Estimating Markov Switching Model Using Differential Evolution Algorithm in Prospective Infectious Disease Outbreak Detection

Rui-tian Xu and Jun Zhang (Corresponding Author)
Department of Computer Science, Sun Yat-sen University
Key Laboratory of Digital Life, Ministry of Education
Key Laboratory of Software Technology, Education Dept. of Guangdong Province, P.R. China
junzhang@ieee.org

ABSTRACT

Prospective infectious disease outbreak detection has long been a major concern in public health. Using time series analysis method for the outbreak detection, a nonlinear Markov switching model is better than linear models in modelling time series, due to its ability to describe the switching process of time series variables in different states. However the estimation difficulty of Markov switching model hinders the model's extensive application in practice. The paper proposes using Differential Evolution (termed DE) algorithm to obtain maximum likelihood estimator of Markov switching model in consideration of DE's good global optimization ability. In addition, to effectively reduce negative impact of label switching problem on disease outbreak detection validity of the estimated model by maximum likelihood estimation (termed MLE) method, the paper introduces identifiability constraint on estimation parameters constructed with the heuristic information about difference between durations of different states into MLE using DE. Encouraging experimental study has demonstrated the effectiveness and efficiency of DE in maximizing likelihood function of the studied Markov switching model as well as the effectiveness of the proposed identifiability constraint on improving disease outbreak detection validity of the estimated Markov switching model by MLE.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search --- *Heuristic methods*; G.3 [Probability and Statistics]: Time series analysis

General Terms: Algorithms

Keywords: Differential Evolution, identifiability constraint, label switching problem, Markov switching model, maximum likelihood estimation, prospective infectious disease outbreak detection, time series analysis

GECCO'12, July 7–11, 2012, Philadelphia, Pennsylvania, USA.

1. INTRODUCTION

Prospective infectious disease outbreak detection has long been a major concern in public health and important for infectious disease controlling. It aims at determining whether disease outbreaks currently on the premise that only time series of observations made prior to the time of detection are available to the detection algorithm [1]. Time series analysis is an important quantitative method in the disease outbreak detection. When using the method, the time series modeling is a key point. In the current available time series models, linear models like Autoregressive model, Moving Average model, Autoregressive Moving Average model and Autoregressive Integrated Moving Average model are easy to use but not as good as nonlinear models for practical complex time series [2]-[4]. Markov switching model is an excellent nonlinear time series model to describe the switching process of time series variables in different states which is a complex dynamical evolution process [5]. It is a valuable model for prospective infectious disease outbreak detection. However, the difficulty in model estimation hinders the model's extensive application in practice. In the current available model estimation methods, the Least Square (termed LS) method is suitable for linear model estimation, but not for estimating Markov switching model which includes state variables and is highly nonlinear [5]. The Bayesian Parameters Estimation (termed BPE) method considers estimation parameters as stochastic variables which have certain prior probability distributions. It estimates the values of parameters using observed samples through transforming the prior probability density into posterior probability density based on Bayesian Theorem [6]. However the BPE method assumes that the prior probability distributions of the estimation parameters are known and requires that the posterior distributions of the estimation parameters are valid or able to be imitated using standard probability distributions. Therefore it has seldom been used in practice. The Maximum Likelihood Estimation (termed MLE) method considers estimation parameters as determinate variables with unknown values. The method obtains the estimator that maximizes probability of generating the observed samples, as the best estimator of parameters. MLE has many outstanding characteristics. The maximum likelihood estimator is an asymptotically unbiased estimator. And the method is easier than other methods like BPE to understand and implement so that it is extensively applied in practice [7]. But there are two big challenges in MLE of Markov switching model, i.e. on the one hand, the likelihood function is not easy to solve for obtaining the maximum likelihood estimator; on the other hand, when using the

estimated model by MLE to detect disease outbreak, the label switching problem [8] has a negative impact on detection validity.

Focusing on the above two challenges of MLE, this paper presents a study on using the Differential Evolution (termed DE) algorithm [9] to obtain the maximum likelihood estimator of a Markov switching model. DE is a simple but efficient evolutionary algorithm for global optimization, and has been successfully applied in many practical optimization problems derived from diverse domain like other computational intelligence algorithms [10]-[13]. DE is suitable for MLE of a Markov switching model due to its good global search ability and little restriction on the optimization function. Moreover, this paper introduces a new identifiability constraint on estimation parameters into MLE using DE, to reduce the negative impact of the label switching problem. The method of constructing identifiability constraint with the time series heuristic information about the difference between durations of different states in disease outbreak detection is proposed. In experimental study, encouraging results have demonstrated that introduction of the proposed identifiability constraint improves outbreak detection validity of the estimated model.

The remainder of the paper is organized as follows. Section 2 describes prospective infectious disease outbreak detection using Markov switching model and MLE of the model. Section 3 reviews DE algorithm and studies on using DE to obtain maximum likelihood estimator of the model. Section 4 describes the proposed method for reducing the negative impact of the label switching problem on disease outbreak detection. In Section 5 the experimental study is introduced and the results are given and analyzed. Finally, Section 6 draws the conclusions.

2. DISEASE OUTBREAK DETECTION USING MARKOV SWITCHING MODEL

Given observable time series $Y_T = \{y_0, y_1, ..., y_{T-1}, y_T\}$ where y_t is the value of time series variable y (e.g. the count of infected residents) in time t ($t = 0,1,...,T-1,T$), prospective infectious disease outbreak detection is to determine the disease state of current time T, i.e. outbreak or nonoutbreak. The detection method follows a two-step procedure. In the first step, a time series model is estimated to model Y_T. In the second step, the estimated model is used to determine the current state.

