

文章编号: 1003-207(2004)05-0006-06

基于信息扩散原理运用人工神经网络识别股票级别

黄亦潇, 邵培基, 李菁菁

(电子科技大学管理学院, 成都 610054)

摘要: 本文提出采用人工神经网络的方法对股票级别进行识别, 以便辨识股票的优劣。首先建立股票识别的指标体系, 对神经网络的训练样本中的各支股票的各项指标进行初步评级, 然后用因素状态空间上信息扩散的方法对初步评级的结果进行优化处理, 把处理后的结果作为神经网络训练样本的输入和输出因素, 并利用 BP 算法进行网络训练, 当计算达到要求的精度后即完成训练。经训练所得的神经网络即可用于识别股票级别。本文最后给出一个实例, 具体说明运用信息扩散和人工神经网络的方法识别股票级别的过程, 并对结果进行了分析讨论。采用本文所介绍的方法对股票级别进行识别, 可以有效地减少主观因素及市场波动所带来的影响。

关键词: 股票; 评级; 人工神经网络; 信息扩散

中图分类号: F830 文献标识码: A

1 引言

股票评级是股票资信评估的一个重要项目, 它可以为投资者提供股票的信息, 减少投资者获取信息的成本, 为投资者的决策提供重要的依据。股票评级也叫股票等级评估。它通过对与股票有关的因素进行分析和研究, 对股票质量进行评价, 并以简单的记号或说明展示给社会公众和投资者。

针对股票评级问题, 国内外学者进行了广泛深入的研究, 其研究主要是从评级指标和评级方法两个方面进行。在评级指标方面, 应用较多的是美国标准普尔公司采用每股收益和股利两个指标来评价股票, 台湾经济日报对股票的评级采用获利能力、股价波动率、营运能力、股票交易周转率、短期偿债能力五个指标^[1]。但这些指标体系并不完全适用于国内的证券市场, 因此国内的学者也对股票评级指标进行了研究, 如文献[2]提出根据风险下的股票收益, 从不同角度建立指标体系来进行股票评级; 文献[3]提出利用公司综合分析指标与股票交易测定指标来综合评估股票级别; 文献[4]则考虑了上市公司的行业背景以及公司本身的历史经营状况作为其股票评级的重要依据。上述的研究为我国建立自己的股票评级指标体系提供了帮助。

针对股票评级的方法, 近几年出现了将人工智能技术引入到股票评级中的潮流, 如文献[5]提出利用聚类的方法对股票进行分类评级。而神经网络方法则是应用最多的一种。文献[6]利用神经网络对股票市场的整体趋势进行预测。文献[7]则利用神经网络实现对个股的价格变化, 投资回报等指标进行预测和评价。文献[8]将模糊技术与神经网络相结合, 提出利用模糊神经网络用于股票评级。文献[9]以神经网络为基础构建决策支持系统来进行股票评价和预测, 从而提高了神经网络方法应用的系统性。

本文的研究主要是针对股票评级的方法, 由于神经网络具有预测精度高的特点, 因此本文选择神经网络方法作为股票评级方法。与上述有关神经网络进行股票评级的研究不同的是本文强调了对神经网络训练样本的处理, 提出利用因素状态空间上的信息扩散原理对样本股票评级的结果进行优化处理, 从而降低神经网络的训练时间。同时可减少训练样本评价中主观因素的影响, 从而提高训练样本的精度, 进一步提高神经网络评级结果的精确性。

2 因素状态空间上的人工神经网络

人工神经网络 (Artificial Neural Networks, ANN) 简称神经网络, 它是一类模拟生物神经的系统结构, 具有较强的自学习能力, 可在学习过程中不断完善自己, 使经过训练后的网络达到希望的精度要求。本文所介绍的 ANN 是前馈神经网络, 采用

收稿日期: 2003-08-06; 修订日期: 2004-07-20

作者简介: 黄亦潇 (1978-), 男 (汉族), 四川攀枝花人, 电子科技大学管理学院, 博士生, 研究方向: 电子商务、数据挖掘。

的是后向传播(Back Propagation, BP)模型的学习算法,即从输出层向输入层逐层进行权值修正。

在利用 ANN 进行训练和预测的时候,通常会遇到具有模糊信息的样本,如“可行性”,“满意程度”等概念。当利用传统的 ANN,即每一个输入节点对应一个输入变量,每一个输出节点对应一个输出变量,这种结构将难以处理具有模糊信息的样本。

