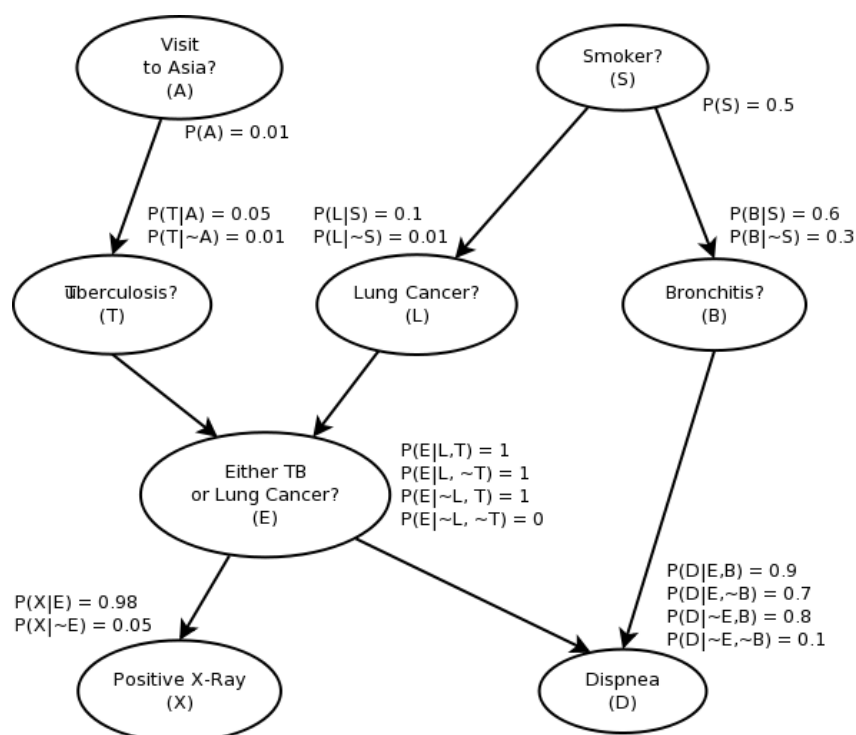


Projeto – Redes Bayesianas
Data de entrega: 30/06



Este projeto é individual. A entrega deve ser realizada via TIDIA, fazendo *upload* dos arquivos na pasta “projeto 1” na ferramenta escaninho.

O objetivo deste projeto é implementar uma rede bayesiana simples e avaliar dois mecanismos de inferência, um exato e outro aproximado. Os resultados dos dois mecanismos devem ser comparados.

O domínio de aplicação deste projeto é o diagnóstico médico, mas especificamente de tuberculose, câncer de pulmão ou bronquite. A rede bayesiana da figura acima foi proposta por Lauritzen e Spiegelhalter [1], e foi descrita da seguinte maneira:

“Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a

single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea."

Este projeto possui duas partes descritas a seguir:

Parte I – Implementação da rede bayesiana (50% da nota final)

A sua implementação da rede bayesiana deve representar a rede e as tabelas de probabilidades condicionais. Você é livre para decidir qual será a representação da rede, e as suas decisões de projeto devem ser apresentadas no relatório que acompanha o projeto.

Você deve implementar um método de inferência exato e um método de inferência aproximado (amostragem). Para o método exato, utilize a inferência por enumeração ou o algoritmo de eliminação de variáveis. Se fizer a opção pelo algoritmo de eliminação de variáveis, inclua no relatório uma discussão sobre a ordem de eliminação das variáveis ocultas. O algoritmo aproximado deve ser o de peso por verossimilhança (*likelyhood weighting*).

Tanto os métodos de inferência quanto a representação da rede podem ser específicos para esse domínio, ou seja, não precisam ser genéricos para qualquer domínio de aplicação ou consulta.

Em particular, estamos interessados nas seguintes consultas:

- $P(T)$: probabilidade de ter tuberculose

- $P(X, +a, +s)$: probabilidade de ter resultado positivo no raio-X dado uma visita recente à Ásia e ser fumante

- $P(S, +x, +d)$: probabilidade de ser fumante dado resultado positivo no raio-X e ter dispneia.

Parte II – Relatório (50% da nota final)

Escreva um relatório curto de 2 a 3 páginas explicando as suas decisões de projeto. Em particular, descreva quão bem o método aproximado é capaz de estimar as probabilidades da resposta à consulta comparado com o método exato. Faça uma análise em função do número de amostras.

Referências

[1] Lauritzen S, Spiegelhalter D (1988). "Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion)". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **50**(2), 157-224.