



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Instituto de Biofísica Carlos Chagas Filho

Concurso para Professor Adjunto

MS-020 – Biologia Computacional

Edital nº 953, de 20 de dezembro de 2019, publicado no DOU nº 248, de 24 de dezembro de 2019 - consolidado com as alterações dos Editais nº 9, de 9 de janeiro de 2020, nº 31, de 03 de fevereiro de 2020, nº 48, de 11 de fevereiro de 2020 e nº 116 de 25 de março de 2020



NOME: DIEGO ENRY BARRETO GOMES

No. fl.

1

04 - MODELAGEM DE REDES BIOLÓGICAS.

O ADVENTO DAS CIÊNCIAS ÔMICAS UNIDO AOS AVANÇOS TÉCNICOS DE ENGENHARIA, E APRIMORAMENTO E DESENVOLVIMENTO DE MÉTODOS E TÉCNICAS DE BIOINFORMÁTICA E BIOLOGIA COMPUTACIONAL DISPONIBILIZAM UMA QUANTIDADE TREMENDA DE INFORMAÇÕES E DADOS SOBRE OS DIVERSOS NÍVEIS ORGANIZACIONAIS DA VIDA, INCLUINDO GENOMAS, TRANSCRIPTOMAS, PROTEOMAS, METABOLOMAS, BEM COMO ABUNDANTES EVIDÊNCIAS EXPERIMENTAIS DE SUAS INTERRELAÇÕES, EM MUITOS SISTEMAS MODELO. A CURADORIA DOS DADOS, ORGANIZADOS E ESTRUTURADOS EM BANCOS DE DADOS RELACIONAIS PERMITE ABORDAGENS "DATA-DRIVEN" FACILITANDO A CONSTRUÇÃO DE MODELOS CONCISOS QUE DESCREVEM E EXPLICAM O FUNCIONAMENTO DOS SISTEMAS E SUBSISTEMAS COMO UM TODO, NUMA SITUAÇÃO IDEAL.

A MODELAGEM COMPUTACIONAL TEM UM PAPEL FUNDAMENTAL PARA NOS AJUDAR A COMPREENDER A COMPLEXIDADE DOS COMPONENTES DOS SISTEMAS (GENES, RNA, PROTEÍNAS, METABOLITOS, ORGANISMOS) BEM COMO SUAS INTERRELAÇÕES, NA FORMA DE REDES DE INTEGRAÇÃO. A MODELAGEM DE REDES BIOLÓGICAS VISA FAVORECER A COMPREENSÃO SISTÊMICA EM VÁRIOS NÍVEIS, E É CALIBRADA

É VALIDADA COM DADOS EXPERIMENTAIS. PARTICULARMENTE ELA PERMITE COMPREENDER A ESTRUTURA DE REDES DE REGULAÇÃO GENÉTICA, REDES DE INTERAÇÃO PROTEÍNA-PROTEÍNA, REDES DE INTERAÇÃO BIOQUÍMICA, REDES METABÓLICAS ETC. TAMBÉM PODE AVALIAR O COMPORTAMENTO DINÂMICO DOS SISTEMAS AO INTEGRAR, POR EXEMPLO, ~~EM~~ DADOS DE ENSAIOS DE ALTO RENDIMENTO EM ESCALA TEMPORAL, PODENDO ENTÃO AVALIAR EFEITOS EM CONDIÇÕES DIVERSAS, EFEITOS DE PERTURBAÇÕES COMO TEMPERATURA, PRESSÃO, CONCENTRAÇÃO DE MOLÉCULAS, ENTRE OUTROS. UM MODELO VALIDADO DEVE TER ~~PO~~ PODER PREDITIVO E PODE SER USADO COMO ALICERCE PARA GUIAR MANIPULAÇÃO DO SISTEMA OU PREVER COMPORTAMENTOS, ESPECIALMENTE EM CONDIÇÕES DIFÍCEIS DE PRODUZIR EXPERIMENTALMENTE. DIVERSAS IMPLICAÇÕES PODEM SURTIR DE UM BOM MODELO, COMO O PLANEJAMENTO DE INTERVENÇÕES TERAPÊUTICAS BASEADAS NO CONHECIMENTO DO SISTEMA, E PLANEJAMENTO DE NOVAS LÓGICAS GENÉTICAS REGULADORIAS QUE PODEM SER INTRODUZIDAS EM ORGANISMOS PARA ALCANÇAR UMA FUNÇÃO DESEJADA.

