

计算机基于时间序列的股价形态分析研究

汤 洋,唐典章

(北京大学 计算机研究所,北京 100871)

摘 要:提出了一种从股价时间序列中提取形态特征的股价研究方案。文中使用该系统对真实股市进行了仿真实验,并取得了很好的结果。系统可以使大家从在大量股票图线中寻找重要技术形态的繁重工作中解脱出来,有针对性地对出现这些形态的股票进行更为深入的指标技术分析和基本面分析,从而提供更为完备、更为准确的投资策略。对于与股市波动直接利益相关的个人投资者而言,自动形态分析系统为他们提供专家级的第一手信息,帮助他们捕捉最佳的买卖时机。研究的算法可以在短时间内从海量的股价数据中搜索大量的形态样本,为相关研究人员提供帮助。

关键词:计算机仿真;股价趋向预测;股市波动

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2009)10-0001-05

The Analysis of Share Price Based on Sentiment Orientation Research of Public

TANG Yang, TANG Dian-zhang

(Institute of Computer Science & Technology of Peking University, Beijing 100871, China)

Abstract: In this paper, offer a research scheme for the forecasting of share price trends based on sentiment orientation of financial information. At the present age, the impact, which caused by Internet financial information, can't be neglected increasingly on financial market. In the face of massive information, most of which are unstructured text data, adopt the current sentiment classification algorithm and then add our value of sentiment orientation to time-series model as external variable. At last, give the forecast of share price's trends and fluctuation. The result shows the relativity between financial information in Internet and the two attributes of share price which prove the scheme is effective.

Key words: computer simulation; forecasting of share price trends; fluctuation of share price

0 引 言

进入新世纪后,中国经济在飞速地发展,在人们生活水平不断提高的同时,国人的理财观点也在悄然发生着变化。随着基金、股票、理财型保险、债券等各式各样的理财产品的出现以及CPI的持续增长,人们不再认为银行是唯一可选择的理财方式,究竟什么更适合中国人的理财习惯?现代人的思维中已将收益高、见效快、风险小的现代理财模式投放在自己的理财计划中,这一理财模式也是现代社会投资理财的主要发展目标。人们通过长时间的观察发现:股票投资是可以实现人们收益高、见效快目标的理财途径之一。

证券业作为金融行业三大支柱之一,其发展成熟

程度介乎于银行与保险业之间,它相较银行业、保险业有资金流动量大、高风险等不同特点。在金融危机席卷世界各地的环境下,证券业如何起到稳定金融市场的作用,对股票投资风险绝对控制是一个迫在眉睫的问题。目前,在互联网各种工具的帮助之下,股民与证券公司已能够获得相关实时信息,并通过机器学习分析信息对股市趋势所造成的影响。现代科学技术目前已涉足金融领域,但网络金融依然是我国金融行业的新鲜事物,它的大力发展会为金融行业注入新的血液,金融行业也将会受益于此而获得质的飞跃。

股票投资其实是一种基本买卖的行为,即低价买进,高价卖出,从中赚取差价。如何判断价格的最低点与最高点成为实现股市交易理想状态的关键所在。目前股票分析方法主要分为两种:基本分析与技术分析。前者研究的方向在于股票自身的内在价值^[2],影响股票内在价值的因素来自于股票市场以外,包括国家所处的经济环境与相关经济政策、行业的发展状况和企业自身的经营状况等等。而后者研究更侧重于对股

收稿日期:2009-02-16;修回日期:2009-05-13

基金项目:国家自然科学基金(70571003);国家863高技术研究发展计划投资项目(2007AA01Z437)

