**Bioinformática**

**Bárbara Fernandes Santana1, Gabriela Karina Pedro Oliveira1, Josiane Aparecida Silva1, Julia Helena Valadares Rodrigues1, Larissa Cristina Barboza Abreu1, Tamires Miranda Milagres Partilho1**

1Universidade Federal de São João Del Rei (UFSJ)

CAP-Ouro Branco-MG-Brasil

barbarafers@hotmail.com, gabykpoliveira@hotmail.com, josianehta@hotmail.com, jujujulia\_55@hotmail.com, larissabarboza07@hotmail.com, tamiresmp@hotmail.com

**1-Bioinformática e sua finalidade:**

A bioinformática é a disciplina que interliga ciência da computação e biologia molecular, utilizando conhecimentos de física, biologia, química, informática, ciência da computação e matemática a fim de se criar diferentes tipos de programas computacionais que possam auxiliar a biologia em si, seja codificando genes, encontrar o fator de inibição da enzima ou em qualquer outro aspecto.

**1.1-Áreas de interesse:**

Aplicações de Informática em Biotecnologia (em especial de Inteligência Artificial, Teoria da Computação, Hardware e Proc. de Imagens em problemas ligados ao genoma).

Principais assuntos veiculados: aplicações de IA, DNA computing, hardware específico para aplicações em genoma, chamadas de conferências nestas áreas e notícias de fontes como as revistas Nature, Science e Sc. American e imprensas nacional e internacional.

**2- A história da bioinformática:**

   Do inicio até os meados do século passado, os geneticistas e químicos se questionavam sobre a natureza química do material genético. Das pesquisas desenvolvidas surgiu a conclusão de que o DNA era molécula que armazenava a informação genética.

Com o surgimento dos sequenciadores automáticos de DNA, houve uma explosão na quantidade de sequência a serem armazenadas, e que exigiam recursos computacionais cada vez mais eficientes. Com isso surgiu a bioinformática. Uma nova ciência que envolveria diversos conhecimentos como a engenharia de softwares, a matemática, a estatística, a ciência da computação e a biologia  molecular.

 O primeiro computador digital foi criado em 1940, mais so começou a funcionar em 1946.

  A descoberta do DNA em 1953 mostrou que a informação genética  é armazenada de forma digital, ou seja, é escrita com um alfabeto quartanário, pois são usadas quatro letras A, C, G, T.

Apesar da estrutura do DNA ter sido  desvendada em 1953, a informação nela contida não podia ser lida, mas em 1995 uma maquina já ler milhares dessas letrinhas.

Com a evolução da informática,fez com que os computadores fossem capazes de armazenar cada vez mais informação,de processá-la cada vez mais rápido,com custo cada vez menor.

 A informática é uma ciência que vem adquirindo seu espaço principalmente no setor da biologia molecular e informática a nível internacional, onde tem por objetivo simplificar as pesquisas cientificas obtendo maior êxito nas sequências das moléculas de DNA e muitas proteínas que compõem o ser vivem.

 Através de uma evolução que parece sincronizada, os computadores já estavam suficientemente poderosos para poder processar os milhões e milhões de letrinhas. Foi assim que nasceu a BIOINFORMÁTICA que se tornou de grande importância para a ciência, que é através dela que pode construir base de dados com informações dos genes  e proteínas dos organismos vivos.

**3- Técnicas da bioinformática:**

A Bioinformática combina conhecimentos de [química](http://pt.wikipedia.org/wiki/Qu%C3%ADmica), [física](http://pt.wikipedia.org/wiki/F%C3%ADsica), [biologia](http://pt.wikipedia.org/wiki/Biologia), [ciência da computação](http://pt.wikipedia.org/wiki/Ci%C3%AAncia_da_computa%C3%A7%C3%A3o), [informática](http://pt.wikipedia.org/wiki/Inform%C3%A1tica) e [matemática](http://pt.wikipedia.org/wiki/Matem%C3%A1tica)/[estatística](http://pt.wikipedia.org/wiki/Estat%C3%ADstica) para processar dados biológicos ou biomédicos.

Buscando tratar os dados, é necessário desenvolver [softwares](http://pt.wikipedia.org/wiki/Software) para, por exemplo: identificar [genes](http://pt.wikipedia.org/wiki/Gene), prever a configuração tridimensional de [proteínas](http://pt.wikipedia.org/wiki/Prote%C3%ADna), identificar inibidores de [enzimas](http://pt.wikipedia.org/wiki/Enzima), organizar e relacionar informação biológica, simular [células](http://pt.wikipedia.org/wiki/C%C3%A9lula), agrupar [proteínas](http://pt.wikipedia.org/wiki/Prote%C3%ADnas) [homólogas](http://pt.wikipedia.org/wiki/Homologia_%28biologia%29), montar [árvores filogenéticas](http://pt.wikipedia.org/wiki/%C3%81rvores_filogen%C3%A9ticas), comparar múltiplas comunidades microbianas por construção de bibliotecas genômicas, analisar experimentos de [expressão gênica](http://pt.wikipedia.org/wiki/Express%C3%A3o_g%C3%AAnica) entre outras inúmeras aplicações.

