Fan Gongxiu Honors College

# Study on Epileptic EEG Signal Based on Multi-level Feature Extraction and SVM Classification

A Keystone Project

Final Report

Authors: Wu Tianhe, Feng Zijian, Yang Zhengwu

Supervised by: Duan Lijuan

Date: 2020.12.31

# Study on Epileptic EEG Signal Based on Multi-level Feature Extraction and SVM Classification

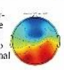
Supervisor: Prof. Duan Lijuan
*Wu Tianhe , Feng Zijian , Yang Zhengwu

## 1. SUMMARY

Epilepsy is a nervous system disorder in which the abnormal discharge of nerve cells causes the central nervous system to malfunction [1]. The main clinical symptoms are cognitive impairment, seizures, and involuntary seating characteristics of mental symptoms [2]. This project aims to classify and predict the EEG signals of patients with epilepsy by using mathematical, machine learning and computer programming related knowledge.

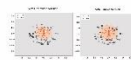
## 2. RESEARCH METHOD

01 Perform a window interception operation on the published data set for a certain period of time to obtain several EEG signal data.



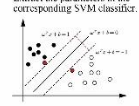
02 Use the mutual information method to extract features from the intercepted data, and obtain several 18 x 18 mutual information matrices.

03 Extract all the elements of the lower triangle in the mutual information matrix as feature data, and train the SVM classifier.



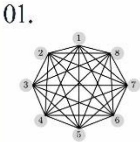
04 Take the method of (1)(2) on the test set data to test the trained SVM classifier.

05 Extract the parameters in the corresponding SVM classifier.



## 3. RESEARCH RESULTS

01.




Realized the algorithm of mutual information matrix.

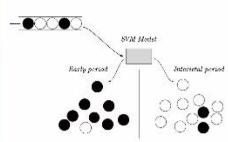
02.

Realized feature extraction of the data set.

03.

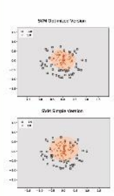
Achieved the classification of the interval and pre-seizure period.

04.



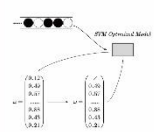

The project successfully used the feature extraction method, and the SVM classifier classified the EEG signals. After research, it is found that the ordinary SVM classifier can only achieve a classification effect of 90% accuracy, and the robustness of the model trained by the training set is poor.

05.



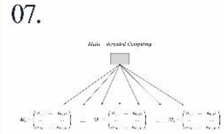
Implemented the algorithm improvement of SVM classifier.

06.




For the ordinary test set and the test set of special circumstances, it cannot be classified correctly. On the basis of this research, we optimized the algorithm of the SVM classifier model for each training, eliminating the dimension value that has the least weight on the model. A model with an accuracy of 92.16% is obtained.

07.



In order to improve the efficiency of calculating the feature matrix, we implemented Python multi-threaded calculation.

## 4. REFERENCES

[1] ORTEGA-MORENO L, GIRÁLDEZ BGSOTO-INSUGA V, et al. Molecular diagnosis of patients with epilepsy and developmental delay using a customized panel of epilepsy genes[J]. PLoS One, 2017.DOI: 10.1371/journal.pone.0188978.

[2] M. Kaleem, A. Guergachi and S. Krishnan, Patient-specific seizure detection in long-term EEG using wavelet decomposition[J]. Biomed Signal Proces.46(2018)157-165.

## 5. ACKNOWLEDGEMENT

Thank you Prof. Duan Lijuan for your support and help to our project. Throughout the process of the project, Prof. Duan has guided the process of the project many times, and also put forward reasonable suggestions for revision many times, and encouraged us. We hope that Prof. Duan is in good health and will have a smoother future work.

# 1 EXECUTIVE SUMMARY

Epilepsy is a nervous system disorder in which the abnormal discharge of nerve cells causes the central nervous system to malfunction. When the symptoms of epilepsy are severe, it can cause patients to lose consciousness or reason. The main clinical symptoms are cognitive impairment, seizures, and involuntary seating characteristics of mental symptoms [1]. The incidence of epilepsy is second only to cerebrovascular disease in neurology, with more than 70 million patients worldwide [2]. The diagnosis of epilepsy is currently done by medical workers based on experience to visually detect the patient's EEG. This work is not only very time-consuming but also affects the accuracy of the results.

