Graduate School Application Evaluation Based on SVM

Wei Li, Yu Zhou, Yi Yang

Introduction

These days, more and more Chinese undergraduates consider pursuing a master's degree in the United States. But admission standards are distinct for each school and it is hard to decide which to apply. School selection alone has already become a big business in China. A list of graduate programs suitable for an applicant can be charged $70 dollars in China [1]In this paper, Support Vector Machine (SVM) and logistic regression are applied for analyzing historical admission data of each university to automatically predict the admission decision for Chinese applications.

SVM METHOD

Support Vector Machine learning algorithm is among the best supervised learning algorithm. It makes no strong assumption about the data and the performance is very good. The downside of it is computational inefficiency, but as is stated later, our sample size is small, which makes SVM the perfect algorithm here.

Problem Formulation

PhD's admission is mainly based on paper publication, research experience and recommendation letters, which are hard to quantize, so this project focuses on master's admission. Moreover, different majors have distinct standard so the data mu be separated by majors. Since the number of EE applicants is among the highest[2] in China, our project focus on prediction of EE master's application. The data was collected from famous realated BBS[3]. Since very few applicants are willing to share their background, only two hundred samples have been colleced, among which the largest data size for a single graduate school is 16. Nonetheless it turns out that even with very limited data the result is quite satisfactory.

Feature Sele]ction

The following popular six features of applicants ware used to train the SVM algorithm: Undergraduate Institute, GPA, GRE verbal score，GRE quantitative score，TOEFL score，and, number of paper publications.

All the applicants in our sample haven't change their major from undergraduate to master, and none of them have an advanced degree. SMO algorithm is applied to solve the constrained optimization problem through both linear and quadratic kernels. All the data have been preprocessed to the same scale before application.

Experiments and Results

Each university has its own admission standard. If the SVM is trained with all the data from all universities and then tested for a particular one, the prediction rate should be

low. TAMU is taken for example:

Kernel	Prediction rate
Linear	62.5%
Quadratic	56.25%

TABLE 1

As a result, it should make more sense to train and test on the same university. In this case, the training data is really scarce. As a result, leave-one-out cross validation has been proposed to test the prediction rate of the learning algorithm.

Here is the test of data from each university with reasonable sample size. All the six features have been used.

University	Sample Size	Prediction rate for Linear Kernel	Prediction rate for Quadratic Kernel
TAMU	16	81%	69%
Purdue	11	64%	46%
NCSU	7	71%	57%
UCSD	6	33%	33%
U Michigan	6	33%	33%
UCLA	6	33%	17%
UT Austin	5	100%	80%
U Pittsburgh	5	60%	60%

TABLE 2

Two important points could be observed from the results.

1. The prediction rate is relatively high for large sample size. As the sample size decreases, the prediction rate would drops down.
2. The prediction rate for linear kernel is better than the Prediction rate for quadratic kernel in all cases.

Since there are six features, which are comparable to some samples, the high prediction error probably results from high dimension of feature space. Forward selection procedure is used to add one feature at a time until reaching the best feature set. The prediction rate is as follows:

University	Sample Size	Prediction rate for Linear Kernel	Prediction rate for Quadratic Kernel
TAMU	16	88%	81%
Purdue	11	91%	73%
NCSU	7	86%	86%
UCSD	6	83%	83%
U Michigan	6	83%	83%
UCLA	6	67%	67%
UT Austin	5	100%	100%
U Pittsburgh	5	80%	80%

TABLE 3

TABLE 3 showed that the prediction rate is better with shrinked feature set. Even when the sample size is as small as 5 or 6, the prediction rate reaches an average of 80%. When the sample size goes above 10, the prediction rate achieves about 90%. Also, linear kernel outperforms quadratic kernel in all cases, which indicates the relationship among the factors could be considered linear. The result here is indeed practical and could serve as a useful indicative for graduate school selection.

Here is best feature set for each university with linear kernel: (1 indicates the feature is included and 0 indicates the it's excluded.)

University	Graduation University	GPA	GRE Quant	GRE Verbal	TOEFL	Paper
TAMU	1	1	1	0	0	0
Purdue	1	1	1	0	0	0
NCSU	0	1	0	0	0	0
UCSD	1	0	0	0	0	0
U Michigan	1	0	0	0	0	0
UCLA	0	1	0	0	0	0
UT Austin	0	1	1	1	0	0
U Pittsburgh	1	0	0	0	0	0

TABLE 4

Although the sample set is small, the results really make sense. It's well known that undergraduate institute and GPA are the most important factors to indicate an applicant's academic background. The main reasons are that undergraduate institute could reflect the overall academic standard of a student's education and GPA could measure his performance under such standard. In most universities' feature selection results, these two features are included. GRE quantitative score was also important for EE applicants since it is an indicator of one's quantitative skills. GRE verbal score and TOEFL score only indicate language skills which are less relevant to EE study so they are not included. Since most master applicants have no paper publications, it would be no surprise to exclude paper for either kernels

In the end is a graph indicating the effects of feature selection and kernel choices on the prediction rate:

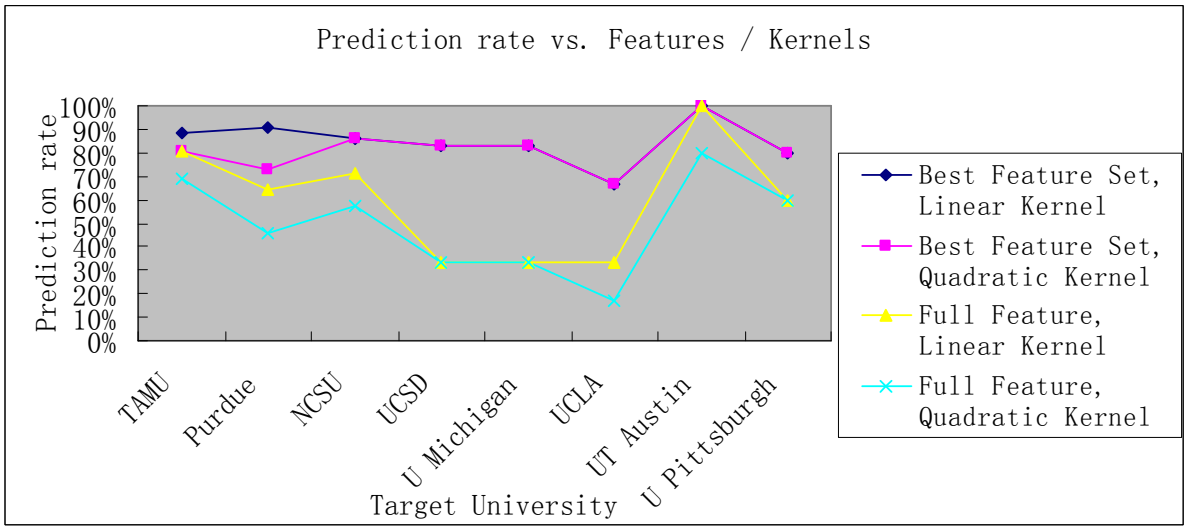

FIGURE 1

Comparison with Logistic Regression

Logistic regression is used in comparison to the SVM learning algorithm.

The learning rate was set 1, batch gradient ascent is used for data separated by schools and tested with 6 features. Here is the result:

University	Sample Size	Prediction Rate
TAMU	16	81%
Purdue	11	72%
NCSU	7	71%
U Michigan	6	83%
UCSD	6	67%
UCLA	6	33%
UT Austin	5	40%
Pittsburgh	5	80%

TABLE 5

It shows that as the sample size got smaller, the prediction rate becomes unstable.

When analyzing the result, some parameters of TOEFL are found negative. Thus, the EE departments might not care much about TOEFL score above minimum requirements. The result becomes much better when excluding TOEFL scores:

University	Sample Size	Prediction Rate
TAMU	16	87%
Purdue	11	91%

NCSU	7	71%
U Michigan	6	83%
UCSD	6	83%
UCLA	6	33%
UT Austin	5	60%
Pittsburgh	5	80%

TABLE 6

Here is a graph comparing the best prediction rate of SVM and logistic regression. The SVM method performs better than logistic regression method.

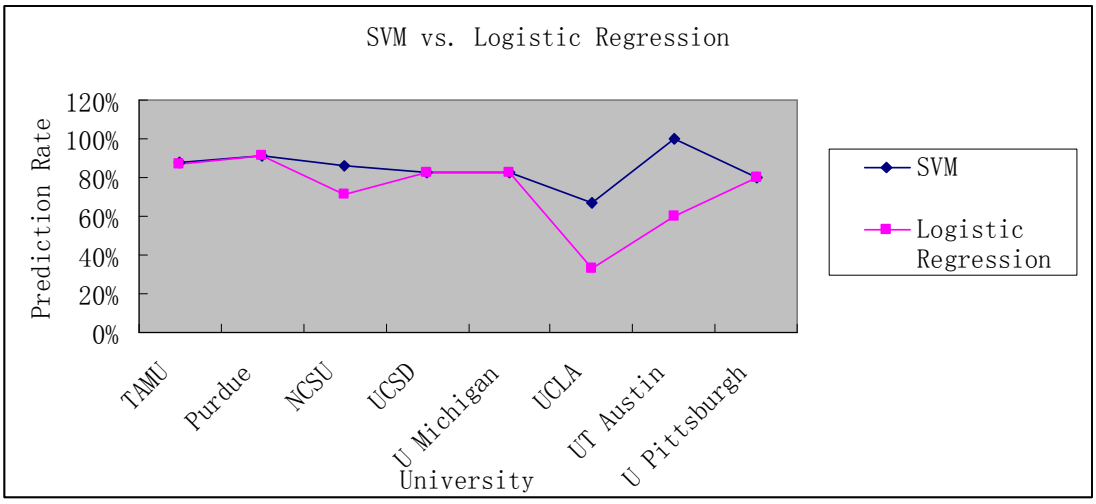

FIGURE 2

Conclusion

SVM with linear kernel can generate good result for graduate admission selection problem. The most significant features are undergraduate institute, GPA, and GRE quantitative scores, while GRE verbal scores, TOEFL, and paper are much less relevant. The only concern is that the data set is too limited. If we can get enough data from department admission office, the SVM should be able to achieve more accurate results. Finally, we compared the result with that of the logistic regression algorithm. And we can see that the SVM algorithm improves the prediction accuracy a lot.

Reference

1. Cloud Apply, http://www.cloudapply.com/Index.do
2. http://liuxue.eol.cn/shenqing_3325/20101029/t20101029_533791.shtml
3. http://bbs.gter.net/bbs/ http://bbs.taisha.org/
4. CS 229 lecture notes, http://www.stanford.edu/class/cs229/materials.html
5. Elements of Statistical Learning, second edition, http://www-stat.stanford.edu/~tibs/ElemStatLearn/