作者简介:汤 洋(1988-),男,北京经济技术职业学院教师,研究方向为电子金融。

票市场内部进行研究,注重对市场本身行为的研究。在股票形势分析的相关活动中,基本分析与技术分析通常组合起来使用,二者形成分析技术的优势互补。

文中主要基于时间序列形态对股市进行研究。形态分析是一种重要的技术分析方法,它的基本思想是通过研究股票的历史价格所走过的轨迹,分析和挖掘出曲线中反映的多空双方力量的对比结果,用于支持投资决策。形态分析的研究对象被称为技术形态,它是股价曲线在某段时间内所显现出来的特定的几何形状,这些几何形状的出现往往预示着股价在未来若干天内的运行趋势和幅度。显然,如果上述论点成立,及时地从股价曲线中识别技术形态,对股票投资的意义是非常重大的。

1 技术形态研究现状

在目前股票分析领域中,股评家们根据市场分析的经验,总结出了上百种不同的技术形态,分析了这些技术形态的几何形状、判别条件和根据它们预测股价的方法。因此,形态理论研究已经得到广泛实践,大量介绍人工识别技术形态方法的文章和书籍涌现出来^[1~6]。

与基本分析相比,技术分析并不关心价值的决定因素和价格形成的原因,而只关心价格走势的最终结果,也就是股票的历史价格。技术分析家们认为,历史股价不仅反映了股票的内在价值,更重要的是反映了短期内多空双方的力量对比,这种供求关系的对比会在未来短时期内决定股价的走势。因此,技术分析主要被运用在预测股价的短期走势上。除了基本面信息影响着多空双方的买卖倾向以外,股价曲线的波动也使股民的心态随之波动,在技术分析家眼里,这种心态波动是有迹可寻的,方法就是通过历史交易数据分析多空双方的力量均衡点,以及这一均衡位置被打破的时间和价格突破的方向。不仅如此,笔者认为基本面信息决定股票价格这种单向关系也是值得商榷的,股价的变化势必会影响投资者的心态,同时也会影响公司管理层的决策行为,从而反过来对基本面信息产生作用。

技术分析在现阶段之所以盛极一时,除了在理论上可行之外,实验证明也是必不可少的。文献[7~10]对股价形态的盈利能力进行了实证分析,证明了某些技术形态确实能够为股价预测提供额外的信息,虽然这些研究并不能提供否定有效市场假说的确切证明,但技术分析的盛行并且已经帮助许多人获取了持续的高额利润的事实意味着市场并不总是有效的。在形态模式识别分析领域中,文献[7]研究了外汇汇率曲线中

的头肩(head and shoulders)形态的识别算法,汇率曲线也是一种被技术分析家们使用得非常多的金融时间序列曲线。文献[6]首先通过设置截断值(cutoff)的方法,搜索汇率曲线中的局部最大值和局部最小值,又称为峰值(peak)和谷值(trough),把这些峰值和谷值用直线连接起来就将汇率曲线转化成一条锯齿形的折线,技术形态就是通过锯齿形曲线来刻画的^[4]。然后用连续的 4 个峰值和 4 个谷值的位置关系来定义头肩形态,最后使用线性扫描的方法来搜索折线中包含的所有头肩形态。

形态理论经过长时间的发展,研究形态已经从最早的 K 线组合形态过渡到现在复杂的技术形态,形态分析也已经成为股评家们最常用的技术分析手段,然而由于技术形态是隐藏在高度复杂的非线性数据中的,并且具有很强的主观性,基于计算机的形态分析应用技术仍然是一个崭新的领域。国内最流行的股票分析软件,如:分析家、乾隆等提供的形态分析功能也仅仅止步于简单 K 线组合形态的识别,而不能自动搜索复杂的技术形态,如三角形、头肩形等。

在长期的股票分析实践过程中,股票分析家们总结出了许多对人们的投资决策有指导意义的技术形态,按照形态形成前后的价格趋势是否相同,这些技术形态可以分成两种类型:整理形态和反转形态。

整理形态表示原有的趋势暂时处于盘整状态,多空双方经过一段时间的拉锯战后,继续保持原有的平衡,股价将维持原来的趋势;