During this project, we first obtained the epilepsy EEG signal dataset from the official website of Boston Children's Hospital. Then the original data is preprocessed into small-scale data through window interception. After that we use the method of mutual information in informatics to extract features of the data, and use multi-threading technology to improve the speed of calculation during the period. Bring the label and data into the SVM to train the model parameters, and use the score elimination algorithm to improve the model calculation speed. The final verification accuracy rate is as high as 92.16%.

Project results can detect the onset of epilepsy patients in time. Reduce the mortality and morbidity of patients with epilepsy, and greatly improve the health and quality of life of patients.

In the future, we may use neural networks as classifiers to classify epileptic EEG signals. At the same time, we will organize project results and apply for software copyright, while exploring more efficient and accurate algorithms.

# 2 BACKGROUND

Epilepsy is a common chronic brain neurological disease. The trigger for seizures is the sudden abnormal discharge of neuronal cells in the brain[3], which spreads to the surrounding normal tissues and causes recurring chronic brain diseases. The patient is having seizures. During the period, there are usually the following abnormal symptoms: convulsions, confusion, rigid short-term movements, sudden syncope etc. [4], the harm of epilepsy to patients is not only It is manifested in the physical and mental damage and memory loss to the patient, and it seriously affects the patient's daily life and even endangers life.

According to the World Health Organization (WHO) survey and research report, there are about 50 million epilepsy patients worldwide. There are about 9 million patients with epilepsy in our country, and the prevalence of epilepsy accounts for about 70% of the total population, of which about 6 million patients are patients with active epilepsy. At the same time, there are approximately 400,000 new epilepsy patients in China each year.

At present, epilepsy can be diagnosed clinically by magnetic resonance spectroscopy, transcranial magnetic stimulation, and EEG analysis. Among them, the EEG analysis method is low in cost, and the EEG examination can classify epilepsy without causing damage to the human body, and is a non-invasive biophysical examination method. The traditional clinical diagnosis is that the doctor performs manual diagnosis on the patient's EEG.

The scalp electroencephalogram (EEG) contains a large amount of physiological and pathological information and plays a very important role in the diagnosis of brain diseases such as epilepsy. At present, the clinical analysis of EEG signals is mainly based on the visual analysis of the clinician, which makes the clinician's task burdensome, and the analysis result has no quantitative standard. Therefore, the automatic classification of epilepsy EEG signals has great potential in current clinical applications.

However, due to individual differences, other brain diseases and neurological abnormalities caused by unknown causes, doctors' judgment based on their own clinical experience is prone to misdiagnosis [5]. Therefore, the automatic analysis and identification by the computer can effectively help doctors to accelerate the prediction and diagnosis of epilepsy diseases, and at the same time, it can improve the accuracy and efficiency.

In 1996, Gabor et al. Applied artificial neural network to the automatic detection of epilepsy for the first time, and proposed the design concept of "feature extraction" combined with "supervised classifier". Excellent feature extraction methods and classifiers have become the focus of inquiry .

Khan and Gotman use wavelet transform to obtain the time-frequency distribution of the signal, and then calculate the energy, coefficient of variation and relative amplitude of each scale signal as features [6]. Oweis and Abdulhay performed Empirical Mode Decomposition (EMD) on EEG signals, and then used weighted frequency to distinguish between seizure and non-seizure EEG for each eigenmode function (IMFs) after decomposition. [7].

Currently, support vector machines technology has been widely applied in machine learning, pattern recognition, pattern classification, industrial engineering, aerospace applications, etc., and its classification results impressive. Zhang Hu et al. [8] used the SVM classification method based on ensemble learning strategy to detect and recognize deceptive information in Chinese text. This is why we chose this classifier.

This project aims to classify and predict the EEG signals of patients with epilepsy by using mathematical, machine learning and computer programming related knowledge.

The research results can help to find the seizure period of patients with epilepsy in time, which has played a great role in the prevention of seizures, which can greatly improve the quality of life of patients and their families. At the same time, strengthening prevention and research on epilepsy has a vital social value and application prospects for improving China's population quality and health levels.

