Change Point Detection based on Call Detail Records

Huiqi Zhang, Ram Dantu
Department of Computer Science and Engineering
University of North Texas
{hz0019, rdantu}@unt.edu

João W. Cangussu
Department of Computer Science
University of Texas at Dallas
cangussu@utdallas.edu

***Abstract*- In this paper we propose a method for combining wavelet denoising and sequential approach for detecting change points on mobile phone based on detailed call records. The Minmax method is used to estimate the thresholds of frequency and call duration for denoising. This work is useful to enhance homeland security, detecting unwanted calls (e.g., spam) and commercial purposes. For validation of our results, we randomly choose actual call logs of 20 users from 100 users collected at MIT by the Reality Mining Project group for a period of 8 months. Simulation data is also used to validate the results. The experimental results show that our model achieves good performance with high accuracy.**

I. Introduction

Change point detection is performed on a series of time ordered data in order to detect whether any changes have occurred. It determines the number of changes and estimates the time of each change. Change point detection problems have many applications, including industrial quality control, reliability, fault detection, clinical trials, finance, environment, climate, signal detection, surveillance and security systems. Analyzing pattern changes of human behavior is an area of increasing interest in a number of different applications. The automatic detection of change points by studying patterns of human behavior is one of them and has recently attracted attention. One of the important applications of change point detection is in the area of homeland security. For example, terrorists or robbers usually do some attack in groups. The leader usually communicates with his members by wireless phones for planning, coordination, and command of the attacks. In this period of time their calling patterns are different from their usual ones. There is more number of calls or longer talk time than those of usual cases among their members; especially the leader. Also they meet at some particular place, especially the attack target place for planning and attacking. Combined with other evidences change point detection of calling patterns can be very useful to prevent security threats.

We use the term *change point* to refer to a large scale activity that changes relative to normal patterns of behavior. To understand such data, we often care about both the patterns of the typical behavior, and detecting and extracting information from the deviations from this behavior.

In this paper we combine the wavelet denoising and sequential detection methods to detect change points based on detailed call records. There is no content in call detail records. We can only use the information such as the time of initiation of a call, number of call in a period of time, a call duration, incoming call, outgoing call, and location. The Minmax is used to estimate the thresholds of number of calls and call duration for denoising. The experimental results show that our method achieves good performance with high accuracy.

This paper is organized as follow. In Section 2 we briefly review the related work. In Section 3 the combination of wavelet denoising and sequential change point detection method is presented. We performed the experiments with both simulation data and the actual call logs and discussed the results in Section 4. The performance analysis of our method is conducted by the actual call logs and described in Section 5. The validation of our model is conducted by the actual call logs and described in Section 6. Finally, we have the conclusions in Section 7.

II. Related Work

There are large amount of previous work on change point detection. In [1] the author proposed efficient on-line and off-line nonparametric algorithms for detecting the change-point based on histogram density estimators. In [2] the authors developed schemes for detecting early change points and frequent change points. They possess a number of desired properties including distribution-consistency which implies convergence of small-sample change-point estimators. The above methods are applied to the temperatures, climate and software engineering quality control data [3]. In [4] the authors develop a Bayesian approach for detecting a single change point at an unknown time of a Poisson process. In [5] the authors propose a Bayesian binary segmentation procedure for detecting multiple change points for a Poisson process. In [6] the authors apply the Bayesian change point method to neural data under the assumption of inhomogeneous Poisson process. In [7] the authors propose the changing linear regression model and obtain the desired posterior densities by iterative Monte Carlo method. In [8] the authors proposed the revision of two-phase linear regression model to apply to climate data. In [9] the authors developed the procedure for detecting change points of a series of varying normal means under the assumption of the known variance. In [10] the authors extended the method in [9] for a series of varying normal mean with unknown variance. These procedures are performed based on the likelihood ratio test. In [11] the authors use a binary procedure combined with Swarz information criterion for testing and locating variance change points in a series of independent normal random variables under the assumption of the known and common mean. In [12] the authors applied reversible jump Markov chain Monte Carlo simulation to estimate variance change points of activation patterns from electromyographic data with assumption of the data to be a zero-mean. The variance is modeled by a step function. In [13] Bayesian approach is applied to detect change points in the time series of annual tropical cyclone counts under assumption of a Poisson process with gamma distribution. In [14] the authors perform direct simulation from the posterior distribution of multiple change point models with the unknown number of change points based on the recursions. The class of models assumes that the parameters associated with segments of data between successive change points are independent of each other. In [15] a scheme for detecting outliers and change-points for non-stationary time series was proposed. The main feature of this scheme is that the outlier is first detected by the model learned in the first stage which repeats the learning process twice and change point is detected by the learned model in the second one. They applied the scheme to AR models and SDAR algorithm as learning modules which is a variant of maximum-likelihood method for online discounting learning of that model, which is adaptive to non-stationary time series. In [16] The Bayesian method is applied for the analysis of change points for the segmentation of micro-array data with implementation of the Bayesian change point method in linear time.

