

稀疏贝叶斯方法在通信告警序列预测中的应用研究*

李彤岩^{1,2}, 李兴明¹

(1. 电子科技大学 宽带光纤传输与通信网技术教育部重点实验室, 成都 610054; 2. 成都信息工程学院, 通信工程系, 成都 610225)

摘要: 针对通信网告警中预示重大故障的告警数量少、不适合用传统预测方法的特点, 提出了一种基于稀疏贝叶斯的通信告警序列预测方法(PBM), 并与支持向量机(SVM)预测方法进行了比较。实验结果表明, PBM 方法非常适用于小样本的通信告警预测, 其不仅具有 SVM 的预测性能, 而且在样本数目增加时的预测误差率要小于 SVM, 具有非常好的预测精度。

关键词: 告警序列; 决策函数; 核函数; 稀疏贝叶斯; 预测精度

中图分类号: TP393 **文献标志码:** A **文章编号:** 1001-3695(2010)04-1427-03

doi:10.3969/j.issn.1001-3695.2010.04.061

Sparse Bayesian method applied to telecommunication alarm sequences forecasting

LI Tong-yan^{1,2}, LI Xing-ming¹

(1. Key Laboratory of Broadband Optical Fiber Transmission & Communication Networks of Ministry of Education, University of Electronic Science & Technology of China, Chengdu 610054, China; 2. Dept. of Communication Engineering, Chengdu University Information Technology, Chengdu 610225, China)

Abstract: The alarm which indicates major failure in communication networks has a small number, in this case it is not suitable for traditional forecasting methods. This paper proposed a PBM, and compared it to SVM estimate method. The results show that, PBM is applicable to the small sample of telecommunication alarms predict. It not only has the predict performance of SVM, but also has lower predict errors than SVM when the number of samples increase, even it has a very good prediction accuracy.

Key words: alarm sequences; decision function; kernel function; sparse Bayesian; prediction accuracy

在通信网告警中, 出现重大故障的告警数量其实并不多, 如果使用关联规则挖掘算法中的寻找频繁项的方法^[1,2], 一定会将这些信息遗漏掉。而告警预测模式又是一类非常重要的相关性知识, 某些告警本身就代表了网络故障的产生, 如链路中断、交换机关闭等, 如果在这些告警产生之前就捕捉到其预测模式, 并采取一定的措施, 就有可能避免故障的产生。在处理告警数据时, 常常使用时间窗口来消除大量出现的冗余告警, 而表示故障的告警出现次数也不会很多, 因此用常规的挖掘方法就难以找到期望的预测模式。告警预测模式的挖掘实际上就是小样本学习过程, 通过有限的样本来达到最佳的学习效果。预测方法的研究在避免重大故障产生、维护系统性能方面有着重要的作用。一些时间序列预测的方法有: a) 2002 年 R. Vilalta 等人^[3]提出了基于序列模式挖掘的告警预测模式挖掘方法 APPM-SPM, 其思想主要是运用关联规则挖掘算法找出在目标告警即故障(表示为 target)发生之前的时间窗口内出现的频繁事件, 然后确定预测规则。这种方法运用在通信告警预测中的问题在于, 表征通信重大故障之前的告警信息并不一定是频繁发生的, 利用这种预测方法仍然会遗漏掉一些非频繁出现但又预示着重大故障的告警事件; b) 基于神经网络的时间序列预测方法。这种方法能够较好地处理非线性问题, 具有灵活的学习方式, 但是神经网络难以解决系统复杂性和推广性的矛盾, 网络结构过于复杂或简单都会影响到预测精度和效率; c) 基于支持向量机(support vector machine, SVM)^[4,5]的方

法可满足结构风险最小化原则, 即使在较少的样本集上, 训练也可得到很好的预测效果, 对小样本的学习效率很高, 但是这种方法会产生较多的核函数, 复杂度也较高。

针对以上预测方法存在的问题, 本文提出了一种应用贝叶斯理论的告警预测方法(prediction with Sparse Bayesian method, PBM)。告警序列是一类典型的时间序列, 利用贝叶斯线性回归分类法对时间序列进行预测, 不仅具有 SVM 的避免过学习的优点, 而且还比 SVM 方法具有更高的预测精度。这种方法针对小样本学习也非常有效。

