Anomaly Detection for Elderly In-home Activity Monitoring

Soon-Chang Poh, Yi-Fei Tan, Soon-Nyean Cheong, Chee-Pun Ooi, Wooi-Haw Tan
Faculty of Engineering, Multimedia University

Article Info

ABSTRACT

The world population of the elderly is expected to have a continuous growth and the number of elderly living in solitude is also expected to increase in the coming years. As our health decline with age, early detection of possible deterioration in health becomes important. Behavioral changes in in-home activities can be used as an indicator of health decline. For example, changes in routine of in-home activities. Past research mainly focused on detecting anomalies in routine of each type of in-home activities individually. In this paper, an anomaly detection model to detect changes in routine of in-home activities collectively for a day is proposed. The experiment was evaluated with an existing public dataset. The experimental results demonstrated that the anomaly detection model performed well on unseen testing data with an accuracy of 94.44%.

Keyword:

Behavioral changes
Changes in routine
Anomaly detection
In-home activities
Elderly care

1. INTRODUCTION

The United Nations predicted the world population of the elderly (65 years old and above) is expected to accelerate in the coming decades [1]. In Malaysia, one out of five of the whole population was predicted to be elderly (60 years old and above) by 2040 [2]. There is an increasing number of elderly end up living in solitude [2]. Hence, it is conjectured that there will be a significant population of solitude elderly in the future. Solitude elderly without care have a higher risk of mortality due to lack of care and monitoring of their health conditions [3]. One of the solutions is to hire caregivers for consistent monitoring and care, but the cost is often quite expensive for long-term care. An alternative option is to use the Internet of Things (IoT) technology in monitoring the elderly's daily activities or condition.

There are several research focusing on in-home activity recognition using machine learning techniques and the data is obtained from the environmental sensors which includes accelerometers and other sensors [4] [5]. The sensors data can be categorized into several in-home activities using the activity recognition software in the format of "Date", "Time" and "Type of activity". By observing one's in-home activities records, his or her usual routine of in-home activities may be modeled and the changes in routine can be detected.

Past works focused on detecting changes in routine of each type of in-home activities individually and the methods includes statistical method [6] and clustering-based anomaly detection [7]. In [6], Forkan et al. demonstrated a statistical method to detect shift in in-home activities routine. They assumed each type of in-home activity at starting time was normal distributed and the distribution is split into several regions of different degree of abnormality. Their technique was performed on synthetic data and it showed an accuracy above 90% generally. In [7], Hoque and colleagues focused on reducing false alarms in clustering-based anomaly detection on in-home activities with rule-based approach. In this work, the authors proposed to cluster each type of activity based on features such as starting time and duration with a clustering algorithm known as DBSCAN. For this method, an activity is classified as abnormal if it is not within 2 standard deviations from centers of all the clusters. They successfully reduced false positives and false negatives by at least 46% and 27% respectively.

Other researchers used in-home activities to detect abnormal behaviour on dementia patients and mild cognitive impairment patients [8], [9]. In [8] and [9], the anomalies to be detected were the sequences of actions which were unique to the dementia patients and these sequences of actions were defined by medical experts.

In this paper, an anomaly detection model based on time interval categorizing is proposed to detect changes in routine of in-home activities collectively for a day. The rest of this paper is organized as follows. In Section 2, details for methodology are discussed. Section 3 shows the experimental results and analysis and Section 4 gives the conclusions and future work.

