Implicit Irregularity Detection Using Unsupervised Learning on Daily Behaviors

Cuijuan Shang, Chih-Yung Chang, *Member, IEEE*, Guilin Chen, Shenghui Zhao, and Jiazao Lin

***Abstract*—The irregularity detection of daily behaviors for the elderly is an important issue in homecare. Plenty of mechanisms have been developed to detect the health condition of the elderly based on the explicit irregularity of several biomedical parameters or some specific behaviors. However, few research works focus on detecting the implicit irregularity involving the combination of diverse behaviors, which can assess the cognitive and physical wellbeing of elders but cannot be directly identified based on sensor data. This paper proposes an Implicit IRregularity Detection (*IIRD*) mechanism that aims to detect the implicit irregularity by developing the unsupervised learning algorithm based on daily behaviors. The proposed *IIRD* mechanism identifies the distance and similarity between daily behaviors, which are important features to distinguish the regular and irregular daily behaviors and detect the implicit irregularity of elderly health condition. Performance results show that the proposed *IIRD* outperforms the existing unsupervised machine-learning mechanisms in terms of the detection accuracy and irregularity recall.**

***Index Terms*—Daily behavior, elderly, homecare, implicit irregularity, unsupervised learning.**

I. Introduction

AGING has been one of the major challenges of the world in the last decades. The need of delivering quality care to a rapidly growing population of elders while reducing the healthcare costs has received much attention but proposed big challenges to the governments and socialites. Since the elderly prefer to spend their later years at home, plenty of researches attempted to support homecare by applying wireless sensor networks which are low-cost and high- reliability [1], [2].

Studies [3], [4] have shown that behaviors are associated with the healthy state of a person, especially an elder who usually lives regularly. Therefore, a variety of sensing systems have been proposed and developed recently to record the basic home activities of an elder [5], [6]. Consequently, a large amount of data can be continuously collected from sensors triggered by an elder, which can be used to recognize the Activities of Daily Living (ADL) of the elder [7]–[9]. Furthermore, the recognized activities can be exploited to assess the cognitive and physical wellbeing of elders. The daily activities of normal elderly people are likely to be steady or unchanging. This implies that indicators of daily activity measurements can specify the quantitative wellbeing of the elderly people. Consequently, when there is a change in the regular activities of an elder, the necessary cares can be taken in advance before any unforeseen situation happened.

Irregularity detection aims to automatically detect the abnormal health condition of an elderly. Since behaviors are associated with the healthy state of the elderly, the irregularity detection can be achieved by analyzing the behaviors of the elderly. In recent years, different statistical and machine learning methods are proposed to detect the irregularities of health condition or abnormal behavior for elderly people in smart home. The irregularity of health condition can be categorized into *explicit type* and *implicit type*. The *explicit irregularity*, such as racing heart or falling, always associates with abnormal biological parameters or actions, which can be directly detected or captured by sensors and should be processed immediately. Another feature of the *explicit irregularity* is that the irregularity can be identified by directly analyzing the explicit sensor data. Unlike the *explicit irregularity*, the features of *implicit irregularity* are implicit, which are not directed correlated to the sensor data. The *implicit irregularity* usually needs more efforts and deep analysis to explore its features which are not directly observed from the sensor data. The *implicit irregularity* of health condition, for instance, the bad mood or discomfort, is likely related to the changes of more than one behaviors.

The detection of *implicit irregularity* is important because the occurrence of irregularity also implies the unsteady of the health status of the elderly. To detect the *implicit irregularity* of the health condition, the analysis on behavior patterns of an elderly is required. For an elder, the behaviors of each day can be regarded as a behavior pattern. The regularity of daily behavior can be found in behavior patterns of the elderly. This also implies that the features of regularity in behavior patterns can be extracted. These important features can be applied to distinguish the regular and irregular behavior patterns or even identify the irregularity of an abnormal behavior pattern. However, the

Manuscript received November 13, 2018; revised January 3, 2019; accepted January 28, 2019. Date of publication February 1, 2019; date of current version January 6, 2020. This work was supported in part by the Doctoral Science Foundation of Chuzhou University (2017qd10), in part by the Chuzhou Science and Technology Program (201712), and in part by the Special Science and Technology Project of Anhui Province (16030901057). *(Corresponding author: Chih-Yung Chang.)*

C. Shang, G. Chen, and S. Zhao are with the School of Computer and Information Engineering, Chuzhou University, Chuzhou 239000, China (e-mail: shangcuijuan@chzu.edu.cn; glchen@chzu.edu.cn; zsh@chzu.edu.cn).

C.-Y. Chang is with the Department of Computer Science and Information Engineering, Tamkang University, Taipei 25137, Taiwan (e-mail: cychang@mail.tku.edu.tw).

J. Lin is with the Department of Information Management, Peking University, Beijing 100871, China (e-mail: linjz84@126.com).

Digital Object Identifier 10.1109/JBHI.2019.2896976

detection of *implicit irregularity* has more challenge since it requires to extract the features from the behavioral patterns, rather than directly analyzing the sensory data.

This paper proposes an *Implicit IRregularity Detection* (*IIRD*) mechanism which can be applied to the elderly living alone by analyzing the behavior patterns and extracting their regularity features. Based on the features, the proposed *IIRD* mechanism adopts unsupervised learning to detect the implicit irregularity and issue a warning alert of the elderly to the caregivers or their children.

The main contributions of this work are listed as follows.

1) **Implicit irregularity detection**: The proposed *IIRD* mechanism detects implicit irregularity, based on analyzing changes of several behaviors, instead of large amounts of original sensor data. The proposed *IIRD* overcomes the challenge of implicit irregularity detection since it might not be reflected directly by sensor data.
2) **Irregularity detection of alternative combination of diverse behaviors**: Since implicit irregularity often carries long-range dependency on mutual effects and changes of behaviors, a set of one day's behaviors is regarded as a basic unit of analysis in this paper. The irregularity of alternative combination of diverse behaviors will be detected by analyzing behavior data for a continuous period of time.
3) **Considering behavior weights based on their different impacts on irregularity**: Since different behaviors have different significances which lead to different impacts on irregularity, this paper explores the weights of different behaviors on implicit irregularity. The design of behavior weights helps explore the features of the implicit irregularity.
4) **Formalization of problem and constraints**: This paper establishes a problem formulation system to represent daily behaviors of an elderly in home for formalizing the problem of irregularity detection. Concepts of distance and similarity between behaviors are also proposed to quantify the difference between daily behaviors.

The rest of this paper is organized as follows. Section II summarizes related work. Section III formalizes the assumptions and problem statement of irregularity detection. In Section IV, two basic calculations including distance and similarity are defined. The distance and similarity play the important roles of features which are used to detect the irregularity in the later Section. In Section V, *Implicit IRregularity Detection* algorithm is introduced. Section VI presents the performance study and Section VII gives conclusions and future work of this study.

II. Related Work

Plenty of researches paid attention to the detection of irregular behaviors in order to help the elders live in their homes safely. In general, the irregular behavior detection mechanisms can be classified into two types: explicit irregularity detection [10]–[17] and implicit irregularity detection [18], [19]. The following briefly reviews these studies. As mentioned before, the explicit irregularity can be detected by analyzing the sensor data since high correlation exists between sensor data and the explicit irregularity. On the other hand, the implicit irregularity has more challenge, as compared with the explicit irregularity, because that it does not directly correlate to the change of sensor data and usually requires to analyze a sequence of changes from sensor data or behaviors.

A class of previous works analyzed the sensor data aiming at detecting the irregularity of health condition. Most of them fall into the explicit irregularity since the irregularity can be detected by directly analyzing the data from biography sensors or motion sensor. Study [10] developed a Heart Rate monitoring approach with the wearable piezoelectric sensor. Study [11] developed home-based wireless ECG monitoring system for appropriate healthcare. To provide a comprehensive service, study [12] proposed a complete and integrated information and communication Technology system to monitor vital signs of the elder at home. Study [12] further identified a minimum set of vital parameters, consisting of electrocardiogram, SpO2, blood pressure, and weight, measured through a pool of wireless, noninvasive biomedical sensors. Moreover, study [13] designed a health monitoring system, which used wearable, unobtrusive clothing with embedded sensors to measure eight health parameters namely heart rate, blood pressure, and so on. If the value of health parameter deviated from the normal range, warning would be sent to the caretaker or caregiver to prevent unexpected fatal medical conditions. However, these studies can only be applied to detect the explicit irregularity.

Some other works devoted themselves to detect the irregularity of health condition by analyzing the data of motion sensor. Fall is a typical explicit irregular behavior for an elder and is the most widespread domestic accidents among the elderly. Therefore, significant research efforts focused on fall detection [14]–[17]. Study [14] proposed a fall detection system based on posture recognition using a visual sensor. The proposed system firstly extracted the feature of human body, and then applied multiclass SVM and rule-based algorithm to detect fall. Using wearable sensors is the most common method for detecting falls and several researchers have explored the use of wearable acceleration sensors which are embedded in wearable devices [15], [16]. They introduced a threshold-based algorithm to analyze the collected changes of angular velocity to make the fall alarm. Study [17] developed a fall detection system to explore the salient spatial-temporal features of falls by pyroelectric infrared (PIR) sensor. It used mask arrays as reference structures to improve the space resolution in the developed Infrared Radiation Changes (IRC) sensing modules. Compared with the wearable acceleration sensors, its advantages are unobtrusive and easily deployed. However, the abovementioned studies that analyzed data of motion sensors or visual sensors can only be applied to the detection of explicit irregularity which has a constraint of high correlations existed between the irregularity and the sensory data. Hence they appeared powerless when some implicit irregularity happens.