FORMALMENTE AS REDES BIOLÓGICAS SÃO CONSTRUÍDAS NA FORMA DE DADOS ESTRUTURADOS CONTENDO COMPONENTES QUE PODEM SER GENES, PROTEÍNAS, METABOLITOS ETC, ATÉ FENOTIPOS COMPLETOS, COMO DOENÇAS. OS COMPONENTES SÃO OS NÓS. AS ARESTAS, QUE CONECTAM OS NÓS, REPRESENTAM AS RELAÇÕES PONDERADAS ENTRE OS COMPONENTES,



AS RELAÇÕES PODEM SER QUANTITATIVAS, CONTENDO INFORMAÇÃO ESTEQUIOMÉTRICA, FORMA E CONSTANTES DE RELAÇÃO, BEM COMO PODE SER QUALITATIVA, ENFATIZANDO CONCEITOS ESPECÍFICOS.

APESAR DE ABUNDANTES OS DADOS EXPERIMENTAIS NEM SEMPRE SÃO PERFEITOS, ELAS PODEM SER INCOMPLETOS PARA A DESCRIÇÃO DA REDE DE INTERESSE OU IMPRECISOS. A CONSTRUÇÃO DAS REDES BIOLÓGICAS É UM PROCESSO COMPLEXO. É METICULOSO. UMA DAS CARACTERÍSTICAS MARCANTE DOS SISTEMAS BIOLÓGICOS QUE DEVE SER CONSIDERADA É A ROBUSTEZ. ESSA CARACTERÍSTICA É ESSENCIAL PARA QUE AS ESPÉCIES SE DESENVOLVAM E EVOLUAM, ELA É A CAPACIDADE DE FUNCIONAR EM MEIO A PERTURBAÇÕES CONSTANTES, DE TEMPERATURA, ~~DE~~ CONCENTRAÇÃO DE MOLECULAS, AMEAÇAS EXÓGENAS E ENDÓGENAS (COMO CANCER), MUTAÇÕES. A ROBUSTEZ SE MANIFESTA POR MECANISMOS DE COMPENSAÇÃO, COMO VIAS ALTERNATIVAS DE METABOLISMO, MECANISMOS REDUNDANTES, ETC.

AS REDES BIOLÓGICAS SÃO COMUMENTE MODELADAS COMO REDES BAYESIANAS, QUE PERMITEM CONSTRUIR REDES QUE RELACIONAM AS EVIDÊNCIAS DE DIFERENTES TIPOS DE DADOS À UMA PROBABILIDADE DE RELACIONAMENTO FUNCIONAL. ESTA PROBABILIDADE DE INTERAÇÃO PODE SER FÍSICA OU REGULATÓRIA. O PONTO FORTE DA ABORDAGEM BAYESIANA É PONDERAR AUTOMATICAMENTE A ACURÁCIA E COBERTURA DE CADA DADO DE ENTRADA.

OS ELEMENTOS CRÍTICOS DA REDE BAYESIANA



NOME: DIEGO ENRY BARRETO GOMES

No. fl.

4

SÃO PORTANTO: A ESTRUTURA DA REDE, DESCRREVENDO OS COMPONENTES DO SISTEMA (NÓS) E SEUS RELACIONAMENTOS (ARESTAS), É A TABELA DE PROBABILIDADE FUNCIONAL QUE MEDE A CONTRIBUIÇÃO DE CADA TIPO DE DADO NA RELAÇÃO.



