisas Espaciais, Laboratorio Associado de Computação e Matemática Aplicada, INPE-LAC - Caixa Postal 515 , 12201-970 - São José dos Campos, SP. Email: pedrob@dpi.inpe.br

Enact is a family of two-dimensional, non- deterministic cellular automata, whose temporal evolution on a periodic background can be described in terms of the metaphor of an artificial-life world where a population of worm-like organisms of arbitrary length undergo an evolutionary process. During their lifetime, the organisms roam around, sexually reproducing, interacting with the environment, and being subjected to a developmental process which includes ageing and death. An organism is formed by a sequence of contiguous cells, só that the cells at each end can be intuitively thought of as its head and tail, whereas the cells in between constitute its body. One single cell of the body is the organism's genotype; the others are phenotype- like cells, since they are subjected to change through environmental interaction. The gene separates the phenotype-like cells into two parts: a single cell between the head and the gene, which we consider the actual organism's phenotype, and the others along the body (towards the tail), which is supposed to be the organism's memetype. Such a distinction comes from the fact that the initial state (during the neonate development) of the phenotypic cell depends on the organism's gene, whereas the state of the memetype does not, being determined through direct parental inheritance. As a consequence, the evolutionary process supported in Enact is in general both genotypic and memetic. The former refers to the evolution of a coordinated movement of the population, while the latter allows to explore the

effects of the genotypic evolution on the state of the organism's memetype. From a computational point of view, Enact can be regarded as a programmable, virtual, parallel machine defined by the artificial- life processes it supports, and relying upon six categories of states which represent environment, the organism's terminals, genotype, phenotype, memetype, and an additional category to allow the organisms to move. Its overall qualitative dynamics depends primarily on the ageing rate of the organisms, its tuning being very straightforward só as to prevent extinction of the organisms or deadlocks due to over-population, and guaranteeing the existence of very long transients. In this talk, I will skim over some aspects of Enact just enough to give a flavour about the computational and conceptual issues underlying its use and design. Time permitting, a video may be shown to illustrate it in action. Most of this research has been carried out at the University of Sussex, Brighton, England, thanks to a grant given by CNPq

SIMULAÇÃO DA DINÂMICA DE POPULAÇÕES DE PLANTAS

Figueira, J.E.C.: Bede, L.C.: Araujo, A.M.; Carvalho, L.M. e Fernandes, G.W.

Depto. Biologia Geral/ICB/Universidade Federal de Minas Gerais, Belo Horizonte, MG Email:

tabarana@brufmq.bitnet

Foi desenvolvido um programa para simular a dinâmica de populações de plantas em linguagem BASIC, para micros da linha IBM-PC. O programa leva em consideração a influencia do tamanho dos indivíduos na produção de sementes, nas distancias de dispersão, nas probabilidades de reprodução e sobrevivência. A estrutura da população e sua taxa de crescimento são mostradas a cada iteração. Parâmetros populacionais como taxas de sobrevivência e fecundidade podem ser facilmente alterados, sendo possível também o uso de equações que expressam a variação destes parâmetros em função da densidade e em resposta a compromissos internos (trade-offs). As curvas de dispersão de sementes e curvas de crescimento das plantas são expressas por meio de equações. Como o algoritmo discrimina cada individuo separadamente, pode projetar graficamente sua posição, crescimento, morte e reprodução a cada iteração. Além disso, diferentes taxas de sobrevivência e crescimento de sementes e plantas podem ser obtidas em diferentes regiões da tela gráfica, simulando microsítios diferenciados no solo. A grande flexibilidade e simplicidade deste programa permitem a simulação da dinâmica de plantas com diferentes características e historias de vida. Trabalho Financiado pela FAPEMIG, processo 1356/90.

UM MODELO PARA O EFEITO DE ANTAGONISTAS DO ESTRÓGENO NA LIGAÇÃO COOPERATIVA DO ESTRADIOL

Porrelli, R.N.¹; Munson, P.J.²; Rodbard, D.²

¹ Núcleo de Informática Biomédica, UNICAMP e ² Laboratory of Structural Biology, Division of Computer Research and Technology, National Institutes of Health, USA Email:

porrelli@ccsun.unicamp.br

Agonistas parciais tais como estriol e estrona são capazes de diminuir ou mesmo eliminar a curvatura do gráfico Scatchard quando competindo com estradiol marcado na ligação ao receptor de estrógeno. Alguns autores interpretam este achado como uma interferência dos agonistas no processo de dimerização do receptor. Para investigar como um segundo ligante

