

特征背离和风险偏好分析的股价态势预测方法

姚宏亮 黄曼 王浩 李俊照

(合肥工业大学计算机与信息学院 合肥 230009)

摘要 由于股价走势与技术指标走势存在不一致性,基于技术特征的股价态势预测算法效果不佳。从特征背离角度提出了一种股价态势预测算法(Deviated Characteristics Predict Algorithm,DCPA),该算法首先进行背离特征的提取,并计算特征的背离程度,然后根据特征的背离程度值和股票的收盘价利用 BP 网络进行股价态势预测。由于当市场风险偏好高时特征背离与股价态势之间相关性很弱,因此在 DCPA 算法的基础上提出了一种风险偏好的股价态势预测算法(Risk Preference Based Deviated Characteristics predict Algorithm,RPDCA)。首先提取与风险偏好相关的特征,利用风险偏好计算模型获得当前的市场风险偏好类型;进而利用贝叶斯网络学习风险偏好、背离特征与股价走势之间的关系,并利用结点非对称信息熵分析风险偏好与背离特征之间的依赖关系;最后根据风险偏好与背离特征之间关系的变化,自适应性地利用 BP 网络预测股价态势。在实际数据上的实验比较与分析结果表明,RPDCA 算法在股市短期预测中具有更高的预测精度。

关键词 背离特征,风险偏好,DCPA 算法,效用函数,RPDCA 算法

中图分类号 TP399 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.3.007

Trend Forecast of Stock Price Based on Deviated Characteristics and Risk Preference

YAO Hong-liang HUANG Man WANG Hao LI Jun-zhao

(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

Abstract Due to the inconsistent tendency of stock price and technical index, the share price trend prediction algorithm based on the technical features performs poorly. A share price trend prediction algorithm from the point of deviated characteristics(DCPA) was proposed. Deviated features firstly are extracted, the degree of swerve is calculated, and then the trend of price is forecasted by BP neural network according to the degree of swerve and the stock's closing price. While the risk appetite is high, the correlation between stock price trend and deviated characteristics is weak, thus a risk preference based share price trend prediction algorithm named RPDCA was put forward on the basis of DCPA algorithm. Firstly, features which are associated with risk appetite are extracted and the current market risk preference type is acquired through the risk appetite computational model. Secondly, by means of Bayesian network, the structural relationship among risk appetite, deviated characteristics and the trend of stock price is learned, and then the interdependent relationship between risk appetite and deviated characteristics is analyzed by using node asymmetric information entropy. Last, the trend of stock price is forecasted self-adaptedly according to the relationship between risk appetite and deviated characteristics by BP neural network. Based on the comparison and analysis on the actual data, the experimental results show that the RPDCA algorithm has higher precision of short-term prediction.

Keywords Deviated characteristic, Risk preference, DCPA algorithm, Utility function, RPDCA algorithm

1 引言

股票市场是一个不稳定的非线性动态变化的复杂系统,很多学者从技术指标角度研究股票价格态势。许兴军等人在基于 BP 神经网络的股价趋势分析中用收盘价、成交量和 KDJ 指标等基本数据作为模型输入^[1];Chih-Ming Hsu 等人在股市行情波段的特征提取中通过对日线平滑处理,将股价走势的波段属性特征提取出来作为预测项^[2],但无法解释指

标与股价有不同走势的情况。通常将指标走势和股价走势之间出现的不一致现象称为指标背离,它是预示市场走势即将见顶或者见底的信号,从而让投资者预测风险和寻找买入机会。

股票特征数据还包含了很多非线性因素,其中投资者的投资心理是一种非常重要的因素,即风险偏好(Risk Preference, RP)^[3],风险偏好体现投资者在风险和收益之间做出抉择的态度。在决策人因拥有财富多寡以及自身性格、意志等

到稿日期:2015-03-13 返修日期:2015-05-25 本文受国家自然科学基金(61175051,61070131,61175033)资助。

姚宏亮(1972—),男,博士,副教授,CCF 会员,主要研究方向为人工智能和知识工程,E-mail:dmicyhl@163.com;黄曼(1989—),女,硕士生,主要研究方向为人工智能和知识工程;王浩(1962—),男,博士,教授,CCF 高级会员,主要研究方向为人工智能;李俊照(1975—),男,博士,讲师,主要研究方向为机器学习和人工智能。

