# 基于时空相关性的数字电影放映场次预测①

付 园，高 强

(北京航空航天大学 电子信息工程学院，北京 100191)

**摘 要**：为提高数字电影流动放映中各院线放映场次的预测精度，基于院线放映场次的时空相关性，提出一种最邻近法和时空序列相结合的预测方法：通过最邻近法选取与待预测院线放映最相关的院线作为预测辅助院线；通过构造神经网络的时空序列将这些相关院线放映场次的时、空特征结合起来，产生更精确的预测模型。实验对比分析了此方法与传统时间序列预测方法的预测结果，证明了该方法具有更高的预测精度。

**关键词**：时空相关性；场次预测；最邻近法；时空序列

## Digital Film Screening Forecast Based on Spatio-Temporal Correlation

FU Yuan, GAO Qiang

(Dept of Electronics and Information Engineering, Beihang University, Beijing 100191, China)

**Abstract**: In order to improve the forecast accuracy of film screenings in each cinema in the mobile digital film projection, this paper proposes a new method combining nearest neighbor method and spatio-temporal sequence together based on the spatio-temporal correlation of cinema screenings. Firstly nearest neighbor method was used to find cinemas which correlated the predicted cinema closely. Secondly, spatio-temporal sequences used for a neural network were constructed to combine the spatio-temporal characteristics together, getting more accurate forecast model. Experiments compared the forecast result of this method with the tradition one, demonstrating its higher accuracy.

**Key words**: spatio-temporal correlation; screening forecast; nearest neighbor method; spatio-temporal sequence

## 1 引言

在我国面向农村地区的数字电影流动放映中，放映活动深入群众，广受好评；但由于流动放映是露天免费放映、缺乏城市影院的各项配套设施及充足的人力物力，因此为了保证放映活动的顺利开展、提高流动放映的服务质量，必须提前进行规划和筹备。如果能够准确预测不同院线不同时期的放映场次，那么规划内容将有章有循、事半功倍；同时准确的预测也对流动放映相关产业（例如流动放映广告产业）经营者的经营决策具有积极指导作用。

流动放映是露天放映，放映活动严重受当地气候、天气、日照时长等因素的制约，因此院线的影片放映场次与这些因素密切相关；另外，国家对流动放映的投资力度以及当地的经济文化环境也会影响院线的放映场次。由于这些影响因素或者具有典型的时间特征、或者具有空间特征、或者兼而有之，因而院线的影片放映场次具有明显的时空相关性。从时间上看，随着电影产业的发展、国家投资力度的增加，影片放映场次呈逐年递增趋势，而每年不同时期的放映场次又随气温、降雨量等因素变化呈现不同趋势；从空间上看，地理位置靠近、自然环境、经济文化环境相似的院线，影片放映场次具有相关性。

流动放映免费不售票的特点使其只能以放映场次作为放映指标，因而在流动放映回传系统的数据库中积累了大量各个院线不同时期放映场次的历史记录。通过分析这些历史记录的统计特征，可以对各个院线未来的放映场次进行预测。传统预测方法中，专家经验预测法、算术平均法、线性回归法、指数平滑法等

① 收稿时间:2011-07-07;收到修改稿时间:2011-08-31



© 中国科学院软件研究所 http://www.c-s-a.org.cn

可以预测未来时间内待测变量变化的大致趋势[1]，但当预测对象具有较强非线性时，建模和求解的难度较大，预测精度不高，对于院线放映场次存在的时空相关性，也不能很好利用；应用人工神经网络进行时间序列预测，具有良好的非线性、高容错性等特点，在很多实际应用领域中都取得了成功[2]，但其也只考虑了预测序列本身在时间上体现的趋势与连贯性，忽略了数据在时空整体上体现的规律性，不适合预测不同院线不同时期的放映场次。

本文基于数字电影放映场次的时空相关性，提出了一种最邻近法与时空序列相结合的方法，对不同院线不同时期的放映场次进行预测。该方法通过最邻近法选择与待预测院线放映最为相关的若干院线作为预测辅助院线；然后将这些相关院线不同时期的放映场次构造时空序列，借助 BP 神经网络计算出预测场次。该方法利用了待预测院线相关院线的放映数据，并将这些数据的时间特征和空间特征结合起来，基于数据的时空相关性提高预测精度，弥补了传统预测方法的不足。

