Article

Leveraging wearable sensors for human daily activity recognition with Stacked Denoising Autoencoders

Qin Ni †, Zhuo Fan †, Lei Zhang *, Chris D. Nugent, Ian Cleland, Yuping Zhang and Nan Zhou

[1] Qin Ni, Zhuo Fan, Yuping Zhang and Nan Zhou are with College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, PR China; niqin@shnu.edu.cn, 1000459457@smail.shnu.edu.cn, yp_zhang@shnu.edu.cn and zn1016807290@outlook.com

[2] C. D. Nugent and I. Cleland are with the School of Computing and Mathematics, University of Ulster, Belfast, BT370QB, U. K; cd.nugent@ulster.ac.uk and i.cleland@ulster.ac.uk

* Correspondence: Lei Zhang is with College of Information Science and Technology, Donghua University, Shanghai 201620, China; lei.zhang@dhu.edu.cn

† These authors contributed equally to this work and should be considered co-first authors.

Version August 5, 2020 submitted to Journal Not Specified

Abstract: Activity recognition has received considerable attention in many research fields, such as industrial and healthcare fields. However, many researches about activity recognition have focused on static activities and dynamic activities in current literature, while, the transitional activities, such as stand-to-sit and sit-to-stand, are more difficult to recognize than both of them. Consider that it may be important in real applications. Thus, a novel framework is proposed in this paper to recognize static activities, dynamic activities and transitional activities by utilizing stacked denoising autoencoders (SDAE), which is able to extract features automatically as a deep learning model rather than utilize manual features extracted by conventional machine learning methods. Moreover, resampling technique (random oversampling) is used to improve problem of unbalanced samples due to relatively short duration characteristic of transitional activity. The experiment protocol is designed to collect twelve daily activities (three types) by using wearable sensors from 10 adults in smart lab of Ulster University, the experiment results show the significant performance on transitional activity recognition and achieve the overall accuracy of 94.88% on three types of activities. The results obtained by comparing with other methods and performances on other three public datasets verify feasibility and priority of our framework. This paper also explores the effect of multiple sensors (accelerometer and gyroscope) to determine the optimal combination for activity recognition.

Keywords: Activity recognition; transitional activities; stacked denoising autoencoders; wearable sensors; resampling technique.

---

1. Introduction

Activity recognition, as an important application, has been received considerable attention and is widely used in many research fields, such as industrial and healthcare fields [1]. In general, activity recognition means that people's daily behavior types are identified by a large amount of human behaviour informations which are collected by a variety of channels which include camera [2], microphone [3], sensor [4], [5] and so on. Wearable sensors which are currently embedded in smart devices are widely used. In this paper, our experimental data is collected by using wearable sensors. Moreover, it is prevailing to utilize wearables which have embedded sensors to collect data, such as smartphones [6], due to the reason that wearable devices are convenient to carry and suitable for long-term monitoring.

The wearable sensors which are used to collect raw data are varied [7], [8], [9], [10]. For example, there are accelerometer, gyroscope, magnetometer, barometer, light, proximity, GPS and so on. In this paper, our experiment utilizes accelerometer and gyroscope to collect data. The data from the two sensors which are tri-axial is three-dimensional [11], which describes meaningful human movement tendency in realistic space. The accelerometer measures acceleration of subject in experiment to analyze human motion states. The gyroscope is commonly used to measure the angle of rotation and its rate of change. Thus, the combination of the two sensors is able to collect informative data which includes more representative features.

The activity recognition is to identify various types of human daily activities which are grouped according to different points of view. For example, the main monitoring activities are classified into static activity, dynamic activity [12] and transitional activity according to activities' characteristics, which is also studied in this paper. Generally, the static activities (e.g., standing, sleeping) and the dynamic activities (e.g., walking, running) are focused in current researches, which have achieved significant performance. But the transitional activities (e.g., stand-to-sit, sit-to-stand) are more complicated due to their short duration, which cause easily loss of performance. Although transitional activities have some overlaps with static and dynamic activities, it is indispensable to detect transitional activities in real life. A variety of activities are usually performed by people in a period of time, containing various transitional activities which include the end of an activity and the beginning of next activity. The transitional activity can also play a huge role in practical application. For patients, monitoring the transitional period makes great sense, which can help diagnosing patient's physical state in real-time [13]. Nowadays, there is already clinical application that using Sit-to-Stand Test to judge safety and reliability with older intensive care unit patients at discharge [50]. Thus, in this paper, we would focus on not only the static and dynamic activities recognition, but also transitional activities recognition. Meanwhile, transitional activities are intermediate and short-term transitions between two activities, thus the number of instances of transitional activities is usually far less than static and dynamic activities in an experiment, which results in drops of accuracy due to the imbalance of sample number. Therefore, the resampling technique [14] is usually used to improve unbalanced problem.

The activity recognition is often conducted by many conventional machine learning algorithms, such as KNN [15], [16], SVM [17], Random Forest (RF) [18] and K-means [19]. The universal procedures which perform activity recognition by applying machine learning algorithms contain data gathering, data preprocessing, data segmentation, feature extraction, classifier training and testing. Feature extraction is the core of procedure, which has a significant influence on classification performance. In this step, features are often extracted manually. Even though conducting activity recognition by applying machine learning algorithms have reached reasonable performance, they need high time-cost and computing resource for extracting features manually. Thus, deep learning methods which are able to extract features automatically have received more widespread attention on human activity recognition [20], [21].

In this paper, the stacked denoising autoencoder (SDAE) [7] which is a kind of deep learning model is utilized to perform activity recognition on account of its advantages. Firstly, it is able to extract features automatically comparing with conventional machine learning methods. Secondly, it is able to reduce data dimension to simplify calculation while preserving information, thus providing better conditions for learning in the final training stage. Thirdly, it utilizes encoder to compress unnecessary data and uses decoder to reconstruct data, which is able to extract useful higher level of representations [22]. The dataset we used is collected by using wearable accelerometer and gyroscope. Moreover, our activities' types in this study contain three static activities (standing, sleeping and watching TV), three dynamic activities (walking, running and sweeping) and six transitional activities (stand-to-sit, sit-to-stand, stand-to-walk, walk-to-stand, lie-to-sit and sit-to-lie).

The main contributions in this study are described as follows:

1. In this paper, besides static and dynamic activities recognition, we also focus on utilizing deep learning models to perform transitional activities recognition which are more difficult and complicated than other two cases.

2. We have improved the problem of unbalanced samples due to relatively short duration characteristic of transitional activity by applying the resampling methods and adopted varied resampling methods to search optimal scenario.
3. A novel framework based on stacked denoising autoencoder is utilized to recognize three types of activities, which has achieved significant performances and compared with other classical methods to verify the effectiveness of SDAE model on activity recognition, especially for transitional activities.

The rest of this paper is organized as follows: Section 2 will introduce existing related works about activity recognition and deep learning methods. Section 3 describes the overall framework about architecture of stacked denoising autoencoder model. Section 4 will design the experimental procedures and propose the methods of data preprocessing. The results of experiment will be also presented and analyzed. Finally, the conclusion will be drawn in Section 6.