前提下研究风险偏好决策^[4]具有较强的主观性。因而,更多的研究者是从量化角度分析风险偏好,如:Markowitz 以证券收益率的方差作为风险,开辟了金融股定量分析的时代。在此基础上研究者又引入了 Markowitz 均值-方差修正模型^[5]、绝对离差模型^[6]及半方差模型^[7]等;胡小文等人在此基础上考虑了投资者的风险爱好,引入风险偏好系数,建立了加权半方差风险度量模型^[8]。风险偏好系数建立在效用函数的基础之上,个体的期望效用降幅越大时,个体的风险承受度越低,风险厌恶程度越高。陈菊花等以“风险资产比重的波动性”作为投资者风险偏好 RP 的代理变量^[9],认为投资者风险偏好程度越高,其风险资产比重的波动性越大。本文将投资者看成一个整体,假定政策等主观影响因素已反映在数据中,选取股市中影响投资者风险偏好的指标因素,通过风险偏好计算模型得出该段股价走势下投资者的风险偏好程度,进而基于风险偏好与背离特征的约束关系建立股价预测模型。

针对股价走势与技术指标走势存在的非一致性,从特征背离角度提出了一种基于 BP(Back Propagation)神经网络^[10]的股价态势预测算法(DCPA)。DCPA 算法提取股市技术指标 MACD 的背离特征,并给出特征背离程度的计算公式,然后根据指标背离程度和股票收盘价预测股价态势。进而在 DCPA 算法的基础之上提出风险偏好分析的 RPDCA 算法,即通过考虑风险偏好对特征背离和股市态势预测的影响,给出了股市风险偏好的计算方法,并利用贝叶斯网络(Bayesian Networks, BNs)^[11]分析风险偏好、背离特征与股价走势之间的结构关系,根据风险偏好程度的变化,自适应性地调整股价预测操作。最后,在实际数据上进行实验比较与分析。

2 BP 神经网络

BP 神经网络是一种按误差逆传播算法训练的多层前馈神经网络,包括输入层、隐层和输出层。设 BP 网络结构为 3 层,输入层节点为 x_i ,隐含层节点为 y_j ,输出层节点为 z_l ,输入节点与隐含层节点间的网络权值为 w_{ji} ,隐含层节点与输出节点间的网络权值为 v_{lj} 。令输出层节点的期望值为 t_l ,则 BP 神经网络算法如下。

1) 网络初始化:给各连接权值分别赋一个区间 $(-1, 1)$ 内的随机数,设定计算精度 ϵ ;

2) 隐含层节点输出为 $y_j = f(\sum_i (w_{ji} x_i - \theta_j))$,输出层节点输出为 $z_l = f(\sum_j (v_{lj} y_j - \theta_l))$;

3) 输出层节点的误差为 $E = \frac{1}{2} \sum_l (t_l - z_l)^2$;

4) 通过输出层误差函数分别对输出层节点权值和隐含层节点权值求导 $(\partial E / \partial v_{lj}, \partial E / \partial w_{ji})$ 以修正权值,用输出层误差函数对输出层节点阈值和隐含层节点阈值求导 $(\partial E / \partial \theta_l, \partial E / \partial \theta_j)$ 以修正阈值;

5) 完成一个训练周期后,如果误差指标 $E < \epsilon$,则训练结束;否则,转到步骤 2),继续下一个训练周期。

3 背离特征股价预测算法 DCPA

指标背离是指指标和股价走势不一致。一般而言,MACD 指标的走势应与股价趋势保持一致,如果出现相反的走势,则意味着市场即将见顶或见底,指标的背离功能可以指导投资者预测风险和寻找买入机会。因此,提取 MACD 两个背离特征,结合当前收盘价建立背离特征股价预测算法 DCPA。

3.1 MACD 能量柱背离 ECD 计算

股价上升阶段的初期,MACD 能量柱和股价同步上升;在后期,常会出现 MACD 能量柱达到最高后开始下降而股价继续上升或维持横盘整理的现象,这种现象称为能量柱背离。

根据大盘走势,MACD 能量柱一般会相对股价提前若干天达到最高值,因此,能量柱的数据取值时段比股价提前 i 天。背离程度表示为:

$$MACD_{decp} = \sqrt{\frac{\sum_{k=l-i}^{m-i} (M_k - \overline{M_k})^2}{\sum_{k=l-i}^{m-i} (X_k - \overline{X_k})^2}} \cdot \frac{\sum_{i=l-i}^{m-i} |\frac{M_{j+1} - M_j}{M_j - M_{j-1}}| \cdot \overline{M_j}}{\sum_{j=l-i}^{m-i} |\frac{X_{j+1} - X_j}{X_j - X_{j-1}}| \cdot \overline{X_j}} \quad (1)$$

其中, l, m 分别表示背离阶段的起始和结束时间; i 是根据实际输入的能量柱相对股价提前达到最高值的天数; $\overline{M_k}, \overline{X_k}$ 分别表示 k 阶段内能量柱和股价数据归一化之后的均值; $\overline{M_j}, \overline{X_j}$ 类似。 $MACD_{decp}$ 取值越小,则背离程度越小。

