Master's thesis

Two years

Datateknik
Computer Science

Applying Data Mining Techniques on Continuous Sensed Data
For daily living activity recognition

Yunjie Li

Mittuniversitetet
MID SWEDEN UNIVERSITY

Campus Härnösand Universitetsbacken 1, SE-871 88. Campus Sundsvall Holmgatan 10, SE-851 70 Sundsvall.
Campus Östersund Kunskapens väg 8, SE-831 25 Östersund.
Phone: +46 (0)771 97 50 00, Fax: +46 (0)771 97 50 01.

Abstract

Nowadays, with the rapid development of the Internet of Things, the application field of wearable sensors has been continuously expanded and extended, especially in the areas of remote electronic medical treatment, smart homes ect. Human daily activities recognition based on the sensing data is one of the challenges. With a variety of data mining techniques, the activities can be automatically recognized. But due to the diversity and the complexity of the sensor data, not every kind of data mining technique can performed very easily, until after a systematic analysis and improvement. In this thesis, several data mining techniques were involved in the analysis of a continuous sensing dataset in order to achieve the objective of human daily activities recognition. This work studied several data mining techniques and focuses on three of them; Decision Tree, Naive Bayes and neural network, analyzed and compared these techniques according to the classification results. The paper also proposed some improvements to the data mining techniques according to the specific dataset. The comparison of the three classification results showed that each classifier has its own limitations and advantages. The proposed idea of combing the Decision Tree model with the neural network model significantly increased the classification accuracy in this experiment.

Keywords: data mining technique, Decision Tree, Naive Bayes, neural network, activities recognition.

Acknowledgements

First of all, I would like to express my sincere appreciation to my supervisor Prof. Tingting Zhang for her careful guidance and patient supervision during my postgraduate study. Thanks to Tingting for all the suggestions and advice in my research work. The whole research procedure could not have been carried out successfully without her help and encouragement.

Secondly, I want to express my sincere thanks to Stefan Forsström and Victor Kardeby for their helpful guidance and assistance in my study. They had a lot of useful suggestions regarding my research work, and patiently answered my various questions.

Finally, I'd like to thank all those who have supported me and helped me during the research, especially my classmates at Mid Sweden University.

Table of Contents

Abstract.....	i
Acknowledgements	ii
Terminology	v
1 Introduction	1
1.1 Background and problem motivation	1
1.2 Overall aim.....	2
1.3 Concrete and verifiable goals	3
1.4 Scope	3
1.5 Outline of the thesis	3
1.6 Contributions	4
2 Theory	5
2.1 Machine Learning	5
2.2 Decision Tree models.....	8
2.3 Naive Bayes Models.....	8
2.4 Neural Networks.....	10
2.5 Hidden Markov models	11
2.6 Other data mining techniques	12
2.7 Principal Component Analysis.....	14
2.8 The Least Squares method	14
3 Methodology.....	16
4 Design and implementation	19
4.1 The input dataset.....	20
4.2 Decision Tree.....	23
4.2.1 Data analysis	23
4.2.2 Data preprocess	26
4.2.3 Linear fitting	27
4.3 Naive Bayes.....	28
4.3.1 Data analysis	28
4.3.2 Data preprocess	28
4.4 Neural network	30
4.4.1 Data analysis	30
4.4.2 Data preprocess	31
4.4.3 The transfer function of neurons.....	32
4.4.4 The structural design of the neural network	33
4.4.5 Other parameters of the neural network.....	34
5 Result	37
5.1 Decision Tree.....	37
5.1.1 Result and analysis.....	37
5.1.2 Advantages and disadvantages.....	38
5.2 Naive Bayes.....	39
5.2.1 Result and analysis.....	39
5.2.2 Advantage and disadvantages	42

5.3	Neural network	43
5.3.1	Classification of the locomotion part.....	43
5.3.2	Classification of the gesture part.....	46
5.3.3	Result and analysis.....	48
5.3.4	Advantage and disadvantages	49
5.4	Summary	50
6	Conclusion.....	52
6.1	Ethical considerations.....	53
6.2	Future work.....	53
	References	55
	Appendix A: The detailed classification result of gesture in Naive Bayes model	59

Terminology

AR	Auto Regressive
ANN	Artificial Neural Network
DT	Decision Tree
HMM	Hidden Markov Model
IOT	Internet of Things
NB	Naive Bayes
PCA	Principal Component Analysis
SVM	Support Vector Machine

1 Introduction

In recent years, along with the rapid development of mobile computing, pervasive computing and other new technologies, the new science and technology makes people's lives more convenient, faster, and also contributes to the rapid development of human-computer interaction technology. The human-computer interaction technology becomes an important part of our daily life. The development of wearable sensors for human-computer interaction has played an indelible role in promoting the realization of human-computer interaction, and provides unlimited possibilities for the development of the Internet of Things [1].

Research based on the wearable sensors for detecting and recognizing the human daily activities attracted a lot of attention from a large number of researchers in recent years. As one of the latest research in the field of motion recognition technology, wearable sensor data based action recognition research is of importance. Due to the characteristics of wearable sensors, such as lightweight, efficient, intuitive, less affected by the external environment, its application field has been expanding and extending. Recently, the human motion recognition technology based on sensors has been widely used in smart homes, handheld mobile devices, remote electronic medical treatment and many other areas. The electronic medical field especially has had a rapid development. By analyzing the collected sensing data of the sensors, which can be worn on the body of the subject, the daily activities of people such as the physical condition, diet, etc., It can be especially useful in the field of detection of falling accidents among the elderly people living alone, daily exercise guidance, dietary guidance and so on.

1.1 Background and problem motivation

With the rapid development of information technology, traditional health care systems are entering a more digital and electronic stage. Our research team has focused on an E-health system for several years; the system is aimed at building a complete electronic health system for people of all ages. In our sensor based E-health system, some data mining analysis for the continuous sensed data is required, in order to achieve the objective of daily activities detection, daily care of patients and fall detection of elderly.

In the current study stage, a variety of data mining techniques can be applied to the study of activities recognition based on the sensor data. However, due to the diversity and complexity of the sensor data, not every kind of data mining techniques can be perfectly applied to the sensor data. In practice, a systematic data analysis and processing is usually required, in order to achieve the desirable recognition goal of the specific data mining methods. Therefore, the premise of activities recognition based on sensor data is to make the right choices and judgments of data mining methods. This requires the researchers to have a good understanding of a variety of data mining techniques, to be able to analyze and improve them depending on the situation in order to obtain satisfactory results.

Based on the E-health system, we can carry out a series of developments and research, including the maintenance and upgrade of the platform, such as the maintenance of the security system, the data query of a distributed system, and processing of the sensing data collected from the sensors which are connected to. The content of this paper is proposed based on the requirements of sensor data processing and analysis of the E-health platform. Continuous sensing data of the sensors distribution in various parts of the human body, requires the processing and analysis of data mining techniques in order to achieve the goal of human daily activities recognition.

1.2 Overall aim

The data mining techniques can be used in different application areas according to their own features, and the selection of the data mining technique is closely related to the classification accuracy. The overall aim of this study is to search for several high-efficient, high-precision data mining techniques which can be applied to large-scale continuous sensed data, design and implement the classifier model for each data mining technique, analyze and compare all of the classifiers according to the classification results, and give some suggestions for improvement of these data mining methods, so that can be applied to the continuous sensed data to achieve the goal of daily human activity recognition. Therefore, the problem I will solve in this thesis is how to analyze and apply several data mining techniques to the continuous accelerometer and inertial measurement data to achieve the goal of daily living activities recognition.

1.3 Concrete and verifiable goals

The concrete and verifiable goals of this thesis includes the followings:

- 1) Based on the analysis of the sensor data, research several kinds of data mining technology, and pick out the ones which can be applied to this massive continuous sensor data for classification and recognition of human daily activities.
- 2) Depending on the differences of the data mining methods, use the appropriate method for data pre-processing, simplify the original dataset, flexibly select different attributes for different data mining methods in order to obtain a higher classification accuracy.
- 3) Summarize and analyze the classification results of various data mining methods, evaluate each approach, pointing out their advantages and disadvantages under these particular experimental conditions and classification goals.
- 4) Propose some improvements for the data mining method, trying to combine different data mining methods to obtain a more desirable classification accuracy, and finally propose a feasible and efficient solution for activities recognition which can achieve a satisfactory classification result.

1.4 Scope

The study focuses on applying several data mining methods, the Decision Tree method, Naive Bayes method and neural network methods, to a large-scale continuous dataset aiming for human daily activities recognition. I will also focus on the analysis of the classification result and the comparison of each method. The study does not focus on the data mining techniques based on the purpose of clustering, frequent pattern mining, or sequential pattern mining, such as K-mean clustering algorithm, association rule mining technique, and Apriori-like algorithm. [2]

1.5 Outline of the thesis

The overall structure of the thesis is as follows. Chapter 1 describes the overall content of the thesis, including the background and motivation, the significance of the overall goal of this study and the main object of study, as well as the

arrangements of the thesis structure. Chapter 2 describes some works and achievements related to this research from other scholars. Chapter 3 proposes the innovation and challenges of this study on the basis of the last chapter. It briefly discusses the various data mining algorithms, and carries out a detailed feasibility analysis of them. Chapter 4 describes the specific design and implementation of a variety of data mining algorithms, and summarizes the regularities of the dataset during the process of experiment. Chapter 5 lists the classification results of various data mining methods, as well as the evaluation of each method, pointing out advantages and disadvantages of each method by analyzing the results. Chapter 6 is a summary of the entire research work, including the direction of future research, the impact generated by the study, as well as directions for improvement.

1.6 Contributions

During the whole study, one of my contributions lies on the design of how to apply each data mining method to the dataset, and I am also responsible for the implementation of the Decision Tree model and the neural network model. Another contribution of mine is that I made a conclusion and comparison based on the classification result of each data mining technique. Junting Wang, a Bachelor student, contributed on the implementation of the Naive Bayes model.

2 Theory

In this chapter, some related research will be listed, and some data mining techniques will be introduced, as well as some related data preprocess techniques.

2.1 Machine Learning

Many other researchers have explored activity recognition using different Machine Learning [3] methods, using the accelerometers and the inertial measurement data. These methods, include the Decision Tree model, Naive Bayes model, neural network (NN), hidden Markov model (HMM), and support vector machine (SVM).

Yu Jin Hong et al. [4] focus on recognizing the Activity of Daily Living. The paper proposed a novel method to recognize a user's activities of daily living with accelerometers and a RFID sensor. In their research, two wireless accelerometers are used for the classification of 5 human body states using Decision Tree, and more information about the gestures can be offered by detection of RFID. In this paper, they used Bluetooth based wireless triaxial accelerometers and iGrabber which is a glove type RFID reader. Their experiments show that the Decision Tree method can be applicable to continuous sensor data in this environment.