1 告警时间序列的数学模型

本文研究的是告警序列, 首先需要构造出告警预测的数学模型。文献[3,6,7]中介绍了告警预测数学模型的建立, 其中文献[7]将告警序列划分成三个时间窗, 将一条告警看做时间 E 在某时刻 t 的一次发生, 用 E_t 表示, 历史告警数据就是由单个告警组成的有序事件序列 $S = \{E_{t_1}, \dots, E_{t_n}\}$, $t_1 \leq t_i \leq t_n$, $i = 1, \dots, n$ 。同时定义 $S(t_1, t_2)$ 是 S 中发生在 t_1 和 t_2 之间的告警组成的子序列。数学模型中, 告警的预测可分为三个阶段: a) 检测期, 长度为 M , 系统记录出现在检测期的所有告警, 将其与事先建立的告警预测模型进行模式匹配, 若匹配成功, 则在 t 时刻预测的 target(在告警的故障预测模型中, target 表示故障)将会发生; b) 预警期, 长度为 W , 作用是在 target 发生之前留

给一定的时间采取必要的保护措施;c) 预测期, 长度为 Q , 就是预测 target 所应该出现的时间区间, 精确到时间点的预测可以通过调整预警期和预测期的长度实现。如果时刻 t 的预测结果与预测期的实际情况相吻合, 则是一次正确的预测, 否则是一次错误的预测。模型如图 1 所示。



图1 告警预测模型

假设 $E_{t_1}, E_{t_2}, \dots, E_{t_n}$ 是发生在 $t-M$ 时刻与 t 时刻之间的告警序列 $S(t-M, t)$, 将 target 的预测归纳为如下二分类的决策函数:

$$f(E_{t_1}, E_{t_2}, \dots, E_{t_n}) = \begin{cases} 1 & \text{target} \in S(t+W, t+W+Q) \\ 0 & \text{target} \notin S(t+W, t+W+Q) \end{cases} \quad (1)$$

预测模式挖掘的任务就是基于历史告警数据构造出决策函数 f , 使其有尽可能好的预测性能。

2 构造训练样本

要构造固定维数的特征向量, 采用 RICE 的方法^[8], 用 $(tf_1, tf_2, \dots, tf_n)$ 表示输入样本的特征。其中: $tf_i = \text{freq}(E_i)$, E_i 表示第 i 种告警类型 ($i = 1, \dots, n$), $\text{freq}()$ 表示某类告警在监测窗内发生的次数。在滑动窗口中同一种告警出现多次只算一次, 用于消除冗余告警信息, 这样就将样本输入确定为 n 维向量。对 target 发生与否采用哑变量编码, 发生就记为 1, 反之则为 0。 $(tp_1, tp_2, \dots, tp_m)$ 表示输出样本的特征向量, $tp_j = Y(F_j)$, $Y()$ 表示哑变量值; F_j 表示第 j 种 target 的类型, $j = 1, \dots, m$ 。提取样本值的过程举例如图 2 所示, 其中假定 E 为 target。

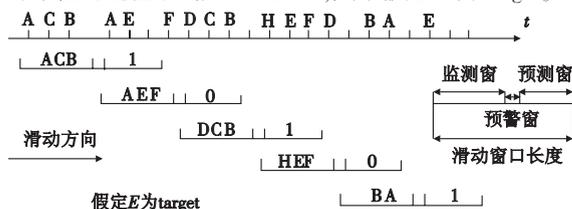


图2 滑动窗口取样

根据在检测期和预测期内的告警数据观测, 可以确定出训练样本值 (x_i, y_i) , $i = 1, \dots, N$ 。其中: x_i 是第 i 个输入样本的特征向量; y_i 为第 i 个样本的响应值 0 或 1, 通过使用这些训练数据来构造预测规则。要从数据的训练过程中找出规则, 不但要求构造一定数目的正样本, 还要构造一定数目的负样本, 从而找出一个决策函数, 使得输入一组新的数据后, 可以根据构造的预测规则来推断出输出值。在告警时间序列里, 按照上面的构造过程, 将 target 不发生的样本即响应值为 0 的数据作为负样本, 取样过程同图 2, 通过滑动窗口的滑动就可以得到一组正负样本值 $(ACB, 1), (AEF, 0), \dots, (BA, 1) \dots$ 。

