Earthquake Forecasting Using Hidden Markov Models

Table of Contents

Introduction

Earthquake is one of the severe catastrophic natural hazards that must be evaluated carefully in engineering project due to the seriously damaging effects on human-made structures and human life. Therefore, the identification of seismicity patterns is an important task to evaluate the risk of damage, especially in some earthquake-prone areas. In this project, the main objective is to determine which Hidden Markov Models are the best model to predict earthquakes. Besides, our group would compare the models with different hidden states and check the accuracy of the parameter learning. Furthermore, our group would try to interpret the meaning behind different hidden states and some possible application of this model.

Data and Data Pre-processing

In this project, the dataset was chosen from *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*. The original data was collected from the U.S. Geological Survey, and it was downloaded on 25 July 2007. The dataset contains the yearly number of major earthquakes (magnitude 7 or greater) from 1900 to 2006. The dataset shows its high completeness because of no missing value.

To properly evaluate the different HMM models, the complete dataset was split into a training set and a test set. The yearly number of major earthquakes from 1900 to 2005 will be treated as a training set and the rest will be considered as test set. Figure 1 shows the actual dataset of the yearly counts and the frequency of number of major earthquakes from 1900 to 2005.

To better understand the target datasets, we performed an exploratory data analysis. From Table 1, it is shown that the sample variance, ($S^2 = 51.392$) is much larger than the sample mean, ($\bar{x} = 19.44$) that indicates a strong over-dispersion relative to Poisson distribution. Besides, the distribution of the observed data is skewed to the right. The over-dispersion and skewness of the data show that HMM is a suitable tool for the earthquake data. Therefore, our group determined to use a mixture Poisson model to deal with over-dispersed observation with a multimodal distribution.

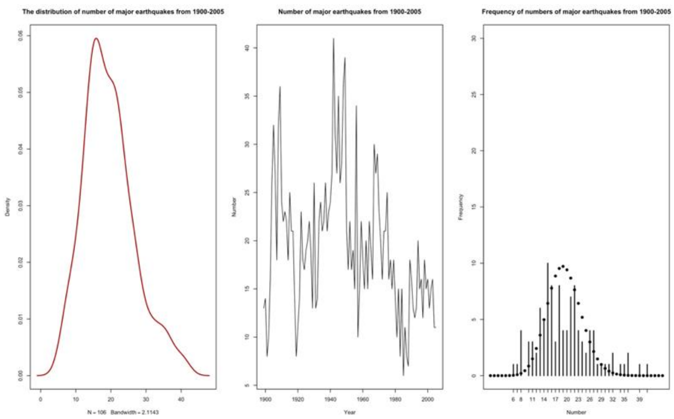

Figure 1: The distribution, yearly counts, and the frequency of numbers of major earthquakes

Variable	Sample mean	Sample Variance	Skewness
No. of earthquakes	19.44	51.392	0.6717067

Table 1: Descriptive statistics of yearly counts of numbers of major earthquakes

Methodology

- **Poisson Hidden Markov Model**

In this project, our group would apply the Poisson Hidden Markov Model which is a discrete-time bivariate stochastic process (X, Y) =(X_t,Y_t). X is a finite-state homogeneous Markov chain and X_t is the state of the underlying system at time t. Y is an observable process that depends on probabilistically Markov chain X which can be discrete or continuous. Each observation is derived from one of the N Poisson distributions. The rate parameters are called hidden states of finite-state homogeneous Markov Chain X. According to the state-dependent probability distribution, the formula of the conditional distribution of Y_t given X_t is as follows:

$$P(Y_i = y|X_i = \lambda_i) = e^{-\lambda_i \frac{(\lambda_i)}{y!}}, if\ y = 0,1,2 \ldots$$

- **Forward and Backward Probability**

Forward and backward probabilities are the tools for EM algorithm, decoding, and state prediction (Walter, Iain, Roland, 2016). Forward probability is a joint probability while backward probability is conditional probability. By multiplying them together, they could be applied in the EM algorithm, or the global decoding, Viterbi algorithm because of the conditional independence.

$$\alpha_t(i)\beta_t(i) = \Pr\left(\boldsymbol{X}^{(T)} = x^{(T)}, C_t = i\right)$$