Some other studies [18], [19] paid attention to the detection of implicit irregularity which cannot be directly reflected by the changes of sensor data. These studies mostly analyzed long term sensor data collected from various sensors by using

statistic methods or machine learning algorithms. Study [18] developed a monitoring system and utilized the sensor data to extract behavior patterns and detect implicit irregularity. It combined three approaches, the cosinor analysis, histogram-based approach and probabilistic model, to interpret daily routines and to detect irregular behavior. For each interested behavior, a probabilistic density function was built to serve as a model for normal behavior and used to check whether or not a behavior is irregular. However, the performance of irregularity detection is significantly influenced by the training data set of one behavior. For a complex irregularity which requires observing the changes of a sequence of behaviors, this scheme cannot be applied.

Study [19] proposed a model based on statistical analysis of some routine behaviors with context-aware information and detected the irregularity of behavior. Based on a large set of samples with normal pattern, Gaussian distribution of some particular routine activities was measured to summarize the normal lifestyle pattern and to identify any behavioral shift in regular activities. Meanwhile, a fuzzy rule-based method was used to help make the final decisions such as whether or not a behavior is a true irregularity. However, the challenge of this study is to convert huge amount of context-aware information into a prior knowledge base and to establish fuzzy rules. Furthermore, similar to [18], the proposed scheme cannot be applied to the complex irregularity, which only can be detected by analyzing a sequence of behavior changes.

The abovementioned studies built models for either detecting explicit irregularity by analyzing the change of sensor data or detecting implicit irregularity by analyzing the change of single behavior. However, most of them did not take into the effect of behavior combination on irregularity. This paper proposes an *Implicit IRregularity Detection* (*IIRD*) mechanism using unsupervised learning for feature extraction. The proposed *IIRD* algorithm takes a day's behaviors as a unit of analysis and considers the impact of alternative combination and weights of diverse behaviors on the irregularity. A comparative summary of the characteristics of the previous related works and the proposed *IIRD* is given in Table I. As shown in Table I, the proposed *IIRD* exhibits several good features, including being able to detect implicit irregularity and analyzing a set of behaviors or their combinations. In addition, the proposed *IIRD* belongs to the category of unsupervised learning which can learn without labeled data.

III. Assumptions and Problem Statement

This section initially introduces the network environment and assumptions. Then the problem statement and restrictions are described.

A. Network Environment

Given a home where an elder lives alone, the daily behavior monitoring system described in [5] can be deployed to collect sensor data. Further, behaviors of each day can be identified based on sensor data by applying the activity recognition algorithm proposed in [8]. Consider a number of consecutive days.

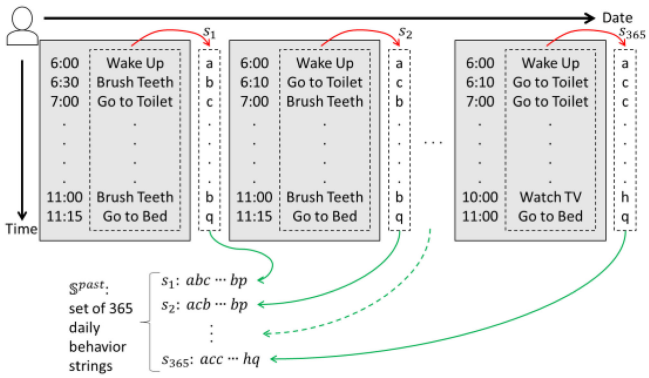

Fig. 1. Representation of daily behaviors.

A day is said to be a *regular day* if the behaviors of the considered day is similar to most of the other days. On the contrary, a given day is said to be *irregular day* if the behaviors of a given day have big changes as compared to the other remaining days. Herein, we assume that the number of regular days is much larger than that of the irregular days.

Given a number of ordered behaviors which have been partitioned based on the occurred date, this paper aims to identify the irregular days of the elderly using machine learning methods. In this study, we assume that the set of behaviors in each day will be a primitive set for observation. Let $\mathbb{B} = \{b_1, b_2, \ldots, b_p\}$ denote the set of p interested behaviors that might occur by the elderly. For instance, some well-known behaviors are taking shower, having lunch, wake up, and so on. Assume that an elder will act plenty of behaviors during a day and each behavior is denoted by a character. For example, characters '*a*' and '*b*' represent *WakeUp* and *BrushTeeth*, respectively. Let string s denote a set of behaviors occurred in a day. This is, string s is a behavior pattern of the elder. Let $\mathbb{S}^{past} = \{s_1, s_2, \ldots, s_m\}$ denote the set of behaviors occurred in the past m days. For the ease of presentation, the string s_i associated with a regular day is said to be a regular string. On the contrary, the string s_i is said to be an irregular string if its associated day is an irregular day. This paper aims to build up a model to present the regularity features of the behavior patterns to detect the abnormal health condition of the elderly.

Fig. 1 depicts an example where there are a number of behaviors occurred in the past one year. Each behavior is represented by a unique character and hence a string represents the set of behaviors of each day. For example, strings $s_1 = abc \ldots bp$ and $s_2 = acb \ldots bp$ indicate the sets of behaviors of the first two days.

Assume that the home space can be divided into several regions, such as living room, shower room and so on. Let $\mathbb{L} = \{l_1, l_2, \ldots l_q\}$ denote the set of all partitioned regions in a home. Each behavior b_i can be represented as a three-tuple list $\Gamma_i = (b_i, [t_i^{start}, t_i^{end}], l_i)$, where notations t_i^{start}, $t_i^{end}$and l_i denote the starting time, ending time and location associated with the behavior b_i.

In the next subsection, the goal and its associated constraints of the investigated problem will be presented.

TABLE I
COMPARISON OF THE MAIN CHARACTERISTICS OF THE PROPOSED *IIRD* WITH THE EXISTING RELATED WORKS

Studies	Implicit Irregularity	Analyzed Data Type	Behavior Combination	Algorithms	Dependence on Hardware/Analysis
[10]-[13]	No	Sensor data	No	Threshold-based	Sensor Hardware
[14]	No	Sensor data	No	Multiple SVMs and Rule-based	Sensor Hardware
[15], [16]	No	Sensor data	No	Threshold-based	Sensor Hardware
[17]	No	Sensor data	No	Mask Strategy	Sensor Hardware
[18]	Yes	Sensor data	No	Combination of Regression Analysis, Histogram and Probabilistic Model	Sensor Data Analysis
[19]	Yes	Behavior data	No	Gaussian Distribution and Fuzzy Fule-based	Behavior Analysis
Our Proposed *IIRD*	Yes	A Day's Behavior Data	Yes	Unsupervised Learning	Behavior Analysis

B. Problem Statement

Given a set of behaviors occurred in the past *m* days $\mathbb{S}^{past} = \{s_1, s_2, \ldots, s_m\}$, this paper aims to build up a model M^{opt} which can present the regularity features of the behavior patterns of the elderly and can be used to identify the irregular situation for a given daily behavior string in the future.

Assume that the living habit of an elder in home would not be changed drastically during a short period. Furthermore, it is assumed that behaviors of any two regular days are similar to each other. In case that some emergencies or abnormal events occur in someday, the behaviors of that day will have a big change. Since the behaviors of the elderly are likely to be affected by external factors on weekends, in our later discussion, any set of behavior data will exclude the behaviors on weekends.

Recall that $\mathbb{S}^{past} = \{s_1, s_2, \ldots, s_m\}$ denote the set of behaviors occurred in the past *m* days. Let *M* be an *irregularity detection model* which is trained based on the given $\mathbb{S}^{past}$. That is, model *M* takes any behavior string s_i as its input and reports either 'regular day' or 'irregular day'. Let $\mathbb{S}^{past}$ be classified into two sets, including the set of regular days, denoted by R, and the set of irregular days, denoted by $\tilde{R}$. It is obvious that

$$|R| + \left|\tilde{R}\right| = m \tag{1}$$

Let notation $\mathbb{M}^{\mathbb{S}^{past}}$ denote the set of all possible *irregularity prediction* models which analyze a given set of behaviors of one day in the future and report whether the day is regular or irregular.

Exp. (2) defines the function f^M which operates on model $M \in \mathbb{M}^{\mathbb{S}^{past}}$ and returns 1 if the daily behavior s_i is regular.

$$f^M(s_i) = \begin{cases} 1, & s_i \in R \\ 0, & s_i \in \tilde{R} \end{cases} \tag{2}$$

For a considered detection model *M*, let Boolean variables γ_i and β_i denote two types of errors, *False Positive* and *False Negative*, respectively. That is, the values of γ_i or β_i will be set to 1 if false positive or false negative occur, respectively. Herein, we notice that a nonzero value of γ_i or β_i represents that the reports of the designed model *M* occur error. This also indicates that the optimal model expects to have zero values of γ_i and β_i for any given daily behavior s_i in the future.