本文第 2 节详细阐述了针对数字电影流动放映特点设计的不同院线不同时期放映场次的预测方法。第 3 节给出了中国部分院线不同时期放映场次的预测实例，通过与实际数据的比较，验证该方法的可行性。第 4 节将该方法的预测结果与传统方法进行比较，给出性能分析。第 5 节对全文进行总结。

## 2 数字电影放映场次预测方法

基于时空相关性的数字电影放映场次预测包括两个过程：首先通过最邻近法选择预测辅助院线；然后根据选择结果构造时空序列、利用神经网络预测放映场次。

### 2.1 最邻近法选择预测辅助院线

最邻近法就是通过比较要判断的案例与每一个案例 $\mathbf{x}$ 之间的距离，来找出与新案 $\mathbf{x}$ 例最相近的案例[3]。因此，最邻近法选择预测辅助院线首先需要确定各院线彼此在特征空间的距离。根据引言对院线影片放映场次的规律分析，本文规定院线 $a$ 、$b$ 的特征空间距离由它们放映场次的时间相似度（式(1)）[4]和院线间的空间距离（式(2)）两部分组成（式(3)）。

根据式(3)，可以定义院线 $i$ 的特征向量 $\mathbf{x}_i=\left(t_i^1,\cdots,t_i^m,\cdots,t_i^{M-1},s_i^x,s_i^y\right)$，其中 $t_i^m$ 即式(1)中的 $t_i^m$，$N_i^{y,m}$ 表示院线 $i$ 第 $y$ 年第 $m$ 个时间段内的放映场次数，$Y$ 为考查总年数；$s_i^x$ 、$s_i^y$ 的定义如式(3)中所示，$lon_i$ 、$lat_i$ 分别表示院线 $i$ 的经、纬度，$lon_0$ 、$lat_0$ 是参考经、纬度，通常选取院线集合中心点的经纬值。从而计算任意两院线在特征空间的距离转化为计算两院线特征向量的欧氏距离。

$$d^t(a,b)=\sqrt{\sum_{m=1}^{M-1}\left(t_a^m-t_b^m\right)^2},$$
$$\text{其中}t_i^m=\frac{1}{Y}\sum_{y=1}^{Y}\frac{N_i^{y,m+1}-N_i^{y,m}}{\max\limits_{1\le m\le M-1}\left(N_i^{y,m+1}-N_i^{y,m}\right)},\quad i\in\{a,b\} \tag{1}$$

$$d^s(a,b)=\sqrt{\left(x_a-x_b\right)^2+\left(y_a-y_b\right)^2}$$
$$\text{其中}\begin{cases}x_i=R\cdot\left(lon_i-lon_0\right)\cdot\cos\left(lat_0\cdot\pi/180\right)\\ y_i=R\cdot\left(lat_i-lat_0\right)\end{cases},i\in\{a,b\} \tag{2}$$

$$\begin{aligned}d(a,b)&=\sqrt{K_t\cdot d^t(a,b)^2+K_s\cdot d^s(a,b)^2}\\&=\sqrt{\sum_{m=1}^{M-1}\left(t_a^m-t_b^m\right)^2+\frac{K_s}{K_t}\cdot\left[\left(x_a-x_b\right)^2+\left(y_a-y_b\right)^2\right]}\cdot\sqrt{K_t}\\&=\sqrt{\sum_{m=1}^{M-1}\left(t_a^m-t_b^m\right)^2+\left(s_a^x-s_b^x\right)^2+\left(s_a^y-s_b^y\right)^2}\cdot\sqrt{K_t}\quad,\end{aligned} \tag{3}$$
$$\text{其中}s_i^x=R\cdot\sqrt{K_s/K_t}\left(lon_i-lon_0\right)\cos\left(lat_0\cdot\pi/180\right),$$
$$s_i^y=R\cdot\sqrt{K_s/K_t}\left(lat_i-lat_0\right),\quad i\in\{a,b\}$$

采用传统的最邻近法，需要计算待预测院线 $i'$ 与其它每个院线特征向量的欧氏距离，通过比较选择其中距离最短的 $h_{sel}$ 个院线作为预测辅助院线。虽然简单易行，但存在两个问题：第一，因为没有控制策略，判断非常盲目，必须将被判断的案例与其它每个案例都进行比较，把很多计算都用在和它不类似的案例上，导致计算量很大[5]；第二，如何选取才能既保证选择的预测辅助院线和待预测院线具有较高相关性，又保证选择的相关院线没有遗漏。