03 - FERRAMENTAS ESTATÍSTICAS EM BIOLOGIA COMPUTACIONAL.

AS ANÁLISES ESTATÍSTICAS PERMEIAM TODAS AS ÁREAS DA BIOLOGIA COMPUTACIONAL, SENDO FERRAMENTAS FUNDAMENTAIS PARA APOIAR A CONSTRUÇÃO DE MODELOS, TOMADA DE DECISÃO, ENCONTRAR RELAÇÕES ENTRE COMPONENTES DE UM SISTEMA COMO GENES E SUAS VIAS DE REGULAÇÃO, TRANSCRITOS E PROTEÍNAS, RECEPTORES E LIGANTES. ALÉM DISSO SÃO ESSENCIAIS PARA VALIDAR O PROJETO DE EXPERIMENTOS, GUIANDO O USO EFETIVO DE RECURSOS PARA UMA AMOSTRAGEM SUFICIENTE E SIGNIFICATIVA DO PROCESSO/FENÔMENO QUE QUEREMOS OBSERVAR. AS FERRAMENTAS ESTATÍSTICAS NOS PERMITEM CORRELACIONAR VARIÁVEIS, REDUZIR A COMPLEXIDADE DE PROBLEMAS MULTIVARIADOS, FACILITANDO A RACIONALIZAÇÃO DOS PROBLEMAS, BEM COMO FACILITA O AGRUPAMENTO E CLASSIFICAÇÃO DE INDIVÍDUOS (AMOSTRAS).

COM O USO MASSIVO DE DADOS ÔMICOS MUITOS MÉTODOS E FERRAMENTAS ESTATÍSTICAS FORAM DESENVOLVIDOS E IMPLEMENTADOS EM PACOTES DE SOFTWARE COMO O MATLAB ("STATISTICS TOOLBOX") E NA LINGUAGEM "R", NESTA ÚLTIMA SE DESTACA A ENORME COMUNIDADE DE DESENVOLVIMENTO EM BIOINFORMATICA E BIOLOGIA COMPUTACIONAL POR TRAZ DO REPOSITÓRIO "BIOCONDUCTOR" COM FERRAMENTAS E PROTOCOLOS DIVERSOS PARA AS MAIS DIVERSAS "PERGUNTAS BIOLÓGICAS".

A "PERGUNTA BIOLÓGICA" PRECISA SER FORMULADA COM MUITO CUIDADO E TRANSFORMADA EM HIPÓTESE ESTATÍSTICA DE FORMA ADEQUADA, POR EXEMPLO



NUM EXPERIMENTO DE ANÁLISE DE CITOCINAS EM RESPOSTA A UMA INFLAMAÇÃO, É CORRETO BUSCAR CORRELAÇÕES EM RESPEITO A ALTERAÇÕES NA TRANSCRIÇÃO GENÉTICA, NO ENTANTO CASO A INVESTIGAÇÃO FOSSE ANALISAR AS EM RESPOSTA A UM ESTRESSE OXIDATIVO, A METODOLOGIA SERIA INVÁLIDA. MESMO QUE EVENTUALMENTE EXISTA UMA CORRELAÇÃO, ESTE EXEMPLO EMBLEMÁTICO ENFATIZA UM ENSINAMENTO COMUM NA ÁREA: CORRELAÇÃO DE QUALQUER TIPO NÃO IMPLICA CAUSA.

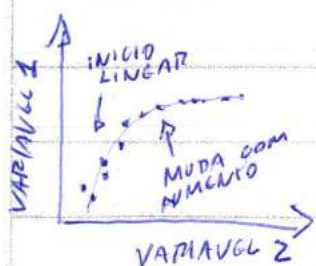
AS FERRAMENTAS ESTATÍSTICAS PODEM SER AGRUPADAS EM DUAS GRANDES CLASSES DE MÉTODOS - OS UNIVARIADOS EM QUE DUAS SITUAÇÕES (VARIÁVEIS) ESTÃO SENDO INVESTIGADAS E OS MULTIVARIADOS, QUANDO É PRECISO LIDAR COM MÚLTIPLAS VARIÁVEIS.