3.2 MACD 白线背离 WD 计算

股价上升阶段的初期,MACD 的白线和股价同步上升;在后期,常会出现 MACD 的白线达到最高后开始下降而股价继续上升或维持横盘整理的现象,这种现象称为 MACD 白线背离。

白线与股价的涨跌时段几乎一致,因此二者数据取值均是 l 至 m 天内。背离程度表示为:

$$DIF_{decp} = \sqrt{\frac{\sum_{k=l}^m (D_k - \overline{D_k})^2}{\sum_{k=l}^m (X_k - \overline{X_k})^2}} \cdot \frac{\sum_{j=l}^m |\frac{D_{j+1} - D_j}{D_j - D_{j-1}}| \cdot \overline{D_j}}{\sum_{j=l}^m |\frac{X_{j+1} - X_j}{X_j - X_{j-1}}| \cdot \overline{X_j}} \quad (2)$$

其中, $\overline{D_k}, \overline{X_k}$ 分别表示 k 阶段内白线和股价数据归一化之后的均值; $\overline{D_j}, \overline{X_j}$ 类似。 DIF_{decp} 取值越小,则背离程度越小。

从 2010 年 10 月 8 日开始,股价、MACD 能量柱、MACD 白线上涨。如图 1、图 2 所示,MACD 能量柱用 MACD, MACD 表示,MACD 白线用 MACD, DIF 表示。能量柱在 10 月 19 日达到最高值后快速下降,而股价在上涨之后保持横盘整理,此时能量柱与股价走势产生背离。另外,10 月 20 日以后,MACD 白线走势也与指数走势出现背离现象。本文选取的实验样本数据周期基本是在 1 个月以内,即针对股价走势进行短期预测。

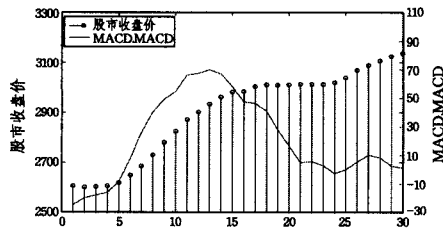


图 1 MACD 能量柱背离案例

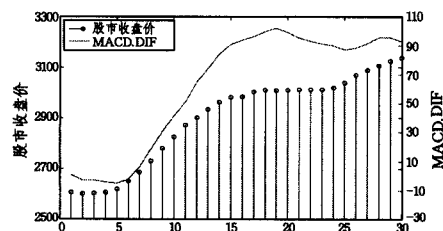


图 2 MACD 白线背离案例

3.3 DCPA 算法

DCPA 算法的基本思想是:提取一段股价走势中的 MACD 能量柱和白线背离特征,结合当前股市收盘价作为算法输入,将次日收盘价作为输出,建立基于背离特征的股价预测算法 DCPA。

算法过程可以描述如下:

1. 输入一段股价走势 $\forall x \subseteq (ECD \cap WD)$;
2. 提取背离阶段股价、MACD 能量柱和白线数据;
3. 将各数据代入式(1)、式(2),得到该段股市下 MACD 能量柱背离程度值 ECD 和白线背离程度值 WD;
4. 获取当前收盘价 CP_{now} 及下一时间收盘价 CP_{next} ;
5. 将训练样本数据的 ECD、WD、 CP_{now} 和 CP_{next} 分别作为 BP 神经网络的输入和输出进行学习;
6. 将测试样本数据代入模型,输出在 DCPA 算法下该段股票走势的 CP_{next} 值。

4 股市风险偏好程度计算

仅通过背离特征预测股价走势的准确率不高,因为特征背离程度受到市场风险偏好的影响。在股市中,风险偏好大小可以理解为对股市技术指标背离的容忍程度,其可以通过股市的相关特征来体现。本文选取了投资者关注的预测股价走势的常用特征因素,将特征因素进行筛选融合成为风险偏好的 5 个综合因素,在 BP 神经网络上利用背离特征和 5 个综合因素来预测股价走势。

4.1 股市风险偏好因素的提取和量化

1) K 线强度 KIR:长阳(实体涨跌幅不低于 2%) L_{ya} 个数,连阳 D_{lya} 天数,缺口 GAP 个数。具体如表 1 所列。

$$KIR = count(L_{ya}) + count(D_{lya}) + count(GAP)$$

表 1 K 线强度离散化

强度等级	划分范围	离散值
低强度	$0 < KIR \leq 8$	1
中等强度	$8 < KIR \leq 10$	2
高强度	$10 < KIR \leq 12$	3