- **EM algorithm**

The EM algorithm is iterative, and it is commonly used in finding the maximum likelihood estimation when having missing data. It could be separated into the Expectation step (E-step) and the Maximizing step (M-step). For E-step, the Q-function is the conditional expectation that integrates the missing data in the complete-data log-likelihood function. It is computed by using the conditional distribution of missing data given the observed data and the current parameters. For M-step, the Q-function is maximized with respect to parameters from the current iteration. After choosing the initial parameter values of the first iteration, the above steps are repeated iteratively until the convergence criterion is achieved. The estimated parameter value of the last iteration is then a stationary point of the likelihood of the observed data. (Pierre, 2013)

Since the sequence of states is not observed, it is difficult to obtain the exact maximum likelihood estimation directly. Therefore, the sequence of states is treated as missing data and the EM algorithm to the HMM model is applied. (Bilmes, 1998) The random variables of the sequence of states will be replaced by their conditional expectations given the observation, which are calculated by the forward and backward probabilities. After substitution of the conditional expectation, it can maximize the complete-data log-likelihood function, with respect to parameters, such as initial distribution, transition probability matrix, and state-dependent distribution (lambda).

- **Model Selection**

Model selection is the process of selecting one final machine learning model among all machine learning models for a training dataset. Akaike information criterion (AIC) and Bayesian information criterion (BIC) are two of the most common methods of model selection. AIC weighs the ability of the models to predict the observed data against the number of parameters that are required to reach the level of precision. The model selection criterion simplifies to the Akaike information criterion: $AIC = -2 \log L + 2p$[1].

The aim of the Bayesian approach to model selection is to select a family which is estimated to be most likely true. The formula of Bayesian information criterion: $BIC = -2 \log L + p \log T$, where T is the number of

[1] p denotes the number of parameters of model and log L is the log-likelihood of the fitted model.

observations.

- **Residual Diagnostics**

Before using the model to forecast the number of earthquakes for the following year, we have to check the pseudo-residuals to determine whether a model has adequately captured the information in the data. Model adequacy can be checked by testing the pseudo-residuals for normality and its sample ACF. If there were substantial autocorrelation in the residuals, that would suggest the independent error model was incorrect. (Walter, Iain & Roland, 2016)

The pseudo-residuals for discrete observations (Poisson, binomial, etc.) are specified as an interval (or segment) rather than a value. The normal version of pseudo-residuals has the advantage that the absolute value of the residual increases with increasing deviation from the median. Besides, it can be used to identify the outlier more easily on a normal scale. In this project, the discrete mid-pseudo-residuals would be used for model validation. The formula for ordinary pseudo-residuals (Normal pseudo-residual segments) is shown below:

$$[z_t^-;\ z_t^+] = [\Phi^{-1}(u_t^-), \Phi^{-1}(u_t^+)]$$

- **Global Decoding- Viterbi Algorithm**

Global decoding is the main objective in the application that determines the most probable states at every time point. The Viterbi Algorithm, which is an efficient dynamic programming algorithm of global decoding, is used to determine the most likely sequence of states. (Viterbi, 1967) With the decoding process, we could have better and more substantive interpretations for the states and investigate the performance of the fitted model.

$$\epsilon_{t,i} = \begin{cases} P(C_1 = i, X_1 = x_1) = \delta_i p_i(x_1)\ for\ t = 1\ and\ i = 1, \dots, m \\ \max_j\{\epsilon_{t-1,j}\gamma_{ji}\}\, p_i(x_t)\ \ for\ t = 2, \dots, T\ and\ i, j = 1, \dots, m \end{cases} \quad (1)$$

$$where\ T\ is\ the\ no.of\ observation, m\ is\ the\ no.of\ states, \delta\ is\ initial\ probability,$$

$$\gamma\ is\ transition\ matrix\ and\ p(x) is\ the\ pdf\ of\ x$$

$$i_t = \begin{cases} argmax_i\ \epsilon_{T,i}\ for\ t = T\ and\ i = 1, \dots, m \\ argmax_i\ \{\epsilon_{t,i}\gamma_{i,i_{t+1}}\}\ for\ t = T-1, \dots, 1\ and\ i = 1, \dots, m \end{cases} \quad (2)$$

The procedure of determination of the most likely states is initialized with the calculation of the joint probability of the first observations(t=1) and each possible state (i), denotes as $\epsilon_{1,i}$. Starting from the second observation, the joint probability $\epsilon_{t,i}$ is evaluated by the equations (1). After computing the $T \times m$ matrix of values $\epsilon_{t,i}$, the most likely states i_t could be determined recursively and backwardly from equation (2).