COMPUTAR A CORRELAÇÃO ENTRE DUAS SITUAÇÕES NOS FORNECE INSIGHTS SOBRE O RELACIONAMENTO ENTRE ELAS. O COEFICIENTE DE CORRELAÇÃO DE PEARSON É UMA DAS FERRAMENTAS MAIS USADAS NESTE CONTEXTO. ELE FORNECE UMA RESPOSTA FACILMENTE INTERPRETÁVEL, CORRELAÇÃO POSITIVA OU NEGATIVA, BEM COMO A INTENSIDADE. POR EXEMPLO NUM EXPERIMENTO EM QUE O PESQUISADOR DESEJA SABER A RELAÇÃO ENTRE TRANSCRITOS (mRNA) E PROTEÍNAS, UM RESULTADO DE CORRELAÇÃO POSITIVA INDICA QUE A PRODUÇÃO DE PROTEÍNAS ACONTECE SOB DEMANDA, ~~PARTE QUE~~ E QUE ~~NÃO~~ HÁ POUCO MECANISMO REGULATÓRIO, ENQUANTO UMA CORRELAÇÃO NEGATIVA INDICA



UM PROVAVEL EFEITO DE "FEEDBACK" NEGATIVO.

A CORRELAÇÃO DE PEARSON NO ENTANTO POSSUI A FRAGUESA DE MEDIR CORRELAÇÕES LINEARES, PARA SUPE-
RAR ESTE TIPO DE LIMITAÇÃO, NOS CASOS ADEQUADOS,
A CORRELAÇÃO DE "RANK" DE SPEARMAN DEVE SER USADA
ELA TRABALHA COM "RANKS", OU SEJA A POSIÇÃO DE UM DADO,
NUMA LISTA ORDENADA, DESSA FORMA NÃO É NECESSARIO QUE
O FORMATO DO RELACIONAMENTO SEJA ASSUMIDO PREVIAMENTE.
A CORRELAÇÃO DE SPEARMAN É ADEQUADA POR EXEMPLO
PARA EXPERIMENTOS DE SATURAÇÃO. ONDE AO LONGO DAS
CONCENTRAÇÕES A CURVA MUDA DE FORMATO. (FIGURA ABAIXO)



ENTRE AS ANÁLISES MULTIVARIADAS UMA FERRAMEN-
TA QUE SE DESTACA É A ANÁLISE DOS COMPONENTES
PRINCIPAIS (PCA). QUANDO TEMOS DADOS ^{MULTIVARIADOS} ~~SIMPLES~~
O VETOR DE DADOS SIMPLES NÃO É SUFICIENTE E TEMOS UMA
REPRESENTAÇÃO EM VETORES DE COLUNAS, POR EXEMPLO NUM
EXPERIMENTO DE SÉRIE TEMPORAL DE TRANSCRITOS E PROTEÍ-
NAS CADA COLUNA TEM UM TEMPO DIFERENTE.

O PROBLEMA A SER ENFRENTADO PARA TENTAR ENCONTRAR
RELAÇÕES NESTE TIPO DE DADO MULTIVARIADO É COMPLEXO
É REFERIDO COMO MALDIÇÃO DA DIMENSIONALIDADE



NOME: DIEGO ENRY BARRATO GOMES.

No. fl.