2. METHODOLOGY

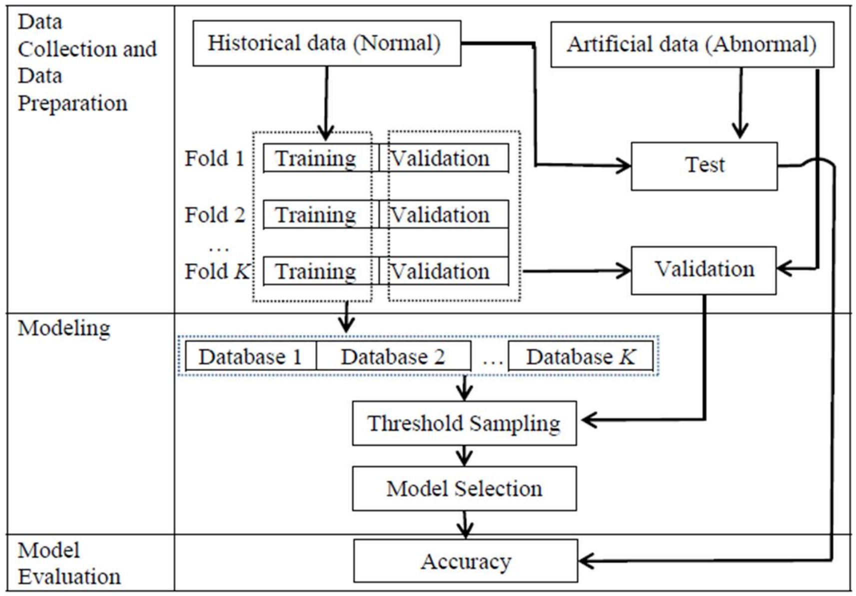

Figure 1. Framework of Analysis

Figure 1 illustrates the procedure designed to build the anomaly detection model using historical data and artificial data. The following sections discuss each steps of the procedure including data collection, data preparation, modeling and model evaluation.

2.1 Data Collection

The public dataset used are from CASAS, Washington State University [10]. It contains 220 days of sensors data of a volunteer adult from November 2010 to June 2011. Figure 2 shows an excerpt of the dataset. It consists of several attributes such as date, time, sensor events, types of activity and state of activity. An example of sensor event is "M003 ON" which means a motion sensor with id, M003 is on. Besides that, the state of activity comes with two labels which are "begin" and "end". "begin" means beginning of an activity and "end" means ending of an activity. In Figure 2, it is shown that some parts of the dataset are annotated with activities such as "Sleeping" and "Bed_to_Toilet. In total, there are 11 types of activities including "Meal preparation", "Relax", "Eating", "Work", "Sleeping", "Wash dishes", "Bed_to_Toilet", "Enter home", "Leave home", "Housekeeping" and "Resperate".

2010-11-04 05:40:43.642664 M003 OFF Sleeping end
2010-11-04 05:40:44.223548 M003 ON
2010-11-04 05:40:45.939846 M005 ON
2010-11-04 05:40:46.310862 M003 OFF
2010-11-04 05:40:51.303739 M004 ON Bed_to_Toilet begin

Figure 2. Excerpt of CASAS dataset

In this research, collection of all the activities happened in a day is considered as a single data instance. For anomaly detection, the data instances need to be categorized into two classes namely "normal" and "abnormal". The public dataset is considered as historical data (normal data). On the other hand, abnormal data is artificially generated. Each abnormal data instance is generated by circular shifting each in-home activities of a normal data instance by 4 hours to simulate changes in routine of in-home activities collectively for a day.

2.2 Data Preparation

Data preparation has 3 steps which are data processing, noise removal and data partitioning. Data processing is a step which processes the data into desired format such that it can fulfill the model requirement. Noise refers to data which has characteristics uncommon to the rest of the dataset. Some portion of the data collected contains unwanted noise. If this type of data is included in building a model, then they will affect the performance of the model. Thus, this type of data needs to be removed. During data partitioning, the processed and cleaned data are partitioned into training, validation and test set for modeling.

2.2.1 Data Processing

The type of data needed to build the anomaly detection model is historical records of in-home activities. Therefore, some part of the dataset which are without activity annotation were removed. The remaining data consists of only four attributes including date, time, types of activity and state of activity as shown in Figure 3.

```
2010-11-04   00:03:50.209589   Sleeping        begin
2010-11-04   05:40:43.642664   Sleeping        end
2010-11-04   05:40:51.303739   Bed_to_Toilet   begin
2010-11-04   05:43:30.279021   Bed_to_Toilet   end
```

Figure 3. Excerpt of data in desired format

2.2.2 Noise Removal

There are 3 types of noise in this dataset. For a small portion of the dataset, the "begin" and "end" of an activity are within "begin" and "end" of another activity or overlaps with "begin" and "end" of another activity. In addition, the length of the data instances is varying, and some data instances have lengths smaller by an order of 10 compared to the rest. Lastly, the activity "Resperate" only appeared six times in the dataset. These types of noises were removed resulting in only a 176 days dataset or 176 data instances remained.