2. Related work

2.1. Activity recognition

In activity recognition, the widespread usage of portable and wearable smart devices such as smartphones and smartwatches which are embedded with sensors have enabled the human activity data to be gathered large-scale. Before smart devices became ubiquitous, human activity recognition was cumbersome by attaching multiple sensors to a user's body [23]. However, with the proliferation of affordable smart devices, it has become easier to collect information about user activity through sensors embedded in the device. Various methods of detecting human activities using acceleration sensors of smartphones have been proposed [24]. In this study, the data we used came from accelerometer and gyroscope, which included three types of activities (static, dynamic and transitional activities). Transitional activity recognition has greater complexity owing to short duration. Jorge-L took feature vectors as input for activity prediction using an SVM and utilized a filtering module to deal with transitions and fluctuations on the SVM module output [25]. Machine learning methods are widely used to predict transitional activities, such as SVM, KNN, Random Forest and so on. Junhuai Li proposed a method based on conventional machine learning algorithms, they first utilized K-means to distinguish between basic activities and transitional activities and then used Random Forest classifier to recognize the two types of activities accurately [26]. In this paper, we exploit a deep learning model, stacked denoising autoencoder, to recognize transitional activities. Feature extraction influences the algorithm performance, computation time and complexity [27]. Conventional machine learning classification methods are often used to complete tasks based on the extracted features. For instance, a paper exploited discrete cosine transform to extract effective features and used PCA to reduce the dimension of feature [28]. At last, they applied multi-class Support Vector Machine (SVM) to recognize human activity. Decision Tree (DT) classifier [29] has the preferable performance in recognizing daily activities. Deep neural language model is exploited for the discovery of interleaved and overlapping activities [30]. The model builds hierarchical activities and captures the inherent complexities in activity details. Backpropagation techniques was used to train feedforward neural for complex human activity recognition in smart home environment [31]. Although the algorithm outperformed the Hidden Markov Model [32], DT and SVM, it requires a combination of manual feature extraction to achieve high performance accuracy. However, for these conventional machine learning algorithms, there have been a drawback that they need to extract features manually. There is no doubt that extracting features manually increases the time consuming, energy and model building procedure. Thus, in this direction, deep learning model which extracts features automatically has received more considerable attention.

2.2. Deep learning methods

With the emergence of deep learning and increased computation powers, deep learning methods are being widely adopted for automatic feature learning in diverse areas like natural language processing, image classification and computer vision. Recently, it is used to extract feature automatically from data collected by mobile and wearable sensors, and then classify human activity. In many researches, Convolutional Neural Network (CNN) [33], [34], [35] and Denoising Autoencoder [36], Restricted Boltzmann Machines [37] were used to extract features from data automatically. In this study, stacked denoising autoencoder technique is utilized for activity recognition. Denoising autoencoder was proposed P. Vincent, et al [38] and then A. Wang, et al [36] proposed denoising autoencoder technique to learn underlying feature representation in sensor data and then integrate it with a classifier trained into single architecture to obtain powerful recognition model. Stacked autoencoder was introduced to analyze human activity from data of motion sensors, which has a high performance [39]. Therefore, autoencoder methods have demonstrated significant approaches for automatic feature representation to learn latent feature representation for human activity monitoring and detection approach. What' more, stacked denoising autoencoder was utilized to recognize 8 activities, which achieved high accuracy [7]. However, it also only focused on static and dynamic activities not transitional activities. Generally, stacked autoencoder provides compact feature representation from continuous unlabelled sensor streams to achieve robust and seamless implementation of human activity recognition system [27]. In addition to the above mentioned, deep learning methods also include Recurrent Neural Network (RNN) [40], [41], Long Short-Term Memory (LSTM) [42], Deep Belief Networks (DBN) [43] and so on. In Table 1, some references that utilized deep learning methods were listed. These deep learning models also have the advantage that extract features automatically.

Table 1. References of applying deep learning methods on activity recognition. (The Acc is accelerometer, the Gyr is gyroscope, the Mag is magnetometer and the Bar is barometer.)

Reference	Method	Sensor	Activity classes	Accuracy
Gu et al. [7]	SDAE	Acc+ Gyr+ Mag+ Bar	stilling, running, walking, upstairs, downstairs, upElevator, downElevato, falsemotion	94.34%
Charissa et al. [33]	CNN+ tFFT	Acc+ Gyr	walking sitting, upstairs, downstairs, standing, laying	95.75%
Song-Mi et al. [34]	1D-CNN	Acc	run, walk, still	92.71%
Mario [35]	CNN	Acc	walking, sitting, jumping, lying, climbing_up, standing, running, climbing_down	94%
Masaya Inoue et al. [40]	RNN	Acc	standing, sitting, downstairs, laying, walking, upstairs	95.42%
Yao et al. [41]	CNN+ RNN	Acc+ Gyr	standing, climbStair-down, biking, walking, sitting, climbStair-up	94.20%
Yu et al. [42]	LSTM	Acc+ Gyr	walking, upstarirs, standing, sitting, downstairs, laying down	93.79%
Zhang et al. [43]	DBN	Acc	walking, running, standing, sitting, upstairs, downstairs, lying	98.60%

It is obvious that these experiments which adopted deep learning models in Table 1 achieved acceptable accuracy. Thus, deep learning methods not only extract features automatically, but also improve the experiments' performance. Moreover, compared to these deep learning models, SDAE model is able to improve effectively the problem of gradient disappearance and compress data by utilizing encoder, which extracts more representative features. Moreover, the SDAE model can reduce dimension of data to simplify calculation and improve operation efficiency. Thus, in this study, we would focus on stacked denoising autoencoders to extract feature and recognize human activity.

3. Materials and Methods

In this section, the overall framework of experiment and the method (SDAE) are described in detail, which are used to distinguish three types of activity on human activity recognition. The SDAE is a kind of deep learning model which extracts features automatically. The framework includes three parts: Data preprocessing and SDAE model, which is illustrated in Figure 1.

3.1. Data preprocessing

After collecting data, data preprocessing is a critical procedure, which includes resampling and standardized as is shown in Figure 1. In our experiment, the data is collected in a controlled lab environment by using wearable accelerometer and gyroscope. Ten healthy adults were chosen to perform twelve activities. At first, data segmentation was carried out to improve recognition accuracy. Then, owing to scarcity of transitional activity samples, the problem of unbalanced samples will drop recognition accuracy. In order to improve this problem, the resampling technique was applied. Moreover, data standardization was also an important procedure to promote performance of experiment.

The dataset we collected contains three primary types of activities: static activity, dynamic activity and transitional activity. The static activity includes standing, sleeping and watching TV. The dynamic activity includes walking, running and sleeping. The transitional activity includes stand-to-sit, sit-to-stand, stand-to-walk, walk-to-stand, lie-to-sit and sit-to-stand. All experimental data is collected by leveraging accelerometer and gyroscope which have three-axis, $x_i = \{ac_i^x, ac_i^y, ac_i^z, gy_i^x, gy_i^y, gy_i^z\}$. The preprocessing procedure is needed to be carried out before the experimental data become input of the model.

