

# 基于 Bayes 网的时间序列预测<sup>\*</sup>)

秦 鹏 邢永康 杨华丽

(重庆大学计算机科学与技术学院 重庆 400044)

**摘 要** 时间序列预测问题在气象、天文、电力、医学、生物、经济、金融和计算机等各个领域有着广泛的应用。本文将 Bayes 网模型用于该领域,提出并建立了一套基于 Bayes 的时间序列预测模型——静态 Bayes 网预测模型,动态 Bayes 网预测模型和分类静态 Bayes 网预测模型。实验表明,这三个模型能更准确地描述用户在 Web 上的浏览特征,在预测准确率和存储复杂度方面都显著地优于传统的时间序列预测模型。

**关键词** 时间序列预测, Bayes 网络, 用户浏览预测

## Time-Series Forecast Based on Bayes Network

QIN Peng XING Yong-Kang YANG Hua-Li

(College of Computer Science and Technology, Chongqing University Chongqing 400044)

**Abstract** Time Series Forecast problem is widely used in weather, astronomy, electric power, medicine, biology, economy, finance, computer etc. Now we use Bayes Network in Time Series Forecast, put forward and established a set of based on Bayes time-series forecasting model -- static Bayes network forecasting model, dynamic Bayes network forecasting model and the classification static Bayes network prediction model. Experiments have shown that these three models can more accurately describe users browsing features on Web. They are significantly better than traditional time-series forecasting model in the forecast storage complexity and the accuracy rate.

**Keywords** Time series forecast, Bayes networks, Users browse forecast

## 1 引言

时间序列是指同一种现象在不同时间上的相继观察值排列而成的一组数字序列。时间序列预测方法的基本思想是:在预测一个现象的未来变化时,用该现象的过去行为来预测未来。即通过时间序列的历史数据揭示现象随时间变化的规律,将这种规律延伸到未来,从而对该现象的未来做出预测。随着计算机技术的发展,特别是自从著名统计学家 Box 和 Jenkins 于 1968 年发展了一整套随机时间序列的模型识别、参数估计和诊断检验的经典建模<sup>[1]</sup>方法以来,时间序列预测的发展非常迅速,在气象、天文、电力、医学、生物、经济、金融和计算机等各个领域已有广泛的应用,并日益显示出强大的生命力和重要性,已成为一门独立的重要数学分支。迄今已有不少专家、学者从事这方面的研究,并建立了一套比较完整的理论和应用体系<sup>[2,3]</sup>。

时间序列预测模型可以分为以下两种:线性模型和非线性模型。在线性模型中以自回归求和移动平均模型(ARIMA)、指数平滑方法模型等为代表;而非线性模型则以临界 AR 模型、广义自回归条件异方差模型(GARCH)<sup>[4]</sup>、神经网络模型等为代表。这些预测模型的预测准确度都很好,但是都存在着对预测数据单一性、选择性、存储复杂度高等问题。本文在传统时间序列预测方法和 Bayes 网络的基础上提出了三种基于 Bayes 网络的时间序列预测模型:静态 Bayes 网预测模型,动态 Bayes 网预测模型,分类静态 Bayes 网预测模型。最后再将这三种模型用于对 Web 服务器的日志文件进行预测。发现三个模型的预测准确度都很好,而且在很大程

度上降低了存储复杂度并且对实验数据的适应性比较好。

本文第 2 节中介绍了 Bayes 网的条件概率表的学习方法;第 3 节中提出了静态 Bayes 网预测模型;第 4 节中提出了动态 Bayes 网预测模型;第 5 节中提出了最后一种模型分类静态 Bayes 网预测模型;第 6 节中将 Web 服务器的日志文件用于这三中模型并做了比较;最后进行了总结。

## 2 Bayes 网简介

Bayes 网是一种基于概率论和图论的不确定知识表示模型,由 pearl<sup>[5]</sup>等人首先提出。Bayes 网常用来表示变量间连接概率的图形模式,它提供了一种自然的表示因果信息的方法,用来发现数据间的潜在关系。在这个网络中,用节点表示变量,有向边表示变量间的依赖关系。它用概率测度的权重来描述数据的相关性,用图形方法描述数据间的相互关系。对信度网的研究已成为当今人工智能领域研究的热点之一。