frio poderia provocar tal efeito, criamos um modelo de ação de massas onde receptores univalentes, que se ligam a dois diferentes tipos de ligantes, podem formar dímeros. Focalizamos nossa atenção nas manifestações de cooperatividade (tais como curvatura do Scatchard) para determinar como elas poderiam ser modificadas pela competição com um segundo ligante. Ao invés de discutirmos matematicamente as equações do modelo, criamos uma base de dados contendo soluções do modelo para virtualmente todas as combinações dos parâmetros, de forma a conter os pontos críticos e intervalos de interesse. Tal abordagem computacional permitiu uma nova interpretação dos achados experimentais, indicando que o efeito dos agonistas parciais na ligação de estradiol marcado ao receptor não se explica pela interferência com a dimerização do receptor ou pela diminuição da cooperatividade. Ela é aplicável a outros modelos, constituindo-se em alternativa a discussão das equações, desde que o número de parâmetros, pontos críticos e intervalos de interesse não exceda a capacidade de armazenagem do computador.

Inteligência Artificial e Reconhecimento de Padrões

REDES NEURAIS E ALGORITMOS GENÉTICOS EM BIOLOGIA

Blinder, P.B.

Departamento de Matemática Aplicada, Instituto de Matemática, Estatística e Ciência da Computação (IMECC), e Centro de Lógica e Epistemologia (CLE) da Universidade Estadual de Campinas. Email: blinder@ime.unicamp.br

As Redes Neurais Artificiais e os Algoritmos Genéticos têm sido utilizados numa infinidade de aplicações práticas, desde análise de crédito financeiro até o controle de fábricas. Este é um enfoque da metodologia (o prático). No entanto, pretendemos observar suas aplicações científicas, tais como classificação através de redes neurais e estudo de sistemas evolutivos por meio de AG. Após abordar os mecanismos básicos de como funcionam tais métodos, e exibir algumas aplicações recentes em Biologia (e se possível em Computação) pretendemos observar se é possível utilizá-los como fator de síntese em Ciência, e não apenas de modo analítico, uma vez que tais técnicas se baseiam em princípios realmente básicos das Ciências Biológicas, como o da Evolução. Pretende-se também realizar uma extensão dos conceitos, através de Lógica Nebulosa, de como podemos representar a vagueza e o conhecimento incompleto da realidade na modelagem de sistemas (eventualmente biológicos). Feito isto, teremos então uma visão global de como podemos modelar sistemas complexos com conhecimento parcial da realidade, que é onde residem os "problemas do mundo real", como é o caso em Biologia.

BEHAVIOR-BASED ACTIVE VISION

Pinhanez, C.S.

University of São Paulo, Institute of Mathematics and Statistics, Dept. of Computer Science, Cx. Postal 20570, São Paulo, SP, CEP: 01498-970 E-mail: pinhanez@ime.usp.br

A vision system was built using a behavior-based model, the subsumption architecture. This is a new artificial intelligence paradigm especially useful for designing artificial agents (normally robots) able to autonomously exist in the real world, dealing with noise, real-timeness and unpredictable environments. The so called active eye moves the camera's axis through the environment, detecting áreas with high concentration of edges, with the help of a kind of saccadic movement. Particular attention is given to the fovea-like sensor structure which enables the active eye to efficiently use local information to control its movements, including an implementation of a kind of saccadic movement. A measure for the eyes behavior is developed, and employed to evaluate the incremental building process and the effects of the saccadic movements on the whole system. A higher level behavior was also implemented, with the purpose of detecting long straight edges in the image, producing pictures similar to humans' hand-drawings. Robustness and efficiency problems are also addressed.

CLASSIFICAÇÃO TAXONÔMICA DE BACTÉRIAS ATRAVÉS DE REDES NEURAIS

Oliveira, Y.G.; Senna, A.L.; Meira Junior, W.; Bunte de Carvalho, M.L.