3.1.1. Segmentation

Data segmentation is an essential procedure in human activity recognition, which is directly related to the recognition performance. In realistic world, human activities are continuous behaviors so that a single data sample at a time point can't reflect enough tendency of an activity. Thus, it is necessary to segment data before training neural network. More representative features of each activity are extracted after data segmentation. In our experiment, the sliding window with 50% overlap is utilized to segment dataset, which has a significant influence on recognition performance. Thus, the dataset is segmented by integrating n samples as a sequence according to sampling rate, namely

$$x_{ac} = \{ac_k^x, ..., ac_{k+n-1}^x, ac_k^y, ..., ac_{k+n-1}^y, ac_k^z, ..., ac_{k+n-1}^z\}. \tag{1}$$

$$x_{gy} = \{gy_k^x, ..., gy_{k+n-1}^x, gy_k^y, ..., gy_{k+n-1}^y, gy_k^z, ..., gy_{k+n-1}^z\}. \tag{2}$$

$$x_i = \{x_{ac}, x_{gy}\}. \tag{3}$$

where n is set to 256 in this paper, which is decided by sample rate of 102.4Hz and time interval of 2.5 seconds. A sliding window whose size is 256 with a 50% overlap is utilized to complete this process. After segmenting, every instance corresponds to a specific activity class. By this way, we acquired more powerful performance of classifier.

3.1.2. Resampling

After segmentation, we then choose whether to apply resampling technique or not according to the distribution of samples. That's because the accuracy of model drops when the distribution of samples is unbalanced. In general, in order to improve the unbalanced samples problem, the common used method is resampling technique, such as random undersampling, random oversampling and Synthetic Minority Oversampling Technique (SMOTE) [44], [45], [46].

Random undersampling is able to promote running efficiency by decreasing the number of training dataset when dataset is too big, but it may discard some important information, which results

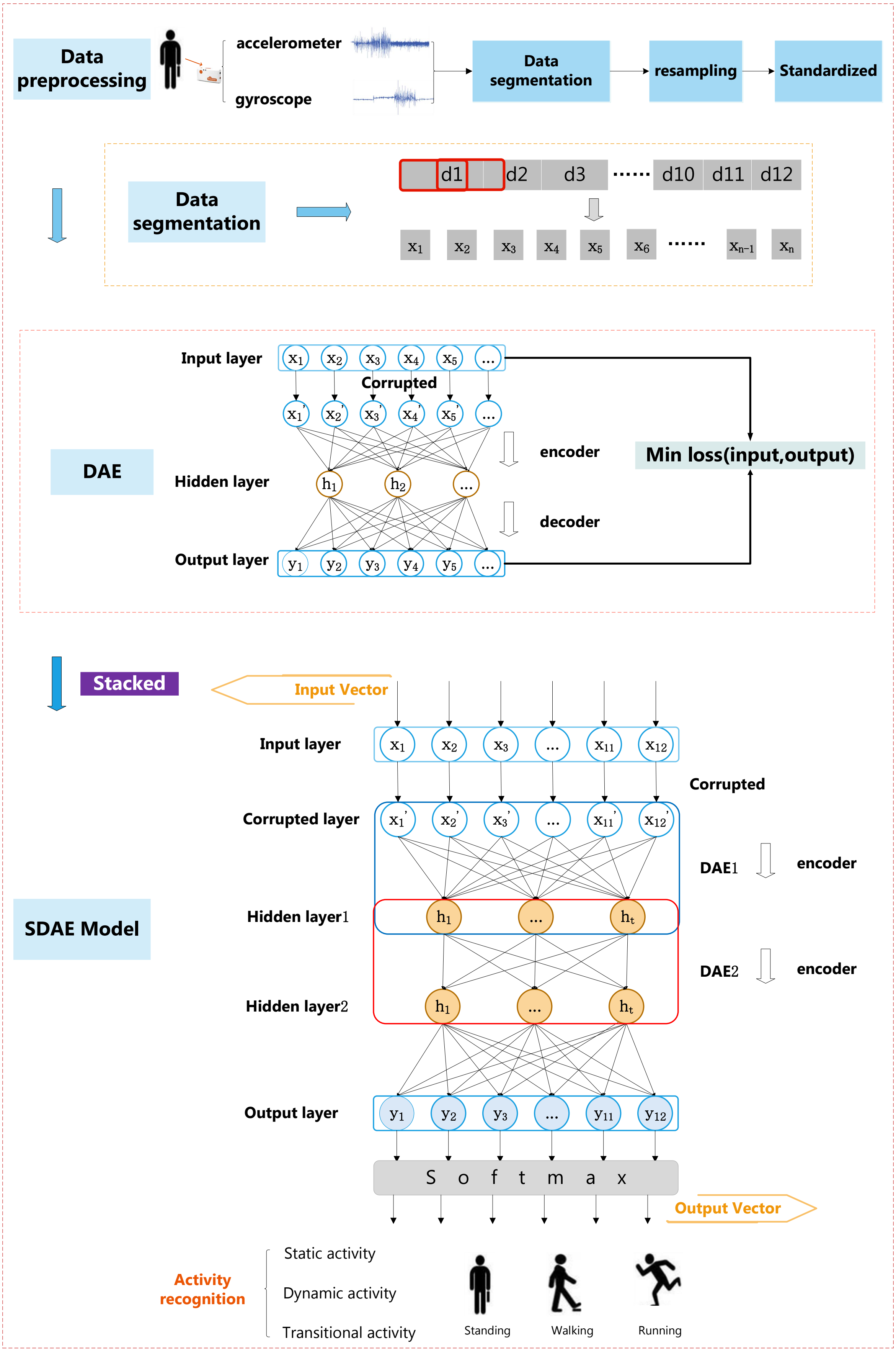

Figure 1. The overall framework of SDAE-based activity recognition

in inaccuracy of model. Random oversampling is able to duplicate some classes which just has few samples to increase the number of minority class and make sample size become balanced, but it may cause overfitting problem due to random duplicate. Despite all this, the SDAE model we utilized is able to decrease noise and therefore mitigate overfitting problem. The SMOTE method analyzes samples of minority class and synthesize new instances, then adds these synthetic samples to dataset. More specifically, for each sample in minority classes, their near neighbors can be acquired according to euclidean distance between them and other samples. Then some samples are chosen randomly to create new samples by applying equation (4). These examples would be added to original dataset.

$$x_{new} = x + rand(0,1) * \mid x - x_n \mid . \tag{4}$$

where x_n is a sample in minority class and x is a random nearest neighbor sample of x_n. $rand(0,1)$ represents a random number between 0 and 1, $\mid x - x_n \mid$ calculate euclidean distance between x and x_n. By this way, it alleviates overfitting problem effectively and not discards meaningful information. However, the SMOTE has also its faults, it may generate extra noise when creating new instances.

These methods are able to improve unbalanced samples problem and have specific drawbacks. We would discuss and analyze their performance in next section. Moreover, in order to improve the accuracy, data standardization was carried out before it was taken as input of model. In this step, each value is mapped to the 0-1 range by applying mean and standard deviation:

$$x^* = \frac{x - \mu}{\sigma}. \tag{5}$$

where x^* represents the input data, x represents concrete value, μ is mean and σ is standard deviation. In the end, the whole dataset was divided into training set, validation set and testing set with the ratio of 6:2:2 randomly.