None of the previous work focuses on the specific problem we study here, combining the wavelet denoising and sequential detection method by studying the calling pattern based on call detail records to detect change points that reflect the human activities.

III. Model and Methodology

In change point detection we need to analyze and classify categorical data, either in an exploratory or in a confirmatory context. Exploratory analysis of such kind of data often has to be with extracting relevant hidden knowledge from large dataset. We need to develop and use robust and flexible classification methods. In this paper we first use wavelet de-noising method to process the data and then applied the modified method in [3] for detecting change points based on number of calls and call durations.

A. Wavelet Overview

Wavelet is associated with building a model for a signal, system or process with a set of special signals. The wavelet transform provides a time-frequency representation of the signal. The wavelet transform uses multi-resolution technique by which different frequencies are analyzed with different resolutions.

A generalized wavelet in normalized form is defined by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \qquad (1)$$

where $a, b \in R$, $a > 0$, a is scale parameter, b is translation parameter and satisfy:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \ (2) \text{ and } \int_{-\infty}^{\infty} |\psi(t)|^2 \, dt < \infty \ (3)$$

The scale parameter, a, indicates the level of analysis defined as 1/frequency and corresponds to frequency information. Scaling either dilates (expand) or compress a signal.

The wavelet transform calculates wavelet coefficient. The Continuous Wavelet Transform (CWT) is given by

$$W_{a,b} = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt \qquad (4)$$

where $W_{a,b}$ is wavelet coefficients, x(t) is the signal to be transformed, $\psi(t)$ is the mother wavelet or basis function.

The wavelet coefficients are measures of the goodness of fit between the signal and the wavelet. Large coefficients indicate a good fit.

The inverse transform of CWT used to compute original data is given as:

$$x(t) = \sum_a \sum_b W_{a,b}\psi_{a,b}(t) + \sum_a c_{a,b}\phi_{a,b}(t) \qquad (5)$$

where $\phi_{a,b}(t)$ denotes the scaling function, $c_{a,b}$ denotes scaling coefficients which are defined by

$$c_{a,b} = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\phi\left(\frac{t-b}{a}\right)dt \qquad (6)$$

In Discrete Wavelet Transform (DWT) the scale factors between levels are usually chosen to be powers of 2. Widely used a and b parameters are set to $a = 2^j$ and $b = 2^j k$, $j, k \in Z$. For DWT the mother wavelet is defined as:

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \qquad (7)$$

The DWT is given by

$$W_{j,k} = 2^{-j/2}\int_{-\infty}^{\infty} x(t)\psi(2^{-j}t - k)dt \qquad (8)$$

where $W_{j,k}$ is wavelet coefficients, x(t) is the signal to be transformed, ψ(t) is the mother wavelet or basis function.

The inverse transform of DWT used to compute original data is given as:

$$x(t) = \sum_{j=1}^{\infty}\sum_{k\in Z} W_{j,k}\psi_{j,k}(t) + \sum_{j=1}^{\infty} c_{j,k}\phi_{j,k}(t) \qquad (9)$$

where $\phi_{j,k}(t)$ denotes the scaling function, $c_{j,k}$ denotes scaling coefficients which are defined by

$$c_{j,k} = 2^{-j/2}\int_{-\infty}^{\infty} x(t)\phi_{j,k}(2^{-j}t - k)dt \qquad (10)$$

The inverse wavelet transform (IWT) reconstructs the signal from its coefficients.

B. Wavelet Denoising

Generally, for the de-noising the wavelet scaling function should have properties similar to the original signal. The general wavelet de-noising procedure follows 3 steps:

(1) Apply wavelet transform to the noisy signal to produce the noisy wavelet coefficients to the level which we can properly distinguish the change occurrence;

(2) Select appropriate threshold limit at each level and threshold method (hard or soft thresholding) to best remove the noises and (3) Inverse wavelet transform of the threshold wavelet coefficients to obtain a denoised signal.