反转形态一般在主力完成大幅拉升,从而进入出货过程而形成的(顶部反转形态),或者在股价经历了一轮大跌行情之后,主力为了尽可能地在低价位吸筹,对股价进行边打压边吸货而形成的(底部反转),反转形态表示原有的平衡将被打破,股价的趋势与原来的趋势相反。

根据股价的趋势方向,又可以将整理形态分为上升整理形态和下降整理形态两类;将反转形态分为顶部反转形态和底部反转形态两类。

文中挑选了 18 种典型的技术形态作为研究对象^[4],它们是:

上升趋势中的对称三角形、下降趋势中的对称三角形、上升三角形、下降三角形、上升矩形、下降矩形、上升旗形、下降旗形、上升楔形、下降楔形、头肩顶、双重顶、三重顶、V 形顶、头肩底、双重底、三重底和 V 形底,其中前 10 种是整理形态,后 8 种是反转形态,如图 1 和图 2 所示。

技术形态最吸引人的地方主要是它们所反映的市场行为。部分学者认为^[10],技术形态是消息灵通的投

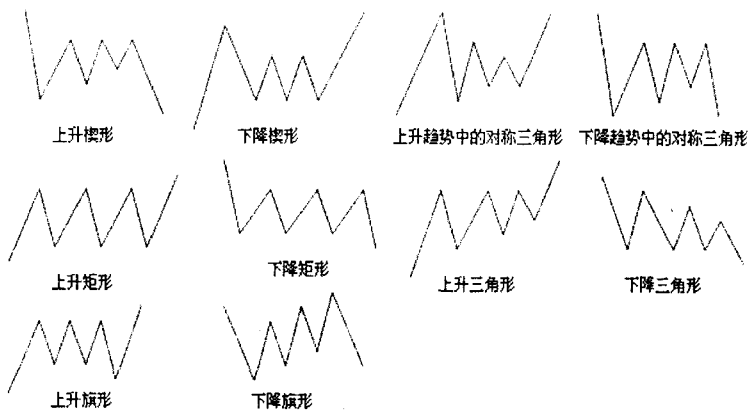


图 1 整理形态

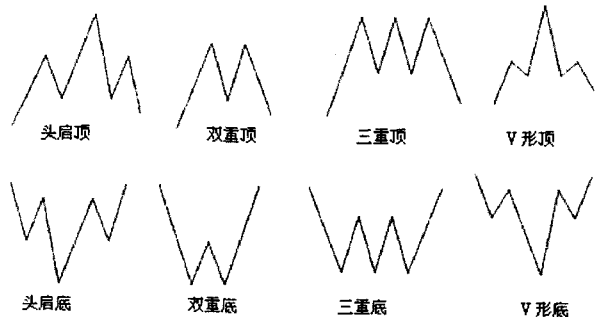


图 2 反转形态

投资者的足迹,只要人们追随这些足迹,就能增加盈利、减少损失。以上升趋势中的对称三角形形态为例,它反映了买卖双方的一个对抗过程。在形态形成之前,买方的力量占据优势,股价被拉升到某个高位,在短期的回落之后很快又被买方拉升上来,但是买方对后市的把握不大,交易量也没有明显放大,因此股价在前期高位的下方再次回落,随着成交量的进一步萎缩,股价在前期低位的上方再次上扬,于是,市场的观望气氛和买卖双方力量的均衡使得股价在一个较小的区域内来回振荡,从而形成了对称三角形形态。当买卖双方的力量发生变化后,股价最终会打破平衡,选择向上或者向下突破。

如何让技术形态的搜索更为简单迅速?如何让普通投资者也能像股票分析家一样准确地识别形态的有效算法。如何在第一时间把握可靠形态指示的投资时机?在文中研究中除了提供了自动识别技术形态的有效算法,还开发了一套股价预测系统,并用该系统对真实股市进行了仿真,取得了很好的实验结果。