3.2. Stacked Denoising Autoencoder

The stacked denoising autoencoder stacks input layers and hidden layers of multiple denoising autoencoders (DAE). Its architecture and overall design in this paper are illustrated in Figure 1. DAE includes an input layer, a hidden layer and an output layer, which utilizes encoder and decoder to acquire output. The encoder can encode input data to learn hidden features. Then the decoder is able to decode the output of encoder to restructure data. Its goal is to minimize the loss between input and output and make them keep consistent as far as possible. What's more, in order to avoid the problem that output is a direct copy of raw input, a denoising factor is used to corrupt input. The training process of stacked denoising autoencoder consists two steps: (i) pretraining; (ii) fine-tuning.

3.2.1. Pretraining

In this step, the stacked denoising autoencoder utilizes greedy layer-wise training to update parameters. In other words, every denoising autoencoder in stacked denoising autoencoder are trained separately. Moreover, every DAE's output will act as input of subsequent denoising autoencoder until the whole network is trained.

There are two processes in training process of denoising autoencoder: encoding and decoding. In the encoding process, our training data from accelerometer and gyroscope will be mapped to hidden layer by applying a sigmoid function, which is used to compress data, and then it is reconstructed by applying a sigmoid function in the decoding process. The error between the input data and the output data would be minimized by using gradient descent.

Formally, let x_i represents the input data, which have already been preprocessed and segmented. x_i is a 1*N vector. The N equals 3072 in this paper. The whole dataset is represented as follows:

$$S_n = \{x_0, x_1, ..., x_n\}. \tag{6}$$

Firstly, each x_i must be corrupted by a denoising factor a, which obtain $x_i^{'}$. The probability of each node lost in the layer is a. Then in the encoding section, $x_i^{'}$ is mapped to hidden layer by a sigmoid function f, namely

$$y = f(W_1 x_i^{'} + b_1). \tag{7}$$

where W_1 and b_1 are the weight matrix and bias respectively. Then, the y is mapped to output layer by a sigmoid function g, namely

$$z = g(W_2 y + b_2). \tag{8}$$

where W_2 and b_2 are the weight matrix and bias respectively. Then, the error between the output data and raw data is computed. The loss function we utilize is cross entropy which is written as:

$$L(x_i, z) = -\sum_{i=0}^{n}(x_i log(z) + (1 - x_i) log(1 - z)). \tag{9}$$

Last, the parameters of each layer are updated by applying gradient descent. After that, the pretraining process is done.

3.2.2. Fine-tuning

In this step, a softmax layer is added at the top of network, which is used to identify current type of activity. Then the whole network would be trained liked a Multilayer Perceptron (MLP) in a supervised manner by using labeled data. It is noted that parameters of pretraining progress are shared with fine-tuning progress. Then parameters of each layer are fine-tuned by applying backpropagation and gradient descent. In training progress, the whole dataset is divided into training set, validation set and test set in proportion. Training set is used to train model, validation set is used to detect optimal model specifications and test set is used to test the performance of model. The pseudo-code of human daily activity recognition in this paper is described in Algorithm 1.

Algorithm 1 Human activity recognition method with SDAE

Input: Raw dataset D
Output: Activity types of testing dataset

```
Data Preprocessing:
Segment the dataset according to sampling frequency
Apply the random oversampling
Standardize the dataset to obtain input vector x_i
Divide the dataset into training dataset D_train, validation dataset D_validate, test dataset D_test
Pretraining:
while l <=Hidden layers N_l do
    The dataset x_i in D_train is corrupted into x_i' by adding a denoising factor. Then let x_i' as input to
    train l-th layer of stacked denoising autoencoder.
    The output of l-th layer will be the input of l+1-th layer
    l+=1
end while
Fine-tuning:
Fine-tune the whole network by applying backpropagation. Utilize labeled dataset D_train to train
softmax layer.
Test:
Use the D_train and D_validate to train model and validate performance of model respectively.
Recognize the activity type of test data D_test.
```

3.3. Experimental design

In order to acquire efficient and useful information of activities, this experiment was carried out in a controlled laboratory environment. Ten healthy adults who came from the Ulster University were chosen to perform the twelve activities which were grouped into three classes: static activities, dynamic activities and transitional activities. These activities contain standing, sleeping, watching TV, walking, running, sweeping, stand-to-sit, sit-to-stand, stand-to-walk, walk-to-stand, lie-to-sit and sit-to-lie. Their brief descriptions are given in Table 2.

Table 2. The concrete descriptions of twelve human daily activities

Class	Activities	Description
Stationary Activities	standing	The subject stands still and maintains 5 minutes
	sleeping	The subject sleeps on the sofa for 5 minutes and is allowed to do some small movements, such as changing the lying posture
	watching TV	The subject watches TV for 5 minutes when he sits on the sofa in a comfortable position. And changing sitting posture is allowed
Dynamic Activities	walking	The subject walks on treadmill at constant speed for 5 minutes
	running	The subject runs on treadmill for 5 minutes
	sweeping	The subject sweeps in room with vacuum cleaner for 5 minutes
Transitional Activities	stand-to-sit	Standing for 15s , and then sitting on the sofa, repeat 15 times
	sit-to-stand	Sitting on the sofa for 10s, and then standing up, repeat 15 times
	stand-to-walk	Standing for 15s, and then walking for 15s, repeat 15 times
	walk-to-stand	Walking for 15s, and then standing for 15s, repeat 15 times
	lie-to-sit	Sitting on the sofa for 15s, and then lying down, repeat 15 times
	sit-to-lie	Lying on the sofa, and then sitting on the sofa, repeat 15 times

In designing progress, every subject was attached a tri-axial accelerometer and a tri-axial gyroscope on their left wrist which is one of the most common placements in activity detection. Then they performed twelve activities according to the description of Table 2. Moreover, the researcher informed subjects what activity to perform but not limit strictly how to perform. The raw data were usually collected by a Shimmer wireless sensor platform which contains an on-board tri-axial accelerometer and gyroscope with a configurable sampling rate up to 1 kHz and an amplitude range up to $\pm 6g$. In reality, our original data were collected with a sampling rate of 102.4Hz and an amplitude range of $\pm 2.0g$ (g = 9.8 m/s2), which was conducted by accelerometer and gyroscope sensors. Therefore, 4020288 samples were recorded over ten hours in the end.

4. Results

In this section, we will analyze the experimental results. After data preprocessing, the segmented dataset is taken as input of model to train. Meanwhile, various experimental methods will be used. In this step, we will discuss results in terms of different perspectives.

4.1. Experimental result without resampling

In our study, there are 12 activities, which contains standing, sleeping, watching TV, walking, running, sweeping, stand-to-sit, sit-to-stand, stand-to-walk, walk-to-stand, lie-to-sit and sit-to-lie. To analyze the performance of experiment in detail, according to the experimental result without resampling, the recognition accuracy of every activity as is shown in Table 3.