Results & Analysis

- **Model Selection**

Since it seems that the observed data followed a Poisson distribution, the Poisson Hidden Markov Models were performed. In this project, four different Poisson Hidden Markov models were built, including one-state, two-state, three-state, and four-state models. The parameters for Poisson will be larger if a higher state is assumed to have a higher frequency level of earthquakes. Furthermore, the hidden transition matrix tends to be stationary, which means that if the model remains in a specific state, the probability of the model remaining in that state for the next period is high.

After training the above Poisson Hidden Markov Models, the Akaike information criterion (AIC) was used to identify the Parsimonious Model, which means the model requires minimal resources to provide good fitting to the data. According to Table 2, the three-state Poisson hidden Markov models were selected because this model has the smallest Akaike information criterion value.

No of hidden states	***Poisson parameter(lambda) of different states***	***Hidden Transition matrix***	***AIC***	***BIC***
I	19	1	777.3	779.9
II	15 25	0.9 0.1 0.1 0.9	687.9	701.2
III	10 20 30	0.8 0.1 0.1 0.1 0.8 0.1 0.1 0.1 0.8	674.3	703.6
IV	10 15 20 30	0.85 0.05 0.05 0.05 0.05 0.05 0.05 0.85 0.05 0.05 0.85 0.05 0.05 0.05 0.05 0.85	685.8	736.3

Table 2: The parameter, hidden states information criterion of 4 PHMM Models

- ## Accuracy of the Parameter Learning

After selecting the best model among four trained Poisson Hidden Markov Models, Mean Square Error (MSE) was used to check the accuracy of the parameter learning and discover the performance of the model setup. From Table 3, we found that the model initialization is close to the fitted result.

Parameter	***Initial Parameter***	***Fitted Parameter***	***ΓMSE***
Transition matrix	0.8 0.1 0.1 0.1 0.8 0.1 0.1 0.1 0.8	0.94 0.03 0.03 0.04 0.91 0.05 0 0.19 0.81	0.02 0.004 0.004 0.003 0.01 0.002 0.01 0.008 0
Poisson parameter(lambda)	10 20 30	13.2 19.7 29.7	10.2 0.08 0.08

Table 3: The accuracy checking of the lambda and transition matrix for 3-state PHMM Models

- ## Residual Diagnostics

Based on the AIC criteria, a three-state HMM model is selected; however, residual diagnostics is done for further investigation and comparison on the performance of models. From the Q-Q plots of residual provided in the Figure 2, it is clear that both two, three and four-state model provide an acceptable fit while the normal pseudo-residuals of the one-state model (a single Poisson distribution) deviate strikingly from the standard normal distribution. According to the the sample ACF plots of residuals. Besides, we observed that the ordinary pseudo-residuals of the three-state and four-state models are independent and uncorrelated. A hypothesis testing, Kolmogorov Smirnov test (K.S. test), is a goodness-of-fit test which is also constructed to test for the normality of the residuals. From Table 4, it could be concluded that only one-state model fails the normality hypothesis at a significant level of 0.05 while the other three models perform similarly. Based on the residual plots, sample ACF, K.S. test on the residual and AIC value of the model, the three-state could be considered as the adequate model since it achieved the best performance among our models.

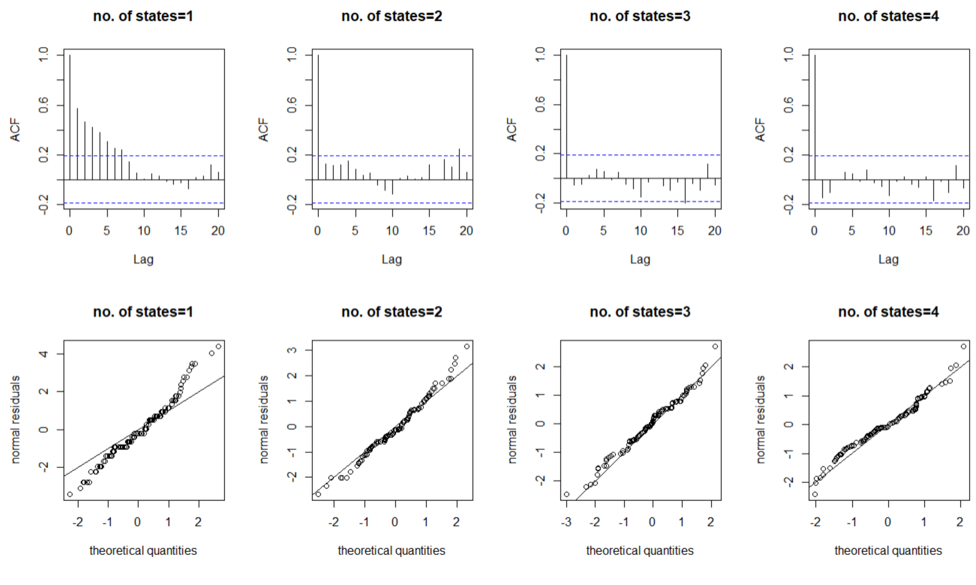

Figure 2: Autocorrelation functions and quantile-quantile plots of the normal pseudo-residuals