Departamento de Ciências da Computação da Universidade Federal de Minas Gerais, Belo Horizonte, MG Email: yuri@dcc.ufmg.br, senna@dcc.ufmg.br, meira@dcc.ufmg.br, mlbc@dcc.ufmg.br

A identificação taxonômica de organismos isolados é feita, em muitos casos, por análises visuais e posterior utilização de chaves de classificação obtidas a partir de testes bioquímicos, entre outros estudos, de exemplares conhecidos. Tais testes podem ser divididos em dois grupos: 1) quanto ao tipo de reação indicada, que pode ser negativa ou positiva; 2) quanto à natureza dos possíveis valores de resposta, podendo estar contidos em faixas de valores contínuos ou discretos. A classificação abordada neste trabalho se baseia em percentuais de reações positivas a um conjunto de testes previamente realizados. A primeira dificuldade para realizar essa classificação advém do fato de que vários organismos apresentam resultados bastante semelhantes no espectro do conjunto de testes. Outro fato importante é que nem sempre um determinado organismo responde positivamente a um certo teste. Assim, a classificação é feita com base nas probabilidades de resposta positiva de determinada espécie de organismo aos diversos testes. Uma implementação computacional eficiente para a solução deste problema é a utilização de Redes Neurais Artificiais, visto que elas são dotadas de uma grande flexibilidade e tolerância a presença de ruídos nos dados de caracterização, ou mesmo a ausência de parte destes. Utilizou-se uma rede neural artificial treinada com o algoritmo de aprendizado conhecido por Back Propagation, para identificar um conjunto de vinte e três espécies encontradas mais freqüentemente em ambientes clínicos da família Enterobacteriaceae. Para isso, foram fornecidos como entrada para a rede os percentuais de um conjunto dos vinte e seis testes, considerados os mais úteis na classificação taxonômica desses exemplares em específico. Além disso, foram desenvolvidas algumas ferramentas de

tratamento estatístico para análise dos dados antes destes serem fornecidos à rede, objetivando organizar as espécies de modo a facilitar seu aprendizado. A rede implementada consegue uma precisão aproximada de 90% no reconhecimento das bactérias. Foram feitos testes preliminares com entradas incompletas e contendo ruídos, que tem apresentado bons resultados, confirmando as expectativas. Como perspectivas futuras pode-se considerar a construção de árvores taxonômicas e a identificação dos testes de maior significância.

USO DE REDES NEURAIS EM TAXONOMIA NUMÉRICA BASEADA EM DADOS HETEROGÊNEOS

Senna, A.L.; Oliveira, Y.G.; Meira Junior, W.; Bunte de Carvalho, M.L.

Departamento de Ciências da Computação da Universidade Federal de Minas Gerais, Belo Horizonte, MG Email: senna@dcc.ufmg.br, yuri@dcc.ufmg.br, meira@dcc.ufmg.br

A diversidade de organismos vivos em nosso planeta torna indispensável o desenvolvimento de ferramentas matemáticas e computacionais capazes de auxiliar o trabalho de biólogos e outros especialistas que se dedicam a árdua tarefa de classificá-los. Uma boa ferramenta deve ser capaz não só de identificar a que grupo pertence um organismo, quando este lhe é apresentado, mas também de dar uma noção da similaridade de tal organismo com os representantes dos demais grupos. Um dos maiores problemas enfrentados durante a abordagem computacional desse problema é a forma como os dados - informações relativas às diversas espécies, obtidas por especialistas, que formarão o conjunto de conhecimento da rede - são apresentados ao sistema. Foram tratados, neste trabalho, dados cujo formato varia da seguinte forma: 1) faixa contínua: entradas tipicamente obtidas pela realização de testes onde organismos conhecidos responderam com variados valores dentro da faixa; 2) dados discretos: entradas cujo valor pode variar dentro de um conjunto finito de valores; normalmente são respostas a análises visuais onde o organismo apresenta uma dentre várias estruturas possíveis; 3) dados binários: entradas do tipo verdadeiro ou falso; tal tipo de dado e um caso particular do acima citado, seu destaque se deve a simplicidade de sua representação. As redes neurais artificiais se mostraram uma boa ferramenta na taxonomia alfa - classificação individual de organismos - devido a sua inerente capacidade de tratar com ruídos nos dados de entrada; uma rede neural artificial do tipo 'backpropagation' foi utilizada para tratar do problema de classificação de micro-organismos com resultados encorajadores: o método se mostrou eficiente e extremamente robusto. Para validar o sistema de classificação desenvolvido no Departamento de Ciência da Computação da UFMG - Neuraltaxon - foram utilizadas duas bases de dados taxonômicas distintas: a primeira contendo características de 26 espécies de pulgas; a segunda com dados de 20 protozoários da família Trypanosomatidae. Foi dimensionada e treinada uma rede neural para cada base. Verificou-se que ambas conseguem classificar representantes com precisão de aproximadamente 95% e são resistentes a possíveis deficiências na entrada de dados: se o usuário deseja identificar um organismo, mas dispõe de apenas 30% das informações padrão sobre o grupo, ainda assim o sistema emite uma resposta com nível de precisão em torno de 80%. Além disso, erros experimentais distorcendo 70% das entradas, por exemplo, ainda permitem uma taxa de acerto de aproximadamente 60%.