**定义 1** 给定一个(离散型)随机变量的集合  $X = \{X_1, X_2, \dots, X_N\}$ , 建立在该集合上的联合概率分布  $p(x_1, x_2, \dots, x_n)$  可以表示为一个 Bayes 网  $B = \{B_s, B_p\}$ , 其中:

$B_s$  为 Bayes 网的结构。它是一个有向无环图。图中的每一个结点唯一地对应一个随机变量,结点的状态对应于随机变量的值。图中的有向边表示变量(结点)之间的条件(因果)依赖关系。

$B_p$  为 Bayes 网的条件概率表集合。每个结点  $X_i$  都有一个条件概率表,用来表示  $X_i$  对于其父结点  $parent(x_i)$  的条件概率  $p(x_i | parent(x_i))$ 。

Bayes 网的独特结构隐含着一种很强的条件独立性。当

<sup>\*</sup>)本研究得到国家自然科学基金青年基金资助(60403009)。邢永康 博士后,副教授,硕士生导师,主要研究领域为人工智能。秦 鹏 硕士生,研究方向为时间序列预测,数据挖掘。杨华丽 硕士生,研究方向为图像处理。

一个接点的父亲结点给定后,除去该结点的后代结点,它与其他的结点之间条件独立。利用这种条件独立性可以将联合概率分布简化为:  $p(x_1, x_2, \dots, x_n) = \prod_{i=0}^n p(x_i | \text{parent}(x_i))$ 。从 Bayes 网的定义可以知道,它综合应用了概率理论和图论理论;同时联合概率分布保证了整个系统的一致性,图形结构又提供了计算的直观性。正因为这一点 Bayes 网不但能用于预测问题,而且还可以将数据本身和影响数据变化的各种因数结合起来,充分发挥它在不确定知识表示及推理中的作用。

### 3 静态 Bayes 网预测模型

静态 Bayes 网预测模型对时间序列做如下的假设。

**假设 1(条件独立性假设)** 假设在时间序列中各个序列值之间是相互独立的。即离散随机变量  $X$  的值域是由自然数所构成的一个集合  $\{1, 2, 3, \dots, n\}$ , 一个时间序列值就是随机变量  $X$  所构成的一个随机序列,且序列值之间是相互条件独立的。

**假设 2(预测相关性假设)** 一般情况下对时刻  $t$  的序列值进行预测时,离时刻  $t$  越近的序列值对预测结果的指导性越大,相反离时刻  $t$  越远的序列值对预测结果的指导性越小,所以我们假设整个预测过程中只有  $n$  个序列值对预测结果有影响,而这  $n$  个序列值以前的所有序列值对预测结果都没有影响。

根据以上假设,对静态 Bayes 网预测模型的定义如下:

**定义 2** 静态 Bayes 网预测模型可以表示为一个五元组,即  $M = \langle X, B_s, B_p, \text{SSearch 算法}, N \rangle$ , 其中:

- $X$  为离散随机变量,值域为  $\{x_1, x_2, \dots, x_m\}$ 。
- $B_s$  为 Bayes 网的结构。它是一个有向无环图。网络中的每一个结点唯一地对应一个随机变量,结点的状态对应于随机变量的值。该网络的结构是一种类倒三角结构。
- $B_p$  为 Bayes 网中结点  $X_0$  的条件概率表。其中  $P_{(1,2,\dots,n,0)_i} = P(X_0 = x_0^i | X_1 = x_1^i, X_2 = x_2^i, \dots, X_n = x_n^i)$ , 它所表示的是在  $X_0$  的所有父亲结点序列的取值为  $\{x_1^i, x_2^i, \dots, x_n^i\}$  时  $X_0 = x_0^i$  的条件概率。
- $\text{SSearch}$  算法为搜索算法。其思想为:将当前序列值的后  $N$  项与条件概率表中的序列值相比较,如果序列值相等且  $P$  最大时,则返回该条记录中的  $X_0$  和  $P$ ; 其中  $X_0$  是预测值,  $P$  是  $X_0$  出现的概率。
- $N$  为 Bayes 网中上层结点的数目。