2.1 Adopted Markov Switching Model

The Markov switching model used to model Y_T [14] is as follows:

$$y_t = \alpha_{s_t} + \beta_{s_t} y_{t-1} + \xi_t \tag{1}$$

$$s_t \in \{0,1\} \tag{2}$$

$$p(s_t = j \mid s_{t-1} = i, s_{t-2} = k, ...) = p(s_t = j \mid s_{t-1} = i) = p_{i,j} \tag{3}$$

$$\xi_t \sim N(0, \sigma^2) \tag{4}$$

where (1) defines how the unobservable state variable s_t (0 for nonoutbreak and 1 for outbreak) controls the dynamics of y_t. α_{s_t} and β_{s_t} are values of parameters α and β in the corresponding state s_t. The model presumes that in each state, the dynamics of y can be described as a stationary one-order Autoregressive model where $-1 < \beta_{s_t} < 1$ [15]. The state variable evolves following a first-order Markov process as indicated in (3) that the impact of previous states on s_t is only through s_{t-1} with a transition probability $p_{i,j}$ which is the conditional probability of s_t equaling j ($j = 0,1$) when s_{t-1} equals i ($i = 0,1$). The fluctuation ξ_t in time t has a normal distribution with mean 0 and variance σ^2 where $\sigma > 0$. The introduction of state variable is helpful for the model to control the underlying time series dynamics and model endogenous structural changes. Therefore the model is suitable for disease outbreak detection based on the time series containing different states [14].

2.2 MLE of Adopted Model

To estimate the model with MLE, the likelihood function should be established first. Defining vector of estimation parameters as $\boldsymbol{\theta}$, the form $f(\bullet \mid *; \boldsymbol{\theta})$ as the conditional probability density of $\bullet$ given $*$ regarding to $\boldsymbol{\theta}$, $P(\bullet \mid *; \boldsymbol{\theta})$ as the probability of $\bullet$ given $*$ regarding to $\boldsymbol{\theta}$, where $\bullet$ and $*$ are both vectors of time series variable y and state variable s separated by comma, and Y_t as the vector consisting of y_0, y_1, ..., y_{t-1}, y_t, the conditional logarithmic likelihood (termed likelihood for short) function $L(\boldsymbol{\theta} \mid Y_T)$ of the model regarding to $\boldsymbol{\theta}$ is established as (5),

$$L(\boldsymbol{\theta} \mid Y_T) = \sum_{t=1}^{T} \lg f(y_t \mid Y_{t-1}; \boldsymbol{\theta}) \tag{5}$$

For the model, (6) and (7) are valid,

$$f(y_t \mid Y_{t-1}; \boldsymbol{\theta}) = \sum_{i=0}^{1} \sum_{j=0}^{1} f(y_t, s_t = j, s_{t-1} = i \mid Y_{t-1}; \boldsymbol{\theta}) \tag{6}$$

$$\begin{aligned} &f(y_t, s_t = j, s_{t-1} = i \mid Y_{t-1}; \boldsymbol{\theta}) \\ &= f(y_t \mid s_t = j, s_{t-1} = i, Y_{t-1}; \boldsymbol{\theta}) \times P(s_t = j, s_{t-1} = i \mid Y_{t-1}; \boldsymbol{\theta}) \end{aligned} \tag{7}$$

Because ξ_t has a normal distribution $N(0, \sigma^2)$, (8) is established,

$$f(y_t \mid s_t = j, s_{t-1} = i, Y_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[y_t - \alpha_j - \beta_j y_{t-1}]^2} \tag{8}$$

And (9) to (10) can be established based on Bayesian Theorem,

$$\begin{aligned} &P(s_t = j, s_{t-1} = i \mid Y_{t-1}; \boldsymbol{\theta}) \\ &= P(s_t = j \mid s_{t-1} = i; \boldsymbol{\theta}) \times P(s_{t-1} = i \mid Y_{t-1}; \boldsymbol{\theta}) \\ &= p_{i,j} \times P(s_{t-1} = i \mid Y_{t-1}; \boldsymbol{\theta}) \end{aligned} \tag{9}$$

$$P(s_t = j \mid Y_t; \boldsymbol{\theta}) = \sum_{i=0}^{1} \frac{f(y_t, s_t = j, s_{t-1} = i \mid Y_{t-1}; \boldsymbol{\theta})}{f(y_t \mid Y_{t-1}; \boldsymbol{\theta})} \tag{10}$$

For the model, the estimation parameters are only α_0, α_1, β_0, β_1, $p_{0,0}$, $p_{1,1}$ and σ. $p_{0,1}$ and $p_{1,0}$ can be obtained by $p_{0,1} = 1 - p_{0,0}$ and $p_{1,0} = 1 - p_{1,1}$. To obtain $L(\boldsymbol{\theta} \mid Y_T)$ using (6) to (10), the values of $P(s_0 = 0 \mid Y_0; \boldsymbol{\theta})$ and $P(s_0 = 1 \mid Y_0; \boldsymbol{\theta})$ are needed to be initialized. We adopt the method given in [15] to initialize the two values. The method infers the probabilities of the state variable having different values unconditionally according to the characteristic of ergodicity in the Markov chain. Defining that $\Delta_{2\times 1} = \left[P(s_0 = 0 \mid Y_0; \boldsymbol{\theta}) \quad P(s_0 = 1 \mid Y_0; \boldsymbol{\theta})\right]'$, state transition

probability matrix $T_{2\times2}=\begin{bmatrix} p_{0,0} & p_{1,0} \\ p_{0,1} & p_{1,1} \end{bmatrix}$, $E_{2\times2}$ is a 2×2 identity matrix, $F_{3\times1}$ is the third column of the 3×3 identity matrix, $I_{1\times2}$ is a 1×2 matrix with each element is 1, $P(s_0=0\,|\,Y_0;\boldsymbol{\theta})$ and $P(s_0=1\,|\,Y_0;\boldsymbol{\theta})$ can be obtained by (11),

$$\Delta=(A'A)^{-1}A'F=\begin{bmatrix} \dfrac{1-p_{1,1}}{2-p_{1,1}-p_{0,0}} & \dfrac{1-p_{0,0}}{2-p_{1,1}-p_{0,0}} \end{bmatrix}' \tag{11}$$

where $A_{3\times2}=\begin{bmatrix} E-T \\ I \end{bmatrix}$. More details can be referred in [15].

