ATLANTIS PRESS

Advances in Computer Science Research, volume 77

8th International Conference on Management and Computer Science (ICMCS 2018)

Growth Enterprises Identification with Artificial Intelligence

Liang Wang[1, a]

[1]School of Economics, Shanghai University, Baoshan 200444, China

[a]dataframe@i.shu.edu.cn

Keywords: Artificial intelligence; Logistics industry; Growth potential; Random forest

Abstract. The rise of big data and artificial intelligence has revolutionized many industries, including logistics. From the long-term development of the enterprise, growth plays an important role. Based on the 222 observations from logistics enterprises, two kinds of strategies are adopted and machine learning algorithm models such as artificial neural network, support vector machine and random forest are employed. To sum up, on one hand, it is feasible to establish growth index by reducing dimension with principal component analysis then classify on growth financial indicators. On the other hand, by comparison, the random forest algorithm model can identify the growth state of the enterprise in accuracy.

Introduction

In recent years, with the increasing market demand, the logistics industry has also developed rapidly, and a large number of logistics enterprises have emerged [1,2]. These new enterprises whether can achieved rapid and healthy development in the future market, which should be taken into account by enterprise itself. Similarly, for enterprise investors, it is important to consider whether the enterprise is worth investing. That is to say, whether investors can obtain the positive net present value. One kind of the characteristics of enterprise which links the two events referred to former is growth. However, there does not cover research approaches on how to identify the growth of an enterprise, especially under the background of artificial intelligence [3].

Methodology

Independent and dependent variables. The datasets are financial data from the wind database about logistics enterprises ranging from 2013 to 2016.

Based on the growth financial indicators, including operating income year-on-year growth rate, operating profit year-on-year growth rate and net assets year-on-year growth rate, the growth index of enterprise is established as dependent variable. There are two strategies available:

a) if operating income year-on-year growth rate, operating profit year-on-year growth rate and net assets year-on-year growth rate are greater than zero simultaneously in the same year, thus call it growth enterprise and mark it discrete value 1; if operating income year-on-year growth rate, operating profit year-on-year growth rate and net assets year-on-year growth rate are smaller than zero simultaneously in the same year, thus call it recession enterprise and mark it discrete value -1; and in the rest of situation, call it maintenance enterprise and mark it discrete value 0;

b) first of all, principal component analysis is employed on operating income year-on-year growth rate, operating profit year-on-year growth rate and net assets year-on-year growth rate by the same year, and in further weighted principal component score is available, which is called integrated score, while the weight is the proportion of the characteristic value of the corresponding principal component and the sum of characteristic value. Thus, ordering the integrated score by descending, if an enterprise is in the top 20%, also call it growth enterprise and mark it discrete value 1; if an enterprise is in the bottom 20%, call it recession enterprise and mark it discrete value -1; and in the rest of situation, call it maintenance enterprise and mark it discrete value 0.

The independent variable is taken from the rest of financial indicators.

Copyright © 2018, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Training set and test set. The idea is to use the data from 2013 to 2015 as a training set to train the model and the data in 2016 is used as a test set for model validation. An instruction could be essential. Assuming that the current period is in 2015, thus characteristics of enterprises in current period and past two years are used to interpret the growth of enterprises to dig the factors hidden [4]. The correlation can be written as:

$$growth_{i,t} = X_{i,t}\beta_t + X_{i,t-1}\beta_{t-1} + X_{i,t-2}\beta_{t-2} + growth_{i,t-1} + growth_{i,t-2} + \varepsilon_{i,t} \quad (1)$$

While, $growth_{i,t}$ represents the growth of enterprise i in current period t, $growth_{i,t-1}$ and $growth_{i,t-2}$ represents the growth of enterprise i in the past two years $t-1, t-2$ respectively, $X_{i,t}$, $X_{i,t-1}$ and $X_{i,t-2}$ represents the characteristics of enterprise i in current period t and past two years $t-1, t-2$ respectively, β_t, β_{t-1} and β_{t-2} represents the influence of characteristics on growth of enterprise i in current period t and past two years $t-1, t-2$ respectively, $\varepsilon_{i,t}$ represents other factors that influence the growth of enterprise.