RECONHECIMENTO DE SEQÜÊNCIAS COMPORTAMENTAIS UTILIZANDO REDES NEURAIIS ARTIFICIAIS

Sabbatini, R.M.E.

Núcleo de Informática Biomédica da Universidade Estadual de Campinas Email:

sabbatini@ccvax.unicamp.br

A identificação, isolamento e quantificação dos padrões e seqüência presentes no comportamento de um animal, são pontos centrais da metodologia etológica. Diversas técnicas estatísticas foram desenvolvidas com essa finalidade, tendo como base modelos de Markov, gramaticais, e outros; todos eles bastante trabalhosos quanto à programação, e de resultados duvidosos quando o fluxo do comportamento tem probabilidades de transição instáveis. A abordagem apresentada neste trabalho parte de um pressuposto inteiramente diferente. Como o nosso sistema nervoso (e o dos animais que tem que reconhecer padrões comportamentais comunicativos complexos em seus conspecificos) funciona tão bem nessa tarefa? A resposta está na emulação das redes neurais que são capazes de efetuar os procedimentos fundamentais de reconhecimento e segmentação de padrões a partir do fluxo contínuo do comportamento. Desenvolvemos um simulador de redes neurais artificiais de tipo perceptron, com três camadas, capazes de aprender tarefas de reconhecimento de padrões através do algoritmo da retropropagação de erros. Este programa (NEURONET, para microcomputadores de 16 bits), foi treinado a reconhecer padrões (perfis não seqüenciais) e seqüência comportamentais de exemplo, classificadas previamente por um observador humano; até atingir o critério de 98.7 % de acertos. Utilizando perfis e seqüência de teste, a rede neural foi capaz de identificar o padrão correto em cerca de 90 % das mesmas. Outras possibilidades abertas com o uso de redes neurais são: a) a obtenção de um gráfico temporal mostrando as probabilidades de ocorrência das classes de comportamento ao longo do tempo. Estas curvas podem ser segmentadas (detecção de gradientes) através de uma segunda rede neural, que simula a organização da inibição lateral na retina. b) a quantificação do grau de estereotipia dos padrões e seqüência comportamentais; c) a detecção e interpretação automática de vocalizações e outros sinais contínuos emitidos por um animal; d) a implementação de sistemas de observação e categorização do comportamento totalmente automáticos (sem necessidade de observador humano) através da análise de imagens gravadas em filme e vídeo, por redes neurais artificiais. O presente trabalho pode demonstrar de forma definitiva a viabilidade e grande utilidade dos sistemas computacionais neuromorficos em etologia, utilidade esta que devera aumentar com o aparecimento de neurocomputadores hiperparalelos verdadeiros.

Aplicações Educacionais

USO DE PROGRAMAS PARA ENSINO DE GENÉTICA

Matioli, S.R.

Depto. de Biologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP. Email:

srmatiol@fox.cce.usp.br

O uso de computadores para simulação de fenômenos genéticos remonta praticamente ao início do emprego de computadores digitais em pesquisa científica, havendo registros deste emprego desde 1957. A Genética, embora seja uma área eminentemente biológica, compartilha muita característica das Ciências exatas, uma vez que, desde os trabalhos de Mendel, os fenômenos de transmissão hereditária podem ser, na maioria dos casos, tratados de forma quantitativa. Em 1983, começamos a utilizar computadores para ensino de Genética de Populações em cursos de graduação para Ciências Biológicas, na época através do emprego de terminal de computador de grande porte. Os programas foram escritos em FORTRAN e simulavam os processos de seleção natural e de deriva genética. As versões seguintes foram escritas em linguagem BASIC para microcomputadores da linha Sinclair, Apple e IBM-PC. A última versão incorpora além das simulações de deriva genética, equilíbrio para genes ligados ao sexo e migração entre populações. Este programa (GENPOP2) tem sido distribuído para várias Universidades do país e tem sido utilizado em cursos de graduação. A partir de 1990, através de colaboração com Dr. Paulo A. Otto, foi desenvolvido o programa POPGEN1, que, através de um sistema de coordenadas triangulares homogêneas, permite que as frequências gênicas sejam visualizadas simultaneamente com as frequências genotípicas, possibilitando que diferentes sistemas de cruzamentos possam ser estudados através de interpretação gráfica. Além do desenvolvimento de programas para o ensino de Genética de populações, também foi iniciado o desenvolvimento de programas para o ensino de Genética básica, por enquanto restrito ao programa HEREDO, que simula quatro tipos de herança em três gerações representadas por heredogramas. O programa HEREDO tem sido utilizado em cursos de graduação em caráter experimental.

