# 基于 ABS 算法的可视化金融数据挖掘平台开发

## 戴鹏1,黄珊2

(1.同济大学 CAD 中心,上海 201804;2.长江上海通信管理局,上海 200010)

摘要:基于 GDI+显示技术和图像匹配技术的理论,研究了有关金融数据挖掘的相关技术,在 Visual Studio2008 编译环境下开发可 视化金融数据挖掘平台。为用户提供直观详尽的金融信息,并提供交互式的分析功能。

关键词:可视化;数据挖掘;ABS;交互式

中图分类号:TP311 文献标识码:A 文章编号:1009-3044(2011)13-2977-02

### Visual Financial Data Mining Platform Development Based on ABS

DAI Peng<sup>1</sup>, HUANG Shan<sup>2</sup>

(1.CAD Center of Tongji University, Shanghai 201804, China; 2.Bureau of Changjiang Shanghai Communication Administration, Shanghai 200010, China)

Abstract: Based on GDI plus and image matching theory, researched the technology about financial data mining, developed the visual financial data mining platform in Visual Studio 2008. Provided visual and particular financial data and interactive analysis for the users.

Key words: visual; data mining; ABS; interactive

金融管理研究的一个显著特点是数据分析量大、不确定性因素多,面对当今时代的海量金融数据,基于传统统计技术建立的模 型假设条件多,实际应用难以奏效。数据挖掘是 20 世纪 90 年代中期兴起的新技术,是发现数据中有用模式的过程,其目的在于使 用所发现的模式帮助解释当前的行为或预测未来的结果,以人们容易理解的形式提供有用的决策信息。如何对这些海量的数据进 行管理,并从中提取潜在的有价值的信息,成为企业在激烈的市场中占据优势地位的关键点。因此数据挖掘技术应运而生,并且在各 个行业领域中应用,取得了广泛和重大的进展。

本文通过可视化挖掘平台对海量金融数据进行有教师类型的特征提取、形成特征样本库、在此基础上运用 ABS 图像匹配算法 进行特征提取、识别,以达到对海量数据进行分析的目的。

### 1 可视化金融数据挖掘平台的总体框架

本平台采用 CDI+作为数据显示技术,GDI(Graphics Device Interface)是 Windows 系统的图形设备接口,主要负责系统与绘图程序之间的信息交换。而 GDI+作为 GDI 的继承者,对前者进行了优化,并添加了许多新功能。GDI+使得应用程序开发人员在 输出屏幕和打印机信息的时候无需考虑具体显示设备的细节,只需调用 GDI+库中类 的一些方法即可完成图形操作,阵子的绘图工作由特定的设备驱动程序来完成,使得 图形硬件和应用程序相互隔离,更加符合面向对象的编程思想。

在显示平台的基础上,本平台建立了日 K 线特征库,用户通过平台的特征识别 这一功能,对当前所显示的金融数据进行特征匹配,从中挖掘出有意义的信息,为用 户提供相应的参考意见。平台整体框架如图 1 所示。

## 2 基于 ABS 算法的图像匹配技术

#### 2.1 基干灰度的图像匹配

图像匹配是一种重要的图像分析与处理技术,常见的图像匹配技术分为两类:基

于灰度的匹配和基于特征的匹配。目前基于灰度的图像匹配技术相对较为成熟,原理也相对简单,易于硬件实现,而基于特征的图 像匹配算法较为复杂,图像噪声会对其有交大影响[1]。

灰度型算法包括 MAD(平均绝对差)算法、ABS(Absolute Balance Search,绝对差搜索)算法和 Nprod(归一化互相关)算法等;基 于特征的算法包括图像矩算法、边缘特征算法、物理特征算法等。为了提高图像的匹配精度还可以将多种类型的算法结合起来,成 为复合型算法以。算法越复杂,匹配效果越好,但计算时间也随之变长。在实际应用中,图像噪声比较明显,对匹配结果的影响较大, 因此本平台采用基于灰度的图像匹配。