2 相关工作

2.1 采集数据

实验第一步就是从实际的股价数据中寻找特定的技术形态作为实验样本。要使实验结果在统计意义上是显著的,实验样本的数目必须是相当大的。在文中

的试验中,主要借助沪深两市 A 股个股日线数据来进行相关研究。通过自己编写的一个爬虫程序将数据信息从互联网上采集下来,然后经过一些列的处理过程,如:格式整理,数据转换,将数据存储为系统所支持的数据文件。

2.2 股价时间序列化

所研究基于形态理论的技术形态实际上是股价序列的图表中反复出现的模式,股价序列的图形表示主要有 K 线图和棒线图。

这两种图表表示出每个交易日中的四个重要的股价,分别为开盘价、收盘价、最高价和最低价。在文中,主要考察股票的收盘价,它是一种典型的时间序列数据,文中后面提到的股价时间序列,如无特殊说明,就是指股票收盘价时间序列。时间序列是将某个统计指标在不同时间上的数值,按照时间先后顺序排列而成的数列。时间序列数据挖掘方法是近年来的一个研究热点,典型的时间序列分析方法有趋势分析、循环性分析、季节性分析和不规则性(随机性)分析等。形态理论将股价时间序列分解成多个趋势片断,每个片断用一条折线段来表示,这条折线段代表了股价在特定时间内的基本趋势:如果折线段向上,则表示股价在上升;如果折线段向下,则表示股价在下降。

股价时间序列可以用平面坐标轴上的一条曲线表示,如图 3 所示,横坐标表示交易日,纵坐标表示收盘价,将每相邻两个交易日的收盘价用直线段连接起来就形成如图 3 所示的一条曲线。图中是大元股份(600146)从 2002 年 8 月 28 日至 2006 年 3 月 16 日一共 1205 个交易日的收盘价曲线。

2.3 提取形态特征

通过上面部分将股价时间序列划分为趋势片断后,整个股价时间序列就由这些片断的端点唯一确定了。技术形态是由若干个连续的趋势片断组成的,以每个趋势片断为单位,使用一定长度的时间窗口可以从趋势片断序列中分割出若干片断序列,每个片断序列称为一个样本。例如,假设长度为 T 的股价时间序列可以被划分为 m 个趋势片断,此时片断序列可以表示为 (s_1, s_2, \dots, s_m) ,其中 $s_i (1 \leq i \leq m)$ 表示第 i 个片断,当时间窗口长度为 w 时,可以从这个片断序列中提取 $m - w + 1$ 个样本,分别表示为 (S_1, S_2, \dots, S_w) , $(S_2, S_3, \dots, S_{w+1})$, \dots , $(S_{m-w+1}, S_{m-w+2}, \dots, S_m)$ 。形态理论关心的是股价运行过程中有意义的高点和低点之间的位置关系,而上一章节中划分的片断的端点恰好就是股价的局部高点和低点,根据这些端点股价可以计算出描述样本的一组特征。