Bayes 网的结构(BS)如下:

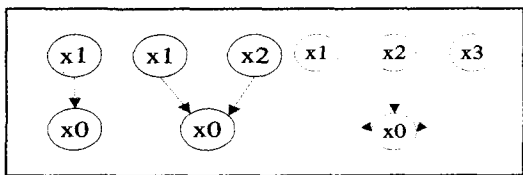


图 1

Bayes 网的条件概率表(Bp)为:

表 1 Bayes 网的条件概率表

$X_1$	$X_2$	...	$X_n$	P	$X_0$
$x_1^1$	$x_2^1$	...	$x_n^1$	$P_{(1,2,\dots,n,0)_1}$	$x_0^1$
$x_1^2$	$x_2^2$	...	$x_n^2$	$P_{(1,2,\dots,n,0)_2}$	$x_0^2$
.....					
$x_1^m$	$x_2^m$	...	$x_n^m$	$P_{(1,2,\dots,n,0)_m}$	$x_0^m$

### SSearch 算法

输入:序列值  $\{x_1^i, x_2^i, \dots, x_n^i\}$

输出:  $X_0$  的状态值 result,  $X_0$  发生的概率值 resultp

```

For(int i=0; i<Bp.length; i++) {
If(Bp[i]的前 n 项值与序列值
{x1, x2, ..., xn}相等) {
If(resultp <= Bp[i].P) {
result = Bp[i].X0
resultp = Bp[i].P
}
}
}
Return result, resultp
结束
    
```

根据条件独立性假设和容斥原则<sup>[6]</sup>可以得到:

$$P(X_0 = x_0^i | X_1 = x_1^i, X_2 = x_2^i, \dots, X_n = x_n^i) = \sum_{j=0}^n \eta_j | \text{其中} | \eta_j = P(X_0 = x_0^i | X_i = x_i^i)$$

用序列值  $\{x_1^i, x_2^i, \dots, x_n^i\}$  表示到离  $t$  时刻最近  $n$  个序列值,将该序列值输入 SSearch 算法中,该算法将从模型的条件概率表中找出前  $n$  个与以上的序列值相等的且  $X_0$  发生的条件概率最大的  $X_0$  取值以及相应的概率值。当然  $X_0$  的取值也就是我们所要的预测结果。

### 4 动态 Bayes 网预测模型

在静态 Bayes 网预测模型中,我们只将时间序列中近期的有限的几个序列值用于预测。也就是说该模型认为只有离当前时刻最近的少数的序列值对预测结果有影响,而其它的先前的序列值对预测结果没有影响。这样的预测虽然已经能够得到很高的预测精度,但是在某些情况下预测也应该将所有的序列值都纳入预测模型,才能使预测到达预期的效果,所以也就有了以下的动态 Bayes 网预测模型。

这里同样要利用静态 Bayes 网预测模型中的假设一(条件独立性假设)。不同的序列值之间会有一定的差异,序列的长度也不一样。这里我们将相同长度的序列归为一类,然后对每类序列都建立和静态 Bayes 网预测模型中一样的 Bayes 网络。在预测的时候根据状态序列的长度来选择相应的 Bayes 网络进行预测。

动态 Bayes 网预测模型的定义如下:

**定义 3** 动态 Bayes 网预测模型可以表示为一个四元组,即  $M = \langle X, B_s, B_p, \text{DSearch 算法} \rangle$ , 其中:

- $X, B_s, B_p$  的定义见定义 2。
- $\text{DSearch}$  算法为搜索算法。其思想为:先将条件概率表集合根据其序列的长度建立一个平衡二叉树。然后根据输入的序列值的长度,在平衡二叉树中查找出属于该条记录的条件概率表。将当前序列值与条件概率表中的序列值相比较,如果序列值相等且  $P$  最大时则返回该条记录中的  $X_0$  和  $P$ ; 其中  $X_0$  是预测值,  $P$  是  $X_0$  出现的概率。