In experiment, the estimation metrics this paper adopted are *Accuracy*, *Precision*, *Recall* and *F1 score*. The *Accuracy* is the proportion between number of true labels model predicted and all true labels. The *Precision* of an activity is proportion between the number of correct predictions and the number of corresponding activity's label model predicted. The *Recall* of an activity is the proportion between the number of true label model predicted and the number of this activity's realistic label. The *F1 score* is the combination of the *Precision* and the *Recall*, which can be defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Table 3. Performance of each activity class recognition without resampling

	Accuracy(%)	**Precision(%)**	**Recall(%)**	**F1 score(%)**
standing	97.03	94.72	97.03	95.86
sleeping	97.32	**98.37**	97.32	97.84
watching TV	96.82	92.54	96.82	94.63
walking	86.75	85.71	86.75	86.25
running	95.73	91.81	95.73	93.73
sweeping	88.92	82.52	88.92	85.60
stand-to-sit	62.16	63.01	62.16	62.59
sit-to-stand	51.19	70.49	51.19	59.31
stand-to-walk	35.14	39.39	35.14	37.14
walk-to-stand	**26.51**	51.16	26.51	34.92
lie-to-sit	73.33	68.75	73.33	70.97
sit-to-lie	58.73	58.73	58.73	58.73

According to the Table 3, it shows that the static and the dynamic activities recognition acquire high performance in experiment, and static activities recognition yields higher accuracy than dynamic activities. But the performance of transitional activities recognition is much lower than static and dynamic activities. For example, the walk-to-stand activity recognition obtained a worst accuracy of 26.52% which is bolded and the sleeping activity recognition achieved best precision of 98.37% which is bolded. The main reason for this result is that our dataset are unbalanced. The maximum gap has been up to 246074 between transitional and other activities. The relatively short duration characteristic of transitional activities make the number of transitional activities be far less than static and dynamic activities. Meanwhile, it can be concluded that the problem of unbalanced samples will drop accuracy apparently.

Therefore, in order to promote the performance of experiment, the resampling techniques are utilized to improve the problem of unbalanced samples.

4.2. Performance enhancement with resampling

The primary reason of poor performance is that the distribution of samples in our study is extremely unbalanced. Compared to static and dynamic activities, the number of transitional activities is too little. As mentioned above, the maximum gap has been up to 246074 between transitional activity and dynamic activity, which decreases the accuracy obviously. Thus, in order to avoid the drops of accuracy, the resampling technique is proposed. The random oversampling, SMOTE and random undersampling are applied to enhance performance. Table 4 shows the sample size of raw data and the instance size after segmentation and applying resampling technique. Obviously, the resampling methods have made the size of each class be same by increasing or decreasing samples. Figure 2 demonstrates the final performance of model by applying resampling technology.

Table 4. The sample size of raw data and the instance size after segmentation and applying resampling technique

	Initial number	**After segmentation**	**Undersampling**	**SMOTE**	**Oversampling**
standing	307061	1198	238	1200	1200
sleeping	307109	1200	238	1200	1200
watching TV	306228	1196	238	1200	1200
walking	300457	1174	238	1200	1200
running	294676	1151	238	1200	1200
sweeping	302052	1179	238	1200	1200
stand-to-sit	61173	239	238	1200	1200
sit-to-stand	61035	238	238	1200	1200
stand-to-walk	61881	242	238	1200	1200
walk-to-stand	61640	242	238	1200	1200
lie-to-sit	61454	240	238	1200	1200
sit-to-lie	62089	242	238	1200	1200

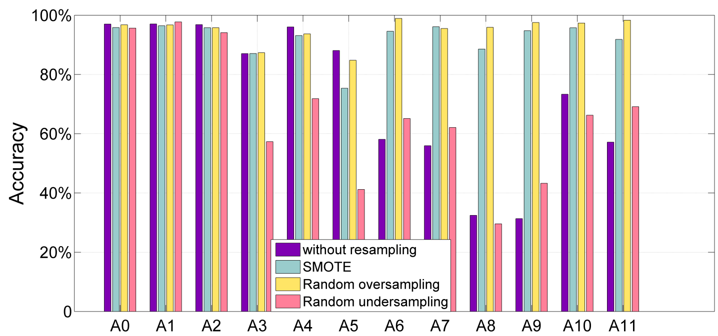

Figure 2. Accuracy of each activity class by applying random oversampling, SMOTE and without resampling. (A0-standing, A1-sleeping, A2-watching TV, A3-walking, A4-running, A5-sweeping, A6-stand-to-sit, A7-sit-to-stand, A8-stand-to-walk, A9-walk-to-stand, A10-lie-to-sit and A11-sit-to-lie)

According to Table 4, the sample size of different activities have large gaps and three resampling techniques have improved the imbalance sample problem. In Figure 2, it is obvious that performance of recognition has significant improvements after applying random oversampling and SMOTE. But the random undersampling has a negative impact on the recognition accuracy, which may be due to discard of too much useful information. Moreover, the random oversampling technique in this study has more powerful influence. Thus it may be known, the problem of unbalanced samples does have significant influence on results. When sample size is balanced, the accuracy has been greatly improved. Furthermore, it is noted that dynamic activities recognition performance has a slight drop. That's because dynamic activities are more complex than other activities and more likely to be misrecognized.

After resampling, the final classification performance of method in this study reached 94.88% as measured by *Accuracy*, 94.88% as measured by *Precision*, 94.88% as measured by *Recall*, 94.86% as measured by *F1 score*. Meanwhile, the classification performance of each activities is shown in Table 5.

Table 5. The classification performance of each activity by using random oversampling technique

	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
standing	96.75	95.61	96.75	95.73
sleeping	96.74	98.79	96.74	97.75
watching TV	95.77	98.37	95.77	96.85
walking	87.34	89.46	87.34	88.39
running	93.70	97.61	93.70	95.62
sweeping	**84.81**	89.97	84.81	87.31
stand-to-sit	**98.92**	95.80	98.92	97.34
sit-to-stand	95.53	96.07	95.53	95.80
stand-to-walk	95.92	93.39	95.92	94.64
walk-to-stand	97.53	93.92	97.53	95.69
lie-to-sit	97.34	96.32	97.34	96.83
sit-to-lie	98.31	93.57	98.31	95.88

From Table 5, the SDAE model acquires a significant performance and the recognition of every activity contributes high accuracy. The balanced samples have played a great role in recognition performance. And transitional activity recognition has achieved high accuracy by utilizing SDAE model. Furthermore, it is obvious that the performance of transitional activities precedes dynamic activities. The sweeping and walking reach lowest accuracy than other activities. We assume that compared to dynamic activity, the transitional activity has more obvious changes in a short time, which make the activities' features more representative. According to bolded results, the sweeping achieves a lowest accuracy of 84.81%, the stand-to-sit reaches a highest accuracy of 98.92%.

As is shown in Figure 3, a confusion matrix is given, the most obvious errors are that 15 instances of walking are incorrectly recognized as sweeping, and 19 instances of sweeping are misrecognized as walking. It is concluded that walking activity and sweeping activity have similar characteristics to some extent. Thus, the two activities are easily misrecognized and reach lowest accuracy of recognition.

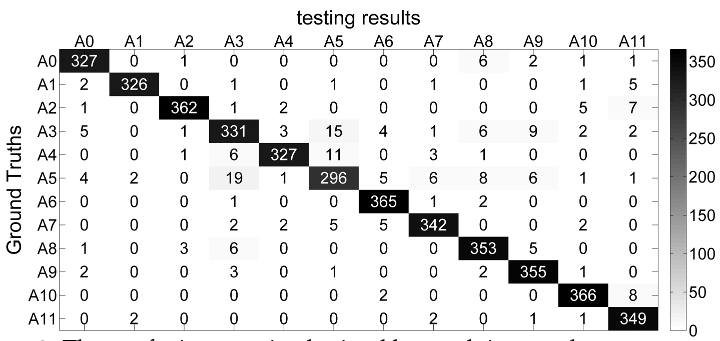

Figure 3. The confusion matrix obtained by applying random oversampling.