2) 放量空间 VS:该段时间内最高成交额对应的股价,与上一次出现该最高值 10% 幅度范围内的成交额对应的股价相比,有一定的指导作用,如果上次比这次股价高,则预测股价之后有上涨的可能,即定义放量空间为 1,否则为 0。具体如表 2 所列。

$$VS = \langle \frac{CP}{l \leq t \leq m} | V_m = \text{MAX}(V(t)) \rangle - \langle \frac{CP}{t < l} | V = (1 \pm 10\%) * V_m \rangle$$

表 2 放量空间离散化

空间	划分范围	离散值
上涨	$VS \geq 0$	1
下降	$VS < 0$	0

3) 调整幅度 AR:用涨跌幅的方差表示调整的平稳性,方差小则表示平稳,即积聚力量后上涨或下跌的强度大。具体如表 3 所列。

$$AR = \frac{[\sum_{i=1}^m (IDR_i - \overline{IDR})]^2}{m-l+1}$$

表 3 调整幅度离散化

变化幅度	划分范围	离散值
稳定	$0 < AR \leq 1$	2
不稳定	$1 < AR \leq 2$	1
震荡	$2 < AR \leq 3$	0

4) 调整强度 AI:分别取盘整阶段前 3 个大的阴阳实体涨幅,比较加和之后的符号,符号正负表明其具有相应方向的量能。具体如表 4 所列。

$$AI = \text{MAX}3(IDR_{ya}) - \text{MAX}3(IDR_{yt})$$

表 4 调整强度离散化

能量	划分范围	离散值
正能量	$AI \geq 0$	1
负能量	$AI < 0$	0

5) 布林线中轨支撑 BS:盘整阶段 K 线实体 KTT 在中轨上方的比重大会支撑上涨,比重在 50% 以上离散值为 1,50% 以下为 0。具体如表 5 所列。

$$BS = \text{pro}(KTT \geq BOLL)$$

表 5 布林线中轨支撑离散化

支撑	划分范围	离散值
支撑上涨	$BS \geq 50\%$	1
支撑下跌	$BS < 50\%$	0

4.2 股市风险偏好的计算

根据风险偏好的因素和效用函数计算股市的风险偏好。

(1) 风险效用函数

效用函数用于描述行为和效果之间的关系^[12],不同行为主体的风险偏好不同,因而风险偏好具有多种效用函数^[13]。选取幂函数作为股市效用函数,因为其二阶导函数仍是幂函数,损益值达到一定水平时风险偏好类型不会转变,且符合股市中收益值可能为负的情况。

这里的幂函数 $U(x) = a(x+c)^b$,其中 a, b, c 分别为尺度参数、形状参数和位置参数,最大和最小收益值分别记作 x_0, y_0 ,令 $u(x_0) = 0, u(y_0) = 1$ 。对于任意 $a(0 < a < 1)$,如果存在 $[x_0, a, y_0] \sim \tau_0$,则 τ_0 满足 $x_0 < \tau_0 < y_0$,且 $u(\tau_0) = a \cdot u(x_0) + (1-a) \cdot u(y_0) = a$ 。

将影响风险偏好的 5 个指标因素作为模型输入,其中用 K 线强度 KIR 和放量空间 VS 特征来修正 x_0 ,用调整强度 AR、调整幅度 AI 和布林线支撑 BS 特征来修正 y_0 。代入方程:

$$\begin{cases} a(x_0+c)^b = 0 \\ a(\tau_0+c)^b = a \\ a(y_0+c)^b = 1 \end{cases} \quad (3)$$

可求出效用函数参数,又由 $U'(x) = ab(x+c)^{b-1}$ 和 $U''(x) = ab(b-1)(x+c)^{b-2}$ 可知:

当 $b > 1$ 时, $U''(x) > 0$, 函数曲线为凹形,决策者为风险追求型;

当 $b = 1$ 时, $U''(x) = 0$, 函数曲线是一条直线,决策者为风险中立型;

当 $b < 1$ 时, $U''(x) < 0$, 函数曲线为凸形,决策者为风险规避型。

(2) 风险偏好的计算

具体计算过程可描述如下。

1) 输入一段股价走势 $\forall x \subseteq (ECD \cap WD)$ 。

2) 提取股市指标 $\text{extract}(x) \rightarrow KIR, VS, AR, AI, BS$ 。

3) 获取该段股价收盘价最低值 CP_{min} 及最高值 CP_{max} 。

4) 修正 CP_{min} 作为未来收盘价可能出现的最小值 $x_0 = CP_{min} + \sigma_{KIR, VS}$, 具体为:

① 为消除数量级差异,将指标因素 KIR, VS, AR, AI, BS 归一化 ϵ_i ;

②利用因子分析法学习 5 个特征因素对股价走势的影响权重 $w_i (1 \leq i \leq 5)$;

