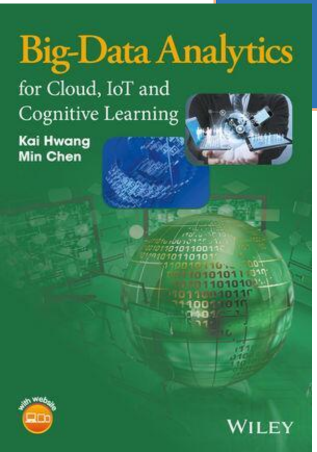
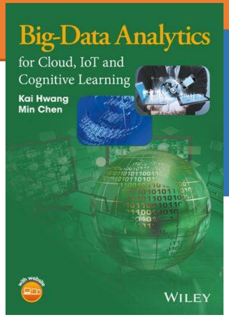
Big Data Analytics for Cloud, IoT and Cognitive Computing

Part 3 Big Data Analytics for Health-Care and Cognitive Learning

Chap. 7 Machine Learning for Big Data in Healthcare Applications



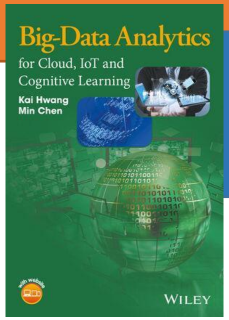
Companion Website: http://www.wiley.com/go/hwangIOT

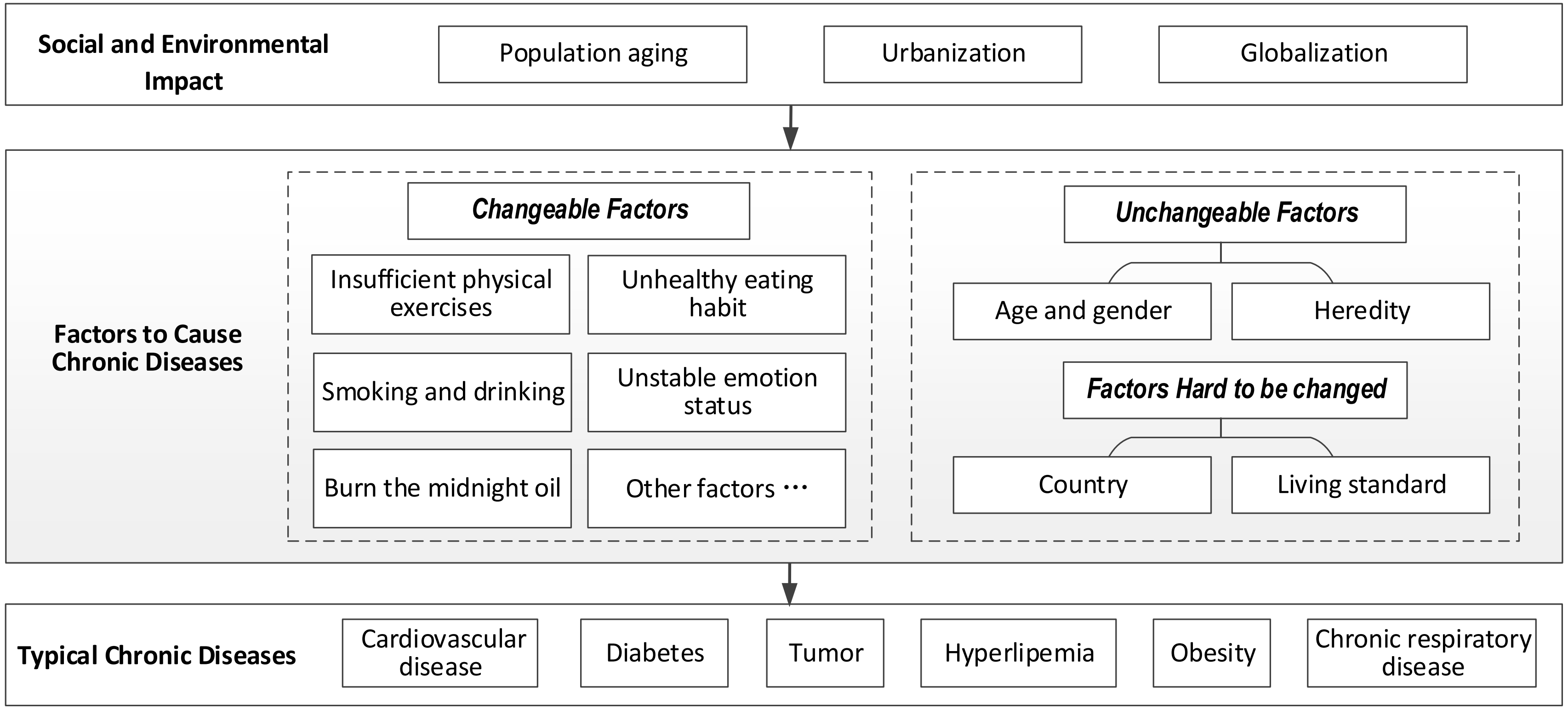


Summary

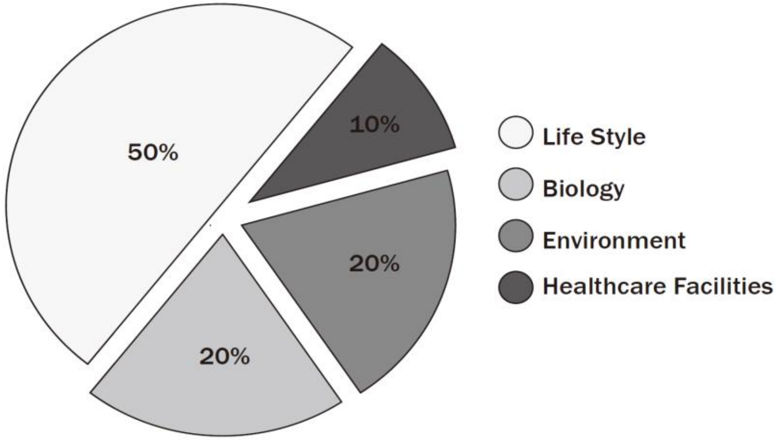
This chapter is devoted to studying how machine learning and data analytics can be applied to healthcare applications. We start with an introduction of the medical and healthcare problem environments. Then we present IoT-assisted healthcare monitoring systems. In data analytics for medical applications, we focus on cloud-based detection of chronic diseases. This chapter is highly relevant to those machine learning algorithms presented in Chapters 6 and 8. We will not repeat the theories behind those ML algorithms, only biomedical and healthcare applications are studied in this chapter, including both system settings and reported performance. In particular, we present an intelligent human-machine interface approach with smart clothing and robotic interaction healthcare clouds.

Healthcare and Chronic Disease Detection Problem

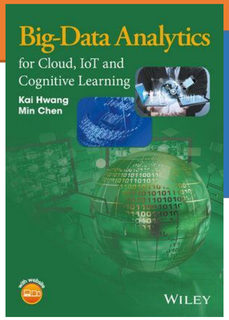




Factors that affect the detection accuracy in detecting chronic diseases



Determinants of health (statistics from centers for disease control in 2003)



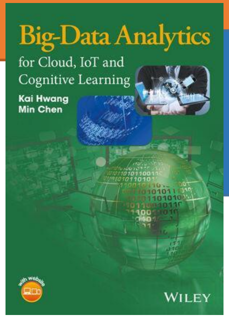
Software Libraries for Machine Learning Applications

Here, we identify a few software toolkits that can analyze dataset help users put together running programs. Surprisingly, many of these ML packages are from open source. Readers may check the developer websites for more details of the functionality and capability of those programs or runtime support systems provided.

We identify other software libraries and toolkits that can analyze dataset and help users put together programs for running deep learning tasks. Only brief introductory information is given here. However, we will dig much deeper with Google TensorFlow framework in subsequent sections. The Spark and TensorFlow libraries have enriched our capability to develop new ML or DL applications. A lot of cognitive activities which humans (even a newborn baby) can perform easily but not always with certainty, now we can train the computer to handle those screening and filtering tasks routinely to save us time and augment our decision-making processes with better evidence and support.

Table 7.2. Commonly Available Machine Learning Toolkits

Toolkit or Framework, Language, Web site of Developer	Short Description of Functionality and Capability
Scikit-learn, Python, **http://scikit-learn.org/stable/**	Built with NumPy and Matplotlib and provides simple and efficient mathematical tools for data mining and big data analysis.
Shogun, C++, **http://www.shogun-toolbox.org/**	SWIG interfaces enables communication between C++ and target languages Python, Octave, R, Java, C#, etc., focusing on SVM kernel functions.
Accord, Aforge.net,.NET, **http://accord.codeplex.com/** **http://www.aforgenet.com/framework**	Applied for audio/image processing in face detection and image stitch on SIFT, supporting real-time mobile computing with ANNs or decision trees
Mohout, Hadoop, **https://mahout.apache.org/**	Using MapReduce to run on a single or multiple nodes of a Hadoop cluster, greatly improving the data volume. （Details in Chapter 9）
MLlib, Spark **http://spark.apache.org/mllib/**	MLlib is designed to enable many ML algorithms to run fast on a large cluster. It supports personalized ML code design. (Details in Chapter 9)
Cloudera, Hadoop, **http://www.cloudera.com/**	Provided by Cloudera Hadoop distribution, enabling machine learning models to run on real-time data flow, such as spam email filtering.
GoLearn, Go, **https://github.com/sjwhitworth/golearn**	Developed by Go and developed with Google to support customized code design with simple tools to extend data structure and source code.
Weka, Java， **http://weka.wikispaces.com/**	Weka is designed for data mining, preprocessing, classification, regression and clustering applications with visualization support.
CUDA-Convnet, C++， **https://code.google.com/p/cuda-convnet**	CUDA is a speed-up toolkit of GPU, while CUDA-Convnet is a machine learning library for ANNs based on using fast GPU clusters.
ConvNetJS, JavaScript， **http://www-cs-faculty.stanford.edu/people/karpathy/convnetjs/**	An online training service for deep learning, which helps the user to understand the algorithms intuitively by showing some simple demos to users.
FBLearner Flow, Python， **https://code.facebook.com/posts/1072626246134461/**	This platform reuses many algorithms in different products, by stretching into thousands of customized experiments of simulation. It also offers automatic generation of user interface experiences from Python codes.

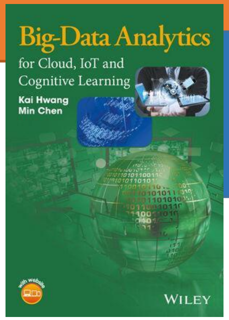


IoT-based Healthcare Systems and Applications

Healthy Internet of Things (Health-IoT) is an important path to solve medical health problems, and also has an important realistic meaning for promoting the development of the medical health industry and improving people’s quality of life. Compared with the traditional things-centered IoT, the Health-IoT is “human-centered”, and all network accesses, data analyses and services are conducted surrounding humans; the sensor at the data collection layer is not a common sensor, but a human body sensor for collecting physiological health parameters, and network accesses, data analyses and services are all conducted based on the “human-centered” idea.

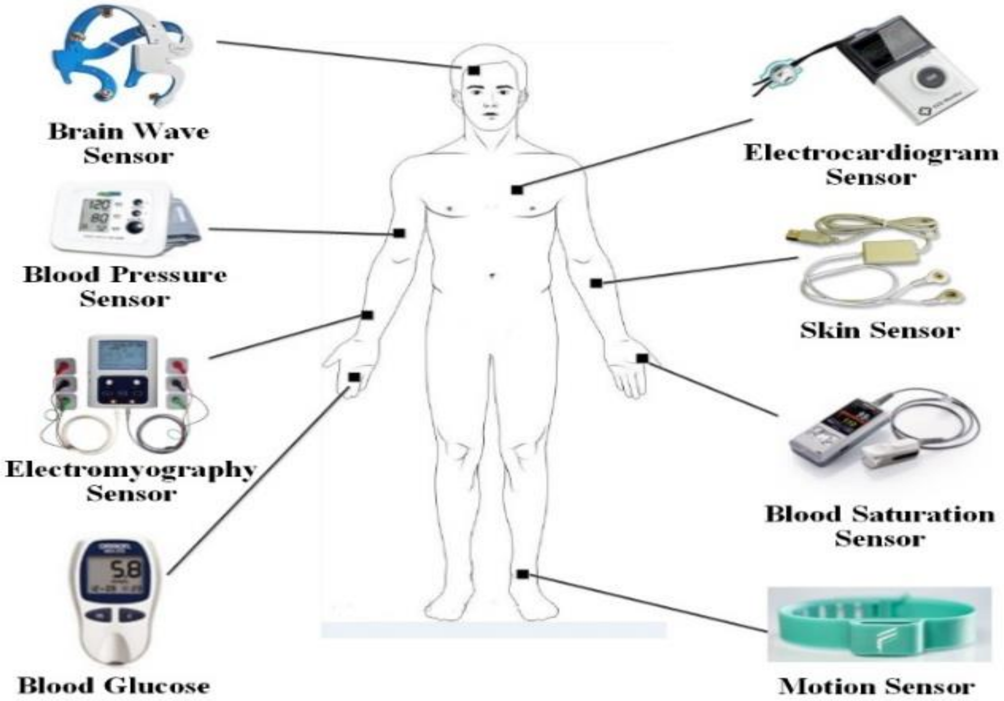
The previous Health-IoT emphasized the design of a human body sensor and the collection of human body physiological data, but did not fully consider the users’ mobility. Therefore, it is inconvenient to use in daily life and may even adversely affect daily life. The development of the mobile internet brings the integration of physical world, virtual world and social network, thus generating Cyber-Physical Society Systems (CPSS). Integrating the Health-IoT into CPSS, allows users to obtain the services and convenience brought by mobile health and mobile medical treatment, while users are under highly mobile conditions in the physical world and the social network space is an inevitable trend for development of Health-IoT.

