Change Point Detection Using Multivariate Exponentially Weighted Moving Average (MEWMA) for Optimal Parameter in Online Activity Monitoring

Naveed Khan[1(✉)], Sally McClean[1], Shuai Zhang[2], and Chris Nugent[2]

[1] School of Computing and Information Engineering, Ulster University, Coleraine, UK
khan-n5@email.ulster.ac.uk, si.mcclean@ulster.ac.uk
[2] School of Computing and Mathematics, Ulster University, Jordanstown, UK
{s.zhang, cd.nugent}@ulster.ac.uk

Abstract. In recent years, wearable sensors are integrating frequently and rapidly into our daily life day by day. Such smart sensors have attracted a lot of interest due to their small sizes and reasonable computational power. For example, body worn sensors are widely used to monitor daily life activities and identify meaningful events. Hence, the capability to detect, adapt and respond to change performs a key role in various domains. A change in activities is signaled by a change in the data distribution within a time window. This change marks the start of a transition from an ongoing activity to a new one. In this paper, we evaluate the proposed algorithm's scalability on identifying multiple changes in different user activities from real sensor data collected from various subjects. The Genetic algorithm (GA) is used to identify the optimal parameter set for Multivariate Exponentially Weighted Moving Average (MEWMA) approach to detect change points in sensor data. Results have been evaluated using a real dataset of 8 different activities for five different users with a high accuracy from 99.2 % to 99.95 % and G-means from 67.26 % to 83.20 %.

Keywords: Multiple change points · Activity monitoring · Genetic algorithm · Accelerometer

1 Introduction

In the current era, the world is changing very fast in almost every aspects of life. Hence, the capability to detect, adapt and respond to change performs a key role in all aspects of life. The number of real world problems such as fault detection and diagnosis (monitoring) [1], quality control [2], natural catastrophic event prediction like earthquakes [3] and monitoring of context aware systems requires sequential detection of a change in the process. Activity monitoring is a key element of context aware systems. The primary use of such system within healthcare is to detect daily life activities and to monitor these over time. Body movement can be captured through wearable sensors such as accelerometers, gyrometers etc. to detect different transitions of movement from one

C.R. García et al. (Eds.): UCAmI 2016, Part I, LNCS 10069, pp. 156–165, 2016.
DOI: 10.1007/978-3-319-48746-5_16

activity to another for activities such as walk, sit, stand, and run. The key issue is often the detection of abrupt change points from one activity level to another in real time systems [4]. Such abrupt change points can be very fast with respect to the sampling period of measurement. In real time system, online change detection algorithms are used to observe, monitor and evaluate data as soon as it becomes available. The objective of an online change detection algorithm is to detect a change in a sequence of random observations from a probability distribution. Also, online change detection algorithms are generally required to be fast and minimize false alarms. In activity monitoring such automatic change point detection is still a challenging task for researchers. The sensor data arrives continuously as sequential and fast data streams. Therefore, the algorithm should be lightweight to detect changes in sensor data and work efficiently under resource limited constraints [5]. Such in-time response can be useful in various scenarios such as to generate real world datasets or observing patient vital signs for example heart rate against various activities [6]. The detection problem can be evaluated on the basis of relative time constants of the process to be monitored on the sampling data. Moreover, the probability distribution over past and present intervals of time series data can be compared using change point detection techniques. However, specific strategy use in different techniques for detection of a particular change point to prompt an alarm as two distributions becomes significantly different [7].

In this paper, we evaluate our proposed technique Multivariate Exponentially Weighted Moving Average (MEWMA) on detecting multiple changes in activities on data collected from real sensor for various subjects on a number of different activities. In [8], we tuned different parameters of MEWMA manually to detect the change point between two user activities. However, in current work, we evaluate the algorithm for detecting multiple changes in user activities for various subjects. Also, the MEWMA incorporates an appropriate fitness function using a GA to automatically identify the optimal set of parameters for detecting multiple changes in the sensor data. The various parameters such as λ, window size and significance value are evaluated using genetic algorithm to find the optimal parameter set for MEWMA. The performance of the extracted optimal parameter for accurate change point detection were analyzed using different metric measures like accuracy, specificity, precision, sensitivity and G-Means. The remainder of this paper is structured as follows: in Sect. 2 Related work is presented. In Sect. 3 we provide an overview of our proposed approach and the experimental setup with results presented in Sect. 4. Finally, Conclusion and Future Work are presented in Sect. 5.

