



(12) 发明专利申请

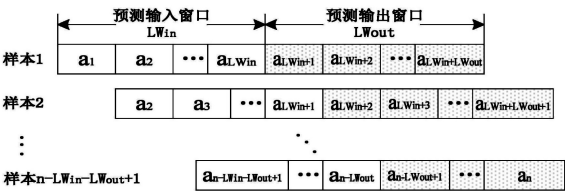
(10) 申请公布号 CN 104766137 A
(43) 申请公布日 2015. 07. 08

(21) 申请号 201510139813. 3
(22) 申请日 2015. 03. 27
(71) 申请人 中国人民解放军信息工程大学
地址 450052 河南省郑州市高新区科学大道
62 号 109 信箱
(72) 发明人 汪永伟 张红旗 杨英杰 常德显
(74) 专利代理机构 郑州睿信知识产权代理有限公司 41119
代理人 胡泳棋
(51) Int. Cl.
G06Q 10/04(2012. 01)

权利要求书2页 说明书9页 附图1页

(54) 发明名称
一种基于证据理论的网络安全态势预测方法
(57) 摘要

本发明涉及一种基于证据理论的网络安全态势预测方法,属于网络安全评估技术领域。本发明根据网络安全态势组合预测模型中各子模型算法预测得到的网络安全态势结果和实际网络安全态势构建样本集;确定样本集各子模型算法的评价指标和指标权重;利用证据理论对得到的各子模型算法中的指标权重进行融合,以得到各子模型算法的组合权重;根据得到的组合权重带入组合预测模型中进行组合预测,实现组合预测模型对网络安全态势的综合预测。本发明将能够精确刻画不同特征曲线的典型态势预测子模型进行组合,同时采用证据理论对组合模型中各子模块算法的指标权重进行权重融合,得到组合权重,实现对各子模型预测的综合,从而提高预测的精确性。



1. 一种基于证据理论的网络安全态势预测方法,其特征在于,该预测方法包括以下步骤:

1) 根据网络安全态势组合预测模型中各子模型算法预测得到的网络安全态势结果和实际网络安全态势构建样本集;

2) 根据所构建的样本集确定各子模型算法的评价指标和指标权重;

3) 利用证据理论对得到的各子模型算法中的指标权重进行融合,以得到各子模型算法的组合权重;

4) 根据得到的组合权重带入组合预测模型中进行组合预测,实现组合预测模型对网络安全态势的综合预测。

2. 根据权利要求1所述的基于证据理论的网络安全态势预测方法,其特征在于,该方法还包括在组合预测完成后计算指标可信度,采用证据折扣法对指标权重进行优化调整,以降低低可信度指标的负面影响。

3. 根据权利要求1或2所述的基于证据理论的网络安全态势预测方法,其特征在于,所述步骤2)中各子模型算法的评价指标包括相对误差、趋势拟合度和拟合稳定度,所述的指标权重包括相对误差权重、趋势拟合权重和拟合稳定性权重。

4. 根据权利要求3所述的基于证据理论的网络安全态势预测方法,其特征在于,所述步骤3)中的组合权重的确定过程如下:

A. 将得到的组合模型中各子模型算法的相对误差权重、趋势拟合权重和拟合稳定性权重作为组合模型中各子模型算法的三种权值分配方式;

B. 将上述权值分配方式转化为信任分配,得到权重证据矩阵;

C. 通过证据理论对得到的权重证据矩阵进行融合,获取的权重分配即为组合权重。

5. 根据权利要求4所述的基于证据理论的网络安全态势预测方法,其特征在于,所述步骤C中组合权重的计算公式为:

$$m(A) = \begin{cases} 0, A = \phi \\ (1-k)^{-1} \sum_{\cap A_i = A} \prod_{i=1}^n m_i(A_i), A \neq \phi \end{cases}$$

其中, $k = \sum_{\cap A_i = \phi} \prod_{i=1}^n m_i(A_i)$, $m_i(A_i)$ 对应于权重证据矩阵中的各个向量, $m(A)$ 的计算结果即为权重向量,将其赋值给 $Fw = (m_{c1}, m_{c2}, m_{c3})$ 。

