MPML REPORT

Qiu Wei

April 26, 2017

Contents

Abstract

This document describes the experiment result of python ver. Data analysis will be added soon.

1 Trainning Result

The trainning result is as following:
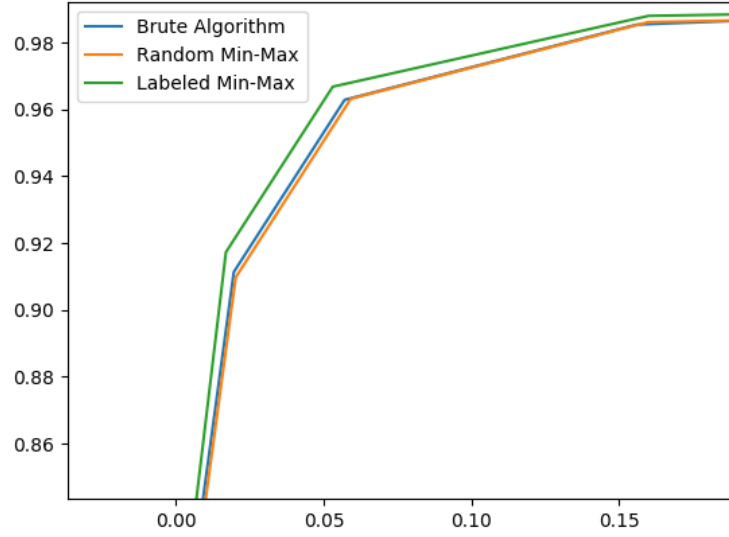
No.	Algorithm	?	Time/s	Accuracy	F1 value	AUC value
1	brute svm	\	34.04	96.37167%	0.92404	0.48782
2	random min-max	serialized	305.47	96.26846%	0.92194	\
3	random min-max	parallelized	140.08	96.26846%	0.92184	0.48307
4	labeled min-max	serialized	204.35	96.71042%	0.93101	\
5	labeled min-max	parallelized	204.35	96.69719%	0.93074	0.48567

The random min-max algorithm separate the input data into 5 parts randomly(5*5 models). The labeled min-max seperate the input data via the first two letters(4*12 models).

The total contains the time of load data, save model and other IO operations. Parsing is finished before the program runs.

The ROC Graph is as following:

The time cost between serialized min-max and parallelized min-max is:

No.	Parallelized?	Algorithm	Trainning time/s	Testing time/s	Total time/s
1	Yes	labeled	59.37346	144.54046	203.93383
2	No	labeled	115.96271	371.83840	489.00508
3	Yes	random	54.49303	84.16756	138.69087
4	No	random	103.98813	200.19163	305.47026

Test environment is Ubuntu, 4 kernal. Python version is 3.5.