1. Wavelet Selection

The differences among different mother wavelet functions (e.g. Haar, Daubechies, Coiflets, Symlet, Biorthogonal and etc.) consist in how these scaling signals and the wavelets are defined. The choice of wavelet determines the final waveform shape. To best characterize the change points in a noisy signal, we should select the wavelet to better approximate and capture the transient change points of the original signal. The choice of mother wavelet can be based on correlation between the signal of interest and the wavelet denoised signal.

2. Threshold Selection

After a suitable wavelet basis function is chosen, the DWT decomposition of a signal will compress the energy of the signal into a small number of large magnitude wavelet coefficients. The DWT transforms the Gaussian white noise in any one orthogonal basis into wavelet coefficients of small magnitude. This property of the DWT allows the suppression of noise by applying a threshold which retains wavelet coefficients representing the signal and removes low magnitude coefficients which represent noise.

The waveShrink method [19] is widely used to estimate signal *x*. The two commonly used shrinkage functions are hard and soft thresholding functions defined as

$$S_\delta^{hard}(x) = \begin{cases} x & |x| > \delta \\ 0 & |x| \le \delta \end{cases}$$

$$S_\delta^{soft}(x) = \begin{cases} x-\delta & x > \delta \\ \delta - x & x < -\delta \\ 0 & |x| < \delta \end{cases}$$

where $\delta \ge 0$ is the threshold.

The threshold is usually determined in one of the following four ways.

1) Universal Threshold

The universal threshold is defined as [18]:

$$\delta_U = \sigma\sqrt{2\log(n)}$$

where σ is standard deviation of the noise and n is the sample length. It uses a single threshold for all wavelet coefficients. One drawback is signals of short duration. Other methods may provide more accurate results.

2) SURE Threshold

The SURE threshold is based on Stern's Unbiased Risk Estimator [17]. This method minimizes the SURE function to determine an optimal threshold. The SURE threshold is defined as

$$\delta_{SURE} = \min_\delta SURE(\delta, \frac{x}{\sigma})$$

where SURE() is defined as

$$SURE(\delta, X) = n - 2m_{\{i:|X_i| \le \delta\}} + \sum_{i=1}^{n} [\min(|X_i|, \delta)]^2$$

where δ is the candidate threshold, x_i is the wavelet coefficient , n is the data size, and m is the number of the data point less than δ. This method has the sparse wavelet coefficient problem and often combined with universal threshold in a hybrid method.

3) Hybrid Threshold

The hybrid threshold method is a combination of the universal and SURE threshold methods [18]. This method uses the universal threshold if the signal to noise ratio (SNR) is low with sparse wavelet coefficient, otherwise uses SURE threshold method.

4) Minmax Threshold

The Minmax threshold method uses a fixed threshold selected to produce minmax performance for the mean square error (MSE). The Minmax uses a single threshold for all wavelet coefficients. It is defined as [19]

$$\delta_{\min i \max} = \inf_\delta \sup_\mu \left\{ \frac{R_\delta(\mu)}{n^{-1} + \min(\mu^2, 1)} \right\}$$

where $R_\delta(\mu) = E(S_\delta(x) - \mu)^2$, $x \sim N(\mu, 1)$.

C. Change Point Detection

Change point analysis method attempts to find a point along a distribution or trend of values where the characteristics of the values before and after the point are different.

Let $X = (X_1, X_2, ..., X_\theta)$ be a process.

The multiple change points for the process is defined as [3]

$X_1 = (x_1, ..., x_{\tau_1}) \sim f_1$

$X_2 = (x_{\tau_1+1}, ..., x_{\tau_2}) \sim f_2$

....

$X_\theta = (x_{\tau_{\theta-1}+1}, ..., x_{\tau_\theta}) \sim f_\theta$

where $f_1, f_2, ..., f_\theta$ are either known or unknown probability density functions or trends; $\tau_1, \tau_2, ..., \tau_{\theta-1}$ are change points.

We first use wavelet denoising method to pre-process the data, and then use sequential estimation scheme in [3] for detecting multiple change points which chooses increasing subsamples and finds one change point at a time until all change points are found.

The general procedure for change point detection is as follows.