6. 根据权利要求5所述的基于证据理论的网络安全态势预测方法,其特征在于,所述步骤4)的综合预测结果为:

$$P_c = m_{c1}P_1 + m_{c2}P_2 + \dots + m_{cn}P_n$$

其中 P_c 为得到综合预测结果, $P_1, P_2 \dots P_n$ 分别为组合模型中各子模型算法的态势预测值, $m_{c1}, m_{c2} \dots m_{cn}$ 分别是相应子模型算法的权重。

7. 根据权利要求5所述的基于证据理论的网络安全态势预测方法,其特征在于,步骤1)中样本集的构建是通过滑动窗口方式实现。

8. 根据权利要求5所述的基于证据理论的网络安全态势预测方法,其特征在于,所述步骤2)中相对误差权重的计算公式为:

$$Ew_i = \frac{E_{APEi}}{\sum_{j=1}^n E_{APEj}}$$

其中 Ew_i 为第 i 种子模型预测算法的相对误差权重, E_{APEi} 为第 i 种子模型预测算法的相对误差, n 为组合模型中子模型算法的个数。

9. 根据权利要求 5 所述的基于证据理论的网络安全态势预测方法, 其特征在于, 所述步骤 2) 中趋势拟合权重的计算公式为:

$$Tw_i = \frac{TF_i}{\sum_{j=1}^n TF_j}$$

其中 Tw_i 为第 i 种子模型预测算法的趋势拟合权重, 为第 i 种子模型预测算法的趋势拟合度, n 为组合模型中子模型算法的个数。

10. 根据权利要求 5 所述的基于证据理论的网络安全态势预测方法, 其特征在于, 所述步骤 2) 中拟合稳定权重的计算公式为:

$$Sw_i = \frac{FS_i}{\sum_{j=1}^n FS_j}$$

其中 Sw_i 为第 i 种子模型预测算法的拟合稳定度权重, 为 FS_i 第 i 种子模型预测算法的拟合稳定度, n 为组合模型中子模型算法的个数。

一种基于证据理论的网络安全态势预测方法

技术领域

[0001] 本发明涉及一种基于证据理论的网络安全态势预测方法,属于网络安全评估技术领域

背景技术

[0002] 通过态势评估过程所得到的态势评估结果是对当前和历史网络状况的评价。由于态势数据源信息的获取和态势信息融合处理过程都需要经过一定的时间,因此,所获得的态势评估结果总是具有一定的滞后性,态势评估结果产生的同时,网络态势早已发生了变化。从某种意义上说,态势评估的结果更多的是表现为一种“事后补偿”措施,无法达到“事前预防”的管理预期。

[0003] 态势感知是对系统状态的总体评价,是一种状态一种趋势,旨在“看现在,知未来”。因此,预知态势的未来发展趋势是态势感知的本质要求之一。态势评估过程的持续运行积累了大量的态势评估结果,为发现网络状态的演变规律,进行态势预测提供了必要的信息支持。通过态势预测,能够推演安全措施的有效性,并对网络安全状态的发展趋势进行事前预判,以辅助或指导管理人员的策略决策,提前采取应对措施,及时有效地对复杂多变的网络状态做出快速响应,“防患于未然”。因此,态势预测可有效改善态势评估“事后补偿”机制的不足。

[0004] 当前态势预测方法大多基于单一态势预测模型,一般仅能做到对具有某种特质的态势曲线进行高精度预测,在应用范围上具有一定的局限性。

[0005] 王慧强等人提出了基于遗传神经网络模型的态势预测方法,通过遗传算法动态调整参数值,寻找出最优的神经网络参数组合,使实际输出值与期望输出值趋于一致,提高态势预测的精度。

[0006] 卓颖等人提出了利用广义回归神经网络进行网络态势预测的方法,首先,历史数据进行分类,针对每类数据建立广义神经网络模型,进行态势预测;随着态势数据的变化,更新历史数据的输入和输出向量,实现对态势变化的动态预测。

[0007] Gu 等人提出了基于支持向量机 SVM 的态势预测方法,该方法首先通过遗传算法对 SVM 参数进行优化,然后通过 SVM 对态势的最优预测。

[0008] 孟锦提出了基于径向神经网络的预测方法,该方法首先采用递阶遗传算法与最小二乘法相结合的方法确定径向神经网络的结构和参数,然后通过优化的径向神经网络模型实现态势预测。