针对这两个问题，本文设计了一种改进算法：将院线集合采用聚类算法进行前期处理，使得聚到同一类的院线具有较高相关性，称为相关院线（或者同类院线），相关院线之间相互可作为对方预测辅助院线的候选院线；当需要为某个待预测院线选择预测辅助院线时，只需要在它的相关院线集合中采用最邻近法搜索最相近的院线，这样既避免了盲目搜索，又保证了选择的院线具有较高的相关性。

为了保证该方法选择的预测辅助院线在所有院线



© 中国科学院软件研究所 http://www.c-s-a.org.cn

中也是距离待预测院线最近的，本文按照如下方式进行聚类（将院线的特征向量看作空间中的案例）：

1) 从空间中任意选取案例作为球心，以 $D$ 为半径作球形，落在球内的所有案例标记为一类。

2) 任选球内另外任意一案例为球心，仍以 $D$ 为半径作球形，将落在球内尚未作标记的案例标记为这一类。

3) 重复第 2)步，直至再没有新的案例标记为这一类，于是转至第 4)步。

4) 选取空间中尚未标记的案例作为新的一类，重复第 1)至第 3)步，至到所有的案例都分配完毕。

其中 $D$ 是关于院线特征向量空间距离的参数，可以经过反复实验修正得到。

在对院线进行聚类之后，便可为每一个待预测院线选择预测辅助院线。具体操作过程如下：假设待预测院线的相关院线有 $h_{cor}$ 个，预测辅助院线最多不超过 $h_{max}$ 个；如果 $h_{cor} \le h_{max}$，则待预测院线的所有相关院线都被选作预测辅助院线，即 $h_{sel} = h_{cor}$；如果 $h_{cor} > h_{max}$，则计算待预测院线与每个相关院线特征向量的距离，选择其中距离最近的 $h_{max}$ 个院线作为预测辅助院线，即 $h_{sel} = h_{max}$。

### 2.2 构造时空序列预测放映场次

传统的基于 BP 神经网络的时间序列预测方法，对单一的时间序列用训练样本部分和检验样本部分的部分重叠方法进行预测。例如 $x_1, x_2, \cdots, x_n$ 为单一的时间序列，要预测 $x_{n+1}$ 的值，可以把 $x_k, x_{k+1}, \cdots, x_{k+s}$ 过去 $s+1$ 个历史数据作为神经网络的第 $k$ 个输入样本，相应的用 $x_{k+s+1}$ 作为第 $k$ 个样本对应的输出，其中 $k+s+1 \le n$，$k = 1, 2, \cdots, n-s-1$，从而对神经网络进行训练，以构造预测模型。

由于待预测院线和预测辅助院线各自的放映场次可以按照时间顺序构成时间序列，而这组时间序列相互具有相关性，因此可以按照如下方式构造时空序列。

假设院线 $i_1, \cdots, i_u, \cdots, i_h (1 \le u \le h)$ 为预测辅助院线，院线 $i'$ 为待预测院线。$N_{i_1}^{y,m}, \cdots, N_{i_u}^{y,m}, \cdots, N_{i_h}^{y,m}$ 和 $N_{i'}^{y,m}$ 分别为预测辅助院线和待预测院线第 $y$ 年、第 $m$ 个时间段内的放映场次，$1 \le y \le Y$，$1 \le m \le M$，$Y$、$M$ 分别是考查的总年数和每年的总时间段。

从时间上看，任何院线的放映场次都可组成时间序列 $N_i^{1,1}, N_i^{1,2}, \cdots, N_i^{1,M}, N_i^{2,1}, N_i^{2,2}, \cdots, N_i^{2,M}, N_i^{Y,1} \cdots, N_i^{Y,M}$；从空间上看，待预测院线和其预测辅助院线的放映场次可以看成向量 $\mathbf{N}^{y,m} = \left\{ N_{i_1}^{y,m}, \cdots, N_{i_u}^{y,m}, \cdots, N_{i_h}^{y,m}, N_{i'}^{y,m} \right\}$。因此综合考虑，将这些相关院线连续 $L$ 年同一时间段的影片放映场次和待预测院线第 $(y+L)$ 年同一时间段的影片放映场次组成一个相应的时间序列[6]，即将 $\mathbf{N}^{y,m}, \mathbf{N}^{y+1,m}, \cdots, \mathbf{N}^{y+L-1,m}, N_{i'}^{y+L,m}$ 作为一个时间序列，其中 $\mathbf{N}^{y,m}, \mathbf{N}^{y+1,m}, \cdots, \mathbf{N}^{y+L-1,m}$ 作为 BP 神经网络的输入因素，相应的 $N_{i'}^{y+L,m}$ 作为 BP 神经网络的输出因素。