Traditional IoT has been widely applied in the traffic, logistics and retail industries. With its maturity, the IoT attracts people’s attention in the field of healthcare. However, lots of applications, which promote health services to families or individuals by utilizing IoT technology, later were proven to be unsuccessful. Due to its importance to improve medical treatment quality and service efficiency, the Health-IoT is a milestone in health information development. It will play an important role in improving people’s health levels and enhancing their quality of life.

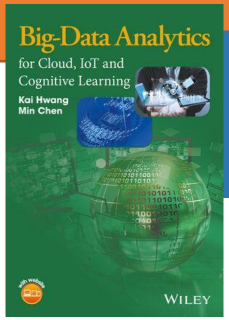


IoT Sensing for Body Signals

The layout of common human body sensors



- **Emedded Sensors:** The MHS device adopts dedicated a sensor, and possesses the advantage of high collecting precision, while it also features by disadvantages including high cost and insufficient portability and usability. This kind of device possesses the following features:
- **Wearability:** Most MHSs must be located on the human body so as to collect data precisely for them to take vital signs of humans as collecting targets. Therefore, almost all existing medical health collecting devices take wearability as the basic requirement. In this case, the users comfort can be improved and the accuracy of the collected data can be guaranteed during the collecting procedure.
- **Long working time**: The method of the dedicated medical health collecting device is different to universal mobile collecting devices. The purpose of the former is to collect data from the human body over a relatively long time period, which requires a highly power supply capability of MHS.
- **Stability:** MHS still can collect data normally when users are under strenuous exercise or in an extreme environment.
- **Low participation degree of users:** Different to the method of GMD, the functions of MHS are relatively independent, and most MHS devices do not need the intervention of users during the data collecting procedure, and users only need to start up the power source, and the MHS will start collecting.
- **Possessing data interim storage mechanism:** The weight and dimensions of MHS may be limited strictly to meet the wearable feature. Therefore, most MHS devices will not integrate the data transmission module, but will select the data storage module with relatively small dimensions, and adopt the data interim storage mechanism to store the collected data in advance, and then transmit the data through other network access devices.

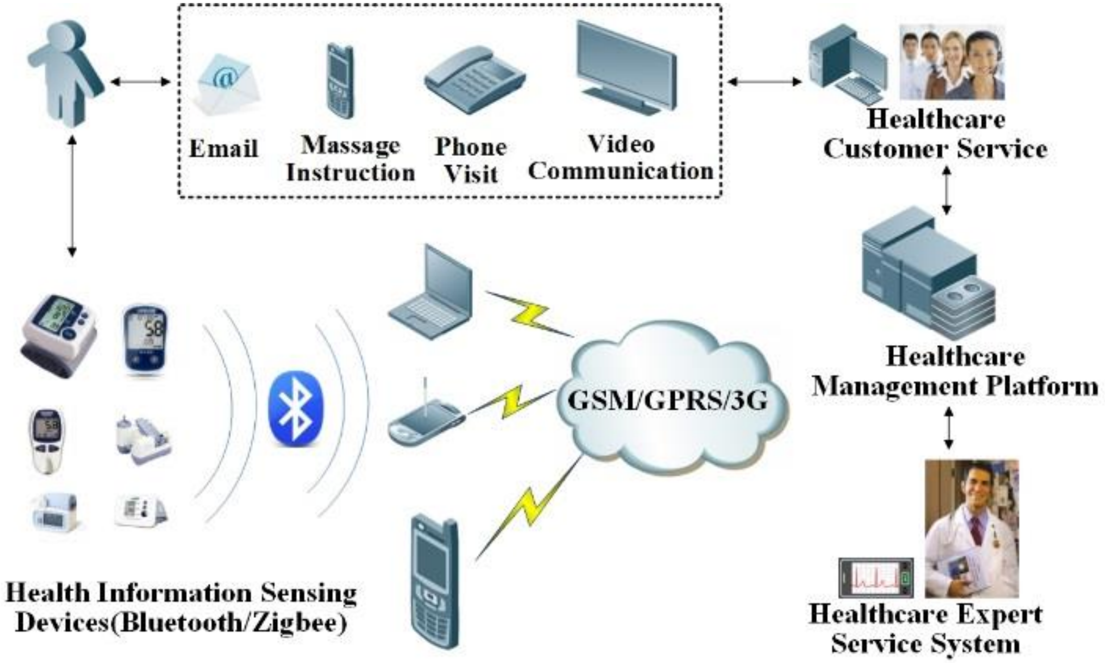


Healthcare Monitoring System

Common health monitoring services

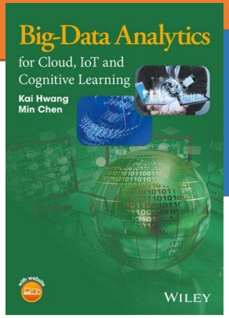
No.	Service Contents	Service method
1	Providing 24h remote ECG/blood pressure/blood glucose/blood oxygen/pulse/respiration/sleeping	Real-time monitoring service
2	Providing real-time warning of monitoring abnormality	Short message
3	Providing the service of notifying relatives of monitoring information	Short message
4	Providing the service of booking expert consultancy	Video or short message
5	Providing the emergency calling and aid service	Automatic telephone calling
6	Providing the family positioning service	Positioning
7	Regular health assessment report service	Short message or email
8	Regular health promotion care service	Short message or email
9	Regular follow-up service	Telephone
10	Life-long health record management service	Website inquiry
11	User data self-help inquiry service	Website inquiry
12	Providing 24h consulting hotline service	Telephone

Common health monitoring system based on community services.



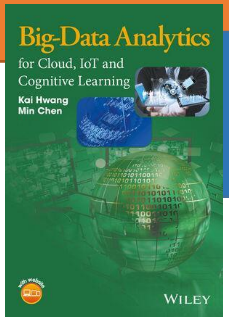
Several common health monitoring devices

Name of Device	Monitoring content	Additional functions
Blood Pressure Monitor	Blood pressure	Recording the historical blood pressure data
e-health Cloud Blood pressure monitor	Blood pressure	Integrating the cloud platform, historical data curves and transmitting distress information
Sunstudy GPS LBS	Tracking the old	Mobile-phone communication, SOS distress help-seeking alarm, and successively dialing three numbers; uploading the tracking position regularly and low-power alarm.
Smart blood pressure device	Blood pressure and heart rate	The blood pressure and heart rate monitoring may avoid atrial fibrillation by contacting doctors to obtain the right treatment and know situations of other patients.
jWatch wristwatch	Blood pressure and heart rate	Data analysis and manual calling center
Remote infant monitoring	Monitoring infants	Monitoring infants at distance and add other guardians.



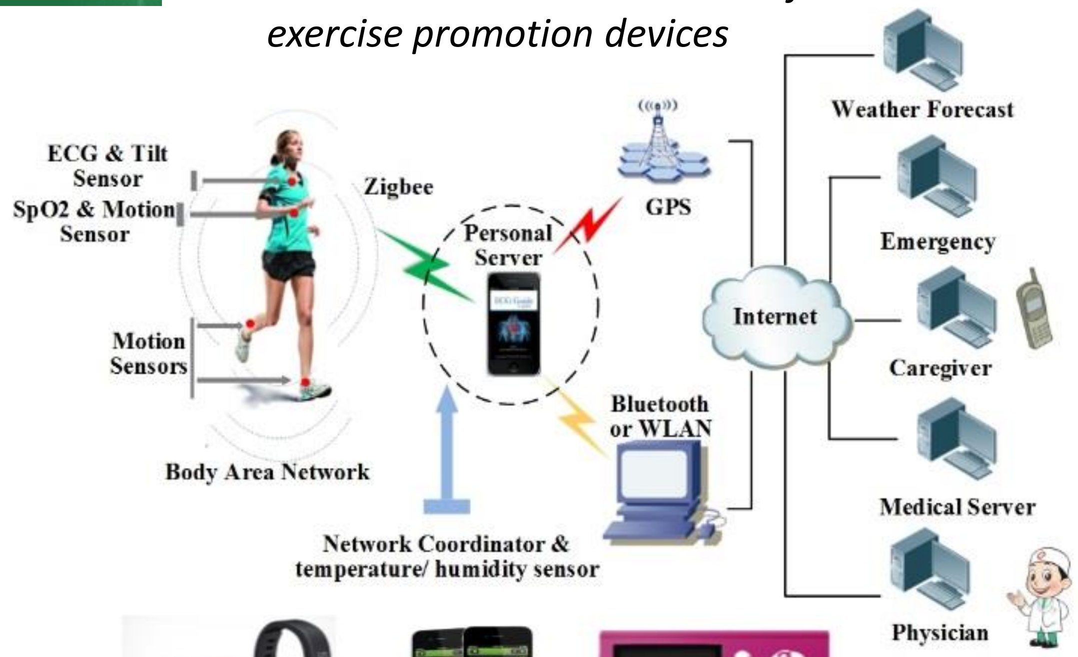
Healthcare Monitoring System

- **Health Cyber-Physical System**: Health-oriented mobile Cyber-Physical System (CPS) plays a vital role in existing medical monitoring applications, such as diagnosis, disease treatment and emergency rescue, etc. Some electronic medical intelligent network systems suitable for a large number of patients have been designed. The End-to-End delay of medical information delivery is the main concern, especially in the event of an accident, or in the period when there is epidemic disease outbreak.
- **Mobile Health Monitoring**: Several years ago, a mobile health monitoring system based on portable medical equipment and smart phones was proposed. Smart phones are used to collect physiological signals of the human body from a variety of health monitoring devices by virtue of dedicated smart phone application software. Then those physiological signals are transmitted to medical centers. If necessary, they can also notify caregivers and medical emergency institutions using the short message service of mobile phone.
- **Wearable Computing for Health Monitoring**: Over a long period, wearable devices and wearable computing are the key research topics to enable health monitoring. As new kinds of body sensor nodes, smart phones and smart watches are adapted to measure SpO_2 and heart rate; however, such measurement data has low accuracy, few signal types and limited medical uses.
- **Health Internet of Things**: Health IoT is another way to provide health monitoring service. The mobile sensing, localization and network analysis based on IoT technologies can be used for healthcare.
- **Ambient Assisted Living**: Ambient Assisted Living (AAL) aims at improving the life quality of patients, and it can notify relevant relatives, caregivers and healthcare experts. AAL-related technologies include sensing technology, physiological signal monitoring, home environment monitoring, video-based sensing, smart home technology, pattern analysis and machine learning.
- **Body Area Network based Health Monitoring**: Existing work on body area network (BAN) focuses on sensor node's energy saving, intra-BAN network design, implantable micro-sensors, physiological signal acquisition, etc. Portable smart wearable health monitoring system based on BAN has been developed. However, stability, sustainable and reliability of the system need to be improved.



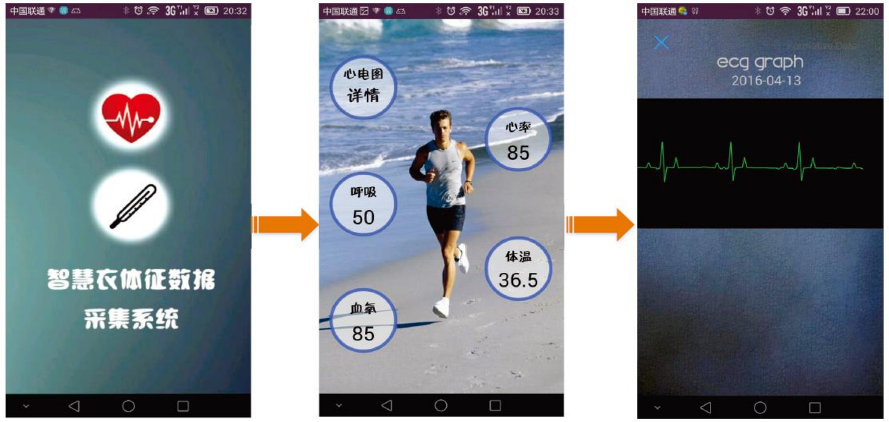
Physical Exercise Promotion and Smart Clothing

Communication architecture of exercise promotion devices

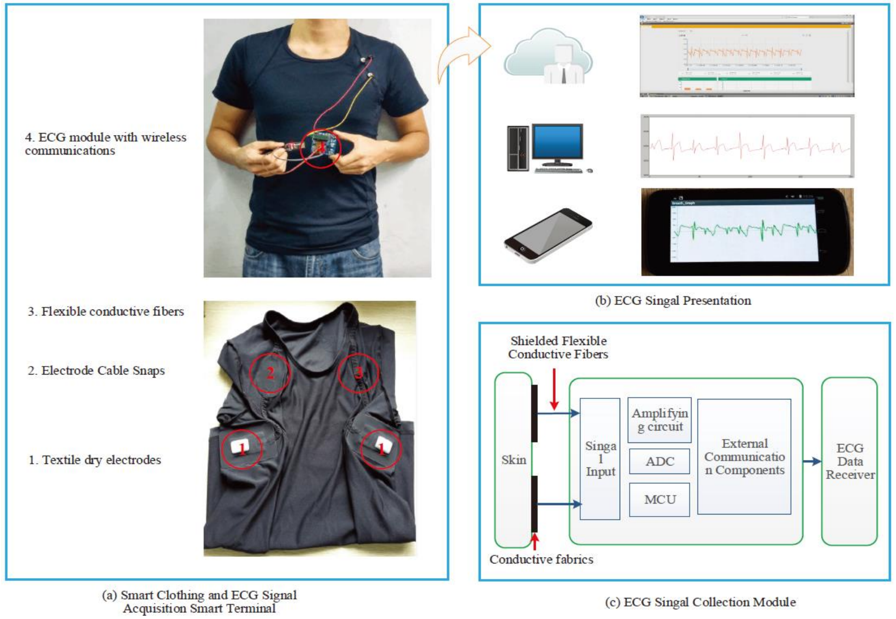


Exercise promotion products available in 2016

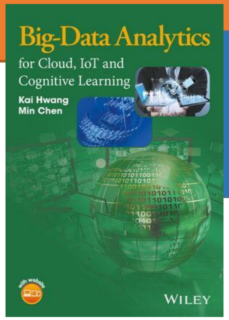
Smart clothing application software and testbed settings