The following presents the design of γ_i or β_i. Herein, we want to emphasize that the occurrence of *False Positive* indicates that $f^M(s_i)$ reports 0 (predicting irregular day) but string s_i is regular. Exp. (3) presents the design of *False Positive* γ_i.

$$\gamma_i = 1 - f^M(s_i) \tag{3}$$

Let $\tilde{s}_j$ denote an irregular string. Variable β_i is set to 1 if the false negative occurs. That is, $f^M(\tilde{s}_j)$ reports 1 (predicting regular day) but $\tilde{s}_j$ is actually irregular. Similar to the design of γ_i, Exp. (4) further designs *False Negative* β_i.

$$\beta_i = f^M(\tilde{s}_j) \tag{4}$$

This paper aims to develop an optimal model M^{opt} based on $\mathbb{S}^{past}$, which can effectively identify the irregular days of the elderly using unsupervised learning. Let notation $\mathbb{S}^{future}$ denote the set of behaviors in the future *m* days. The goal of this paper is to prevent the detection results in any false positive and false negative. That is, for any given regular string s_i and irregular string $\tilde{s}_j$ in the future, the γ_i and β_i are expected to have zero value. Exp. (5) presents the objective function which aims to minimize the number of two type errors.

$$Min\left(\sum\nolimits_{s_i \in R, \tilde{s}_j \in \tilde{R}} (\gamma_i + \beta_j)\right) / \left|\mathbb{S}^{future}\right|, \ s_i, \tilde{s}_j \in \mathbb{S}^{future} \tag{5}$$

C. Assumptions and Constraints

This subsection presents a couple of assumptions and constraints that make up the premise of our work.

Given a big set of behavior strings, this paper assumes that the number of the regular days is much larger than that of the irregular ones. Exp. (6) reflects this assumption.

$$|R| \gg \left|\tilde{R}\right| \tag{6}$$

When constructing the detection model, features play important roles since they can be applied to distinguish the different classes. The difference between the two strings s_i and s_j can vary from small to large. This paper additionally assumes that the difference between two regular strings is small while the difference between regular and irregular strings is very large. However, the difference between two irregular strings is uncertain. Let $d(s_i, s_j)$ denote the difference between string s_i and string s_j. Based on the abovementioned assumptions, the following presents the constraints which should be satisfied when developing the optimal model M^{opt}.

1) Feature Constraint: The maximum difference between any two regular strings should be smaller than the minimum

difference between the pair of any regular string and irregular string. That is

$$\max_{\forall s_i, s_j \in R} d(s_i, s_j) < \min_{\forall s_i \in R, \forall \tilde{s}_j \in \tilde{R}} d(s_i, \tilde{s}_j) \tag{7}$$

This paper treats each behavior as the basic element which can be used to distinguish the given day is regular or irregular. One additional constraint is that the occurrences of behaviors in any given day should be independent. For any given time, there exactly one behavior occurs and the behavior can only happen at one space in home. The following time and space constraints reflect these restrictions. The time constraint reflects that the time duration of any two behaviors, say b_j^i and b_k^i, cannot be overlapped.

2) Time Constraint:

$$\text{if } \forall b_j^i \subset s_i,\ \forall b_k^i \subset s_i \text{ and } j < k, t_j^{end} < t_k^{start} \tag{8}$$

Let $\rho_{i,j}$ denote the Boolean variables, representing whether or not behavior $b_i \in \mathbb{B}$ happens in the region $l_j \in \mathbb{L}$. That is,

$$\rho_{i,j} = \begin{cases} 1, & b_i \text{ occurs in } l_j \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

The space constraint restricts that one behavior can only occur in one certain region. For example, if behavior b_i happens in region l_1, we have $\rho_{i,1} = 1$ and $\rho_{i,j} = 0$, for all $j \neq 1$.

3) Space Constraint:

$$\sum\nolimits_{l_j \in \mathbb{L}} \rho_{i,j} \leq 1,\ \forall b_i^k \in s_k,\ \forall s_k \in \mathbb{S}^{past} \tag{10}$$

IV. Distance and Similarity Calculations

This paper aims to build up an implicit irregularity detection model which can detect the implicit irregular behaviors of the elderly. Meanwhile, it aims to minimize the number of detection errors when the model is applied. Two basic calculations in the proposed algorithm are the *Distance* and *Similarity Calculations*. This chapter first presents the distance calculation of two strings. Based on the distance calculation, the definition of similarity and its calculation are introduced. The similarity calculation will be used in the next Section which presents the main algorithm of *Implicit Irregular Detection*.

A. Distance Calculation

Given a set $\mathbb{S}^{past}$ of behavior strings, $\mathbb{S}^{past}$ should be cleaned up according to the assumptions and restrictions discussed in above Section. Given a string $s_i \in \mathbb{S}^{past}$ which is associated with the behaviors of the *i*th-day of the input data, a main task for identifying the regularity of this day is to compare the degree of differences between string s_i and all the other strings. The difference between any two strings is represented by the distance between them.

The distance between two strings s_i and s_j will be measured by the number of basic operations required for the change from string s_i to string s_j. The following introduces some basic string operations. Based on these operations, the difference between any given two strings s_i and s_j can be defined.

Let Φ denote the set of basic string operations, including insertion, deletion and modification. Let $\phi^{op}(s_i, j, *)$ denote the execution of basic string operation *op* at *j*-th position of string s_i. The three basic string operations are given in what follows.

1) $\phi^{insert}(s_i, j, b)$: insert character *b* at the *j*-th position of string s_i.
2) $\phi^{delete}(s_i, j)$: delete the *j*-th character from s_i.
3) $\phi^{modify}(s_i, j, b)$: replace the *j*-th character with character *b*.

That is, the set of basic string operations is $\Phi = \{\phi^{insert}, \phi^{delete}, \phi^{modify}\}$. The basic operations ϕ^{insert}, ϕ^{delete} and ϕ^{modify} can be associated with different weights, ω^{insert}, ω^{delete} and ω^{modify}, respectively, representing different importance on the change of behavior. In the daily life of the elder, the change of some behavior might be treated as an important event than the changes of other behaviors. For example, 'taking medicine' is an important behavior which should be emergently reminded if the behavior has been forgotten. On the contrary, the behavior of 'tooth brushing' is not so important, as compared with behavior 'taking medicine'. To emphasize the change of the important behavior, each behavior *b* will be associated with a weight μ_b, for any behavior $b \in \mathbb{B}$. Putting the weight of basic operation and the behavior together, let c^{op} denote the importance of behavior change by executing ϕ^{op}. If any string operation ϕ^{op} involves two behaviors b_i and b_j, the total cost of this operation can be expressed by Exp. (11), where $\phi^{op} \in \Phi$ and $b_i, b_i \in \mathbb{B}$.

$$c^{op} = \omega^{op} * (\mu_{b_i} + \mu_{b_j}) \tag{11}$$

Otherwise, if only one behavior b_i is involved in operation ϕ^{opt}, Exp. (11) will be degraded into Exp. (12), where $\phi^{op} \in \Phi$ and $b_i \in \mathbb{B}$.

$$c^{op} = \omega^{op} * \mu_{b_i} \tag{12}$$

The importance of behavior change can be treated as the cost. A higher cost of the behavior change on string s_i indicates that the change likely leads to a irregular day.

The following further formally presents the definition of distance of two given strings s_i and s_j. Let s_i^y denote the resultant string by applying *y* times basic string operations on string s_i. Let string s_j can be obtained after applying *k* times basic string operations on string s_i. That is, we have $s_j = s_i^k$. Let $c_{i,k}^{op}$ denote the cost for applying the *k*-th basic string operation *op* on string s_i. Let $C_{i,k}^{op}$ denote the accumulated cost for applying *k* times basic string operations on string s_i. It is obvious that we have

$$C_{i,k}^{op} = \sum\nolimits_{x=1,k} c_{i,x}^{op}.$$

There are various combinations of the *k* basic operations. In the conceptual level, the distance of strings s_i and s_j can be measured by the minimal cost of the combinations. The following formally defines the distance of strings s_i and s_j.

Definition: $d(s_i, s_j)$, distance of strings s_i and s_j

The distance $d(s_i, s_j)$ of two strings s_i and s_j is the minimal cost of changes from string s_i to string s_j. That is,

$$d(s_i, s_j) = Min(C_{i,k}^{op}) = Min\left(\sum\nolimits_{x=1,k} c_{i,x}^{op}\right). \tag{13}$$

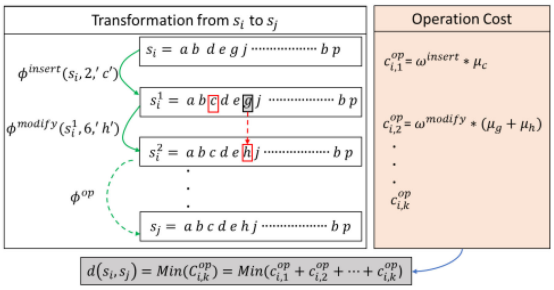

Fig. 2. Distance calculation of two daily behaviors.