# 3 OBJECTIVES

1. Research the processing form and corresponding characteristics of public data sets.
2. Use specific methods to extract features in the data sets.
3. Analyze which features are useful for distinguishing epileptic EEG signals and which are meaningless.
4. Study the mathematical principle and implementation of SVM classifier.
5. Put the extracted features as a data set into the SVM classifier for training.
6. Put the test set into the trained SVM to get the accuracy of the SVM classifier.
7. Improved algorithm time complexity of SVM classifier.
8. Use Python multi-process to speed up the calculation of the feature matrix.

# 4 METHODOLOGY

At the beginning of the project, we read a lot of literature and finally determined the research plan of the project.

The data for our project comes from the data set of epilepsy EEG signals published by Boston Children's Hospital.

Boston Children's Hospital detected 18 parts of the human brain for 10 minutes, which corresponded to 18 channels. The EEG signals of the channels were transformed by discrete wavelet transform to obtain 256 data messages per second, each of which is 153600 data messages.

We use the window interception operation to perform segmentation and interception every 4 seconds to obtain 150 windows. Each window is regarded as a matrix, the number of rows of the matrix is 18, corresponding to 18 channels, and the number of columns is 1024, corresponding to the data extracted in 4 seconds.

After we divided the huge data set into a lot of small data, we chose the method of mutual information in informatics among the many feature extraction methods to perform feature extraction on the data.

We calculate the probability of each data of each channel, obtain the Shannon entropy of each channel, and calculate the joint probability density of every two data of every two channels, and calculate the joint Shannon entropy of every two channels Finally, we can obtain the mutual information of any two channels according to the mutual information definition of every two pieces of information, and construct the data structure of the adjacency matrix to store the mutual information value between each two channels, and this matrix is defined as mutual information The matrix is a matrix of size 18x18.

When we tested the project, the time complexity of our algorithm was too high. It took too long to calculate 150 matrices. So, we consulted a lot of academic materials and found that we can use the Python multi-process method to improve our calculation time for feature proof. So, we created 4 processes and let the computer calculate the matrices in these 150 windows concurrently. This modification greatly improves the time efficiency, from the original 25 hours of processing data to 6 hours.

After feature extraction, we constructed the SVM classifier model, obtained the lower triangular part of the mutual information matrix just processed, as each feature dimension in the classification, trained it, and finally used the test set to verify the accuracy of the model. At first, our classification accuracy was only 10.21%, which means that a certain part of the algorithm in our project has a problem. Later, it was discovered that there was a problem with the algorithm when we built the SVM model. Ideally, the data is linearly discretely distributed, and we can find a linear hyperplane to divide the data. In reality, the data is non-linearly distributed, and we cannot find a linear hyperplane to divide the data. Therefore, after consulting a lot of information. We have optimized the features of SVM classification again. When linearity is inseparable, we will upgrade the data. According to the data information of the existing dimensions, we will construct new dimensional data, and we can linearize it in the new dimension. Hyperplane division. There are many ways of constructing this kind of operation, among which the more perfect algorithm is the Gaussian kernel function. After modifying this algorithm, our project is basically finished.

# 5 KEY RESULTS AND ACHIEVEMENTS

## 5.1 KEY RESULTS AND DISCUSSIONS

The research results of our project team include the following contents:

1) Realized the algorithm of mutual information matrix.
2) Realized feature extraction of the data set.
3) Implemented the algorithm improvement of SVM classifier.
4) Achieved the classification of the interval and pre-seizure period.
5) The project successfully used the feature extraction method, and the SVM classifier classified the EEG signals. After research, it is found that the ordinary SVM classifier can only achieve a classification effect of 90% accuracy, and the robustness of the model trained by the training set is poor.
6) In order to improve the efficiency of calculating the feature matrix, we implemented Python multi-threaded calculation.
7) For the ordinary test set and the test set of special circumstances, it cannot be classified correctly. On the basis of this research, we optimized the algorithm of the SVM classifier model for each training, eliminating the dimension value that has the least weight on the model. A model with an accuracy of 92.16% is obtained.
8) Wrote the code of the project algorithm and optimized the algorithm.
9) Compared with other classification models, our own mathematical model is very robust.
10) The feature extraction algorithm has been improved, and the dimensionality in the classifier has been improved.