8

OUDE TEMOS UM NÚMERO ALTO DE VARIÁVEIS E POSSIVELMENTE ALTAS CORRELAÇÕES ENTRE ELAS, OU SEJA, SÃO REDUNDANTES, * ALÉM DISSO MUITAS DELAS PODEM NÃO CONTER INFORMAÇÃO ÚTIL, SEGUINDO O EXEMPLO, mRNA e ou ~~PRO~~ PROTEÍNAS QUE NÃO TEM VALORES ALTERADOS OU DE BAIXÍSSIMA ALTERAÇÃO ENTRE OS DIFERENTES ESTADOS AMOSTRADOS. A PRESENÇA DESSAS VARIÁVEIS PODE RESULTAR EM SOBREENJUSTE (OVERFIT) DAS FUNÇÕES ~~SEJA~~, LEVANDO A CONCLUSÕES INCORRETAS.

A ANÁLISE DOS COMPONENTES PRINCIPAIS É UM MÉTODO LINEAR NÃO PARAMÉTRICO (SEM PARÂMETROS) BASEADO NA VARIÂNCIA DOS DADOS. ESTA ABORDAGEM VISA CRIAR UMA NOVA REPRESENTAÇÃO DOS DADOS EM UM NÚMERO REDUZIDO DE DIMENSÕES ENQUANTO MANTÉM A VARIÂNCIA. ASSIM, CADA DIMENSÃO REPRESENTA UMA PORCENTAGEM DO TODO, OU SEJA EXPLICA A VARIÂNCIA. OS EIXOS IDEALMENTE SÃO DESCORRELACIONADOS, COM ESPERANÇA DE PRODUTO NULO ENTRE ELAS.

A PCA PORTANTO PROJETA OS DADOS EM UM ESPAÇO DIMENSIONAL MENOR, FACILITANDO A VISUALIZAÇÃO BEM COMO PERMITINDO ANÁLISES SEGUINTE COMO ~~AS~~ OS AGRUPAMENTOS (CLUSTERING).

OS MÉTODOS DE AGRUPAMENTO COMBINAM INDIVÍDUOS COM CARACTERÍSTICAS SIMILARES EM GRUPOS DISCRETOS ATRAVÉS DE MEDIDAS DE DISTÂNCIA ENTRE OS INDIVÍDUOS. ~~AS DISTÂNCIAS~~ AS DISTÂNCIAS ^{DEPENDEM} DERIVAM DE MEDIDAS DE SIMILARIDADE

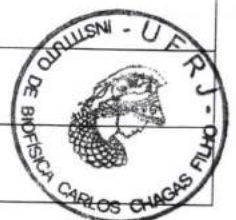


NOME: DIEGO ENRY BARRETO GOMES

No. fl.

9

A TAREFA PORTANTO É ENCONTRAR OU APRENDER PARÂMETROS CAPAZES DE DEFINIR A DISTRIBUIÇÃO DE PONTOS E PROBABILIDADES DOS INDIVÍDUOS PERTENCEREM A CADA UM DOS "CLUSTERS" BEM COMO ESTABELECEER O NÚMERO DE CLUSTERS ÓTIMO PARA DIVIDIR AS CLASSES, QUANDO ESTE NÚMERO NÃO É CONHECIDO. O "K-MEANS" É UMA DAS FERRAMENTAS MAIS USADAS PARA ESTE PROBLEMA, COM SUCESSO, ~~DE~~ O MÉTODO GERA SELECÇÕES ALEATORIAS DE PONTOS NA DIMENSÃO DE BUSCA E CALCULA A MEDIANA, USADA PARA CALCULAR A VARIÂNCIA ENTRE OS PONTOS DAQUELE GRUPO, ~~DE~~ O PROCESSO SE REPETE DE FORMA ITERATIVA VISANDO AJUSTAR O NÚMERO DE CLUSTERS QUE MELHOR REPRESENTA A VARIÂNCIA DOS INDIVÍDUOS SELECIONADOS.



09- FUNDAMENTOS TEÓRICOS DE ALGORITMOS DE BIOLOGIA COMPUTACIONAL E BIOINFORMATICA.