(a) Mobile application software

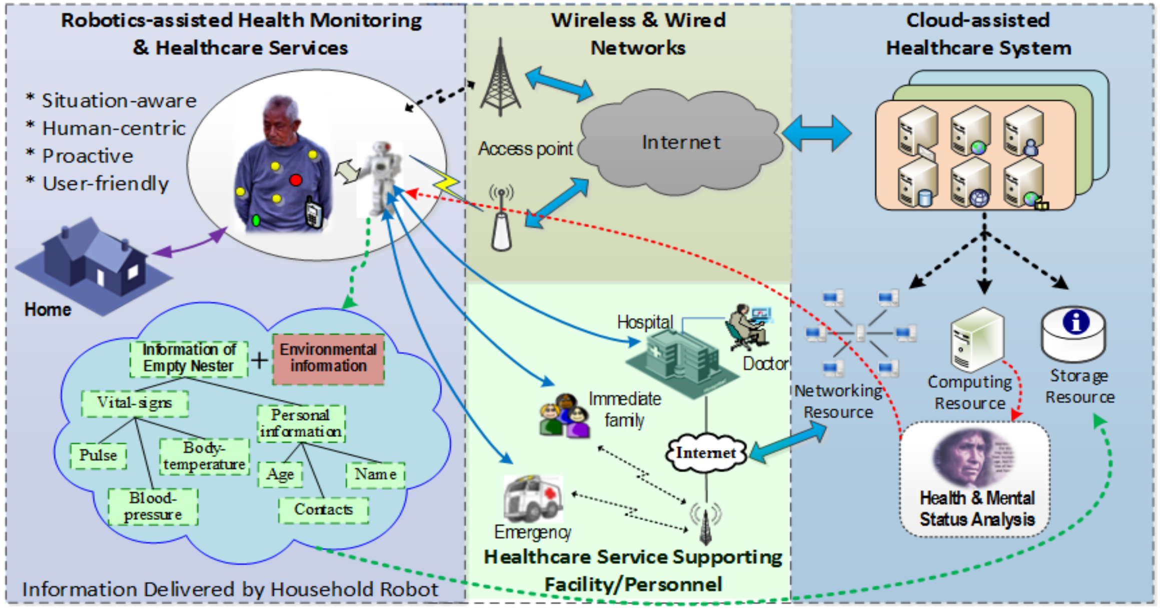


(b) The testbed settings



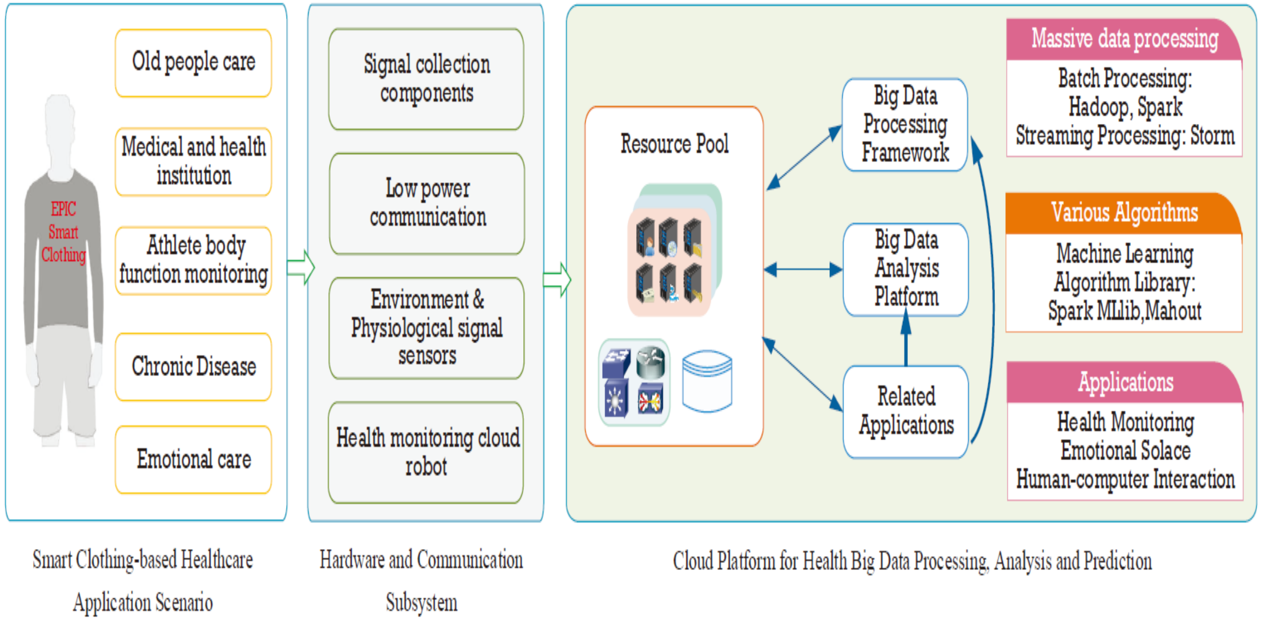
Healthcare Robotics and Mobile Health Cloud

Cloud computing is a new type computing and service mode based on the Internet. Through this method, pooling hardware and software resources and information can be supplied to the service requestor on a basis of requirement. The traditional robots are always restricted in the hardware and software functions where there are serious problems. But the cloud computing, as a good support to the robot technology can easily combine the cloud computing with the robot technology to build a cloud robot.

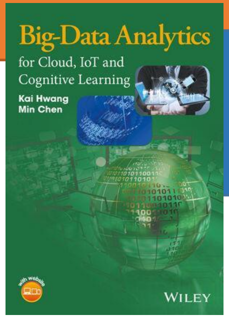


Example: Robotics and cloud-assisted healthcare system

A typical health monitoring system built with smart clothing and backend cloud



As a front end equipment, the robot takes charge in signal collection, specific action performance and some simple tasks of analysis and processing while the more complicated tasks which need a large-scale computing cluster will depend on the cloud. The cloud itself has strong storage and calculating ability, training, learning and building effective models by advanced machine learning algorithm and transmitting the results of calculation or analysis back to the robots. In that way, the robot will be provided with the second-wise brain with the help of cloud's strong analysis and processing abilities.



Big Data Analytics for Healthcare Applications

Example: Predictive Disease Diagnosis using Logistic Regression

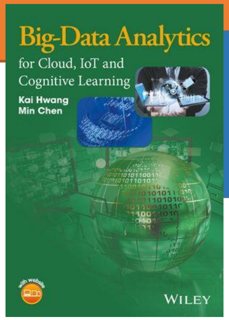
Table 7.5 has listed a data set of triglyceride, total cholesterol content, high-density lipoprotein, low-density lipoprotein, and hyperlipemia or not (1 for yes and 0 for no). These are collected from health examination data in a hospital, in Wuhan, China. Let's attempt to conduct preliminary judgment on whether the person has hyperlipemia, if his or her health examination data are {3.16, 5.20, 0.97, 3.49} in a sequence.

In this example, we need to judge whether an unknown person who received a health examination has hyperlipemia. As per data in the table, it is known that this problem is a dichotomy problem (1 for hyperlipemia or 0 for healthy) with four attributes (features). Therefore, we may conduct prediction and classification by use of logistic regression. Firstly, extract four attributes and combine them into one attribute, as $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$

Where x_1, x_2, x_3, x_4 stand for triglyceride, total cholesterol content, high-density lipoprotein and low-density lipoprotein respectively, z stands for the feature after combination. Secondly, estimate weight β by use of maximum likelihood method, adopt software MATLAB here, and conduct iteration solution to likelihood equation set with Newton-Raphson Method.

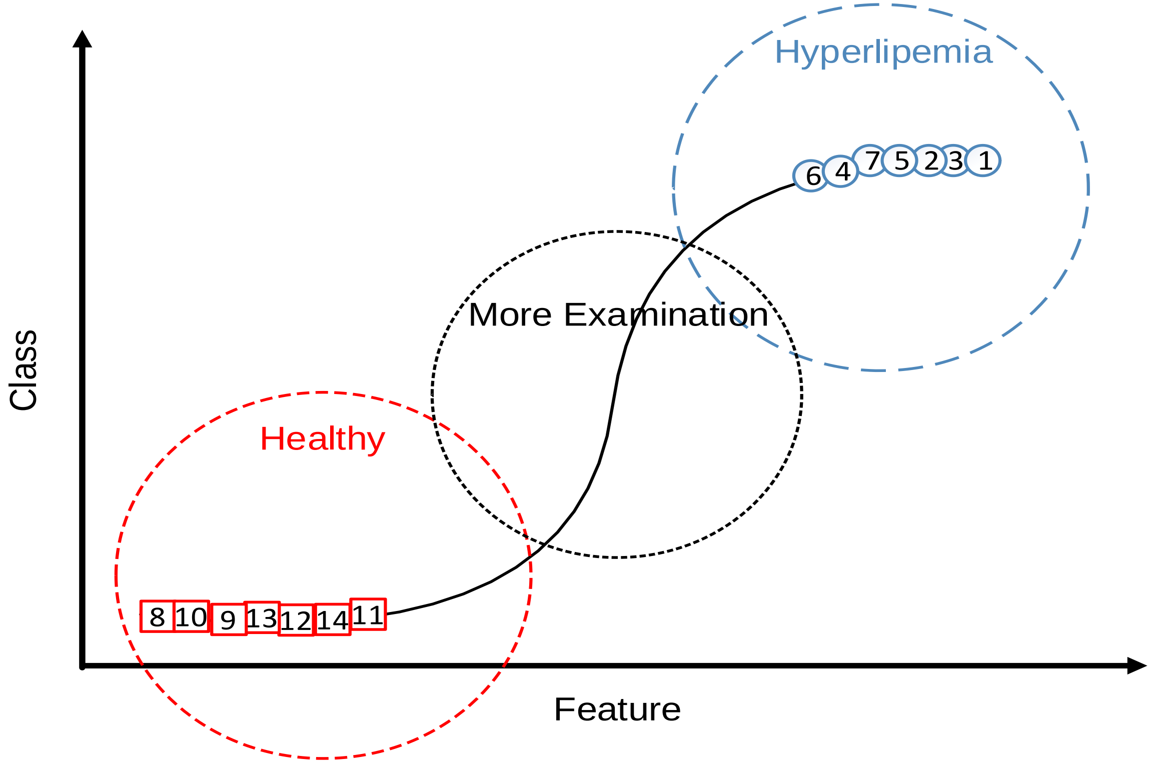
Health examination data for patients with hyperlipemia

Patient ID	Triglyceride	Total Cholesterol	High-Density Lipoprotein	Low-Density Lipoprotein	Whether hyperlipemia or not
1	3.62	7	2.75	3.13	1
2	1.65	6.06	1.1	5.15	1
3	1.81	6.62	1.62	4.8	1
4	2.26	5.58	1.67	3.49	1
5	2.65	5.89	1.29	3.83	1
6	1.88	5.4	1.27	3.83	1
7	5.57	6.12	0.98	3.4	1
8	6.13	1	4.14	1.65	0
9	5.97	1.06	4.67	2.82	0
10	6.27	1.17	4.43	1.22	0
11	4.87	1.47	3.04	2.22	0
12	6.2	1.53	4.16	2.84	0
13	5.54	1.36	3.63	1.01	0
14	3.24	1.35	1.82	0.97	0



Big Data Analytics for Healthcare Applications

Example: Predictive Disease Diagnosis using Logistic Regression

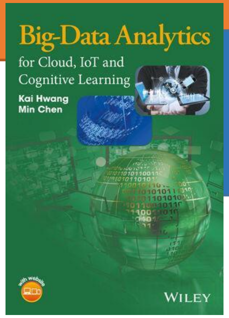


Classification results using logistic regression in Example

In accordance with results above, β_2 is relatively larger; thus it can be seen that whether one person has hyperlipemia or not is largely influenced by total cholesterol content in the health examination. Then figure out class for each sample in training dataset by use of sigmoid function. The results are $class = [1,1,1,1,1,1,1,0,0,0,0,0,0,0]$

$$\beta_0 = -132.3\,,\ \beta_1 = -3.1\,,\ \beta_2 = 39.6\,,\ \beta_3 = -2.9\,,\ \beta_4 = 3.2$$

The number in the figure stands for id of the person tested, and the circles in dotted line stand for class. It can be seen from the figure that the accuracy of classification with logistic regression in this instance is 100%, thus this model could be adopted for prediction. Lastly, let's predict whether a person whose data are {3.16, 5.20, 0.97, 3.49} respectively has hyperlipemia. Adopt the model above and conduct solving-equation-by-substitution, then $class = 1$. Therefore, that person is predicted to have hyperlipemia.



Big Data Analytics for Healthcare Applications

Example: Use of Baysian Classifier in Diabetic Analysis and Prediction

This example analyzes diabetic patients and predicts whether they have acquired the disease. The prediction is based on training from sample data on labelled patients on their obesity and blood sugar content. The sample data are given in Table 7.6. Here, *Yes* stands for obesity or diabetic patient and *No* for normal weight or healthy persons.