Obviously, here in training set, t is 2015, and in test set, t is 2016.

Training model. In this part**,** training set would be imported to algorithm models, including artificial neural network, support vector machine and random forest [5].

Model validation. The purpose of the model validation is to compare the advantages and disadvantages of different training models who are capable of identifying the growth of enterprise under different growth index definition strategies. The performance statistics used as model measurement include confusion matrix or Accuracy indicator, ROC curve or area under ROC curve.

Data

Growth index. Following two strategies on growth index definition, Table 1 and Table 2 display the distribution of growth enterprises ranging from 2013 to 2016.

Table 1 Distribution of enterprise state under strategy a)

Year	Growth	Maintenance	Recession
2013	86	131	5
2014	105	110	7
2015	98	114	10
2016	90	125	7

Generally speaking, the classification results are stable, and the number of growth enterprises is less than the maintenance enterprises, recession enterprises account for at least.

Strategy b) request to obtain weighted integrated score after dimension reduction by principal component analysis on three financial indicators. Above all, the number of principal component should be determined. Fig. 1 tells that two principal components have achieved requirement. And the distribution of enterprise state under strategy b)can be seen in Table 2.

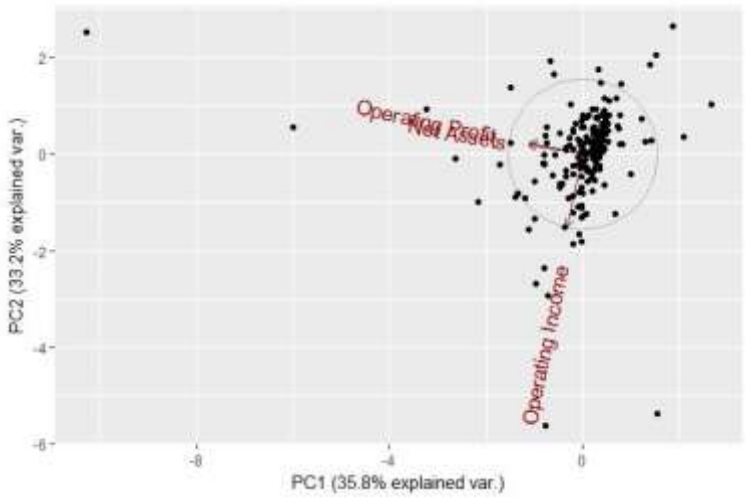

Fig. 1 Principal component and Constant density ellipse

Table 2 Distribution of enterprise state under strategy b)

Year	Growth	Maintenance	Recession
2013	45	132	45
2014	45	132	45
2015	45	132	45
2016	45	132	45

Data description. Taking 2015 as an example, the complete datasets cover variables: ROE, ROA**,** Operating Profit Ratio, Operating Profit Ratio, Accounts Receivable Turnover Ratio, Current Turnover Ratio, Debt Assets Ratio, Quick Ratio, Operating Income Growth Ratio, Operating Profit Growth Ratio, and Net Assets Grow Ratio.

Table 3 Data description

Variables (%)	Mean	S.D.	Min	Max
ROE	3.913	27.57	-324.2	35.29
ROA	6.086	6.736	-24.52	38.08
Operating Profit Ratio	11.43	21.85	-91.43	152.5
ART Ratio	85.36	887.8	0.785	13,126
Current Turnover Ratio	1.678	1.878	0.0574	16.79
Debt Assets Ratio	43.62	20.72	2.552	96.52
Quick Ratio	2.023	3.638	0.0794	38.47
Operating Income Growth Ratio	31.93	296.2	-81.06	4,336
Operating Profit Growth Ratio	-96.72	2,111	-25167	6,921
Net Assets Grow Ratio	73.41	480.8	-82.58	7,011

Results

Research results. As is described above in methodology, data from 2013 to 2015 is used as training set to train the model and the data in 2016 is used as a test set for model validation. Table 4 demonstrates the performance of different algorithm models under two kinds of strategies.