由此可以看出，时空序列的构造方式取决于预测辅助院线选择结果。如果待预测院线没有预测辅助院线，则该时空序列退化成传统时间序列。

为了使神经网络更加容易训练和学习，需要对样本进行归一化处理：

$$\hat{N}_i^{y,m} = \left( N_i^{y,m} - \min_{i \in \mathbf{S}} (N_i^{y,m}) \right) \Big/ \left( \max_{i \in \mathbf{S}} (N_i^{y,m}) - \min_{i \in \mathbf{S}} (N_i^{y,m}) \right) \quad (4)$$

其中 $\mathbf{S} = \{ i_1, \cdots, i_u, \cdots i_h, i' \}$，$1 \le u \le h$。使用归一化的数据训练神经网络，可以避免出现学习结果不收敛的情况。

一个 BP 神经网络由一个输入层，多个隐含层和多个输出层组成。研究表明[7]，一个三层人工神经网络模型就可解决一般非线性函数的拟合、逼近问题，故本文采用三层 BP 神经网络结构。根据预测辅助院线的选择结果，输入节点共有个，隐含层节点数按经验值设为输入节点个数的 1/2,即。Kurt Hornik 等人早已证明[8]：若输入层和输出层采用线性转换函数，隐含层采用 Sigmoid 转换函数，则含一个隐含层的人工神经网络能够以任意精度逼近任何有理函数。因此隐含层使用 Sigmoid 函数：。该神经网络结构如图 1 所示。

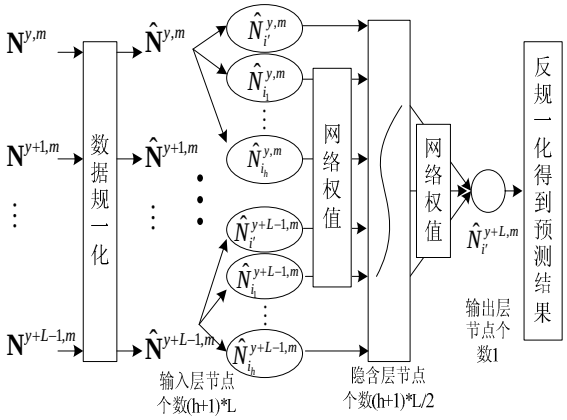


图 1　神经网络结构图



© 中国科学院软件研究所 http://www.c-s-a.org.cn

网络学习过程由信号的正向传播与误差的逆向传播两个过程组成，并通过迭代地处理一组训练样本、将每个样本的网络预测与实际值进行比较，修改网络权值使得网络预测和实际值之间的均方误差逐渐减少到可接受的程度或达到设定的学习次数为止[9]。利用训练好的神经网络模型，按照训练样本的方式输入预测相关数据，计算预测结果。

## 3 实例分析

本节通过预测实例进一步说明最邻近法和时空序列相结合的数字电影放映场次预测方法操作流程并验证其可行性与预测效果。

以中国东部的丽水时代院线、南部的莆田新星院线、中部的武汉春晓院线为预测对象，利用其 2005 至 2010 年的放映数据预测 2011 年 1 至 5 月各月的放映场次，并与实际值进行比较。表 1 列出了丽水时代院线部分历史放映数据。

表 1 丽水时代院线部分历史放映数据

| 月 / 年 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2010 | 63 | 34 | 51 | 193 | 253 | 324 | 385 | 447 | 459 | 474 | 418 | 291 |
| 2009 | 58 | 26 | 42 | 159 | 214 | 249 | 318 | 378 | 391 | 389 | 347 | 241 |
| 2008 | 46 | 21 | 34 | 127 | 170 | 213 | 244 | 286 | 301 | 304 | 275 | 192 |
| 2007 | 36 | 19 | 29 | 95 | 124 | 158 | 179 | 223 | 257 | 236 | 204 | 142 |
| 2006 | 23 | 11 | 18 | 64 | 86 | 105 | 127 | 145 | 161 | 154 | 132 | 79 |
| 2005 | 15 | 7 | 11 | 41 | 55 | 69 | 82 | 97 | 112 | 103 | 89 | 62 |