MLE takes the estimator $\hat{\boldsymbol{\theta}}$ with the maximum of $L(\boldsymbol{\theta}\,|\,Y_T)$ as the best estimator of $\boldsymbol{\theta}$. To search for $\hat{\boldsymbol{\theta}}$, the common differentiation method for MLE and the iterative search algorithms based on derivative like Gauss-Newton method [16] and Newton-Raphson method [17] are not suitable and convenient. Hamilton proposed a particular form of the Expectation-maximization (termed EM) algorithm [15] which has a fast convergence speed. However it is possible that the algorithm converges to a local maximum instead. This paper proposes using DE algorithm to search for $\hat{\boldsymbol{\theta}}$. As an evolutionary algorithm, DE has good global optimization ability and no restriction on differentiability, continuity and unimodality of the objective function. It is suitable to maximizing $L(\boldsymbol{\theta}\,|\,Y_T)$.

2.3 Disease Outbreak Detection Using Estimated Model

After obtaining the maximum likelihood estimator of $\boldsymbol{\theta}$, i.e. $\hat{\boldsymbol{\theta}}$, s_T can be determined by calculating the value of $P(s_T=1\,|\,Y_T;\hat{\boldsymbol{\theta}})$ (aliased alarm score below). If the obtained alarm score is larger than a prescribed threshold (termed alarm threshold) ς, then s_T equals 1, i.e. T is a disease outbreak time, and a alarm will be triggered to alert the monitors. Otherwise s_T equals 0, i.e. T is a nonoutbreak time, and no alarm will be triggered.

3. DE ALGORITHM FOR MLE

3.1 Review of DE Algorithm

DE searches for a global optimum in a feasible search space maintaining a population, i.e. a set of parameter vector $\{\boldsymbol{x}_i^g=[x_{i,1}^g,x_{i,2}^g,...,x_{i,D}^g],i=1,2,...,NP\}$, where NP is the size of the population, g denotes the index of the current generation, and D is the number of parameters in the vector also the dimension of the search space. DE initializes the population in the beginning of evolution when g=0. The jth component of the ith parameter vector $x_{i,j}^0$ is set by (12),

$$x_{i,j}^0=xmin_j+\text{rand}(0,1)\times(xmax_j-xmin_j) \tag{12}$$

where $xmax_j$ and $xmin_j$ are prescribed upper and lower bounds of the jth component of parameter vector respectively, rand(0,1) is a random number generation function that returns a uniform random real number in the range $[0,1]$. After initialization, DE involves the general evolution operators in order, i.e. mutation, crossover and selection, to iteratively evolve the population for optimizing the objective until meeting the end condition.

1) Mutation: The operator is used on each parameter vector $\boldsymbol{x}_i^g$ (aliased target vector below) to generate a corresponding mutant vector $\boldsymbol{v}_i^g$. Although various sophisticated mutation strategies have been proposed, there are five most frequently used basic mutation strategies, i.e. DE/rand/1, DE/rand/2, DE/best/1, DE/best/2 and DE/target-to-best/1 [10]. Due to the page limit of this paper, (13) describes DE/rand/2 only. More details about other mutation strategies can be referred in [10],

$$\boldsymbol{v}_i^g=\boldsymbol{x}_{r_1}^g+F\times(\boldsymbol{x}_{r_2}^g-\boldsymbol{x}_{r_3}^g)+F\times(\boldsymbol{x}_{r_4}^g-\boldsymbol{x}_{r_5}^g) \tag{13}$$

where indices r_1, r_2, r_3, r_4 and r_5 are distinct random integers uniformly sampled in the rage $[1,NP]$ and all different from i. Scale factor F is a positive control parameter to amplify the difference vectors.

2) Crossover: The operator is used to determine which components of $\boldsymbol{v}_i^g$ and $\boldsymbol{x}_i^g$ form the trial vector $\boldsymbol{u}_i^g$ by (14),

$$u_{i,j}^g=\begin{cases} v_{i,j}^g, & \text{if rand}(0,1)\le CR \text{ or } j=j_{rand} \\ x_{i,j}^g, & \text{otherwise} \end{cases} \tag{14}$$

where $u_{i,j}^g$, $v_{i,j}^g$ and $x_{i,j}^g$ denote the jth component of the ith trial vector, mutant vector and target vector respectively in the current generation. Function rand(0,1) returns a random real number in the range $[0,1]$. Crossover probability CR ($0\le CR\le1$) is another control parameter of DE to determine the fraction of components in the mutant vector which can be used to generate the trial vector. j_{rand} is a random integer uniformly selected in the range $[1,D]$. The branch condition stated in (14) ensures that the trial vector has at least one component inherited from the mutant vector in order to enhance the population diversity.

3) Selection: The operator is performed after mutation and crossover to determine whether the trial vector can survive to the next generation. For every target vector, if the fitness of the corresponding trial vector isn't worse than its, it will be replaced by the trial vector for the next generation. The determination mechanism can be formulated as (15),

$$\boldsymbol{x}_i^{g+1}=\begin{cases} \boldsymbol{u}_i^g, & \text{if fitness of } \boldsymbol{u}_i^g \text{ isn't worse than that of } \boldsymbol{x}_i^g \\ \boldsymbol{x}_i^g, & \text{otherwise} \end{cases} \tag{15}$$

3.2 MLE Using DE Algorithm

Figure 1 gives the pseudo code of DE algorithm with DE/rand/2 mutation strategy for MLE. Although in the literature there are various alternative mutation strategies for us to choose, we consider adopting DE/rand/2 for the below considerations: On the one hand, if the simplest DE with basic mutation strategy would give good result, we are confident in the sense that more sophisticated mutation strategy could only work better regardless of time consuming. On the other hand, DE/rand/2 is one of the mutation strategies that bear strongest exploration capability to avoid premature convergence or stagnation in the five most frequently used basic mutation strategies [18].

In the algorithm shown by Figure 1, the dimension of parameter vector D is 7 and the components of each parameter vector correspond to α_0, α_1, β_0, β_1, p_{00}, p_{11} and σ in sequence. Because $-1<\beta_0,\beta_1<1$ and $0\le p_{00},p_{11}\le 1$, so for β_0 and β_1, the upper bounds are both set to be 1 and the lower bounds -1, and for p_{00} and p_{11}, the upper bounds are both set to be 1 and the lower bounds 0. For α_0, α_1 and σ, the upper and lower bounds can be set according to the observed time series in the specific scenario. For each parameter vector, the fitness evaluation function is $L(\boldsymbol{\theta}\,|\,Y_T)$ denoted by (5). Therefore the larger fitness a parameter vector has, the better it is.