**Project planning**

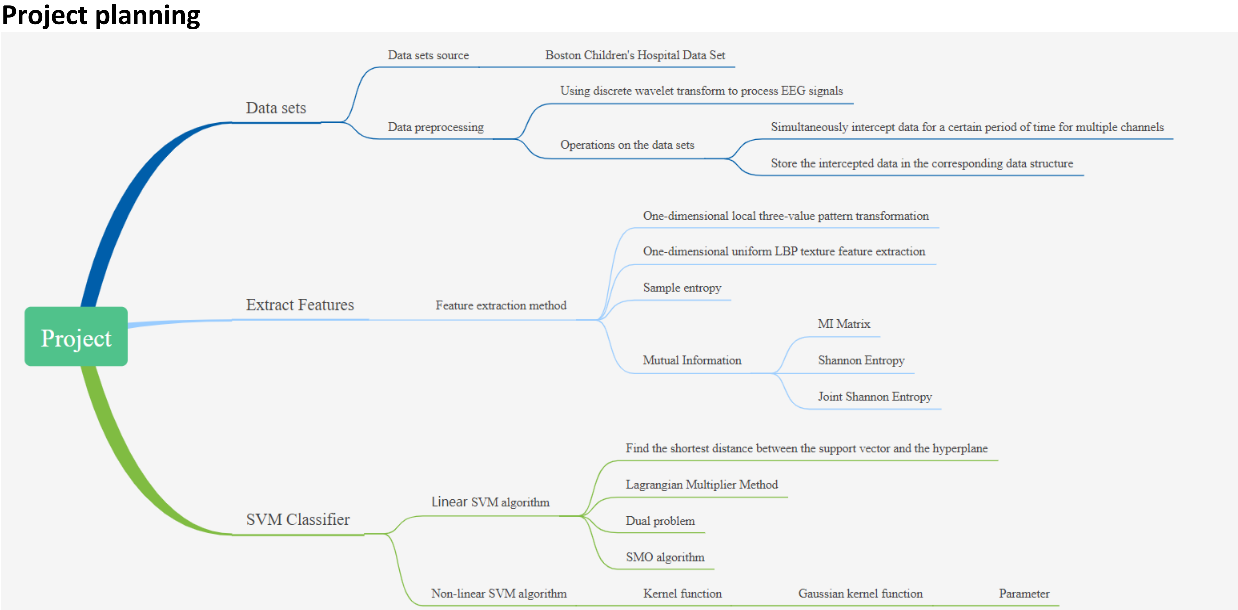



Figure 1 Project Planning

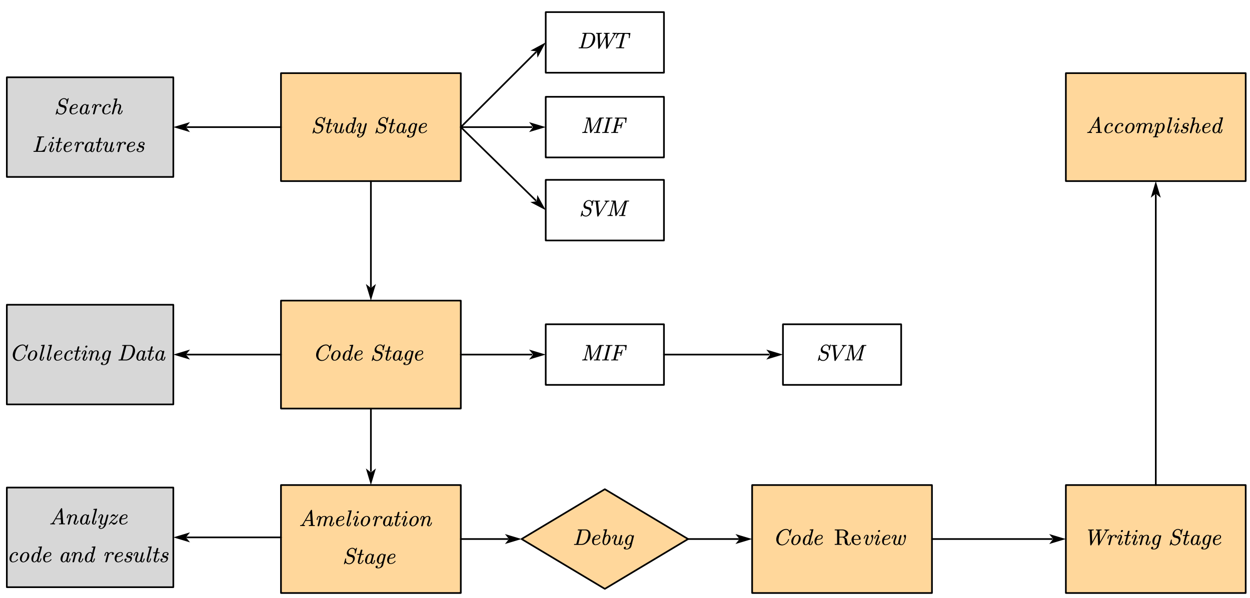
Our project experienced many difficulties in the early stage of the project. I don't know the mathematical principle of the discrete wavelet transform, nor whether the selected method is correct. The research on machine learning is not very in-depth. The lack of mathematics led to the wrong feature extraction method we chose at the beginning. Our project has not been advanced in the feature extraction part. Later, under the prompt of the teacher, we learned the method of mutual information.

