

Vigilância de doenças usando um modelo Markov HMM

1. [Abstração da problemática](#)
2. [Metodologia HMM \(Hidden Markov Model\)](#)
3. [Avaliação e testes de performance do Modelo](#)
4. [Avaliação da percentagem de falsos alarmes](#)
5. [Motivação e Conclusão](#)
6. [Referência](#)

Abstração da problemática

Efetuar uma boa detecção com capacidade de análise e reconhecimento de sinais iniciais de possíveis surtos de doenças locais, utilizando dados proveniente da vigilância regular da recolha de notificações de doenças.

Com o exponencial aumento do uso de tecnologias de vigilância em saúde, bem estar e em sua vez o enorme volume de dados eletrônicos gerados por tais sistemas, cada vez mais se é apercebido do potencial benefício em automatizar e melhorar os métodos de tratamento de dados, levando assim a uma melhor e mais rápida detecção de casos anormais no tempo e em certo espaço. Estes métodos espaço-temporais, exibem uma melhor habilidade de detecção para pequenas regiões e localidades.

Este artigo visa demonstrar uma implementação simples e genérica de um HMM(Hidden Markov model) para uma análise e vigilância de uma pequena área.

Metodologia HMM (Hidden Markov Model)

O método HMM utilizado foi baseado em um HMM temporal já existente que foi posteriormente modificado para incorporar referência de dados espaciais, tendo estas áreas espaciais sido separados pela área do código postal, permitindo

identificar zonas específicas e as suas zonas vizinhas permitindo uma interação da ilustração de uma doença em uma zona com as suas zonas vizinhas.

Este modelo requer um alcance válido de vizinhos propriamente identificado pelos seus códigos postais, tendo então sido alocadas as populações em ilhas no Leste da Austrália como vizinhas do código postal da ilha principal sendo que as suas populações são quase inexistentes. São então registados e representados os habituais movimentos entre as zonas.

Para simplicidade foi utilizado um modelo HMM de primeira ordem onde são representados dois estados: um estado endêmico (sem surto de doença) e um estado epidêmico (com surto de doença). Estados que são suportados por pesquisas anteriores.

O modelo utilizado denomina $x[t,i]$ como o somatório de casos de doença em cada zona $y[t,i]$ em cada área de vizinhos $y_n[t,i]$, sendo t o número do estágio em dias ex: $t = 1, \dots, t$ (dias) e i a área ex: $i = 1, \dots, n$, para os dois estados usando um modelo de Poisson. Este modelo de Poisson implica que a variância de $x[t,i]$ é igual a sua média esperada. Estudando dados históricos da evolução de Hepatite A, verificou-se uma grande dispersão e flutuação para $x[t,i]$ sendo um problema na automatização do sistema. Para tal resolução introduziu-se uma distribuição a que os dados deveriam obedecer antes de ser efetuada a média para cada estado, sendo assim descrita essa variância que anteriormente não era exprimida por Poisson.