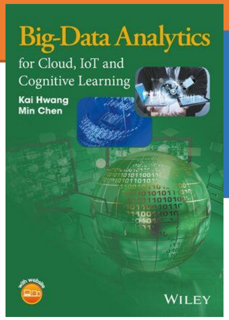
Table 7.6 Health examination data of diabetics patients

id	Obesity(A)	Blood Sugar Content(B) (mmol/L)	Diabetics Patient or Not
1	No	14.3	Yes
2	No	4.7	No
3	Yes	17.5	Yes
4	Yes	7.9	Yes
5	Yes	5.0	No
6	No	4.6	No
7	No	5.1	No
8	Yes	7.6	Yes
9	Yes	5.3	No

Table 7.7 Probabilistic results on patient obesity and blood sugar content

Diabetics	Obesity		Blood Sugar Content (mmol/L)	
	Yes	No	Mean Value	Variance
Yes	3/4	1/4	11.83	18.15
No	2/5	3/5	4.94	0.07

For simplicity, we denote attribute A for obesity and attribute B for blood sugar content. Based on the statistics from Table 7.6., we obtain the following probability distributions on patient obesity and blood sugar contents in Table 7.7



Big Data Analytics for Healthcare Applications

Example: Use of Baysian Classifier in Diabetic Analysis and Prediction

To predict class label of a person who received health examination, if X = (A=Yes, B=7.9), the Calculation of $P(Yes|X)$ and $P(No|X)$ is required. Using statistics data in Table 7.7, we have:

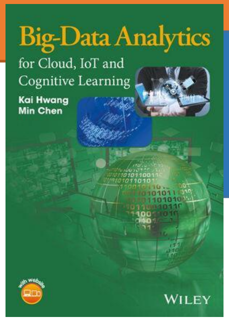
$$\begin{cases} P(A=Yes|Yes)=\frac{3}{4} & P(A=No|Yes)=\frac{1}{4} \\ P(A=Yes|No)=\frac{2}{5} & P(A=No|No)=\frac{3}{5} \end{cases} \qquad \begin{cases} P(Yes)=\frac{4}{9} \\ P(No)=\frac{5}{9} \end{cases}$$

As for index of blood sugar content, if class is *Yes*, then:

$$\begin{cases} \bar{x}_{yes}=\frac{14.3+17.5+7.9+7.6}{4}=11.83 \\ s_{yes}^2=\frac{(14.3-11.83)^2+(17.5-11.83)^2+\cdots+(7.6-11.83)^2}{4}=18.15 \end{cases}$$

If the class is *No*, then:

$$\begin{cases} \bar{x}_{yes}=\frac{4.7+5.0+4.6+5.1+5.3}{5}=4.94 \\ s_{yes}^2=\frac{(4.7-4.94)^2+(5.0-4.94)^2+\cdots+(5.3-4.94)^2}{5}=0.07 \end{cases}$$



Big Data Analytics for Healthcare Applications

Example: Use of Baysian Classifier in Diabetic Analysis and Prediction

With Gaussian distribution in blood sugar content, we have

$$\begin{cases} P(B{=}7.9 \mid Yes) = \dfrac{1}{\sqrt{2\pi} \times \sqrt{18.15}} e^{-\frac{(7.9-11.83)^2}{2\times 18.15}} = 0.062 \\ P(B{=}7.9 \mid No) = \dfrac{1}{\sqrt{2\pi} \times \sqrt{0.07}} e^{-\frac{(7.9-4.94)^2}{2\times 0.07}} = 9.98\times 10^{-28} \end{cases}$$

At the moment, conduct classification for *X* with naive Bayesian classification method,

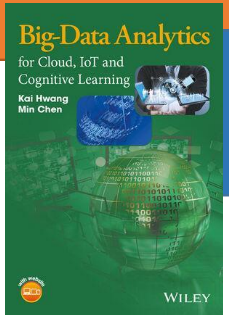
$$P(X \mid Yes) = P(A=Yes \mid Yes)P(B=7.9 \mid Yes) = \frac{3}{4}\times 0.062 = 0.0465$$

In a similar way, the probability of $P(X \mid No)$ is obtained as follows with errors estimated.

$$P(X \mid No) = P(A=Yes \mid No)P(B=7.9 \mid No) = \frac{2}{5}\times 9.98\times 10^{-28} = 3.99\times 10^{-28}$$

$$\begin{cases} P(Yes \mid X) = \dfrac{P(X \mid Yes)P(Yes)}{P(X)} = \varepsilon \times \dfrac{4}{9}\times 0.062 = \varepsilon \times 0.0276 \\ p(No \mid X) = \dfrac{P(X \mid No)P(No)}{P(X)} = \varepsilon \times \dfrac{5}{9}\times 3.99\times 10^{-28} = \varepsilon \times 2.218\times 10^{-28} \end{cases} \qquad \varepsilon = \frac{1}{P(X)}$$

We get $P(Yes \mid X)P(X) = 0.0276 > 2.218\times 10^{-28} = P(X)P(No \mid X)$. Therefore, the class of the person is *Yes* if $X = (A{=}Yes,\ B{=}7.9)$. Thus, the person has acquired diabetes.



Example: Selection for Hyperlipemia Detection Methods over Medical Data

Table 7.8 Labeled samples from examination reports of 20 hyperlipemia patients.

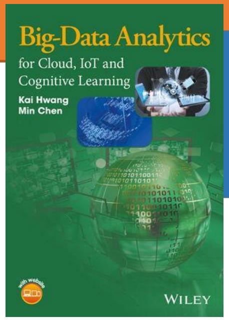
Patient ID	Triglyceride (mmol/L)	Total Cholesterol (mmol/L)	High-Density Lipoprotein (mmol/L)	Low-Density Lipoprotein (mmol/L)	hyperlipemia or not
1	3.07	5.45	0.9	4.02	1
2	0.57	3.59	1.43	2.14	0
3	2.24	6	1.27	4.43	1
4	1.95	6.18	1.57	4.16	1
5	0.87	4.96	1.36	3.61	0
6	8.11	5.08	0.73	2.05	1
7	1.33	5.73	1.88	3.71	1
8	7.77	3.84	0.53	1.63	1
9	8.84	6.09	0.95	2.28	0
10	4.17	5.87	1.33	3.61	1
11	1.52	6.11	1.29	4.58	1
12	1.11	4.62	1.63	2.85	0
13	1.67	5.11	1.64	3.06	0
14	0.87	3.45	1.25	1.92	0
15	0.61	4.05	1.87	2.05	0
16	9.96	4.57	0.53	1.73	1
17	1.38	5.61	1.77	3.62	0
18	1.65	5.1	1.77	3.16	0
19	1.22	5.71	1.53	3.93	1
20	1.65	5.24	1.47	3.41	1

In order to determine whether students suffer from hyperlipidemia, physical examination is conducted to measure the triglyceride, total cholesterol, high-density lipoprotein, and low-density lipoprotein. and other projects. Table 7.8 lists the results of 20 students under testing. Here, those students detected to have acquired hyperlipidemia are marked by a “1” and those who have not acquired hyperlipidemia by a “0” in the rightmost column.

By observing the sample data sets, we know all data have class labels, so this can be solved by a supervised classification method. Table 7.9 summarizes the memory demand, training time and accuracy measured in using the three candaidate machine learning methods. By accuracy demand, obvious KNN and SVM methods are perfect to serve the purpose. If memory demand and training time are important, the SVM method is even a better choice.

Table 7.9 Measured performance of three competing classifer choices

ML Algorithm	Memory Demand (in KB)	Training Time (in second)	Accuracy
Decision Tree	1,768	1.226	90%
KNN	556	0.741	100%
SVM	256	0.196	100%



Performance Analysis of Five Disease Detection Methods

Big data can be applied to predict whether a person is among the high-risk population of a certain chronic disease, based on their personal information such as age, gender, the prevalence of symptoms, medical history and living habits (e.g., smoking or not, etc.). Figure 7.11 lists 5 distinct machine learning methods, namely naive Bayesian (NB), k-nearest neighbor (KNN), SVM, neural network (NN) and decision trees (DT), that we evaluate for disease detection.

We apply the naive Bayesian (NB), k-nearest neighbor (KNN), SVM, artificial neural network (ANN), decision trees (DT) models to predict the risk of chronic disease. The model's basic framework is shown in Figure. We randomly divided the data into training data and test data, and the ratio of the training set and the testing set was 3:1. The method mentioned above was used to train the model.

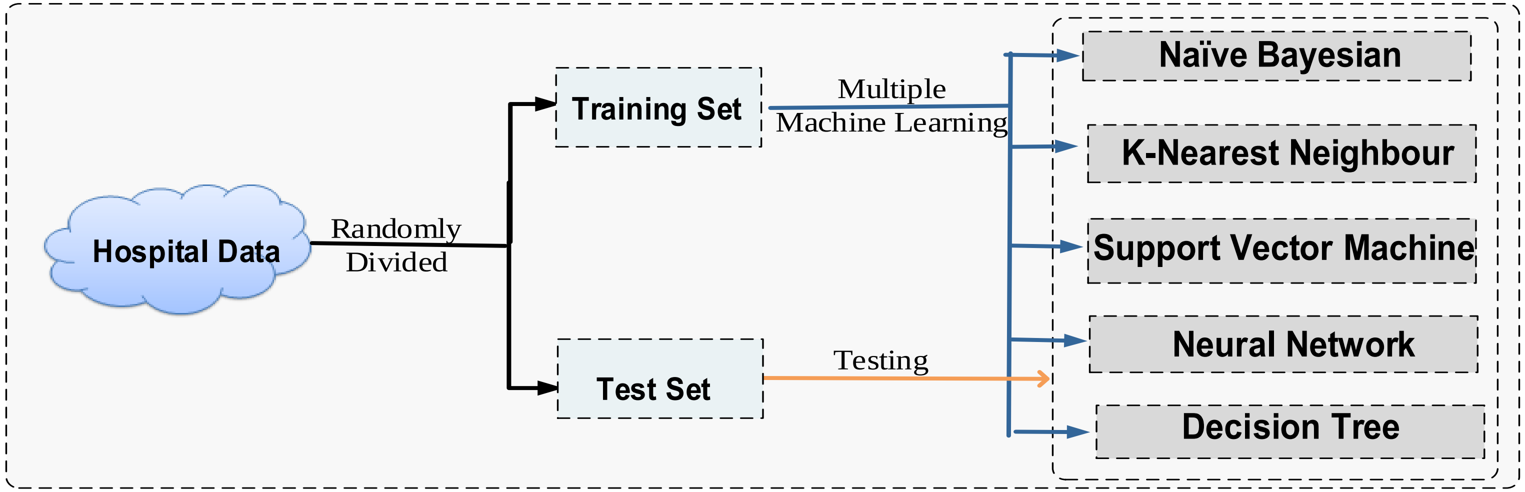
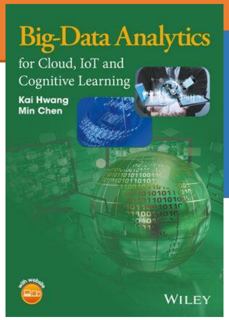


Figure 7.11 Five machine learning models for disease prediction based on medical big data



Performance Analysis of Five Disease Detection Methods

A. Prediction Using Nearest Neighbor Algorithm

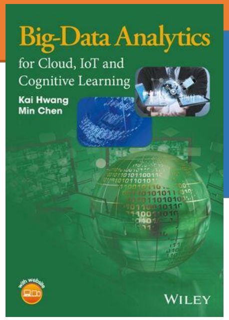
NB classification is a simple probabilistic classifier presented in Section 6.3.3. Based on a patient's input feature vector $\mathrm{x} = (x_1, x_2, \cdots, x_n)$, we can calculate $\mathrm{p}(\mathrm{x} \mid c_i)$ and the prior probability distribution $\mathrm{p}(c_i)$. Bayesian theorem, $\mathrm{p}(c_i|\mathrm{x}) = \frac{\mathrm{p}(c_i)\mathrm{p}(\mathrm{x} \mid c_i)}{\mathrm{p}(\mathrm{x})}$ is applied to obtain the posteriori probability distribution, $\mathrm{p}(c_i \mid \mathrm{x})$. Through solving the problem of $\mathrm{argmax}_{c_i} \mathrm{p}(c_i \mid \mathrm{x})$, the NB classifier can predict the disease of a patient.

B. Risk Prediction Using the Nearest Neighbor Algorithm

KNN was discussed in Section 4.3.2. In this example, we use Euclidean distance. Based on the medical big data, $\mathrm{x} = (x_1, x_2, \cdots, x_n)$ and $\mathrm{y} = (y_1, y_2, \cdots, y_n)$ are the characteristic vectors of two given patients, with each of the vectors containing n characteristics. The Euclidean distance between two patients is calculated as follows: $\mathrm{d}(\mathrm{x}, \mathrm{y}) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$. The parameter K is sensitive to the model performance. We choose from 5 to 25 in typical healthcare applications. For the dataset we used, when K = 10, the model exhibits the highest performance. Thus, we set K to 10.

C. Prediction Using Support Vector Machine

SVM was studied in Section 4.3.4. It is used to find a max hyperplane to divide an n-dimensional space into subspaces. In typical medical applications, the patient's characteristics vector $\mathrm{x} = (x_1, x_2, \cdots, x_n)$ is linearly inseparable. To map the data to a transformed feature space, using kernel-based learning, it is easier to classify the linear decision surfaces and, therefore, to reformulate the problem so that the data are mapped explicitly to this space. The kernel function can have many forms. Here, we use the radial basis function (RBF) kernel. The SVM classifier can be implemented using the LibSVM library.