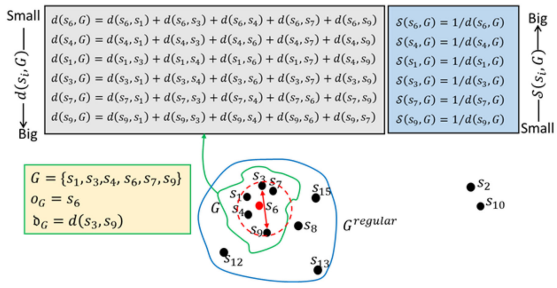

Fig. 3. Origin o_G and diameter $\mathfrak{d}_G$ of group G.

It is observed that, the more similar the two behavior strings, the smaller the distance between them.

Fig. 2 gives an example to elaborate the computation of $d(s_i, s_j)$. Assume that the weights of insertion and modification operations are $\omega^{insert} = 1$ and $\omega^{modify} = 1$, respectively. String s_j can be obtained after applying k basic string operations on string $s_i ='' abdegj \ldots bp''$. The first two operations are $\phi^{insert}(s_i, 2,' c')$ and $\phi^{modify}(s_i, 6,' h')$. That is, the first operation is to insert a character 'c' after the second character of s_i and generates a temporary string s_i^1. Assume that the weight of operation on behavior 'c' are $\mu_c = 3$. Concurrently, the cost of the first operation is calculated as follows by applying Exp. (12). That is, $c^{op}_{i,1} = \omega^{insert} * \mu_c = 1 * 3 = 3$. Then, the second operation replaces the character 'g' with 'h' at the 6-th position of the temporary string. Assume that the weights of operation on behaviors 'g' and 'h' are $\mu_g = 3$ and $\mu_h = 1$, respectively. The operation cost is computed by applying Exp. (11) as follows.

$$c^{op}_{i,2} = \omega^{modify} * (\mu_g + \mu_h) = 1 * (3 + 1) = 4$$

After k times basic operation, s_j turns out and the distance between s_i and s_j is the sum of each operation cost.

$$d(s_i, s_j) = Min(C^{op}_{i,k}) = Min(3 + 4 + \cdots + c^{op}_k)$$

B. Similarity Calculation

The previous subsection has introduced the distance calculation of given two strings. Two strings with a long distance indicate the behaviors in the associated two days have a big change and hence an irregular day might exist in the two days. Herein, we notice that a day is irregular if its behaviors have big changes, as compared to all the other days. To identify the irregular day, the similarity of a given string to all the other strings should be calculated. This subsection presents the definition and calculation of the similarity of a given string to a given group of strings.

Let G be a group of behavior strings. For any $s_i \in G$, let $d(s_i, G)$ denote the sum of distances between string s_i and all the other strings in G. That is, the $d(s_i, G)$ is the accumulated distance between s_i and every string $s_j \in G$. That is,

$$d(s_i, G) = \sum\nolimits_{s_j \in G} d(s_i, s_j) \text{ for any } s_i \in G \quad (14)$$

The key concept behind the distance $d(s_i, G)$ is that if s_i has a small value of $d(s_i, \mathrm{G})$, s_i is similar to most of the other strings in group G. Let notation $\mathcal{S}(s_i, G)$ denote the similarity between s_i and all the other strings in G. Exp. (15) further defines $\mathcal{S}(s_i, G)$.

$$\mathcal{S}(s_i, G) = \begin{cases} \text{Max (Integer)}, & d(s_i, G) = 0 \\ 1/d(s_i, G), & \text{otherwise} \end{cases} \quad (15)$$

Next, we notice that the value of $d(s_i, G)$ is zero if all strings in G are identical. If this is the case, the $\mathcal{S}(s_i, G)$ should take the maximum value of Integer, denoted by *Max*(Integer).

Herein, we expect to present G in a multi-dimensional space. In group G, each string can be treated as a point in the space. The following defines origin o_G of group G.

Definition: o_G, origin of group G.

The point that has the largest similarity to group G will be considered as the original o_G of group G. That is,

$$o_G = \arg\max_{s_i \in G} \mathcal{S}(s_i, G) \quad (16)$$

To present the range that covers all points in G, the following defines diameter $\mathfrak{d}_G$ of group G.

Definition: $\mathfrak{d}_G$, diameter of group G.

The diameter $\mathfrak{d}_G$ of group G is defined by the maximum distance between any two behavior strings in G. That is,

$$\mathfrak{d}_G = \max_{s_i, s_j \in G} d(s_i, s_j) \quad (17)$$

Fig. 3 depicts an example for calculating the origin o_G and diameter $\mathfrak{d}_G$ of group G. Assume that group G is made up by six regular strings, namely $G = \{s_1, s_3, s_4, s_6, s_7, s_9\}$. As shown in Fig. 3, each string in G is represented by a point that is enclosed by the green contour line. In order to obtain the origin o_G and diameter $\mathfrak{d}_G$ of group G, distance $d(s_i, G)$ and similarity $\mathcal{S}(s_i, G)$ will be calculated for each element $s_i \in G$, according to Exps. (14) and (15), respectively. The calculations of distance $d(s_i, G)$ and similarity $\mathcal{S}(s_i, G)$ of string s_i are shown in Fig. 3. As a result, s_6 has the largest similarity and will play the role of origin o_G, as marked with the red color in Fig. 3. Since s_3 and s_9 has the largest distance in group G, the value of diameter can be evaluated by $\mathfrak{d}_G = d(s_3, s_9)$. As shown in Fig. 3, the diameter is marked by the red line.

V. The Proposed Implicit Irregularity Detection Mechanism

This section presents the proposed ***Implicit IRregularity Detection*** (***IIRD***) mechanism for daily behaviors of the elder.

Given a behavior string set $\mathbb{S}^{past}$ which consists of the behavior strings of the past *m* days, the proposed *IIRD* mechanism aims to construct a regular group of behavior strings. Then the irregularity detection model will be built up.

Initially, the proposed *IIRD* mechanism constructs a basic regular group, say *G*. One important feature of this group is that all strings belonging to *G* are regular. This also indicates that all strings in group *G* are similar to each other. After constructing the basic regular group, group *G* will be expanded by adding one regular string at a time. The group expansion will be repeatedly executed until the stop criteria are satisfied. The following firstly presents the construction of the basic regular group. Then the expansion of the basic group is presented.

A. Construction of Basic Regular Group (CBRG)

Given a behavior string set $\mathbb{S}^{past}$, the first task is to construct the basic regular group *G*. As mentioned before, one important feature of *G* is that all strings in G should be regular strings. Recall that a large value of similarity of string s_i indicates that string s_i has the smallest accumulated distance to all the other strings in *G*. That is, a string with bigger similarity is much similar to all the other strings and should be prior included in group *G*. The similarity of each string in *G* will be calculated according to Exp. (15). The string which has the maximum value of similarity $\mathcal{S}(s_i, \mathbb{S}^{past})$ is definitely a regular behavior string and should be the first one to be included in group G. Let s^{best} denote the string which has largest similarity. That is,

$$s^{best} = \arg \max_{s_i \in \mathbb{S}^{past}} \mathcal{S}\left(s_i, \mathbb{S}^{past}\right)$$

The group G is thus initialized by $G = \{s^{best}\}$ and string s^{best} also plays the role of origin of *G* while the diameter $\mathfrak{d}_G$ is zero obviously.

However, if the expansion of group G begins from one element group, the calculation complexity is large. A method to speed up the construction of group *G* is presented in the following.

Let $\mathfrak{S}^{\mathcal{S}} = \{\mathcal{S}(s_1, \mathbb{S}^{past}), \mathcal{S}(s_2, \mathbb{S}^{past}), \ldots, \mathcal{S}(s_m, \mathbb{S}^{past})\}$ denote the set of similarities of all strings in $\mathbb{S}^{past}$. Let *Median*$(\mathfrak{S}^{\mathcal{S}})$ is a function which sorts the elements in $\mathfrak{S}^{\mathcal{S}}$ in a nonincreasing order and then returns the value of middle item in the sorted set. Let *median similarity*, denoted by $\mathcal{S}^{median}_{\mathbb{S}^{past}}$, be the value returned by function $Median(\mathfrak{S}^{\mathcal{S}})$. Exp. (18) calculates the value of $\mathcal{S}^{median}_{\mathbb{S}^{past}}$.

$$\mathcal{S}^{median}_{\mathbb{S}^{past}} = Median\left(\mathfrak{S}^{\mathcal{S}}\right) \tag{18}$$

Recall the assumptions and constraints discussed in Subsection III.C. The number of regular days is much larger than that of the irregular ones and meanwhile, the maximum difference between any two regular strings should be smaller than the minimum difference between the pair of any regular and irregular strings. That is to say, any regular string $s_i \in \mathbb{S}^{past}$ has a bigger similarity $\mathcal{S}(s_i, \mathbb{S}^{past})$ than that of any irregular string. This also implies that string s_i satisfying the following condition can be included in group *G*.

$$\mathcal{S}\left(s_i, \mathbb{S}^{past}\right) > \mathcal{S}^{median}_{\mathbb{S}^{past}}$$

Consequently, we have

$$G = \left\{s_i \left| \mathcal{S}\left(s_i, \mathbb{S}^{past}\right) > \mathcal{S}^{median}_{\mathbb{S}^{past}}, s_i \in \mathbb{S}^{past}\right.\right\}.$$

The origin o_G and diameter $\mathfrak{d}_G$ of group *G* can be accordingly calculated by applying Exps. (16) and (17), respectively.

