INTERNATIONAL JOURNAL OF ADVANCE RESEARCH AND DEVELOPMENT

IJARnD

***(Volume 4, Issue 3)***

Available online at: www.ijarnd.com

# Effective performance of hidden markov model for epidemiologic surveillance

**Hla Hla Myint**

*Student, Universities of Computer Studies, Magway, Myanmar*

**ABSTRACT**

***The public health surveillance system is one of the most important for the detection of the seasonal influenza epidemic. We introduced different kinds of surveillance data for early detection of a disease outbreak. Hidden Markov model has been recognized as an appropriate method to model disease surveillance data. In this work, we proposed a hidden Markov model (HMM) to characterize epidemic and non-epidemic dynamic in a time series of influenza-like illness incidence rates and presents a method of influenza detection in an epidemic. ILI is defined as an illness marked by the presence of a fever (100.5°F or greater) and either a cough or sore throat within 72 hours of ILI symptom onset, or physician-diagnosed ILI. ILI incidence rate is based on surveillance data and activity state. HMMs have been used in many areas, including automatic speech recognition, electrocardiographic signal analysis, the modelling of neuron firing and meteorology. A two-state HMM is applied on incidence time series assuming that those observations are generated from a mixture of Gaussian distribution. Bayesian inference method is calculated to obtain the probability of an epidemic state and non-epidemic state every week. The various influenza dataset applied the methodology.***

***Keywords****—Hidden Markov model, Influenza, Epidemics*

## 1. INTRODUCTION

Influenza is an acute respiratory illness that spreads quickly from person to person and occurs every year. Global influenza surveillance has been conducted through WHO's Global Influenza Surveillance and Response System (GISRS) since 1952. [6]. Our climate changes cause influenza inflections. The epidemiological surveillance of the influenza activity is supported by symptoms of ILI patients. ILI is defined as an illness marked by the presence of a fever (100.5°F or greater) and either a cough or sore throat. Influenza-Like Illness data collected on weekly basic and presented in the form of the epidemiologic indicator at regular time intervals (week, month). The prospective detection of the beginning of the epidemic period has been done by a variety of statistical methods such as regression technique, time series methods, a method of statistical process control and also on statically multivariate analysis using multiple data source [1].

Timeliness of a public health surveillance system is one of its most important characteristics given that is a crucial for the capacity of a timely invention and timeliness will be considered as the time elapsed from the disease onset to the generation of an automated alert. [1]. For detection of influenza epidemics, we suggested HMMs for disease surveillance, using a mixture of Gaussian Distribution to model ILI incidence and demonstrated the detection accuracy using collected data and two measures of interest: the ILI incidence rate and influenza activity state.

HMMs which used the hidden variable to eliminate the need for explicit modelling of trend and seasonal effects that can introduce detection bias were found to produce fewer false alarms and be more robust to variations in the data [2]. We aim to evaluate the performance of a Bayesian HMM, that analyses the detection of epidemic state requires little baseline data for modifiable inflections disease surveillance. HMM will provide a natural way of modelling epidemic and non- epidemic periods to assign a different probability to these two states.

## 2.RELATED WORK

The epidemiological surveillance system is one of the challenging objectives of the early detection of outbreak disease. We introduce a framework of a model for the early detection of the onset of an influenza epidemic. This framework is based on a HMMs. HMMs have been used in many areas, including automatic speech recognition, electrocardiographic signal analysis, the modelling of neuron firing and meteorology. To determine the timing of epidemic periods, ILI incidence rate will be used to two state HMMs (epidemic and non-epidemic rate).

The HMMs were first applied to ILI surveillance data in 1999 by Le Strat and Carrat [4]. Two state HMMs (epidemic and non-epidemic) were described using the weekly ILI rate of each state. This paper has used a Bayesian inference for detection of epidemic state. We need a model that enables the use of covariates with forecasting capacity of incomplete data collected by influenza surveillance system these data are now casting the current weekly ILI rate and the respective influenza activity state. We proposed the use of an HMM that allows the covariate to model the weekly ILI rate and state transition probability, where transition probability from the epidemic to non-epidemic and non-epidemic to the epidemic.