K.S. test ($\alpha = 0.05$)	*One-state Model*	*Two-state Model*	*Three-state Model*	*Four-state Model*
P-value	0.0010	0.8581	0.7212	0.7188

Table 4: Result of Kolmogorov Smirnov test (K.S. test) for normality checking of residuals

- **Performance of Selected Model (Viterbi Algorithm)**

The path obtained by the Viterbi Algorithm (Viterbi path) was shown in Figure 3. By comparing the Viterbi path with the series of earthquake data, it is shown that the trend of the series of earthquake data is similar to that of the Viterbi path. For further investigation of the model performance, a sequence of 10,000 observations was simulated from a three-state model with parameter estimates from the EM algorithm, and then the conditional probability of the inferred state given the true state is calculated. From Table 5, we had a partial check of the performance of each state. According to Table 5, we observed that the accuracies in decoding the true states for states 1, 2, and 3 are 93.5%, 92.4%, and 86.0% respectively. It is also observed that there is a 13.6% chance to identify the observation as state 3 when the true state is 2. It is therefore concluded that the three-state model performed well, except for the boundary between states 2 and 3.

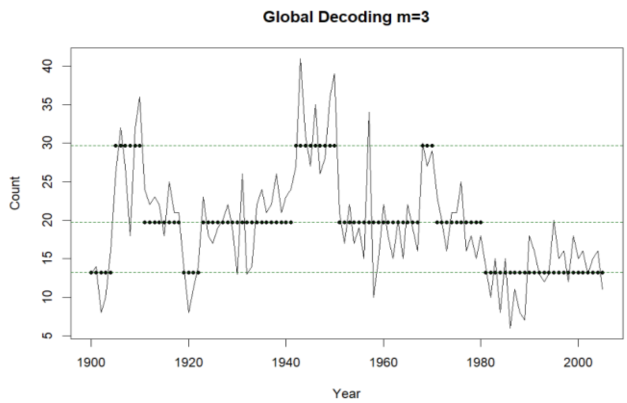

Figure 3[2]: Earthquakes data: global decoding according to three-state PHMM (m=3) model.

[2] The green horizontal dotted line indicates the state-dependent means. The black dot indicates the annual most likely state which is inferred from Viterbi Algorithm.

True / Inferred	i=1	2	3	Sum
j=1	**0.9355**	**0.0639**	**0.0005**	**1**
2	**0.0424**	**0.9239**	**0.0338**	**1**
3	**0.0041**	**0.1361**	**0.8598**	**1**

Table 5: The matrix of P (inferred state = j | true state = i) of simulated data from three-state PHMM model

Discussion

- ## Possible Interpretation and Application of the Model

After global decoding and parameter learning, it could have a clear and substantive interpretation of different states. According to Figure 3, the state could be interpreted as the frequency level of earthquakes. The higher the state is, the more earthquakes it has. For instance, state 1 is interpreted as the lowest frequency level with 13.2 counts on average. State 2 is the middle frequency level with 19.7 counts on average and state 3 is the highest frequency level which also indicates the peak period of earthquakes with 29.7 counts on average.

There are two potential applications of our model. On the one hand, our model could be used to predict the year with high-frequent earthquakes so that residents could prepare the materials earlier and the government could also have enough time to take appropriate measures. On the other hand, it could be used for pattern recognition, and it could also provide statistics and evidence for geography research propose.