2 Related Work

In a real time scenario, the data is evolve continuously over time and can be analyzed adequately and appropriately as it become available. Nowadays, smart devices are used more frequently for online data collection in various domains. However, online data learning and evaluation of such data create considerable challenges for associated learning algorithms. Moreover, accurate change detection and in-time decision making from observed data is still an important problem that needs to be addressed. A numbers of algorithms have been discussed in the literature to detect changes in health sensor data. The Hidden Markov Models (HMM) have been used in [9] to detect change in streaming data. In this approach, an appropriate threshold value has been used for automated change detection. In the first step, the interrelationship is formed between two data streams through a time-invariant sequence of linear dynamic model. In the second step, the estimated parameters are modelled using HMM to evaluate the likelihood ratios for the new parameters. Finally, a change is flagged if the likelihood ratio is less than the given threshold. This technique requires quick probability estimation in two consecutive windows to detect changes in streaming data which comes from different distribution. The authors [10] have used a semiparametric log-likelihood criterion (SPLL) to detect change in multivariate streaming data. Additionally, the two well-known criterion Kullback-Leibler (K-L) distances and Hotelling's T-square have also been used together with SPLL. The experiments were performed on 30 real datasets for detecting change. The results have shown that SPLL performs better than K-L and Hotelling's T-square on both normalized and un-normalized data. A reactive clustering algorithm have been proposed in [11] to detect change in multivariate streaming data. The two overlapping windows were used to identify the change. The first window used a reference window to form a cluster and the second window used to capture the new incoming data. The distance between the incoming data and centre of the cluster is then calculated, and if it is greater, the new data point considered as a change. A drawback of this technique is the proper selection of window size according to the corresponding data. Moreover, the feature space, hypersphere and hypersphere radius have been used in support vector change point detection (SVCPD) algorithm [12] to identify the location and to detect change in the data stream. The advantage of SVCPD algorithm is to monitor and analyze each data point, and also compare the distance with current hypersphere models to accurately classify change points in the data stream. The authors have used Cumulative Sum Control Chart (CUSUM) [13] to monitor and detect small shifts in cardiovascular events using the process mean. Also, a number of primary methods have been used such as a process control approach, biometric methods and an online recognition approaches to analyze and evaluate physiological monitoring. Such multivariate analyses are crucial to investigate because the problem involves more than one variable which are correlated and observed simultaneously. Moreover, numerous parameter tuning has also been used to improve monitoring and detection of changes in user activity. Optimization is used for tuning input parameters to find the best solution from all feasible solutions. In the literature, the Genetic algorithm (GA) [14] has been used extensively for optimization problems to find and identify the optimum solution. The analysis of the literature review reflects that most algorithms requires prior knowledge of the characteristics of change points and the data stream's underlying distribution(s) which can be ineffective for our target application.

3 The Proposed Approach

3.1 Multivariate Exponentially Weighted Moving Average (MEWMA) Change Point Detection Algorithm

The Multivariate Exponentially Weighted Moving Average (MEWMA) is a statistical control method to monitor simultaneously two or more correlated variables and also provide sensitive detection of small and moderate shifts in time series data. The MEWMA statistic incorporates information of all prior data including historical and current observation with a user-defined weighted factor [15, 16]. Moreover, MEWMA can be used to detect shift of any size in the process. The MEWMA has achieved better performance to detect small and moderate changes than other multivariate control chart like the T-Square control chart [17]. It is described by the following equation.

$$\mathbf{Z}_i = \Lambda\mathbf{X}_i + (1 - \mathbf{\Lambda})\mathbf{Z}_{i-1} \qquad i = 1, 2, 3\ldots n \tag{1}$$

where $\mathbf{Z}_i$ is the i^{th} MEWMA vector, $\mathbf{\Lambda}$ is the diagonal matrix with elements $\boldsymbol{\lambda}_i$ for $i = 1, \ldots, p$ where p is the number of dimensions and $0 < \boldsymbol{\lambda}_i \leq 1$, and $\mathbf{X}_i$ is the i^{th} input observation vector, $i = 1, 2, 3\ldots n$. The out-of-control signal is defined in Eq. 2.