Algorithm: DE Algorithm for MLE

1. initialize index of current generation g:=0 and the population
2. evaluate the parameter vectors in the population and record the searched optimum
3. **while** (number of fitness evaluations is smaller than the prescribed value) **do**
4. **for** i:=1 to NP **do**
5. generate mutant vector $\boldsymbol{v}_i^g$ according to the DE/rand/2 mutation strategy //Mutation
6. **for** j:=1 to D **do**
7. **if** $v_{i,j}^g > xmax_j$ or $v_{i,j}^g < xmin_j$ **then** $v_{i,j}^g := \text{rand}(xmin_j, xmax_j)$ **End if**
8. **End for**
9. j_{rand} := a random integer uniformly selected in the range [1, D]
10. **for** j:=1 to D **do** //Crossover
11. **if** j= j_{rand} or rand(0,1) $\le CR$ **then** $u_{i,j}^g := v_{i,j}^g$ **else** $u_{i,j}^g := x_{i,j}^g$ **End if**
12. **End for**
13. evaluate the fitness of $\boldsymbol{u}_i^g$ and update the searched optimum if necessary
14. **if** $\boldsymbol{u}_i^g$ is not worse than $\boldsymbol{x}_i^g$ **then** $\boldsymbol{x}_i^{g+1} := \boldsymbol{u}_i^g$ **else** $\boldsymbol{x}_i^{g+1} := \boldsymbol{x}_i^g$ **End if** //Selection
15. **End for**
16. g:=g+1
17. **End while**
18. output the searched optimum as $\hat{\boldsymbol{\theta}}$

Figure 1. Pseudo code of DE algorithm for MLE

4. MLE WITH PARAMETERS IDENTIFIABILITY CONSTRAINT

In the context of disease outbreak detection, a serious drawback of MLE is the label switching problem, i.e. for a certain parameters estimator $\boldsymbol{\theta}_1$, a new parameters estimator $\boldsymbol{\theta}_2$ is obtained by respectively switching the values between α_0 and α_1, β_0 and β_1, $p_{0,0}$ and $p_{1,1}$ in $\boldsymbol{\theta}_1$. It can be proved according to (5)-(10) that $L(\boldsymbol{\theta}_1\,|\,Y_T) = L(\boldsymbol{\theta}_2\,|\,Y_T)$, $P(s_T=1\,|\,Y_T;\boldsymbol{\theta}_2) = P(s_T=0\,|\,Y_T;\boldsymbol{\theta}_1)$ and $P(s_T=0\,|\,Y_T;\boldsymbol{\theta}_2) = P(s_T=1\,|\,Y_T;\boldsymbol{\theta}_1)$. Details of the proof are not demonstrated due to the page limit. Although $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ share the equal likelihood, they have opposite detection result according to a certain alarm threshold. Therefore the existence of label switching problem can cause detection error. A common response to the problem is to impose an identifiability constraint on parameters [8]. In MLE of the studied model, the paper imposes identifiability constraint on the values of p_{00} and p_{11}, which is constructed exploiting the heuristic information of the difference between durations of different states. For the outbreak state, the expected duration D_{outbreak} is obtained by (16),

$$D_{\text{outbreak}} = \lim_{n\to\infty}(1 + p_{1,1} + p_{1,1}^2 + p_{1,1}^3 + \cdots + p_{1,1}^n) = \frac{1}{1-p_{1,1}} \quad (16)$$

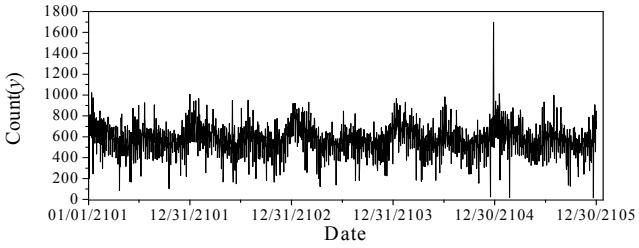

Figure 2. Data set used for experiment

For the nonoutbreak state, the expected duration $D_{\text{nonoutbreak}}$ is obtained by (17),

$$D_{\text{nonoutbreak}} = \lim_{n\to\infty}(1 + p_{0,0} + p_{0,0}^2 + p_{0,0}^3 + \cdots + p_{0,0}^n) = \frac{1}{1-p_{0,0}} \quad (17)$$

Generally for the studied disease outbreak detection, the nonoutbreak state has a longer duration than the outbreak state. Therefore (18) is established,

$$\frac{1}{1-p_{1,1}} < \frac{1}{1-p_{0,0}} \Rightarrow p_{0,0} > p_{1,1} \quad (18)$$

Due to the simplicity of DE, it is easy to implement the identifiability constraint shown by (18) in MLE. The modifications of the algorithm showed in Figure 1 are only made in the initialization of the population in Line 1 and before evaluating the fitness of $\boldsymbol{u}_i^g$ in Line 13. For each parameter vector

in the initialized population and $\boldsymbol{u}_i^g$, if the values of the components corresponding to p_{00} and p_{11} do not satisfy (18), the two values should be swapped. Comparatively speaking, the identifiability constraint is difficult to be implemented and satisfied in the EM algorithm proposed in [15].