A PREVISÃO DAS INTERAÇÕES MOLECULARES ENTRE PARES DE OBJETOS BIOLÓGICOS COMO ~~DE~~ PROTEÍNAS, LIGANTES, METABOLITOS, ÁC NUCLEICOS ETC É UMA DAS TAREFAS ESSENCIAIS DA BIOINFORMATICA E BIOLOGIA COMPUTACIONAL. O RECONHECIMENTO MOLECULAR NO FINAL ~~DE~~ DAS CONTAS É O QUE VAI DETERMINAR A FORÇA DAS INTERAÇÕES BEM COMO PARTICIPAR DA DETERMINAÇÃO DA CINÉTICA DOS PROCESSOS E REAÇÕES QUE PODER SER POR EXEMPLO MODELADOS EM REDES BIOLÓGICAS. O RECONHECIMENTO E SUA PREVISÃO TEM SE DESTACADO NAS ÁREAS DE PREVISÃO DE INTERAÇÕES PROTEÍNA-PROTEÍNA, BEM COMO NA PREVISÃO DE COMPLEXOS RECEPTOR-LIGANTE COMUNMENTE APLICADO A FÁRMACOS E SEUS ALVOS. OU COMPLEXOS NANOCARREADORES DE FÁRMACOS (SISTEMAS HOST-GUEST) COMO CICLODEXTRINAS E CURCUBITURIL.

O DOCKING MOLECULAR É UM DOS PRINCIPAIS MÉTODOS DE ESCOLHA PARA REALIZAR TAIS PREDIÇÕES. O DOCKING É UM PROBLEMA MULTIVARIADO QUE ENVOLVE ~~DOIS~~ ^{DOIS} PRINCIPAIS DESAFIOS: RESOLVER UM PROBLEMA DE OTIMIZAÇÃO PARA A BUSCA DE CONFORMAÇÕES DOS COMPONENTES DO PROCESSO EX RECEPTOR E LIGANTE, E PONTUAR AS "POSES" DO DOCKING ADEQUADAMENTE, DE FORMA A ENCONTRAR A FORMA NATIVA DO COMPLEXO E QUE ESTA SEJA A MELHOR ~~RAIO~~ CLASSIFICAÇÃO (RANQUEADA) POR SUA FUNÇÃO DE PONTUAÇÃO



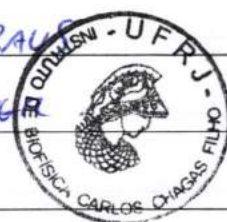
NOME: DIEGO ENRY BARRETO GOMES.

No. fl.

11

A TAREFA DE BUSCA ENVOLVE A DESCRIÇÃO DE TODOS OS GRAUS DE LIBERDADE DOS MEMBROS DO COMPLEXO. OS GRAUS DE LIBERDADE PODEM SER DIVIDIDOS EM INTRAMOLECULARES E INTERMOLECULARES. OS INTRAMOLECULARES SÃO OS ESTIRAMENTOS DAS LIGAÇÕES QUÍMICAS, ÂNGULOS ENTRE ÁTOMOS. GERALMENTE NA ETAPA DE BUSCA APENAS OS ÂNGULOS DE TORÇÃO SÃO CONSIDERADOS ROTÁVEIS, UMA ESTRATÉGIA INTELIGENTE DE REDUÇÃO DE DIMENSIONALIDADE JÁ QUE SÃO OS QUE MAIS IMPORTAM NESTA SITUAÇÃO DE BUSCA CONFORMACIONAL. ESSA REDUÇÃO É NECESSÁRIA PARA EVITAR "EXPLOSAO COMBINATORIA" QUANDO SÃO INCLUIDOS OUTROS GRAUS DE LIBERDADE. OUTRA ESTRATÉGIA QUANTO AS TORÇÕES É DISCRETIZAR O ESPAÇO DE BUSCA, LIMITANDO A VARIAÇÃO DOS ÂNGULOS A INTERVALOS PLAUSÍVEIS, DE ACORDO COM A CAPACIDADE COMPUTACIONAL DISPONÍVEL, EM OUTRAS PALAVRAS AS TORÇÕES PODEM "RODAR" EM ÂNGULOS DETERMINADOS, QUE ADICIONALMENTE, PORÉM NÃO FREQUENTEMENTE, PODEM SER LIMITADOS POR MÉTODOS HEURÍSTICOS, BASEADOS EM CONHECIMENTOS DE CONFORMAÇÕES BIOATIVAS, OU BASEADOS EM CAMPO DE FORÇA...