Gamze Uslu et al. [5] focus on the activity monitoring of elderly people, especially on falling detection in order to provide emergency assistance. The paper presents a composite action description and detection model for activity monitoring. The model accomplishes real time continuous monitoring of composite actions by detecting the transitions from one simple action to another and determining the types of actions. It uses Naive Bayes classifier for the classification of simple actions: walk, sit, stand and lie. The fall tends to happen when the elderly people walk, sit and lie, so they mainly focus on the recognition of these activities. The limitation of this study is that it only concerns simple activity recognition.

Lei Gao et al. [6] proposes a system for activity recognition using multi-sensor fusion and the Naive Bayes model. In this system, four sensors are attached to the waist, chest, thigh, and side of the body. The paper gives some comparisons of the recognition results using different numbers of sensors and they also compared the differences of parameters definition and the choice of coordinate system. The Naive

Bayes classifier using multi-sensor fusion is adopted in this paper. However, they only focus on four kinds of coarse classifications of the body's motions.

Annapurna Sharma et al. [7] uses the neural network to classify human activities. The paper presents the designing of a neural network for the classification of human activity. A tri-axial accelerometer sensor, located in a chest worn sensor unit, has been used for capturing the acceleration of the associated movements. All three axis acceleration data were collected, processed and classified using MATLAB. They discuss the choice of some parameters of the neural network. The work shows a detailed description of the designing steps for the classification of human body acceleration data. They designed a 4-layer back propagation neural network, with Levenberg-marquardt algorithm for training. The main problem of this work is that the multiple layers network is usually not very stable.

Rubaiyeat H.A. et al. [8] build an auto-regressive model for the real-time recognition of daily human activities using a single tri-axial accelerometer. They compared the recognition accuracy in the online condition and the offline condition. Extra information such as number of steps, energy expenditure, activity duration, etc. were also calculated in their study.

There are also some other researchers combining two or more data mining methods to classify certain activities of daily living. There are also researchers comparing several different data mining methods, from the classification result analysis, they give their conclusion of each methods.

In the works of Liu Rong and Liu Ming, the recognition of daily human activity is based on the collect multi-sensor data; this paper applies an information fusion algorithm based Naive Bayes to obtain higher-level contexts from a small number of sensors. Firstly, the obtained sensing data is classified by the C4.5 Decision Tree algorithm, then, in the sensor fusion part, the corresponding posterior probability will be calculated at the classifier level. The final classification result is based on the highest posterior probability. The classification result indicates that the classifier fusion strategy based on Naive Bayes technique can achieve a higher correct classification. Their work offers a feasible idea of increase the classification accuracy by efficiently fusing the accelerometer data. [9]

The works of Xi Long et al. compares two data mining methods of the Decision Tree model and Naive Bayes model. In their study, a single tri-axial accelerometer is placed at the waist. The accelerometer sensing data was recorded from 24 subjects

performing daily real-life activities in a natural environment without the researchers' intervention. Their purpose was to recognize the target activities such as walking, running, cycling, driving, and sports. This study compared a Bayesian classification with that of a Decision Tree based approach. Principal components analysis was applied to remove the correlation among features and to reduce the feature vector dimension. Their conclusion was that the Bayes classifier is more extensible on the aspect of classifier update when compared with the Decision Tree model. [10]

Si-Jung Ryu and Jong-Hwan Kim constructed a classification model for hand gesture recognition using several 3-axis accelerometers by the hidden Markov model. The two tri-axis accelerometers are attached to both forearms, the sensing data are collected, transferred and processed by the HMM model. At the first stage, they will classify some short-term motions, and then classify the long-term motions, which is a sequence of the short-term motions. They also compared their work with several other algorithms, such as artificial neural networks, support vector machine, k-nearest neighbor algorithm and k-means clustering. However, with the accelerometers on the forearms, the recognition is limited to the gesture classification. [11]

Researchers from College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand and Faculty of Computing, Engineering and Technology, Staffordshire University, Stafford, UK use a single accelerometer on the wrist for automatic activity classification. Two data mining methods of Decision Tree C4.5 and Artificial Neural Network is performed in their study. Their study focuses on achieving their target by analyzing the mean, minimum energy and sample data differences. They realized the recognition of the locomotion states such as sitting, standing, lying, walking and running by a single sensor on the wrist; it should work better if sensors were added on the lower limbs. [12]

Researchers from Universidad Panamericana, Aguascalientes, Mexico realized the classification of hand gestures by hidden Markov model and Artificial Neural Networks using accelerometer data. They proposed a classification system by analyzing the Fourier spectrum of different circular motion paths in the Euclidean space. Their analysis is intuitional by drawing some figures. After the data is processed by the Fast Fourier Transform, it will be transferred to the HMM and ANN classification system. The novelty of this study is the figure analysis of classification. [13]

Researchers from the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China set up an classification system combining the auto-regressive (AR) model with SVM. The tri-axial accelerometer data is processed by the AR model instead of calculating the traditional frequently used time domains features such as mean, standard deviation, energy and correlation of acceleration data, and the AR coefficients as the input features of the SVM. [14]

2.2 Decision Tree models

Decision Tree [15] is a classification model used for prediction or analysis. The path from the root node to a leaf node gives the attribute values of an instance. The tree is either a binary tree or a multi-tree. Each leaf node leads to a kind of classification. Decision Tree learning is a machine learning technique for analyzing the data, and producing a Decision Tree from the results of the analysis. Each node of the tree represents an object, and each of its branches indicates the classification of certain attribute, this means, that it is determined by the value of an attribute about which node to move towards in the next step, from current node to its child nodes. Decision Tree is a top-down architecture. The path from root to leaf node is more of a set of rules; using the rules, we can find the result of classification on the leaf node.

Decision Tree has been proved to perform very well in the field of medical care; for classifying a patient's disease or analyzing the disease. It is also used in the industry to detect machine failures, etc.

2.3 Naive Bayes Models

Among the numerous classification models, the two most widely used classification models are the Decision Tree model and the Naive Bayes model [16]. Compared to the Decision Tree model, Naive Bayes classifier originated in classical mathematical theory, and has a solid mathematical foundation and the stability of the classification efficiency. Meanwhile, the Naive Bayes model requires few parameters, it is also less sensitive to missing data as the algorithm is relatively simple. Theoretically, comparing Naive Bayes model with other classification methods, the Naive Bayes model usually has the lowest error rate. But in reality, that is not always the case, this is mainly because of the simple but not realistic assumptions which the Naive Bayes model is based on. The assumption is that the presence or absence of a particular

feature is unrelated to the presence or absence of any other feature. Although the attributes are not independent in this dataset, recent studies of supervised learning indicates that even if the necessary conditions for independence is not satisfied, the Naive Bayes model can still get a desired result of classification which is as good as the Decision Tree model performs in some areas. In many practical applications, parameter estimation for the Naive Bayes model uses the method of maximum likelihood; in other words, one can work with the Naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

The Naive Bayes model provides a method for probabilistic reasoning, it considers the data to be classified to follow a certain probability distribution, and according to these probabilities and the known data to infer, in order to make optimal classification decisions. The main features of the Bayesian methods are:

- 1) Incremental learning: the final probabilities are decided by the prior knowledge together with the known dataset.
- 2) Prediction uncertainty: for a new instance to be classified, numbers of possible probabilities will be calculated and a classification decision will be made based on the results of the probabilities.

Now, I will give a brief introduction of the Bayesian formula. Given the dataset D , the Bayesian formula provides a method to calculate its most likely hypothesis in space H . Using Bayes' theorem, this can be written as:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Here, $P(h)$ stands for the initial probability of h , it reflects the Initial probability that one hypothesis is correct. It does not matter if the the priori probabilities are not known, it can simply be set the same for each candidate hypothesis, and it will be amended during the training session. $P(D)$ stands for the probability of dataset D . $P(D|h)$ stands for the probability of D while the hypothesis h is correct. $P(h|D)$ stands for the probability of h being correct while specifies dataset D .

From the Bayesian formula, it is easy to know that $P(h|D)$ increases with $P(D|h)$ and $P(h)$, while it decreases with $P(D)$. This is mainly because if the higher probability of D independent of h , the lower probability that D supports h .

2.4 Neural Networks

The artificial neural network [17][18] is a kind of mathematical model or computational model, its structure and function are imitations of the biological neural networks. Usually, the multilayer neural network consists of three layers: the input layer, the hidden layer and the output layer. The general view of a neural network is shown in Figure 2.1. The neural network is commonly used to model complex relationships between inputs and outputs, or to explore the patterns of data.

Figure 2.1 The architecture of neural network

Artificial neural networks, to some extent, can be seen as a simplification, an abstraction and a simulation of the human brain neural system. It attempts to establish a mathematical model to reflect how our human brain neural system processes the information. A neural network model including a number of neurons, that are connected with each other. Each neuron represents a particular output function. This output function is also called activation function. The weight of connection between two neurons is adjustable, this weight reflects the importance of the input information's effect. It is used to record the whole process of learning in the neural networks. The output of the neural network is related how the neurons are connected, and it also changed by the different weight and different activation functions.

Artificial neural network is mainly used in the field of classification or regression. The trained network is a kind of algorithm or function approximation. At the same time, it expresses the logic strategy of the network to find the best approximation model from the input to the output.

There are a lot of neural network models, and the BP neural network is the most widely used. BP neural network [19] is short for error back propagation neural network. It proposed the solution of the weight learning problem in the hidden layer. The BP neural network is a kind of neural network which can be adjusted to establish a model to fit the specified input and output. If the output is not the same as the expected output, the weight will be revised until the expected output is obtained. In general, the error back propagation and signal forward propagation forms the BP neural network learning process. The process of signal forward propagation is as follows: the input layer receives input data samples, and transmit it to the hidden layer. The processed signal will be transmitted to the output layer. Error back propagation occurs when the resulting output is inconsistent with the expected output. The process of error back propagation is: the output error will be back propagated to the input layer, and the output error will be apportioned to each layer. The weight is also adjusted by these output errors. The training process is actually the process of constantly adjusting the weights. The training process will not be over until the scheduled training times are obtained or the accuracy requirements of the error are met.

In the learning process of the neural networks, as long as we input enough data samples, the BP neural networks can create a nonlinear mapping between n-dimensions input and m-dimensions output through the neural network learning. Another advantage of the BP neural network is the strong fault tolerance. The huge data sample contains the correct regularity; even if there is a small amount of error in the input data samples, they will have little effect on the result of the neural network. Therefore, the BP neural network is widely used in the technical field.