B. Expansion of Regular Group (ERG)

This subsection details the expansion procedure using the basic regular group G. As mentioned before, group *G* will be expanded by adding one regular string at a time until the termination criteria is satisfied.

Let G^{unpro} denote the set of strings excluded by *G*. For instance, in Fig. 3, $G^{unpro} = \{s_2, s_8, s_{10}, s_{12}, s_{13}, s_{15}\}$. The strings in G^{unpro} will be checked one by one in the ascending order of distance $d(s_j, \mathbb{S}^{past})$ or in an descending order of similarity $\mathcal{S}(s_j, \mathbb{S}^{past})$, for determining whether or not it can be joined to *G*. The following presents the expansion procedure executing on string s_j. The expansion procedure mainly consists of two steps. The first one is to calculate the distance from s_j to the origin o_G of group *G*. In the second step, based on the results of previous step, a decision can be made to determine whether or not string s_j can be actually included in *G*. When string s_j has joined group *G*, the new origin and diameter of the new group will be calculated. Otherwise, expansion procedure will be terminated.

In step 2, the decision should be made to determine whether or not string s_j can be actually included in *G*. The following presents the criteria for making such a decision.

Inclusion Criteria:

A string s_j can be included in a regular group *G* only if the following two criteria are satisfied.

1) $s_j = \arg\max_{s_i \in G^{unpro}} \mathcal{S}(s_i, \mathbb{S}^{past})$
2) $d(s_j, o_G) \leq \mathfrak{d}_G$

In the inclusion criteria, the first condition for s_j is that s_j has the maximum value of similarity in G^{unpro}. This condition guarantees that string s_j has the maximal potential to be included in *G*, as compared with all the other strings in G^{unpro}. The second condition is that the distance $d(s_j, o_G)$ is no more than the diameter of the group *G*.

If the *Inclusion Criteria* are satisfied, string s_j will be actually included in group G, that is, $G = GU\{s_j\}$. Meanwhile, the new origin o_G and new diameter $\mathfrak{d}_G$ of new group G will be calculated accordingly. Otherwise, strings s_j will be regarded as an irregular string and the expansion procedure of the group *G* will be terminated. If it is the case, the regular group $G^{regular}$ which we aim to construct is the final group *G*. As shown in Fig. 3, the cyan contour presents the group $G^{regular}$ which should include all the regular strings.

Based on the group $G^{regular}$ with the origin $o_{G^{regular}}$ and diameter $\mathfrak{d}_{G^{regular}}$, any future string of daily behavior can be examined. As a result, the irregular detection model, denoted by $M_{G^{regular}}$ can be constructed based on $G^{regular}$.

Given any string $s_j \in \mathbb{S}^{future}$, the constructed model $M_{G^{regular}}$ is able to determine whether or not s_j belongs to group $G^{regular}$ according to the distance between s_j and $o_{G^{regular}}$, as

Algorithm

Inputs: A set of behavior strings $\mathbb{S}^{past}$

Output: A group $G^{regular}$ with the center $o_{G^{regular}}$ and diameter $\mathfrak{d}_{G^{regular}}$

1. Set G, $\mathfrak{S}^{\mathcal{S}}$, o_G to be null, r_G=0;
2. $G^{unpro} = G^{unpro} \cup \mathbb{S}^{past}$;

/* Calculate the distance for each string in $\mathbb{S}^{past}$ */

3. for (i=1,i≤ $|\mathbb{S}^{past}|$, i++){
4. $\mathcal{S}(s_i, \mathbb{S}^{past}) = 1/(\sum_{s_j \in \mathbb{S}^{past}, s_i \neq s_j} d(s_i, s_j))$;
5. $\mathfrak{S}^{\mathcal{S}} = \mathfrak{S}^{\mathcal{S}} \cup \mathcal{S}(s_i, \mathbb{S}^{past})$; }
6. $\mathcal{S}^{median}_{\mathbb{S}^{past}} = Median(\mathfrak{S}^{\mathcal{S}})$;

/*Find regular behavior strings from G^{unpro} */

7. for (i=1,i≤ $|G^{unpro}|$, i++){
8. if $\mathcal{S}(s_i, \mathbb{S}^{past}) > \mathcal{S}^{median}_{\mathbb{S}^{past}}$ {
9. $G = G \cup s_i$;
10. $G^{unpro} = G^{unpro} \setminus s_i$; }}
11. $o_G = \arg\min_{s_i \in \mathbb{S}^{past}} d(s_i, G)$;
12. $\mathfrak{d}_G = \max_{s_i, s_j \in G} d(s_i, s_j)$;

/* Process of rest of G^{unpro} */

13. **While** TRUE{
14. $s_j = \arg\max_{s_i \in G^{unpro}} \mathcal{S}(s_i, \mathbb{S}^{past})$;
15. if $d(s_j, o_G) \leq \mathfrak{d}_G$ {
17. $G^{unpro} = G^{unpro} \setminus s_i$
18. $G = G \cup s_j$;
19. $o_G = \arg\min_{s_i \in G} d(s_i, G)$;
20. $\mathfrak{d}_G = \max_{s_i, s_j \in G} d(s_i, s_j)$;}
21. else{
22. Break;}}
23. $G^{regular} = G$, $o_{G^{regular}} = o_G$, $\mathfrak{d}_{G^{regular}} = \mathfrak{d}_G$;
24. **Return** $G^{regular}$

Fig. 4. The proposed *IIRD* algorithm.

expressed by Exp. (19).

$$d(s_j, o_{G^{regular}}) \leq \mathfrak{d}_{G^{regular}} \tag{19}$$

In case that the condition given in Exp. (19) holds, s_j should be included in $G^{regular}$, which indicates that s_j is regular. Further, the $G^{regular}$ can be adaptively updated by recalculating the new origin and diameter. Otherwise, s_j is irregular and the subsequent actions such as warning or notification to the relevant caregivers should be initiated.

C. The Proposed IIRD Mechanism

This subsection summarizes the proposed *IIRD* mechanism. The proposed *IIRD* mechanism mainly consists of two procedures, including of *CBRG* and *ERG*. The procedure *CBRG* mainly constructs a basic regular group *G,* according to the computation of distance and similarity for each given string. Once group *G*, containing one or more regular strings, is constructed, it begins to be expanded by applying procedure *ERG* until the *Inclusion Criterion* is not satisfied.

Fig. 4 depicts the detailed steps of the proposed *IIRD* algorithm. In steps 1 and 2, variables including G, $\mathfrak{S}^{\mathcal{S}}$, o_G, r_G as well as G^{unpro} are initialized. Then the procedure *CBRG* is implemented from steps 3 to 12. Steps 3 and 5 calculate the distance for each string while step 6 calculates the median distance of all strings. After that, steps 7 to 10 build up the basic regular group G and update the group G^{unpro} accordingly. Steps 11 and 12 further calculate the origin and diameter of G. Steps 13 to 22 describe the process of procedure *ERG* which repeats to check strings in G^{unpro} for determining whether or not a string should be included in group G. Step 14 gives the *Inclusion Criterion* which specifies the stop condition. Finally, steps 23 to 24 construct the group $G^{regular}$ with the origin $o_{G^{regular}}$ and diameter $\mathfrak{d}_{G^{regular}}$.

TABLE II
SIMULATION PARAMETERS

Notation	Description	Values
N	Number of derived strings for one basic behavior string	500
α	Ratio of the number of irregular string to the number of a dataset	0.1~0.3
ε	Ratio of the number of changed characters to the length of basic regular strings for regular strings	0.05~0.3
δ	Ratio of the number of changed important characters to the length of basic regular string for irregular strings	0.15~0.4
w_b	Ratio of minimum weight of important behavior and maximum weight of ordinary behavior	⩾1

VI. SIMULATION

This section investigates the performance of the proposed *IIRD* mechanism against K-means, Hierarchical Cluster (HC) and Density-Based Spatial Clustering of Application with Noise (DBSCAN) in terms of the accuracy of irregularity detection, which is measured by the number of False Positive (FP) and False Negative (FN). In addition, the robustness of the compared mechanisms are investigated. The following firstly shows the data preparation and then discusses the performance results.

A. Data Preparation

This subsection illustrates the preparation of the simulation data. There are 15 elders interviewed for collecting their daily behaviors. According to the behavior patterns of the 15 elders, there are three types of different patterns, noted by s_X, s_Y and s_Z. Then, we select one behavior pattern, say s_X, to be considered in the experiments. The selected s_X represents the behaviors of 8 elders whose living styles are similar. Then representative behaviors will be picked out and organized as a behavior set which is further considered in the experiments. Next, based on the interviews, the behaviors which might affect the implicit irregularity are picked out and regarded as the important behaviors. Let important behavior set associated with s_X, is denoted by $B_X^{important}$.