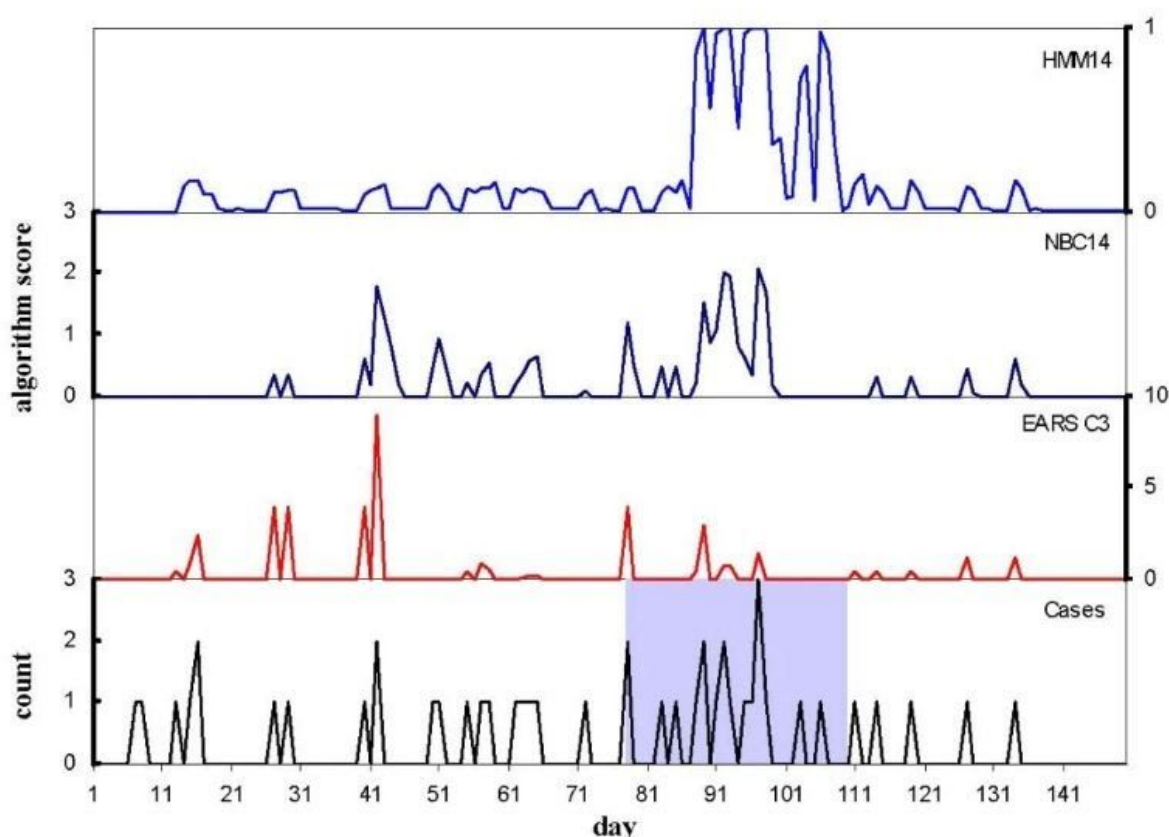
Movimento entre os dois estados é governado por uma matriz de transição estacionária que determina as probabilidades de transição. Nas probabilidades de transição inicial foi utilizada uma distribuição Dirichlet sem informação, aplicando uma distribuição beta generalizada cujo K assume valores entre 0 e 1. Nas transições subsequentes foi utilizada uma distribuição equivalente a Dirichlet. Foram usadas distribuições cujas médias para os dois estados descrevem também uma pequena probabilidade de casos mais extremos.

Com o objetivo de desenvolver uma HMM simples e de baixa manutenção, não foram incluídos no modelo variáveis, assumindo que a população em risco continua estável ao longo do tempo. Para limitar os recursos computacionais o HMM avalia somente 7 pontos de dados temporais variáveis em 1-7 dias de análise até aos dias 144-150.

O modelo foi testado em MCMC (Markov Chain Monte Carlo) tendo sido corridas 1000 iterações sem monitorização de parâmetros e posteriormente 1000 com monitorização chegando a uma análise do tempo e hardware necessário para produzir resultados estáveis de uma análise dos resultados diários para as 383 áreas.

Avaliação e testes de performance do Modelo

O modelo foi avaliado com dados semi-sintéticos simulados com base em dados históricos do surto de Hepatite A. Como comparação foram usados os algoritmos existentes EARS C1,C2 e C3 para estabelecer uma base padrão. Tendo sido efectuadas várias simulações e utilizada a média da percentagem de detecção.



Notificação dos casos de Hepatite A tendo o surto ocorrido na área sombreada e analisada pelos algoritmos EARS C3, Negative binomial 14 dias e o discutido HMM, sendo o seus scores o desempenho de cada algoritmo em detectar o surto.

Avaliação da percentagem de falsos alarmes

Um modelo de detecção que põem em risco a saúde pública tem de obedecer a estritos padrões de viabilidade, sendo um deles a sensibilidade do algoritmo em apresentar um falso alarme de surto, para tal foram seleccionados aleatoriamente dois cenários, um pequeno e outro maior para quais foram executadas várias simulações para este e vários outros algoritmos.

