Research on the Growth Force of Listed Companies on GEM

Han-Qing HU
School of Economics and Management, Beijing University of Technology, China
huhanqing2013@163.com

Keywords: Innovative start-ups, Growth force, Data mining, Feature extraction, SVR.

Abstract. Innovative start-ups combine the characteristics of high growth and high risk, so the research of innovative start-ups growth force is of great significance. This paper takes Ningbo as example, selects a large number of the basic information of the financial indicators and other quantifiable as evaluation indexes, and reduces the dimension of index, determines the key indicators of growth force of different industries by using the feature extraction algorithm in data mining. At the same time, this paper establishes evaluation model of firm growth force by SVR regression algorithm, and evaluates the prediction effect of the model.

Introduction

As the most active part of national economy, enterprise plays a very important role in economic and social development. Economy and society cannot develop without healthy growth of enterprise. Evaluation and analysis of the growth of enterprise is particularly important. The growth of enterprise is an interactive process of scale expansion and structure upgrading. The growths of enterprise include expanding of enterprise’s organization, increasing in market share and improved business managerial practice.

The potentials and creativities of innovative startups are far more potential than other startup; meanwhile they tend to die early. According to figures from the U.S. Bureau of Labor Statistics only 26.8% innovative startup have survived since foundation in 2001 to 2015. With China’s reform reaching crucial period, more awareness about innovative startups’ growth is significantly important for deepening reform in all fields and improving the survival rate of startups.

Economic development requires new driving force. Now many innovative startups gradually grow into motivational driving force. China has been endeavoring to promote economic transition and innovative startup’s development. Innovation also gets the strong support of many regional governments and wide attention of foreign researchers.

Methods

Confirming the Indicators

Financial indicators are selected in general financial data based on simple and easy access principle. We will maximize the use of financial indicators for the follow-up analysis, since financial information of enterprise in all industries is the most complete.

Feature Extraction Based on SVM

After analyzing the target enterprise in profitability, the quality of assets, the debt risk, Operating growth and other ratio, we can get many indicators on different dimensions of
enterprise. But these indicators are not all necessarily effective in predicting the enterprise growth; we need to screen out the useful characteristic indexes from many indicators. Feature extraction can be seen as a kind of optimization, it is important to build an evaluation criterion to distinguish redundant or irrelevant feature combination.

In the framework of financial indicators, with total asset growth rate, net profit growth rate, revenue growth rate as the key indicator respectively evaluating enterprise growth, with SVM (Support Vector Machine) extracting key indicators character, a model flow chart can be build.

Support vector machine (SVM) is a new machine learning method based on statistical theory proposed by Vapnik in 1990s, it is suitable to solve the problem of machine learning with nonlinear high dimensional features, so it can be extended to a variety of linear and nonlinear function simulation. SVM has been applied in regression field that is developed based on classification support vector machine; it keeps the merits of good promotion ability of classification support vector machine and is often used in regression prediction.

There are many feature selection based on SVM. SVM-RFE (Recursive Feature Elimination based on Support Vector Machine) method, also known as SVM based regression feature elimination method is used in this paper. Just as its name implied, it uses SVM as a classifier in the process RFE, sets the standard for feature ordering based on the various parameters of the trained SVM.

SVM-RFE algorithm has priority to keep optimal feature subset in the process of feature ordering applying RFE algorithm, it sorts of features by discriminating information in a function with SVM, the weight of classifier of it is used as the feature ranking criteria of RFE.

SVM-RFE is an iterative process of backward elimination, each step in the iterative process includes the following three steps:

1. Using the current data set to train the classifier to obtain the information about the feature of the classifier.
2. Calculating the standard values of all features, such as the cost function, according to the predefined evaluation criteria.
3. Removing the features of minimum ranking criteria in the current data set.

The features of minimum ranking criteria will be removed in each cycle, and then SVM will retrain the remaining features to get a new feature ordering. The SVM-RFE method iteratively performs the process, and finally gets a list of a feature ranking. The ranking list defining a number of nested feature subsets is used to train SVM. The correct rate of prediction of SVM is used to evaluate the pros and cons of the subsets to obtain the optimal feature subset.