- ## Prediction for the number of major earthquakes in 2006

The finalized model was used to forecast the discrete distribution of earthquake count in 2006, which ranges from 0 to 50. $x^{(-t)}$ was defined as all observations at all times except t.

$$x^{(-t)} \equiv (x_1, \dots, x_{t-1}, x_{t+1}, \dots, x_T)$$

$d_i(t)$ was also defined as the product of the i th entry of $\alpha_{t-1}\Gamma$ and the i th of β_{t-1}, where α_{t-1} is the vector form of forward probabilities at time t and β_{t-1} is the vector form of backward probabilities at time t. Γ is the hidden transition matrix of the Hidden Markov model (Walter et al., 2016).

$$d(t) = \alpha_{t-1} \Gamma \beta_t$$

The probabilities of earthquake count at time t given all observations except time t:

$$\Pr\left(X_t = x \mid X^{(-t)} = x^{(-t)}\right) = \sum_{i=1}^{m} w_i(t) p_i(x), where\ w_i(t) = d_i(t) / \sum_{j=1}^{m} d_j(t)$$

By using 3-state PHMM Model, the forecast distribution of the number of earthquakes in 2006 was shown in Figure 4. The actual number of earthquakes in 2006 is 11 while the expected value of forecast distribution is 13.98711 and the mean squared error is 8.922826. Therefore, our group may conclude that its performance is acceptable and reasonable.

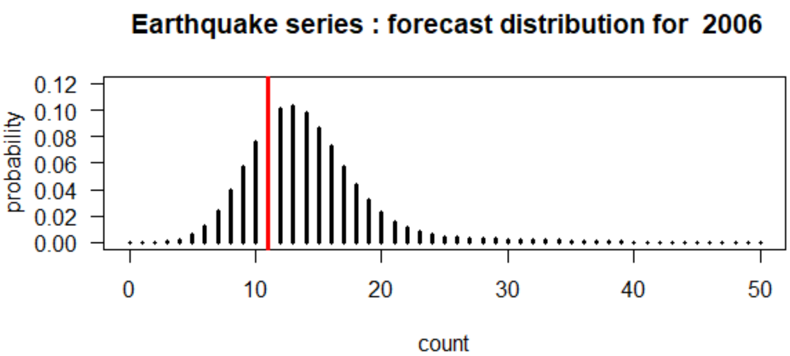

Figure 4: Forecast distribution: earthquake count of 2006 by three-state HMM (m=3).

Conclusion

In this project, Poisson HMM models are applied to analyze the earthquake data from 1900 to 2006. As a result, the model with three states was selected, which obtained the lowest value of AIC during the model selection and its pseudo-residuals are normally distributed. By global decoding, it was found that the hidden states match well with our HMM model. As for the performance of decoding, there was more than 85% accuracy in decoding each state. Besides, we obtained a credible forecasting of frequencies for earthquakes by using the Poisson HMM model which is powerful to predict frequencies and patterns of earthquakes. We are looking forward that these models can make an important contribution to minimizing earthquake damage.

Limitation and Suggestion

In this project, the training dataset contains 106 observations only. However, it may not be enough to have a good fit for the hidden Markov model. As a result, we may consider gathering more data on the number of major earthquakes, such as collecting data from 2007 to 2022. It may improve the performance of the Hidden Markov model by getting more data from the training set.

Besides, our group applied the Poisson Hidden Markov Model only in this study because it seems that the 3. However, this assumption may not be perfectly correct. As a result, we may consider other Hidden Markov Models based on different distributions, such as Gaussian, Gamma, or self-defined discrete distributions for further study.

For selecting the number of states in Hidden Markov Models, four different hidden Markov models are built for model selection, but with those models having similar state interpretations (higher state means a higher probability of the highest frequency level of an earthquake and a transition matrix close to stationary), it may not be comprehensive for the model selection. Therefore, our group may consider the hidden Markov models with more state interpretations such as the transition matrix, which is not close to being stationary.

Reference

1. Bilmes, J.A. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Berkeley, CA: International Computer Science Institute. 7-13.
2. Can, C. E. and Ergun, G. and Gokceoglu C. (2014) Prediction of Earthquake Hazard by Hidden Markov Model. Central European Journal of Geosciences
3. Ephraim, Y. and Merhav, N. (2002). Hidden Markov processes. *IEEE Trans. Inform. Th.* 48,1518-1569.
4. Pierre R.Bushel (2013) Poisson Hidden Markov Model (PHMM)
5. U.S. Geological Survey (USGS) (25 July 2007) Search Earthquake Archives. Retrieved from http://neic.usgs.gov/neis/eqlists
6. Viterbi, A.J. (1967). Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. IEEE Trans. Inform. Th. 13, 260–269.
7. Zucchini, W. and MacDonald, I.L. and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*.

8. Zucchini, W. and MacDonald, I.L. (1999). Illustrations of the use of pseudo residuals in assessing the fit of a model. In Statistical Modelling. Proceedings of the 14th International Workshop on Statistical Modelling, Graz, July 19– 23, 1999, H. Friedl, A. Berghold, G. Kauermann (eds), 409–416.