Performance Analysis of Five Disease Detection Methods

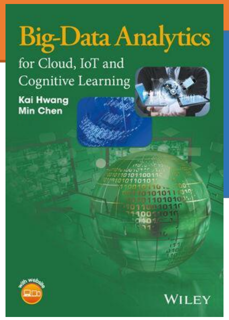
D. Prediction Using Neural Network

NN classifiers were invented by mimicking biological neural networks. In this example, we need to set parameters first: i) the number of layers. The NN model contains four layers generally, including an input layer, two hidden layers and an output layer and; ii) the number of neurons in each layer. Here, the dimension of the input layer is equal to the number of patient's characteristics. The input is denoted by $\mathrm{x} = (x_1, x_2, \cdots, x_n)$. In this example, we set 10 neurons in the first hidden layer, while setting 5 as the number of neurons in the second hidden layer. The output only has two results, i.e. high-risk or low-risk. Thus, the output layer only contains two neurons.

After constructing the structure of NN, we need to train the model. For each connection weight w and bias b in each layer, we use the back propagation algorithm. For the activation function, we apply the sigmoid function.

E. Prediction Using Decision Tree

Decision Trees (DT) based classification was presented Section 4.3.1. Its basic idea is that an object is classified by minimizing the data impurity, which is determined by the use of information gain. The information gain is based on the concept of entropy, whose definition is as follows: $\mathrm{H}(\mathrm{S}) = -\sum_i p_i \log p_i$, in which $p_i = |C_{i,s}| / |S|$ is the non-zero probability of C_i. The expected information required for the classification of S according to attribute A is denoted by $H_A(S)$. Then, we can obtain $H_A(S) = \sum_{v \in V} |S_v| / |S| H(S_v)$, where v represents the v subsets divided from S according to attribute A. We can then obtain the information gain as follows: $\mathrm{Gain}(\mathrm{S}, \mathrm{A}) = \mathrm{H}(\mathrm{S}) - H_A(S)$



Performance Analysis of Five Disease Detection Methods

E. Prediction Using Decision Tree

To improve the model, the 10-fold cross-validation method is used on the training set, where data from the testing participant are not used in the training phase. Let TP, FP, TN and FN be the true positive (the number of legitimate instances correctly predicted), false positive (the number of legitimate instances incorrectly predicted), true negative (the number of negative instances correctly predicted) and false negative (the number of negative instances incorrectly predicted), respectively. We define four measurements: accuracy, precision, recall and F1-Measure as follows:

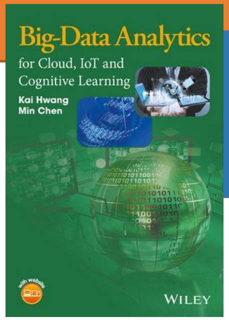
$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + \text{FN}}$$

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

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

$$\text{F1} - \text{Measure} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

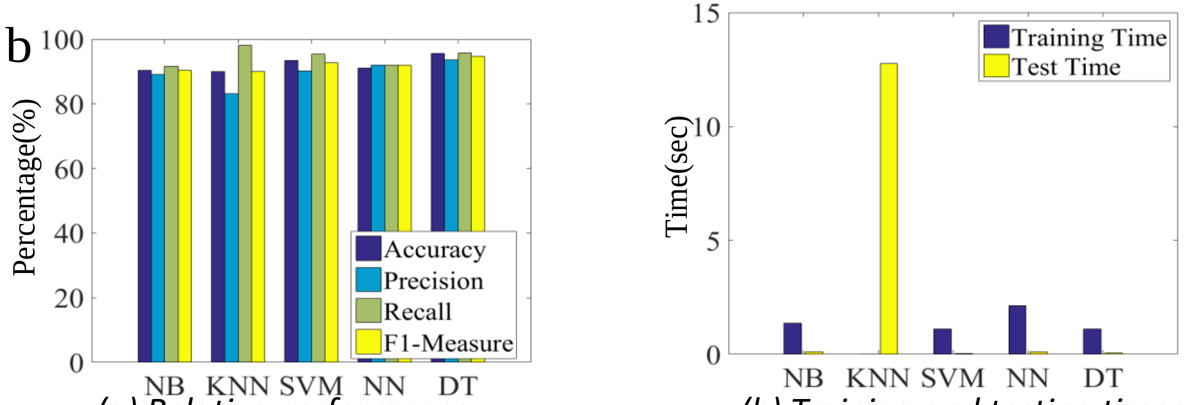
The F1-Measure is the weighted harmonic mean of the precision and recall and represents the overall performance. In addition to the evaluation criteria above, we most often use the receiver operating characteristic (ROC) curve and the area under the curve (AUC) to evaluate the pros and cons of the classifier. The ROC curve showed the trade-off between the true positive rate (TPR) and the false positive rate (FPR), in which $\text{TPR} = TP/(TP + FN)$, $\text{TFR} = \text{FP}/(\text{FP} + \text{TN})$.When the area is closer to 1, the better the model.



Big Data Analytics for Healthcare Applications

Example: Prediction of High-Risk Disease with Five Machine Learning Algorithms

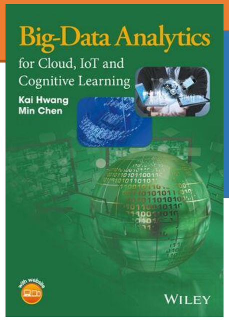
The inputs to the model are the attribute values of the patient, denoted by $x = (x_1, x_2, \cdots, x_n)$ The output value is $C = \{c_0, c_1\}$, where c_0 indicates whether the patient is amongst the hyperlipemia high-risk population class, and c_1 indicates whether the patient is amongst the hyperlipemia low-risk population class. We are concerned about the accuracy, precision, recall and F1-Measure of the hospital's data set. The DT had the highest accuracy in the training set and the test set. The relative performance and training time of five machine learning models are given in Fig. 7.12



(a) Relative performance *(b) Training and testing times*

Figure 7.12 Relative performance of 5 machine learning methods for disease prediction

Figure 7.12(a) plots the accuracy, precision, recall rate, and F1 performance of all 5 prediction methods. Based on the datasets we have processed, they all perform in the same range between 82% and 95%. Considering accuracy alone, the SVM and DT methods are higher at around 92%, while the other 3 methods stay at around 90%. By precision measurements, we find that NN and DT are better and KNN is lowest at around 80%. In terms of recall rate, the KNN method is the worst and the remainder is about at the same level above 90%. Finally, the DT has the highest F1 measure of 95%, while others stay around 90%.



Big Data Analytics for Healthcare Applications

Example: Prediction of High-Risk Disease with Five Machine Learning Algorithms

In summary, in terms of training time, as plotted in Figure 7.12(b), we find KNN takes a much longer time to be trained, while the rest have much lower training times. Based on these results, we rank the DT method as the highest in performance and the KNN method as the lowest in overall scores. However, we have to indicate that this ranking result is by no means the same in general situations. The relative performance is very sensitive to the dataset size and characteristics. By ROC results, we find that SVM exhibits high performance for high-dimensional cases, whereas the DT works better for low-dimensional cases (Figure 7.13). Finally, we summarize the pros and cons of using these five machine learning models in Table 7.10.

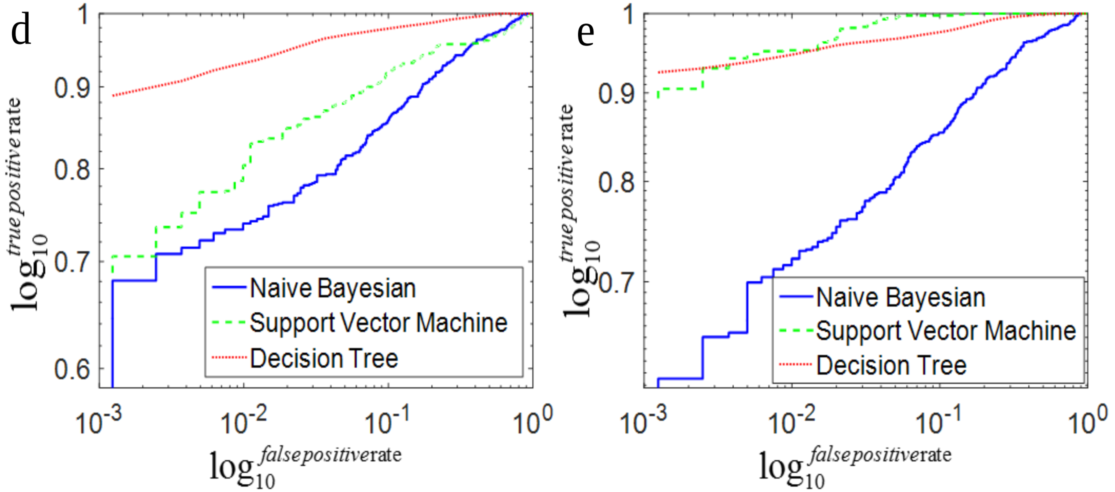
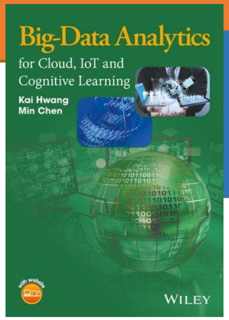


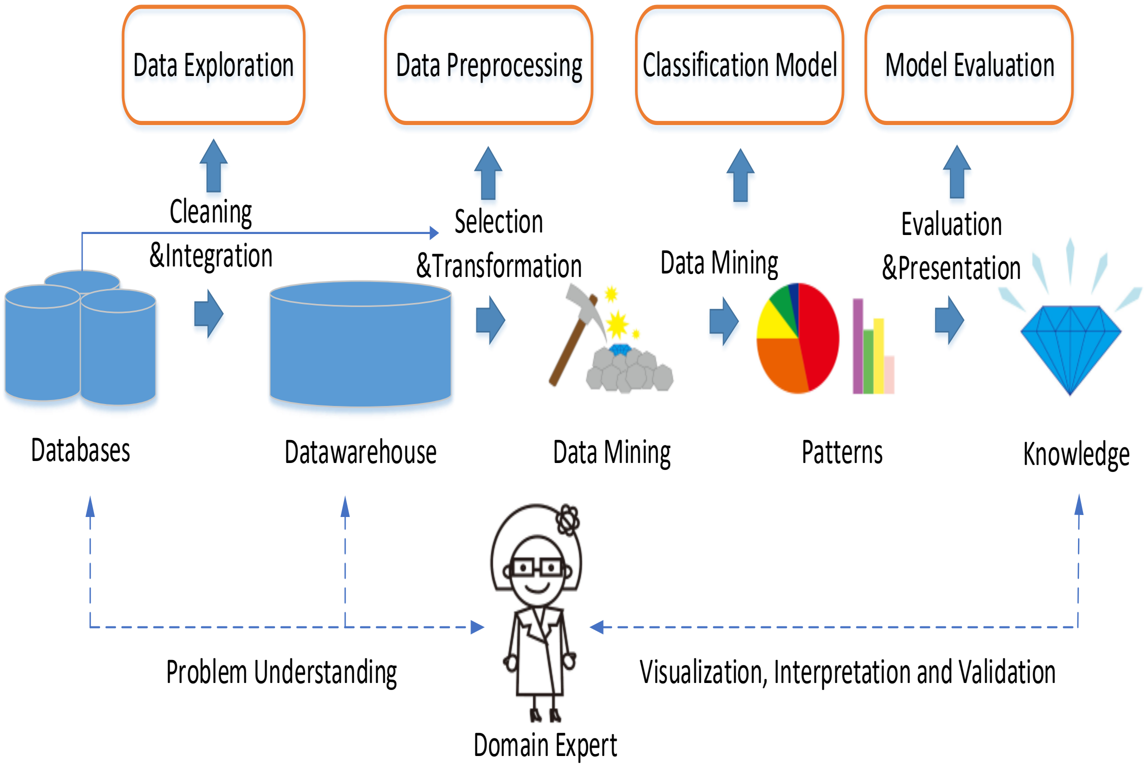
Figure 7.13 ROC curve of the disease prediction results using hospital data

Table 7.10 Strength and weakness of disease detection methods

Algorithm	Pros	Cons
Naïve Bayesian	Easy to implement; has strong robustness to the independent attribute and noise points; the training time is fast.	Attribute assumption of the data set occurs independently from one another; generally, the accuracy of the classification is not that high.
K-Nearest Neighbour	Easy to understand; there is no assumption about the distribution of the data set; the data can be multi-dimensional.	Classification speed is slow; all training sets are stored in memory and are faced with the problem of storage; sensitive to noise.
Support Vector Machine	Can handle high-dimensional data; generally, the accuracy is high; the abnormal value has good processing abilities.	With high dimension, it is necessary to choose a good kernel function; the time to train is longer; and demands for storage and CPU are both high.
Neural Network	Handle multiple feature data; classification speed is fast; it can address redundant characteristics.	Training time is relatively long; the training of the concentration noise is relatively sensitive.
Decision Tree	Has no potential distribution assumption for the data set; the classification of the data sets is fast; comparisons are easy to explain.	Is prone to the problem of data fragments; the best DT is difficult to identify.