## 2.2 ABS 算法

图像的灰度信息包含了图像记录的所有信息,因此基于灰度的匹配算法通常直接利用政府图像的灰度信息建立两幅图像之间

收稿日期:2011-03-06

作者简介:戴鹏(1987-),男,湖北十堰人,同济大学在读硕士,主要研究方向为 CAD。

■■■■■■■ 数据库与信息管理 ■ ■ 2977

可视化金融数据挖掘平台

图 1 平台整体框架

的相似性有关的度量,然后采用某种搜索方法寻找使该度量值最大或最小的变换模型的参数值,进而寻找匹配点 $^{\square}$ 。ABS 算法较为简单也很直接,因此计算时间也较短,这也是本平台采用此算法的重要原因。ABS 算法用实时图与基准图逐个进行像素灰度值的差运算,并用像素灰度值的差来表示两者的相关性。计算 ABS 值有四种算法,公式分别为

$$D(u,v) = \sum_{i=1}^{N_1} \sum_{k=1}^{N_2} |x_{j+u,k+v} - y_{j,k}|,$$
(1)

$$D(u,v) = \frac{1}{N_1 N_2} \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} \left| x_{j+u,k+v} - y_{j,k} \right| , \qquad (2)$$

$$D(u,v) = \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} (x_{j+u,k+v} - y_{j,k})^2,$$
(3)

$$D(u,v) = \frac{1}{N_1 N_2} \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} (x_{j+u,k+v} - y_{j,k})^2 \, . \tag{4}$$

采用 ABS 法进行图像匹配时,用极小值原则判定正确匹配点,即在度量值 D(u,v)出现极小值处认为是正确匹配点。

#### 3 日 K 线曲型特征库

#### 3.1 K 线图原理

K 线是由开盘价、最高价、最低价、收盘价所构成,也是记录买方和卖方实战的过程;如果收盘价高于开盘价则以实体红线表示,收盘价低于开盘价则以黑色实体表示,最高价和最低价则以影线表示。这样的表示方法具有直观、立体感强、携带信息量大的特点,预测后市走向较为准确,是现今应用较为广泛的技术分析手段。

#### 3.2 K 线图特征库

本平台为永和提供了 71 种典型的 K 线形态及相应的形态分析和操作建议。部分特征数据及分析如表 1 所示。

## 4 可视化金融数据挖掘平台的实现案例和案例分析

首先平台为用户提供了详尽并直观的股票信息,用户通过输入通用的股票代码来查看感兴趣的股票信息,包括了 K 线, MACD, KDJ, RSJ 等, 此外用户可通过键盘操作符对显示数据的范围清晰度进行调节, 平台右边的信息栏为用户提供了详尽的当日信息,包括开盘收盘价,成交量等常用信息。如图 2 所示。

然后用户在点击分析键 后可通过鼠标操作截取希望 通过平台分析的数据段,最后 平台会从 K 线特征库中为用 户寻找与该数据段最为匹配 的典型特征,并为用户提供相 应的数据分析和建议操作。如 图 3 所示。

## 5 可视化金融数据挖掘平 台的结论

当前有众多股票显示软件,但尚未有基于用户交互的



图 2 平台显示界面

## 表 1 K 线特征及其分析 在上升趋势中,出现一根长阳线

在上升趋势中,出现一根长阳线,在此长阳之后,出现一群实体短小的阴阳线,显示先前趋势所面临的一些压力。一般而言,这些盘整的线形大多为黑线,但最重要的一点是,这些线形的实体必须处在第一天长白线的高、低价范围内,包括影线在内。最后一根阴阳线(通常为第五天)的开盘价位于前一天收盘价之上,并且收盘价创出新高。进而维持了原来的趋势。

下降三法为上升三法的对应形态。市场处于下降趋势,一根长 黑线的出现使其跌势得到加强。随后三天则为实体短小的线形,其 走势与既定趋势相反。如果这些盘整线形的实体为白色,则情况最 佳。必须注意,这些短小的实体全部位于第一根长黑线的高、低价 范围内。最后一天开盘价应该在前一天的收盘价附近,收盘则创出 新低,宜告市场休息时间结束。

黄昏之星的情况与早 晨之星正好相反,是上升趋势出现反转的组合形态。黄昏之星的开始是一根长阳线,它加强了原有的上升趋势。第二天价格向上跳空出现新高,交易发生在小的范围内,收盘同开盘接近持平。这个小实体显示了不确定性的开始。第三天价格跳空低开,收盘更低。显示反转已经发生。



图 3 分析结果

分析平台。基于本平台用户可选择自己感兴趣的数据段进行分析,更加符合用户对于金融数据分析平台的需求,也为今后类似平台的开发提供了可行的技术路线和方法。

## 参考文献:

- [1] 祝明波.灰度相关图像匹配算法噪声适应性仿真[J].海军航空工程学院学报,2009(1).
- [2] 张红源.图像匹配经典算法及其改进方法研究[J].软件开发与应用,2008,27(11).
- [3] 冈萨雷斯.数字图像处理[M].北京:电子工业出版社,2003.
- [4] 李金迎.金融行业的数据挖掘技术研究[J].现代管理科学,2009(8).
- [5] 张婷娟.数据挖掘技术及其应用[J].山西广播电视大学学报,2008(1).