As técnicas utilizadas para o aprendizado das máquinas utilizadas na bioinformática são: redes neurais artificiais, máquinas de vetores suporte, árvores de decisão, algoritmos genéticos, algoritmos de agrupamento, cadeias escondidas de Markov (HMMs).

  Redes Neurais: são funções não-lineares complexas com muitos parâmetros. Seus parâmetros podem ser aprendidos a partir de dados ruidosos, e elas são usadas em milhares de aplicações. As redes neurais artificiais possuem algumas características que as tornam satisfatoriamente aplicáveis a problemas não adequados à computação tradicional:

-**Generalização:** Podem aprender através de conjuntos de exemplos e apresentar respostas coerentes para entradas não vistas durante o treinamento.

-**Adaptabilidade:** Podem adaptar-se ao seu novo ambiente através de alterações em seus pesos sinápticos, no caso de operarem em outro ambiente. Também podem ser projetadas para alterarem seus pesos em tempo real ou para operarem em ambientes que variem com o tempo.

-**Informação Contextual:** Processam as informações contextuais de forma natural, uma vez que o processamento de um neurônio é afetado pelo processamento de outros neurônios da rede.

-**Uniformidade:** A mesma notação é utilizada em diferentes domínios de  aplicações; os neurônios são encontrados em todas as redes neurais; é possivel utilizar os mesmos algoritmos de aprendizagem e teorias em diversas apllicações; através da integração homogênea de módulos, podemos construir redes modulares.

 A modelagem de uma rede neural depende da análise consistente de um sistema  muitas vezes complexo, implicando em dificuldades para definir qual arquitetura melhor responde às necessidades do problema proposto e na escolha de quais dados são verdadeiramente relevantes para o processamento. Além da entrada, também devemos definir de forma ideal os parâmetros de aprendizagem, os pesos sinápticos e os níveis de bias, os quais são de severa importância para o processo de aprendizado.

Outra dificuldade encontrada seria a extração de regras justificativas da decisão tomada pela rede, as quais representariam o conhecimento adquirido durante o treinamento. Por este motivo, as redes neurais são apelidadas de "Caixas Pretas".

**Máquinas de vetor suporte**: Tem boa capacidade de generalização e robustez diante de dados de grande dimensão, como o os presentes em grande parte das aplicações envolvendo o reconhecimento de gêneses e análise  de dados de expressão.

**- Árvore de decisão:** As árvores de decisão organizam as informações em uma estrutura composta de nós e ramificações. Cada nó representa um atributo e está associado a um teste relevante para a classificação de dados, ou a uma classe, no caso do nó ser folha. Uma árvore de decisão alcança sua decisão executando uma seqüência de testes. Os ramos representam cada um dos testes aplicados. A classificação de uma instância se dá seguindo os nós e as ramificações formado até que uma folha seja alcançada.

Uma das vantagens das ADs está na compressibilidade das regras de classificação    geradas.   Em aplicações médicas, por exemplo, é necessário não somente predizer uma doença, mas também especificar quais sintomas determinaram seu diagnóstico.

- **Algoritmos genéticos:** São técnicas inteligentes baseadas nos mecanismos do processo de evolução biológica e da genética. Segundo a teoria da evolução , os organismos que se adaptam melhor a seu ambiente possuem maiores chances de terem suas características reproduzida em uma nova geração. Seu funcionamento básico consiste na alternância de vários ciclos de geração: geração de população, avaliação dos seus indivíduos e aplicações dos operadores genéticos sobre aqueles considerados aptos.

**- Algoritmo de Agrupamento:** faz os agrupamentos de dados de padrões. O primeiro consiste em agrupar classes mais similares em cada nível da árvore multiclasses com o algoritmo k-m´edias, adotando k igual a 2. No segundo, as classes s˜ao divididas de acordo com a distância média de seus dados á origem. No terceiro método, as classes são divididas em subconjuntos para os quais a diferença no núumero de exemplos é mínima.

-**Cadeias Escondidas de Markov (HMMs):** fornece uma estimativa razoáveis das probabilidades posteriores verdadeiras em uma rede e podem lidar com redes muitos maiores do que os algoritmos exatos.

 **4-Bioinformática no Brasil:**

 A Bioinformática chegou no Brasil através de Neshich, de origem sérvia, que foi idealizador do projeto BrasilianBioNet que era uma rede de usuários da bioinformática, que permitiram os primeiros contatos de cientistas brasileiros com o programa de analises de sequência de DNA, gratuitamente, através de um computador da Embrapa Recursos Genéticos e Biotecnologia.