Table 4 Performance of different algorithm models under different strategies

	Strategy a)		Strategy b)	
	Accuracy	AUC	Accuracy	AUC
Artificial Neural Network	60.36%	63.59%	94.59%	97.63%
Support Vector Machine	**67.12%**	54.72%	95.05%	95.54%
Random Forest	65.32%	**65.16%**	**96.85%**	**98.13%**

As can be seen in Table 4, strategy b) is superior to strategy a) exactly. Furthermore, random forest gives splendid predicting performance with accuracy 96.85% and AUC 98.13%. Random forest actually is an algorithm that integrating a number of trees through the thought of ensemble learning. The basic unit itself is decision tree, and it in essence is an Ensemble Learning algorithm which is a branch of machine Learning.

Actual validation. Under strategy b), random forest algorithm is used to identify the growth of logistics enterprises. Among them, the enterprise with the most growth potential is the Yuan Tong Express (600233.SH).

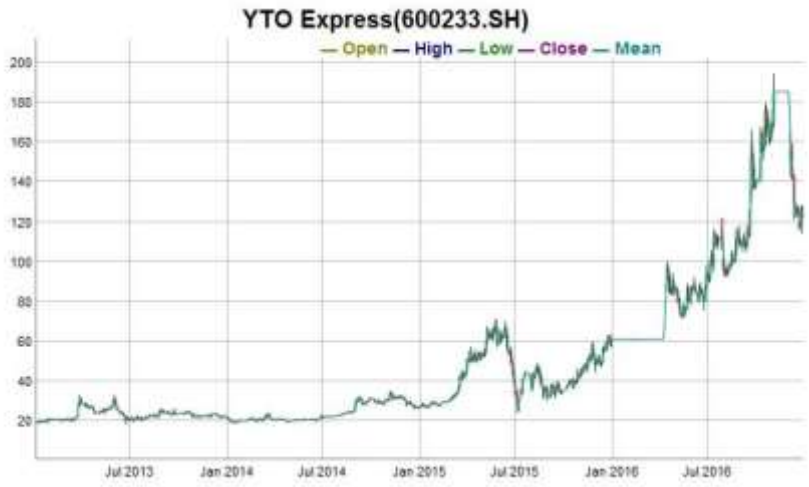

Fig. 2 Share price of Yuan Tong Express

From Fig. 2 it can be found that the actual situation of the Yuan Tong Express (600233.SH) is roughly in line with the results that the model predicts, which is very comforted.

Summary

Growth potential is the guarantee of healthy and sustainable development of logistics enterprises [6]. It is of great significance to identify the growth state of logistics enterprises in current period, whether for enterprises themselves or external investors. Our Research constructs the growth index and verifies it with the background of artificial intelligence. The results above show that it is feasible to establish growth index by reducing dimension with principal component analysis then classify on growth financial indicators. And the random forest algorithm model can identify the growth state of the enterprise in accuracy.

References

[1] Z.H. Jiang, J. Liu: *Contemporary Logistics in China* (Springer, German 2016), p.179.
[2] J. Xiao: *Contemporary Logistics in China* (Springer, Singapore 2017), p.139.
[3] L. Qi, T. Na: *International Conference on Mechatronics and Intelligent Robotics* (Kunming, China, May 20-21, 2017) Vol. 1, p.140.
[4] A. B. Bouazza, D. Ardjouman and O. Abada: American International journal of Social science,

Vol. 4 (2015) No.2, p.101.
[5] E. Alpaydin: *Introduction to Machine Learning* (MIT press, America 2014).
[6] J. Fernie, L. Sparks: *Logistics and retail management: emerging issues and new challenges in the retail supply chain* (Kogan page publishers, Australia 2014).