## 3.BACKGROUND

Routine surveillance of disease notification data can enable the early detection of localized disease outbreaks [3]. Using novel surveillance methods and historical data to estimate future trends of ILI can lead to early detection of influenza activity increase and decreases [5]. Therefore public's health professionals give more time to prepare and increase prevention efforts by anticipating surges. HMMs were applied to ILI surveillance data and this model has been recognized as an appropriate method to disease surveillance. The potential benefits of applying automated motoring method to population health data being increasing realized, particularly data that are routinely collected by population health surveillance system. Although HMMs have been proposed for disease surveillance and suitable for use small area count data. In this model, we demonstrated the influenza activity state and correspondent ILI rate of each week, from the week (40/2017) to week (39/2018) in Texas ILI Net data and nowcasted the previous week of ILI using HMM. For example, week 40/2017 was now casted from 40/2016 to week 39/2017, week 41/2017 was nowcasted from 40/2016 to week 40/2017, and so on.

Now, we applied to the entire time series, from week 40/2017 to week 39/2018, parameters estimates of each week in one of the markov chain state (epidemic or non-epidemic).

## 4.METHOD

Data were obtained in Texas ILINet from week 40/2017 to week 39/2018, for analysis ILI activity in every week and then we will be forecasted on ILI activity for subsequent 20 weeks. We developed a simple Bayesian HMM for the surveillance of weekly reported case of ILI in Texas. HMMs provide a natural way of modelling epidemic and non-epidemic periods to assign a different probability distribution to ILI activity state. We described the ILI detection system by following the system flow diagram figure 1.