③上涨阶段因素 KIR, VS 对后来股价的贡献为 $\sigma_{KIR, VS} = \sum_{i=1}^2 w_i \epsilon_i$ 。

④未来收盘价可能的最小值为该段股价最低值及特征因素贡献值的加和 $x_0 = CP_{\min} + \sigma_{KIR, VS}$ 。

⑤同理,修正 CP_{\max} 作为未来收盘价可能出现的最大值 $y_0 = CP_{\max} + \sigma_{AR, AI, BS}$ 。

⑥选取当前收盘价作为决策人对风险事件愿意接受的最保守确定性收益 $\tau_0 = CP_{now}$;

⑦取 $a = 1/2$ 的等可能涨跌趋势评价 $[x_0, a, y_0] \sim \tau_0$, 代入式(3), 求得的 b 值即为风险偏好值 RP 。

5 RPDCA 股价预测算法

利用贝叶斯网络学习节点间的结构关系, 利用非对称信息熵计算背离特征与风险偏好之间的相关程度, 从而建立特征背离和风险偏好分析的股价态势预测算法 RPDCA。

5.1 相关关系分析

在网络结构简单的情况下, 利用 K2 算法评分搜索算法: 先定义一种评价网络结构优劣的评分函数, 再从一个网络开始, 根据事先确定的最大父节点数目和节点次序, 选择分值最高的节点作为该节点的父节点。K2 算法使用后验概率作为评分函数:

$$p(D|B_i) = \prod_{i=1}^n score(i, pa_i) \quad (4)$$

$$其中, score(i, pa_i) = \prod_{j=1}^{q_i} \left[\frac{\Gamma(\partial_{ij})}{\Gamma(\partial_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\partial_{ijk} + N_{ijk})}{\Gamma(\partial_{ijk})} \right]$$

利用 K2 算法学习出背离特征、风险偏好、股价走势的贝叶斯网络结构关系, 如表 6、图 3 所示, 其中有向边表示变量之间的依赖或因果关系。

表 6 各指标对应的节点标号

序号	特征因素
1	MACD 能量柱背离
2	MACD 白线背离
3	风险偏好
4	股价走势

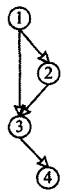


图 3 引入风险偏好后的贝叶斯网络结构

通过网络结构中边的非对称信息熵来判定背离特征与风险偏好之间的相关程度; 非对称信息熵是衡量一个节点的变化对网络其它节点的影响程度, 影响越大说明与该节点相关性越强。

非对称信息熵定义如下:

$$H_{NS}(X) = \sum_Y (-P(X \rightarrow Y) * \log(P(X \rightarrow Y)) - (1 - P(X \rightarrow Y)) * \log(1 - P(X \rightarrow Y))) \quad (5)$$

其中, X, Y 为网络中的任何节点。

利用 Koivisto 提出的 exact method 计算贝叶斯网络中边的概率, 然后利用式(5)计算出两种背离特征与风险偏好的相

关强度分别为 0.4464 和 0.3853, 由此可知二者之间相关关系较强。

经多次实验, 当风险偏好程度较强时, 背离特征对股价走势预测的作用较小, 此时风险偏好成为影响股价的主导作用; 风险偏好程度较弱时, 背离特征对股价涨跌影响较大, 此时将背离特征作为股价预测的主要成分。因此根据风险偏好程度的不同选取不同输入项来建立风险偏好和背离特征分析的股价态势预测算法 RPDCA。

5.2 RPDCA 算法

RPDCA 算法的基本思想是: 利用 4.2 节的风险偏好计算模型将当前该段股价走势的投资者风险偏好表示出来, 并在 5.1 节背离特征和风险偏好的约束关系结论下将风险偏好程度进一步划分, 当风险偏好值高于 3.5 时, 属于风险强追求状态, 风险偏好成为影响股价的主导力量, 此时将风险偏好及其组成因素作为算法输入, 结合当前收盘价 CP_{now} 预测股价未来走势; 当风险偏好值低于 0.5 时, 属于风险强规避状态, 背离特征对股价涨跌影响较大, 此时将背离特征与当前收盘价作为预测输入; 否则综合选取当前收盘价、背离特征、风险偏好及其组成因素共同作为算法的输入项。

算法过程可以描述如下:

1. 输入一段股价走势 $\forall x \subseteq (ECD \cap WD)$;
2. 提取背离特征 $extract(x) \rightarrow ECD, WD$ 和指标因素 $extract(x) \rightarrow KIR, VS, AR, AI, BS$;
3. 利用 4.2 节风险偏好计算模型求出当前股市风险偏好程度 RP ;
4. 根据风险偏好的程度选取算法不同输入项:
 - if $(RP > 3.5) x_{KIR, VS, AR, AI, BS, RP} + CP_{now} \rightarrow forecast$
 - else if $(RP > 3.5) x_{ECD, WD} + CP_{now} \rightarrow forecast$
 - else $x_{ECD, WD, KIR, VS, AR, AI, BS, RP} + CP_{now} \rightarrow forecast$
5. 输出该段股价未来涨跌趋势。