[0009] Zheng 等人提出了利用相对误差的信息熵获得组合权重、构建组合模型的方法。该方法将具备不同预测能力的单一模型进行组合,组合权重依据不同模型的信息熵进行设置,实现对态势的多模型组合预测。

[0010] Zhang 等人提出了基于小波变换、ARIMA 模型和径向神经网络模型的组合预测模型,该模型首先利用小波变换将态势预测目标转化为低频信号和高频信号,然后分布使用 ARIMA 模型和径向神经网络预测低频信号和高频信号。

[0011] Saima Hassan 等人提出了利用神经网络构建预测模型的方法,该方法首先利用神经网络进行多次训练,获得最佳神经网络模型,然后利用最佳神经网络模型构建组合预测模型。

[0012] 王慧强等人提出的基于神经网络模型的态势预测方法、卓颖等人提出的广义回归神经网络进行网络态势预测的方法、Gu 等人提出的基于支持向量机 SVM 的预测方法、孟锦的基于径向神经网络的预测方法属于典型的单一模型预测方法,都是针对目标曲线的某种特征而设计的,因此,单一态势预测模型一般仅能做到对具有某种特质的态势曲线进行高精度预测,在态势预测的适应性和应用范围上具有一定的局限性。

[0013] Zheng 等人、Saima Hassan 等人提出的预测方法建立在多种单一预测模型之上,属于多模型组合预测方法,组合权重都是依据误差精度进行调整,即依据单一指标建立组合模型,在模型的组建过程中并未考虑,稳定性、拟合性等特征。然而,在一些情况下,单一的误差精度指标不能全面反映组合模型中各子模型的性能优劣程度。

发明内容

[0014] 本发明的目的是提供一种基于证据理论的网络安全态势预测方法,以解决现有网络安全态势预测精度低以及无法适应多特征态势变化曲线的问题。

[0015] 本发明为解决上述技术问题而提供一种基于证据理论的网络安全态势预测方法,该预测方法包括以下步骤:

[0016] 1) 根据网络安全态势组合预测模型中各子模型算法预测得到的网络安全态势结果和实际网络安全态势构建样本集;

[0017] 2) 根据所构建的样本集确定各子模型算法的评价指标和指标权重;

[0018] 3) 利用证据理论对得到的各子模型算法中的指标权重进行融合,以得到各子模型算法的组合权重;

[0019] 4) 根据得到的组合权重带入组合预测模型中进行组合预测,实现组合预测模型对网络安全态势的综合预测。

[0020] 该方法还包括在组合预测完成后计算指标可信度,采用证据折扣法对指标权重进行优化调整,以降低低可信度指标的负面影响。

[0021] 所述步骤 2) 中各子模型算法的评价指标包括相对误差、趋势拟合度和拟合稳定度,所述的指标权重包括相对误差权重、趋势拟合权重和拟合稳定性权重。

[0022] 所述步骤 3) 中的组合权重的确定过程如下:

[0023] A. 将得到的组合模型中各子模型算法的相对误差权重、趋势拟合权重和拟合稳定性权重作为组合模型中各子模型算法的三种权值分配方式;

[0024] B. 将上述权值分配方式转化为信任分配,得到权重证据矩阵;

[0025] C. 通过证据理论对得到的权重证据矩阵进行融合,获取的权重分配即为组合权重。

[0026] 所述步骤 C 中组合权重的计算公式为:

$$[0027] \quad m(A) = \begin{cases} 0, A = \phi \\ (1-k)^{-1} \sum_{\cap A_i = A} \prod_{i=1}^n m_i(A_i), A \neq \phi \end{cases}$$

[0028] 其中, $k = \sum_{\cap A_i = \phi} \prod_{i=1}^n m_i(A_i)$, $m_i(A_i)$ 对应于权重证据矩阵中的各个向量, $m(A)$ 的计算结果即为权重向量, 将其赋值给 $Fw = (m_{c1}, m_{c2}, m_{c3})$ 。