1) sequential detection

A widely used change point detection method based on Page's cumulative sum (cusum) rule is defined by

$$T(a) = \inf\{n : S_n \ge a\}$$

where

$$S_n = \max_{1 \le k \le n} \sum_{i=k+1}^{n} \log \frac{g(x_i)}{f(x_i)}$$

is maximum likelihood ratio based cumulative sum, *a* is a threshold, f and g are pre- and post-change density functions. When the density functions are unknown, the best estimates for each value k of a change point are used, the cusum is computed by

$$\hat{S}_n = \max_{1 \le k \le n} \sum_{i=k+1}^{n} \log \frac{\hat{g}(x_i)}{\hat{f}(x_i)}$$

and the stopping rule is defined by [1]

$$\hat{T}(a) = \inf\{n : \hat{S}_n \ge a\}$$

The call log data is not classical change point model type. We use the method in [3] to detect trend changes, either linear or exponential, in each segment between change points. The computation of the above cusum change point detection method will be more complicated if there are unknown parameters in segments [3].

We use leased-squares method for the nonlinear trends. The trend which is linear or exponential is decided by the lower sum of squares for each segment. In this way the maximum likelihood ratio is replaced by the minimum weighted sum of squared residuals

$$\hat{S}_n = \min_k \{\sum_{i=1}^{k} \sqrt{\hat{f}_i}(x_i - \hat{f}_i) + \sum_{i=k+1}^{n} \sqrt{\hat{g}_i}(x_i - \hat{g}_i)\}$$

and the stopping rule is defined as

$$\hat{T}(\alpha) = \inf\{ n : p_n \le \alpha \}$$

where p_n is a p-value testing significant based on $\hat{S}_n$ of a change point at k.

2) post-estimation

The detected change point must be estimated after it is detected by a stopping rule. The change point estimator we use is based on the cusum stopping rule $\hat{T}$ and the minimum p-value

$$\hat{\tau} = \arg\min_{1 \le k < \hat{T}} p(k, \hat{T}, X)$$

where $p(k, \hat{T}, X)$ is the p-value of the likelihood ratio test comparing $X_1 = (x_{1,...,}x_k)$ and $X_2 = (x_{k+1,...,}x_{\hat{T}})$.

3) significance tests

To eliminate false change point, we use ANOVA F-type tests. If the test is significant, we repeat step 1-3 to search for next change point, else, it is a false change point and the search continues based on initial sequence after the last significant change point.

For fitting linear trend $E(x_i) = a + bt_i$, we use the standard least square estimates

$$\hat{b} = \frac{\sum_i (x_i - \bar{x})(t_i - \bar{t})}{\sum_i (t_i - \bar{t})^2} \quad \text{and} \quad \hat{a} = \bar{x} - \hat{b}\bar{t}$$

and for fitting exponential trend $E(x_i) = \exp(a + bt_i) - 1$, we use

$$\hat{b} = \frac{\sum_i \log(1 + x_i)(t_i - \bar{t})}{\sum_i (t_i - \bar{t})} \quad \text{and} \quad \hat{a} = \overline{\log(1 + x)} - \hat{b}\bar{t}$$

for initial approximation.

When the preliminary estimator of a change point is obtained, a refinement of this estimator is performed by least square fitting from the segment in the neighborhood of the preliminary estimator. If the change points are not significant for the chosen level α, they are removed and the corresponding segments are merged.

After the iterations end, all the change points are significant at the chosen level α.

D. Real-life Data Sets and Parameters

***Real-life traffic profile*:** In this paper, actual call logs are used for analysis. These actual call logs are collected at MIT [21] by the Reality Mining Project group for a period of 8 months. This group collected mobile phone usage of 100 users including their user IDs (unique number representing a mobile phone user), time of calls, call direction (incoming and outgoing), incoming call description (missed, accepted), talk time, and tower IDs (location of phone users). These 100 phone users are students, professors and staff members. The collection of the call logs is followed by a survey of feedback from participating phone users for behavior patterns such as favorite hangout places, service providers, talk time minutes, phone users' friends, relatives and parents. We used this extensive dataset for our event analysis in this paper. More information about the Reality Mining Project can be found in [21]. The following three measurements are considered for the experiments conducted here.

Day of week: Everyone has his/her own schedule for working, studying, entertainment, traveling and so on. The schedule is mainly based on the day of the week.