For basic behavior string s_X, 500 daily behavior strings are simulated based on the behaviors in s_X and the associated important behavior set $B_X^{important}$. According to the investigation, the regular string has relatively fixed patterns of important behaviors while the irregular string is opposite.

Table II lists the simulation parameters for the derivations of each basic behavior string. There are 500-day behavior strings simulated while the ratio of the number of irregular strings to that of all strings ranges from 0.1 to 0.3. Based on the previous investigation, regular behavior strings are simulated by changing a small number of the ordinary behavior characters and the

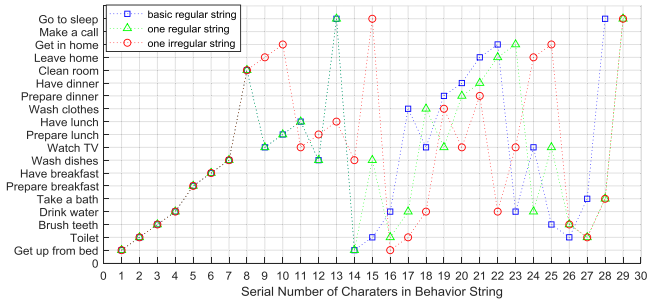

Fig. 5. Example of the derivations of one regular and one irregular strings from basic regular string.

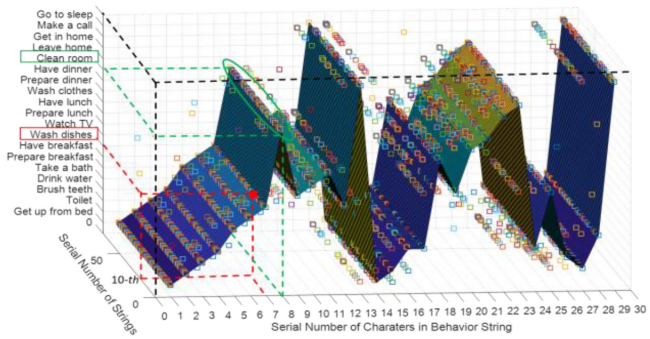

Fig. 6. The derivations regular strings which are 75 irregular strings selected from 365 days.

ratio of the number of changed ordinary characters to the length of basic regular strings ranges from 0.05 to 0.3. The irregularity is caused by skipping or increasing important behaviors and the ratio of the number of changed important characters to the length of basic regular string ranges from 0.15 to 0.4. Since the operations on strings are not the focus of this paper, the weights of all the operations, including insertion, deletion, and modification, are set to 1. The behavior weight will be set according to the investigation. Table II summarizes the parameter settings.

Fig. 5 depicts a basic regular behavior string s_X. Based on s_X, two strings are derived, including one regular and one irregular strings, with $\varepsilon = 0.05$ and $\delta = 0.15$. The basic regular string s_X and the derived regular and irregular string are marked with blue, green and red colors, respectively. Compared with the s_X, the derived regular string is changed by inserting a 'Wash dishes' behavior at the 15-*th* serial number. The red irregular behavior string, as compared with s_X, has two more important behaviors insertions, namely 'Leave home' and 'Get in home', at the 9th and 10th serial numbers, respectively. In addition, after 'Prepare dinner' at the 21-*th* serial number, two insertions 'Drink water' and 'Watch TV', and one deletion 'Have dinner' are changed. It likely indicates the bad mood or indigestion of the elder's irregularity features. The distance between the regular string and the basic regular string and distance between the irregular string and the basic regular string can be calculated by Equ. (14), which results in 1 and 18, respectively.

Fig. 6 further depicts the randomly selected 75 derivations of regular strings. The related parameters are set at $\varepsilon = 0.05$ and $\delta = 0.15$. In Fig. 6, three dimensions, including of 'Serial Number of Characters in String', 'Serial Number of Stings' and

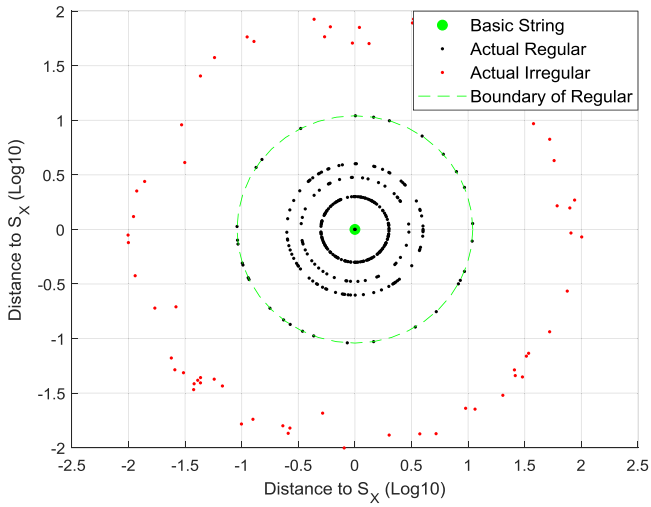

Fig. 7. Projecting strings onto a two-dimensional plane.

'Behavior', present time sequence, day ID, considered behavior set, respectively. Each small square (also called 'point') shown in the three-dimensional space represents the behavior occurred in some certain day. For instance, the 7-*th* char(behavior) occurred in 10-*th* day is 'Wash dishes', as the point marked with red solid square in Fig. 6. Another example shown in Fig. 6 is that points grouped by a green oval represent the behavior 'Clean room' almost occurred in every day of the 75 selected sampling days, which indicates the behavior regularity of the sampled elder. As shown in Fig. 6, there exists a general trend where similar behaviors are found, as lying on the colorful band.

Fig. 7 shows the logarithm of distances between 365 derived strings and the basic regular behavior string s_X. The related parameters are set at $\alpha = 0.2$, $\varepsilon = 0.15$, $\delta = 0.15$ and $w_b = 2$. Based on the distance from s_X, each string is projected onto a two-dimension plane. As shown in Fig. 7, the s_X is projected onto the origin, marked by a green dot. The derived regular strings, marked as black dots, are similar to the s_X and have smaller distances than derived irregular strings, as marked with red dots. According to the aforementioned assumptions, the boundary, which is marked with green color, includes all the regular strings and excludes all the irregular strings. That's the boundary that the proposed *IIRD* algorithm aims to find.

In the experiments, the detection of irregularity is achieved by two designs, called Model-based Detection and Direct Detection. In the experiment design of Model-based Detection, the derived strings are divided into two sets: training data and testing data. The Model-based Detection experiment is achieved in two stages. In the first stage, several unsupervised clustering mechanisms, including the proposed *IIRD*, DBSCAN, HC and K-means, are applied to build up two clusters: regular and irregular clusters. This result will be treated as a classification model which will be used to detect (or classify) if a given testing data string falls into the regular cluster. In the second stage, the testing data set is used to verify the accuracy of the detection

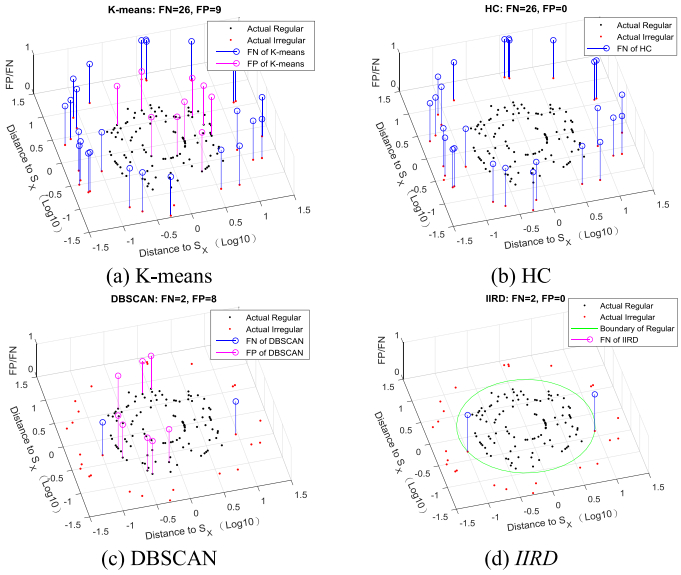

Fig. 8. False positive/false negative of four mechanisms.

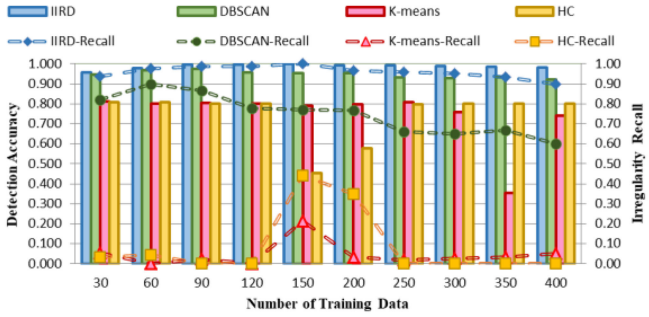

Fig. 9. Detection accuracy and irregularity recall (vs number of training data).

model. Different from the Model-based Detection design, the Direct Detection design treats all data as the inputs of the unsupervised clustering mechanisms, including the proposed *IIRD*, DBSCAN, HC and K-means. Herein, we notice that each data string has been already labeled with 'regular' or 'irregular' before the experiment. However, the labels are not used for training the features of the built model in our experiments, but used to verify the accuracy of the experimental results of the compared algorithms. Among the compared mechanisms, the DBSCAN is operated based on two parameters: *eps* and *minPts*. In the experiments, the two parameters are set at *eps* = 5 and *minPts* = 5 according to the length of string s_X and the value of ε.