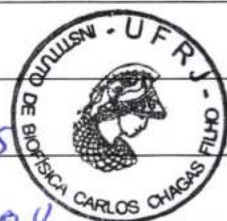
A SEGUNDA ETAPA DA BUSCA ENVOLVE OS GRAUS DE LIBERDADE DE TRANSLACIONAIS E ROTACIONAIS DAS MOLÉCULAS COMO UM TODO, ESTES PODEM SER DISCRETIZADOS. ~~COMO UM TODO~~ APENAS UMA DAS MOLÉCULAS PRECISA CONSIDERAR ESTES GRAUS DE LIBERDADE, POIS É SUFICIENTE PARA DESCREVER TODO O ESPAÇO DE BUSCA, O ESPAÇO DE BUSCA



É COMUMMENTE RESTRITO A UMA ÁREA DE INTERESSE, EX O SÍTIO DE LIGAÇÃO CONHECIDO, UM SÍTIO ALOSTÉRICO, UMA REGIÃO DE INTERFACE PROTEÍNA-PROTEÍNA, ESSA RESTRIÇÃO DEVE SER CUIDADOSA E BASEADA EM CONHECIMENTO, QUANDO BEM UTILIZADA ABRANGE TODA A ÁREA DE INTERESSE, EVITANDO DESPERDÍCIO DE CÁLCULO NUM ESPAÇO DE BUSCA ESPARSO, E REDUZINDO A CHANCE DE PRODUIR PREVISÕES INCORRETAS.

UMA VEZ DEFINIDO O ESPAÇO DE BUSCA (GRAUS DE LIBERDADE, SÍTIO DE INTERESSE DELIMITADO), MÉTODOS DE AMOSTRAGEM SÃO USADOS PARA EXPLORAR AS POSSÍVEIS CONFORMAÇÕES.

AS CARACTERÍSTICAS DO PROBLEMA MULTIVARIADO E COM TANTOS GRAUS DE LIBERDADE GERA UMA COMBINAÇÃO DE POSSÍVEIS ESTADOS QUE DIFICILMENTE PODE SER AMOSTRADA POR ABORDAGENS SISTEMÁTICAS. PORTANTO, MÉTODOS ESTOCÁSTICOS SÃO OS MAIS ADEQUADOS PARA A BUSCA DOS COMPLEXOS. DENTRE ELAS OS ALGORITMOS EVOLUTIVOS COMO O ALGORITMO GENÉTICO ~~LAMARCKIANO~~ SE DESTACAM. OS ALGORITMOS GENÉTICOS TRABALHAM COM POPULAÇÕES DE ESTRUTURAS ONDE AS CARACTERÍSTICAS GENÉTICAS ESTÃO CODIFICADAS COM OS GRAUS DE LIBERDADE DAS MOLÉCULAS (POSIÇÃO NO ESPAÇO, TORÇÕES, ROTAÇÃO) E SOFREM OPERAÇÕES GENÉTICAS COMO CRUZAMENTO E MUTAÇÕES. EVOLUINDO DÍVERSAS POPULAÇÕES, PONTUADAS PELA FUNÇÃO DE PONTUAÇÃO (SCORE), ATÉ UM CONJUNTO FINAL DE SOLUÇÕES, QUE COSTUMAM SER OTIMIZADAS LOCALMENTE POR UMA FUNÇÃO DE PONTUAÇÃO OU



BASEADA EM CAMPO DE FORÇAS.