2.5 Hidden Markov models

A hidden Markov model (HMM) [20] is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. The HMM is a sequence of two random variables, one is a changeable and unobservable state sequence, and the other is a sequence of observable symbols generated by the unobservable state sequence. Each state has a probability

distribution over the possible output tokens. Hidden Markov models are especially known for their application in temporal pattern recognition, such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

There are two assumptions of the HMM:

- 1) The probability of the state transition, when the current state transit to the next state, only related to the current state, regardless of the previous history.
- 2) The probability of outputting the observed value at the current time, only depends on the current state, regardless of the previous states.

Based on these two assumptions, the hidden Markov model is usually used for prediction because of its independence of the previous states.

There are three types of problems that can be solved by HMM, their detailed descriptions and the solutions are listed as follows:

- 1) Given the observation sequence $O = O_1, O_2, \dots, O_T$, and parameters $\lambda = (A, B, \pi)$ of the model, how to efficiently compute the posterior probability of the observed sequence $P(O|\lambda)$; The Forward algorithm is suitable for solving this kind of problem.
- 2) Given the observation sequence $O = O_1, O_2, \dots, O_T$, and parameters $\lambda = (A, B, \pi)$, how to select an optimal state sequence $Q = q_1, q_2, \dots, q_T$; In this case, the Viterbi algorithm is applicable.
- 3) How to adjust the parameters $\lambda = (A, B, \pi)$, so that the posterior probability $P(O|\lambda)$ is maximized. This kind of problem can be settled by the Baum-Welch algorithm or Reversed Viterbi algorithm.

The first step of the HMM is to analyze and determine what is the overall aim of the problem, and find a suitable solution according to the analysis.

2.6 Other data mining techniques

In this research field of daily activities recognition on the continuous sensed data, some other data mining techniques can also be applied.

The supervised learning models of Support Vector Machine (SVM) [21] analyze the data and recognize patterns, used for classification and regression analysis. It might be difficult to classify all the data in a low dimension, but with the SVM, it tries to classify the data by mapping it in a high-dimension with a kernel function. The most important factor of the SVM model is the selection of kernel function. Commonly used kernel functions include linear kernel, Gaussian kernel, polynomial kernel and string kernel. Assume there is a dataset which has attribute set A , and the amount of data is N . The kernel function selection can be based on Table 2.2. If there are negative attribute values, the polynomial kernel should not be used, which is why the polynomial kernel method is not commonly used.

Kernel function	$N < 100$	$N > 1000$	$N > 107$
$A > 10$	linear kernel	-	-
$A < 10$	-	Gaussian kernel	linear kernel

Table 2.2. Selection guide of kernel function in the SVM

The k-means the algorithm [22] is based on the theory of clustering. In the clustering method, the similarities are usually judged by the mean value of the instances in the clusters. The processing of k-means the algorithm is to divide the n of objects into k partitions, according to their properties. In the algorithm, the value of parameter k is artificially selected according to the experiment experience. When the number of clusters k is determined, k instances will act as the cluster centers. The rest of the data will be assigned to the nearest cluster center by calculating the Euler distance. The process is stopped when there is no change in the k-means. The k-means algorithm does not guarantee to get a global optimal solution, because the final solution depends largely on the quality of the initialization clusters. As the value of k is a tentative value, the improper k has a relatively large impact on the clustering system. This is the drawback of the k-means algorithm.

When the main purpose is to find frequent patterns from sensing data in wireless sensor network, the Apriori algorithm [23] may be a good choice. Apriori is a typical association rule based algorithm. An association rule mining algorithm is one kind of frequent pattern mining technique. The two main steps of association classification are: to find the association rules and to build up a classifier based on the rules. The name of this algorithm is Apriori, this is mainly because it is based on prior knowledge. A valid rule means that the confidence and support of the rule meet a certain threshold range. In the practice, the Apriori generates the rules based on the frequent item sets, which enhanced the efficiency. The disadvantage of Apriori is that

a lot of frequent item sets will be generated, thus leading to a growth of the space complexity.

2.7 Principal Component Analysis

Principal component analysis (PCA) [24] is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Principal component analysis is often used to reduce the dimension of the dataset, while maintaining the characteristics which contributed the most to the variance of the dataset. Principal component analysis can effectively reduce the dimensions of the original large-scale dataset, by keeping the lower level principal components, ignoring higher level principal components, and these lower level components usually retain the most important aspects of the original dataset.

Principal component analysis is achieved by extracting some bigger eigenvalues and their corresponding eigenvectors. This depends on the mathematical sense of eigenvalues and eigenvectors. In simple terms, the PCA algorithm tries to find the dimensions which have relatively large variances on the whole dimensions, while ignoring relatively average dimensions. For example, if matrix X , which all the k -th elements of its eigenvectors, has constant value a , then this column of data can be ignored, because it is useless to distinguish it from the others. On the contrary, we are looking for the columns which have a relatively wide range of values and to extract these values.

2.8 The Least Squares method

Least squares method [25] is a mathematical optimization technique. It minimizes the sum of the squares of the differences between the observed values and the estimated values, and then finds the best matching functions.

Considering the function model:

$$Q = \sum_{i=0}^{N-1} (aX_i + b - Y_i)^2$$

If a and b can fit the minimum value of Q , then it is the result we required. In order to get the maximum and minimum values of Q , we just need to find the solution of the following two equations:

$$\begin{aligned}\frac{\partial Q}{\partial a} &= 0 \\ \frac{\partial Q}{\partial b} &= 0\end{aligned}$$

If the solution exists, it means that the maximum or minimum value exists. But is it a maximum value, or a minimum value? Here, the minimum value means that the target straight line is the closest to the experimental spots. The maximum value on the other hand means it is the value farthest away from the experimental spots.. Since the second case does not exist, if the solution satisfies the above conditions, then it must be the optimal straight line.

3 Methodology

The main task of this study is to filter out data mining methods and algorithms suitable for human daily activity recognition. As for the complexity and particularity of the sensor data, it still requires some improvement for the corresponding data mining method, so that it can find a complete match and ultimately obtain the desired classification results. Studying the related research, they focus on either the locomotion activities recognition or the gesture activities recognition. But in this thesis, both of them were taken into consideration, which made it more challenging. The challenges in this thesis also lies on the selection of the data mining techniques, since there are so many kinds of data mining techniques. Another challenge lies in how to efficiently deal with large-scale data.

- Analysis: For the selection of the dataset, the data of the tri-axial accelerometer sensor and inertial measurement sensor provides a large amount of useful information about how people moved. The sensed data were collected and recorded periodically, this makes it possible to carry out the activities recognition with this sensed dataset. The large-scale dataset increased the complexity of the activities recognition; multidimensional scaling (MDS), Linear Discriminant Analysis (or Fisher's linear discriminant) (LDA), Isomap, Principal Component Analysis (PCA) are common methods for dimensionality reduction in high-dimensional data. The PCA algorithm, with the superiority of selecting arbitrary number of key attributes, was my first choice algorithm.
- Design demonstration: According to the characteristics of each data mining technique, the classifiers were designed. The Decision Tree method is a kind of classification depending on the attributes' value, because for the sensors distributed on various parts of the body makes it possible for the sensing value be distinguished by their different ranges by this methods. The Naive Bayes method and the neural network method based are rather on mathematical theory. The Naive Bayes method calculates the probability of the sample belonging to each classification, and it can be used to select the key attributes. The neural network is actually a kind of regression of the data sample. These two techniques can be implemented with the help of programming and mathematical tools. Unlike the other methods, the hidden Markov model is aimed at the prediction of data; it can be used to predict the next possible action of a person. Because of the specificity of the HMM, the prediction result varies from person to person. There might be a learning section before the HMM can be

used for activities recognition in practice. However, the HMM classifier was not a priority in this study.

- Experiment: The classification accuracies of all the activities were manipulated and the results were analyzed. For each classifier, the classification result of locomotion activities and the gesture activities were ensured and obtained for the following analysis.
- Evaluation: Research to determine whether all the classifiers followed the prescribed procedures and achieved the stated outcomes. According to the result analysis, evaluations of each classifier were proposed. The focus was to find the correlation regularities of the result.
- Comparison: Three kinds of data mining techniques were studied to determine their similarities and differences. This comparison focused on the analysis of the advantages and disadvantages of each classifier, and possible improvements for future work.

According to the concrete and verifiable goals, the research was carried out step by step.

Achieve the goal selection of data mining technique: in order to select the appropriate data mining technique, it is necessary to be familiar with a variety of data mining algorithms and to understand their application field, and search for help by reading related books and papers. Focus was on the study of the general rules of attributes selection in the classification algorithms model, their application areas and study of the effect on the classification result of parameter selection in the experiment.

In order to achieve the goal analyzing and preprocessing the dataset, an understanding of the relevant data processing methods which can be applied to the study is required, including the selection of data dimensionality reduction method, selection of the data linear fitting algorithm, as well as graphical analysis method, providing a fast and accurate approach for the study. Due to the complexity of the original dataset, the problem of multiple dimensions can be solved by applying the Principal Component Analysis algorithm. The Least squares method can be used for the linear fitting in this dataset.

To achieve the goal evaluating each classifier model, constantly discovering the regularity of experimental datasets is required, as well as constantly proposing improvement and adjustment of the classification model. The final classification results were listed in the result chapter, based on the comparison of these results, advantages and the disadvantages of each data mining technique were discussed.

The goal summarize the whole work achieved based on this research work. The suggestions for improvements were proposed according to the classification result. The summary combined the theoretical knowledge with the practical results and pointed out the possible improvement.

4 Design and implementation

The whole procedure of this thesis includes the following parts: The input dataset is the same for the three classification models. According to the different data mining techniques, the selection of the attributes is flexible and changed based on the analysis of each technique. After having decided on attributes, it may need to go through the procedure of data preprocessing and the linear fitting depending on the data mining technique. Finally, the processed data will be transferred to the classification model, then the result analysis procedure will be carried out based on the classification result.

The whole procedure of this thesis work can be presented by Figure 4.1:

Figure 4.1: The architecture of this thesis

4.1 The input dataset

During the whole research process, I used the OPPORTUNITY dataset, which is also called Activity and Context Recognition with Opportunistic Sensor Configurations. It is a FET-Open (Future and Emerging Technologies Open Call) project under the Information and Communication Technologies theme of the 7th Framework Program of the European Commission. [26][27]

This database contains natural human activities recorded in a sensor rich environment: a room simulating our daily life environment, including a small kitchen, a table, a chair, some household appliances and two doors, where subjects performed daily activities.

The distribution of the on-body sensors is shown by Figure 4.2 below.

