Fichamento

Os autores propuseram o uso de HMMs para a identificação de genes com funções similares em dados de expressão gênica (*Microarrays*). Com esta finalidade, eles partiram do princípio que genes com funções semelhantes compartilham, em condições iguais, padrões de expressão parecidos.

Os autores notaram que teriam problemas similares aos encontrados na utilização de HMMs para a detecção de homologia em proteínas. Então optaram por utilizar o Profile HMM, que é uma série lienar de estados que se transformam sequencialmente partindo do estado anterior.

O algorítimo proposto trabalha bancos de dados como o KEGG (Kyoto Enciclopedia of Gene and Genomes), para vias metabólicas e GO (Gene Onthology), para genes anotados.

Primeiramente é feita uma seleção de genes com um coeficiente de correlação entre genes (pairwise) acima de um threshold pré-delimitado, isto filtra genes que a relação não foi bem demonstrada no *microarray* selecionado; logo após é utilizado um novo algorítimo para selecionar um subgrupo relevante de experimentos da matriz do microarray, que tenha os genes que melhor representem este array; após esta etapa treinou-se o HMM com o subgrupo selecionado e com grupos randômicos de mesmo tamanho do que os do *microarray* de interesse. Todos os genes do *microarray* de interresse passam pelo HMM e recebem um score. Os scores dos genes selecionados randomicamente são utilizados para estimar a Parzen Density Function (PDF).

Como resultado, os autores chegam a conclusão que seu algoritmo trabalha melhor ,com dados complicados de *Microarray,* que o Signature Algorithm e que comparações baseadas em correlações, apenas.