2.2.3 Data Partitioning

The 176 data instances are normal data which were divided into training, validation and test set. Firstly, 10% of the normal data instances were partitioned into test set. Then, the rest of the normal data instances were partitioned using K fold cross validation. In K fold cross validation, the data instances are randomly partitioned into K partitions and each partition should contain an equal number of data instances. Each of the partition can be used as validation data once and the remaining $K-1$ data instances can be used as training data. This results in K random folds or combinations of normal training and validation data.

Table 1. Details of normal data partitioning

Fold	Training size	Validation size	Test size
1	138	20	18
2	138	20	
3	138	20	
4	138	20	
5	138	20	
6	139	19	
7	139	19	
8	139	19	

In the experiment, K=8 was used and the details of normal data partitioning is listed in Table 1. Firstly, 18 data instances were randomly chosen and partitioned into test set for each of the fold. The size of remaining data is 158 and not evenly divisible by K=8. As a result, the ratio of training to validation size is different for some of the folds. For example, the ratio for fold 1 to 5 is 138:20 and that of the rest is 139:19. However, the difference in ratio does not affect the results as the sizes are only different by one. The sizes of abnormal validation and testing data instances are equal to those of normal validation and testing data instances.

2.3 Modeling

As shown in Figure 4, the proposed anomaly detection model consists of two components including database and anomaly detector. Each anomaly detection model has 2 parameters which are the fold of training set used to train its database and the threshold τ of the anomaly detector. In this research, 80 models were built using different folds of training set and varying threshold choices. The performance and reliability of an anomaly detection model varies with its parameters. The purpose of modeling is to get the best model out of all the 80 models. Subsection 2.3.1 includes the description of the anomaly detection model and its training process. Subsection 2.3.2 gives the method to systematically sample the threshold choices to build different models. Subsection 2.3.3 discusses the method to choose the best model.

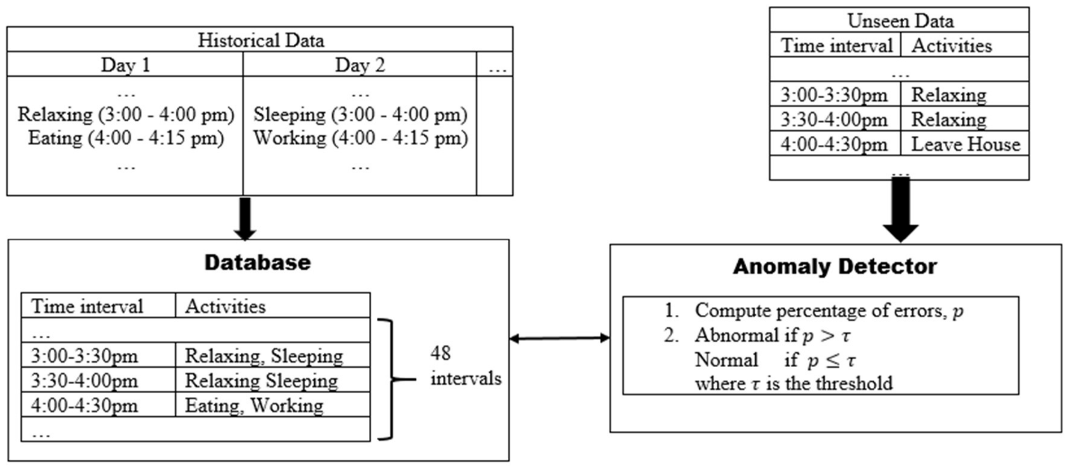

Figure 4. Overview of the anomaly detection model

2.3.1 Training Process

The first component of the proposed model is the database. The database models a user's normal daily routine based on time interval in a day. The database has 48 time interval categories, each for a unique 30-minutes interval in a day (e.g., 9:00-9:30 am). During the training, each of the activities in the training data instances is categorized based on the time intervals in a day and then saved into the database. For an activity that spans across several time intervals, it is saved into every time interval category it spans across. For example, in Figure 5, the relaxing activity spans across two 30-minutes time intervals, this activity will be saved into both the time interval categories, 3:00-3:30 pm and 3:30-4:00 pm. In this research, we used K=8 and trained 8 different databases with different folds of training set.