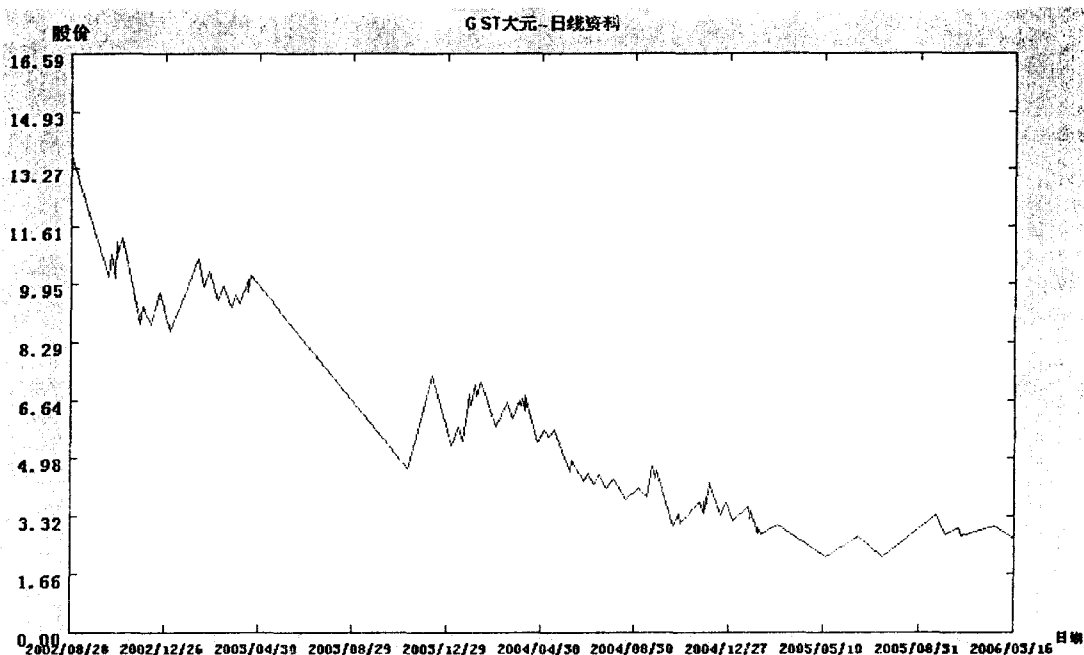


图 3 大元股份(600146) 收盘价曲线

如图 4 所示,设窗口长度为 6,将连续的 6 个片断的端点依次标记为 0,1,2,⋯,6,这 6 个片断组成一个样本,将各个端点的股价记为 $y_0, y_1, y_2, \dots, y_6$ 。一般来讲,股价数据是有偏斜的,也就是说,不同的股票有不同的价格变动范围。股价的偏斜会对形态的识别产生影响,消除偏斜的一般方法是将股价规格化到统一的区间(例如[0,1])上,另外还可以通过求比例的方法来获取股价的相对值,这正是文中所采用的办法,对长度为 6 的样本,定义如下 6 个特征:

$$p_1 = \begin{cases} 1, & \text{如果 } y_1 - y_0 \geq 0 \\ -1, & \text{如果 } y_1 - y_0 < 0 \end{cases} \quad (1)$$

$$p_i = \frac{y_i - y_{i-1}}{y_{i-1} - y_{i-2}}, i = 1, 2, \dots, 6 \quad (2)$$

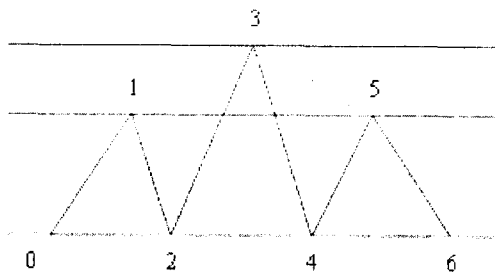


图 4 标记片断端点

下面对这 6 个特征作简要的说明: p_1 表示形态的初始趋势,如果 p_1 等于 1,表示初始趋势向上,否则表示初始趋势向下; $p_2 \sim p_7$ 表示了相邻两个片断的股价变化量的大小关系,由于相邻两个片断的方向是相反的,因此 $(y_i - y_{i-1})$ 和 $(y_{i-1} - y_{i-2})$ 的符号也相反,于是 $p_2 \sim p_7$ 都为小于零的实数,而且当 $p_i < -1$ 时,第 i 个

片断的股价变化量小于第 $i-1$ 个片断的股价变化量,当 $p_i > -1$ 时,第 i 个片断的股价变化量大于第 $i-1$ 个片断的股价变化量;给定一个长度为 6 的样本,就可以用这 6 个特征的有序集合 (p_1, p_2, \dots, p_6) 来表示该样本。

2.4 定义识别形态的规则

在目前很多相关的书籍中,都可以找到许多典型技术形态的文字描述,从这些文字描述中可以总结出形态必须满足的条件。并且通过观察发现,这些必要条件与股价的具体数额是无关的,而只是与股价的局部高点和低点(也就是趋势片断的端点)之间的位置关系有关,于是可以根据这些条件的描述,使用上面章节中定义的特征来定义识别形态的规则。