6 实验分析与比较

6.1 BP 神经网络的构建

实验收集了 2006 年 3 月 27 日到 2014 年 9 月 30 日包括上证指数等具有类似特征的数据共 112 个, 其中 56 个作为学习样本, 另外 56 个作为测试样本。为消除输入数据间数量级的差别, 避免出现因变量数量级差别较大而造成模型预测误差较大的现象, 本文对输入向量利用线性函数转化法进行了归一化处理。以第一上涨阶段和盘整阶段中相应背离特征数据、风险偏好及其组成因素、当前收盘价作为输入向量, 以第二上涨阶段的初始收盘价作为输出层创建 BP 神经网络。其中设置一个隐含层, 隐含层节点数经过多次实验、比较误差和迭代步数后确定为 12 个。网络隐含层及输出层的传递函数均为 logsig, 训练函数为 traingdx, 设置相关参数如下:

```
bpnet=newff(pr,[12 1],{'logsig','logsig'},'traingdx','learnngdm');
net.trainParam.epochs=2000;%最大训练步数 2000
net.trainParam.goal=0.001;%训练目标误差 0.001
net.trainParam.show=100;%每间隔 100 步显示结果
net.trainParam.lr=0.05;%学习速率 0.05
```

6.2 实验结果

对 4 种情况下的股价预测算法进行实验, 其中 DCPA 算法仅利用背离特征预测; DFPA 算法在此基础上加入能表示风险偏好的因素参与预测; DRPA 算法通过风险偏好计算模型将风险偏好组成因素合成为一个风险偏好节点预测; RPD-

CA 算法将风险偏好与背离特征的约束作用融入实验。

图4—图7 示出同种样本数据集下4种算法的预测结果。通过与预测前收盘价相比较得知,DCPA算法、DFPA算法、DRPA算法、RPDCA算法的预测准确率分别为58.93%、66.07%、71.43%、73.21%。这说明在背离特征下引入风险偏好及二者间的相互关系是有必要的;在RPDCA算法中,提取各训练样本风险偏好程度,其中大于3.5的样本共14个,在这些样本中,经常出现背离特征不明显而股价之后依然强势上涨的现象,此时将风险偏好及其组成要素作为算法输入;小于0.5的共9个,此时股价明显会根据背离特征出现相应涨跌,因而选取背离特征参与预测;剩下的33个样本综合利用背离特征、风险偏好及其组成要素作为输入项,即根据风险偏好程度的不同来选取不同因素作为输入指标,其在准确率4种算法中是最高的。

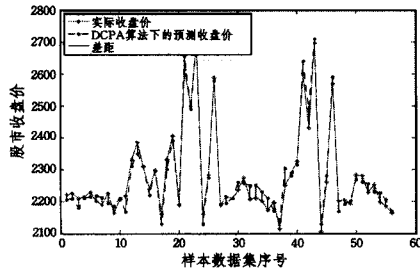


图4 DCPA算法的预测结果

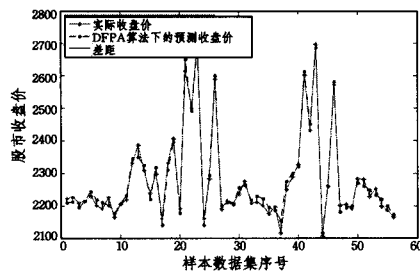


图5 DFPA算法的预测结果

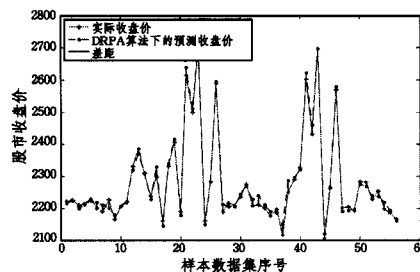


图6 DRPA算法的预测结果

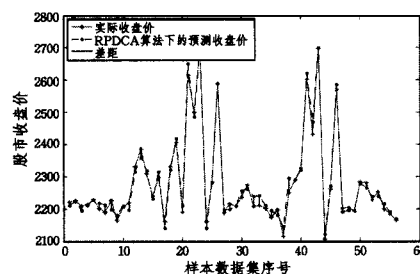


图7 RPDCA算法的预测效果

6.3 对比实验

本节将RPDCA算法与基于情感分析和技术的股价预测模型SATA^[14]进行对比,按照SATA算法要求选取与

RPDCA算法实验样本相应时段的若干博文,对博文信息进行情感特征提取,并选取相应MA、BISA、RSI指标作为模型输入,在BP神经网络下进行实验,准确率为62.5%。部分结果对比如图8所示。