Mobile Big Data for Disease Control

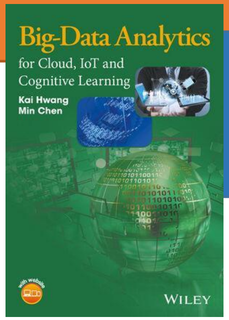


Methods for the high-risk patient prediction process encompassing exploration, preprocessing and evaluation stages

This study used general data mining, including data pre-processing, data mining models and data post-processing. Medical big data must be discussed with the doctor to obtain an understanding of the problem and the data. The hospital's data were stored in the cloud. To protect the user's privacy and security, we created a security access mechanism. We first pre-processed the data, including the processing and dimension reduction of missing values, repeat values and exception values. According to the doctor's opinion to extract feature values, we used machine learning algorithms to evaluate the patient's risk model; and finally, the best model was selected via evaluation using the mathematical method.

Aggregation was performed on the training data by implementing the demographics, risk factors, vulnerability on which the pre-processing was performed and transformation of the input data. Data cleaning included cleaning and pre-process the data by deciding which strategies to use to handle missing fields and to alter the data according to the requirements. We first identified uncertain, inaccurate, incomplete or unreasonable medical data and then modified or deleted them to improve the data quality.

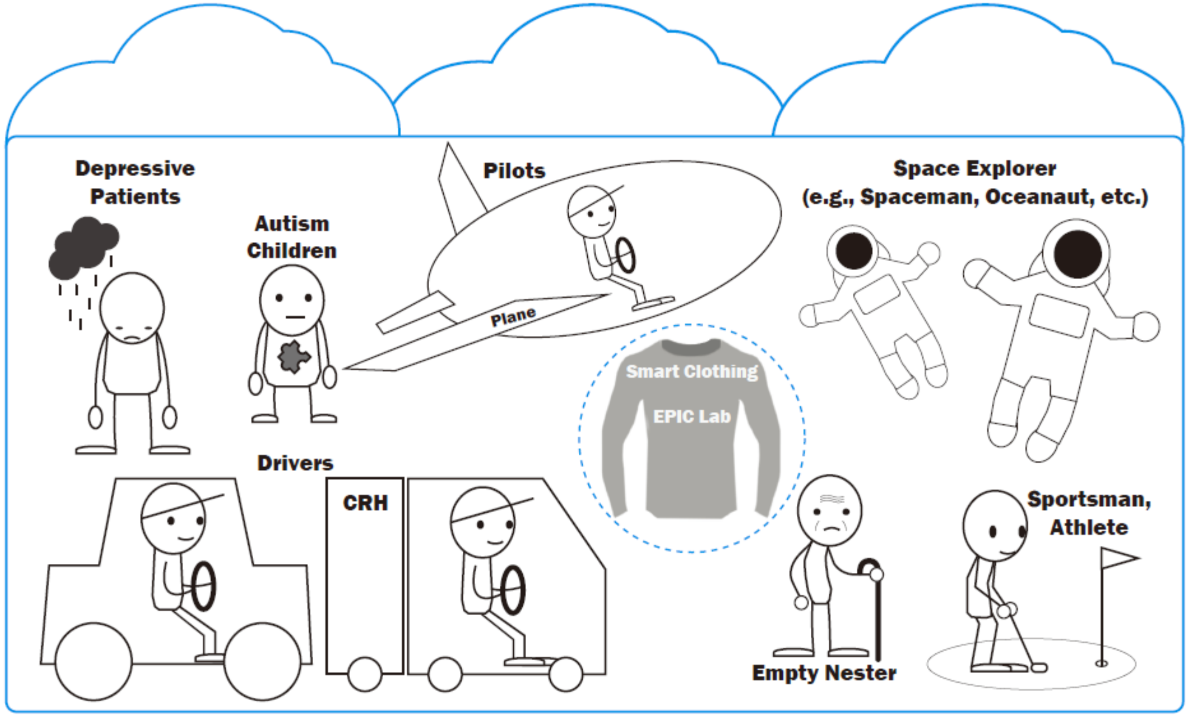
In the clean-up process, we examined the format, integrity, reasonableness and limitations of the data. Data cleaning is of vital importance for maintain the consistency and accuracy of the data analysis. The accuracy of risk prediction depends on the diversity feature of the hospital data. We can integrate the medical data to guarantee data atomicity, i.e., we integrated the height and weight to obtain BMI. According to the discussion with a domain expert and Pearson's correlation analyses, we extracted the user's statistical characteristics and some of the characteristics associated with hyperlipemia and living habits (such as smoking).



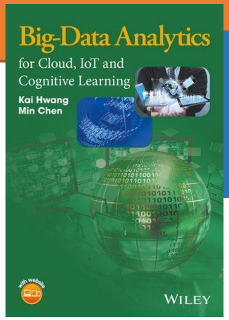
Emotion-Control Healthcare Applications

To provide proper and effective emotion care, we need to develop an emotion model based on physiological data training in the cloud. The system should establish unique responses for different user emotion patterns. For example, to use ECG (Electrocardiography), the signal is transmitted to the cloud via wisdom clothing with ECG acquisition and transmission function. When the cloud receives the ECG data, it will conduct analysis and processing in real time. Next, according to the user's unique identification, the user's emotional state is predicted by the trained model, while the other data collected from mobile terminal can assist emotion prediction.

When detecting that the user has negative emotions, an immediate call is made to the relevant equipment and resources to emotionally interact with users. For example, with a sadness emotion, in order to play music which can ease the grief for the user, the system can even send a command to a robot in the home and let the robot emotionally interact with user through a series of methods of actions, voice, etc. And finally, the system realizes the effect of emotional care. The population that needs emotional care includes empty-nest people, depressive patients, autism children, long-distance drivers, pilots and spacemen, prisoners or slaves, etc.



Mental healthcare for special groups of populations

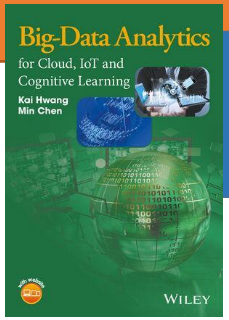


Data Collection and Feature Extraction

Wearable devices and mobile phones are used to collect data every 30 minutes. The data collected are then categorized into physical data, cyber data and social network data. Physical data consists of physiological data, activity level, location information, environmental, phone screen on/off and body videos. Cyber data includes phone call logs, SMS logs, emails logs and application usage logs. Social network data includes SNSs. On the other hand, the user's emotional status is obtained mainly through the following two methods: i) self-label by the user; and ii) label through transfer learning. Table 7.11 shows the data collection in detail. Data preprocessing mainly contains the following four aspects: data cleaning, eliminate redundancy, data integration and time series normalization.

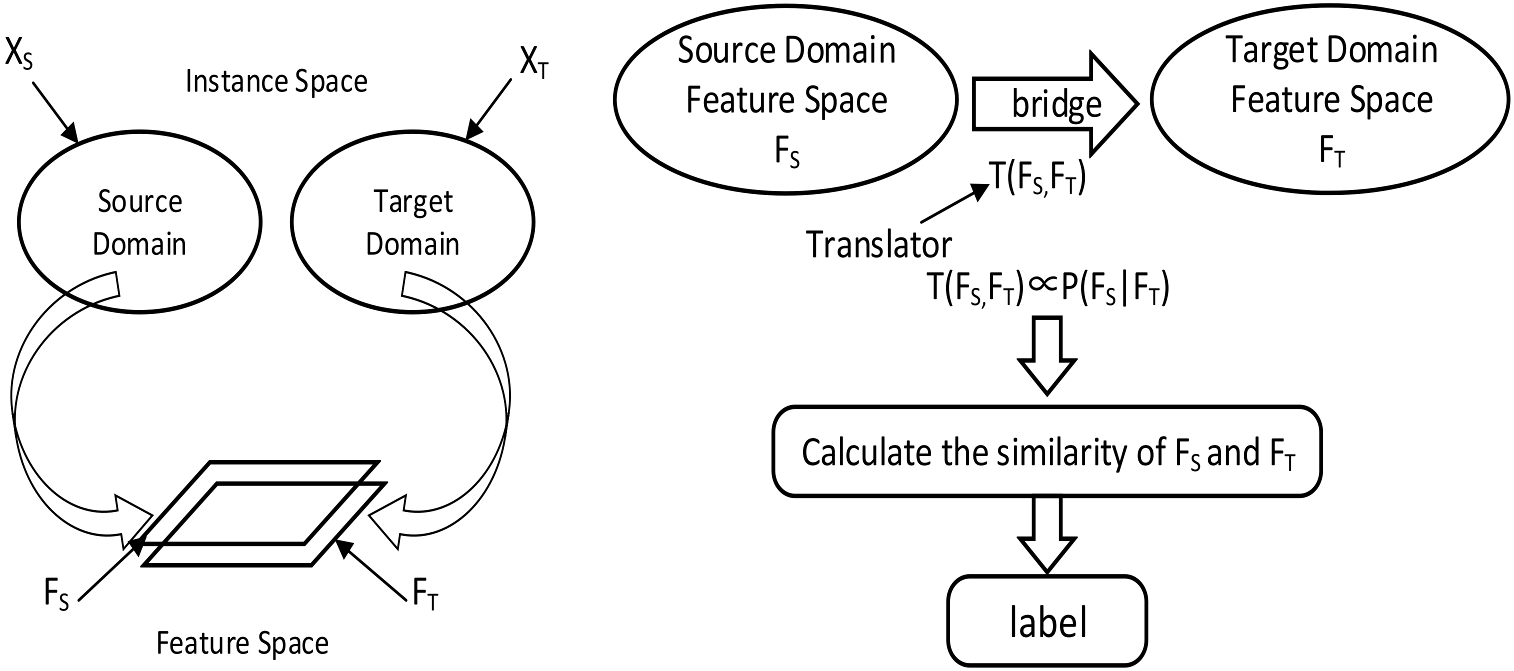
Table 7.11 Various data types in providing emotion control services

Data Style	Data Type	Usage Cue
Physical Data	Physiological data	Heart rate, Breathing rate, Skin temperature, Duration time of sleep
	Activity level	Static, Walking, Running
	Location	Latitude and longitude coordinates, User retention time
	Environmental	Temperature, Humidity
	Phone screen on/off	The time screen on/off
	Body video	Facial expression video, Head movement video, Eye blink video, Behavioral video
Cyber Data	Calls	No. of incoming/outgoing calls, Average duration of incoming/outgoing calls, No. of missed calls
	SMS	No. of sent/receive messages, The length of the messages, Content of each SMS
	Emails	No. of sent/receive emails,
	Application	No. of uses of Office Apps, No. of uses of Maps Apps, No. of uses of Games Apps, No. of uses of Chat Apps, No. of uses of Camera App, No. of uses of Video/Music Apps,
Social Network Data	SNS	The user ID and screen name, No. of friends, Content post, repost and comment, Image post, repost and comment, Content or Image create time



Transfer Learning based Labeling for Emotion Detection

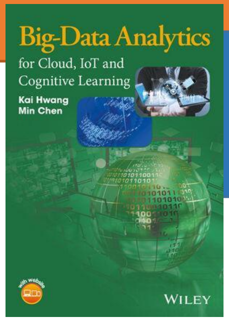
Typically, each person has his/her own behavioral pattern in terms of behaviors state and living habits, i.e. different people may have different physiological signals and living habits under the same emotions. As shown in Figure 7.16, various people express their emotion of happiness by difference behaviors, which can be sensed by multimodal person-centric data. One key penetrating point is to match a single type of emotion with various user's behaviors through transfer learning. The concept of transfer learning is illustrated below. Various data types are given in Table 7.11 in providing emotion control services.



(a) The instance space *(b) The concept of transfer learning*

Figure 7.16 Instance space and feature space for transfer machine learning

Let X_S be the source instance space, i.e. the data collected which have mood label, and let X_T be the target instance space, i.e. the data collected which does not have mood label. F_S and F_T are the feature spaces corresponding to X_S and X_T, respectively. As shown in the Figure 7.16(a), C denotes the label space of a number of emotional modes: {happiness, sadness, fear, anger, disgust, surprise}. The transfer learning model applies a Markov chain ($c \rightarrow f_s \rightarrow f_t \rightarrow x_t$), where $x_t \in X_T$, $f_t \in F_T$, $f_s \in F_S$ and c ∈ C.



Transfer Learning based Labeling for Emotion Detection

Our goal is to estimate the conditional probability $p(c \mid x_t)$. First, we need to find a translator $T(f_t, f_s) \propto p(f_t \mid f_s)$ to link the two feature spaces. The similarity of features is used to judge the similarity of feature domains. As shown in the Figure 7.16(b), we link the feature f_s and f_t through the following equation:

$$\mathrm{D}_{JS}(P_T \,\|\, P_S) = \frac{1}{2}(D_{KL}(P_T \,\|\, M) + D_{KL}(P_S \,\|\, M))$$

where $M = 1/2(P_S + P_T)$ and the KL-divergence DKL is defined as：

$$D_{KL}(P_T \,\|\, P_S) = \sum_{x \in X} P_T(x) \log \frac{P_T(x)}{P_S(x)}$$

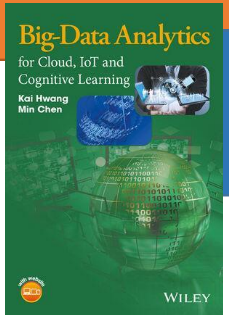
For time-series data, we first normalize them into [0, 1]. Time series collected from source domain and target domain are denoted by M_S and M_T, respectively. Dynamic time warping (DTW) is used to measure the similarity of M_S and M_T as

$$D(i, j) = d(m_i, n_j) + \min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\}$$

where $d(m_i, n_j) = \sqrt{(m_i - n_j)^2}$ $m_i \in M_S, n_j \in M_T$. With the decreasing of $D(i, j)$, M_S and M_T become more similar. So we take the low-n similar sequences out. Now we link the f_t and f_s, so we can calculate the top-N most probability sequences label.