In the part of preprocessing data, we thought of the window interception operation, and after discussion among the group members, we planned to use the Python programming language to implement the functions of our entire project. In the part of preprocessing data, our main idea is divide and conquer. Divide large-scale problems into several sub-problems, calculate and study them one by one. Solve the original problem.

In the middle of the project, we learned the basic mathematical principles of the SVM classifier during the epidemic, studied the relevant algorithms of the SVM classifier, and tried to write the prototype code of the SVM classifier. After comparing the realization of neural network, it is found that the realization of neural network is very complicated, and it cannot be mastered and realized in a short time. Therefore, we did not give up the implementation of the SVM classifier. Instead, it chose to optimize the SVM classifier algorithm to a certain extent.

In the later stage of the project, we integrate all the codes. Constructed the structure of the entire project system. The data is processed, and the data is divided into multiple dimensions for classification. In the end, the expectations of our project were realized, and the algorithm in the project was optimized, and the time complexity and space complexity of the algorithm were improved.

Our project stage:



## 5.2 THE ACHIEVEMENT AGAINST ACTIVITIES AND MILESTONES

| Scheduled date | Expected results | Participant |
|---|---|---|
| March20 to August31 | Learning EEG related knowledge | Tianhe Wu<br>Zhengwu Yang<br>Zijian Feng |
| September1 to September15 | Linear classification learning | Tianhe Wu<br>Zhengwu Yang<br>Zijian Feng |
| September 16 to September 30 | Maximized interval learning | Tianhe Wu<br>Zhengwu Yang<br>Zijian Feng |
| October 1 to October 12 | Solve the problem of convex quadratic programming | Tianhe Wu<br>Zhengwu Yang<br>Zijian Feng |
| October 12 to October 28 | Solve the dual problem | Tianhe Wu<br>Zhengwu Yang<br>Zijian Feng |
| October 29 to November 15 | Learn SMO algorithm | Zijian Feng<br>Zhengwu Yang |
| November 15 to November 30 | Understand the origin and use of kernel functions | Tianhe Wu<br>Zhengwu Yang<br>Zijian Feng |
| December | Realize SVM by applying High - level programming language | Tianhe Wu<br>Zhengwu Yang<br>Zijian Feng |

# 6 IMPACTS

## 6.1 SCIENTIFIC IMPACTS:

In the current research environment, there are few studies on EEG, especially for patients with epilepsy. In our keystone project, our research on EEG of epileptic patients not only combines medical knowledge, but also combines the knowledge of signal science and machine learning in the field of computer. Through these interdisciplinary links, we systematically studied the EEG signals of epileptic patients, and finally distinguished the disease period of epileptic patients by SVM classifier in machine learning. We believe that this will have a certain enlightening contribution to the field of EEG. At the end of our project, we will also put our report and code on GitHub, an open-source community, for others' reference.