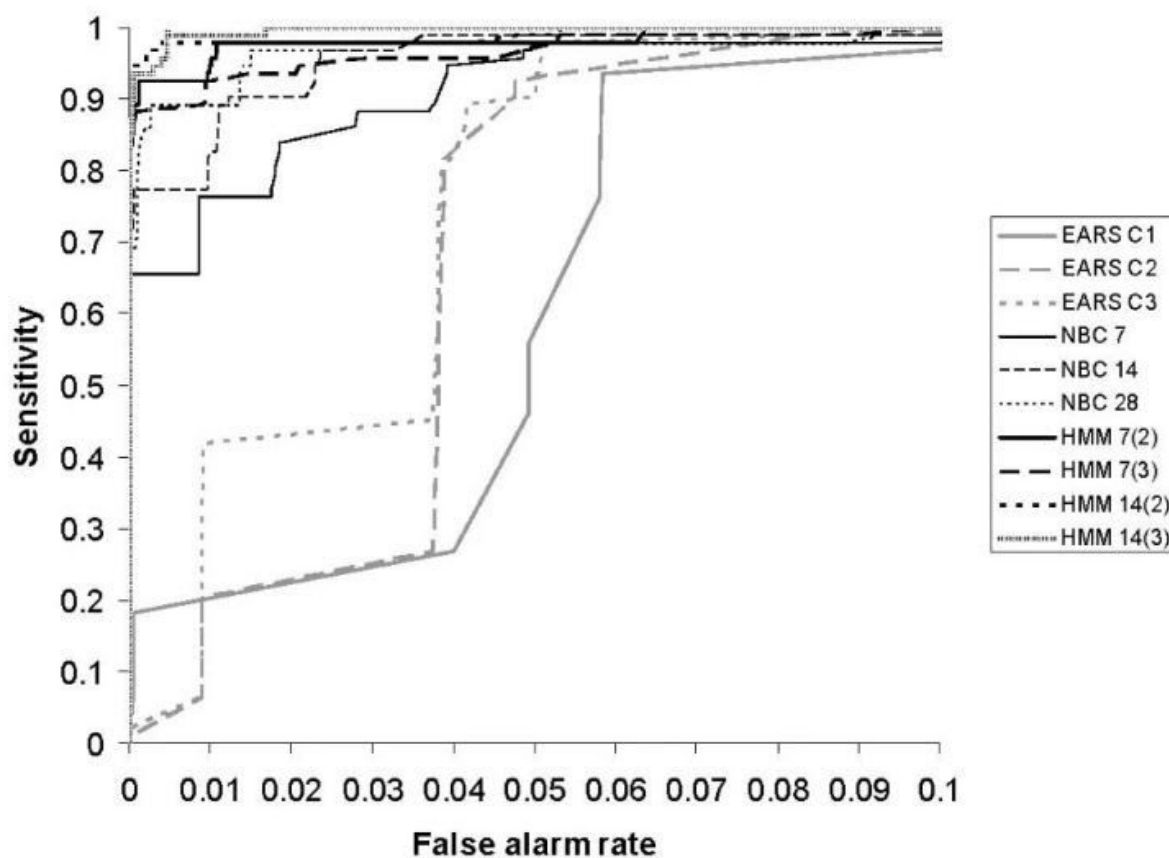


Figura acima demonstra as várias percentagens de falsos alarmes para diferentes valores de sensibilidade para cada algoritmo. Tendo HMM para 28 dias sido desprezado por não diferenciar de 14 dias

Motivação e Conclusão

A escolha deste artigo vai em direção aos eventos em que vivemos atualmente, avaliando de um ponto de vista pertinente à nossa formação e procurando responder a questões da importância de modelos e algoritmos eficazes em detecção e previsão de surtos de doenças com baixa percentagem de falsos positivos. O avanço e implementação de algoritmos para tratamento de grandes quantidades de dados é um tema atual proveniente da crescente ligação entre a vida humana e várias tecnologias, sendo especialmente importante o tratamento de dados de saúde para prevenção e melhor adaptabilidade em situações futuras.

O modelo HMM utilizado supera os métodos temporais previamente estabelecidos, utilizando uma metodologia espaço-temporal particularmente baixa em falsos alarmes.

Referência

- Watkins, R.E., Eagleson, S., Veenendaal, B. et al. Disease surveillance using a hidden Markov model. BMC Med Inform Decis Mak 9, 39 (2009).
<https://doi.org/10.1186/1472-6947-9-39>