Expert Systems with Applications 38 (2011) 6424–6429

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Early warning of enterprise decline in a life cycle using neural networks and rough set theory

Yu Cao [a], Xiaohong Chen [a,*], Desheng Dash Wu [b,*], Miao Mo [c]

[a] School of Business, Central South University, Changsha 410083, PR China
[b] Reykjavik University, Kringlunni 1, IS-103 Reykjavík, Iceland and RiskLab, University of Toronto, 1 Spadina Crescent, Toronto, ON, Canada M5S 3G3
[c] Xiangya School of Medicine, Central South University, Changsha 410083, PR China

ARTICLE INFO

Keywords:
Enterprise life cycle
BP neural network
Rough set
Early warning
Decline

ABSTRACT

Early warning of whether an enterprise will fall into decline stage in a near future is a new problem aroused by the enterprise life cycle theory and financial risk management. This paper presents an approach by use of back propagation neural networks and rough set theory in order to give an early warning whether enterprises will fall into a decline stage. Through attribute reduction by rough set, the influence of noise data and redundant data are eliminated when training the networks. Our models obtained favorable accuracy, especially in predicting whether enterprises will fall into decline or not.

1. Introduction

The enterprise life cycle is an important part of enterprise theory which views an enterprise from the longitudinal perspective, where enterprises may move through a fairly predictable sequence of developmental stages. Over the past three decades, a considerable study has been conducted on enterprise life cycles, which is usually linked to the study of organizational growth and development (Snyder, 2003). Since the majority of enterprise life cycle research is on investigation of the number of stages an enterprise can experience and what characteristics an enterprise exhibits in different stages, the approach of classifying different life cycle stages, is a crucial issue for enterprise life cycle researches.

The number of enterprise life cycle stages proposed in existing works has varied: some researchers have identified up to ten different stages of an enterprise life cycle, while others may only prefer three stages. A widely used enterprise life cycle usually includes four or five stages that can be encapsulated as start-up, growth, maturity, decline, and death (or revival). During enterprises' different life cycle stages, the decline stage is much more crucial than other stages. After experiencing original birth, rapid growth and a relative steady maturity stage, many enterprises fall into decline and confront crises in many aspects. Even though Miller and Friesen (1984) point out that there are wide variety of transitional paths open to enterprises and there is no definite, singular and irreversible sequence of phases, progressing from birth, to growth, to maturity, and then onto revival or perhaps decline, a number of declining enterprises cannot walk out of the woods and perished, implying that the decline stage of an enterprise is much more crucial through the total enterprise life cycle stages.

Existing literatures pay much attention to develop various early warning models to predicting the enterprises' financial crises, however, little research have done on early warning the enterprises falling into decline stage. Since this topic is important for the reasons that a firm falls into decline compose a common source of financial crises, and enterprise's managers and shareholders can adopt measures ahead to prevent the enterprise falling into decline, this paper we focus on construct an early warning model to predicting and prewarning the enterprise falls into decline.

One clue of this topic is from the research of enterprise life cycle theory. Previous studies have approached various methods to solve the problem of classifying enterprise life cycle stages (Hwang, 2004). Generally speaking, the existing life cycle classification methods can be divided into two categories in accordance with the employed techniques. Most of life cycle studies were performed by applying simple self-assessment questionnaires to top managers of an enterprise, in an effort to ascertain where their firms were positioned within the enterprise life cycle framework. Another category of method of classification uses various financial and non-financial measures. The most common measures are age, sales, and growth rate with growth rate defined as income growth, asset growth, or sales growth (Hanks, Watson, Jansen, & Chandler, 1993; Miller & Friesen, 1984; Olson & Terpstra, 1992). For example, Miller and Friesen (1984) have classified a firm with a decrease in sales growth is in the decline stage.

The mathematic methods used in previous life cycle researches are commonly cluster analysis and factor analysis, etc., while another similar field, financial crises early warning models commonly

* Corresponding authors. Tel.: +1 416 8805219.
E-mail addresses: dash@risklab.ca, dwu@rotman.utoronto.ca (D.D. Wu).

0957-4174/$ - see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2010.09.138

adopt many much more complicated methods, such as discriminate analysis, case-based reasoning and expert systems, et al. to solve the problem. At present years, a new method which combined the rough set theory and neural network to prediction enterprises' financial crises arose and shows fine performance. This paper, referring to the measures used in financial crises, a multilayer back propagation network (BP), combined with one tool of soft theory - rough set are used to construct an enterprise-decline early warning model. This measure can integrate each advantage of neural network and rough set, for BP network is often widely applied in many areas such as predicting stock price, early detecting financial risks and many others, because of its ability to study and remember the relation between inputs and out-puts as well as to approach any types of function, and for rough set can reduce the influence due to the drawbacks of BP such as low training speed and easily affected by noise and weak interdependency data through attribution reduction process (Xiao, Ye, Zhong, & Sun, 2009). In other words, we take advantages of rough set – it could deal with incomplete, uncertain and ambiguous data and complex data containing mass of variables; moreover, it can abstract knowledge or patterns from data. After attribution reduction, the noise data and weak interdependency data are eliminated, so the influences they have to BP during the initialization, study and training process are avoided, and then the accuracy of predictions is developed and proved.

2. Rough sets and neural network

2.1. Rough sets

Rough sets theory (RST) is a machine-learning method, which is introduced by Pawlak (1991) in the early 1980s, has proved to be a powerful tool for uncertainty and has been applied to data reduction, rule extraction, data mining and granularity computation (Yeh, Chi, & Hsu, 2009).