文献[10]提出了构造因素状态 ANN,利用因素状态空间上信息扩散原理对 ANN 的输入输出因素进行优化处理,从而解决模糊信息的问题。下面,我们将简要介绍因素状态神经网络和信息扩散方法的原理。

2.1 因素状态神经网络

因素空间就是一个以布尔代数 $F = F(\wedge, \vee, c, 10)$ 为指标集并包括空集且组合运算定义为集合笛卡尔运算的集合族 $\{X_f\}(f \in F)$ 。 f 称为因素空间,而 X_f 称为 f 的状态空间。因素空间为概念描述提供了一般的数学框架,便于人们处理较为抽象和难于度量的因素。

定义 1 设 u_j 是一神经元, $\{X_f\}(f \in F)$ 是一因素空间。若存在 $f \in F$, 使神经元的输入 $X_j = X_f$, 则称 u_j 是一输入型的因素神经元; 若存在 $g \in F$, 使神经元的输出 $Y_i = X_g$, 则称 u_j 是一输出型的因素神经元。

仅是输入型和输出型的因素神经元称单侧因素神经元, 双侧因素神经元称为完全因素神经元, 而输入输出节点是因素神经元的 ANN 称为因素神经网络。

定义 2 设 X_f 是 f 的状态空间, u_j 是一神经元。设 $x \in X_f$ 是 f 的某一状态。如 x 出现确定了 u_j 的输入(输出)值为 1, 而 x 不出现确定了 u_j 的输入(输出)值为 0, 则称 u_j 是因素 f 关于状态 x 的因素状态神经元。

输入输出节点由状态神经元构成的 ANN 称为因素状态神经网络。

2.2 信息扩散方法

因素状态神经网络的输入输出节点采用的是具有模糊性的因素, 对这些因素我们采用信息扩散的方法对其进行优化处理^[11]。

设 $W = \{w_1, w_2, \dots, w_m\}$ 是知识样本, V 是基础论域。记 W_j 的观测值为 V_j , 设 $x = \varphi(V - V_j)$, (V 是控制状态) 当样本 W 为非完备样本时, 存在函数 $\mu(x)$, 使 V_j 点获得的量值为 1 的信息可按 $\mu(x)$ 的量值扩散到 V 中去, 且扩散所得的信息分

布:

$$Q(V) = \sum_{j=1}^m \mu(x) = \sum_{j=1}^m \mu(\varphi(V - V_j))$$

该信息分布能够更好地反映 W 所在总体的规律。

根据扩散估计的定义, 设 $x = \varphi(V - V_j) = \frac{V - V_j}{\Delta m}$ (其中 Δm 为窗宽, 则函数

$$f_m(V) = \frac{1}{m \cdot \Delta m} Q(V) \\ = \frac{1}{m \cdot \Delta m} \sum_{j=1}^m \mu\left(\frac{V - V_j}{\Delta m}\right)$$

为扩散估计。

本文所述的因素状态神经网络采用正态扩散来进行优化处理, 其扩散函数为:

$$\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$$

设窗宽为 Δm , 令 $h = m \cdot \Delta m$, 则扩散估计为:

$$f_m(V) = \frac{1}{\sqrt{2\pi} \cdot h} \sum_{j=1}^m \exp\left[-\frac{(V - V_j)^2}{2h^2}\right]$$

2.3 信息扩散与因素状态人工神经网络的结合

因素状态神经网络的输入输出元素是基于因素空间的元素, 具有模糊信息的性质。利用信息扩散的方法可以解决模糊信息样本所带来的知识不完备性, 同时可以避免在神经网络的训练过程中出现过拟合的情况。

设输入输出因素分别为 f, g , 其中 f 是可测因素, g 是不可测因素。 f, g 的状态集为 U_f 和 O_g 。

$$U_f = \{u_1, u_2, \dots, u_N\}$$

$$O_g = \{o_1, o_2, \dots, o_M\}$$

取 f 的控制状态集为 $S = \{s_1, s_2, \dots, s_N\} n < N$;