Call frequencies: The call frequency is the number of incoming or outgoing calls in a period of time. The more the number of incoming or outgoing calls in a period of time, the more socially close the caller and callee relationship.

Call duration: The call duration is how long both caller and callee want to talk each other. The longer the call duration is in a period of time, the more socially close the caller and callee relationship.

We select Coiflets5 wavelet and Minimax threshold method to denoise the data by the principles described above, and then apply the sequential change point detection method in [3].

We use both simulation data and the real data from the data set of four months, a semester since the communication members were relatively less changed in a semester for students.

The simulation data sets are randomly generated based on $X = (X_1, X_2, ..., X_\theta)$, $X_i \sim N(\mu_i, \sigma_i^2)$ for i=1, 2, …, θ.

IV. Experimental Results and Discussion

First we conduct simulation using synthetic data by the above method, and then apply this method to real data.

A. Simulation

Many workers have weeks when they are very busy (we refer to them as "busy weeks") and others where the work to be performed is less than the regular load (we refer to them as "easy weeks"). The expectation is that during the busy weeks the worker makes and receives less number of calls and has shorter talk time than those of usual weeks. During the easy weeks there is a tendency that he/she will make and receive more number of calls with longer talk time than those of usual weeks. We randomly generated multiple simulation data sets of number of calls and call duration for 120 days for simulation. Fig. 1 and 2 are two of them.

Fig. 1 and 2 show the change points for the number of calls and call duration of dataset1 and dataset2, where the x-axis indicates the days and y-axis indicates the number of calls and call duration (minutes) respectively. The blue color dotted curve and green color solid curve indicate original and denoised ones, respectively. The vertical lines indicate the change points. The three detected change points happened on the 56^{th}, 60^{th} and 84^{th} days which match the change points of the curves.

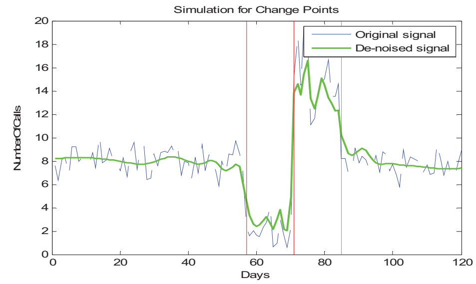

Fig. 1 The change points based on number of calls for simulation data.

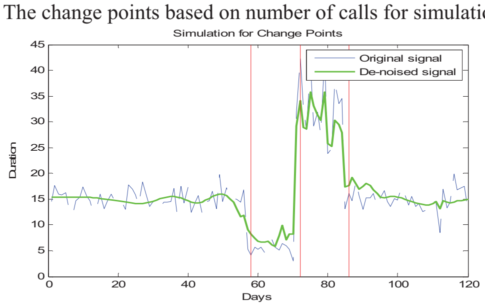

Fig. 2 The change points based on call duration for simulation data.

B. Real Data

After multiple experiments, we select Coiflets5 wavelet and Minimax threshold method by the principles described in Section 2 to denoise the data, and then apply the sequential change point detection method in [3].

Fig. 3 and 4 show the change points for the number of calls and call duration of user3, where the x-axis indicates the days and the y-axis indicates the number of calls and call duration respectively. The dotted curve and the solid curve indicate, respectively, original and de-noised data. The vertical lines indicate the change points. There are 3 change points identified at the 32^{nd}, 48^{th}, and 64^{th} day which correspond to Friday, Sunday, and Sunday respectively.

From the 1^{st} day to the 32^{nd} day, the user3 visited to New York City, world trade center and Harvard University. The average of number of calls is 8 and the average of duration is 7.5 minutes per day. Between the 32^{nd} and 48^{th} day, the user's activities were at local areas. The average of number of calls is 12 and the average of duration is 10 minutes per day. From the 48^{th} day to the 64^{th} day, the user's activities were at local areas. The average of number of calls is 9 and the average of duration is 6 minutes per day. From the 64^{th} day to the 108^{th} day, the user's activities were at local areas. The average of number of calls is 2 and the average of duration is 2 minutes per day.

From the results we found that most of change point days are associated with weekends when most people tend to have some leisure time and which consequently changes their calling pattern. Although a sophisticated technique is not needed to find out change points over the weekends it should be noticed that our goal here is to show that our technique can identify change points based on call detail records. The actual change point can represent other behaviors, for example, if the individual under observation is a potential threat to public security.