B. Experimental Results

Fig. 8 adopts Model-based Detection design to compare the performances of the compared four mechanisms in terms of the numbers of the FP and FN. The related parameters are set at $N = 500$, $\alpha = 0.2$, $\varepsilon = 0.15$, $\delta = 0.15$ and $w_b = 2$. The training data contains 365 strings where 73 strings are irregular and 292 strings are regular. The testing data consists of 27 irregular strings and 108 regular strings. In comparison, the proposed *IIRD* algorithm outperforms DBSCAN, HC and K-means in terms of numbers of FN and FP. It occurs because that the proposed *IIRD* takes advantages of the behavior weights that can widen the distance between the regular and irregular strings. The DBSCAN has better performance than the K-means and HC, but not as good as the proposed *IIRD* algorithm. It occurs because that DBSCAM considers noise in advance and thus is robust to outliers. The K-means and HC have worse performance, as shown in Figs. 8(a) and 8(b), respectively. In particular, almost all the derived irregular strings are recognized as regular strings by the K-means and HC mechanisms, which leads to large values of FN.

The following defines *detection accuracy* $\mathcal{A}^{\mathfrak{M}}$ of an algorithm $\mathfrak{M}$, which is an evaluation parameter measuring the number of occurrences of FN and FP. Recall that notations γ_i and β_j denote the numbers of FP and FN, respectively, and notation $\mathbb{S}^{future}$ denotes the set of testing data. Exp. (20) presents the definition of $\mathcal{A}^{\mathfrak{M}}$.

$$\mathcal{A}^{\mathfrak{M}} = 1 - \left(\sum_{s_i \in R, \tilde{s}_j \in \tilde{R}} (\gamma_i + \beta_j) \Big/ \left| \mathbb{S}^{future} \right| \right) \quad s_i, \tilde{s}_j \in \mathbb{S}^{future} \tag{20}$$

The $\mathcal{A}^{\mathfrak{M}}$ has a value between 0 and 1. A large value of $\mathcal{A}^{\mathfrak{M}}$ indicates the total number of FP and FN obtained by applying algorithm $\mathfrak{M}$ is small.

Another evaluation parameter, called irregularity recall $\mathcal{R}^{\mathfrak{M}}$ which aims to measure the number of FN occurrences for a given set of irregular strings. The irregularity recall $\mathcal{R}^{\mathfrak{M}}$ represents the ratio of the correctly detected irregular strings to all the derived irregular strings. Exp. (21) presents the calculation of $\mathcal{R}^{\mathfrak{M}}$.

$$\mathcal{R}^{\mathfrak{M}} = \left(\alpha \times \left| \mathbb{S}^{future} \right| - \sum_{s_i \in R, \tilde{s}_j \in \tilde{R}} \beta_i \right) \Big/ \left(\alpha \times \left| \mathbb{S}^{future} \right| \right) \quad s_i, \tilde{s}_j \in \mathbb{S}^{future} \tag{21}$$

Fig. 9 compares four mechanisms in terms of the detection accuracy $\mathcal{A}^{\mathfrak{M}}$ and irregularity recall $\mathcal{R}^{\mathfrak{M}}$ by adopting Model-Based Detection design. The related parameters are set at $N = 500$, $\alpha = 0.2$, $\varepsilon = 0.15$, $\delta = 0.15$ and $w_b = 2$. The number of training data is ranging from 30 to 400. In general, the proposed *IIRD*, DBSCAN and K-means have a common trend that both the detection accuracy and irregularity recall increase first and then decrease with the number of training data, as shown in Fig. 9. This occurs because that the model constructed using a small set of training data inclines to be sensitive with the changes of data while the overfitting problem tends to appear when the number of training data is too large. As shown in Fig. 9, a reliable model can be constructed if the behavior data of recent 90 week days are provided. However, the detection accuracy of HC and K-means present some extreme values because that HC tends to obtain local optimum while the result of K-means sometimes relies on random selection of the initial cluster centers, which will lead to the result of low detection accuracy. In comparison, the proposed *IIRD* outperforms DBSCAN, HC and K-means in terms of detection accuracy and irregularity recall in all cases. It occurs because that the proposed *IIRD* extracts the feature of

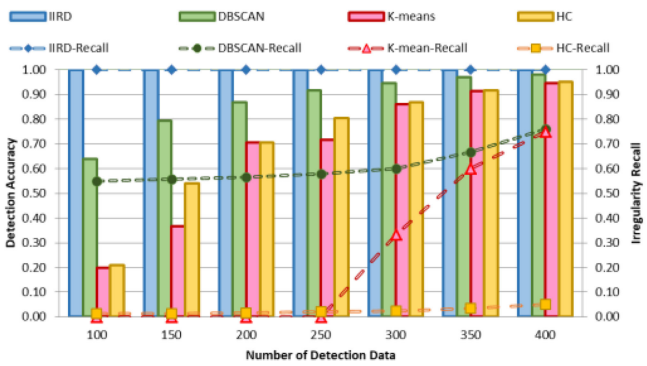

Fig. 10. Detection accuracy and irregularity recall (vs number of detection data).

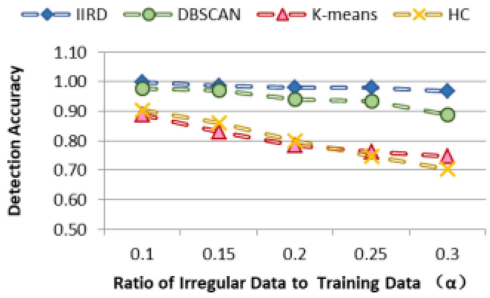

Fig. 11. Detection accuracy (vs ratio of irregular data to training data).

regularity from given data and then presents the feature as the distance between data. Based on the distance, the basic behavior group can be clearly identified. After that, the proposed *IIRD* finds the boundary between the regular and irregular data, which perfectly separates data into regular and irregular groups.

Fig. 10 adopts Direct Detection design, instead of Model-Based Detection design, to compare the four mechanisms in terms of the detection accuracy and recall. The settings of related parameters are identical to those in Fig. 9. In general, compared with Fig. 9, Fig. 10 has an opposite trend that the detection accuracy and irregularity recall increase with the number of detection data. This is because when the experiment adopts the Direct Detection design in which all data participate in the regular and irregular clustering, the features of regular and irregular data become easier to be identified when the number of detection data increases. In comparison, the proposed *IIRD* still outperforms the other mechanisms in terms of detection accuracy and irregularity recall in all cases. Combining Figs. 9 and 10, in general, the performances of the four compared mechanisms adopting the Model-Based Detection design is better than those adopting the Direct Detection design. Therefore, the following experiments will be carried on by applying Model-based Detection design.

Fig. 11 compares the four mechanisms in terms of detection accuracy with varied ratio α by applying Model-based Detection design. The numbers of training data and testing data are set at 365 and 135, respectively. The related parameters are set at $\varepsilon = 0.15$, $\delta = 0.15$ and $w_b = 2$. The parameter α, the ratio of irregular data to training data, varies ranging from

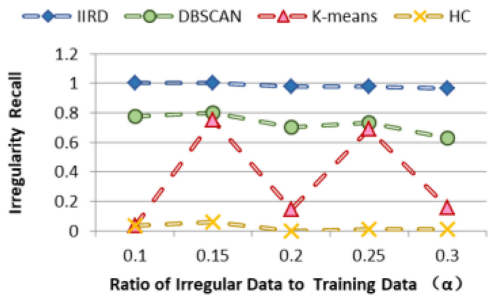

Fig. 12. Irregularity recall (vs ratio of irregular data to training data).

TABLE III
DISTANCE FEATURE OF DERIVED STRINGS

β	0.1	0.15	0.2	0.25	0.3
d_r^m	4.52	4.46	4.6	4.51	4.73
d_i^m	10.16	10.14	10.22	10.28	10.15
d_{ir}^m	5.64	5.68	5.62	5.77	5.42

0.1 to 0.3. In general, the detection accuracy and irregularity recall decrease when the ratio of irregular data to training data grows. This occurs because that the increase of α value indicates that the number of regular data is reduced. However, the similarity analysis and feature identification highly depend on the regular data. Therefore, the decreasing number of regular data leads to low detection accuracy. In comparison, the proposed *IIRD* outperforms the other compared mechanisms. This occurs because that *IIRD* assigns important behaviors with higher weights, sorts the data strings based on similarity distance as well as creates the basic regular group, which can better distinguish the regular and irregular strings.