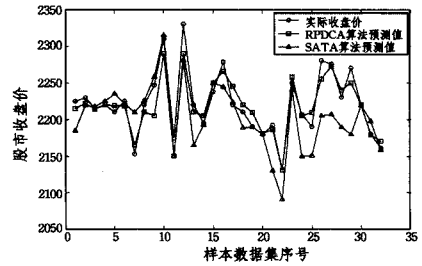


图8 两种算法实验结果对比

6.4 评价标准

均方误差:

$$MSE = \sqrt{\frac{\sum_{i=1}^n (y - y')^2}{n}} \quad (6)$$

平均绝对误差:

$$MAE = \frac{\sum_{i=1}^n |y - y'|}{n} \quad (7)$$

其中, N 表示预测集的样本个数, y 是真实值, y' 是预测值,MSE、MAE用来表示预测值偏离实际值的大小,值越小表明偏离度越小,即预测结果的精确度越高。

6.5 结果分析

在6.4节的评价标准下将本文4种情况下的预测结果及对比实验结果与实际收盘价进行对比,结果如表7所列。

表7 误差结果

标准	DCPA	DFPA	DRPA	RPDCA	SATA
MSE	26.7924	23.2018	20.0584	17.7295	21.3546
MAE	23.6942	18.0424	17.7237	14.5696	18.5781

由表7可知,利用风险偏好计算模型将风险偏好组成因素融合成一个风险偏好节点后,预测精度相比前两个算法有所提升,RPDCA预测算法考虑了风险偏好与背离特征的约束关系,预测精度是最高的,且SATA算法在准确率和精确度上均没有RPDCA算法好,结合6.2节的准确率分析可证明RPDCA算法的有效性。

结束语 基于股市中指标背离的严重性及重要性,给出MACD能量柱和白线背离特征的定义,基于背离特征建立股价预测算法DCPA;在此基础上选取股价走势中影响投资者风险偏好的若干指标,将其融合成5个风险偏好组成因素;进而将组成因素通过风险偏好计算模型融合成一个风险偏好节点;最后考虑风险偏好与背离特征的约束关系,建立风险偏好和背离特征分析的股价态势预测算法RPDCA;分别通过具体实验及相应评价标准证明了DCPA算法和RPDCA算法对短期股市预测的有效性。

参考文献

- [1] Xu Xing-jun, Yan Gang-feng. Price trend analysis based on the BP neural network[J]. Zhejiang Finance; Capital&Insurance, 2011(11):57-61(in Chinese)
- 徐兴军,颜钢锋.基于BP神经网络的股价趋势分析[J].浙江金融;证券保险,2011(11):57-61