$$\mathbf{T}_i^2 = \mathbf{z}_i' \Sigma_i^{-1} \mathbf{z}_i < h \tag{2}$$

where $\sum_i$ is the variance covariance matrix of $\mathbf{Z}_i$ and h (>0), chosen to achieve a specified in-control signal. The analysis of MEWMA is used to evaluate and monitor simultaneously two or more correlated variables and the inter-relationship among these variables. In multivariate analysis, the data points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\ldots\mathbf{x}_n$ is a subsequence of a data stream where n is the length of the subsequence. The data points in the data stream may be from various distributions, for example, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\ldots\mathbf{x}_{i-1}$ and $\mathbf{x}_1, \mathbf{x}_{i+1}\ldots\mathbf{x}_n$ can be from distributions D1 and D2 respectively. The algorithm's objective is to find and evaluate the location of the change points i in the time series data. MEWMA is applied to each data stream to calculate the exponentially weighted moving average of multivariate input observations in order to evaluate and find the location for change points. In our experiments, we incorporate sliding windows to perform sequential analysis of input data incremented by one within each window. Likewise, the MEWMA vector is calculated using input vectors and is represented by $\mathbf{Z}_i$ as shown in Eq. 1. Moreover, in order to find the T-squared, the variance-covariance matrix of $\mathbf{Z}_i$ is calculated recursively and is represented by $\sum_i$ as shown in Eq. 2. However, the significance value is used to classify the confidence of the entire window.

3.2 Genetic Algorithms

The Genetic Algorithm (GA) is a well-known heuristic search algorithm that inspired by the process of evolution in nature. The GA starts with a random number of variables that intelligently exploit the random search of individual solutions to solve the optimization problems. The GA searches to identify the fittest value among the individuals over successive generation of solutions [14]. Optimization is used for tuning input parameters to find the best solution form all feasible solutions. The GA used fitness function to find the optimal solution for a system. In the current scenario of our proposed work, the different combination of the three variables, namely λ, window size and significance value, is used to identify a single point in the population. The individual solutions “evolve” over consecutive generation to find the optimal solution. The fitness function is used in each iteration by the GA to evaluate the quality of all the proposed solutions to the problem in the current population. The fitness function evaluates how good a single solution in a population. The fitness function is the core component of a GA, which identifies the optimal fitness value after evaluating each individual in the population. In our fitness function, we initialize the population of vectors whose elements contains the λ value, window size and significance value. Here, our fitness function is chosen as G-means with a given range of input values. G-means is used as the measure to find the ratio and overall efficiency of an activity by combining sensitivity and specificity. It is defined as follows:

$$\mathrm{G}\_means_{max} = max_{(\lambda, win\_size, sig\_value)}(G\_means_{MEWMA}) \tag{3}$$

where, λ ranges from 0.1 to 1 for each activity with corresponding significance values of 0.05, 0.01, 0.025, 0.005 and window sizes of 1 s, 2 s, 3 s, 4 s, 5 s and 6 s.

The following GA parameters are used in our experiments to maximize our fitness function and identify the optimal parameter set with maximum accuracy as shown in Table 1.

Table 1. GA Parameters

Parameters	GA
Population size	50
Selection	Stochastic uniform
Crossover rate	0.8
Mutation	Gaussian
Crossover	Heuristic
Generations	200

Our proposed model used Eq. 3 as the fitness function by initializing upper and lower bounds of the three parameters to find the maximum G-means with the optimal parameter set. After exploration with different GA parameter setting, the optimal GA parameter settings were chosen as shown in Table 1. These GA parameters are then used to maximize our fitness function and return the optimal best parameters with maximum G-Means.

4 Experimental Setup

A real accelerometer dataset was used to perform quantitative evaluation of change detection algorithm in human activity. The data set was collected from 5 participants consists of 3 males and 2 females. The shimmer sensing platform [18] integrated with three-axis accelerometer was used to collect the data. Each participant wore an accelerometer on the right ankle to collect the data from the accelerometer signals for various activities with a sample frequency of 102.4 Hz. The sampling rate is chosen by the Shimmer sensing platform. Also, the storage is supported by the Shimmer sensing platform and the data is transferred periodically to the computer via Bluetooth communication protocol. The change points detection and the evaluation were implemented in Matlab. To facilitate future work, additional shimmers were also placed on the participants' sternum and lower limb to facilitate evaluation and analysis of the optimal sensor placement for detecting changes in human activity.