取 g 的控制状态集为 $I = \{i_1, i_2, \dots, i_m\} m < M$ 。

利用正态信息扩散, 确定状态 u_i 扩散给控制状态 s_j 的状态函数即扩散估计为:

$$\mu(s_j, u_i) = \frac{1}{\sqrt{2\pi} \cdot h} \exp\left[-\frac{(s_j - u_i)^2}{2h^2}\right] \quad (1)$$

采用如下的公式来计算 h :

$$h = \begin{cases} 1.6987d/N & 1 \leq N \leq 5 \\ 1.4456d/N & 6 \leq N \leq 7 \\ 1.4230d/N & 8 \leq N \leq 9 \\ 1.4208d/N & 10 \leq N \end{cases} \quad (2)$$

其中 $d = \|a - b\|$, a 和 b 分别为样本观测到的最大和最小状态。 N 为样本容量。

将(1)式进行归一化处理,得:

$$\mu'(s_j, u_i) = \mu(s_j, u_i) \setminus \sum_{u_f} \mu(s, u_i) \quad s \in S \quad (3)$$

u_i 对 S 中的各控制状态进行信息扩散,可得 u_i 的状态函数:

$$\mu'(u_i) = (\mu'(s_1, u_i), \mu'(s_2, u_i), \dots, \mu'(s_n, u_i))$$

将各输入样本的状态函数落影到 I 上,即将输出相同的样本进行矢量叠加:

$$\mu'_i = \sum_{o_i=i} \mu'(u_i) \quad (4)$$

经归一化处理得:

$$\mu_{i_i} = \mu'_i \setminus \sum_{m \in I} \mu' I_n \quad (5)$$

经过处理后最后得到的样本输入为 μ_{i_i} , 输出为 $(0, \dots, 0, 1, 0, \dots, 0) I = I_i$ 处为 1。

当输入和输出因素都是可测因素时,其计算与此相类似,本文在此不再介绍。

3 因素状态人工神经网络识别股票级别的步骤

在识别股票级别的研究中,经过初步评级后所得的结果具有模糊信息。前文所介绍的信息扩散原理和因素状态神经网络方法对模糊信息具有较好的处理效果,可以利用该方法进行股票评级。

传统的股票评级方法通常是对股票的历史数据进行考察、分析,以便得到评级结果。这种方法受市场波动的影响较大。利用 ANN 进行股票评级,将由历史数据所得的评级结果作为 ANN 的输出,将当期数据(可根据需要选择当年、当月或当日数据)作为 ANN 的输入。即投资者可根据当期数据判断股票的级别,使对股票的当期市场表现评价与长期趋势趋于一致。利用要本评级方法减少了评级结果中主观因素和市场波动带来的影响,使评级结果具有较好的客观性。

利用因素状态 ANN 识别股票级别,其具体步骤如下:

(1) 选择 ANN 的训练样本,通过对样本进行初步评级,得到 ANN 的输入和输出因素。

对每支股票的当期数据作一次评价,得到一个分值型结果,这一结果反应的是每支股票在当期的市场表现。将这个结果作为 ANN 的输入因素。

对每支股票的长期数据(这里可根据具体的评级要求选用适当的时间段)做一次评价,得到每支股票的级别,将这一结果作为 ANN 的输出因素。

以上两次评价中的评级方法均可采用传统的评级方法,通过建立指标体系来进行股票评级。但在数据选择的时间段上不同,在最后结果的表示上,当期数据评价的结果被折算为一个 10 以内的分值,而长期数据的评价结果则以股票级别来表示。

(2) 利用信息扩散原理,对(1)中所得的结果即 ANN 的输入输出因素进行优化处理,将所得的结果用于网络训练,经过训练达到要求的精度后,就可以利用训练好的 ANN 进行股票评级。