5. EXPERIMENTAL STUDY

5.1 Experimental Design

In this section, an experiment is performed to validate the effectiveness and efficiency of DE on obtaining the maximum likelihood estimator of the Markov switching model as well as the effectiveness of the proposed identifiability constraint on reducing the negative impact of label switching problem on disease outbreak detection. The experiment uses a disease outbreak detection scenario based on a data set developed by International Society for Disease Surveillance [19]. The adopted data set is "OTC-train-01" available in the Zip file "ISDS_contest_training_data.zip" downloadable in http://isds.wikispaces.com/Resources and contains a daily time series of 5 years with explicit disease state relating to each day for ease of test. The data set is plotted in Figure 2, where the horizontal ordinate denotes the dates marked starting from "01/01/2101" to "12/31/2105", the vertical ordinate denotes the time series variable y , i.e. the count of aggregated OTC antidiarrheal and antinauseant sales. The time series contains an outbreak starting from "12/14/2104" to "02/02/2105". For ease of experiment without loss of generality, in the used scenario, the disease outbreak detections operated on the days starting from "11/19/2104" to "02/28/2105" are studied where there are 51 outbreak days and 51 nonoutbreak days consisting of 25 days before the outbreak and 26 days after the outbreak. On each detection day, the time series starting from "01/01/2101" up to the day are available to aid in detection. For MLE of the model using DE, the upper bounds of α_0 and α_1 are both set to be 10000, and the lower bounds 0. For σ the upper bound is set to be 20000, and the lower bound 1. Due to the fact that DE is a nondeterministic algorithm, for the purpose of experiment, DE independently runs 5 times in MLE of the studied model on each detection day, and then the resultant estimator $\hat{\boldsymbol{\theta}}$ with the maximum likelihood in 5 runs is used to determine the disease state relating to the day.

5.2 Study on Effectiveness and Efficiency of DE Algorithm for MLE

To study the effectiveness and efficiency of DE for MLE of the Markov switching model, the experiment tests the obtained $\hat{\boldsymbol{\theta}}$ by DE. Because the search space is changed by introduction of identifiability constraint on estimation parameters, the experiment is carried for both MLE with identifiability constraint (termed MLEic) and MLE without identifiability constraint (termed MLEn). The results obtained by DE are compared with a Particle Swarm Optimization (termed PSO) algorithm, a well-known swarm intelligent based numerical optimization algorithm [20]. Furthermore, the EM algorithm is implemented for MLEn. Due to the fact that PSO is also a nondeterministic algorithm and the randomness of EM in initializing estimation parameters, PSO and EM both run 5 times on each detection day like DE.

In each independent run, the maximum fitness evaluations of DE and PSO are both 50000 and the sizes of population 20. For DE its scale factor F is 0.5 and the crossover probability CR is 0.4 which are determined by a prior trial-and-error procedure. For PSO its inertia weight is linearly decreasing from 0.9 to 0.4 generation by generation, self-cognitive and social-influence weights are both 2, and the upper bound of speed for each parameter is 20% of each parameter's domain size. The setting of the control parameters for PSO is determined by a prior trial-and-error procedure and can be referred in the literature [20]-[22].

Figure 3 and Figure 4 respectively gives a comparison on results of DE and PSO in MLEic and those of DE, PSO and EM in MLEn for each detection day T. In the figures, the average value and standard deviation of the obtained optimal $L(\boldsymbol{\theta}\,|\,Y_T)$ which has a 0.01 precision in 5 runs are calculated and presented by error bar graph for DE. And for PSO and EM, the maximum of obtained optimal $L(\boldsymbol{\theta}\,|\,Y_T)$ in 5 runs is chosen and plotted. It is obvious that DE obtains better estimator than PSO and EM for each detection day. PSO and EM converge to a local maximum on most detection days. The values of obtained optimal $L(\boldsymbol{\theta}\,|\,Y_T)$ in DE's 5 runs are equal for most days excluding 12 days with clear top and bottom error bars presented in Figure 3 and 4 days in Figure 4. In the study on convergence speed of DE and PSO, we learn that DE has a higher convergence speed than PSO in both MLEic and MLEn which is not demonstrated in detail due to the page limit. Furthermore we find that the maximum values of obtained optimal $L(\boldsymbol{\theta}\,|\,Y_T)$ in DE's 5 runs for MLEic and MLEn are equal for each detection day T. These make us confident of that DE works well in maximizing the likelihood function for the Markov switching model.