[0029] 所述步骤 4) 的综合预测结果为:

$$[0030] \quad P_c = m_{c1}P_1 + m_{c2}P_2 + \dots + m_{cn}P_n$$

[0031] 其中 P_c 为得到综合预测结果, $P_1, P_2 \dots P_n$ 分别为组合模型中各子模型算法的态势预测值, $m_{c1}, m_{c2} \dots m_{cn}$ 分别是相应子模型算法的权重。

[0032] 步骤 1) 中样本集的构建是通过滑动窗口方式实现。

[0033] 所述步骤 2) 中相对误差权重的计算公式为:

$$[0034] \quad Ew_i = \frac{E_{APEi}}{\sum_{j=1}^n E_{APEj}}$$

[0035] 其中 Ew_i 为第 i 种子模型预测算法的相对误差权重, E_{APEi} 为第 i 种子模型预测算法的相对误差, n 为组合模型中子模型算法的个数。

[0036] 所述步骤 2) 中趋势拟合权重的计算公式为:

$$[0037] \quad Tw_i = \frac{TF_i}{\sum_{j=1}^n TF_j}$$

[0038] 其中 Tw_i 为第 i 种子模型预测算法的趋势拟合权重, 为第 i 种子模型预测算法的趋势拟合度, n 为组合模型中子模型算法的个数。

[0039] 所述步骤 2) 中拟合稳定权重的计算公式

$$[0040] \quad Sw_i = \frac{FS_i}{\sum_{j=1}^n FS_j}$$

[0041] 其中 Sw_i 为第 i 种子模型预测算法的拟合稳定度权重, 为 FS_i 第 i 种子模型预测算法的拟合稳定度, n 为组合模型中子模型算法的个数。

[0042] 本发明的有益效果是: 本发明根据网络安全态势组合预测模型中各子模型算法预测得到的网络安全态势结果和实际网络安全态势构建样本集; 确定样本集各子模型算法的评价指标和指标权重; 利用证据理论对得到的各子模型算法中的指标权重进行融合, 以得到各子模型算法的组合权重; 根据得到的组合权重带入组合预测模型中进行组合预测, 实现组合预测模型对网络安全态势的综合预测。本发明将能够精确刻画不同特征曲线的典型态势预测子模型进行组合, 利用不同子模型各自在态势预测上的优势, 取长补短, 同时采用证据理论对组合模型中各子模块算法的指标权重进行权重融合, 得到组合权重, 实现对各子模型预测综合, 从而提高预测的精确性。

附图说明

[0043] 图 1 是态势预测训练样本集构建示意图;

[0044] 图 2 曲线分段拟合的示意图。

具体实施方式

[0045] 下面结合附图对本发明的具体实施方式作进一步的说明。

[0046] 网络安全态势随时间并未表现出显著的规律性特征,在某些情况下表现平稳,在另外一些情况下则表现出周期变化或随机变化,因此,只用一种模型很难够精确地刻画预测过程。本发明将能够精确刻画不同特征曲线的典型态势预测模型进行综合,利用不同模型各自在态势预测上的优势,取长补短,从而获得更为精确的预测结果。证据理论具有较强的多目标融合决策能力,能够实现对综合多源信息,从而获得更为精确的决策;因此,本发明提出了一种基于证据理论网络安全态势预测方法,该预测方法包括。其具体的实施过程如下:

[0047] 1. 构建样本集

[0048] 根据网络安全态势组合预测模型中各子模型算法预测得到的网络安全态势结果和实际网络安全态势构建样本,具体的样本集构建如下:

[0049] 假设,已采集的态势样本序列为 a_1, a_2, \dots, a_n , 为了保持预测算法始终是基于最近的数据进行预测,保持采样数据的新鲜性,采用滑动窗口动态生成样本集,预测输入滑动窗口大小为 LW_{in} , 预测输出滑动窗口为 LW_{out} , 则第 1 条训练样本为 $a_1, a_2, \dots, a_{LW_{in}}, a_{LW_{in}+1}, a_{LW_{in}+2}, \dots, a_{LW_{out}+LW_{in}}$, 预测输入滑动窗口和预测输出滑动窗口依次向右滑动,完成样本集构建,样本集的构建过程如图 1 所示。