(3) 选择需要评级的股票,对这支股票的当前数据按步骤(1)中的方法计算,将所得的分值型结果代入 ANN,所得的输出即为该股票的级别。

(4) 对股票评级的结果进行分析,得出相应的投资建议。

4 应用实例

利用本方法对股票评级,对样本的选择要考虑股票分析的稳定性,信息扩散过程中信息传递的准确性及神经网络的特性。本文选用沪市的股票进行分析,每支股票都对其 1998 年 - 2003 年的年报数据进行统计。其中 1998 - 2002 年的数据作为长期数据,2003 年的年报数据作为当期数据。本文的数据主要来源于中国证券监督管理委员会网站,中国证券报网站,泰阳证券网站以及其他一些提供年报数据的网站。由于 1999 年以前的年报较难寻找并且不够完整,全部 6 年年报齐全的股票共找到 183 支。由于本文的研究要求样本在一定时间内具有稳定性,因此本文对上述 183 支股票进行筛选,首先剔除在 1998 - 2003 年出现违规情况的股票,如红光(600083),然后对剩余各支股票的基本财务数据进行时间序列分析,剔除掉波动情况异常显著的股票,最终选用 112 支股票作为 ANN 的训练样本。

(1) 对样本进行初步评级。

根据中国证券业协会在 1994 年制定的用于分析和评估我国上市公司状况的指标体系,在兼顾全面性、可比性和可行性三个原则的基础上,并参考现有的实际效果较好的评级方法,在征求了相关专家的意见后,本文选择以下 10 个指标构成评价指标体系:

主营业务增长(%), 净利增长(%), 股东权益增长, 每股净资产(元), 每股未分配利润(元), 每股资本公积(元), 每股现金流量(元), 净资产收益率, 每股收益(元), 主营业务利润率共 10 个指标。

首先对股票的当期数据进行计算,得到其分值。

这里我们以 2003 年的数据作为当期数据。这个结果将作为 ANN 的输入因素, 是一个可测的因素。

然后对样本中各支股票的长期数据进行评价。

本文选择 1998~ 2002 五年的年报数据综合进行评

价。评级结果用 $\{a, b, c, d\}$ 来表示{优, 良, 中, 劣}四级。将评级结果作为神经网络的输出因素, 是一个不可测因素。

经计算后所得的数据如表 1 所示:

表 1 训练样本记录表

NO	U	O	NO	U	O	NO	U	O
1	5.65	c	41	7.30	b	81	6.70	b
2	6.70	b	42	5.00	d	82	5.90	c
3	5.80	b	43	5.00	d	83	4.60	d
4	7.40	b	44	4.75	d	84	7.90	d
5	5.00	d	45	5.75	c	85	6.80	b
6	6.30	b	46	6.75	b	86	5.50	d
7	5.90	c	47	5.85	c	87	5.60	c
8	7.25	c	48	5.50	c	88	6.90	c
9	8.20	c	49	5.50	d	89	6.60	b
10	6.50	b	50	5.30	c	90	5.70	c
11	7.50	d	51	6.80	b	91	5.30	c
12	4.30	d	52	6.10	b	92	6.30	c
13	5.75	c	53	5.90	c	93	7.40	b
14	5.30	c	54	6.60	b	94	5.70	c
15	6.50	b	55	7.75	a	95	6.80	b
16	5.50	d	56	5.00	d	96	7.50	a
17	6.30	b	57	6.00	c	97	4.90	d
18	6.25	b	58	8.35	a	98	5.10	d
19	6.90	a	59	5.60	c	99	5.70	c
20	5.00	d	60	4.80	d	100	6.50	b
21	5.00	d	61	5.70	b	101	5.80	c
22	4.55	d	62	5.20	c	102	6.85	b
23	7.60	b	63	6.80	b	103	5.90	c
24	6.75	c	64	7.80	a	104	4.80	c
25	4.75	d	65	7.20	a	105	5.45	d
26	7.80	a	66	8.20	a	106	7.80	a
27	8.25	a	67	6.10	c	107	6.80	b
28	6.00	c	68	5.25	d	108	7.20	b
29	7.50	b	69	6.80	b	109	7.50	a
30	4.80	d	70	7.20	b	110	6.90	b
31	5.75	c	71	7.30	b	111	6.80	c
32	5.80	c	72	6.50	c	112	6.60	b
33	4.30	d	73	5.50	c			
34	7.80	a	74	6.30	d			
35	7.20	b	75	6.50	c			
36	5.00	d	76	7.10	a			
37	6.00	c	77	7.90	b			
38	8.50	a	78	4.90	c			
39	5.20	d	79	5.20	d			
40	5.40	c	80	5.00	c			