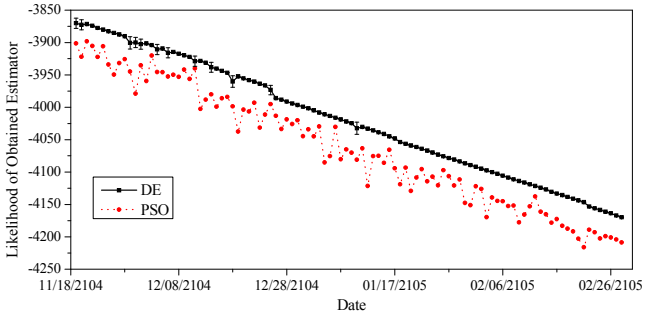

Figure 3. Results Comparison of DE and PSO in MLEic

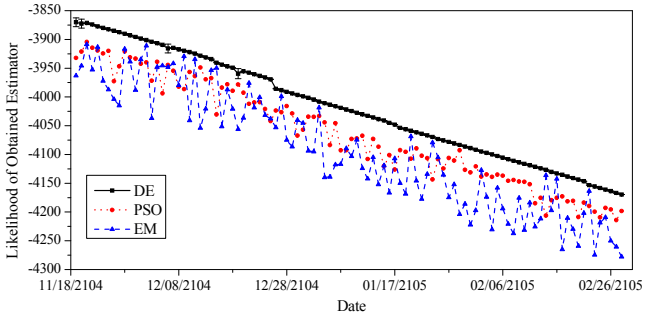

Figure 4. Results Comparison of DE, PSO and EM in MLEn

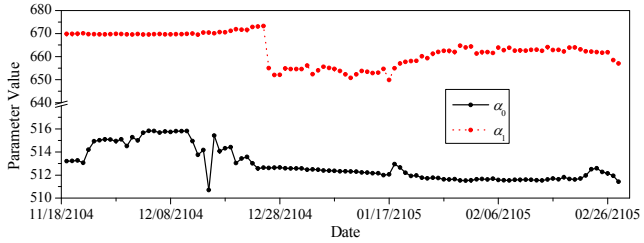

Figure 5. Values of α_0 and α_1 obtained by DE in MLEic

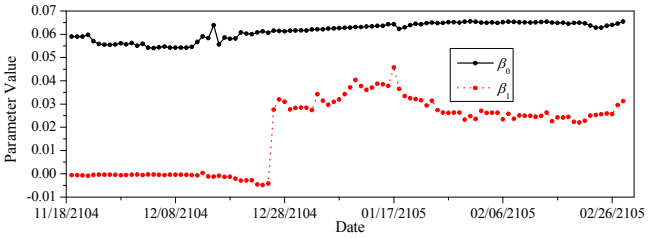

Figure 6. Values of β_0 and β_1 obtained by DE in MLEic

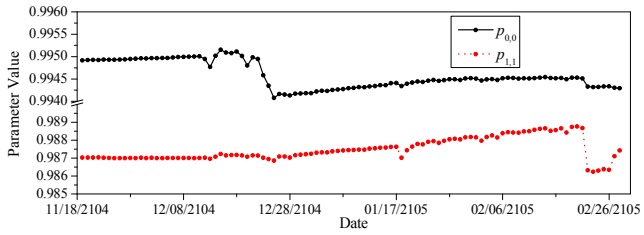

Figure 7. Values of $p_{0,0}$ and $p_{1,1}$ obtained by DE in MLEic

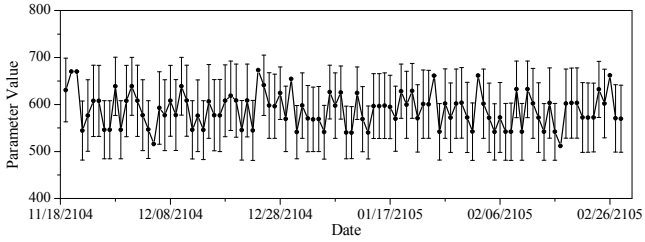

Figure 8. Values of α_0 obtained by DE in MLEn

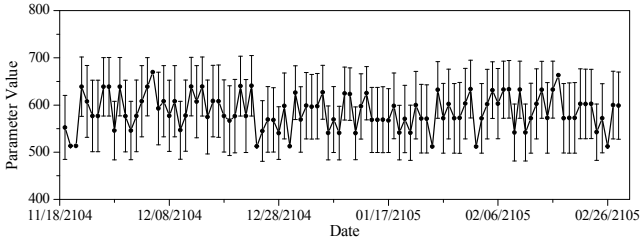

Figure 9. Values of α_1 obtained by DE in MLEn

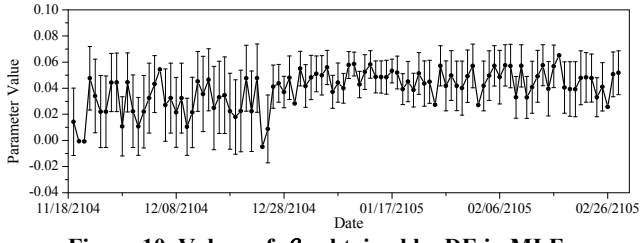

Figure 10. Values of β_0 obtained by DE in MLEn

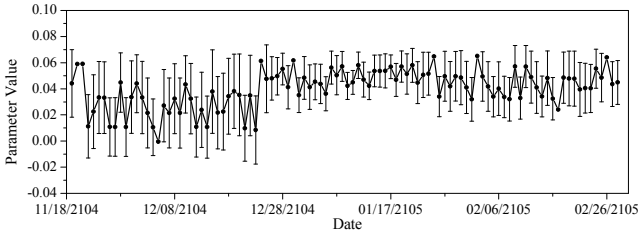

Figure 11. Values of β_1 obtained by DE in MLEn

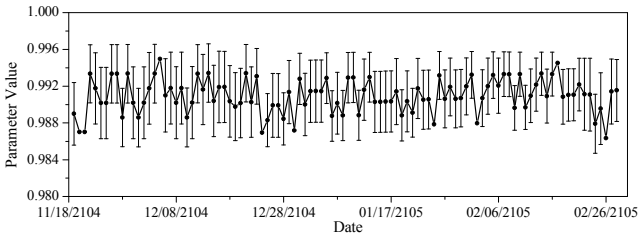

Figure 12. Values of $p_{0,0}$ obtained by DE in MLEn

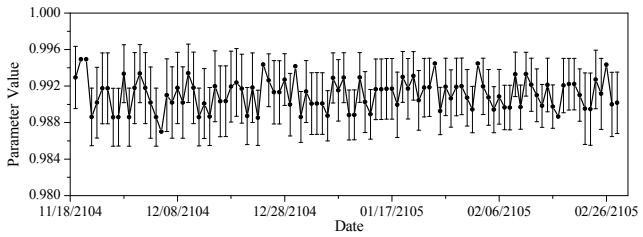

Figure 13. Values of $p_{1,1}$ obtained by DE in MLEn

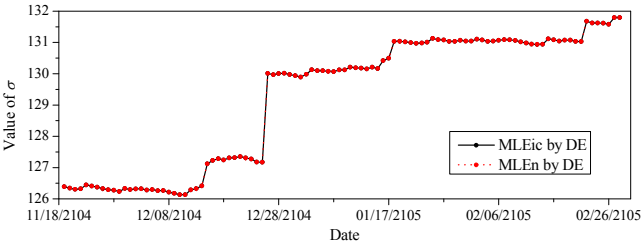

Figure 14. Values of σ obtained by DE in MLEic and MLEn

To closely learn the estimators obtained by DE in MLEic and MLEn, on each detection day, we have a statistics on parameters values of the resultant estimators with the maximum likelihood in 5 runs. For each parameter, the average value and standard deviation are calculated and presented by error bar graph. Figure 5 presents the error bars of α_0 and α_1 in MLEic, and Figure 6 β_0 and β_1, Figure 7 $p_{0,0}$ and $p_{1,1}$. Figure 8 to Figure 13 respectively present the error bars of α_0, α_1, β_0, β_1, $p_{0,0}$ and $p_{1,1}$ in MLEn. Figure 14 presents the error bars of σ in both MLEic and MLEn. In MLEic, each parameter has a standard deviation equals or approximates to 0 in value which can be obviously observed in Figure 5 to Figure 7 and Figure 14.