3 稀疏贝叶斯方法训练样本

3.1 贝叶斯方法

在贝叶斯方法中, 根据数据模型中的未知量的先验分布, 利用先验知识和似然函数来估计给定数据后的后验分布。

根据前面给定的观测数据 $x = [x_1, \dots, x_n]$, 用 θ 表示产生数据模型的参数向量, 则 x 服从条件分布 $f(x|\theta)$ 。

3.2 稀疏贝叶斯线性回归模型

假定输入向量 $X = (E_{t_1}, E_{t_2}, \dots, E_{t_n})$, 并且希望预测实数

值输出 Y 。线性回归模型如下:

$$f(X_n, \beta) = \sum_{j=1}^m \varphi_j(X_n) \beta_j = \beta^T \varphi(X_n) \quad (2)$$

$$Y_n = f(X_n) + \beta_0 \quad (3)$$

其中: β 是未知的可调参数, 回归函数 $E(Y|X)$ 是线性的; 变量 $\varphi(X_n)$ 是非线性的基函数, Y_n 是非线性目标输出, f 是决策函数; β_0 是与 $f(X_n)$ 相互独立的噪声, 假设为服从零均值的高斯分布。

给定一系列输入向量和相应的输出, 通过给定的训练数据和先验知识来设计一个预测系统, 使得设计出的系统对新的输入 X , 可以预测出其输出 Y 。训练过程就是确定参数 β 的过程。

用 $\varphi(X) = (\varphi_1(X), \dots, \varphi_n(X))^T$ 表示一组非线性基函数, 可看做是特征向量。 $f(X_n, \beta)$ 是基函数 $\varphi(X_n)$ 的线性组合。核回归函数表示为

$$\varphi(X) = [1 \quad K(X, X_1) \dots K(X, X_n)]^T \quad (4)$$

其中: 内积函数 $K(X, X_i)$ 表示核函数。

训练数据学习的目的是为了得到参数 β , 而经典的 Bayes 方法是求 β 的点估计。由于训练数据和参数很多, 就不可避免产生过学习现象。采用稀疏 Bayes 方法为了使其稀疏化, 就给参数加上一个约束函数, 即认为 β 的先验分布服从 Laplace 分布:

$$P(\beta|\alpha) = \prod_{i=1}^k \frac{\alpha}{2} \exp[-\alpha |\beta_i|] = \left(\frac{\alpha}{2}\right)^k \exp[-\alpha \|\beta\|_1] \quad (5)$$

其中: α 为假设的参数向量, $\|\beta\|_1 = \sum_j |\beta_j|$ 是 β 的一次范数。

运用稀疏 Bayes 的过程可以使大部分的参数趋于零, 其相应的核函数和训练样本被剔除, 只保留了少数的参数不为零, 保留下来的训练样本被称为相关向量, 类似于支持向量量中的支持向量。此外, 采用稀疏 Bayes 方法可以得到分类或回归结果的后验概率, 并根据训练样本值的变化来自动进行调整。

4 PBM 的算法实现

时间序列经过数据预处理, 然后根据前面的时间窗口滑动得到一定数量的正负样本数据。将时间序列样本分成两部分, 前一部分作为预测模型的训练样本, 其余部分作为测试数据来检验模型的正确性和有效性。PBM 算法的步骤如下:

- a) 输入一部分告警时间序列样本。
- b) 构建稀疏贝叶斯预测模型。
- c) 将正负样本数据输入预测模型中进行训练, 以确定参数 β 。
- d) 利用构造好的预测模型对剩下的一部分告警样本值进行检验, 并按照评价标准对预测精度及准确度进行评判。