[0050] 2. 获取态势预测各子模块的评价指标和权重分配

[0051] 本发明所用到的评价指标包括相对误差、趋势拟合度和拟合稳定度,根据上述指标得到的权重分配分别为相对误差权重、趋势拟合权重和拟合稳定性权重。下面对各评价指标和权重分配计算过程进行详细说明。

[0052] 1) 相对误差和相对误差权重的计算

[0053] 态势预测一般使用误差来描述预测模型的精度,相对误差指的是预测输出值与实际输出值之间的相对差值,具体公式如下:

$$[0054] \quad E_{APE} = \left| \frac{p_i - a_i}{a_i} \right| \times 100\%$$

[0055] 其中 E_{APE} 为模型预测的相对误差, p_i 表示第 i 种子模型算法的态势预测值, a_i 表示网络安全态势的实际值。 E_{APE} 值越大,表明预测算法的预测精度越低,在组合预测算法中应为其分配越低的权重;反之,应为其分配越高的权重。

[0056] 计算出每种子模型预测算法的相对误差后,对相对误差进行归一化处理,可得每种子模型预测算法的相对误差权重 $Ew(\text{Error Weight})$ 。

$$[0057] \quad Ew_i = \frac{E_{APEi}}{\sum_{j=1}^n E_{APEj}}$$

[0058] 其中为第 i 种子模型预测算法的相对误差权重,为第 i 种子模型预测算法的相对误差, n 为组合模型中子模型的个数。

[0059] 2) 趋势拟合度及趋势拟合权重的计算

[0060] 从数学形式上看,可用连续型时变曲线 $z = f(t)$ 描述态势的演化过程。对态势数

据按照时间间隔 τ 进行离散采样,得到由采样点 (t_k, z_k) 所形成的离散型时间序列,如图 2 所示。假设, $F(i, n)$ 表示从 t_i 时刻开始的由采样点构成的折线子图。每条线段的斜率 g_k 可表示为:

$$[0061] \quad g_k = \frac{z_{k+1} - z_k}{t_{k+1} - t_k}$$

[0062] 其中, z_k 表示 t_k 时刻的态势值,依据上式,预测曲线的斜率序列 (Prediction Serial, PS) 和实际曲线的斜率序列 (Actual Serial, AS)。

$$[0063] \quad PS = \{P_i, P_{i+1}, \dots, P_{i+n-1}\}$$

$$[0064] \quad AS = \{A_i, A_{i+1}, \dots, A_{i+n-1}\}$$

[0065] 计算 PS 与 AS 转置的乘积,可得乘积向量 $MS = \{M_i, M_{i+1}, \dots, M_{i+n-1}\}$ 。即:

$$[0066] \quad MS = PS \times AS^T$$

[0067] 向量 MS 中的元素为正值表示预测曲线与实际曲线变化趋势相同, MS 中的元素为负值表示预测曲线与实际曲线变化趋势相反。MS 中的元素为正值的个数 m 越大,表明预测曲线在大多数时间的变化趋势与实际曲线相似;反之, m 越小,表明预测曲线在大多数时间的变化趋势与实际曲线相反, MS 中正值的个数从某种意义上反映了预测算法对真实值的预测能力。为此,本发明中预测序列的趋势一致量用相量 MS 中正值个数来表示,即:

$$[0068] \quad TCA = |\{\forall M_i | M_i > 0\}|$$

[0069] 其中, $|\{\forall M_i | M_i > 0\}|$ 表示乘积向量 MS 中正值元素所构成的集合, $|\{\forall M_i | M_i > 0\}|$ 为该正值集合的势,表示其中元素的个数。

[0070] 预测曲线的趋势拟合度用预测序列与实际序列中趋势相同序列个数与序列的势之间的比率来表示。

$$[0071] \quad TF = \frac{TCA}{|PS|}$$

[0072] 其中, TCA 表示乘积向量 MS 中正值个数, $|PS|$ 称为序列 PS 的势,表示序列 PS 中元素的个数。预测序列与实际曲线的趋势拟合度越高,说明预测算法的预测精确性高的可能性越大,因此在组合预测时应赋予其较高的组合权重;反之,在组合预测时为其分配较低的组合权重。