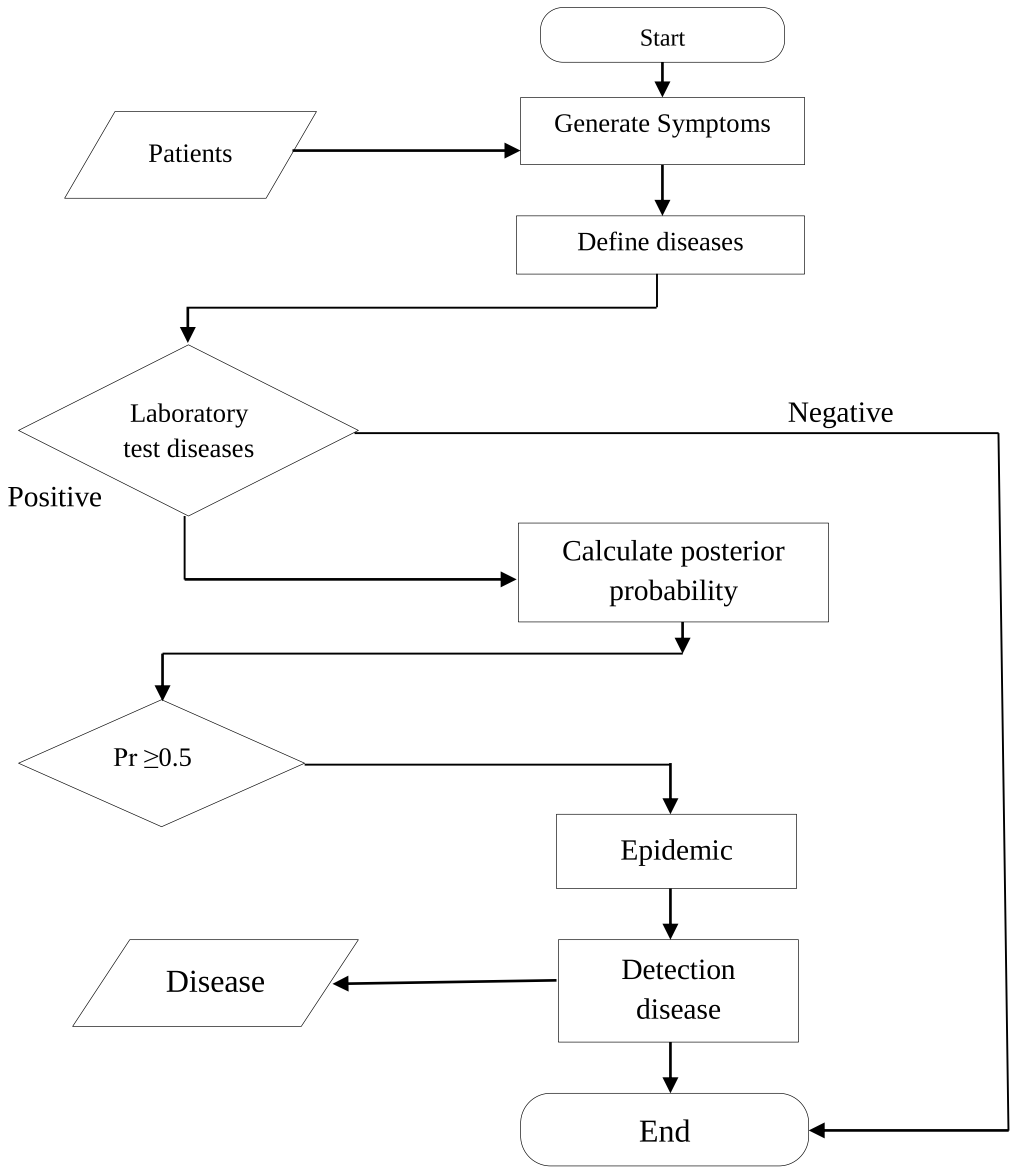


**Fig. 1: System flow diagram of ILI detection**

## 5.INFLUENZA-LIKE ILLNESS

Influenza-Like Illness data were obtained from Texas ILI Net providers. ILI is defined as an illness marked by the presence of a fever (100.5°F or greater) and either a cough or sore throat within 72 hours of ILI symptom onset, or physician-diagnosed ILI [2]. We have considered the data collected between week 40/2017 to week 39/2018. One of these data we describe the ILI incidence week 40/2017 to week 07/2018 data as figure 2. In addition, we described the ILI rate of 10 weeks as a 95% confidence interval (table1). ILI rate is found most of weeks 49 in table 1.