The total eight activities were performed by each participant as presented in Table 2.

Table 2. Overview of activities in dataset

Activity Seq.	Label	Type	Description
1	Sit	Static	Sit for 5 min (m)
2	Sleep	Static	Lie on sofa for 5 m
3	Stand	Static	Standing still for 5 m
4	Stand to walk	Transitional	Stand for 10 s and walk for 5 s
5	Walk	Dynamic	Walk on treadmill at constant speed of 5 m
6	Run	Dynamic	Run on treadmill at constant speed of 5 m
7	Watch TV	Static	Sit on sofa for 5 m
8	Vacuum	Dynamic	Vacuum for 5 m

Each activity was carried out in the predefined order above with a rest period of one minute observed between each activity. The different activities were classified as either static, representing that each participant was asked to remain comfortably still such that small natural movements were allowed, or dynamic representing purposeful human movement. The change points were manually labelled in a controlled environment by a human expert. Moreover, if a participant was unable to complete the tasks in succession they were allowed to rest with the start and end time recorded and the relevant sensor data subsequently removed from the dataset. The resultant dataset contains a continuous data stream of approximately 35 min for each participant with the activities performed according to the order in Table 2. There are 7 labelled changes for each participant as a result of the change of activities.

4.1 Results Evaluation

This section evaluates the performance of the MEWMA for change detection where GA was applied on the real dataset to identify the optimal parameter set. The MEWMA algorithm is applied and the calculated vector is used to detect changes in the data stream. The different performance measures such as accuracy, precision, sensitivity and specificity and G-means were used to evaluate change point detection in activity monitoring. The detected change point is classified as true if in the data the index i, $i \in \{z - (f/4), \ldots\ldots z + (f/4)\}$ where f is the sampling frequency in Hz and z indicates the index of manually label change in the data stream. The target of our proposed approach is to find and detect the primary change point in different activities such as sit, stand, walk, sleep, run, vacuum and watching TV as shown in Fig. 1.

The x, y and z acceleration magnitude is calculated form input observation and used as input to MEWMA. The different parameters of MEWMA were analyzed initially; λ (0.1 to 1), window size (1 s, 2 s, 3 s, 4 s, 5 s. 6 s) and significance values (0.005, 0.01, 0.025, 0.05) in order to evaluate and find the accurate change point. Moreover, following this, in offline mode, the GA is used to find the optimal parameter set for MEWMA.

The positive and negative detection is defined as; true positive (TP) is the correctly classified change points, false positive (FP) is non-transitional point but algorithm detect it as a change. The True negative (TN) is the non-transitional points and not labelled as change while false negative (FN) is the transitional change points missed by detection algorithm. The real dataset example of subject 3 for change detection using MEWMA for different activities are shown in Fig. 1.