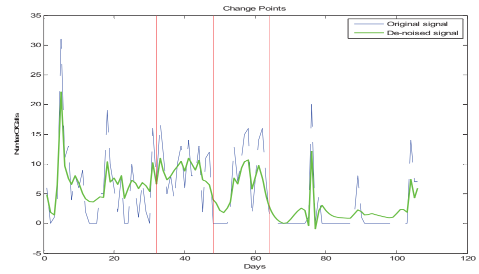

Fig. 3 The change points based on number of calls for user3.

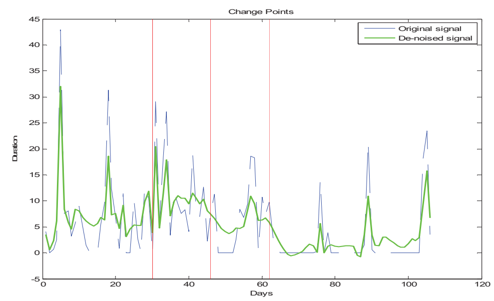

Fig. 4 The change points based on duration for user3.

V. Performance Analysis

To evaluate the accuracy of our method, we conducted experiments for randomly choosing 20 phone users from actual call logs of 100 phone users. The significant level and the latency are two important measurements when considering the accuracy in the change point detection. The lower the significant level, the more sensitive is the strategy. This can avoid false positives. In all data sets we used a significant level of 0.01.

The shorter the latency, the better is the indication of the accuracy of the strategy. So a minimization of the latency is desired. In this study a minimum of 10 data points were chosen to start the detection of a new change point. A total of 81 change points were identified for 20 users. 78% of the change points were identified using about 12 data points and 10% using about 15 data points shown in Fig. 5. In only a very few cases, the change points were identified later than the 20^{th} points as shown in Fig. 5. This is a very good result of the accuracy of the strategy.

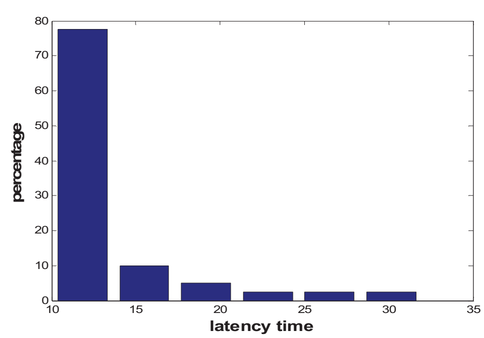

Fig. 5 Latency in the change point detection for 20 users.

VI. VALIDATION

To evaluate the accuracy of our model, we used actual call logs of 100 phone users and randomly choose 20 phone users. These users include students, professors and staff members. The best way to validate the results is to contact the phone users to get feedback. But because of the privacy issues it is almost impossible to obtain such information. Therefore, in order to validate our model, we hand-labeled the change points based on the number of calls, duration of calls in the day, history of call logs, location, time of arrivals, and other humanly intelligible factors. The results based on the hand-labeled data are a good indication of the accuracy of the proposed approach.

VII. CONCLUSION

In this paper we combine wavelet denoising and the sequential change point detection method for detecting change points based on mobile phone call detail records. The data is pre-processed to reduce noise influence and the de-noised data is used for the sequential change point detection to identify any new pattern in the call log data.

This work is useful for enhancing homeland security, detecting unwanted calls (e.g., spam), communication presence, and marketing among other applications. The detection of change points in the calling pattern is an indication of the occurrence of an event in someone's life. The knowledge about certain events can be used, for example, for security aspects. Hearing all conversations of all suspects in all potential terrorist attack is almost impossible. However, the efforts can be more concentrated if a change in behavior is observed for a certain group or individual. The experimental results show that our model can indeed detect certain events based on the observation of existing calling patterns. The use of both synthetic data as well as real data has shown promising results with respect to performance (number of detected change points) and with accuracy (number of false positives and latency).

In our future work we plan to detail the change point classification, analyze and use some criterion to optimize the detection process. The validation procedure is also going to be expanded. The simulation data accounted for only two events ("busy" and "easy" week) and more events will be added to diversify the synthetic data. Also, due to the tedious and time consuming effort required for hand-labeling only twenty individuals have been selected for the experiments conducted here. We plan to do the hand-labeling for all elements to further validate the proposed approach.