A FUNÇÃO DE PONTUAÇÃO PORTANTO É DETERMINANTE PARA A ESCOLHA DE QUAIS CONFORMAÇÕES DOS COMPLEXOS DEVEM SER PRIORIZADAS ^{APÓS} ~~EM~~ CADA ETAPA DA BUSCA. AS FUNÇÕES DE PONTUAÇÃO SÃO DESENVOLVIDAS TENDO DOIS OBJETIVOS PRINCIPAIS EM MENTE: A PREVISÃO DA AFINIDADE E A CLASSIFICAÇÃO DOS COMPLEXOS. ~~CADA~~ IDEALMENTE UMA FUNÇÃO DEVE SER ACURADA E PRECISA PARA ENCONTRAR A CONFORMAÇÃO NATIVA DO COMPLEXO BEM COMO A AFINIDADE DE LIGAÇÃO.

AS FUNÇÕES DE PONTUAÇÃO PODEM SER DIVIDIDAS EM DUAS CLASSES PRINCIPAIS, AS BASEADAS EM FÍSICA, ~~GERAL~~ GERALMENTE FÍSICA CLÁSSICA, E AS BASEADAS EM CONHECIMENTO, INCLUINDO FUNÇÕES EMPÍRICAS E DE APRENDIZADO DE MÁQUINA. AS DUAS ÚLTIMAS ENVOLVEM O TREINAMENTO E TESTE ~~DE~~ ~~FORA~~ CONJUNTOS DE DADOS E FUNÇÕES OU MODELOS. FOCADOS NA PREVISÃO DAS INTERAÇÕES E/OU AFINIDADES.

EM COMUM ^{PODEM SER} ~~SE~~ USADOS DADOS EXPERIMENTAIS DE AFINIDADE DE LIGAÇÃO, COEFICIENTES DE PARTIÇÃO, DADOS ESTRUTURAIS DE COMPLEXOS CONHECIDOS, VARIAÇÕES DE ÁREA ACESSÍVEL AO SOLVENTE, NA CALIBRAÇÃO (TREINAMENTO) DOS MODELOS E FUNÇÕES DE ACORDO COM O OBJETIVO E RECURSO COMPUTACIONAL DISPONÍVEL, O ALGORITMO PODE PRIORIZAR POR EXEMPLO A CLASSIFICAÇÃO DE COMPOSTOS ATIVOS E INATIVOS, ESTE TIPO DE ABORDAGEM PODE FAVORECER ENSAIO DE TRIAGEM VIRTUAL DE COMPOSTOS



NOME: DIEGO ENRY BARRATO GOMES

No. fl.

14

ONDE A VELOCIDADE DO CÁLCULO, PARA VASCUJAR MILHÕES DE COMPOSTOS E EXTRAIR OS MAIS INTERESSANTES É A PRIORIDADE. NESTE EXEMPLO O PESQUISADOR PROGREDI O ESTUDO PARA UMA ETAPA SUBSEQUENTE USANDO UMA FUNÇÃO MAIS ROBUSTA QUE ENTÃO CLASSIFICA COMPOSTOS COM FOCO NA CORRELAÇÃO COM A ATIVIDADE EXPERIMENTAL, O TREINAMENTO DAS FUNÇÕES POR SI SÓ TAMBÉM É UM PROBLEMA ~~DE~~ MULTIVARIADO, QUE PODE ENVOLVER A SELEÇÃO DE CARACTERÍSTICAS, LEVANDO EM CONTA AS SUAS RELEVÂNCIAS PARA EXPLICAR O MODELO ~~COMPLEXO~~ EX A CONFORMAÇÃO DO COMPLEXO OU AFINIDADE.