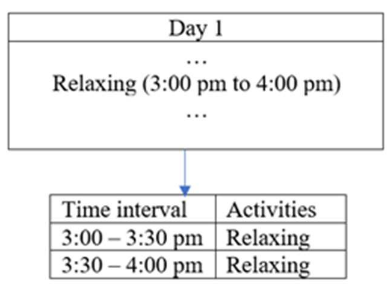

Figure 5. An activity that spans across two time interval categories

The second component of the model is the anomaly detector. Anomaly detector decides whether an unseen data instance is normal or abnormal. The first step of anomaly detection for is to check for errors in the unseen data instance by comparing it with the database. An error is defined as an activity that happens during a time interval in the data instance but it is not recorded in the respective category in the database. Then, the percentage of errors, p for the unseen data instance is computed using following equation:

$$p = \frac{\text{Number of Errors}}{\text{Total number of activities}} \times 100\%$$

The data instance is categorized as abnormal if its percentage of errors p is more than the threshold τ. On the other hand, it is categorized as normal if its p is less than or equal to τ. Threshold τ is one of the parameters of the model and the ways to sample threshold is given in the next subsection.

2.3.2 Threshold Sampling

The threshold τ of the model is a percentage of errors, p which separates percentages of errors of normal data instances and abnormal data instances. A percentage of errors, p is computed for each data instance in the validation set. The minimum and maximum of the calculated percentages of errors of every data instances in the validation set are taken as minimum and maximum of the range of threshold τ. Then, T values ($\tau = \{\tau_1, \tau_2, \ldots\ldots, \tau_T\}$) are linearly sampled from that range as threshold choices.

2.3.3 Model Selection

We used T=10 for threshold selection and trained K=8 databases. A total of 80 different models using different database and threshold τ were created. To choose the best model, an evaluation metric, $F1$ score is used. The $F1$ score of each model is derived from Confusion Matrix which was adopted in this research to study model performance.

Table 2. Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	True positive, *TP*: An abnormal data instance correctly classified as abnormal	False negative, *FN*: An abnormal data instance misclassified as normal
	Negative	False positive, *FP*: A normal data instance misclassified as abnormal	True negative, *TN*: A normal data instance correctly classified as normal

With assistance of the definition of Confusion Matrix in Table 2, 2 evaluation metrics including precision and recall can be derived as below:

$$\text{precision} = \frac{TP}{TP+FP}$$

$$\text{recall} = \frac{TP}{TP+FN}$$

where precision is a ratio of true positive to predicted positive and recall is a ratio of true positive to actual positive of a model. $F1$ score is a balanced combination of precision and recall and it can be calculated using the following equation:

$$F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

We calculated $F1$ score of all the 80 models (different fold and threshold τ) and select the model with the highest $F1$ score as the best model.

2.5 Model Evaluation

Model evaluation is the step of evaluating performance of the best model on an unseen test set using the evaluation metric, accuracy. It is computed using following equation:

$$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

3. RESULTS AND ANALYSIS

For each of the 80 models, a percentage of errors, p was calculated for each of the data instance in the validation set. Then, the minimum and maximum of all the calculated percentages of errors are 0% and 14.89%. 10 values were linearly sampled from the range [0%,14.89%] and they are 0%, 1.65%, 3.31%, 4.96%, 6.62%, 8.27%, 9.93%, 11.58%, 13.24% and 14.89%

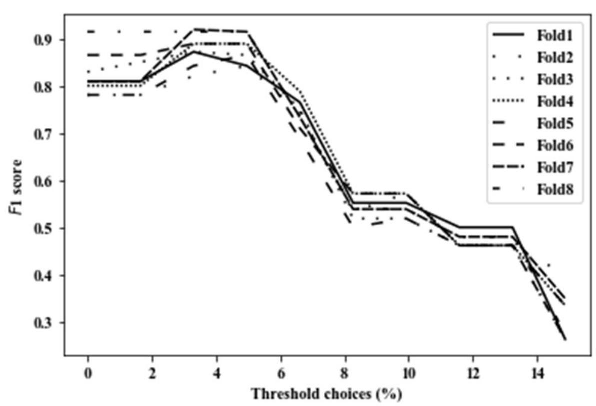

Figure 6. $F1$ score vs. threshold choices for K=8

Figure 6 shows the $F1$ score plot for each of the 80 models with different fold and τ. From eye estimation, the models with threshold choices between 3% and 6% has the among the highest $F1$ scores. The best model has parameters fold = 7 and τ = 3.31% and the highest $F1$ score, 91.89%.