注: 表中 NO 表示样本编码, U 代表当期评价的结果, O 代表长期评级的结果。

(2) 利用信息扩散的原理对输入输出因素进行优化处理。

表 1 中 U 列的最大值为 8.50, 最小值为 4.30。综合考虑计算的工作量及信息扩散过程中信息传递的准确性, 故选取 15 个点作为输入因素的控制状态点, 步长为 0.3。则控制状态集为:

$S = \{s_1, s_2, \dots, s_{15}\} = \{4.3, 4.6, \dots, 8.2, 8.5\}$
 因为输出因素是非可测因素, 则输出因素的状态集即为其控制状态集:

$$I = \{i_1, i_2, i_3, i_4\} = \{a, b, c, d\}$$

由以上两式确定 ANN 的构成为 15 个输入节点和 4 个输出节点。其中 $\{a, b, c, d\}$ 四个级别分

别用向量(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1) 来表示。

根据不同的 i_k 可以将表 1 中的样本 W 分为四个子样本, 得 $W(ik)$, $k = 1, 2, 3, 4$ 。信息扩散窗宽 h 由各子样本 $W(ik)$ 的平均域宽 d 和平均样本数目 N 决定。

$$d = \frac{1}{4} \sum_i (a(ik) - b(ik)) = 1.7625$$

$$N = \frac{1}{4} \sum_i N(ik) = 28$$

由(2) 式得 $h = 1.4208d/N = 0.08943$ 。由式(1), (3), (4), (5) 可求得各控制状态的落影 $\mu_i(S_j)$, 并将其转化为 ANN 的训练样本。

例如对第 5 个样本 $u_s = 5.00$, 其扩散给控制状态 S_j 的信息为:

$$\mu(s_j, u_s) = \frac{1}{\sqrt{2\pi} \cdot h} \exp\left[-\frac{(s_j - u_i)^2}{2h^2}\right]$$
$$= 4.4609 \exp(-62.5178(s_j - u_s)^2)$$

$$\mu(u_s) = \{0, 0.0002, 2.3873, 0.3659, 0, \dots, 0\}$$

归一化处理得:

$$\mu'(u_s) = \{0, 0, 0.867, 0.133, 0, \dots, 0\}$$

同理对所有的样本进行计算, 并将输出为同一级别的输入样本按照公式(4), (5) 叠加, 例如对 $O_i = I_i = a$, 有:

$$\mu'_{I1} = \{1, 3.6294, 10.5834, 7.3918, 4.248, 0.0144, 0, \dots, 0\}$$

经归一化处理得:

$$\mu_{I1} = \{0.038, 0.135, 0.394, 0.275, 0.158, 0, \dots, 0\}$$

其对应的输出为 $\{1, 0, 0, 0\}$ 。则 μ_{I1} 是输入为 $\{1, 3.6294, 10.5834, 7.3918, 4.248, 0.0144, 0, \dots, 0\}$, 输出为 $\{1, 0, 0, 0\}$ 的一个学习样本。

同理可求得用于 ANN 训练的 4 个样本。

(3) 用于学习的 ANN 具有 15 个输入节点, 4 个输出节点。根据经验, 选择 ANN 的隐层数为 2, 第一个隐层的节点数为 7, 第二个隐层的节点数为 5。将这 4 个样本对神经网络进行训练。本文采用 Matlab6.1 作为编程工具, matlab6.1 中的神经网络工具箱包括多个用于神经网络初始, 训练, 仿真的函数, 利用这些函数和 Matlab 提供的基本命令可以方便、快捷地进行神经网络的设计, 完成复杂的工作^[12]。本文所设计的 ANN 的学习步长为 $\eta = 0.8$, 惯量系数 $\alpha = 0.6$, 期望误差 $\epsilon = 0.01$ 。经过 3071 次

训练, 可使误差小于期望值, 从而完成网络的训练。

(4) 在完成了网络的训练后, 就可以用该 ANN 来进行股票级别的识别。利用被评价股票的当前指标来求得 u_i , 将 u_i 值扩散到 ANN 的 15 个控制点上, 将所得的分布作为 ANN 的输入, 进行训练后即可求得网络的输出。输出的峰值即为该支股票的级别。