The basic concept in rough set theory is an information system which can be expressed by a 4-tuple $S = (U,A,V,f)$, where $U = \{x_1,x_2,\ldots,x_n\}$ is a finite set of objects, called the universe; $A = C \cup D$ is a finite set of attributes, which is a union of the condition attributes set C and decision attributes set D with $C \cap D = \varnothing$; $V = \cup_{a\in A} V_a$ is a domain of attribute a, and $f : U \times A \rightarrow V$ is an information function to determine each object x_i's attribute value in set U that is: $f(x_i,a) \in V_a$, for $\forall x_i \in U$, $a \in A$.

In rough set theory, the objects in universe U can be described by various attributes in attributes set A. When two different objects are described by the same attributes, then these two objects are classified as one kind in the information system S, thus we call their relationship is indiscernibility relation. In mathematical word, an indiscernibility relation $IND(B)$ generated by attribute subset $B \subseteq A$ on U, is defined as follows:

$$IND(B) = \left\{(x_i,x_j) \in U \times U | f(x_i,a), = f(x_j,a), \forall a \in B\right\}$$

The partition of U generated by$IND(B)$ is denoted by $U/IND(B) = \{C_1,C_2,\ldots,C_k\}$ for every C_i is an equivalence class. For $\forall x \in U$ the equivalence class of x in relation to $U/IND(B)$ is defined as follows:

$$[x]_{U/IND(B)} = \{y \in U | f(y,a) = f(x,a), \forall a \in B\}$$

Let $X \subseteq U$ be a target set and $P \subseteq A$ be a attribute subset, that we wish to represent X using attribute subset P. In general, X cannot be expressed exactly, because the set may include and exclude objects which are indistinguishable on the basis of attributes P. However, Pawlak (1991) present a method to approximating the target set P only by the information contained within P by constructing the P-lower and P-upper approximations of X, which is respectively defined as:

$$P\text{-lower approximations of } X : P_*X = \left\{x | [x]_{U/IND(P)} \subseteq X\right\}$$

$$P\text{-upper approximations of } X : P^*X = \left\{x | [x]_{U/IND(P)} \cap X \neq \varnothing\right\}$$

The P-lower approximation, also called the positive region, is the union of all equivalence classes in $[x]_{U/IND(P)}$ which are contained by (i.e., are subsets of) the target set X. In another word, the lower approximation is the complete set of objects in $U/IND(P)$ that can be positively (i.e. unambiguously) classified as belonging to target set X.

The P-upper approximation is the union of all equivalence classes in $[x]_{U/IND(P)}$ which have non-empty intersection with the target set, that is the complete set of objects that in $U/IND(P)$ that cannot be unambiguously classified as belonging to the complement $(\overline{X})$ of the target set X. In other words, the upper approximation is the complete set of objects that are possibly members of the target set X.

One of the most important aspects in rough set theory is the discovery of attribute dependencies, that is, we wish to discover which variables are strongly related to which other variables. For this purpose, given two attribute subset P, $Q \subseteq A$, Then, the dependency of attribute set Q on attribute set P, $\gamma_P(Q)$, is given by

$$\gamma_P(Q) = \frac{card(\cup_{X\in U/IND(Q)} P_*X)}{card(U)}$$

where $\cup_{X\in U/IND(Q)} P*X$ can be denoted as $POS_P(Q)$, which means that the objects in it can be classified to one class of the classification $U/IND(Q)$ by attribute P.

An attribute a is said to be dispensable in P with respect to Q, if $\gamma_P(Q) = \gamma_{P-\{a\}}(Q)$; otherwise a is an indispensable attribute in P with respect to Q.

Let $(S = U,A,V,f)$ be a decision table, the set of attributes $P(P \subseteq C)$ is a reduce of attribute,C if it satisfied the following conditions:

$$\gamma_P(D) = \gamma_C(D), \quad \gamma_{P'}(D) \neq \gamma_C(D) \quad \forall P' \subset P$$

A reduction of condition attributes C is a subset that can discern decision classes with the same accuracy as C, and none of the attributes in the reduced can be eliminated without decreasing its distrainable capability (Pawlak, 2002).

Though it is a kernel concept in rough set, it is difficult to calculate the reduction if the size of information system is large. Many scholars proposed a variety of attribute reduction algorithm, such as: consistency of data (Mi, Wei-zhi, & Wen-Xiu, 2004; Pawlak, 1991), dependency of attributes (Wang, Hu, & Yang, 2002), mutual information (Skowron & Rauszer, 1992), discernibility matrix (Jue

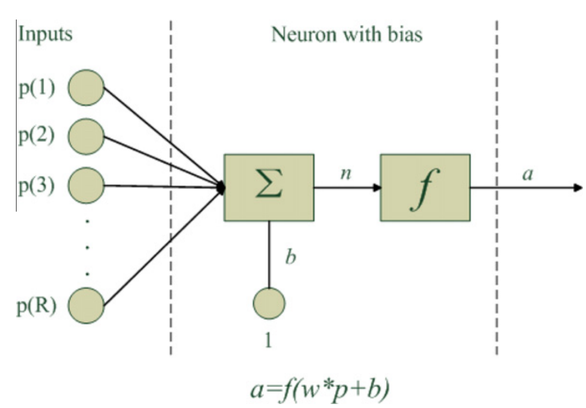

Fig. 1. A neuron of BP.