- [2] Hsu C M. Forecasting Stock/futures Prices by Using Neural Networks with Feature Selection[C]//Proceedings of 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference(ITAIC 2011). 2011;1-7
- [3] Ma Li-li, Li Quan. Chinese Investors' Risk Preferences in China [J]. Statistical Research, 2011, 28(8): 63-72(in Chinese)
马莉莉, 李泉. 中国投资者的风险偏好[J]. 统计研究, 2011, 28(8): 63-72
- [4] Duan Jing, Liu Yong-fang, He Qi. The Effects of Decision Makers' Roles and Related Variables on Risk Preferences[J]. Acta Psychologica Sinica, 2012, 44(3): 369-376(in Chinese)
段婧, 刘永芳, 何琪. 决策者角色及相关变量对风险偏好的影响[J]. 心理学报, 2012, 44(3): 369-376
- [5] Chen Xu-hong. Extension and Application of Markowitz Mean-Variance Model[D]. Shanghai: Shanghai Jiaotong University, 2009(in Chinese)
陈许红. Markowitz 均值-方差模型的推广及应用[D]. 上海: 上海交通大学, 2009
- [6] Xu Xu-song, Chen Yan-bin. Portfolio choice model based on mean absolute deviation and it's simulated annealing algorithm[J]. Journal of Management Sciences in China, 2002, 5(3): 79-85(in Chinese)
徐绪松, 陈彦斌. 绝对离差证券组合投资模型及其模拟退火算法[J]. 管理科学学报, 2002, 5(3): 79-85
- [7] Huang Fu-kai. An Empirical Study of Downside Risk and Value Premium of A-Share Market; Based on Semi-Variance Analysis [J]. South China finance, 2014(3): 57-62(in Chinese)
黄赋凯. 我国 A 股价值溢价实证研究-基于半方差方法的分析[J]. 南方金融, 2014(3): 57-62
- [8] Hu Xiao-wen, Hui Jun. Weighted semi-variance risk measuring model for portfolio investment[J]. Journal of Hefei University of Technology(Natural Science), 2007, 30(4): 518-520(in Chinese)
- 胡小文, 惠军. 加权半方差风险度量模型[J]. 合肥工业大学学报(自然科学版), 2007, 30(4): 518-520
- [9] Chen Ju-hua, Sui Shan-shan, Wang Jian-jiang. Compensation Regulation, Risk Preference and Enterprise Investment Efficiency: An Analysis Based on Meeting Effect Theory[C]//Proceedings of the Fourth Conference on Accounting on Both Sides of the Taiwan Straits. 2012(in Chinese)
陈菊花, 隋姗姗, 王建将. 薪酬管制、风险偏好与企业投资效率: 基于迎合效应的分析[C]//第四届海峡两岸会计学术研讨会论文集. 2012
- [10] Zhao Yan. Research and Application on BP Neural Network Algorithm[C]//Proceedings of 2015 International Industrial Informatics and Computer Engineering Conference (IIICEC 2015). 2015: 1444-1447
- [11] Wang Shuang-cheng. Bayesian Network Learning, Reasoning and Application[M]. Shanghai: Lixin accounting press, 2009(in Chinese)
王双成. 贝叶斯网络学习、推理与应用[M]. 上海: 立信会计出版社, 2009
- [12] Dong Hong-hui, Jia Li-min, Tian Yin, et al. Quantitatively Analysis of Train Communication Network Based on Multi-attribute Utility Function[C]//Proceedings of 17th International IEEE Conference on Intelligent Transportation Systems I. 2014: 2330-2335
- [13] Chen Xiao-wei, Park G K. Expected Uncertain Utility Function and Its Risk Averse Analysis[C]//Proceedings of the Thirteenth International Conference on Information and Management Sciences. 2014: 305-309
- [14] Wu Jheng-long, Su Chen-chi, et al. Stock Price Predication using Combinational Features from Sentimental Analysis of Stock News and Technical Analysis of Trading Information[C]//Proceedings of International Conference on Economics, Business and Management(ICEBM2012). 2012: 39-43

(上接第 7 页)

- [52] Li Bo-duo, Edward M, Diao Yan-lei, et al. A platform for scalable one-pass analytics using MapReduce[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data(SIGMOD'11). Athens, Greece, 2011: 985-996
- [53] Blanas S, Jignesh P, Ercegovic V, et al. A comparison of join algorithms for log processing in MapReduce[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data(SIGMOD'10). Indianapolis, Indiana, USA, 2010: 975-986
- [54] Qin Xiong-pai, Wang Hui-ju, Du Xiao-yong, et al. Big Data Analysis-Competition and Symbiosis of RDBMS and MapReduce [J]. Journal of Software, 2012, 23(1): 32-45(in Chinese)
覃雄派, 王会举, 杜小勇, 等. 大数据分析-RDBMS 与 MapReduce 的竞争与共生[J]. 软件学报, 2012, 23(1): 32-45
- [55] Meng Xiao-feng, Ci Xiang. Big Data Management: Concepts, Techniques and Challenges[J]. Journal of Computer Research and Development, 2013, 50(1): 146-169(in Chinese)
孟小峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(1): 146-169
- [56] Guo J, Li Y, Du L, et al. Research on Distributed Data Mining System Based on Hadoop Platform[C]//Proceedings of International Conference on Computer Science and Information Technology. SpringerIndia, 2014: 629-636
- [57] Skillicorn D, Talia D. Mining large data sets on grids: Issues and prospects[J]. Computing and Informatics, 2012, 21(4): 347-362
- [58] Sakaeda T, Kadoyama K, Yabuuchi H, et al. Data mining of the public version of the FDA adverse event reporting system[J]. International Journal of Medical Sciences, 2013, 10(7): 796-803
- [59] Tong Y, Chen L, Yu P S. UFIMT: an uncertain frequent itemset mining toolbox[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012: 1508-1511
- [60] Weisrock D W, Smith S D, Chan L M, et al. Concatenation and concordance in the reconstruction of mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics [J]. Molecular Biology and Evolution, 2012, 29(6): 1615-1630
- [61] Koren Y. Collaborative filtering with temporal dynamics [J]. Communications of the ACM, 2010, 53(4): 89-97
- [62] Feng Deng-guo, Zhang Min, Li Hao, et al. Big Data Security and Privacy Protection[J]. Chinese Journal of Computers, 2014, 37(1): 246-258(in Chinese)
冯登国, 张敏, 李昊, 等. 大数据安全与隐私保护[J]. 计算机学报, 2014, 37(1): 246-258