The subject wore a custom-made motion jacket composed of 5 commercial RS485-networked XSense inertial measurement units [28]. In addition, 12 Bluetooth accelerometer sensors [29] on the limbs and commercial InertiaCube3 inertial sensors located on each foot.

Figure 4.2: The distribution of the on-body sensors [30]

The format of the dataset is shown by Table 4.3 and Table 4.4. There are 116 columns of data in total, and they can be divided into four parts: The first part indicates time, the sensing data is collected every 33 milliseconds; The second part indicates the values of accelerometers, including 36 columns of data, there are 12 accelerometer sensors on the body, mainly distributed on the limbs; The third part indicated of Inertial Measurement Unit values, including 77 columns of data, there are 7 set of inertial measurement sensors; The last part of the dataset is the two-columns label of the human actions, the labels marked the actions as one locomotion state or one gestures.

Name	Time	Accelerometer	Inertial Measurement Unit	Labels		Total Number
				Locomotion	Gestures	
Columns	1	36	77	1	1	116

Table 4.3: The numbers of the column

Column	1	2	3	4	...	37
Name	Time MILLISEC	Acc. RKN^ accX	Acc. RKN^ accY	Acc. RKN^ accZ	...	Acc. RH accZ
Column	38	39	40	...	113	114
Name	Inertial sensor BACK accX	Inertial sensor BACK accY	Inertial sensor BACK accZ	...	Locomotion	Gesture

Table 4.4: The details of each column

Each accelerometer has three values in the three directions on the x, y and z axis. But the inertial measurement sensors are more complicated, which not only include the inertial values on the three directions but also includes the gyroscope values and the magnetic values. The inertial measurement sensors on the feet also include compass information.

The locomotion has 4 different states, these are stand, walk, sit and lie respectively. There are 17 different gestures, which are open or close door1, door2, fridge, dishwasher, drawer1, drawer2, drawer3, clean table, drink coffee, and toggle switch respectively. In fact, these 17 gestures can be summarized in 4 different categories: the open gesture, the close gesture, clean table and drink cup, because in practice, the

gestures open door, drawer, fridge, dishwasher or drawers, have a lot in common, and we can easily distinguish these objects by some other method, for example, using the Radio Frequency Identification. As for the close gestures, there is no need to classify them separately.

The main part of the data pre-processing includes the following two aspects:

1) Alternate missing values of the attributes:

There is a small number of unknown values in the attribute values, they are marked as Not a Number (NaN), which should be replaced to make sure that they will not produce unwanted effects to the classification results.

2) Remove the useless attributes:

In this Opportunity Dataset, I noticed that some values of attributes cannot be acquired. The NaN data has no contribution to the construction of the classifiers. At the same time, these data will lead to memory overflow during the classification process. So the useless attributes must be removed. In this Opportunity Dataset, these useless attributes are: Accelerometer RWR accX, Accelerometer RWR accY, Accelerometer RWR accZ, Accelerometer RH accX, Accelerometer RH accY, Accelerometer RH accZ.

Through the above analysis about the principal component analysis, I found that it is suitable for the Opportunity dataset which I have adopted for the reasons listed below:

1) The Opportunity dataset contains large amounts of data, making up a 116-column and more than 50,000 row dataset. Such a large amount of data causes issues in the processing of data, e.g. excessive calculation, result out of range, some unpredictable miscalculation. Applying the PCA algorithm to this dataset, I can reduce the dimension of the original dataset, and the benefit is obviously, that it becomes much easier and more feasible to handle the data.

2) Another significant benefit of applying the PCA algorithm to this situation is that I can easily find which column of data is the most important component, because the new lower-dimension dataset is constructed by mapping the original high-dimensional vector onto new lower-dimensional vector, not simply by cutting

off some vectors. This will guarantee the accuracy of the original data is not lost while reducing the dimension.

3) The data of the Opportunity dataset is collected by recording the sensors' value, as a result, in the process of signal transmission, the channel is not ideal. The signal received will inevitable be influenced by some noise. So in this case, how to effectively remove the effect of noise in the original data, is also an aspect which I should take into consideration. One more benefit of the PCA is that it is possible to reduce the noise and redundancy, while also reducing the possibility of overfitting.

4.2 Decision Tree

In this section, the detailed design and implementation of the Decision Tree model will be proposed, including the analysis of the dataset, and description of how the related data preprocess algorithms will be applied to the dataset.

4.2.1 Data analysis

In order to find out the regularity of attributes in each human action. I started by drew figures for each attribute of each action, some of the results can be seen in Figure 4.5 ~ 4.8. From these figures we can find some regularities, and we can also find out what the relationship is between all the attributes.

After I studied all sensing data, I picked out four important parts of the attributes; they are sensor RKN[^], HIP, LUA_ and RUA_. Here, each figure contains 500 rows of data of six different kinds of human actions. The actions are: the locomotion states stand, sit, walk, and the gestures clean table, drink cup, open and close. These curves reflected the differences between different actions in the range of its values and the trend of the curves. This makes it possible to classify all of these actions, if adding the linear fitting part to this data, by calculating the slope and intercept.

Figure 4.5: The figure of accelerometer RKN^

From the data of the sensor on the subject's right knee, it can be seen that when the subject walks, the data will exhibit relatively large fluctuations, while the other actions shows a flat trend. But they can still be distinguished by the range of their values; when people sit, opens or closes a certain object, the sensing value is usually around 1,000, while the value for other actions is much lower.

Figure 4.6: The figure of accelerometer HIP

There are also some regularities when observing the data of the sensor on the subject's hip. There are some repetitions when the subject is cleaning the table and sitting down. Walking will still cause the fluctuation of the sensing data, while

drinking and opening or closing is relatively flat, but has a significant gap in their values.

Figure 4.7: The figure of accelerometer LUA^

When paying attention to the sensors on the left arm, I find that there are more fluctuations in most actions except the opening and closing gestures; this means that the subject usually uses his left hand to hold the cups while drinking, and the flat trend of opening and closing might mean that he does not usually use his left hand to do this type of gesture. Notice that the repetitions of drinking might show the gesture of putting up and down the cup.

Figure 4.8: The figure of accelerometer RUA_

The sensing values are much more complicated on the right arm, which may make it may be less suitable to be used to classify these actions, but some assumptions can be proved by analyzing the sensor data of the right arm. The dramatic fluctuations while opening and closing the object illustrates that this man indeed uses his right hand to complete these types of gestures.

So from the analysis of these sensors' data, I found some useful information and regularities. In order to classify the locomotion states, I will focus on the sensors bound to the lower body of the subject, as well as the sensors bound to the upper limbs of the subject while classifying the gestures. There is also one problem that might occur in the process, the state of walk usually has a lot of similarities with the state of standing, in fact, standing is one special type of walking; the most obvious difference will be reflected on the sensing data of the feet, which may need more information of the inertial measurement sensor to accurately classify these two states.

4.2.2 Data preprocess

In this case, in order to apply the Decision Tree method to the Opportunity Dataset, the original complicated dataset must be simplified. Since Decision Tree is based on the information gain, it means the first thing I should do is to find the key attributes. The original dataset has 113 attributes in total, but not all of them are necessary, sometimes, too many useless attributes will actually lead to a deterioration of accuracy. In fact, 12 sets of accelerometers distributed on the limbs of the subject, using the sensing data of accelerometers is entirely adequate to classify the target actions theoretically.

According to the idea of data fusion [31], there are three levels of data fusion, data level fusion, feature level fusion and decision level fusion respectively. Intermediate level of fusion is feature level fusion, firstly, the original information gathered from the sensors will extract the feature (here, feature may be the edge of a target, direction, or speed, etc.), and then carry out a comprehensive analysis of the characteristics and processing of the information. Using the PCA algorithm in this dataset is actually a feature level data fusion, and the fusion of the original data is acting on the tri-axial directions. After having adopted the PCA algorithm, the tri-axial data will be mapped into one-axial data. So the result is that I mapped the 36 columns of data into 12 columns; this preprocessing makes the next step much easier and more efficient.

Since each row of data in the Opportunity Dataset is recorded every 33 milliseconds, this being a very short time slot, which means that the whole procedure of one action will take a dozen rows of data. Due to this particularity of the Opportunity Dataset, one way of classifying these actions by Decision Tree is to take several rows data into consideration at one time, otherwise single row data will not make sense in such a short time. So in the implementation, I calculated 33 rows of data at a time, then I was able to figure out the action's classification during the late second.

4.2.3 Linear fitting

According to the analysis above, it is necessary to find the approximate function to present the relationship between an attribute's value and time. In this case, I applied the least squares method.

The result shows that least squares method performs very well in this case, the result can be seen in Figure 4.9:

Figure 4.9: The result of linear fitting

The linear fitting result i got was the following:

1) Result of linear fitting of the blue curve is:

$$y1 = -0.11947 x + -7.68254, r = 0.44758$$

2) Result of linear fitting of the red curve is:

$$y_2 = -0.10845x + -539.55943, r = 0.38920$$

3) Result of linear fitting of the green curve is:

$$y_3 = 0.01493x + 985.25168, r = 0.03995$$

I also drew the three straight target lines in the figure, which are marked by the color black, in this way, I can learn directly about the slope and the intercept. For example, the slope of line y_3 is 0.01493, and the intercept of line y_3 is 985.25168. From the figure, it can be proved that we can get a desired linear fitting result using the least squares method.

4.3 Naive Bayes

In this section, the detailed design and implementation of the Decision Tree model will be proposed. The Naive Bayes model is a simple and efficient classifier model; here, I focus on the data preprocess of this classifier.

4.3.1 Data analysis

In this Naive Bayes classifier, the selection of attributes is not as complicated as the Decision Tree model. I preferred to use the sensor value of 12 accelerometers as the input of the classifier for the following reasons: Firstly, all of the sensor values have the same magnitude, which means the classification result is prevented from being influenced by the different magnitude. Secondly, the distribution of the sensors covered the various parts of the body, all of their values make the classification result more credible.

4.3.2 Data preprocess

In order to reduce the complexity of the original dataset, I also applied the principal component analysis algorithm to each set of accelerometers. The pre-process of data is not the same as in the Decision Tree model. In the Decision Tree model, I made the classification decision by checking the range of the linear fitting results, but in Naive Bayes model, the classification is related to each attribute's value.