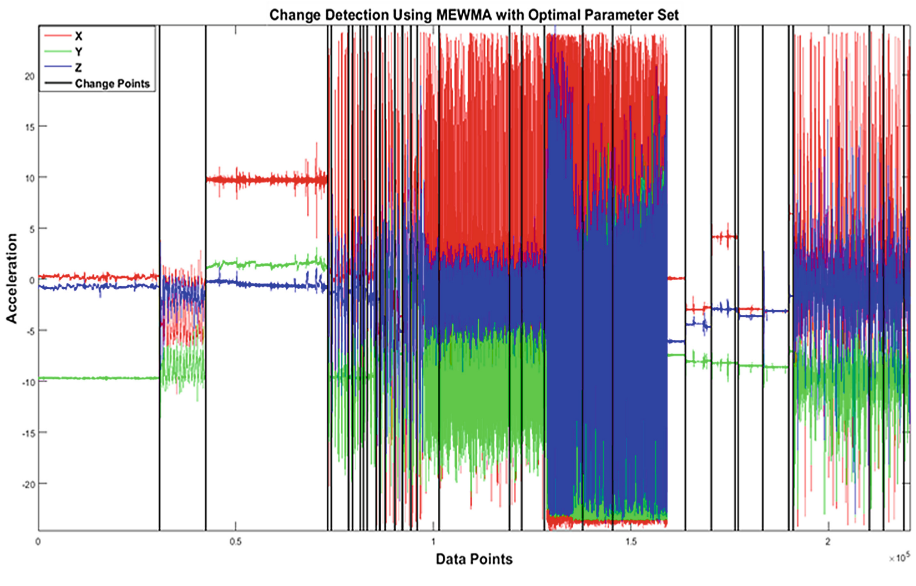

Fig. 1. The real dataset example of subject 3 for change point detection using MEWMA in different activities. The vertical black lines identify change points detected by our algorithm.

Table 3 presents the results of the model performances and the best combination of optimal parameter set (λ, window sizes and significance values) for MEWMA using the GA for 5 subjects performing 8 different activities. The values presented here are the optimal values derived from the GA that optimize our fitness function.

The accuracy, precision, specificity, sensitivity and G-means metrics were used for evaluation of optimal parameter selection for MEWMA algorithm. The accuracy is the ratio of correctly classified data point over the total data points. However, the precision is the ratio of true positive over true positive plus false positive.

The highest accuracy achieved is about 99.2 % to 99.5 % of window size (3 s, 4 s and 5 s), λ (0.6 & 0.7) and p = 0.05 for the optimal parameter set using GA of all subjects with 8 different activities. However, the maximum precision range from 20 % to 33 % for the same optimal set of parameters as presented in Table 3. The reason for low precision is due to the high number of occurrences of false alarms as our algorithm is very sensitive and detects possible change points even if they are small. Likewise, the specificity and sensitivity are used to find and measure the proportion of correctly classified negative and positive detection in the data.

The highest specificity and sensitivity achieved is about 91.50 % to 97.50 % and 55 % to 71 % of window size (3 s, 4 s and 5 s), λ (0.6 & 0.7) and p = 0.05 for the optimal parameter set using GA as presented in Table 3. The specificity is used to identify the performance of an algorithm with regard to accurate prediction. The results show that our proposed approach achieved higher specificity and made accurate prediction. However, the sensitivity was dropped slightly because our algorithm missed few true change points which were classified as false negative. Moreover, G-means is the combination of sensitivity and specificity and is used to find the ratio of positive and negative accuracy of the data and calculated using Eq. 4.

$$\sqrt{sensitivity \times specificity} = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \tag{4}$$

The highest G-means achieved is about 67.26 % to 83.20 % of the window size (3 s, 4 s and 5 s) and p = 0.05 for the optimal parameter set using GA as presented in Table 3. The G-means are used as fitness functions for the GA to find the optimal parameter set for MEWMA algorithm.

Table 3. MEWMA optimal parameter set using GA for 5 different subjects on real dataset

Optimal Parameters by GA				Model Performances				
Participants	Win Size	Significance value	λ	Accuracy %	Precision %	Specificity %	Sensitivity %	G-Means %
Subject1	3 s	0.05	0.7	99.92	25	91.50	55	70.94
Subject2	4 s		0.7	99.94	20	90.50	50	67.26
Subject3	5 s		0.6	99.95	33	97.50	71	83.20
Subject4	4 s		0.6	99.94	30	96.25	62	77.25
Subject5	5 s		0.7	99.95	28	95.75	57	73.90

Overall, we have achieved good results for different metrics except for the low precision. Our focus is to find the change points so our proposed approach detects every potential change point which can cause a relatively large number of false alarms resulting in low precision. This might not be a major issue for us but we still plan to address it further in future work. This issue of low precision is partly due to the class imbalance problem in our dataset, which we plan to explore in our future work via the online Bagging and Boosting algorithm [19]. Hopefully such an approach can improve the precision while maintaining the good performance on our performance metrics.