4.3. Hyperparameter analysis

When adopted different hyperparameters, the classifier had different performance. Therefore, some specific parameters were configured in experiment to obtain the best performance. These parameters we discussed contains iterations of training progress, pretraining learning rate and the number of hidden layer.

a) Iterations of training progress

In order to analyze effect of iterations, the accuracy changes of each activity are shown in Figure 4. With the increasing of iterations, the performance of each activity has no obvious changes. There are only three dynamic activities, walking, running, sweeping, that have changed and have worse performance than static and transitional activities. But it is noted that changes of different iterations are so small, which cannot transmit enough and useful information. Therefore, it is concluded that the number of iterations have not a significant influence on the performance of model. Moreover, considering the accuracy, time-cost and computational complexity, the choice of 200 iterations is a balance point. For each iteration, the training dataset is first used to train network, then the loss of validation set and testing set can be obtained based on current parameters to detect recognition performance. This progress will repeat 200 times.

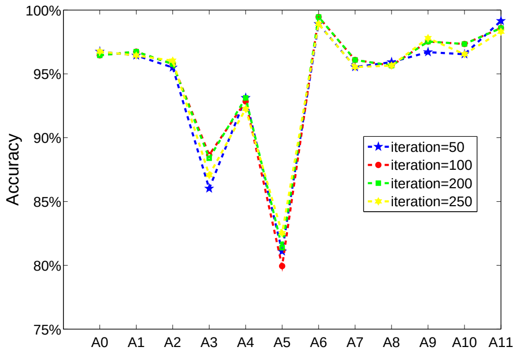

Figure 4. The recognition performance of each activity when using different iterations. (learning rate is set to 1×10^{-7}, number of hidden layer is set to 2)

b) Pretraining learning rate

Figure 5 indicates the influence of pretraining learning rate. When different pretraining learning rates are chosen, the performance had obvious changes. To analyze the influence of pretraining learning rate, the performance of every activity was tested when learning late is set to 1×10^{-3}, 1×10^{-4}, 1×10^{-5}, 1×10^{-6}, 1×10^{-7}, 1×10^{-8} respectively. As is shown in Figure 5, adopting different learning rates can lead to enormous fluctuations. Moreover, it demonstrates that the best performance is achieved at 1×10^{-7} level by comparing six parameters though the pretraining learning rate is set very low. Because the goal of pretraining is to acquire proper initial values of parameters and ensure that the fine-tuning process has better astringency and performance. Thus, the pretraining learning late doesn't need to be too high.

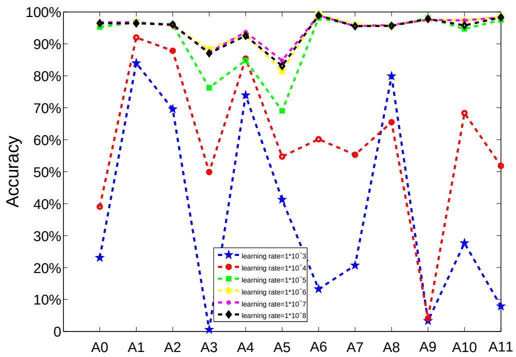

Figure 5. The recognition performance of each activity when using different learning rates. (Iterations is set to 200; number of hidden layer is set to 2)

c) The number of hidden layer

In Figure 6, the performance fluctuations are exhibited when selecting different numbers of hidden layer. It is obvious that the main disparity concentrate on walking and sweeping. The both activities recognition achieve lowest accuracy than other activities, which is also as mentioned above. With growth of hidden layer number, dynamic and transitional activities recognition performance are fluctuant while static activities are stable. Moreover, the fluctuation is not very obvious and computational cost will raise when using more layers. Thus, in order to achieve the trade-off between recognition accuracy and time-cost, we utilize 2 hidden layers to create the network.

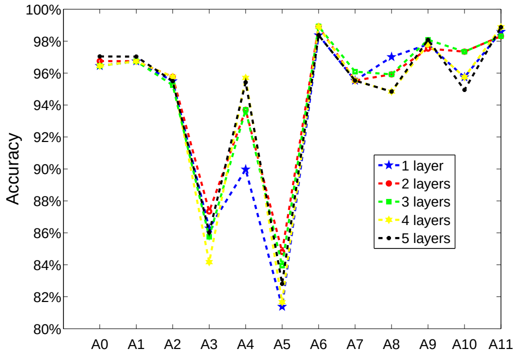

Figure 6. The performance of each activity recognition when selecting different numbers of hidden layer. (Iterations is set to 200, learning rate is set to 1×10^{-7})

Then Table 6 given a list of fixed hyperparameters of neural network architecture. Because some hyperparameters have a little influence on recognition accuracy according to experiment, thus we have discussed above three hyperparameters.

Table 6. Optimal hyperparameters for neural network

Hyperparameters	values
number of hidden layers	2
number of units per layer	500
pretraining learning rate	1×10^{-7}
fine-tuning learning rate	0.01
iteration	200
denoising factor	0.5
data segment size (seconds)	5

4.4. The influence of different sensors

In this experiment, the data we collected was from accelerometer and gyroscope. In general, adopting more types of sensors has a stronger influence on results. To analyze how the two sensors affect the performance of model, the Figure 7 is given to demonstrate the changes of specific activity when using different sensors. The performance of accelerometer and gyroscope is shown in Table 6.

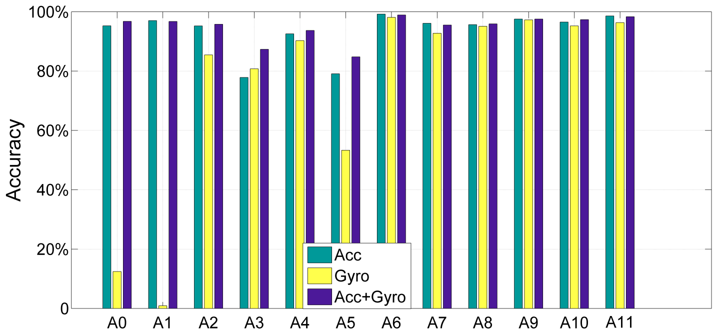

Figure 7. The performance changes of each activity recognition when adopting different sensors.

From Figure 7, the gyroscope improved indeed the accuracy of dynamic activities (walking, running, sweeping) recognition comparing to the single accelerometer. Moreover, it is obvious that when exploiting the data collected by gyroscope identified the transitional activities, the performance of dynamically activity recognition is acceptable, but the static activity recognition is not efficient. When the data collected by accelerometer is used, the accuracy of static and transitional activity recognition was acceptable, but the dynamic activity wasn't recognized precisely. However, when the data collected by accelerometer and gyroscope together is used, all classes of activity recognition obtained a significant accuracy. Thus, utilizing two sensors has a good influence on result than one sensor.