In this Opportunity Dataset, all sensing data is collected every 33 milliseconds, so it is continued. Since the attributes are continuous variables, the method applicable for calculating discrete variables cannot be used to calculate the probability in this case. This means I should find one efficient approach to deal with the continuous attributes. Usually, there are two kinds of common solutions:

1) Assuming the probability distribution of variables can be calculated:

Supposing that the training data contains a continuous attribute x , the mean and variance of x for each classification is calculated. In a certain classification c , let the μ_c stands for the mean of x , and σ_c^2 stands for the variance of x . Then, the probability density of a value is presented by $P(x = v|c)$, and it can be calculated by the equation for a Normal distribution parameterized by μ_c and σ_c^2 .

$$P(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

This is the density value used to reflect the relative likelihood of each value. With this probability density function, the training data can be substituted, and the density values for each instance can be calculated.

Another similar possibility is that the training data obeys a certain distribution, the probability distribution is already known or can be calculated.

2) Dividing the attributes' values into several intervals:

The second approach is actually a discretization method, which can split the continuous variables into discrete variables by the range of their values. A condition for using this processing method is that the data sample should be sufficient enough, since dispersing a small number of data samples will result in a substantial loss of classification accuracy.

In general, the probability distribution method is a better choice if the data sample is relatively small, or the data obeys a known probability distribution, and its probability can be calculated by the known distribution function. The discretization method tends to perform better if there is a large amount of training data, because it will learn to fit the distribution of the data.

In this case of the Opportunity Dataset, the discretization method is preferred because of the large amount of continuously changed data samples, which makes it impossible to calculate the probability distribution, and another reason is that more computationally expensive models can generally achieve better accuracy.

In details, segment each attribute into 10 intervals. First, find out the maximum and minimum values for each attribute in the training dataset, then divide it according to the maximum and minimum values. During the training process, get the original value and then map it into the corresponding interval and calculate the probability of each possible classification with the mapped attributes' values.

4.4 Neural network

In this section, the detailed parameter selection of neural network will be discussed, including the selection of transfer function, the structure of the neural network, and some other parameters of the network.

4.4.1 Data analysis

Compared to the previous two data mining methods, the difference of neural network model lies on the choice of training dataset. In the Decision Tree model, too many attributes makes it more difficult to construct the tree classification model. In the Naive Bayes model, the increasing attributes will lead to a heavier calculation. Therefore, it is not appropriate to use the amount of attributes to classify these actions in the two methods. So I used the PCA algorithm to reduce the dimension of the original dataset, merge the data in the three directions of the x , y , z axes into a new attribute. MATLAB is a powerful engineering calculations and mathematical analysis visualization software. The MATLAB platform has a great advantage in terms of matrix operations when compared to the other data processing tools, which makes it possible to use more attributes as input during the process of the neural network classifier. Moreover, MATLAB itself has a functional neural network library, therefore, this neural network experiment is carried out on the MATLAB platform with more detailed attributes information. Later on, I will propose a detailed description of the structural and parametric design, including the specific parameters on the neural network.

4.4.2 Data preprocess

Data normalization is necessary in this classification network. Normalization is a data preprocessing method. Its main idea is to preprocess the data sample (via algorithms) in order to limit the data sample to the range you need. One of the benefits of normalization is providing a convenience for the following step, secondly, it ensures that the program is running to accelerate the convergence.

There are two types of normalization methods; one is mapping the original data sample into the specific interval, usually the (0,1) interval. This is aimed at providing convenience for the data processing. The other is mapping the dimensionless expressions into nondimensional expressions. This is aimed at simplifying the calculation; the nondimensional expressions guarantees convenient computing, and also highlights the essential meaning of physical quantities.

The common solutions for data normalization and its characteristics can be seen as Table 4.10 [32]:

Name	Formula	Usage
Standard score	$\frac{X - \mu}{\sigma}$	To normalize error, when the parameter is known.
Student's t-statistic	$\frac{X - \bar{X}}{s}$	Used to normalize residuals, when the parameter is unknown or used for estimate
Studentized residual	$\frac{\hat{\epsilon}_i}{\hat{\sigma}_i} = \frac{X_i - \hat{\mu}_i}{\hat{\sigma}_i}$	Used to normalize residuals, when the parameter is the estimation value
Standardized moment	$\frac{\mu_k}{\sigma_k}$	For normalizing the center distance, using standard deviation as a measurement of the scale
Coefficient of variation	$\frac{\sigma}{\mu}$	For normalizing dispersion, it is a normalization measurement of the discrete probability distribution.
0-1 Normalization	$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$	Used to map the data sample into the specific interval

Table 4.10: Characteristics of the data normalization methods

The data normalization is essential because there are some singular values in this Opportunity Dataset. The singular value refers to the particularly large or small samples with respect to the other input samples. The existence of singular values will lead to the increasing the training time, or even worse, it may make the network unable to converge.

I used the 0-1 normalization method in this BP neural network. The process of training will be repeated until the network is stable, which is a symbol of the convergence. The number of training times reflect the speed of convergence. Without the data normalization, the training times reached more than one million, and it usually cannot achieve the desired error rate. However, the network with the normalized data sample will rapidly converged, and the error rate is satisfactory after only five thousands times of training. The result shows that the convergence comes faster and the network becomes more stable after the normalization.

4.4.3 The transfer function of neurons

The transfer function is the determinant part of the neural network; the network model will vary with the different transfer functions.

There are four types of transfer functions [33]; the non-linear transfer function, the transfer function with the threshold, the probability transfer function and the linear transfer function.

Sigmoid transfer function is a typical non-linear transfer function. The definition of unipolar and bipolar sigmoid transfer function is:

$$f(x) = \frac{1}{1 + e^{-x}}$$

$$f(x) = \frac{2}{1 + e^{-x}} - 1 = \frac{1 - e^{-x}}{1 + e^{-x}}$$

transfer function with the threshold is also called hard limit transfer function. The unit step function is the commonly used transfer function with the threshold. The definition of the unit step function is:

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Sign function is an example of the bipolar transfer function with the threshold, and the definition of it is:

$$\text{sgn}(x) = \begin{cases} 1, x \geq 0 \\ -1, x < 0 \end{cases}$$

The probability transfer function requires a probability function to describe the states of the neurons.

The linear transfer function is also called purelin transfer function, the definition of it is:

$$f(x) = \begin{cases} 0, x \leq 0 \\ cx, 0 < x \leq x_c \\ 1, x_c < x \end{cases}$$

The choice of transfer function has a significant effect on the final result of the neural network, which depends on the specific conditions. Usually, the output has less limitations when adapting the purelin transfer function. However, if we choose the sigmoid transfer function, the output will be limited in a smaller range of values. In order to achieve the goal of classification in this dataset, I chose the sigmoid transfer function, so that the output will be more accurate.

4.4.4 The structural design of the neural network

In addition to making the choice of transfer function, there are some other parameters that need to be determined in the BP neural network model: the structural design including the decision of the number of the nodes in the input layer and the output layer, the number of hidden layers, and the number of the nodes in the hidden layer.

The number of nodes in the input layer and the output layer is usually decided by the actual demand. The number of nodes in the input layer is the dimensions of the training sample, and the number of nodes in the output layer is $\log_2 m$. Here m is the number of the different classifications in the classification network. In this Opportunity Dataset, I also will divide the classification task into a locomotion recognition part and a gesture recognition part. The training dataset is a 30-dimensions vector. In order to present the three different output states, the output layer should at least have two nodes. Each node has two possible values, 0 or 1, and

the combination of these two values will be used to identify a certain state. In detail, in the locomotion recognition part, 00 stands for state stand, 01 stands for state walk, 10 stands for state sit, and 11 stands for no defined state. And in the gesture recognition part, 00 stands for gestures of opening and closing, 01 stands for the gesture clean table, 10 stands for the gesture drink cup, and 11 stands for no defined gesture.

The number of hidden layers is flexible. Increasing of the number of hidden layers can obviously increase the accuracy and minimize the errors, but it will lead to an increase of complexity in the whole network. Considering the pros and cons, the priority is to tend a hidden layer network, and the theoretical analysis shows that in most instances, a hidden layer is totally meets the accuracy requirement completely.

However, when it comes to the number of the nodes in the hidden layer, its determinants are related to the training data sample, the complexity of regularity in the data sample, and the quantity of the noise in the data sample. The difficulty of determining the number of hidden nodes consists of the regularity and the noise usually being very hard to measure. Even though there is a formula that we can use to roughly estimate the number, in practice, we should try some different values to find the most suitable number of hidden nodes. The common formulas used to estimate the number of hidden nodes can be expressed as:

$$m = n + l + a$$

Here, m stands for the number of hidden nodes, n stands for the number of nodes in the input layer, l stands for the number of nodes in the output layer, and a is a constant range from 1 to 10.

In this case, since the input is a 30 dimensions vector, and the output is a 2 dimensions vector, the number of hidden nodes should be set to the value range 7 to 16.

4.4.5 Other parameters of the neural network

There remains some other issues to be settled, such as the choice of the learning algorithm, the original weight of the network, and the value of learning rate.

The learning algorithm [34] including gradient descent momentum method(`traingdm`), Levenberg - Marquardt method(`trainlm`), conjugate gradient method(`trainscg`), resilient gradient descent method(`trainrp`), and some features of each algorithm can be seen in Table 4.11:

learning algorithm	Appropriate problem	Convergence	Storage space	Other characteristics
<code>traingdm</code>	Approximation Classification	Rapid converges	Medium	Ensure the reliability of the network
<code>trainlm</code>	Approximation	Rapid converges Error is relatively small	Large	The larger the scale is, the weaker the performance
<code>trainrp</code>	Classification	The quickest converges	Small	The smaller the error is, the weaker the performance
<code>trainscg</code>	Approximation Classification	Rapid converges Stable performance	Medium	Appropriate for the large scale network
<code>trainbfg</code>	Approximation	Rapid converges	Relatively large	The larger the scale causes the large amount of calculation
<code>traingdx</code>	Classification	Slower converges	Relatively small	Appropriate for early stop training

Table 4.11: Detailed features of training algorithms

In general, the original weight of the network is randomly assigned by a small value, it typically ranges from -1 to 1. The advantage of doing this is to make sure that each neuron has a smaller input value, resulting in getting the best performance of the activation function. At the same time, the unlimited growth of the weight after the multiple learning and training process is avoided.

The learning rate should be set low enough to guarantee the stability of the system. A higher learning rate will lead to the instability of the system, while a lower

learning rate will take a longer time to complete the training process. In order to balance the two probable effects, the learning rate is usually set the to value range 0.01 to 1.

5 Result

In this chapter, the final classification model will be presented in the corresponding sections. Based on the classification result, comparison and analysis will be discussed, and the advantages and disadvantages of each classifier will finally be listed.

5.1 Decision Tree

In this section, the detailed analysis of the classification result of the Decision Tree model will be discussed, including the result analysis and the advantages and disadvantages of this classifier.