**Enterprise Evaluation Based on SVR**

According to the above feature extraction method, we can screen out the most effective indicators of the evaluation of enterprise growth from the complex index. A support vector regression model with parameter constraints could be established, key indicators of enterprise growth after reducing dimension as the independent variable, the first principal component of total asset growth rate, net profit growth rate and revenue growth rate as the dependent variable. At the same time, inputting above financial data in the evaluation model, predicting next year, finally the prediction result of enterprise growth could be obtained.

The regression method used here is based on the support vector regression of SVM. SVR is the application of support vector in the field of function regression. The specific algorithm is as follows:
For classification, the support vector machine is equivalent to that the tutorial samples is a finite set. Considering tutorial set is an uncountable set, the training set can be:

\[ S = \{(x_i, y_i), \ldots, (x_i, y_j) \mid x_i \in \mathbb{R}^n, \ y_j \in \mathbb{R}\} \]

So it becomes support vector regression concept.

Support vector regression is also divided into linear regression and nonlinear regression:

**Linear Regression.** For a given set of samples \( S \), and any given \( \varepsilon > 0 \), if there is a hyperplane in the original space \( \mathbb{R}^n \), \( f(x) = \langle w, x \rangle + b \), \( w \in \mathbb{R}^n, \ b \in \mathbb{R} \), makes \( |y_i - f(x_i)| \leq \varepsilon, \ \forall (x_i, y_i) \in S \), then \( f(x) = \langle w, x \rangle + b \) is \( \varepsilon \)-linear regression of sample set \( S \).

**Nonlinear Regression.** For a sample set \( S \) which cannot be linearly separated in the original space \( \mathbb{R}^n \), data \( S \) could be mapped into a high dimensional feature space using nonlinear mapping \( \phi \). That makes \( \phi(S) \) get very good linear regression characteristics in the feature space \( H \). Hence we can do linear regression in space \( H \), and then return to original space \( \mathbb{R}^n \). That is the support vector non-linear regression.

Implementation steps for nonlinear problems

a. Finding a kernel function \( K(s, t) \), making \( K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \).

b. Finding the solution \( \alpha_i, \alpha_i^* \) of the optimization problem

\[
\begin{align*}
\min & \left\{ -\frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) + \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^{l} (\alpha_i + \alpha_i^*)\varepsilon \right\} \\
\text{s.t.} & \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0, \ 0 \leq \alpha_i, \alpha_i^* \leq C, \ i = 1, 2, \ldots, l
\end{align*}
\]

c. Calculating

\[
b = \begin{cases} 
\begin{align*}
& y_j + \varepsilon - \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)K(x_i, x_j), \ \alpha_i \in (0, \ C) \\
& y_j - \varepsilon - \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)K(x_i, x_j), \ \alpha_i^* \in (0, \ C)
\end{align*}
\end{cases}
\]

d. Structuring nonlinear function

\[
f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)K(x_j, x) + b, \ x_j \in \mathbb{R}^n, \ b \in \mathbb{R}
\]

In this paper, model is built and used to analyze enterprise growth force with SVR model based on RBF (radial basis function), key indicators after extracting features as input independent variable, the first principal component of total asset growth rate, net profit growth rate and revenue growth rate as the dependent variable. By parameter tuning and fitting SVR model is obtained, and empirical analysis is also carried. Detailed analysis is as follow:
Results

The public data of corresponding industry companies listed in GEM is used in preliminary analysis. Databases are built on existing financial data to comprehensively evaluate related enterprises. Extracting related feature algorithm in big data, classification model, regression model are all applied to screen out enterprise evaluating indicators and to predicate growth.

In order to verify the feasibility of the evaluation system, we select the data to test and verify. It is important to point out that the industries and indicators use in this paper are not the same as the known standards, with the pure purpose of the test is to verify the validity of the model.

Empirical Analysis

The Internet and Information Services Industry

A. Yoy_assets (total asset growth rate) as the target dependent variable, the selected indicators are cfps (cash flow per share), fcpf (cash flow per share), ocftocf (operating cash flow to cash flow), and ebitdointere (cash interest coverage ratio).

Establishing SVR model and randomly selecting 10% test data.