## 6.2 COMMUNITY IMPACTS：

First of all, when our tutor gave us this subject about EEG signal processing, we understood that this is a scientific research project cooperating with Tongren Hospital, and our research results may be adopted by Tongren Hospital in the end. Secondly, there are as many as 40 million epileptic patients suffering from epilepsy in the world. Our project hopes to analyze the EEG signals of epileptic patients, and ultimately help doctors accurately understand the patient's condition, so that doctors can give patients the most correct medical treatment strategy.

# 7 CONCLUSIONS AND RECOMMENDATIONS

Although our project has been successfully implemented, there are still many shortcomings. For example, the computational efficiency of our algorithm is very low, which leads us to spend a long time in computing data. When we build the model, the code is very lengthy, and there is no good package management, which leads to our inconvenience in code review, as well as the lack of sufficient background research preparation in the early stage, which delays us a very long time in selecting methods. And the most important thing is that after communicating with Mr. Duan, we can use more popular neural network models to classify signals. But because we are only the third-year students, we do not have a good knowledge reserve to write a neural network which is in line with the theme of our project. So, when we apply for software copyright later, we will modify our project. When we classify EEG signals of epileptic patients, we will classify them by neural network. Through this project, we also learned that before carrying out scientific research projects in the future, we must learn more knowledge about the project related fields before carrying out the project, and make sufficient knowledge reserves, so as to calmly deal with the problems encountered in the project and carry out the project in an orderly way.

# 8 ACKNOWLEDGMENTS

Thank you Prof. Duan Lijuan for your support and help to our project. Throughout the process of the project, Prof. Duan has guided the process of the project many times, and also put forward reasonable suggestions for revision many times, and encouraged us. We hope that Prof. Duan is in good health and will have a smoother future work.

# 9 REFERENCES

[1] ORTEGA-MORENO L, GIRÁLDEZ BG， SOTO-INSUGA V, et al. Molecular diagnosis of patients with epilepsy and developmental delay using a customized panel of epilepsy genes[J]. PLoS One, 2017.DOI: 10.1371/journal.pone.0188978.

[2] Wang Ping,Yang Liming,Liao Hongmei,Hu Wenjing,Kang Qingyun.Characteristics of epileptic seizures in children with autoimmune encephalitis and analysis of antiepileptic drug treatment[J].Central South China Pharmaceutical,2019,17(12):2177-2180.

[3] Jiang Jun, Kuang Guangtao, Li Cheng, et al. Quantitative analysis and classification of EEG in typical absence epilepsy[J]. Journal of Huazhong University of Science and Technology: Medical Edition, 2016(3):315-318.

[4] Yan Gang. Research on Automatic Epilepsy Detection Algorithm Based on EEG Signal[D]. 2017.

[5] Pradhan N , Sadasivan P K , Arunodaya G R . Detection of Seizure Activity in EEG by an Artificial Neural Network: A Preliminary Study[J]. Comput Biomed Res, 1996, 29(4):303-313.

[6] Qu H, Gotman J. A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device[J]. IEEE transactions on biomedical engineering, 1997, 44(2): 115-122.

[7]Tzallas A T, Tsipouras M G, Tsalikakis D G, et al. Automated epileptic seizure detection methods: a review study[M]//Epilepsy-histological, electroencephalographic and psychological aspects. InTech, 2012.

[8] Subasi A . Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction[J]. Computers in Biology and Medicine, 2007, 37(2):227-244.

# 10 APPENDIXES

The results figure about our project.