However it is not optimistic for MLEn that each parameter excluding σ has an intolerable large standard deviation in value which can be obviously observed in Figure 8 to Figure 13. However the observation is not surprising because of the existence of label switching problem. Defining the obtained optimal likelihood is $\tilde{L}(\boldsymbol{\theta}\,|\,Y_T)$ in DE's any run on a detection day T, $\boldsymbol{\theta}_1$

and $\boldsymbol{\theta}_2$ share the same likelihood $\tilde{L}(\boldsymbol{\theta}|Y_T)$ where $\boldsymbol{\theta}_2$ is obtained by respectively switching the values between α_0 and α_1, β_0 and β_1, $p_{0,0}$ and $p_{1,1}$ in $\boldsymbol{\theta}_1$, then $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ have a same probability of being the obtained estimator. Therefore there are large fluctuations in values of α_0, α_1, β_0, β_1, $p_{0,0}$ and $p_{1,1}$ in the statistics. The observation also demonstrates the advantage of the proposed identifiability constraint in MLE.

5.3 Study on Effectiveness of Proposed Identifiability Constraint

To study the effectiveness of the proposed identifiability constraint to reduce the negative impact of label switching problem on the estimated model's disease outbreak detection validity, on each detection day, the experiment respectively uses the estimated model with the maximum likelihood in DE's 5 runs in MLEic and MLEn to determine the disease state.

Table 1. Statistics on detection metrics when ς is 0.5

	MLEn	MLEic
FAR	56.9% (29)	39.2% (20)
Per-day Sensitivity	47.1% (24)	82.4% (42)

Table 2. Statistics result under different tolerable FAR

FAR	Per-day Sensitivity	
	MLEn	MLEic
0% (0)	2.0% (1)	17.6% (9)
1.25% (1)	9.8% (5)	25.5% (13)
2.5% (2)	9.8% (5)	35.3% (18)
5% (3)	11.8% (6)	54.9% (28)
7.50% (4)	11.8% (6)	54.9% (28)
10% (5)	11.8% (6)	54.9% (28)

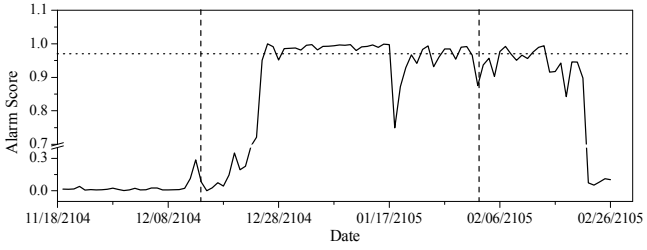

Figure 15. Alarm score of each detection day with MLEic

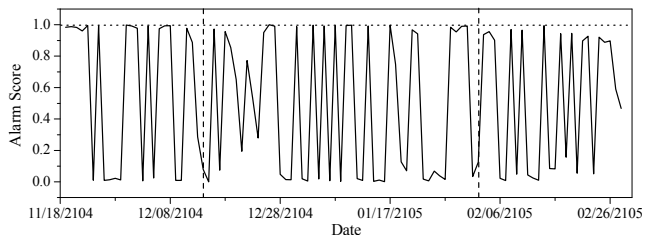

Figure 16. Alarm score of each detection day with MLEn

To measure the detection validity, we choose two intuitive metrics which are used by most disease outbreak detection studies [14], i.e. per-day sensitivity and false alarm rate (termed FAR). Per-day sensitivity is defined as the probability of having alarms on outbreak days, and FAR the probability of having alarms on nonoutbreak days. In practice, too high FAR will cause unnecessary fear and waste of public health resource, too low per-day sensitivity will lead to bad disease control timeliness. Table 1 gives statistics on the two metrics relating to the estimated models by MLEic and MLEn when the alarm threshold ς is prescribed as 0.5. The parentheses in the second row are the numbers of nonoutbreak days having alarms, and those in the third row the numbers of outbreak days having alarms. It is obvious the estimated model by MLEic has better detection validity.

To better understand the difference in detection validity, according to the obtained alarm score on each detection day, we respectively adjust the value of alarm threshold ς for the estimated models by MLEn and MLEic, to restrict FAR. The per-day sensitivity under different tolerable FAR is given in Table 2 where the parentheses in the first column are the numbers of nonoutbreak days having alarms according to the tolerable FAR, and the parentheses in the last two columns are the numbers of outbreak days having alarms. When the tolerable FAR is between 0% and 10%, the per-day sensitivity of the estimated model by MLEn is still obviously worse than that by MLEic.

For each detection day, the obtained alarm score based on the MLEic's resultant estimator is plotted in Figure 15. The two dash lines in the figure marks the start and end dates of the outbreak. The dot line identifies the corresponding alarm threshold ς when the tolerable FAR is 10%. The days starting from "12/29/2104" to "01/17/2105" have corresponding alarm score larger than ς. That is to say, the estimated model by MLEic detects an outbreak lasting 20 days.

Figure 16 plots the obtained alarm score for each detection day based on the MLEn's resultant estimator. The dash lines are defined as same as those in Figure 15. The dot line identifies the corresponding alarm threshold ς when the tolerable FAR is 10% and the value of ς is different from that of Figure 15. To lower the FAR, the value of ς is adjusted to approximate to 1 extremely. In spite of this, under the tolerable FAR, very few outbreak days and no lasting outbreak are detected. The obvious contrast on validity of outbreak detection can be explained by the negative impact of label switching problem. On the detection day T, assume that the maximum of obtained optimal likelihood function values in DE's 5 runs for MLEn is $\tilde{L}(\boldsymbol{\theta}|Y_T)$, and parameters estimators $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ share the same likelihood $\tilde{L}(\boldsymbol{\theta}|Y_T)$ where $\boldsymbol{\theta}_2$ is obtained by respectively switching the values between α_0 and α_1, β_0 and β_1, $p_{0,0}$ and $p_{1,1}$ in $\boldsymbol{\theta}_1$, then the two models respectively determined by $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ have a same probability of being used to determine the current disease state. However, the calculated values of $P(s_T=1|Y_T;\boldsymbol{\theta}_1)$ and $P(s_T=1|Y_T;\boldsymbol{\theta}_2)$, i.e. the alarm score, are quite different and the detection results are opposite according to a prescribed alarm threshold. Suppose that the model determined by $\boldsymbol{\theta}_1$ gives a right detection result, then that determined by $\boldsymbol{\theta}_2$ would give a wrong detection result. The

probability of using the model corresponding to $\boldsymbol{\theta}_2$ for determining the current disease state lowers the disease outbreak detection validity.

