Hand-written Digits Classification and Letter Recognition
Initial Report

Group 10
Sixiang Ma, Yuan Xiao, Xiaokuan Zhang

11/22/2016

Abstract

say something

1 Introduction

1.1 Problem Description

Our group wants to solve the problem of classifying hand-written digits and black-and-white rectangular pixel-displayed letters. We have two datasets. One is hand-written digits dataset [1], from a total of 43 people, 30 contributed to the training set and different 13 to the test set. Another is letter recognition dataset [2]. The character images were based on 20 different fonts. More details can be found by visiting the links.

1.2 Classifying Algorithms

In this report, we test three machine learning algorithms, i.e., Support Vector Machine (SVM), Neural Networks, and Naive Bayes. We will present the detail of algorithms in 3.

1.3 Result Summary

For the first dataset, after tuning the parameters, the three classifiers can achieve , , , accuracy.

2 Background

The hand-written digit dataset was first used in [5]. In the paper they combined different classifiers to obtain a better performance. The letter recognition dataset was first used in [3]. They generated classification rules to distinguish different letters.

3 Methodology

3.1 Support Vector Machine

3.2 Neural Networks

3.3 Naive Bayes

Naive Bayes classifier makes use of the Bayes Theorem with an assumption that all the features are independent. It is basically a conditional probability model. It is one of the simplest machine learning algorithms.

4 Evaluation

In this project, we mainly used `Weka` [4] to test different machine learning algorithms.

4.1 Hand-written Digits Classification

4.1.1 Support Vector Machine

4.1.2 Neural Networks

4.1.3 Naive Bayes

4.2 Letter Recognition

4.2.1 Support Vector Machine

4.2.2 Neural Networks

4.2.3 Naive Bayes

5 Discussion

6 Conclusion

References

[1] Optical recognition of handwritten digits data set. `http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits`.

[2] Optical recognition of handwritten digits data set. https://archive.ics.uci.edu/ml/datasets/Letter+Recognition.

[3] Peter W Frey and David J Slate. Letter recognition using holland-style adaptive classifiers. *Machine Learning*, 6:161, 1991.

[4] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[5] C Kaynak. Methods of combining multiple classifiers and their applications to handwritten digit recognition. Master's thesis, Institute of Graduate Studies in Science and Engineering, Bogazici University, 1995.