For the text, we extract the word which scores between [−1,−0.4] ∪ [0.4, 1]. According to SentiWordNet, the vectors of scores from source and target domains are denoted by V_S and V_T. Now we use cosine similarity to measure the similarity between V_S and V_T as:

$$\cos(\theta) = \frac{\mathrm{V}_S \cdot \mathrm{V}_T}{\|\mathrm{V}_S\| \cdot \|\mathrm{V}_T\|}$$



Transfer Learning based Labeling for Emotion Detection

Now we link the f_s and f_t, so we can calculate the top-N most probability vector label. As shown in the Figure 7.17, the P_S is the probability distribution of f_s from the source domain, for example plot the frequency of body temperature value. For physiological data, call, SMS, email and application, we adopt the same method to estimated distribution. P_T is the probability distribution of f_t from the target domain, since the Jensen–Shannon divergence is widely used form measuring the similarity between two probability distributions. $D_{JS}(P_T \| P_S)$ equals to zero if and only if the two distributions P_T and P_S are identical. So we take the low-n similar distributions out. Now we link the f_t and f_s, so we can calculate the top-N most probability distributions label.

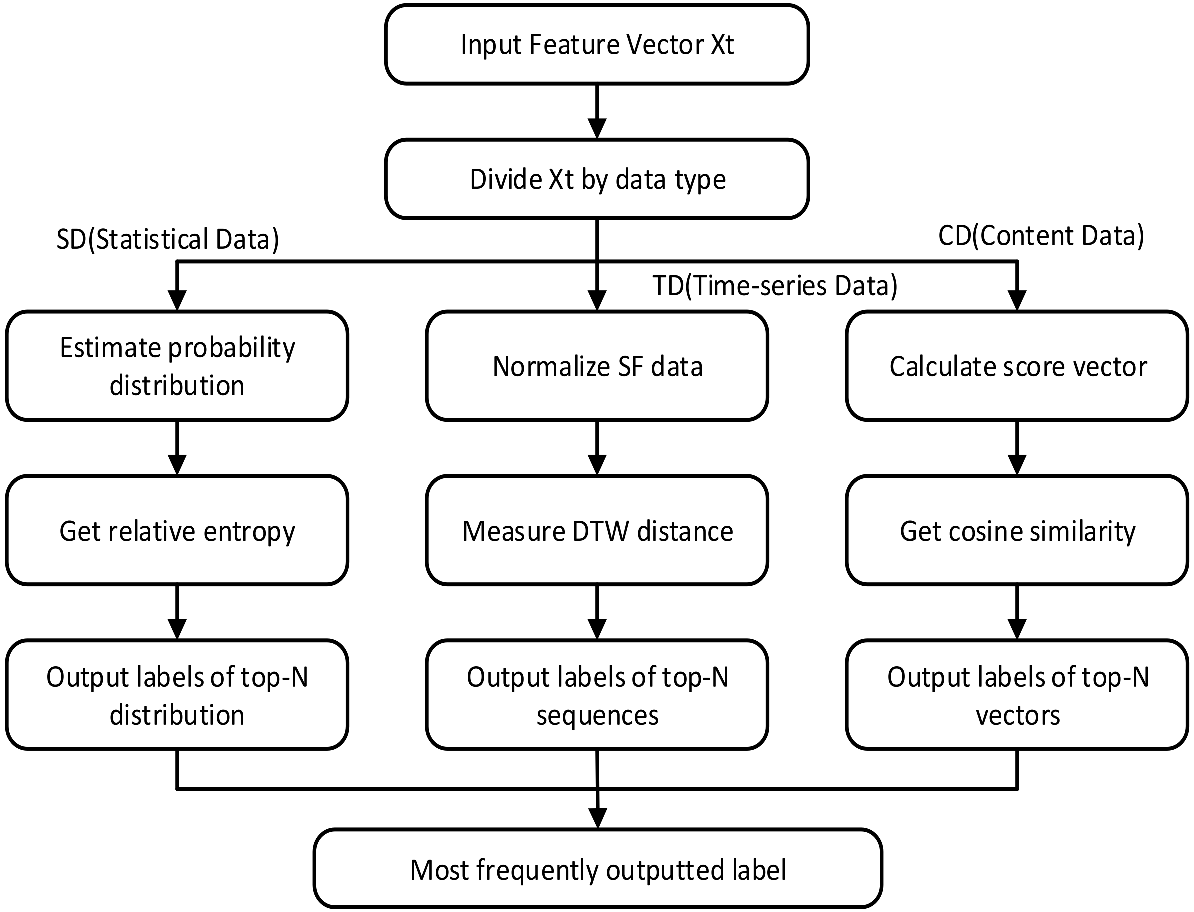
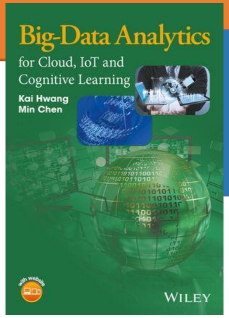


Figure 7.17 The concept of transfer learning for emotion labeling



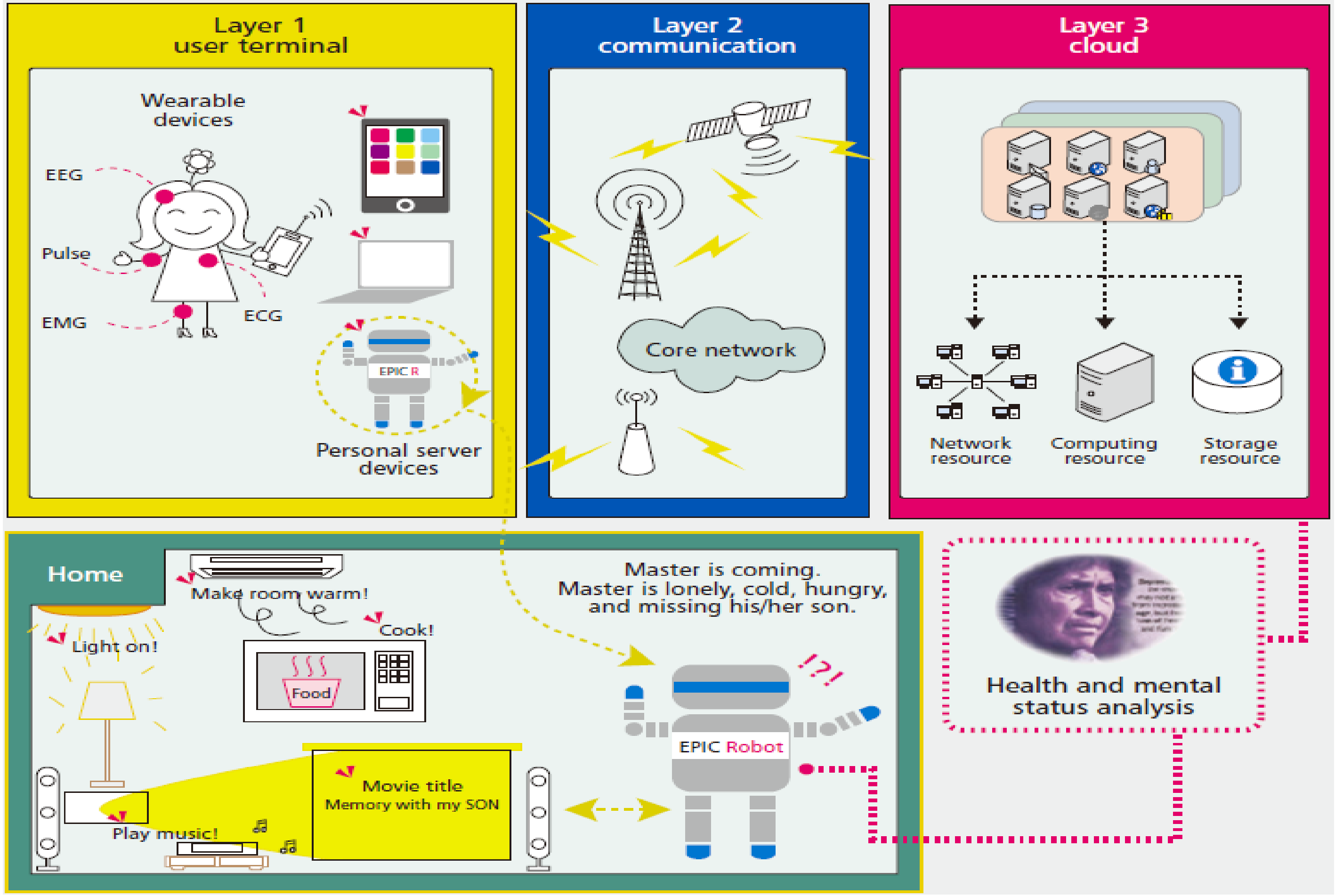
Emotion Interaction through IoT and Clouds

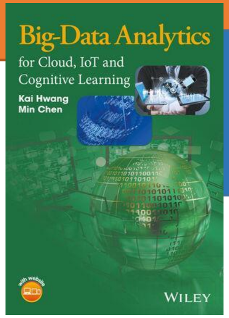
Example 8.7 The AIWAC emotion monitory system Developed at Huazhong University of Science and Technology

Traditional affective prediction results from analyzing one type of emotional data. This may lead to inaccuracy to validate the detection results. To overcome this difficulty, we present an emotion detection architecture, named AIWAC. AIWAC stands for Affective Interaction through Wearable Computing And Cloud. The system collects emotional data from multiple sources: namely the cyber, physical and social spaces. In the physical space, user's physiological data is collected, including various body signals, such as EEG, ECG, electromyography (EMG), blood pressure, blood oxygen.

In the cyber space, we use a computer to collect, store and transfer user's facial and/or behavioral video contents. In the social space, the user's profile, behavioral data and interactive social contents are extracted. With the availability of social networking services, IoT frameworks, and 4G/5G mobile networks, the affective data collected is truly a big data source over a long observation period. The AIWAC provide users with physiological and psychological healthcare support. AIWAC is developed in three layers: (1) user terminal layer with wearable devices for physiology data collection and emotional feedback; (2). communication layer; and (3) cloud layer for affective interaction.

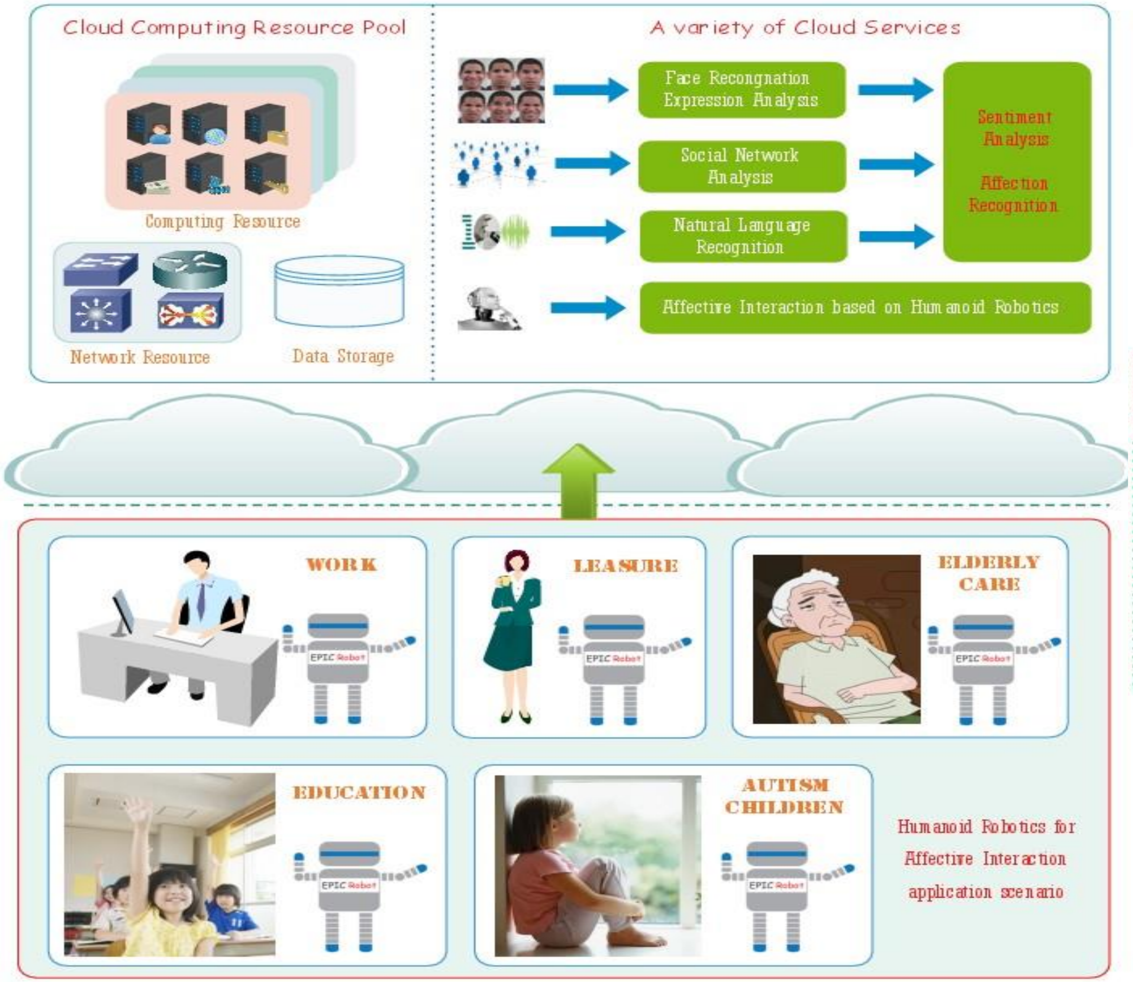
Layered architecture of the AIWAC emotion monitory system (reprinted with permission from Zhang et al., 2015)