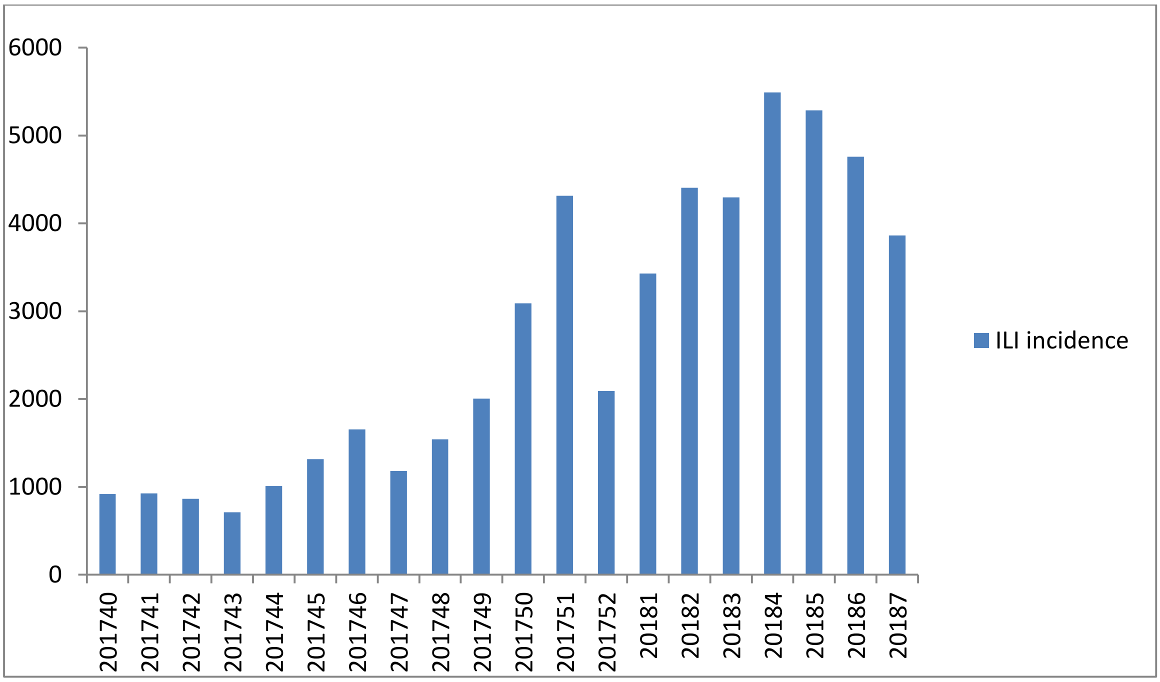


**Fig. 2: ILI incidence of Texas ILI Net from week40/2017 to week07/2018**

**Table 1: 95% confidence interval in week 40/2017 to week 49/2017 in Texas ILI**

| Week | Total ILI patient | Total patients | 95% (CI) |
|---|---|---|---|
| 201740 | 918 | 31871 | 2.7-3.06 |
| 201741 | 928 | 31433 | 2.76-3.14 |
| 201742 | 864 | 31527 | 2.56-2.92 |
| 201743 | 710 | 30309 | 2.17-2.51 |
| 201744 | 1012 | 25857 | 3.68-4.15 |
| 201745 | 1317 | 32080 | 3.89-4.32 |
| 201746 | 1656 | 32907 | 4.8-5.27 |
| 201747 | 1180 | 23582 | 4.73-5.28 |
| 201748 | 1541 | 29259 | 50.1-55.2 |
| 201749 | 2005 | 30858 | 62.7-67.7 |

## 6.HIDDEN MARKOV MODEL

HMMs is a statistical model that was first proposed by Baum L.E.( Baum and Prtried, 1956 ) and use a markov process. HMM can be represented as the simplest dynamic Bayesian network. In sampler Markov model (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probability is the only parameter, while the HMM, the state is not directly visible, but the output, the dependent on the state is visible. HMMs have been used in many surveillance data because HMM can automatically adjust the seasonal, covariant and distribution elements. HMM can be considered a generalization of a mixture model where the hidden variables, which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other.

Generally, an HMM assumes that the observed time series, $y_t$ with t=1, …T, is a realization of a stochastic process {$Y_t$ : t=1, …T}, where the distribution of each $Y_t$ is conditionally determined by an unobserved discrete stochastic process {$S_t$ : t=1, …T}, that assume values in an m-states set {1,2,…m}[2]. For the homogenous case, this unobserved stochastic process is assumed to be an order one Markov chain with stationary transition probability given by $P_{ij}$ =P{ $s_t$=j\ $s_{t-1}$=i},i,j ∈{0,1}. We proposed the application of an HMM stated transaction probability is given by time-dependent matrix $A^t$ with elements $P^t_{ij}$ =P{$s_t$ =j\ $s_{t-1}$=i}, i,j ∈{0,1} and t=2,…T , where 0 and 1 represent, respectively, the epidemic and non-epidemic state of influenza activity.

Formally:

$$A^t = \begin{bmatrix} P^t_{0,0} & P^t_{0,1} \\ P^t_{1,0} & P^t_{1,1} \end{bmatrix}$$

The sequence of the state is unobserved, this Markov- dependent mixture model is called HMM. In this work, we proposed two states HMM, epidemic and non-epidemic state, where each of these states described the weekly ILI rates by a normally distributed.