以上升趋势中的对称三角形为例,它要满足的必要条件有:

(1) 以一轮相当幅度的涨势开始,用形态特征表示就是 $p_1 = 1$ 。

(2) 多空双方在某一价格区域暂时维持平衡,股价在上升到这一区域的高点附近时就会回落,股价在下降到这一区域的低点附近时就会反弹,高点逐步下降,低点逐步上升,用形态特征表示就是 $p_2 \sim p_6$ 都小于 -1 。

同样使用这样一组特征,不同的技术分析家可能给出不同的形态识别规则,这种不同可能体现在规则的结构上,也可能体现在阈值的选取上。当然,不同的规则对同一个样本的判别结果也可能是不同的。人为设计的形态识别规则的问题在于它们的主观性和不完

备性,这两种问题都可能使得规则不能体现技术形态的本质特征,从而不能准确地识别形态。

3 试验结果

经过系统试验,生成了大量的图形结果,系统结果展示了股票在各个时期所形成的形态特征。图形结果基本能反应出各种技术形态的形态特征。例如:图 5 就显示了八一钢铁(600581) 2004 年 2 月 16 日~2004 年 4 月 8 日之间股价收盘价所形成的上升趋势中的对称三角形形态,而系统对所有的实验数据都进行了精确的形态分类。部分图文结果如以下图形所示(见图

5~7),图形对应各支股票在相应时间段所形成的形态特征。

4 结束语

通过以上试验结果可以看到,文中对股价形态的研究取得了较为理想的成果。系统能在短时间内自动从海量的股价数据中识别搜索出大量的精确形态样本,能通过精密的规则识别出肉眼所难以识别的形态,样本结果准确并且清晰。系统基于时间单位,对股价信息进行多“粒度”的形态搜索,对得到的样本形态结果集以股票代码为单位进行了精确的归类,将所有同

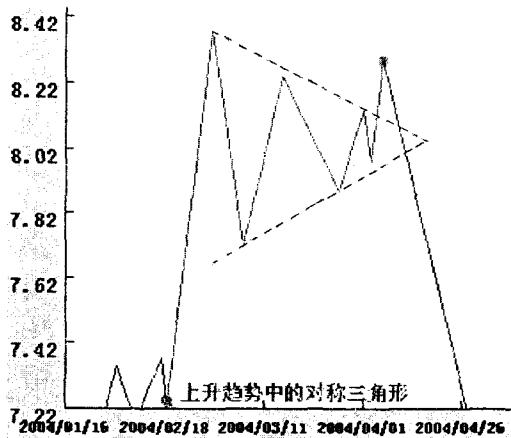


图 5 八一钢铁(600581)

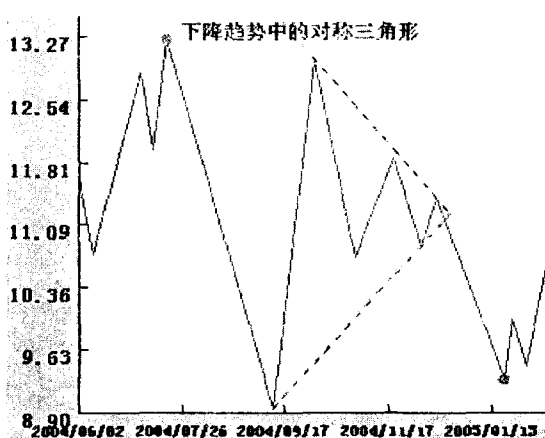


图 6 吉恩镍业(600432)