 Depois disso a biotecnologia foi crescendo e se desenvolvendo no Brasil, para auxiliar esse processo foi criada uma Associação Brasileira de Bioinformática e Biotecnologia Computacional, que visava promover uma maior interação entre vários especialistas de áreas importantes como a ciência da computação, medicina e biologia. Houve também o ONSA (Organization for Nucleotide Sequencing and Analysis), que auxiliou a disseminação da bioinformática no Brasil, viabilizando o estudo dinâmico e virtual, criando vários centros de pesquisa a partir dessa rede, espalhando para diversas instituições por todo o país, como a UNICAMP,USP,UNESP,UFMG, o Laboratório nacional de Computação Cientifica, UFPE,UFRS e o Instituto Ludwig.

 A realização de congressos e seminários desse tema consolidam a imagem do Brasil no cenário internacional, como por exemplo o Congresso Internacional de bioinformática, realizado em Fortaleza em 2006, que teve a presença de dois cientistas do Premio Nobel: Kurt Witchich e Robert Huber, o que proporcionou a troca de ideais e aperfeiçoamento dos cientistas, e em 2007 realizado em São Paulo, onde foram apresentadas pesquisas inovadoras em áreas como genômica, biotecnologia, mapeamento genético, identificação de doenças, visualização e simulação do sistema biológico.

**5-Aplicações da bioinformática:**

 As inúmeras informações geradas pelos pesquisadores e profissionais atuantes da área

 biológica possuem a necessidade de serem guardadas de alguma maneira para que esses

 possam desenvolver seus estudos de forma organizada.

O Sistemas Gerenciadores de Banco de Dados que consiste em um conjunto de programas de computador responsáveis pelo gerenciamento de uma base de dados, na maioria das vezes possui finalidade administrativa, no qual não suporta informações mais complexas.
  A necessidade de utilizar informações mais complexas fez com que  a bioinformática se tornasse uma importante ferramenta para que pesquisadores e profissionais da área possam fazer a realização de seus estudos. Ela se tornou a principal opção pois é através dela que permite localizar com mais eficiência as informações desejadas, mas não basta apenas obter métodos rápidos para a busca é necessário que o usuário da bioinformática tenha uma base de conhecimento biológico para compreender as informações a serem obtidas. Um ponto positivo da bioinformática é a grande capacidade de armazenamento de códigos genéticos.

Para facilitar a compreensão dos usuários da bioinformática é necessário que os métodos e tipos de dados utilizados no banco de dados sejam padronizados e devem haver também a descrição das suas funcionalidades .Por este fato é necessário que seja criado diferentes formas para o armazenamento, acesso e procura desses dados facilitando assim a busca de informações necessárias.

 Primeiramente deve-se escolher um Sistema Gerenciador de Banco de Dados que sustente a criação de novos métodos e tipos de dados. Após isto deve-se adotar sistemas envolvendo Data Warehouses ou armazém de dados, que consiste em uma tecnologia para armazenar e organizar em um banco de dados altos volumes de dados relativos aos conteúdos que se deseja trabalhar.

 Mas, para que as buscas e análises de bancos de dados sejam eficientes, podem ser utilizadas técnicas de inteligência Artificial -  é um ramo da ciência da computação que se propõe a elaborar dispositivos que simulem a capacidade humana de raciocinar, perceber, tomar decisões e resolver problemas - e sistemas no qual foram desenvolvidos para verificar dados em problemas que pretende-se obter a solução, como foi o caso da Joana S. de P. Gonçalves (2007), ganhadora do Prêmio Nacional de Trabalhos de Licenciatura, atribuído pela associação portuguesa para inteligência artificial, que desenvolveu o sistema BiGGEsTS.
          Segundo  Gonçalves (2007) , o BiGGEsTS (Biclustering Gene Expression Time-Series) consiste numa ferramenta de análise, manipulação de diversos conjuntos de dados de expressões genéticas. As suas funções principais são:

 • Realiza a manipulação de dados de expressões genéticas em simultâneo;

  • No sistema foram aplicados diversos algoritmos de pré-processamento(filtragem de genes, preenchimento de valores em falta, normalização,smoothing e discretização), algoritmos de biclustering e cálculo de funções biológicas significativas para análise dos dados;

  • O sistema permite a visualização dos dados ao decorrer do processo de manipulação, utilizando-se representações gráficas inseridas posteriormente, também é feito uma analise detalhada dos resultados obtido.

**6-Considerações finais**

 O texto teve como objetivo definir o termo “Bioinformática”, esclarecer suas aplicações e sua importância no cenário nacional e internacional de pesquisa. Com isso definiu um ponto de partida para os seminários da turma do primeiro período de Engenharia de Bioprocessos, da disciplina de Algoritmo e Estrutura de dados, da Universidade Federal de São João Del Rei.