条件概率表集合  $(B_p)$  为:  $\{B_{p_1}, B_{p_2}, \dots, B_{p_n}\}$ , 其中  $B_{p_i}$  表示的是上层有  $i$  个结点的 Bayes 网络中结点  $X_0$  的条件概率表。  $B_{p_i}$  的内容如表 1 所示。其中

$$P_{(1,2,\dots,j,0)_i} = \sum_{k=1}^j \alpha_k P(X_0 = x_0^i | X_i = x_i^i)$$

$$\alpha_k = 1 + \sum_{k=1}^{j-1} (-1)^k \sum_{k=1}^j \frac{1}{k} P(X_0 = x_0^i | X_i = x_i^i)$$

其实这里的条件就是对  $t$  时刻的所有先前状态条件概率的加权平均。

### DSearch 算法

输入:序列值  $\{x_1^i, x_2^i, \dots, x_j^i\}$

输出:  $X_0$  的状态值 result,  $X_0$  发生的概率值 resultp

首先将条件概率表集合  $\{Bp_1, Bp_2, \dots, Bp_n\}$  构造成一棵平衡二叉树 BPT。然后根据输入序列的长度  $j$  从 BPT 中得到上层有  $j$  个结点的 Bayes 网以及  $X_0$  的条件概率表  $Bp_j$ 。

```
For(int i=0; i<Bp_j.length; i++){
    If(Bp_j[i]的值与序列值{x1, x2, ..., xj}相等){
        If(resultp<=Bp_j[i].P){
            result=Bp_j[i].X0
            resultp=Bp_j[i].P
        }
    }
}
Return result, resultp
结束
```

用序列值  $\{x_1, x_2, \dots, x_j\}$  表示到  $t$  时刻之前最近  $n$  个序列值, 将该序列值输入 DSearch 算法中, 该算法将从模型的条件概率表集合中找出符合该序列的条件概率表, 然后从该条件概率表中找出序列值与以上的序列值相等的且  $X_0$  发生的条件概率最大的  $X_0$  的取值以及相应的概率值。这样  $X_0$  的取值也就是我们所要的预测结果。

## 5 分类静态 Bayes 网预测模型

在以上的两种预测模型中, 我们分别是把预测序列的后  $n$  个状态和预测序列的所有状态用于预测。在第一个模型中只把预测序列的后  $n$  个状态用于预测, 而没有考虑预测序列的前端序列值。这样虽然降低了条件概率表的存储空间和搜索的时间, 但是它没有充分地利用整个预测序列从而影响预测精度。在第二个模型中将预测序列中的每个值都作为一种状态然后将它们用于预测。这样虽然充分地利用了整个预测序列也在一定程度上提高了预测精度, 但是它又增加了条件概率表的存储空间以及搜索的时间复杂度, 增加了搜索的时间。所以综合以上两种模型, 我们又提出了第 3 种预测模型“分类静态 Bayes 网预测模型”。该模型的大体思想就是: 先对时间序列值进行分类, 然后给每一类序列值都建立类似于“静态 Bayes 网预测模型”中的静态 Bayes 网, 再利用该网络进行预测。

这里我们利用 CURE 聚类算法对序列进行聚类得到相类的预测序列分类, 并做如下的假设。

**假设 4(时间序列值分类假设)** 假设根据时间序列值的特征, 可以将所有的时间序列值分为  $K$  类, 同一类别的时间序列值有着相同的特征。如果用  $C = \{c_1, c_2, \dots, c_k\}$  表示时间序列值的类别, 则任意一个时间序列值属于类别  $c_k$  的概率为  $P(C=c_k)$ , 且有  $\sum_{k=1}^K P(C=c_k) = 1$ 。

根据以上假设, 对分类静态 Bayes 网预测模型的定义如下:

**定义 4** 分类静态 Bayes 网预测模型可以表示为一个五元组, 即  $M = \langle X, Bs, Bp, TSSearch \text{ 算法}, N \rangle$ , 其中:

- $X, Bs, N$  的定义见定义 1。

- $Bp$  为 Bayes 网中结点  $X_0$  的条件概率表集合。集合中每一个元素就是一类用户 Bayes 网中  $X_0$  的条件概率表, 其内容如表 1 所示。

- TSSearch 算法为搜索算法。其思想为: 先根据输入的序列值来判定该序列值所属于的类型, 再找出属于该类型的条件概率表。将当前序列值的后  $N$  项与该条件概率表中的序列值相比较, 如果序列值相等且  $P$  最大时则返回该条记录中的  $X_0$  和  $P$ ; 其中  $X_0$  是预测值,  $P$  是  $X_0$  出现的概率。

### TSSearch 算法

输入: 序列值  $\{x_1, x_2, \dots, x_j\}$

输出:  $X_0$  的状态值 result,  $X_0$  发生的概率值 resultp

首先我们通过序列值  $\{x_1, x_2, \dots, x_j\}$  来判断该序列值所属的类型  $c_k$  以及  $P(C=c_k)$ , 然后从条件概率表集合  $Bp$  中选出与类型  $c_k$  相对应的条件概率表  $Bp_k$ ; 最后利用算法 SSearch 算法(SSearch 算法中的条件概率表  $Bp = Bp_k$ ) 得到预测结果。

Return result, resultp

结束

在该模型中对时间序列值分类是至关重要的。好的、准确的分类能使模型的预测结果达到最佳, 相反分类不准确或是对时间序列值的判断不正确则将使预测结果大大的偏离实际情况。所以选用一个良好的分类和一个准确的时间序列值类型判定对提高模型预测性能很有帮助。我们采用以下的方法来判断时间序列值类型:

根据贝叶斯公式, 时间序列值属于类别  $c_k$  的概率为:

$$P(C=c_k | x_1, x_2, \dots, x_j) = \frac{P(x_1, x_2, \dots, x_j | C=c_k)P(C=c_k)}{P(x_1, x_2, \dots, x_j)}$$

上式中, 分母表示序列  $\{x_1, x_2, \dots, x_j\}$  的边际概率, 对于不同的分类, 该值都不会改变, 所以不用计算它, 从而建立如下的判定规则:

$$\text{If } P(x_1, x_2, \dots, x_j | C=c_k)P(C=c_k) = \max_{i=1,2,\dots,k} (P(x_1, x_2, \dots, x_j | C=c_i)P(C=c_i))$$

Then 时间序列值类别为  $c_k$

该判定规则就是基于最小错误率的贝叶斯判定规则, 可以证明按这种规则对时间序列值分类, 其分类结果的错误率最小<sup>[7]</sup>。

## 6 实验分析

为了验证以上三种模型的预测能力, 并进行比较, 我们采用了 EPA WWW 服务器的日志文件作为实验数据, 它记录了 EPA WWW 服务器上连续 24 小时内对服务器的 47748 次 HTTP 请求, 共涉及到 3086 个页面。我们将来自同一个 IP 的请求看作是同一个用户的请求, 并根据每次请求的时间确定其先后顺序, 这样就将这些数据转化为 2266 个用户浏览序列, 每个浏览序列的平均长度为  $47748/2266 \approx 21$ 。我们随机地从中抽去 1500 个作为学习数据来建立模型的条件概率表, 其余的 700 多作为测试数据分别用于这三个模型做了三个实验。

实验一利用静态 Bayes 网预测模型, 建立一个上层有 3 个结点的 Bayes 网络。实验二利用动态 Bayes 网预测模型, 建立了一个上层结点从 1 到 21 的 Bayes 网络集合。实验三则是利用分类静态 Bayes 网预测模型, 先将用户分为了 20 类, 然后为每类用户建立一个上层有 3 个结点的 Bayes 网络。我们用以上三个实验的结果来比较他们的预测准确率, 其结果如图 1 所示。从图表上我们可以看出当用户的浏览序列大于 6 的时候, 三个模型的预测准确度都得到了明显的提高。当浏览序列大于 8 的时候分类静态 Bayes 网预测模型和动态 Bayes 网预测模型的预测准确度大大地超过了静态 Bayes 网预测模型。产生这种特征的原因是: 用户浏览序列增长时, 分类静态 Bayes 网预测模型对用户的类别的判断将更加的准确, 反之当浏览序列还不够长时对用户类别判定的出错率将