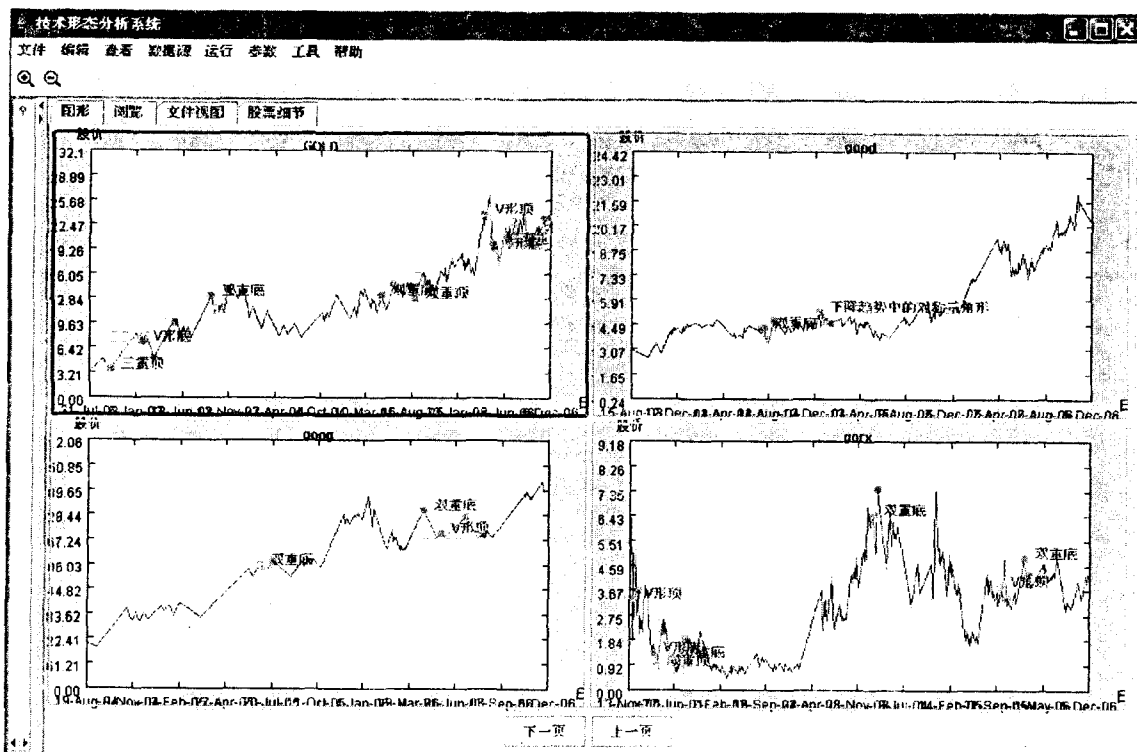


图 7 技术形态分析系统不同的形态特征的股票收盘价形态展示

(下转第 10 页)

地检测和控制网络中的 P2P 流量,且检测精度高、性能好,同时该系统网络架构可以应于多种网络环境,系统具有良好的可扩展性。

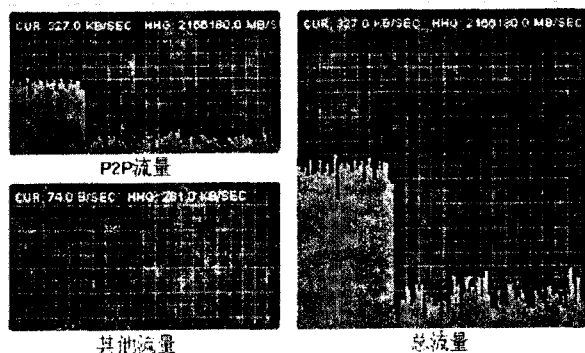


图 5 监控系统检测出的 P2P 流量

3 结束语

近年来 P2P 技术得到飞速发展,同时给网络带来了很大的负担,对 P2P 流量的识别和控制成为当前 P2P 技术研究一大热点。文中设计的 P2P 流量监控系统,采用分布式监控技术,控制策略由系统管理员根据要求制定,而控制策略的具体实施在各个终端上。该系统的实现与现有技术相比,可以避免单点监控系统性能瓶颈,能够根据要求制定控制策略,具有良好的可扩展性。