B. With yoy_profit (net profit growth rate) as the target dependent variable, the follow indicators are selected as a result of characteristic selection: fcffps (Corporate cash flow per share), fceps (Cash flow per share), ebitdaps (Earnings Before Interest, Tax, Depreciation and Amortization per share), netprofitmargin (Net Profit Margin on Sales), npto cost expense (profit cost ratio), profitto (Net profit / total operating income), ope (Operating profit / total revenue), ebitdosal (EBITDA/ total revenue), ocftr (Net cash flow from operating activities / operating income), currentdebtto (current liability ratio), apturn (turnover of payable).

C. With yoy_tr (total revenue growth rate) as the target dependent variable, the follow indicators are selected as a result of characteristic selection: cfps (cash flow per share), fcffps (corporate cash flow per share). Fcftocf (Net cash flow from financing activities), debttoassets (asset liability ratio), catur (current assets turnover), assets turn (total assets turnover).

Establishing SVR model and randomly selecting 10% test data, the prediction is shown in figure 4 with an average error 7.9%.

Overall, the enterprise growth force indicators of the internet and information services industry mainly focus on cash flow and asset liability related to the management model of Internet industry. Internet and information services industry is the light asset industry, enterprises have no heavy machinery and plant, with main assets as computer and other electronic office equipment and a little real estate. The main cost is labor and management. From the point of view of management model, after obtaining the project the enterprise needs to put in the labor cost. But enterprises have to wait to get payment after acceptance of project, and it often takes a long time. Accordingly these operating liabilities (staff salary, etc.) will lead to weak cash recovery capacity. That is the key to effect enterprise growth of internet and information services industry. Enterprise will be more and more dynamic if it is able to manage these indicators well.

Special Chemical Industry

A. With yoy assets (total asset growth rate) as the target dependent variable, the follow
indicators are selected as a result of characteristic selection: bps(net asset per share), fcffps (corporate free cash flow per share), tangibleassets/totalassets (proportion of tangible assets), ebitdatointerest (EBITDA/ interest expenses), assetsturn (asset turnover).

Establishing SVR model and randomly selecting 10% test data.

B. With yoy profit (total asset growth rate) as the target dependent variable, the follow indicators are selected as a result of characteristic selection: grps(operating income per share), cfps(cash flow per share), fcfeps(free cash flow per share), roic(return on capital), grossprofitmargin (gross profit rate), ocftocf (proportion of operating cash flow), currentdebt/todebt (proportion of current asset).

Establishing SVR model and randomly selecting 10% test data.

C. With yoy_tr (revenue growth rate) as the target dependent variable, the follow indicators are selected as a result of characteristic selection: ebitps(earnings before interest and tax per share), fcffps(free cash flow per share), ebitdas (EBITDA per share), assetstoequity (equity multiplier), arturn(turnover of account receivable), yoyprofit(net profit growth rate).

Establishing SVR model and randomly selecting 10% test data.

Special chemical industry products cover a wide range, including special fluoride, lubricating oil additives, liquid crystal chemicals, charge regulator, etc. The key indicators of the special chemical industry are mainly cash flow related rates. As a traditional heavy asset industry, the operating activities of chemical industry enterprises are greatly influenced by cash flow which is crucial to production and business activities. The results indicate that the model is very good for identifying relevant indicators.

Summary

Different environment and industry type innovative startups possess result in different operation model and asset attributes, which makes the crucial indicators of enterprise growth different. For internet and information services industry, indicators of enterprise growth are mainly focused on asset and liability because of weak cash recovery capacity. Good cash flow capacity is required in special chemical industry which has high barrier in market. The results of prediction are based on the random selection of the sample number 10%. Asset growth rate, net profit growth rate and revenue growth rate are used as dependent variables. The prediction of enterprise growth force is not same in different models in which asset growth rate, net profit growth rate and revenue growth rate are used as dependent variables. The prediction results have a certain imprint effect on the historical data and from it production management characteristics of different industry and the general trend of enterprise growth can be seen. The main innovation of this paper is to introduce the feature extraction and SVR regression algorithm based on data mining which makes the evaluation model of enterprise growth more accurate and the forecast results more precise. Taking GEM companies in internet and information services and special chemical industries as example, by extracting feature for different types of enterprises the critical factors of innovative enterprise growth obtained are more practical. It should be pointed out that, as sample rates are of 10%, it is necessary to increase the amount of data to enhance the training effect of the model.
References