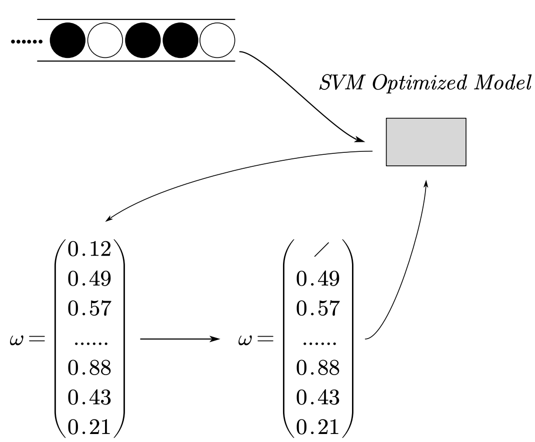



Figure 2 SVM-REF

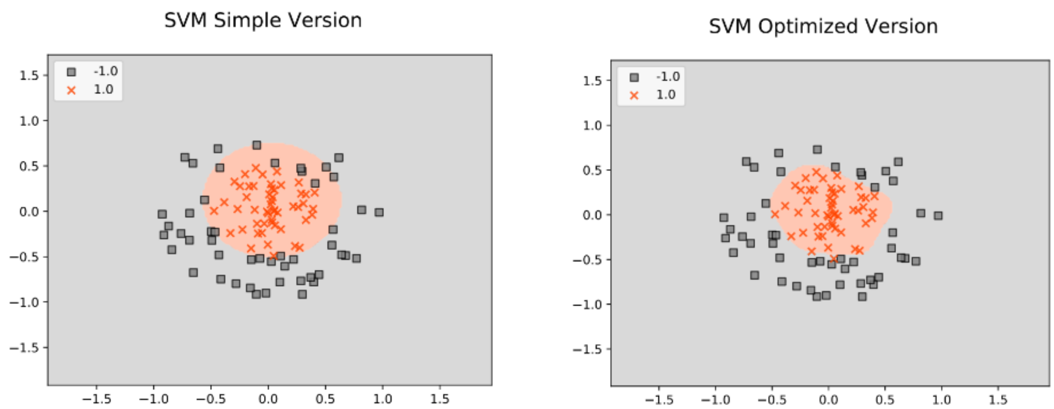



Figure 3 SVM Version

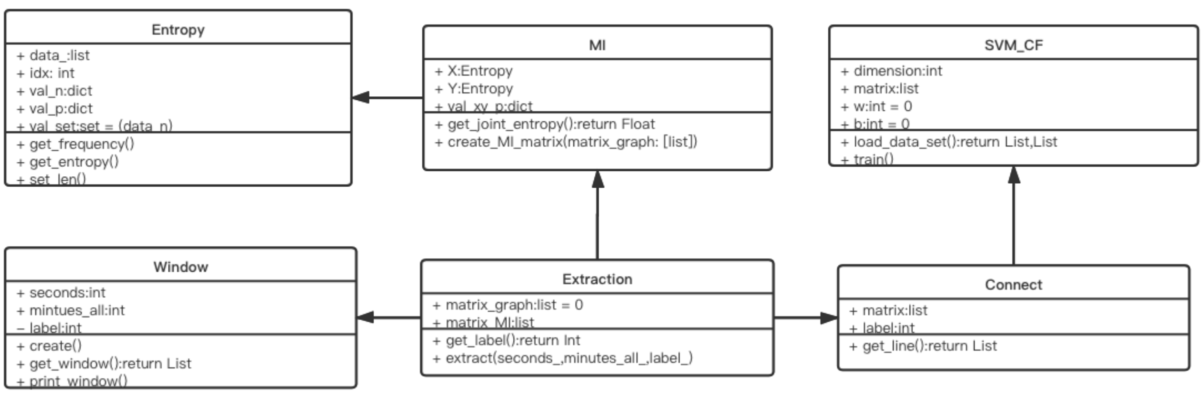



Figure 4 Project UML

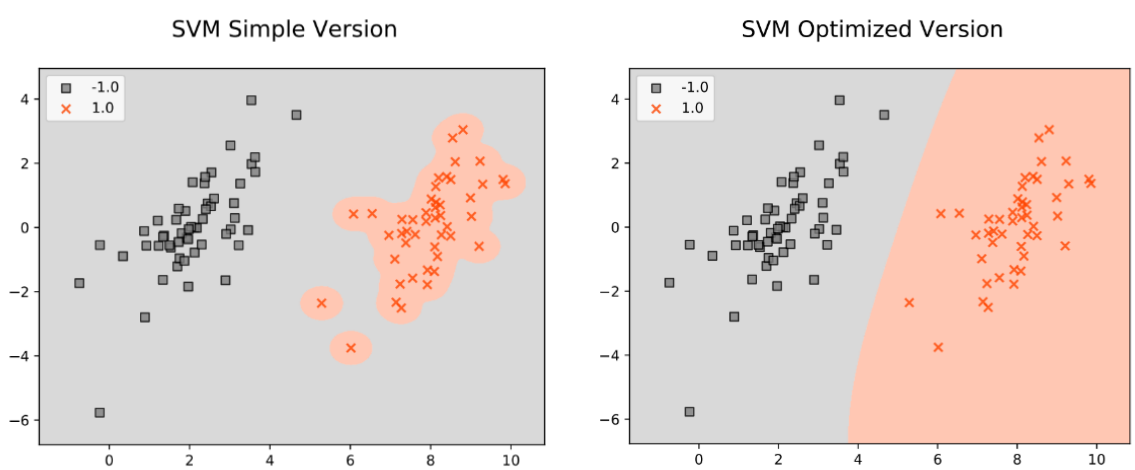



Figure 5 SVM Version

```
# Get the window_graph
def get_window(self) -> [[list]]:
   self.create()
   # Get the Data file
   file_object = open('D:\wutianhe_document\Keystone_Code\data2.txt',
encoding='utf-8')
   # Process the data
   for line in file_object.readlines():
      line = line.strip('\n').split('\t')
      # Deal the idx for every window
      window_pos = 0
      # Second list in window_graph
      idx_plus = 0
      # Create small window list
      cur_list = [ ]

      # Put each window (Two-dimensional matrix) into a big list
      # Each row in window graph means one window
      for i in range(len(line)):
         cur_list.append(line[i])
         window_pos += 1
         if window_pos >= self.window_num:
            self.window_graph[idx_plus].append(cur_list)
            idx_plus += 1
            window_pos = 0
            cur_list = [ ]

   return self.window_graph
```

Code 1 Window

```
class MI:

def __init__(self, X_: Entropy, Y_: Entropy):
        self.X = X_
        self.Y = Y_
        self.val_xy_p = {}

    def get_joint_entropy(self) -> float:
        print('start compute the joint entropy')
        joint_entropy_val = 0
        for x in self.X.val_set:
            for y in self.Y.val_set:
                p_num = len(np.where(np.in1d(np.where(self.X.data == x)[0],
np.where(self.Y.data == y)[0]) == True)[0])
                p_xy = p_num / self.X.len_data