6. CONCLUSION

The paper proposes using DE algorithm to obtain the maximum likelihood estimator of Markov switching model for prospective infectious disease outbreak detection. Due to its simplicity, good global optimization ability and no restriction on differentiability, continuity and unimodality of the objective function, DE is suitable and effective to optimize the complex likelihood function for Markov switching model. The encouraging experimental results make us confident of that DE provides a good and easy solution for the challenge of estimating Markov switching model.

To reduce negative impact of the label switching problem on the disease outbreak detection validity of the estimated Markov switching model by MLE, the paper introduces the identifiability constraint constructed with the heuristic information about the difference between durations of different states on the estimation parameters. The identifiability constraint is easy to implement in MLE using DE due to DE's simplicity in algorithm design. And the contrastive experimental study has demonstrated the effectiveness of the proposed identifiability constraint on improving the detection validity of estimated model.

The work reported in the paper suggests a promising future of the use of DE in time series analysis. We plan to extend our study to other more complex time series models and explore opportunities to study other time series application areas beyond prospective infectious disease outbreak detection.

7. ACKNOWLEDGMENTS

This work was supported in part by the National Science Fund for Distinguished Young Scholars No.61125205, National Natural Science Foundation of China (NSFC) No.61070004 and NSFC Joint Fund with Guangdong under Key Project U0835002. The authors are with the Key Lab of Digital Life, MoE, China, and the Key Lab of Software Tech., Edu. Dept. of Guangdong. The corresponding author is Jun Zhang, email: junzhang@ieee.org.

8. REFERENCES

[1] Hu, P. H., et al. 2007. System for Infectious Disease Information Sharing and Analysis: Design and Evaluation. *IEEE Trans. Info. Tech. Bio.,* 11, 4 (Jul. 2007), 483-492.

[2] Box, G., Jenkins, G. M., and Reinsel, G. 1994. *Time Series Analysis: Forecasting and Control.* 3rd Edition, Prentice Hall.

[3] Zhang, J., et al. 2004. Time Series Prediction Using Lyapunov Exponents in Embedding Phase Space. *Comput. Electr. Eng.,* 30, 1 (Jan. 2004), 1-15.

[4] Zhang, J., Chung, S. H. H., and Lo, W. L. 2008. Chaotic Time Series Prediction Using a Neuro-Fuzzy System with Time-Delay Coordinates. *IEEE Trans. Know. Data Eng.,* 20, 7 (Jul. 2008), 956-964.

[5] Hamilton, J. D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica,* 57, 2 (Mar. 1989), 357-384.

[6] Mafred, O. and David, H. 1991. Generalization Performance of Bayes Optimal Classification Algorithm for Learning a Perceptron. *Phys. Rev. Lett.,* 66, 20 (May. 1991), 2677-2680.

[7] Duda, R. O., Hart, P. E., and Stork, D. G. 2000. *Pattern Classification.* 2nd Edition, John Wiley and Sons.

[8] Matthew, S. 2000. Dealing with Label Switching in Mixture Models. *J. R. Statist. Soc.,* 62, 4 (2000), 795-809.

[9] Storn, R. M. and Price, K. V. 1997. Differential Evolution-A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *J. Global Optim.,* 11, 4 (Dec. 1997), 341–359.

[10] Das, S. and Suganthan, P. N. 2011. Differential Evolution: A Survey of the State-of-the-Art. *IEEE Trans. Evol. Comput.,* 15, 1 (Feb. 2011), 4-31.

[11] Zhang, J., et al. 2011. Evolutionary Computation Meets Machine Learning: A Survey. *IEEE Comput. Intell. Mag.,* 6, 4 (Nov. 2011), 68-75.

[12] Zhang, J., Chung, S. H. H., and Lo, W. L. 2007. Clustering-Based Adaptive Crossover and Mutation Probabilities for Genetic Algorithms. *IEEE Trans. Evol. Comput.,* 11, 3 (Jun. 2007), 326-335.

[13] Chen, W. N. and Zhang, J. 2009. An Ant Colony Optimization Approach to a Grid Workflow Scheduling Problem with Various QoS Requirements. *IEEE Trans. Syst., Man, Cybern., C,* 39, 1 (Jan. 2009), 29-43.

[14] Lu, H. M., Zeng, D., and Chen, H. 2010. Prospective Infectious Disease Outbreak Detection Using Markov Switching Models. *IEEE Trans. Know. Data Eng.,* 22, 4 (Apr. 2010), 565-577.

[15] Hamilton, J. D. 1994. *Time Series Analysis.* Princeton University Press.

[16] Nocedal, J. and Wright, S. J. 2000. *Numerical Optimization.* Springer.

[17] Ypma, T. J. 1995. Historical Development of the Newton-Raphson Method. *SIAM Rev.,* 37, 4 (Dec. 1995), 531-551.

[18] Qin, A. K., Huang, V. L., and Suganthan, P. N. 2009. Differential Evolution Algorithm with Strategy Adaptation for Global Numerical Optimization. *IEEE Trans. Evol. Comput.,* 13, 2 (Apr. 2009), 398-417.

[19] ISDS, *https://wiki.cirg.washington.edu/pub/bin/view/Isds/TechnicalContest,* accessed, Dec. 2011.

[20] Shi, Y. H. and Eberhart, R. C. 1998. A Modified Particle Swarm Optimizer. In *Proceedings of IEEE International Conference on Evolutionary Computation,* 69-73.

[21] Zhan, Z. H., et al. 2009. Adaptive Particle Swarm Optimization. *IEEE Trans. Syst., Man, Cybern., B,* 39, 6 (Dec. 2009), 1362-1381.

[22] Chen, W. N., et al. 2010. A Novel Set-Based Particle Swarm Optimization Method for Discrete Optimization Problems. *IEEE Trans. Evol. Comput.,* 14, 2 (Apr. 2010), 278-300.