From Table 3, it is shows that the precision, recall, $F1$ score and accuracy of the best model evaluated on validation and unseen testing data are at least 88.89%. In addition, accuracy of the model on validation and testing data are at least 92.11% and has a small difference of 2.33%. This demonstrates that the model performs excellently with a high accuracy and generalizes well to unseen data. Generalizing well to unseen data is important as it guarantees consistent performance of the model when it is deployed.

Table 3. Evaluation metrics for the best model

Evaluation metrics	Validation	Testing	Difference
Precision	94.44%	100%	5.56%
Recall	89.47%	88.89%	0.58%
$F1$ score	91.89%	94.12%	2.23%
Accuracy	92.11%	94.44%	2.33%

4. CONCLUSIONS AND FUTURE WORK

For monitoring behavioral changes of the elderly living alone at home, an anomaly detection model which can detect changes in routine of in-home activities was proposed. The experiment conducted with CASAS public dataset reveals excellent performance in terms of accuracy, precision and recall. This demonstrated that the anomaly detection model is effective in finding anomalies due to changes in routine of in-home activities collectively in a day. Currently the proposed method is mainly trained based on the activity in certain intervals of time from day to day. The correlation between the activities in the consecutive time interval may be investigated and studied to improve the model.

ACKNOWLEDGEMENTS

This work was funded by the TM Research & Development (TM R&D) Grant. The dataset was obtained from public dataset from CASAS, Washington State University.

REFERENCES

[1] United Nations (2015). World Population Ageing. [online] Available at: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf

[2] The Star (2017). Growing demand for retirement villages. [online] Available at: https://www.thestar.com.my/news/nation/2017/07/09/growing-demand-for-retirement-villages-the-demand-for-retirement-villages-in-malaysia-can-only-go-up/.

[3] T. P. Ng, A. Jin, L. Feng, M. S. Z. Nyunt, K. Y. Chow, L. Feng and N.P. Fong, "Mortality of older persons living alone: Singapore Longitudinal Ageing Studies," *BMC Geriatrics,* vol. 15, no. 1, pp. 126-126, 2015.

[4] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Communications Surveys and Tutorials,* vol. 15, no. 3, pp. 1192-1209, 2013.

[5] A. Bux, P. Angelov and Z. Habib, "Vision Based Human Activity Recognition: A Review", *Advances in Computational Intelligence Systems. Advances in Intelligent Systems and Computing*, vol. 513, pp. 341-371, 2016.

[6] A. Forkan, I. Khalil, Z. Tari, S. Foufou, and A. Bouras, "A Context-aware Approach for Long-term Behavioural Change Detection and Abnormality Prediction in Ambient Assisted Living," *Pattern Recognition,* vol. 48, no. 3, pp. 628-641, 2015.

[7] E. Hoque, R. F. Dickerson, S. M. Preum, M. Hanson, A. Barth and J. A. Stankovic, "Holmes: A Comprehensive Anomaly Detection System for Daily In-home Activities," *2015 International Conference on Distributed Computing in Sensor Systems, Fortaleza*, pp. 40-51, 2015

[8] D. Arifoglu, and A. Bouchachia, "Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks," *Procedia Computer Science*, vol. 110, pp. 86-93, 2017.

[9] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and R. Helaoui, *et al*.,"SmartFABER: Recognizing Fine-grained Abnormal Behaviors for Early Detection of Mild Cognitive Impairment," *Artificial Intelligence in Medicine*, vol. 67, pp. 57-64, 2016.

[10] D. Cook, Aruba, Washington State University, Center for Advanced Studies in Adaptive Systems, 2011. Available at: http://casas.wsu.edu/datasets/.