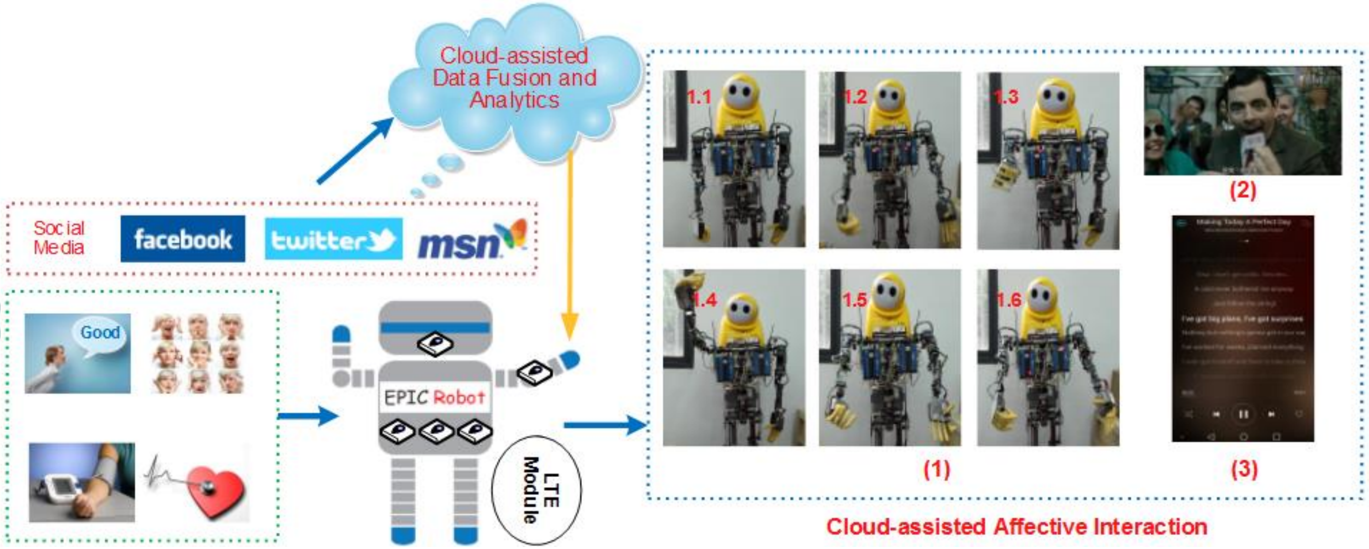


Emotion-Control via Robotics Technologies

Humanoid robotics for affective interactions between AIWAC and clients.

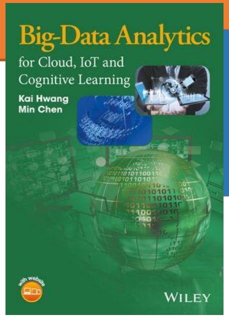


The humanoid robot has made great progress, but is also facing many technical challenges to make it fully integrated into human life, among which to equip the human robot with emotional interaction ability is one of the most challenging problems.



Robot affection interaction based on cloud computing

With cloud computing technology, there is no need for users to understand every detail of the cloud computing infrastructure, the corresponding professional knowledge or the direct control.



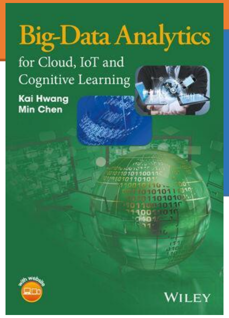
A 5G Cloud-Centric Healthcare System

The concept of the smart cognitive system is illustrated with the following features:

- ◆ Through 5G future telecommunication technologies, sensors, cognitive devices and robots interact smoothly with ultra-reliable low latency communications.
- ◆ The design of networking is enhanced so that it can move data quickly. For the retrieval or access of stored big data, 5G networks connect terminal devices and datacenters at a very fast rate, facilitating quick learning response.
- ◆ Learning from data is the heart of cognitive computing, and a cloud datacenter is the main hardware facility for advanced learning.
- ◆ Cognitive computing requires a wealth of data available, as clouds are implemented and configured to store and process those data.

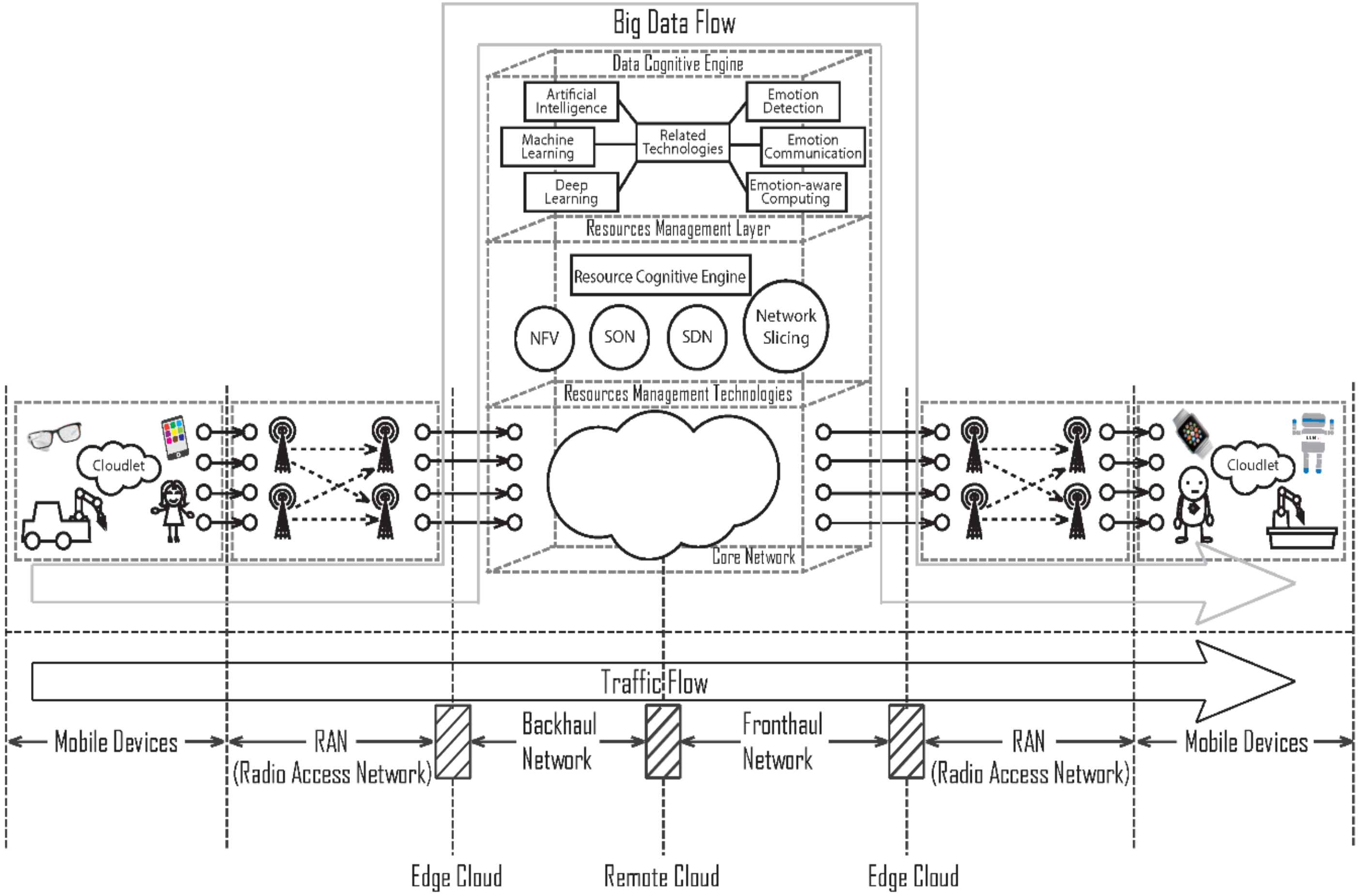
To build a smart cognitive system in the 5G era, the system needs to include three functional components:

1) Behavioral interaction terminal: cognitive behaviors in a cognitive system should be displayed in terminals; in order to achieve this, robots of varied types and increasingly powerful functions are favorable alternatives;
2) Environmental perception component: realization of cognition should be based on big data, and the cognitive component should realize the comprehensive perception of hearing, vision, touch and human emotion;
3) Cognitive reasoning component: the intelligent cognitive reasoning model can effectively simulate human cognitive process, and related technologies including AI, machine learning, deep learning, cloud computing and other effective tools utilized to establish cognitive reasoning model.



A 5G Cloud-Centric Healthcare System

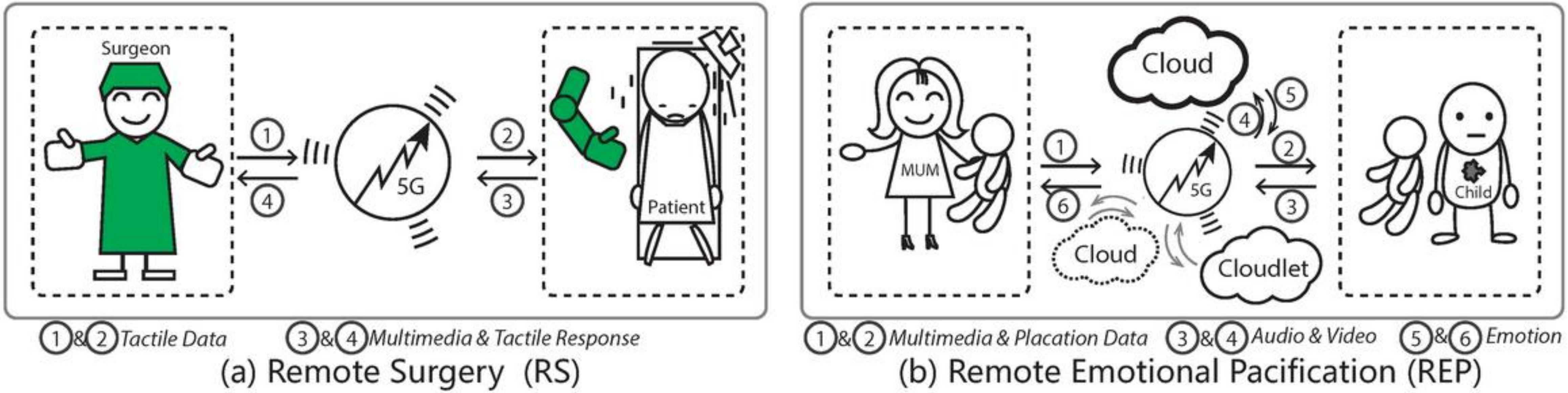
The system is divided into three layers: The first layer is built with smart terminals, cloud-based RAN and cloud-based Core Network. The heterogeneous access networks interconnect smart terminals, such as smart phones, smart watches, robots, smart cars and other devices. The edge cloud and remote cloud are the infrastructures to support the realization of cognitive functions in terms of storage and computing resources. The second layer is for resource management to support a resource cognitive engine to achieve resource optimization and high energy efficiency. The third layer provides data cognitive capability. In data cognitive engine, AI and big data learning techniques are employed for cognitive big data analytics, such as in the domain of healthcare. The big data flow represents the process of massive data collection, storage and analysis with the support of cloud or IoT. The traffic flow consists of packets and control messages during users' end-to-end communications.



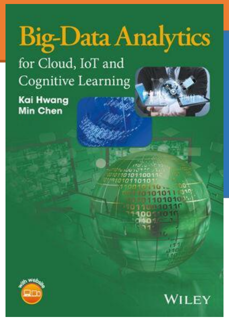
Architecture of a smart cloud/IoT/5G based cognitive system

A 5G Cloud-Centric Healthcare System

Here, we show two archetypal applications of the smart cognition system. In telemedicine, remote surgery may be designed to save life in the domain of healthcare. Using the 5G network, the critical operation action and haptic perception of the surgeon will be mapped to the robot arm in the remote operating table with very short delay and high reliability. In addition, all vital data of the patient can be processed with analytics tools at remote cloud in real time to guide the rescue team to carry out some preliminary life-saving operations before transporting the patient to the hospital. The second archetypal application is to detect human emotions with the help of smart robots, which interact with clouds to execute some responsive actions to calm patients. A lot of research experiments have been suggested in the past. The cloud/IoT based system may help to solve emotion control problems in the future.



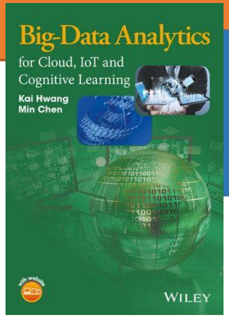
Two applications of a smart cognitive system



Conclusion

In this chapter, we have focused on big data application in the bio-medical and healthcare areas. However, the datasets we have tested in the illustrated example cases are not sufficiently large in scale to draw a general conclusion on TB or PB datasets. This chapter needs the background from previous chapters. Big data and clouds both demand a major overhaul of our educational programs in science and technology. There is no unique or general solution to big-data problems, due to heavy dependence on specific application domains.

We must leverage the use of clouds and big-data analytics in storing, processing and mining big data, which changes rapidly in time and space. The clouds, mobile, IoT and social networks are changing our world, reshaping human relations, promoting the global economy and triggering societal and political reforms on a world-scale. Those machine learning methods may perform differently, if non-medical or non-healthcare datasets are processed or tested. However, learning the machine learning methodologies is more important in general big data science and cloud computing applications.


Big-Data Analytics
for Cloud, IoT and
Cognitive Learning
Kai Hwang
Min Chen
WILEY

Thank You !