Fig. 12 further investigates the irregularity recall of the four compared mechanisms. Herein, we notice that the irregularity recall decreases with the ratio of irregular data to training data. It is interesting that the curve of K-means is zigzagged. The closer observation of data depicted in Table III can help understand this phenomenon. Let notations d_r^m, d_i^m and d_{ir}^m represent the average distance of regular strings, average distance of irregular strings, and the difference between the former two, respectively. The number of FP will be decreased in the following two conditions, leading to the increasing of irregularity recall. The first condition is that the average distance between irregular strings is relative small, leading to the right clustering result where all irregular strings are grouped into the same group. As shown in Table III, the value of d_i^m, as marked with red color in Table (2, 2), falls in the first condition. Another condition is that the distance between regular and irregular strings is large. This condition can also lead to the situation that the regular and irregular strings are separated into the different groups. As shown in Table III, the value of d_{ir}^m, as marked with blue color in Table (3, 4), falls in the second condition.

Fig. 13 further compares the four mechanisms in terms of detection accuracy by adopting the Model-based Detection design. In this experiment, the diversity of training data is changed based on varying the values of parameters ε and δ. The parameter ε ranges from 0.1 to 0.25 while parameter δ ranges from 0.15

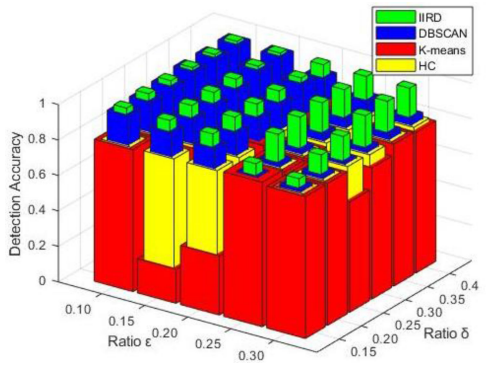

Fig. 13. Detection accuracy of four mechanisms (vs ratio ε and ratio δ).

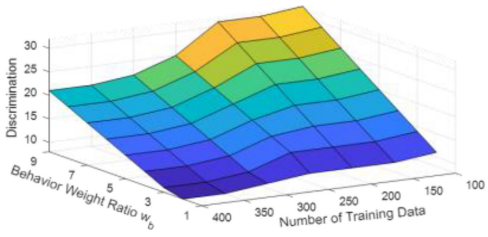

Fig. 14. Discrimination of *IIRD*.

to 0.4. Other related parameters are set at $N = 500$, $\alpha = 0.2$ and $w_b = 2$. The number of training data is 365. In general, there is a trend for the proposed *IIRD* and DBSCAN that the detection accuracy decreases with parameter ε and increases with the parameter δ. It occurs because that the difference between regular and irregular strings diminishes when the difference between ε and δ becomes small. In comparison, the proposed *IIRD* outperforms the other three mechanisms in all cases. This verifies that the robustness of the proposed *IIRD* is high, even though diversity exists in the behavior strings.

From the experiments, we observed that detection accuracy and irregularity recall are related to the distance between the irregular strings and regular strings. Let $D^{\mathfrak{M}}$ denote the discrimination of algorithm $\mathfrak{M}$ for regularity and irregularity of training data. Let $R_{\mathfrak{M}}$ and $\tilde{R}_{\mathfrak{M}}$ denote the regular strings and irregular strings after applying the algorithm $\mathfrak{M}$, respectively. D^{M} is defined as the minimum distance between any string $s_i \in R_{\mathfrak{M}}$ and any string $\tilde{s}_j \in \tilde{R}_{\mathfrak{M}}$ as shown in Exp. (22).

$$D^{\mathfrak{M}} = \min_{\forall s_i \in R_{\mathfrak{M}}, \forall \tilde{s}_j \in \tilde{R}_{\mathfrak{M}}} d\left(s_i, \tilde{s}_j\right) \tag{22}$$

Fig. 14 presents the discrimination of the proposed *IIRD* with varied numbers of training data and behavior weight ratios w_b by adopting the Model-based Detection design. The related parameters are set at $N = 500$, $\varepsilon = 0.15$ and $\delta = 0.15$. The number of detection data ranges from 150 to 400 while the behavior weight ratio w_b ranges from 1 to 9. In general, there is a trend that the discrimination of the proposed *IIRD* increases with the behavior weight ratio w_b. It occurs because that the impact of important behaviors on discrimination is increased with the weight ratio. As a result, large values of w_b lead to large differences between regularity and irregularity. On the other hand, the discrimination of the proposed *IIRD* decreases with the number of training data. It occurs because that a large number of training data increases the diversity of the regular and irregular data and hence leads to small discrimination.

VII. Conclusion and Future Work

This paper proposes a mechanism, called *IIRD,* which aims to detect the implicit irregularity of daily behaviors to support the homecare of the elderly. In order to formulize the irregularity, the distance and similarity between behaviors are defined and applied to extract the regularity features of behavior data. Moreover, the behavior weight is proposed to amplify the difference between regularity and irregularity. The proposed *IIRD* consists of two phases, named *CBRG* and *ERG*. The *CBRG* phase mainly extracts the feature of regularity from given data and then presents the feature as the distance between data. To reduce the computation, a basic regular group is constructed. Then the ERG phase finds the boundary between the regularity and irregularity by gradually expending the regular group. When the expansion of the regular group stops, the detection model is built up by the proposed *IIRD*. Extensive experiments verify the robustness of the proposed *IIRD* and show that *IIRD* outperforms DBSCAN, K-means and HC in terms of the detection accuracy and irregularity recall.

Our future work will integrate the proposed *IIRD* mechanism to the real-life long term activity monitoring system to measure its practical impact on the elderly. Moreover, we will consider statistical approaches to determine the importance of the behaviors. Further, we will consider the multi-dimensional features of daily behaviors and provide refined characterizations of the elderly behaviors.

References

[1] F. Zhou, J. Jiao, S. Chen, and D. Zhang, "A case-driven ambient intelligence system for elderly in-home assistance applications," *IEEE Trans. Syst. Man Cybern.*, vol. 41, no. 2, pp. 179–189, Mar. 2011.

[2] J. J. Liu *et al.*, "BreathSens: A continuous on-bed respiratory monitoring system with torso localization using an unobtrusive pressure sensing array," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1682–1688, Sep. 2015.

[3] T. S. Barger, D. E. Brown, and M. Alwan, "Health-status monitoring through analysis of behavioral patterns," *IEEE Trans. Syst. Man Cybern.*, vol. 35, no. 1, pp. 22–27, Jan. 2005.

[4] D. Naranjo-Hernández, L. M. Roa, J. Reina-Tosina, and M. Á. Estudillo-Valderrama, "SoM: A smart sensor for human activity monitoring and assisted healthy ageing," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3177–3184, Nov. 2012.

[5] C. Tsirmpas, A. Anastasiou, P. Bountris, and D. Koutsouris, "A new method for profile generation in an intemet of things environment: An application in ambient-assisted living," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 471–478, Dec. 2015.

[6] A. Nazábal, P. García-Moreno, A. Artés-Rodríguez, and Z. Ghahramani, "Human activity recognition by combining a small number of classifiers," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 5, pp. 1342–1351, Sep. 2016.

[7] Y. Chen, L. Yu, K. Ota, and M. Dong, "Robust activity recognition for aging society," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1754–1764, Nov. 2018.

[8] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst. Man Cybern.*, vol. 42, no. 6, pp. 790–808, Nov. 2012.

[9] C. Zhu and W. Sheng, "Realtime recognition of complex human daily activities using human motion and location data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2422–2430, Sep. 2012.

[10] G. Valenza *et al.*, "Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1625–1635, Sep. 2014.

[11] N. Dey, A. S. Ashour, F. Shi, S. J. Fong, and R. S. Sherratt, "Developing residential wireless sensor networks for ECG healthcare monitoring," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 442–449, Nov. 2017.

[12] L. Fanucci *et al.*, "Sensing devices and sensor signal processing for remote monitoring of vital signs in CHF patients," *IEEE Trans. Instrum. Measu.*, vol. 62, no. 3, pp. 553–569, Mar. 2013,

[13] A. M. Nia *et al.*, "Energy-efficient long-term continuous personal health monitoring," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 1, no. 2, pp. 85–98, Apr. 2015.

[14] M. Yu, A. Rhuma, S. Naqvi, L. Wang, and J. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1274–1286, Dec. 2012.

[15] J. Wang *et al.*, "An enhanced fall detection system for elderly person monitoring using consumer home networks," *IEEE Trans. Consum. Electron.*, vol. 60, no. 1, pp. 23–29, Feb. 2014.

[16] B. R. Greene, S. J. Redmond, and B. Caulfield, "Fall risk assessment through automatic combination of clinical fall risk factors and body-worn sensor data," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 725–731, Mar. 2017.

[17] Q. Guan, C. Li, X. Guo, and B. Shen, "Infrared signal based ederly fall detection for in-home monitoring," in *Proc. 9th Int. Conf. Intell. Human-Machine Syst. Cybern.*, 2017, pp. 373–376.

[18] D. Elbert *et al.*, "An approach for detecting deviations in daily routine for long-term behavior analysis," in *Proc. 5th Int. Conf. Pervasive Comput. Technol. Healthcare*, 2011, pp. 426–433.

[19] A. R. M. Forkan *et al.*, "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living," *Pattern Recognit.*, vol. 48, no. 3, pp. 628–641, Mar. 2015.