Table 7. Performance comparison of adopting different sensors

Sensors	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
Acc	93.36	93.35	93.36	93.27
Gyro	75.76	74.77	75.76	72.60
Acc+Gyro	94.88	94.88	94.88	94.86

According to the Table 7, although there is not much difference in the total performance of single accelerometer and two sensors, different activity recognition has different performance according to Figure 7. Specifically, combination of accelerometer and gyroscope obtained a best performance. The single accelerometer reached an acceptable result and the single gyroscope achieved an unsatisfactory result. However, it isn't able to conclude that the data from gyroscope is not worthy. In short, the combination of accelerometer and gyroscope contributes better performance than applying single sensor.

4.5. Comparison with other conventional methods

In order to demonstrate the superiority of method of this paper, we compared it with some conventional machine learning algorithms, including Support Vector Machine (SVM), Decision Tree (DT), Naive Bayesian Model (NBM), K-nearest neighbors (KNN), and other deep learning algorithm, such as CNN, LSTM and BiLSTM. In training progress of these conventional machine learning methods, the features we extracted are statistical features, including mean, maximum, minimum, variance and standard deviation. Moreover, the random oversampling technique is also used to balance samples in these methods. By comparing with other methods, their performances are shown in Table 8.

Table 8. Performance comparison between SDAE and other methods

Methods	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
SVM	90.95	90.81	90.95	90.60
DT	88.15	87.53	88.15	87.54
KNN	84.84	84.38	84.84	84.29
CNN	81.33	79.85	81.33	80.27
LSTM	81.63	83.56	81.63	81.62
BiLSTM	84.75	85.23	84.75	84.63
SDAE	94.88	94.88	94.88	94.86

From Table 8, it is obvious that the SDAE model obtains a best performance than other conventional machine learning algorithm (SVM, DT, KNN), which achieves accuracy of 94.88%. Compared to other conventional machine learning algorithm, SDAE model have a better performance. In conventional machine learning methods, SVM acquired the best performance that recognition accuracy exceeded 90% and other methods all have more than 80% of accuracy, which are lower than SDAE model. Moreover, conventional machine learning algorithm need to extract features manually and SADE is able to extract features automatically. Thus, in activity recognition, it is more precise and more convenient to utilize SDAE than conventional machine learning algorithms. Furthermore, compared to CNN and LSTM, SDAE model acquires more accurate result. Generally speaking, CNN and LSTM model can also achieve significant performance. In this situation, we analyze the accuracy of each activity after utilizing LSTM and CNN. We find that the watching TV has achieved the lowest accuracy of 42.39% than other activities which have achieved the accuracy of 76.02% - 95.87% when LSTM model is used. CNN model has faced a similar situation that the recognition accuracy of watching TV is lowest. Thus, we assume that these model can acquire significant performance in some activities (e.g. running), but they can't extract useful representation in some activities (e.g. watching TV) which have more similarities with other activities. We also use BiLSTM to recognize activities. It has a better performance than CNN and LSTM, but also has a lower accuracy than SDAE model. And we also find that they need more time to train than SDAE model. In this point, SDAE model is proper.

4.6. The performance of SDAE model on three public datasets

With the experimental verification, the behavioral dataset we collected reach a significant performance on activity recognition. However, considering a special situation that the result of experiment may depend on the dataset, and not on the model. Therefore, without loss of generality, we also conducted different experiments that exploited three public datasets which include human activity recognition using smartphones data set [47], activity recognition from single chest-mounted accelerometer data set [48] and UCI daily and sports dataset [49]. Table 9 shows their basic informations and their performance in experiments are given in Table 10.

Table 9. Basic informations of the three public datasets

Datasets	People	Classes	Sensors	Transitions
Smartphone	30	6	Acc+Gyro	No
Chest-mounted	15	7	Acc	No
UCI	8	19	Acc+Gyro+Mag	No

Table 10. Activity recognition performance of SDAE model on other public datasets

Datasets	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
Smartphone	97.15	97.19	97.15	97.15
Chest-mounted	89.99	89.96	89.99	89.83
UCI	95.26	95.42	95.26	95.15

According to the Table 10, the SDAE model obtained significant performances on three public datasets. Thus, we are able to conclude that the SDAE mode has strong applicability and the result of our experiment is convincing.

5. Discussion

The preceding results show the significant performance of SDAE model on transitional activities and illustrate that the accuracy will drop when dataset is unbalanced. The resampling technique provides feasible methods to improve this question and the accuracy can be greatly improved. In experiment, we perform multiple sets of contrast experiments which refer to resampling, hyperparameters, different sensors, multiple methods and datasets. According to the resampling experimental result, resampling technique is able to balance sample gap and enhance the recognition performance obviously. And SDAE model can alleviate overfitting problem due to its denoising characteristic. Different sensors have unique effect on activity recognition, the intergration of varied wearable sensors can have a positive influence on performance. Moreover, we have analyzed the reason that CNN and LSTM achieve lower performance than SDAE. We assume that these models cannot apply to all human daily activities and the architecture need to be adjusted according to demand. Furthermore, results on three public datasets based SDAE demonstrate the effectiveness and applicability of our framework. According to these experimental results, it can be concluded that SDAE is proper to recognize three types of activities in this paper, especially for transitional activities.

In activity recognition area, the transitional activity is as important as other activities and can be applied to healthcare filed. For example, it can play a significant role in human fall detection technique, detecting patient state and so on. We believe there will be more research results about transitional activities in different fields in the future.

6. Conclusions

In this paper, we utilized the stacked denoising autoencoder, a deep learning model which is able to extract useful features automatically and compress unnecessary data by encoder to extract more representative features and recognize human daily activities. Moreover, besides static and dynamic activities, we also focused on transitional activities recognition which were more difficult for recognizing. Furthermore, we utilized random oversampling technology to improve problem that samples were unbalanced due to the short-term characteristic of transitional activities. The experimental result we obtained demonstrated the SDAE model was able to achieve significant performance on transitional activities recognition and outperformed other conventional machine learning methods on human daily activity recognition. Meanwhile, we analyzed the influence of sensors respectively and verified the effectiveness of SDAE model on three public datasets for activity recognition, which proved indirectly that the result we obtained on our dataset is convincing.

In the future, we consider studying further special human's activity recognition, which will be able to be applied to specific fields, such as healthcare. Additionally, we also consider utilizing more kinds of sensors to collect data in future experiments. Furthermore, considering various channels of data gathering, the combination of activity recognition and computer vision is also an important research.

Author Contributions: Conceptualization, Qin Ni, Zhuo Fan and Lei Zhang; Data curation, Ian Cleland; Formal analysis, Lei Zhang; Investigation, Qin Ni; Methodology, Qin Ni and Zhuo Fan; Project administration, Qin Ni; Resources, Chris Nugent; Software, Zhuo Fan and Chris Nugent; Supervision, Yuping Zhang; Validation, Qin Ni and Ian Cleland; Visualization, Zhuo Fan; Writing – original draft, Zhuo Fan; Writing – review & editing, Qin Ni, Yuping Zhang and Nan Zhou.