 

## 7. RESULT

Our objective was to detection of epidemic state. To detection, first, we study and validate prediction estimates of ILI activity and ILI incidence rates. From our observation, the influenza epidemic is sustained when the number of ILI cases tested positive for influenza in a certain week is high, e.g 25. On the other hand, if the number of ILI case confirmed for influenza is zero, this state is non- epidemics or influenza activity state is cleared. Based on this, if an increase in the ILI rate is observed without ILI case confirmed for influenza, one cannot assume that this increase is due to an influenza epidemic, but can be related to the circulation of other respiratory viruses. These two covariates that are functions of the early estimate of the ILI rate and the number of ILI cases positive for influenza in the previous week are proposed. The objective of these two covariates is to discriminate either epidemic or non- epidemic period: one covariate to model the response variable in the non- epidemic period,

$$y_{t(t)0} = \begin{cases} y_{t(t)} & y_{t(t)} \leq b_0 \\ 0 & Otherwise \end{cases}$$

And the other to model the response variable in the epidemic period,

$$y_{t(t)1} = \begin{cases} y_{t(t)} & y_{t-1(t)} \leq b_1 \\ 0 & Otherwise \end{cases}$$

In this case, $b_0$ =25 and $b_1$=1

As can be seen, both variables share a common part of the early estimate of the ILI rate $y_{t(t)}$.

For the HMM chain state transition probability, a time-varying matrix will element $P^t_{ij}$ was considered. For a specific week t, $P^t_{0,1}$ and $P^t_{1,1}$ represent the probability that in week t the influenza activity is epidemic given that in the week before the influenza activity was respectively in the non- epidemic state or epidemic state. For example, A week is more likely to belong to the epidemic state if the previous week was in epidemic state ( $P_{1,1}$ posterior mean 0.85) similarly a week is more likely to non-epidemic state if the previous week was in non-epidemic state ( $P_{0,0}$ posterior mean 0.79).This results also means that, ($P_{0,1}$ =0.21 and $P_{1,0}$ =0.15). We expected the posterior probability of chaining in influenza activity state in each week vary over time. The detection of the epidemic was considered on the mean posterior probability of each week of the epidemic and non-epidemic states. For example, if posterior probability is higher or equal to 0.5 in a week, this period was defined as an epidemic state. These results forecasted to detect in the real-time epidemic state.

## 8. CONCLUSION

We have proposed a new model for detection of influenza epidemics in time series of ILI incidence rates using HMMs. HMM was applied to the analysis of ILI incidence rate and then it's differentiated clearly between epidemic and non-epidemic rates. Two states HMM, is applied on incidence time series assuming that those observations are generated from a mixture of Gaussian distribution. We had been described as a result of epidemic and non-epidemic for detection on time series. And then, the system will be continued to detect the epidemic and non- epidemic state using Texas ILINET data 2017-2018. Finally, the exact value of the ILI epidemic in Texas ILINET data using HMM will be calculated. After, Texas ILINET data will analyze for detection continuously.

## 9. REFERENCES


[1] Baltazar Nunes, Isabel Nat´ario, CEAUL, M. Luc´ılia Carvalho, CEAUL, "Nowcasting influenza epidemics using non-homogenous hidden Markov models", Departamento de Epidemiologia, Instituto Nacional de Sa´ude Dr.RicardoJorge, Av. Padre Cruz, 1649-016 Lisboa, Portugal. Telephone: +351217526490 Email:baltazar.nunes@insa.min-saude.p August 17, 2011

[2] Lisa A.Shoubaki, MPH, MSMR Vol.25 No.6-June 2018-Health.mil

[3] RE Watkins, "Disease Surveillance using a Hidden Markov Model", CBI-NIH, 2009

[4] S trat L, Carrat F. Monitoring epidemiologic surveillance data using HMM chains model, Statist.Med.18, 3463-3478(1999)

[5] Susannah Paul, Osaro Mgbere, Raouf Arafat, Biru Yang and Eunice Santos, "Modeling and Forecasting Influenza-like Illness (ILI) in Houston, Texas Using Three Surveillance Data Capture Mechanisms", Online J Public Health Inform.9(2); 2017PMC5630275

[6] https://www.who.int/influenza/gisrs_laboratory/en/