(下转第 193 页)

$m_{i_0j_0} \neq \emptyset$ , 故有  $d(x_{i_0}, D); d(x_{j_0}, D)$  且至少存在一属性  $a \in C - B$ , 使得  $f(x_{i_0}, a) \neq f(x_{j_0}, a)$ , 故有  $[x_{i_0}]_C \neq [x_{j_0}]_C$ . 另一方面有  $[x_{i_0}]_B = [x_{j_0}]_B$  (这是因为  $m_{i_0j_0} \cap B = \emptyset$ ), 即  $[x_{i_0}]_C \cup [x_{j_0}]_C \subseteq Y_0 \in U/B$ . 又由于  $d(x_{i_0}, D) \neq d(x_{j_0}, D)$ , 故一定存在  $D_p \cap [x_{i_0}]_C \neq \emptyset, D_q \cap [x_{j_0}]_C \neq \emptyset, D_p D_q \in U/D, p \neq q$ , 故有  $|D_p \cap [x_{i_0}]_C| |D_q \cap [x_{j_0}]_C| \neq 0$ , 由定理 2 知有  $I(D|B) > I(D|C)$ , 这与条件矛盾, 故假设不成立, 从而命题成立.

**定理 7** 在决策表  $S = (U, C, D, V, f, d)$  中  $B \subseteq C, \forall b \in B$ , 若  $I(D|(B - \{b\})) > I(D|C)$ , 则一定存在  $\emptyset \neq m_{i_0j_0} \in M$ , 使得  $(B - \{b\}) \cap m_{i_0j_0} = \emptyset$ .

证明: 假设结论不成立, 即对  $\emptyset \neq m_{ij} \in M$ , 使得  $(B - \{b\}) \cap m_{ij} \neq \emptyset$ . 由定理 3 的证明过程可知有  $I(D|(B - \{b\})) = I(D|C)$ , 这与条件矛盾, 故假设不成立, 从而命题成立.

**定理 8** 在决策表  $S = (U, C, D, V, f, d)$  中,  $\forall B \in \text{Ner-}w\text{RED}_C(D)$ , 则有  $B \in \text{HRED}_C(D)$ .

证明: 由定理 6 和定理 7 即得.

**定理 9** 在决策表  $S = (U, C, D, V, f, d)$  中, 基于新的信息熵的属性约简与基于 HU 差别矩阵的属性约简是等价的.

证明: 由定理 5 和定理 8 即得.

**结论** 王国胤教授定义了一种新的决策表信息熵计算方法, 在此方法的基础上, 给出了粗糙集理论代数观的一种新的信息观解释, 并证明了这种新的信息观与代数观是等价的. 在本文中, 我们证明了梁吉业教授和罗平教授提出的一种基于新的信息熵的属性约简与基于 HU 差别矩阵的属性约简是等价的. 从而找到了基于 HU 差别矩阵的属性约简的信息观解释. 这一研究思路揭示了属性

约简中代数方法与信息熵方法的联系, 在结合代数方法与信息熵方法优点的基础上, 设计高效属性约简算法是我们进一步的研究方向.