5.1.1 Result and analysis

As long as I get the linear fitting result of a set of pre-processed data, then the necessary parameter is known. The next step is to build up a tree model to classify all of the actions. The classification tree model can be seen in Figure 5.1:

Figure 5.1: The Decision Tree model

The process of the Decision Tree starts with checking the sensing data of sensor RUA_. According to the range of intercept, the classification tree has three branches. The first branch will check the sensing data of sensor HIP, according to the different

intercepts, directed to the two categories stand and walk. The second branch considers the sensing data of sensor RKN[^], according to the different intercepts, directed to the two categories of clean table and open and close. The last branch considers the sensing data of sensor LUA[^], according to the different slopes, directed to the two categories of sit and drink cup. Considering the situations where some data might not belong to any of these six classifications, I also set up an additional state in case this would happens.

In order to test the accuracy of the tree model, I prepared 1,200 test cases in total, and 200 test cases for each classification. The classification result can be seen in Table 5.2:

Activities	stand	sit	walk	drink cup	clean table	open and close
Accuracy /%	72.8	55.6	46.9	34.1	49.2	32.2

Table 5.2: The result of the Decision Tree model

The result shows that the recognition of locomotion actions is better than the recognition of gesture actions. The accuracy of walk is 46.9%, and it is usually mistaken for stand, this is mainly because walking is a special type of standing. Another noteworthy result is that the accuracies of three gestures are less than desirable, all of which are less than 50%. This might be due to the recognition of gesture movements requiring more details. For example, the direction of the movement should be taken into consideration, which hand joints are moving, and we also need to know the habits and regularities of subject's movements. It is difficult to continue the classification of the more detailed ones, because if I continues to build a deeper classification tree model, so that more attributes can be taken into consideration, this will lead to a loss in the classification accuracy. Another difficulty is that increasing the attributes makes it very hard to find the optimal tree model, and it also results in a reduction of the accuracy.

5.1.2 Advantages and disadvantages

The traditional classification tree model is not suitable for this Opportunity Dataset, and the result above is a proof of this. Even so, the Decision Tree model is an important attempt, and the method of combining the classification tree model with

the PCA algorithm and the least square algorithm provides an intuitive guidance for future research of another data mining method.

The advantages of the Decision Tree model are that the process of classification is very simple and intuitive. Each leave node of the tree will lead to a kind of classification, and its traces can easily be tracked. Another advantage is that there are 33 rows of data in each test case as this approach takes into account the continuity of the data sample, and it is a very important factor which should be paid more attention in this kind of continuous data, however, it may cause inaccuracy in the linear fitting result of the data.

The disadvantages of the Decision Tree are also obvious. It is hard to find a way to build the tree model automatically, and artificially constructing a tree model has its own limitations, making it difficult to find the optimal model. In order to obtain detailed classification results, more attributes are required, which makes the complexity of the tree model rapidly increasing. However, in the theoretical analysis, as long as there is a sufficient number of attributes, the classification tree model can be constructed. But, in reality, it is almost impossible to implement because of the limited information offered by this dataset. As the result shows, the accuracy of a three levels tree is not satisfied in terms of the accuracy requirement, and the accuracy will inevitably reduce with the increasing of the depth of the tree, so it is not a good choice to continue revising this classification tree.

In conclusion, we cannot get a desirable result using this method, however, it proposed a feasible direction for the next step, and the main idea of this tree model can be a reference for the other methods. Perhaps I will get a better result if combining the tree model with another method, and this will be a focus for future research.

5.2 Naive Bayes

In this section, the classification result of the Naive Bayes model will be listed, and the advantages and disadvantages of Naive Bayes will also be discussed.

5.2.1 Result and analysis

The process of the Naive Bayes model includes two stages. At the first stage, I try to distinguish the different locomotion states, and at the second stage, to distinguish the

different gestures. Once the data sample is obtained, the program will calculate the probability of this data sample belonging to each classification. The final task is to find the classification with the maximum probability, and this classification result will be recorded as the classification of the data sample.

Table 5.3 shows the detailed classification result of the locomotion part, and Table 5.4 shows the brief result of the gesture part, the detailed result of gesture recognition can be found in Appendix A. Some other information can be found in the two tables, including the number of test cases and the probabilities of each classification.

Locomotion Classify as	Stand	Walk	Sit
Stand	67.73	31.98	0.29
Walk	30.27	66.73	3.00
Sit	8.84	5.28	85.88
The number of test cases	6845	5542	1289
Accuracy /%	67.73	66.73	85.88

Table 5.3: The result of locomotion

Gesture	open door	close door	open fridge	close fridge	open dish washer	close dish washer
Accuracy /%	44.44	56.55	48.86	58.49	39.61	18.32
Gesture	open drawer	close drawer	clean table	drink cup	toggle switch	
Accuracy /%	29.47	45.64	74.89	77.98	48.95	

Table 5.4: The result of gestures

From the result of the data, we can see that the recognition rate of sit is relatively satisfactory, but the accuracies of stand and walk are 67.73% and 66.73% respectively.

This conclusion is similar to the conclusion of the Decision Tree classifier, but in general, the Naive Bayes classifier performs better than the Decision Tree model. The Naive Bayes method assumes that each attribute is independent. When the classifier is constructed, it will ignore the existence of dependencies, this leads to a decline in classification accuracy, while the Decision Tree is based on the information gain, which is associated with the correlation between the attributes. If the attributes are relatively independent, the result of the classification accuracy may decline. However, in this case, the attributes are not independent, the sensors on the upper limb are obviously related to each other, and so are the sensors on the lower limbs, and the sensors on the upper limbs also have some relatively weak connections to the sensors on the lower limbs. Since sitting is a kind of systemic action, while the obvious difference between standing and walking is the lower limb movement, and the sensors on the lower limbs are strongly related, this may be the reason why the recognition of sit is relatively satisfactory when compared it to stand and walk.

The recognition result of the gestures part is much more complicated than the locomotion part because the classification number of locomotion is less than the number of gestures. The result shows the recognition results of clean table(74.89%) and drink cup(77.98%) are more satisfactory than the other gestures. Noteworthy is that the gesture clean table has a 11.74% possibility to be classified as the gesture closing fridge, and drink cup also has a 9.79% possibility of classified as closing fridge, this may be a prompt that the these gestures have some similarities. In order to get a better classifier, more detailed information should be taken into consideration. The distinctions between clean table, drink cup and close fridge mainly lie in the direction of the hand's movements. The sensors will perceive the substantial changes in the horizontal direction when cleaning table and closing the fridge, it will repeatedly change while cleaning table, but the repetition should not occur when closing the fridge, because it is a one-time completed action. The changes will be noticed in the vertical direction when drinking cup. In the theoretical analysis, if I pay more attention to the direction of the hand movements, and in that way get a better classification result. Also noteworthy is that the gesture of opening an object is usually misjudged as the gesture of opening another object or closing the object. For example, the classification result of opening the fridge is 48.86%, but it has a 40.12% possibility of being classified as the gesture of closing the fridge, and the classification result of closing the dish washer is 18.32%, while there is a 27.22% possibility of being classified as the gesture closing drawer, which is even higher than its accuracy. Such a situation illustrates a feasibility that the recognition of the open gestures and the close gestures can be merged at the first stage, and combined

with the RFID and some other details, then realized a further distinction in the second stage. This ideal is already adapted in the process of the Decision Tree method, and the present results also indicates that this is a viable approach.

5.2.2 Advantage and disadvantages

The result of the Naive Bayes method gives me a lot of inspiration as to how to deal with this kind of classification.

The advantages of the Naive Bayes model are reflected in the following aspects. The Naive Bayes model provides a detailed analysis of the classification results, sorts out the most confusing classifications, and the corresponding analysis can be carried out based on these results, and the key attributes can be found, which can be used to construct a more appropriate classification model. The idea of Naive Bayes is very simple, and the process is simple too. The process does not require manual intervention, which belongs to a kind of supervised learning, this makes the results more convincing.

The disadvantages of the Naive Bayes model are reflected in the following aspects. It removes the associations between attributes, mandatorily turns the dependent attributes into independent attributes, which led to reducing the classification accuracy. In the case of a small number of categories, such as the first stage classification of the locomotion part, the classification accuracy is acceptable. However, the second stage classification of the gesture part, with a large increase in the number of categories, its classification accuracy also reduces a lot. In particular, the distinction of open and close gestures would be more difficult, even using the radio frequency identification technology in practical applications can locate different objects, but the Naive Bayes model is not suitable for distinguishing the open and close gestures.

Finally, the conclusion is that after getting a Naive Bayes analysis results, I had a more intuitive and a better understanding of the correlation between these data attributes, and I also learned about how the classification can be misjudged. This attempt provides some valuable information for future work.

5.3 Neural network

In this section, the classification result of the neural network will be listed, as well as the advantages and the disadvantages of neural network will be discussed.

5.3.1 Classification of the locomotion part

The number of hidden nodes in the neural network is set according to the result of the selection criteria formula. Different numbers of hidden nodes will have some impact on the classification accuracy. Specific results are shown in Table 5.5:

Locomotion The number of hidden nodes	sit	walk	stand
10	0.717841727	0.402891028	0.832093123
15	0.807692307	0.485981308	0.842767295
20	0.647482014	0.353920739	0.792934173

Table 5.5: The influence of the different numbers of hidden nodes

As the results show in the above table, when the number of nodes in the hidden layer is 15, the three categories of classification have also achieved maximum accuracy. This is illustrated in this neural network model; the number of hidden nodes should be set to 15. We also noticed that among the classification result of the locomotion part, classification accuracies of sit and stand were 80.77% and 84.28% respectively, this classification accuracy is quite satisfactory. However, the walk state, was only able to reach a 48.60% classification accuracy. This result is coincided with the previous analysis of the Decision Tree model and the Naive Bayes model, in order to prove that the state of walk can easily be misjudged as the state of stand. I analyzed the detailed classification result, and the detailed results are shown in Table 5.6:

Locomotion Classify as	Sit	Walk	Stand
Sit	378	11	34
Walk	37	104	16
Stand	49	91	268
None	4	8	0
Total number	468	214	318
Accuracy /%	80	48	84

Table 5.6: Detailed classification result of locomotions

The test includes a total number of 1,000 test cases, including 468 test cases for the sit state, 214 test cases for the walk state and 318 test cases for the stand state. The results show that 91 test cases of walk were misjudged as the state of stand, while the total number of walk test cases is 214, this probability is as high as 42.52%. This is also a good explanation of why the classification accuracy of the walk state is relatively lower than the other.