ACKNOWLEDGMENT

We would like to thank Nathan Eagle and Massachusetts Institute of Technology for providing us the call logs of Reality Mining dataset.

This work is supported by the National Science Foundation under grants CNS- 0627754, CNS-0516807, CNS-0619871 and CNS-0551694. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] M. Baron. Nonparametric adaptive change-point estimation and on-line detection. *Sequential Analysis*, 19:1–23, 2000.

[2] M. Baron and N. Granott. "Consistent estimation of early and frequent change points". In Y. Haitovsky, H. R. Lerche, Y. Ritov, eds., Foundations of Statistical Inference, Proceedings of the Shoresh Conference 2000, pages 181–192, Springer, 2003.

[3] J. W. Cangussu, M. Baron. "Automatic identification of change points for the system testing process" *The 30th Annual IEEE International Computer Software and Applications Conference (COMPSAC 2006)* Chicago, IL, Sept. 18-21, 2006

[4] A. E. Raftery and V. E. Akman. "Bayesian analysis of a Poisson process with a change-point". *Biometrika* 73: 85–89, 1986.

[5] T. Y. Yang and L. Kuo. "Bayesian binary segmentation procedure for a Poisson process with multiple change points". *Journal of Computational and Graphical Statistics* 10: 772–785, 2001.

[6] Y. Ritov, A. Raz, and H. Bergman. "Detection of onset of neuronal activity by allowing for heterogeneity in the change points". *Journal of Neuroscience Methods* 122: 25–42, 2002.

[7] B. P. Carlin, A. E. Gelfand and A. F. M. Smith. "Hierarchical Bayesian analysis of change point problems". *Applied Statistics* 41: 389-405, 1992.

[8] R. Lund and J. Reeves. "Detection of undocumented change points: A revision of the two-phase regression model". *Journal of Climate* 15, 2002: 2547–2554.

[9] D. M. Hawkins, "Testing a sequence of observations for a shift in Location", *Journal of the American Statistical Association* 72, 1977, pp. 180-186.

[10] K. J. Worsley, "On the likelihood ratio test for a shift in location of normal populations", *Journal of the American Statistical Association* 74, 1979, pp. 365-367.

[11] J. Chen and A. K. Gupta. "Testing and locating change points with application to stock prices". *Journal of the American Statistical Association* 92: 739–747, 1997.

[12] T. D. Johnson, R. M. Elashoff and S. J. Harkema. "A Bayesian change-point analysis of electromyographic data: detecting muscle activation patterns and associated applications". *Biostatistics* 4: 143–164, 2003.

[13] P. Chu and X. Zhao. "Bayesian Change-Point Analysis of Tropical Cyclone Activity: The Central North Pacific Case". *Journal of Climate*, vol. 17, pp.4893-4901, 2004.

[14] P. Fearnhead. "Exact and efficient Bayesian inference for multiple Change point problems". *Stat Comput* 16: 203–213, 2006

[15] K. Yamanishi, J. Takeuchi "A unifying framework for detecting outliers and change points from time series". *IEEE transactions on knowledge and data engineering*, vol. 18, no.4, pp. 482-492, April, 2006.

[16] C. Erdman and J. W. Emerson. **"**A fast Bayesian change point analysis for the segmentation of micro-array data**"** Bioinformatics 2008 24(19):2143-2148

[17] C. Stern. "Estimation of the Mean of Multivariate Normal Distribution". *The Annals of Statistics*. vol. 9, pp. 1135-1151, 1981.

[18] D. Donoho. and I. Johnstone. "Adapting to Unknown Smoothing via Wavelet Shrinkage". *J. Amer. Statist. Association,* vol. 90, pp. 1200-1224, 1993.

[19] D. Donoho and I. Johnstone. "Ideal Spatial Adaptation by Wavelet Shrinkage". *Biometrica,* vol. 81, pp. 425-455, 1994.

[20] I. Johnstone and B. W. Silverman. "Wavelet Thresholding Estimators for Data with Correlated Noise". *Journal of Roy. Stat. Soc.*, Series B, vol. 59, pp. 319-351, 1997.

[21] Massachusetts Institute of Technology: Reality Mining. http://reality.media.mit.edu/ (2008)

[22] H. Zhang, R. Dantu. "Quantifying the presence for phone users". In *Proceedings of the fifth IEEE Consumer Communications & Networking Conference* 2008.