[0073] 在获得每个算法的趋势拟合度之后,对趋势拟合度进行归一化处理,可将其归一化结果作为其趋势拟合权重 Tw (Trend Fitness Weight)。

$$[0074] \quad Tw_i = \frac{TF_i}{\sum_{j=1}^n TF_j}$$

[0075] 其中 Tw_i 为第 i 种子模型预测算法的趋势拟合权重,为第 i 种子模型预测算法的趋势拟合度, n 为组合模型中子模型的个数。

[0076] 3) 拟合稳定度及拟合稳定性权重的计算

[0077] 信息熵标志着所含信息量的多少,是对系统不确定性程度的描述,可以用来衡量信息的不确定性程度。对于一组在 $[0, 1]$ 内的且满足归一化的数据,信息熵可以衡量数据

的集中程度。如果一个信息系统中 n 个不确定事件, 记为 x_1, x_2, \dots, x_n , 每个事件的发生概率记为 p_1, p_2, \dots, p_n , 则信息熵的计算公式为:

$$[0078] \quad H = -\sum_{i=1}^n p_i \log_a p_i$$

[0079] 预测序列中每个分段的相对误差构成相对误差序列 (Absolute Percent Error Serial, $E_{APE}S$):

$$[0080] \quad E_{APE}S = \{E_{APE}S_1, E_{APE}S_2, \dots, E_{APE}S_n\}$$

[0081] 根据信息熵的含义, $E_{APE}S$ 序列的信息熵越大, 预测算法对态势真实值的拟合越趋于稳定, 其对态势真实曲线的拟合能力越强。因此, 能较好的反映预测算法对目标态势曲线的稳定、持续拟合能力。

[0082] 为此拟合稳定度用 MAPES 序列的信息熵来表示, 其计算公式如下:

$$[0083] \quad FS = -\sum_{i=1}^n E_{APE}S_i \log_a E_{APE}S_i$$

[0084] 预测序列与实际曲线的拟合稳定度越高, 说明预测算法的持续精确预测能力越高, 因此, 在组合预测时应赋予其较高的组合权重; 反之, 应为其分配较低的组合权重。

[0085] 在获得每个算法的拟合稳定度之后, 对拟合稳定度进行归一化处理, 可将其归一化结果作为其拟合稳定性权重 Sw (Fitness Stability Weight)。

$$[0086] \quad Sw_i = \frac{FS_i}{\sum_{j=1}^n FS_j}$$

[0087] 其中 Sw_i 为第 i 种子模型预测算法的拟合稳定度权重, 为 FS_i 第 i 种子模型预测算法的拟合稳定度, n 为组合模型中子模型的个数。

[0088] 3. 采用证据理论对上述得到的指标权重进行权重融合, 以得到各子模型算法的组合权重。

[0089] 证据理论是通过多源证据进行融合, 获取对命题一致性描述, 以降低信息不确定的方式。

[0090] 本实施例中给出的组合预测模型中各子模型算法包括二次指数平滑、BP 神经网络模型和 ARIMA 模型, 个模型得到的预测值分别为 P_1, P_2, P_3 , 三种模型的权重分别为 w_1, w_2, w_3 , 识别框架 $\Theta = \{P_1, P_2, P_3\}$, 采样序列的长度为 n , 将权值分配转化为信任分配:

$$[0091] \quad m(P_i) = w_i \quad (i = 1, 2, 3)$$

[0092] 将三种权值分配的结果转化为信任分配, 可获得相对误差证据矩阵 E_e :

$$[0093] \quad E_e = \begin{bmatrix} Ew_{11} & Ew_{12} & Ew_{13} \\ Ew_{21} & Ew_{22} & Ew_{23} \\ \dots & \dots & \dots \\ Ew_{n1} & Ew_{n2} & Ew_{n3} \end{bmatrix}$$