According to the analysis results of this experiment, I came up with a hybrid neural network model, which combines the ideals of the BP neural network together with the Decision Tree. The main idea of this hybrid neural network model is: since the traditional neural network model cannot distinguish the state of walk and stand very well, we can divide the classification task into two stages. In the first stage, make a rough classification of the three locomotion states, which aims to distinguish the sit state, and the walk state and the stand state will be classified as a new category. In the second stage, this new category will be fine classified, using the different values of the key attributes to distinguish walk from stand.

According to the design idea of this hybrid neural network model, the construction of the first layer of the neural network is relatively simple, and the parameters that need to be determined is which attribute value can be used to distinguish the two easily misjudged states. Since I have already taken all the acceleration sensor data into consideration when constructing the neural network model, I will not consider using the acceleration data. There are seven sets of complete inertial measurement sensors on the subject's body, and these data can be used to realize the fine classification in the second stage, to distinguish the walk and the stand state. A pair

of inertial measurement sensors were worn on the subject's shoes, as the significant difference between walking and standing lies in the movement of the feet. When a person is standing, because there is no foot movement, the value of the sensor tend to be more stable, and when walking, the sensor values of the shoes should be changed more frequently. Due to the movement of the foot being a regularly repeated operation, we can choose the sensors either on the left foot or the right foot for analysis. To prove this conclusion, I randomly selected 500 continuous data samples of the walk state and another 500 continuous data samples of the stand state, and make a statistical analysis of the sensing data on the subject's right shoe. The results are shown in Table 5.7.

Locomotion	Walk	Stand
Maximum value	144.2657	-23.3523
Minimum value	6.5869	-88.4466
Mean value	39.2941	-33.0973

Table 5.7: Comparison of walking and standing in sensing data

From the above results, we can find some of the expected regularities. First, when comparing the data of stand, we can observe an obvious fall-and-rise in the data of walk, which reaches 138, while the change range of stand is only 65. Secondly, the ranges of the sensing value of these two different states are different. The typical value of walking is positive, while the typical value of standing is negative. Finally, there are obvious differences in the mean value of these two states, the mean value of the walk data is 39.29, the mean value of stand data is -33.09. Based on the above analysis, it is a feasible option to choose this set of inertial measurement sensing data as the key attribute to distinguish the two states.

So far, the overall framework design of this hybrid neural network model has been completed, the main part of the model is divided into two stages. The first stage is the design of a BP neural network for rough classification of data, the second stage is the fine classification according to the specified value. The procedure is shown in Figure 5.8:

Figure 5.8: The procedure of locomotion classification.

This is the design of the hybrid neural network model for the locomotion classification. In the next section, I will describe the design of a neural network model for gesture classification.

5.3.2 Classification of the gesture part

In the gesture classification part, because I used the same input and output data format as the locomotion classification, therefore, the design of the neural network hidden layer nodes do not need to change. In this part, the network still has 15 nodes in the hidden layer.

In this part, the analysis is similar with the locomotion part. I find the most easily misjudged gesture is the open gesture and the close gesture, so I merged these two gestures in the first stage, and I prepared 1,500 test cases, 500 test cases for each gesture. The classification result is listed in Table 5.9:

Gesture Classify as	Open & Close	Drink Cup	Clean Table
Open & Close	462	8	42
Drink Cup	3	492	49
Clean Table	35	0	409

None	0	0	7
Total number	500	500	500
Accuracy /%	92.4	98.4	81.8

Table 5.9: The result of the tests

From the classification result above, we can see that the classification accuracies are all over 80%. This result is very desirable, especially the accuracy of drink cup which is 98.4%. The next step is trying to distinguish the open gesture and the close gesture. I attempt to construct a neural network model to recognize these two gestures, and the classification result is listed in Table 5.10:

Gesture Classify as	Open	Close
Open	388	376
Close	112	124
Total number	500	500
Accuracy /%	77.6	75.2

Table 5.10: The classification of open and close gestures

According to the analysis, the recognition of gesture part is also divided into two stages: In the first stage, the neural network focus on classifying the drink cup and clean table. In the next stage, focus is on the fine classification of the open and close gesture. The procedure is shown in Figure 5.11:

Figure 5.11: The procedure of gesture classification

Until now, the design of the hybrid neural network for recognizing the locomotions and gestures is completed. In the next section, I will list the final classification result of the hybrid neural network, and make some analyses according to the result.

5.3.3 Result and analysis

According to the design of the hybrid neural network, the final classification result of the locomotion part and the gesture part can be seen in Table 5.12 and 5.13:

Locomotion	Sit	Walk	Stand
Accuracy /%	80.2	84.9	86.1

Table 5.12: The classification result of locomotion

Gesture	Open	Close	Drink cup	Clean table
Accuracy /%	74.2	72.7	96.8	80.5

Table 5.13: The classification result of gesture

The classification result proves that in this case of action recognition, the hybrid neural network works better than the previous work of the Decision Tree model and the Naive Bayes model.

The result also shows that the classification accuracy of the states which have some significant differences to the other is usually higher than the complicated ones. For instance, in the locomotion part, the state of standing has the highest classification accuracy, and standing is the simplest locomotion, it tends to be a stationary state while the other states are accompanied by some movement. This makes the sensing value change a little and easy to be recognize. In the gesture recognition part, the accuracy of the drink cup gesture is much higher than the others. This is mainly because of the movement of drinking is a little different to the other gestures. When a person drinks, the hand moves from down to up, while the other gestures moves from back to front, or from front to back. So, the difference between the directions of hand movements makes the drinking gesture easy to distinguish from the others.

5.3.4 Advantage and disadvantages

So far, the implementation and result analysis of this hybrid neural network models have been completed. Through the entire process of implementation, the advantages and disadvantages of hybrid neural network can be clearly stated.

One of the advantages of this neural network is that it has a strong nonlinear fitting capabilities, which can be used to map arbitrarily complex nonlinear functions. This advantage of the neural network makes it possible to apply in the process of this experiment using a large number of continuous sensing datasets. A neural network has a strong self-learning ability, and learning rules is relatively simple. It can keep adjusting its structure according to the results recorded in the study procedure. Therefore, the neural network model can be used to identify the human daily activities with high feasibility. In the process of my research, using the combined ideas of the neural network model and the Decision Tree model, not only makes the whole classification process more streamlined, but also resulted in a significant improvement of the classification accuracy, compared to the result of the typical neural network model or typical Decision Tree model.

The main drawback of neural networks is that it is not able to explain its reasoning and its inference based. It is not able to feedback regarding problems encountered in the process, so that we can make the appropriate analysis. The most typical problem is that when the training data is not sufficient or there is an error existed in the training data, the neural network model will not work properly. Essentially due to the gradient descent algorithms in the BP neural network, and the complicated objective function to be optimized, these are the main reasons why the learning rate

is relatively slow in the BP neural network model. Function approximation and generalization ability of the network is closely related to the typicality of the training samples, therefore, selection of attributes is a very important factor which eventually affects the performance of the neural network. In this experiment, the dataset included a relatively large amount of information, but I still take the full range of sensor data as the input of the neural network. This may be the second reason for the slow learning rate. Finally, since at the present stage there is no complete and unified theoretical guidance for designing the neural network structure, generally it is only selected by experience, although I made a lot of experiments during the experiment to obtain a better network structure and constantly made adjustments still I am not entirely sure the network structure has achieved the optimal effect.

5.4 Summary

In this section, a summary of the three completed classifiers will be presented.

In the Decision Tree model, the classification accuracy of locomotion activities is much higher than the accuracy of gesture activities, but it is still not satisfactory in terms of expectations. With the analysis of the intercept and slope of the figure, the procedure of the classification is intuitional. The challenge of this model lies in the construction of the classifying tree, and it can be improved by taking more attributes into consideration or with the technique of RFID.

In the Naive Bayes model, the classification accuracy is relatively balanced in the locomotion activities and the gesture activities. Based on the assumption of independent attributes, the classifier mandatorily removed the correlation between attributes. The superiority of this model mainly lies in the concise classification in a supervised learning condition.

During the course of implementing the neural network, some ideas of possible improvements came to mind. The obvious advantages of the high ability of dealing with the large-scale data matrix made the neural network model appropriate to adopt in my study. With the characteristics of neural network, it was possible to cut off the simplification of the data, since each kind of simplification inevitably caused varying degrees of loss of information. Even the neural network performed better than the other classifiers, but there is still rooms for further improvement.

Based on the analysis of all experiments results, the ideas of the hybrid neural network model was designed and implemented, which combined the ideas of the

Decision Tree model and the neural network model. This hybrid neural network fully used the advantages of both data mining techniques, and hid their shortcomings at the same time, achieving a desirable classification accuracy.

6 Conclusion

The research of activities of daily life recognition has been a point of interest for many researchers for a while, however, searching and applying a high-efficient, high-precision data mining technique to deal with the continuous sensing data has posed new challenges. The overall aim of this research is studying and comparing several data mining techniques suitable for the specific data sample, then select and design three kinds of appropriate data mining models. The objective of this research has been to analyze the classification result and evaluate them. In order to achieve the objective, this thesis proposed the design and implementation, also listing the classification accuracy for each data mining model.

Firstly, Goal 1 was to study data mining techniques. During my research, I mainly studied four kinds of data mining techniques, including the Decision Tree method, the Naive Bayes method, the neural network model and the hidden Markov model, analyzing the feasibility of them in experimental conditions. Based on the objective of this thesis, I successfully applied the Decision Tree method, the Naive Bayes method, and the neural network model during the research process.

Secondly, in order to achieve Goal 2, simplifying the original dataset, I searched for appropriate methodologies related to data processing. The application of these methodologies makes carrying out the research smooth. The Principal Component Analysis algorithm plays an important role in the data preprocessing part, and it successfully achieved the objective of reducing the dimensions of the original dataset. The result also showed the Least Square method performed well in the linear fitting part. With the help of these auxiliary algorithms, the classification result also improved.

Thirdly, Goal 3, applying the data mining techniques to this dataset, was achieved by designing and implementing all the data mining classifiers. The evaluation of the three data mining methods is proposed based on the final result. The Decision Tree model is suitable for less complicated classification using the differences between the key attributes' values. The Naive Bayes is the simplest, the classification process is carried out automatically but the result is not desirable. The neural network works well only for relatively simple activities. However, the analysis showed that there is still much room for improvement for each data mining method.

Finally, Goal 4, proposing some improvements of the classifiers, was achieved by designing a hybrid classification model combining the neural network model and the Decision Tree model. The result showed that this model performs much better than the typical neural network or the typical Decision Tree model.