### 参考文献

- 1 Pawlak Z. Rough Sets[J]. International Journal of Computer and information Science, 1982, 11(5): 341~356
- 2 Hu Xiao Hua, Cercone N. Learning in relational databases: a rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-337
- 3 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759
- 4 Wang GuoYin. Rough reduction in algebra view and information view[J]. International Journal of Intelligent System, 2003(18): 679~688
- 5 王国胤. 决策表核的计算方法[J]. 计算机学报, 2003, 26(5): 611~615
- 6 叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报, 2002, 30(7): 1086~1088
- 7 杨明, 孙志挥. 改进的差别矩阵及其求核方法[J]. 复旦学报(自然科学版), 2004, 43(5): 865~868
- 8 胡健, 徐章艳, 杨炳儒. 关于“一种信息系统求核的新方法”的笔记[J]. 计算机工程与应用, 200, 41(27): 51~52
- 9 龚勋, 王国胤. 对代数观 Rough 集理论的信息观解释[J]. 计算机科学, 2006, 33(4): 151~154
- 10 Liang Jiye, Chin K S, Dang Chuangyin, Yam R C M. A new method for measuring uncertainty and fuzziness in rough set theory[J]. International Journal of General Systems, 2002, 31(4): 331~4342
- 11 Luo Ping, He Qing, Shi Zhongzhi. Theoretical study on a new information entropy and its use in attribute reduction [C]. ICCI, 2005. 73~79

(上接第 185 页)

增加也就使预测的准确度降低了. 随着用户浏览序列的增加, 分类和用户类别判定所出现的错误率就会越来越小, 预测的准确率将会越来越高, 这也就是当浏览序列大于 9 之后分类静态 Bayes 网预测模型的预测准确率一直高于其他模型的原因所在. 同样随着用户浏览序列的增加, 用户的浏览状态就越能体现用户浏览的特征, 所以当用户浏览序列增加时三种模型的预测准确度都在上升.

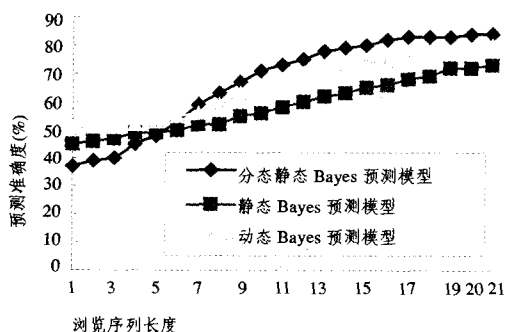


图 1 三种预测模型预测准确度的比较

**总结** 本文针对 Web 上用户的浏览序列这一时间序列问题所提出的三种预测模型. 实验表明以上三种模型与传统的预测模型相比有以下的优点:

① 三种模型都与 Bayes 网络相结合. 充分利用了传统时间序列预测模型和 Bayes 网络的优点. 时间序列预测问题的不确定性本质, 表明该问题本身就是一个典型的不确定性处

理问题. 因此利用 Bayes 网对其进行处理和研究表明能够充分地利用 Bayes 网络的理论提高预测的准确度.

② 实际应用中合理的选用这三种模型, 可以达到降低模型的存储空间, 充分地利用实验数据的优势, 从而提高预测精度和时间复杂度.

当然, 本文所提出的预测模型还有很多未完善的地方需要进一步的研究. 比如在对条件概率的计算和降低条件概率表的存储度以及模型中所提到的三个搜索算法等都需要做进一步的改进.

### 参考文献

- 1 (US) Box G E P, (UK) Jenkins G M. Times Series Analysis; Forecasting and Control[M]. 2<sup>nd</sup> edn. San Francisco: Holden-Day, 1976
- 2 West M, Harrison P J. Bayesian Forecasting and Dynamic Models[M]. New York: Springer-Verlag, 1997
- 3 张孝令, 刘福升, 张承进, 等. 贝叶斯动态模型及其预测[M]. 济南: 山东科学技术出版社, 1992
- 4 Shephard N. Statistical aspects of ARCH and stochastic volatility. In Time Series Models (eds D. R. Cox et al.), London. Chapman and Hall, 1996. 1~87
- 5 Fusion P J. Propagation, and Structuring in Belief Networks. Artificial Intelligence, 1985, 29: 241~288
- 6 Dohmen, K. Improved Bonferroni Inequalities with Applications: Inequalities and Identities of Inclusion-Exclusion Type. Berlin, Germany: Springer-Verlag, 2003
- 7 边肇祺, 张学工, 等. 模式识别(第二版). 北京: 清华大学出版社, 2000