                if (x, y) not in self.val_xy_p and p_xy > 0.0:
                    self.val_xy_p[(x, y)] = p_xy
        for key in self.val_xy_p:
            joint_entropy_val += self.val_xy_p[key] * math.log(self.val_xy_p[key])

        return -joint_entropy_val if joint_entropy_val != 0.0 else 0.0

    def create_MI_matrix(self, matrix_graph: [list]) -> None:
        print('start compute MI')
        MI_val = self.X.get_entropy() + self.Y.get_entropy() - self.get_joint_entropy()
        # MI_val = 0.0 if MI_val <= -0.5e-10 else MI_val
        matrix_graph[self.X.idx][self.Y.idx] = MI_val
        matrix_graph[self.Y.idx][self.X.idx] = MI_val
```

Code 2 MI

```python
class SVM_CF:

    def __init__(self, dimension_: int, matrix_: [list]) -> None:
        self.dimension = dimension_
        self.matrix = matrix_
        self.w = 0
        self.b = 0

    def load_data_set(self) -> list:
        label_mat = []
        data_mat = [[] for i in range(self.dimension)]
        for line in self.matrix:
            line_arr = line
            for i in range(self.dimension):
                data_mat[i].append(float(line_arr[i]))
            label_mat.append(float(line_arr[self.dimension]))
        data_mat_re = [[] for i in range(len(data_mat[0]))]
        for i in range(len(data_mat[0])):
            for j in range(self.dimension):
                data_mat_re[i].append(data_mat[j][i])
        return data_mat_re, label_mat

    def train(self) -> None:
        data_mat_, label_mat_ = self.load_data_set()
        svm = SVC(kernel='linear', random_state=1, gamma='auto', C=1.0)
        svm.fit(data_mat_, label_mat_)
        self.w = svm.coef_
        self.b = svm.intercept_
```

Code 3 SVM