SIMULAÇÃO DA DERIVA GENÉTICA EM POPULAÇÕES SUBMETIDAS OU NÃO A PROCESSOS DE SELEÇÃO: UM PROGRAMA DE APOIO AO ENSINO DE GENÉTICA DE POPULAÇÕES

Jacchier, S.; Silveira, C.H.; Abrantes, E.F. Laboratorio de Biologia Computacional, Departamento de Bioquímica- Imunologia, ICB/UFMG Email: silveira@dcc.ufmg.br

O fenômeno da deriva genética apresenta um caráter probabilístico cujo efeito torna-se mais evidente somente após a realização de um grande número de simulações envolvendo sucessivas gerações. Em populações naturais expostas a processos seletivos, a deriva genética pode determinar a fixação de genes com baixo valor adaptativo. O número de gerações necessárias para que esta fixação ocorra depende do tamanho da população inicial e de sua dinâmica de crescimento. Este programa foi desenvolvido com o objetivo de oferecer uma ferramenta didática para o estudo dos efeitos da deriva genética sobre populações submetidas ou não a pressões seletivas com padrões de crescimento variados. Através de uma interface interativa, o aluno terá condições de planejar e realizar experimentos, criando suas próprias estruturas populacionais. A apresentação gráfica dos resultados obtidos a partir destes experimentos permitira uma melhor visualização das características do processo de deriva em diferentes populações. A estratégia didática adotada baseia-se no princípio de que o aluno deveria ser capaz de propor hipóteses, elaborar simulações e analisar seus resultados. Este programa pretende ser parte de um futuro projeto a ser implantado no ICB/UFMG destinado

ao desenvolvimento de softwares aplicativos orientados para o auxílio do ensino de Biologia em suas mais diversas áreas. APOIO: CNPq, FAPEMIG.

CASOS CLÍNICOS SIMULADOS PARA COMPLEMENTAÇÃO DO ESTUDO DAS ABERRAÇÕES CROMOSSÔMICAS

Masuko, F.K.M.; Hackel, C.; Norato, D.Y.J e Sabbatini, R.M.E.

Departamento de Genética Médica Faculdade de Ciências Médicas e Núcleo de Informática Biomédica da UNICAMP, Campinas. Email: sabbatini@ccvax.unicamp.br

Os autores apresentam uma coleção de casos clínicos simulados desenvolvidos com o auxílio do sistema de autoria MEDTEST, para microcomputadores da linha IBM-PC. O conjunto de casos inclui pacientes com as aberrações cromossômicas mais frequentes na prática médica, com o objetivo de serem utilizados em exercícios para fixação da aprendizagem por alunos de graduação na área de saúde. Deste modo, procurou-se elaborar casos típicos, evitando-se as exceções, com ênfase no quadro clínico, diagnóstico diferencial, solicitação de exames complementares, coleta e interpretação dos resultados. Também constam dos casos as decisões quanto à conduta em relação à criança, a orientação familiar e o aconselhamento genético. Foram elaborados seis casos, compreendendo as síndromes de Down, Edwards, Patau, Turner, Klinefelter e do sítio frágil do cromossomo X. Os casos serão testados por alunos de graduação do curso médico, tendo sido avaliados por dois residentes em genética clínica e considerados de fácil entendimento. Apoio do Fundo de Apoio ao Ensino e Pesquisa da UNICAMP (FAEP).

Outras Aplicações

COMPARAÇÃO DE MÉTODOS DE INFERÊNCIA FILOGENÉTICA

Meyer, D.