Funding: This research was funded by Shanghai Sailing Program (Grant No.19YF1436800).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Y. Chen, L. Yu, K. Ota, M. Dong. Robust Activity Recognition for Aging Society, *IEEE Journal of Biomedical and Health Informatics*, 2018, 22(6): 1754-1764.
2. X. Yang, et al. Super Normal Vector for Human Activity Recognition with Depth Cameras, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(5): 1028-1039.
3. J. A. Ward, et al. Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 2006, 28(10): 1553-1567.
4. Y. Zheng, et al. Unobtrusive Sensing and Wearable Devices for Health Informatics, *IEEE Transactions on Biomedical Engineering*, 2014, 61(5): 1538-1554.
5. L. Chen, et al. Sensor-Based Activity Recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012, 42(6): 790-808.
6. Y. Chen, C. Shen. Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition, *IEEE Access*, 2017, 5: 3095-3110.
7. F. Gu, et al. Locomotion Activity Recognition Using Stacked Denoising Autoencoders, *IEEE Internet of Things Journal*, 2018, 5(3): 2085-2093.
8. Y. Chen, Y. Xue. A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer, *2015 IEEE International Conference on Systems*, Man, and Cybernetics, 2015.
9. M. Zeng, et al. Convolutional Neural Networks for human activity recognition using mobile sensors, *6th International Conference on Mobile Computing, Applications and Services*, 2014.
10. Y. Hsu, et al. Human Daily and Sport Activity Recognition Using a Wearable Inertial Sensor Network, *IEEE Access*, 2018, 6: 31715-31728.
11. D. Tao, et al. Ensemble Manifold Rank Preserving for Acceleration-Based Human Activity Recognition, *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(6): 1392-1404.
12. L. Xie, J. Tian, G. Ding, Q. Zhao. Human activity recognition method based on inertial sensor and barometer, *2018 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*, 2018.

13. R. Ali, L. Atallah, B. Lo and G.-Z. Yang. Transitional Activity Recognition with Manifold Embedding, *22009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, 2009.
14. M. Bolic, P.M. Djuric, S. Hong. Resampling algorithms and architectures for distributed particle filters, *IEEE Transactions on Signal Processing*, 2005, 53(7): 2442-2450.
15. J. Wannenburg, R. Malekian. Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, 47(12): 3142-3149.
16. P. Gupta, T. Dallas. Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer, *IEEE Transactions on Biomedical Engineering*, 2014, 61(6): 1780-1786.
17. Z. Chen, et al. Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM, *IEEE Transactions on Industrial Informatics*, 2017, 13(6): 3070-3080.
18. H. Xu, et al. Activity Recognition Method for Home-Based Elderly Care Service Based on Random Forest and Activity Similarity, *IEEE Access*, 2019, 7: 16217-16225.
19. S. Gaglio, et al. Human Activity Recognition Process Using 3-D Posture Data, *IEEE Transactions on Human-Machine Systems*, 2015, 45(5): 586-597.
20. T. Plötz, Y. Guan. Deep Learning for Human Activity Recognition in Mobile Computing, *Computer*, 2018, 51(5): 50-59.
21. J. Wang, et al. Device-Free Wireless Localization and Activity Recognition: A Deep Learning Approach, *IEEE Transactions on Vehicular Technology*, 2017, 66(7): 6258-6267.
22. P. Vincent, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, *Journal of Machine Learning Research* , 2010, 11(12): 3371-3408.
23. A. M. Khan, et al. A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer, *IEEE Transactions on Information Technology in Biomedicine*, 2010, 14(5): 1166-1172.
24. S. Dernbach, et al. Simple and complex activity recognition through smart phones, *Eighth International Conference on Intelligent Environments*, 2012.
25. Jorge-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra and D. Anguita. Transition-Aware Human Activity Recognition Using Smartphones, *Neurocomputing*, 2015.
26. J. li, T. Ling, et al. Segmentation and Recognition of Basic and Transitional Activities for Continuous Physical Human Activity, *IEEE Access* , 2019, (99):1-1.
27. H. F. Nweke, et al. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges, *Expert Systems With Applications*, 2018, 105: 233-261.
28. Z. He and L. Jin. Activity recognition from acceleration data based on discrete consine transform and SVM, *2009 IEEE International Conference on Systems, Man and Cybernetics*, 2009.
29. M.W. McCarthy, et al. Decision-tree-based human activity classification algorithm using single-channel foot-mounted gyroscope, *Electronics Letters*, 2015, 51(9): 675-676.
30. E. Rogers, et al. Towards a deep learning-based activity discovery system, *Dublin Institute of Technology*, 2016.
31. Ho. Fang, et al. Human activity recognition based on feature selection in smart home using back-propagation algorithm, *ISA Transactions*, 2014, 53(5):1629-1638.
32. K. Safi, et al. Recognition of different daily living activities using hidden Markov model regression, *2016 3rd Middle East Conference on Biomedical Engineering (MECBME)*, 2016.
33. C. A. Ronao and S. Cho. Human activity recognition with smartphone sensors using deep learning neural networks, *Expert Systems with Applications*, 2016, 59: 235-244.
34. S. Lee, S. M. Yoon and H. Cho. Human Activity Recognition From Accelerometer Data Using Convolutional Neural Network, *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017.
35. M. Mario. Human Activity Recognition Based on Single Sensor Square HV Acceleration Images and Convolutional Neural Networks, *IEEE Sensors Journal*, 2019, 19(4): 1487-1498.
36. A. Wang, et al. Human Activity Recognition in a Smart Home Environment with Stacked Denoising Autoencoders, *Web-Age Information Management*, 2016: 29-40.
37. J. Gao, et al. A novel feature extraction method for scene recognition based on Centered Convolutional Restricted Boltzmann Machines, *Neurocomputing*, 2015.
38. P. Vincent, et al. Extracting and composing robust features with denoising autoencoders, *Proceedings of the 25th international conference on machine learning*, 2008: 1096-1103.

39. X. Zhou, et al. Motion Recognition by Using a Stacked Autoencoder-Based Deep Learning Algorithm with Smart Phones, *International Conference on Wireless Algorithms, Systems, and Applications*, 2015: 778-787.
40. M. Inoue, S. Inoue and T. Nishida. Deep Recurrent Neural Network for Mobile Human Activity Recognition with High Throughput, *Artificial Life and Robotics*, 2016.
41. S. Yao, et al. DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing, *Proceedings of the 26th International Conference on World Wide Web*, 2017.
42. S. Yu and L. Qin. Human Activity Recognition with Smartphone Inertial Sensors Using Bidir-LSTM Networks, *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 2018.
43. L. Zhang, X. Wu and D. Luo. Real-Time Activity Recognition on Smartphones Using Deep Neural Networks, *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, 2015.
44. N. V. Chawla, et al. SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research*, 2002, 16(1): 321-357.
45. M. Galar, et al. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012, 42(4): 463-484.
46. L. Abdi and S. Hashemi. To Combat Multi-Class Imbalanced Problems by Means of Over-Sampling Techniques, *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(1): 238-251.
47. D. Anguita, et al. THuman Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine, *Ambient Assisted Living and Home Care*, 2012.
48. P. Casale, et al. Human activity recognition from accelerometer data using a wearable device, *Iberian Conference on Pattern Recognition and Image Analysis*, 2011.
49. K. Altun, et al. Comparative study on classifying human activities with miniature inertial and magnetic sensors, *Pattern Recognition*, 2010, 43(10): 3605-3620.
50. M. TA, D. ACM, et al. The Five Times Sit-to-Stand Test: safety and reliability with older intensive care unit patients at discharge., *Rev Bras Ter Intensiva.*, 2019, 31(1): 27-33.