5 Conclusion

The GA is used to identify an optimal parameter set for MEWMA to successfully detect multiple change points in different user activities. The different parameters of MEWMA are optimized using the GA to find the maximum G-means and identify the optimal parameter set for each subject's activities. The optimal parameter set has achieved good results for most performance metric considered. The potential of the algorithm is to adjust the individual changes and learn through time irrespective of the individual patterns. The collection of data is expansive and time consuming. Moreover, the real data is used in our experiments and trying to evaluate the prototype of the algorithm that gives the promising results which opens up opportunities to be tested on more subjects. In the future we will explore the low precision problem and try to improve it. Also, the class imbalance problem will also be analyzed for the existing dataset. Moreover, different datasets will be used for evaluation with multiple change points.

References

1. Fabien Meinguet, E.S., Kestelyn, X., Mollet, Y., Gyselinck, J.: Change-detection algorithm for short-circuit fault detection in closed-loop AC drives. IET Electr. Power Appl. **8**, 165–177 (2014)
2. Evan, R.H.W., Brooks, B., Thomas, V.A., Blinn, C.E., Coulston, J.W.: On-the-Fly massively multitemporal change detection using statistical quality control charts and landsat data. IEEE Trans. Geosci. Remote Sens. **52**, 3316–3332 (2014)
3. Brunner, D., Bruzzone, L., Lemoine, G.: Change detection for earthquake damage assessment in built-up areas using very high resolution optical and SAR imagery. In: 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2010, pp. 3210–3213 (2010)
4. Basseville, M., Nikiforov, I.V.: Detection of Abrupt Changes: Theory and Application, vol. 104. Prentice Hall, Englewood Cliffs (1993)
5. Gama, J., Gaber, M.M.: Learning from Data Streams: Processing Techniques in Sensor Networks, vol. 21. Springer, New York (2007)
6. Cleland, I., Han, M., Nugent, C., Lee, H., Zhang, S., McClean, S., Lee, S.: Mobile based prompted labeling of large scale activity data. In: Nugent, C., Coronato, A., Bravo, J. (eds.) IWAAL 2013. LNCS, vol. 8277, pp. 9–17. Springer, Heidelberg (2013)
7. Darkhovski, B.S.: Nonparametric methods in change-point problems: a general approach and some concrete algorithms. In: Carlstein, E., Muller, H.-G., Siegmund, D. (eds.) Change-point Problems, vol. 23, pp. 99–107. Institute of Mathematical Statistics, Hayward (1994)

8. Khan, N., McClean, S., Zhang, S., Nugent, C.: Parameter optimization for online change detection in activity monitoring using multivariate exponentially weighted moving average (MEWMA). In: Ubiquitous Computing and Ambient Intelligence. Sensing, Processing, and Using Environmental Information, pp. 50–59 (2015)
9. Alippi, C., Ntalampiras, S., Roveri, M.: An HMM-based change detection method for intelligent embedded sensors. In: The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–7 (2012)
10. Kuncheva, L.I.: Change detection in streaming multivariate data using likelihood detectors. IEEE Trans. Knowl. Data Eng. **25**, 1175–1180 (2013)
11. Tran, D.-H.: Automated Change Detection and Reactive Clustering in Multivariate Streaming Data, arXiv preprint arXiv:1311.0505 (2013)
12. Vlasveld, R.: Temporal Segmentation using Support Vector Machines in the context of Human Activity Recognition, pp. 1–85 (2014)
13. Zhang, S., Galway, L., McClean, S., Scotney, B., Finlay, D., Nugent, C.D.: Deriving relationships between physiological change and activities of daily living using wearable sensors. In: Par, G., Morrow, P. (eds.) S-CUBE 2010. LNICST, vol. 57, pp. 235–250. Springer, Heidelberg (2011)
14. McCall, J.: Genetic algorithms for modelling and optimisation. J. Comput. Appl. Math. **184**, 205–222 (2005)
15. Khoo, M.B.: An extension for the univariate exponentially weighted moving average control chart. Matematika **20**, 43–48 (2004)
16. Pan, X., Jarrett, J.E.: The Multivariate EWMA model and health care monitoring. Int. J. Econ. Manage. Sci. **3**, 176 (2014)
17. Hotelling, H.: Multivariate quality control. Techniques of statistical analysis (1947)
18. Shimmer: Shimmer Wearable Sensing Technology. http://www.shimmersensing.com/. Accessed April 2016
19. Wang, S., Minku, L.L., Yao, X.: A learning framework for online class imbalance learning. In: 2013 IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL), pp. 36–45 (2013)