[0094] 通过证据理论的组合公式对矩阵 E_e 进行融合, 可以获得更为精确的相对误差权值 $Ew_{c1}, Ew_{c2}, Ew_{c3}$ 。

[0095] 依据相对误差、趋势拟合度和拟合稳定度可以为组合算法模型中的算法赋予不同

的权值序列 $Ew_{c1}, Ew_{c2}, Ew_{c3}, Tw_1, Tw_2, Tw_3, Sw_1, Sw_2, Sw_3$ 。三种权值分配方案从不同的角度评价了子模型的预测精度。类似于 Ew 的处理,将三种权值分配结果转化为信任分配,可得组合权重证据矩阵 E_c :

$$[0096] \quad E_c = \begin{bmatrix} Ew_{c1} & Ew_{c2} & Ew_{c3} \\ Tw_1 & Tw_2 & Tw_3 \\ Sw_1 & Sw_2 & Sw_3 \end{bmatrix}$$

[0097] 通过证据理论对证据矩阵 E_c 进行融合,可以获得更为精确的权重分配 $Fw = (m_{c1}, m_{c2}, m_{c3})$,该分配结果即为组合权重。

$$[0098] \quad m(A) = \begin{cases} 0, A = \phi \\ (1-k)^{-1} \sum_{\cap A_i = A} \prod_{i=1}^n m_i(A_i), A \neq \phi \end{cases}$$

[0099] 其中, $k = \sum_{\cap A_i = \phi} \prod_{i=1}^n m_i(A_i)$,【说明,这里共有 3 个证据源,故 $n = 3$; $m_i(A)$ 对应于矩阵中的 $Ew_{c1}, Ew_{c2}, Ew_{c3}, Tw_1, Tw_2, Tw_3, Sw_1, Sw_2, Sw_3$, $m(A)$ 的计算结果即为权重向量,将其赋值给 $Fw = (m_{c1}, m_{c2}, m_{c3})$ 。

[0100] 4. 将得到组合权重带入组合预测模型进行组合预测,得到的预测结果即为本发明采用组合预测模型得到综合预测结果。

$$[0101] \quad P_c = m_{c1}P_1 + m_{c2}P_2 + m_{c3}P_3$$

[0102] 其中 P_c 为得到综合预测结果, P_1, P_2 和 P_3 分别为采用二次指数平滑、BP 神经网络模型和 ARIMA 模型的预测值, m_{c1}, m_{c2} 和 m_{c3} 分别是上述模型的权重。

[0103] 上述过程的伪码描述如表 1 所示。

[0104] 表 1

[0105]

算法 1. 基于 SFM_ETO 的组合预测算法 SFM_ETO_A

输入：样本集

输出：预测值

```

(1)  BEGIN
(2)    Set ( $LW_{in}$ ,  $LW_{out}$ ); //设定滑动窗口值
(3)    Set (ForecastStep); //设定预测步长
(4)    For  $i=1:n-LWin-LWout+1$ 
(5)      SFvector=SForecast();//二次指数平滑法训练
(6)      NFvector=NForecast();//BP 神经网络预测模型训练
(7)      AFvector=AForecast();//ARIMA 模型训练
(8)    End for
(9)    For  $i=1$ : Forecast
(10)     Ee=CalculateEe(SFvector, NFvector, AFvector, realValue);//计算相对误差矩阵
(11)     Ew=DSFusion(Ee);//利用证据理论计算相对误差权重
(12)     TF=CalculateTF(SFvector, NFvector, AFvector, realValue);//计算趋势拟合度;
(13)     Tw=CalculateTw(TF);//计算趋势拟合权重;
(14)     FS=CalculateFS(SFvector, NFvector, AFvector, realValue);//计算拟合稳定度;
(15)     Sw=CalculateSw(FS);//计算拟合稳定性权重;
(16)     Fw= DSFusion(Ew, Tw, Sw);//利用证据理论计算组合权重
(17)     Result=CombineForecast(Fw, SF_result , NF_result, AF_result);//进行组合预测
(18)   End for
(19) End

```

[0106] 本发明采用的相对误差、拟合一致度和拟合稳定度指标对刻画预测模型的性能具有一定的普遍性。但是,由于态势曲线的变化是不确定的,因此,在一些情况下,某些指标对态势曲线的表达会出现较大偏差,从而导致较差的预测精度。