预测模型训练过程是算法的核心, 本文拟采用迭代法来获取参数 β 。参数训练过程如下:

a) $\varphi(X) = [1 \quad K(X, X_1) \dots K(X, X_n)]^T$, $X_i = [x_i, x_{i+1}, \dots, x_{i+m-1}]$ ($i = 1, 2, \dots, n-m$; n 为训练样本的数目), $Y = [y_1, y_2, \dots, y_n]^T$, 令 $\Phi = [\varphi^T(X_1), \dots, \varphi^T(X_n)]^T$ 。

b) 由式 (2) ~ (4) 可将参数 β 初始化为 $\hat{\beta}^{(0)} = (\hat{\beta}_0 E + \Phi^T \Phi)^{-1} \Phi^T Y$, 令 $\hat{\beta}_0 = 10^{-6}$ 。

c) 定义参数矩阵 $M = \text{diag}(|\hat{\beta}_1^{(t)}|, \dots, |\hat{\beta}_n^{(t)}|)$ 。其中 n 表示训练的序列样本数目。

d) 迭代产生新的参数 $\hat{\beta}^{(t+1)} = M(E + M\Phi^T\Phi M)^{-1} M\Phi^T Y$ 。

e) 如果 $\|\hat{\beta}^{(t+1)} - \hat{\beta}^{(t)}\| / \|\hat{\beta}^{(t)}\| < \delta$ 则证明训练结束, 得到的参数即为所求值, 否则就继续执行步骤 c)。

为了评价模型的预测性能, 需要计算预测的准确性。这里

采用文献[8]介绍的均方根相对误差 (root mean square error, RMSE) 模型对预测结果进行统计:

$$e = \sqrt{\left(\frac{1}{N} \sum_{i=1}^N \sum_{d \in [u, v]} (u_{di} - v_{di})^2\right)} \quad (6)$$

其中: N 表示测试样本数目; u_{di} 表示模型的预测值; v_{di} 是时间序列的真实值。

5 PBM 与 SVM 预测方法的比较

实验分别选取某地区通信网络中一周的通信告警数据, 分别根据不同的监测窗、预警窗和预测窗对告警交易取样, 并且用两种预测方法分别建模。

表 1 显示了取样的过程。设滑动窗口为 3S, 预测窗口、预警窗口、预测窗口分别为 3S、0.5S、1.5S, 得到一组样本值。取其中 2/3 样本作为训练集, 1/3 样本作为测试集, 分别针对两种预测方法 SVM 和 PBM 进行预测误差率的测试, 结果如表 2 所示。

表 1 告警取样值

原始告警	1 000	2 000	3 000	4 000	5 000	10 000
交易数目	350	715	1 064	1 372	1 739	3 538
正样本数目	64	86	105	119	138	234
负样本数目	286	629	959	1 253	1 061	3 304

表 2 SVM 和 PBM 的预测误差率比较

样本数	10	20	30	40	50	60
SVM	0.289	0.167	0.059	0.032	0.014	0.010
PBM	0.304	0.237	0.036	0.029	0.009	0.007

从表 2 可以看出, 样本数目较小的时候 SVM 的预测精度略好于 PBM, 当样本数目增加, 则 PBM 的预测误差率明显小于 SVM, 说明 PBM 的适用范围要比 SVM 广泛, 并且可以达到 SVM 的预测效果。

6 结束语

通信网中有些预测重大故障的告警数量并不多, 对这类告

警序列进行预测属于小样本学习过程。本文提出了一种新的基于稀疏贝叶斯的预测方法 PBM, 对通信告警序列进行了建模研究, 并比较了 PBM 和经典预测方面 SVM 的预测误差率。实验结果表明: PBM 方法可以用于时间序列的预测, 并且具有较高的预测精度, 是一种适用于通信网告警的告警预测方法。

参考文献:

- [1] HAN J, PEI J, YIN Y, *et al.* Mining frequent patterns without candidate generation: a frequent-pattern tree approach [J]. *Data Mining and Knowledge Discovery*, 2004, 1(8):53-87.
- [2] PI De-chang, QIN Xiao-lin, GU Wang-feng, *et al.* STBAR: a more efficient algorithm for association rule mining[C]//Proc of the 4th International Conference on Machine Learning and Cybernetics. Guangzhou:[s.n.], 2005: 1529-1533.
- [3] VILALTA R, APTE C V, HELLERSTEIN J L, *et al.* Predictive algorithms in the management of computer systems [J]. *IBM Systems Journal*, 2002, 41(3):461-474.
- [4] TAY F E H, CAO L J. A comparative study of saliency analysis and genetic algorithm for feature selection in support vector machines [J]. *Intelligent Data Analysis*, 2001, 5(3):191-209.
- [5] BURGESS C. A tutorial on support vector machines for pattern recognition [J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121-167.
- [6] WEISS G M, HIRSH H. Learning to predict rare events in events sequences[C]//Proc of the 4th International Conference on Knowledge Discovery and Data Mining. [S.l.]: AAAI Press, 1998: 359-363.
- [7] 单莘, 朱永宣, 郭军. 基于支持向量机的网络告警预测知识发现 [J]. *微电子学与计算机*, 2007, 24(6):35-37.
- [8] WILLIAMS O, BLAKE O, CIPOLLA R. Sparse Bayesian learning for efficient visual tracking [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2005, 27(8):1292-1304.

(上接第 1411 页)

图 3 中假设 $m = 10, p = 0.6, N$ 是变化的, 可以看出, 随着网络中节点的增加, EC-MAC 协议比其他两种协议的能耗都要小, 从而 EC-MAC 协议适合于中大规模的网络。图 4 中假设 $N = 10, p = 0.6, m$ 是变化的, 可以看出, 随着 m 的增大, EC-MAC 协议比传统的 TDMA 和 BMA 更节能。

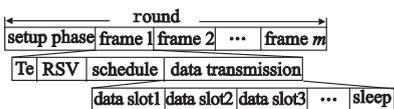


图 1 EC-MAC 协议的帧结构

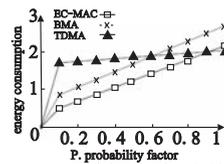


图 2 能量消耗和 p 的关系图

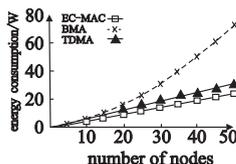


图 3 能量消耗和 N 的关系图

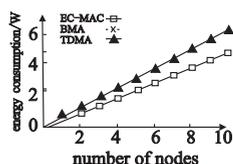


图 4 能量消耗和 m 的关系图

综合图 2~4 结论可以得出, EC-MAC 协议更适合工作在中等规模的无线传感器网络中。

4 结束语

考虑到无线传感器网络的应用环境, 提出了基于簇的节能

MAC 协议, 仿真实验证明, 在中等负载的情况下 EC-MAC 协议比传统的 TDMA 和 BMA 要节能。下一步研究如何将 MAC 层协议与无线传感器网络中其他层结合在一起, 设计基于跨层的数据链路层协议。

参考文献:

- [1] 姜华, 袁晓兵, 王鹏, 等. 无线传感器分簇网络中 MAC 协议的研究 [J]. *计算机应用研究*, 2007, 24(3):265-270.
- [2] BAI Rong-gang, Qu Yu-gui, GUO Yang, *et al.* An energy-efficient TDMA MAC for wireless sensor networks[C]//Proc of the 2nd IEEE Asia-Pacific Services Computing Conference. Washington DC: IEEE Computer Society, 2007:69-74.
- [3] 龚海刚, 刘明, 余昌远, 等. 无线传感器网络环境下基于事件驱动应用的节能 TDMA 协议 [J]. *电子学报*, 2007, 35(10):1843-1848.
- [4] HEINZELMAN W R, CHANDRAKASAN A P, BALAKRISHNAN H. An application-specific protocol architecture for wireless microsensor networks [J]. *IEEE on Wireless Communications*, 2002, 10(4): 660-670.
- [5] LI J, LAZAROU G Y. A bit-map-assisted energy-efficient MAC scheme for wireless sensor networks[C]//Proc of the 3rd International Symposium on Information Processing in Sensor Networks. New York: ACM Press, 2004:55-60.
- [6] RAGHUNATHAN V, SCHURGERS C, PARK S, *et al.* Energy-aware wireless microsensor networks [J]. *IEEE Signal Processing Magazine*, 2002, 19(2):40-50.