但是当前的系统还有缺点:基于 NDIS 的中间层驱动程序与操作系统结合的很紧密,进一步的工作是对其改进实现使其对不同操作系统兼容。此外,本系统只能实现对有特定的 P2P 通信特征字的 P2P 流量进行监控,因此对未知协议的 P2P 流量识别算法研

究也是下一步工作的重点。

参考文献:

- [1] 吴国庆. 对等网络技术研究[J], 计算机技术与发展, 2008,18(7):100-103.
- [2] 蒋海明,张剑英,王青青,等. P2P 流量检测与分析[J]. 计算机技术与发展,2008,18(7):74-76.
- [3] 郭兴阳,高峰,唐朝京. 一种 NDIS 中间层数据包过滤方法[J]. 计算机工程,2004,30(17):102-103.
- [4] Bittorrent[EB/OL]. 2008. <http://www.bitcomet.com>.
- [5] EDonkey[EB/OL]. 2008. <http://www.edonkey2000.cn>.
- [6] PPLive[EB/OL]. 2008. <http://www.pplive.com>.
- [7] Cheng W Q, Gong J, Ding W. Identifying BT-like P2P Traffic by the Discreteness of Remote Hosts[C]//32nd IEEE Conference on Local Computer Networks. Washington, USA: IEEE Computer Society, 2007:237-238.
- [8] Zhang Qi, Piumatti M, Singhal S K. Private Peer-to-Peer Overlay for Real-Time Monitoring of a Deployed Internet-Scale Peer-to-Peer Overlay[C]//Seventh IEEE International Conference on Peer-to-Peer Computing. Galway, Ireland: IEEE Computer Society, 2007: 235-236.
- [9] Schollmeier R, Dr. Schollmeier I G. Why Peer-to-Peer (P2P) Does Scale: An Analysis of P2P Traffic Patterns[C]//Proceedings of the Second International Conference on Peer-to-Peer Computing (P2P'02). Linköping, Sweden: IEEE Press, 2002:112-119.
- [10] Guo Lei, Chen Songqing, Zhen Xiao, A Performance Study of BitTorrent-like Peer-to-Peer Systems[J]. IEEE Journal on Selected Areas in Communications, 2007, 25(1): 155-169.

(上接第 5 页)

类别的股票信息归并到一起集中显示,这样让使用者更清晰地掌握各支股票的信息,能很大程度上协助进行相关分析。

参考文献:

- [1] 梁 循. 数据挖掘—建模、算法、应用和系统[J]. 计算机技术与发展, 2006,16(1):1-4.
- [2] 梁 循,陈 华,杨 健,等. 基于互联网股市信息量和神经网络的股价波动率预测[J]. 金融科技, 2005,5(109):92-96.
- [3] Han J, Kamber M. 数据挖掘原理与技术[M]. 北京:机械工业出版社, 2001.
- [4] 杨 健. 股票市场技术分析手册[M]. 北京:中国宇航出版社, 2002.
- [5] 梁 循. 数据挖掘算法与应用[M]. 北京:北京大学出版社, 2006.
- [6] Hagan M T, Demuth H B, Beale M H. Neural Network Design [M]. [s. l.]: PWS Publishing Company, 1995.
- [7] Osler C, Chang K. Head, shoulders: Not just a flaky pattern [M]. Staff Report No. 4, Federal Reserve Bank of New York, 1995.
- [8] Leigh W, Paz N, Purvis R. Market timing: a test of a charting heuristic[J]. Economics Letters, 2002, 77(1):55-63.
- [9] Leigh W, Modani N, Hightower R. A computational implementation of stock charting: abrupt volume increase as signal for movement in New York Stock Exchange Composite Index [J]. Decision Support Systems, 2004, 37(4):515-530.
- [10] Lo A, Mamaysky H, Wang J. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation[J]. Journal of Finance, 2000, 55(4):1705-1765.