本文在训练样本以外随机选取了五支股票进行评级, 分别为 1 号、2 号、……、5 号。根据其 2003 年年报数据, 计算其分值分别为 6.2, 6.9, 5.5, 4.9, 6.7。代入 ANN 进行训练, 得到评级结果分别为: b, b, c, c, b。也即 1 号、2 号、5 号股票较优, 从评级结果可看出, 该结果反应了股票的长期发展趋势, 使股票的当期市场表现与长期趋势趋于一致, 该评级结果, 将为投资者提供了比较客观的决策支持投资建议。

5 结束语

传统的股票评级方法在确定指标体系中各评价因素的权重上带有一定的主观成分, 所得的结果难以反映客观实际, 并且受市场波动的影响较大, 因此其评级结果与实际情况有一定的偏差。

本方法是在传统评级方法的基础上, 利用信息扩散原理对传统评级方法的结果进行优化处理, 将处理后的结果作为 ANN 训练样本的输入输出因素。经过训练后的 ANN 就可以对股票进行实时评级。通过本方法得到的股票评级结果能更好的反应客观实际, 因此具有较好的投资指导意义。

本方法使对股票的当期市场表现评价与长期趋势趋于一致, 减少了市场波动带来的影响, 但对于异常数据及有效的市场变化则反应不足, 这个问题将是今后进一步研究的方向。

参考文献:

- [1] 胡笑梅. 关于股票评级问题的研究[J]. 南京经济学院学报, 1998, (3): 59- 62.
- [2] 胡荣华. 上市公司股票评级实证研究[J]. 财经问题研究, 2001, (12): 61- 63.
- [3] 魏然. 股票评级方法体系初探[J]. 北京统计, 2000, (1): 25- 26.
- [4] 仝全友. 股票评级的困境与改进[J]. 山西财经大学学报, 2002, 24(6): 144- 145.
- [5] 刘景芳. 聚类分析与股票研究中的应用[J]. 天津纺织工学院学报, 1995, 14(4): 82- 86.
- [6] Kimoto T, Asakawa K. Stock Market Predictions System

- with Modular Neural Networks[J]. Proc IJCNN, 1990, 1 (6): 1- 6
- [7] Kryzanowski, M. Galler, D. W. Wright. Using Artificial Neural Networks to Pick Stocks[J]. Financial Analysis Journal, 1993, 49(7): 21- 27.
- [8] Wong F S, Wang P Z, Goh T. H. Fuzzy Neural Systems for Stock Selection[J]. Financial Analysts Journal, 1992, 48 (1): 47- 53.
- [9] Norio B, Mamoru M, Maki N. A User Friendly Decision Support System for Dealing Stocks Using Neural Network [C]. Proceedings of 1993 International Joint Conference on Neural Network, 1993, 762- 765.
- [10] 王家鼎, 黄崇福. 模糊信息优化处理技术及其应用 [M]. 北京: 北京航空航天大学出版社, 1995: 178- 193.
- [11] 王新洲, 游扬声. 论信息扩散估计的窗宽[J]. 测绘科学, 2001, 26(1): 16- 19.
- [12] 闻新, 周露, 王丹力. MATLAB 神经网络应用设计 [M]. 北京: 科学出版社, 2002, 207- 232.

Using Artificial Neural Networks to Evaluate the Grading of Stock Based on the Principle of Information Spread

HUANG Yi-xiao, SHAO Pei-ji, LI Jing-jing

(School of Management, University of Electronic Science and Technology of China Chengdu 610054, China)

Abstract: This paper proposed the method of applying artificial neural Networks to evaluate the grading of stock. At first, we establish the criteria system to evaluate the grading of stock, and then evaluate the stock index of training sample in ANN. Second, optimize the result based on information spread and use the optimized result to train neural networks with the back propagation. When arriving the necessary precision, the training will be stopped. The artificial neural networks may evaluate the grading of stock. In the end of this paper, we give a practical example and the detailed calculation steps, to explain this method and the grading results of stock. Using this method to evaluate the grading of stock will be helpful to reduce the influences of subjectivity and the market changes.

Key words: stock; evaluating grading; artificial neural networks; information spread