[0107] 为此,在上述预测方法基础上,本发明为了进一步提高预测的精度,在新的预测周期中采用 Shafer 的证据折扣法对指标权重进行优化调整。

[0108] 假设,进行了 n 次精确组合预测(达到设定的精度阈值),若依据某指标获得的权重分配与组合权重分配有 m 次保持一致,则该指标的指标强度可表示为:

$$[0109] \quad IS = \frac{m}{n}$$

[0110] 指标的强度表示了对精确预测的支持程度。指标强度越高,该指标在组合模型中的地位越重要,可信程度越高;反之,该指标在组合模型中的地位越重要,可信程度越低。将指标强度进行归一化处理,以得到指标可信度 IC。

$$[0111] \quad IC_i = \frac{IS_i}{\sum_{i=1}^k IS_i}$$

[0112] 指标可信度较好的刻画了预测指标对预测子模型的评价能力。态势预测模型的优化演进过程应弱化低可信指标的权重分配结果,尽可能的降低低可信指标的负面影响。在组合模型中,指标权重是以证据的形式进行组合的,Shafer 提出的证据折扣法可有效降低低可信证据在融合结论中的影响。因此,每次组合预测完成后,计算证据可信度,在新的预测周期中采用 Shafer 的证据折扣法对指标权重进行优化调整。证据折扣法的表达式如下所示:

$$[0113] \quad \begin{cases} m_i'(A) = w_i m_i(A), \\ m_i'(\Theta) = 1 - w_i + w_i m_i(\Theta), \end{cases} \quad \forall A \subset \Theta$$

[0114] 其中, $w_1, \dots, w_i, \dots, w_n$ 表示证据 i 的权重,可用指标可信度来代替; $m_i(A)$ 表示证据中焦元的信任度,可用指标权重代替。

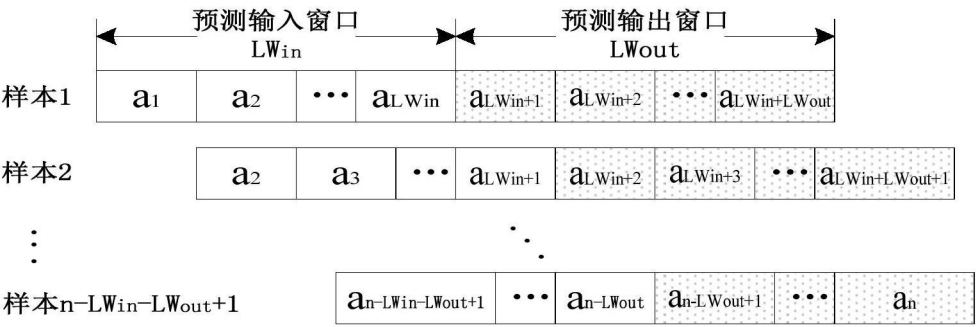


图 1

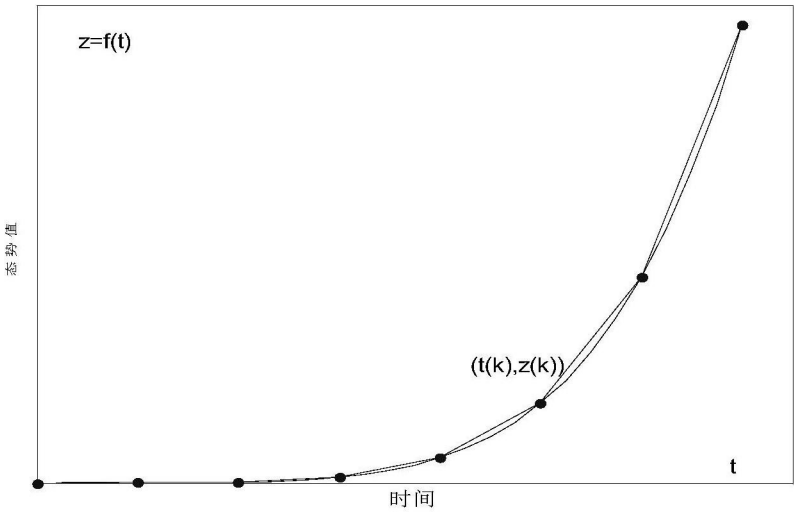


图 2