6.1 Ethical considerations

This thesis work realized the human daily activities recognition based on the wearable sensor values. This research can be used to help people who require daily health care guidance as well as the elderly people who live alone and require emergency accident detection, for example, falling accidents. With the permission of the users, the data of on-body sensors can be acquired and analyzed. However, this research has inherent limitations. In order to obtain the acceleration data of the subject's movement, the sensors should be tied onto the body of the person during the recognition procedure. The sensed value will be transported and recorded periodically, thus having the subject monitored at all times. The problem of security issues of the sensed data also needs to be taken into consideration. Since the activities can be recognized by these data, in order to avoid unnecessary trouble, others should not be able to access them. During the whole experiment, the sensors have to be worn on the body of the subject at all times, this is not possible in a real life.

6.2 Future work

Future work could focus on the following three improvements:

Due to time limitations and my study ability, this thesis work focuses on three data mining techniques: Decision Tree, Naive Bayes and neural network. There are also many other data mining techniques that can be applied in this field, such as the hidden Markov model, the support vector machine, and the regression algorithm. If an appropriate selection of the attributes has been made, and flexibly adjusted to the classification model based on the actual situation, this might work better in this case.

Constructing a sensing dataset based on the requirement analysis in the practical application is another future work of this thesis. The whole procedure of my work is based on a dataset called the Opportunity Dataset. With the help of today's smartphones, wearable sensors or something similar, a similar sensing dataset can be

constructed depending on what is needed. Moreover, the type and quantities of specific activities can be altered as well.

During the study, all experiments were carried out offline. However, in the practice application, the cost of time and calculation should be taken into consideration.

References

- [1] Ashton, Kevin . "That 'Internet of Things' Thing, in the real world things matter more than ideas", RFID Journal, 2009.
- [2] Azhar Mahmood, Ke Shi, Shaheen Khatoon, and Mi Xiao, "Data Mining Techniques for Wireless Sensor Networks: A Survey", International Journal of Distributed Sensor Networks, 2013.
- [3] Ian H. Witten and Eibe Frank, "Data Mining: Practical machine learning tools and techniques", Morgan Kaufmann, ISBN 978-0123748560, 2011, pp. 664.
- [4] Yu-Jin Hong, Ig-Jae Kim, Sang Chul Ahn, Hyoung-Gon Kim, "Activity Recognition using Wearable Sensors for Elder Care", IEEE journal on Future Generation Communication and Networking, vol. 2, 2008, pp. 302-305.
- [5] Uslu G., Altun O., Baydere S., "A Bayesian approach for indoor human activity monitoring", IEEE International Conference on Hybrid Intelligent Systems (HIS), 2011, pp. 324-327.
- [6] Lei Gao, Bourke, A.K., Nelson, J., "A system for activity recognition using multi-sensor fusion", Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE, 2011, pp.7869-7872.
- [7] Sharma, A. ; Young-Dong Lee ; Wan-Young Chung, "High Accuracy Human Activity Monitoring Using Neural Network", Convergence and Hybrid Information Technology, vol. 1, 2008, pp. 430-435.
- [8] Rubaiyeat, H.A. ; Tae-Seong Kim ; Hasan, M.K., "Real-time Recognition of Daily Human Activities Using A Single Tri-axial Accelerometer", Embedded and Multimedia Computing (EMC), 2010 5th International Conference on, 2010, pp. 1 - 5
- [9] Liu Rong, Liu Ming, "Recognizing Human Activities Based on Multi-Sensors Fusion", Bioinformatics and Biomedical Engineering (iCBBE), 2010, pp. 1-4.
- [10] Xi Long ; Bin Yin ; Aarts, R.M., "Single-accelerometer-based daily physical activity classification", Engineering in Medicine and Biology Society, EMBC 2009. Annual International Conference of the IEEE, 2009, pp. 6107-7110.
- [11] Si-Jung Ryu ; Jong-Hwan Kim, "Classification of long-term motions using a two-layered hidden Markov model in a wearable sensor system", IEEE International Conference on Robotics and Biomimetics (ROBIO), 2011, pp.

2975-2980.

- [12] Chernbumroong, S. ; Atkins, A.S. ; Hongnian Yu, "Activity classification using a single wrist-worn accelerometer", IEEE International Conference on Software, Knowledge Information, Industrial Management and Applications (SKIMA), 2011, pp. 1-6.
- [13] Delgado-Mata, C. ; Lee Cosio, B.M., "HMM and NN for Gesture Recognition", Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2010, pp. 56-61.
- [14] Zhen-Yu He ; Lian-Wen Jin, "Activity recognition from acceleration data using AR model representation and SVM", International Conference on Machine Learning and Cybernetics, vol. 4, 2008, pp. 2245-2250.
- [15] Y. Yuan and M.J. Shaw, "Induction of fuzzy Decision Trees". Fuzzy Sets and Systems 69 (1995), pp. 125–139
- [16] John, George H.; Langley, Pat, "Estimating Continuous Distributions in Bayesian Classifiers". Proc. Eleventh Conf. on Uncertainty in Artificial Intelligence. Morgan Kaufmann, 1995, pp. 338–345.
- [17] Lawrence Jeanette, "Introduction to Neural Networks", California Scientific Software Press. ISBN 1-883157-00-5, 1994.
- [18] Bishop, C.M., "Neural Networks for Pattern Recognition", Oxford: Oxford University Press. ISBN 0-19-853849-9 (hardback) or ISBN 0-19-853864-2 (paperback), 1995.
- [19] Pete McCollum, "An Introduction to Back-Propagation Neural Networks", <http://www.seattlerobotics.org/encoder/nov98/neural.html>

Retrieved 2014-04-23.
- [20] Piyathilaka, L.; Kodagoda, S., "Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features," Industrial Electronics and Applications (ICIEA), 8th IEEE Conference on, 2013, pp.567,572, 19–21.
- [21] Hsu, Chih-Wei; Chang, Chih-Chung; and Lin, Chih-Jen, "A Practical Guide to Support Vector Classification", Department of Computer Science and Information Engineering, National Taiwan University, 2003.
- [22] MacQueen, J. B. "Some Methods for classification and Analysis of Multivariate Observations". Proceedings of 5th Berkeley Symposium on Mathematical

- Statistics and Probability 1. University of California Press, 1967, pp. 281–297.
- [23] Rakesh Agrawal and Ramakrishnan Srikant, “Fast algorithms for mining association rules in large databases”, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB,, Santiago, Chile, 1994, pp. 487-499.
- [24] Jolliffe, I.T., “Principal Component Analysis”, second edition (Springer), 2002, pp. 487
- [25] C.R. Rao, H. Toutenburg, A. Fieger, C. Heumann, T. Nittner and S. Scheid. “Linear Models: Least Squares and Alternatives”. Springer Series in Statistics. 1999.
- [26] Roggen, D.; Calatroni, A.; Rossi, M.; Holleczech, T.; Forster, K.; Troster, G.; Lukowicz, P.; Bannach, D.; Pirkel, G.; Ferscha, A.; Doppler, J.; Holzmann, C.; Kurz, M.; Holl, G.; Chavarriaga, R.; Sagha, H.; Bayati, H.; Creatura, M.; del R Millan, J., “Collecting complex activity datasets in highly rich networked sensor environments”, Seventh International Conference on Networked Sensing Systems (INSS), 2010, pp.233-240.
- [27] Lukowicz, P.; Pirkel, G.; Bannach, D.; Wagner, F.; Calatroni, A.; Foerster, K.; Holleczech, T.; Rossi, M.; Roggen, D.; Troester, G.; Doppler, J.; Holzmann, C.; Riener, A.; Ferscha, A.; Chavarriaga, R., “Recording a Complex, Multi Modal Activity Dataset for Context Recognition”, 23rd International Conference on Architecture of Computing Systems (ARCS), 2010, pp. 1-6.
- [28] Dimension Engineering, “A beginner’s guide to accelerometers”, <https://www.dimensionengineering.com/info/accelerometers>, retrieved in 2014-04-20.
- [29] Neil M. Barbour, “Inertial Navigation Sensors”, NATO RTO Lecture Series 232, “Advances in Navigation Sensors and Integration Technology”, 2004.
- [30] Activity and Context Recognition with Opportunistic Sensor Configurations, “challenge Dataset”, <http://www.opportunity-project.eu/challengeDataset>
- Retrieved 2014-05-01
- [31] Liggins, Martin E., David L. Hall, and James Llinas. Multisensor Data Fusion, Second Edition. Theory and Practice, 2008.
- [32] Dodge, Y, “The Oxford Dictionary of Statistical Terms”, OUP. ISBN 0-19-920613-9, 2003.

-
- [33] Wlodzislaw Duch , Norbert Jankowski, "Survey of Neural Transfer Functions", Neural Computing Surveys, 1999.
- [34] Robert Hecht-Nielsen, "Theory of the back-propagation neural network", First Annual I.N.N.S. Meeting, 1988.

Appendix A: The detailed classification result of gesture in Naive Bayes model

$\begin{matrix} G \\ CA \end{matrix}$	OD	CD	OF	CF	ODW	CDW	ODR	CDR	CT	DC	TS
OD	44.44	29.04	0.00	0.31	0.15	1.26	1.42	0.63	1.26	16.34	5.07
CD	19.27	56.55	0.31	0.94	1.26	1.42	2.36	1.42	0.94	12.00	3.47
OF	0.32	1.29	48.86	40.12	1.61	4.53	1.29	0.97	0.32	0.64	0.00
CF	0.00	0.65	26.47	58.49	0.98	2.61	4.57	2.61	2.61	0.98	0.00
ODW	3.27	7.10	12.02	14.48	39.61	6.28	10.38	2.45	3.27	1.09	0.00
CDW	5.08	4.58	6.10	5.85	15.77	18.32	15.52	27.22	0.50	1.01	0.00
ODR	4.01	5.73	5.96	14.10	8.71	0.57	29.47	27.98	1.14	2.29	0.00
CDR	7.02	4.21	2.94	9.55	4.91	1.40	20.92	45.64	2.52	0.84	0.00
CT	0.00	0.00	0.80	11.74	4.85	0.80	4.45	1.61	74.89	0.80	0.00
DC	1.06	0.67	2.61	9.79	0.09	0.00	0.00	0.67	4.75	77.98	2.32
TS	7.69	30.76	0.00	1.39	0.69	0.00	0.69	0.69	0.00	9.09	48.95
A/%	44.44	56.55	48.86	58.49	39.61	18.32	29.47	45.64	74.89	77.98	48.95

Abbreviation:

G - Gesture	OF - Open Fridge	ODR - Open Drawer	TS - Toggle Switch
CA - Classify as	CF - Close Fridge	CDR - Close Drawer	A - Accuracy
OD - Open Door	ODW - Open Dish Washer	CT - Clean Table	
CD - Close Door	CDW - Close Dish Wash	DC - Drink Cup	