Departamento de Biologia, Universidade de São Paulo. Email: diogo@volterra.Stanford.EDU

Métodos de inferência filogenética são ferramentas construídas para gerar filogenias hipotéticas (árvores). Estas ferramentas são receitas numéricas, que encontram a árvore ótima (aquela que minimiza ou maximiza um ou mais parâmetros). Este parâmetro varia conforme o método usado: número de mudanças de caráter na árvore (minimizado por métodos de parcimônia), variância residual (minimizada por métodos de least squares), semelhança global (maximizada por métodos genéticos como UPGMA). Diferentes propriedades de métodos de inferir filogenias podem ser comparadas (Penny et al, 1992. TREE 7:73). 1. Eficiência. Escolhido o parâmetro que desejamos maximizar ou minimizar, o desenvolvimento de um algoritmo que o realize em um tempo computacional razoável é um desafio. Métodos de distância são eficientes - o sucesso de encontrar a melhor árvore é virtualmente garantido. Métodos de máxima verossimilhança encontram árvores ótimas para poucos táxons, mas se tornam

inviáveis para conjuntos maiores. Análises de parcimônia se tornaram mais eficientes com novos algoritmos (por exemplo, métodos de branch-and bound), mas para análises de números maiores de táxons são ineficientes (não há garantia de encontrar a melhor árvore). Melhorias na eficiência de métodos dependem em grande parte da contribuição de cientistas de computação. 2. Potência. À medida que mais dados são utilizados para realizar uma inferência, o método tende a convergir para uma única árvore (ótima para o algoritmo) e a adição de mais informação não o afasta desta topologia. Detectar a taxa sobre as quais métodos convergem para uma árvore ótima à medida que mais dados são incorporados serve como uma orientação para decidir que quantidade de informação deve ser usada para encontrar a árvore ótima associada a determinado método. Nossos experimentos comparam UPGMA, Neighbour-Joining e máxima verossimilhança quanto a suas taxas de convergência. Os resultados indicam que (a) UPGMA converge mais lentamente para uma árvore ótima e (b) aumento do número de taxa leva a taxas de convergência mais lentas. 3. Consistência. A melhor árvore encontrada por determinado método não é necessariamente a árvore correta. Para que a filogenia inferida difira da real basta que o critério utilizado para encontrar a melhor árvore assuma pressupostos do processo evolutivo que sejam violados na natureza. Detectar pressupostos associados a métodos de inferência filogenética constitui uma tarefa difícil. A utilização de métodos de simulação por computadores fornece contribuições para a detecção dos pressupostos necessários para que um método seja consistente (encontre a árvore real). Todos os itens acima podem ser explorados de maneira criativa por evolucionistas e cientistas de computação, com equipamentos acessíveis a Universidades Brasileiras.

CONJUNTO DE ROTINAS PARA AVALIAÇÃO DA CAPACIDADE SECRETORA E SENSIBILIDADE PERIFÉRICA À INSULINA

Porrelli, R.N.; Sabbatini, R.M.E.; Wajchenberg, B.L. Núcleo de Informática Biomédica, UNICAMP e Unidade de Diabetes e Adrenal, Disciplina de Endocrinologia, HCFMUSP Email:

porrelli@ccsun.unicamp.br

Estudos de metabolismo de glicose e sensibilidade a insulina geralmente envolvem técnicas tais como uso de radioisótopos, peptídeo C recombinante e realização de clamp euglicêmico hiperinsulinêmico. As dificuldades técnicas, associadas ao alto custo restringem seu uso. Foram criados vários testes alternativos, geralmente envolvendo infusão de hormônios ou drogas (glicose, insulina, glucagon, tolbutamida, arginina) e medidas seriadas de insulina, glicose, e peptídeo C. De execução mais simples, estes métodos permitem a estimativa da sensibilidade a insulina quando não se dispõe das técnicas citadas. Criamos uma série de rotinas, escritas em BASIC, para a interpretação dos resultados destes testes. Entre eles podemos citar o modelo mínimo de Bergman (avaliação de sensibilidade a insulina), a desconvolução do peptídeo C (taxa de secreção pre-hepática de insulina), a desconvolução de insulina (cálculo da oferta periférica), extração hepática de insulina e estimativa da massa de receptores periféricos. As rotinas foram utilizadas para avaliação da eficácia do tratamento com hipoglicemiantes orais em Diabetes tipo II, bem como da sensibilidade a insulina e capacidade secretora em moléstia que cursam com distúrbios do metabolismo da glicose e/ou ação da insulina.