1) 最邻近法选取预测辅助院线

a) 计算院线的特征向量

设置 $R\cdot\sqrt{K_s/K_t}=0.9$，$M=12$，$lon_0=117.625$，$lat_0=26.825$。根据表 1 计算得到 $t^1_{\text{丽水}}\sim t^{11}_{\text{丽水}}$ 依次等于-0.225、0.141、1.0、0.457、0.443、0.430、0.493、0.287、-0.108、-0.416、-0.947；根据 $lon_{\text{丽水}}=119.92$、$lat_{\text{丽水}}=28.45$ 计算得到 $s^x_{\text{丽水}}$=1.843，$s^y_{\text{丽水}}$=1.463，因此其特征向量：

$$\mathbf{x}_{\text{丽水}}=(-0.255,0.141,1.0,0.457,0.443,0.430,0.493,\\ 0.287,-0.108,-0.416,-0.947,1.843,1.463)$$

选取中国东、南、中部部分地区的 49 家院线作为考查的院线集合，按照同样的方法，分别计算每个院线的特征向量，部分计算结果如表 2 所示。

表 2 部分院线的特征向量

| 院线 | 特征向量 |
|---|---|
| 金华新农村院线 | $\mathbf{x}_{\text{金华}}=(-0.198,0.016,0.574,1.0,0.689,0.402,0.331,0.336,-0.089,-0.387,-0.896,1.618,2.065)$ |
| 衢州新农村院线 | $\mathbf{x}_{\text{衢州}}=(-0.217,0.091,1.0,0.238,0.307,0.314,0.472,0.263,-0.101,-0.414,-0.926,1.008,1.930)$ |
| 温州新农村院线 | $\mathbf{x}_{\text{温州}}=(-0.151,0.109,0.592,1.0,0.228,0.421,0.439,0.286,-0.124,-0.417,-0.907,2.429,1.066)$ |
| 台州新农村院线 | $\mathbf{x}_{\text{台州}}=(-0.249,0.144,1.0,0.975,0.402,0.419,0.445,0.293,-0.116,-0.421,-0.914,3.047,1.651)$ |
| 绍兴新世纪院线 | $\mathbf{x}_{\text{绍兴}}=(-0.252,0.138,0.922,1.0,0.482,0.399,0.463,0.326,-0.107,-0.413,-0.902,2.373,2.866)$ |

b) 根据院线的特征向量对院线进行聚类

设置 $D=1.25$,对考查的院线集合进行聚类，得到部分聚类结果如表 3 所示。

表 3 院线聚类结果

| 预测院线 | 同类院线 |
|---|---|
| 丽水时代院线 | 金华新农村院线、衢州新农村院线、温州新农村院线、台州新农村院线、绍兴新世纪院线 |
| 莆田新星院线 | 宁德新农村院线、福州华益院线、泉州闽南院线、厦门新农村院线 |
| 武汉春晓院线 | 孝感澴川院线、黄冈新农村院线、咸宁新农村院线 |

c) 根据聚类结果选择预测辅助院线

设置 $h_{max}=4$，由表 3 可以看到，莆田新星院线和武汉春晓院线的同类院线个数不超过 4 个，因此它们的同类院线直接作为预测辅助院线；丽水时代院线的同类院线超过 4 个，因此计算它到各个同类院线的距离如表 4 所示。

表 4 丽水时代院线的特征向量 $\mathbf{x}_{\text{丽水}}$ 到同类院线的距离同类院线的特征向量

| 同类院线的特征向量 | $\mathbf{x}_{\text{金华}}$ | $\mathbf{x}_{\text{衢州}}$ | $\mathbf{x}_{\text{温州}}$ | $\mathbf{x}_{\text{台州}}$ | $\mathbf{x}_{\text{绍兴}}$ |
|---|---|---|---|---|---|
| 距离 | 1.002 | 1.001 | 1.026 | 1.759 | 2.557 |



© 中国科学院软件研究所 http://www.c-s-a.org.cn

由表 4 可得：离 $\mathbf{x}_{丽水}$ 最近的 4 个特征向量是 $\mathbf{x}_{金华}$、$\mathbf{x}_{衢州}$、$\mathbf{x}_{温州}$、$\mathbf{x}_{台州}$，因此，其对应的四个院线被选作预测辅助院线。

综上所述，丽水时代院线的预测辅助院线为：金华新农村院线、衢州新农村院线、温州新农村院线和台州新农村院线；莆田新星院线的预测辅助院线为：宁德新农村院线、福州华益院线、泉州闽南院线和厦门新农村院线；武汉春晓院线的预测辅助院线为：孝感澴川院线、黄冈新农村院线和咸宁新农村院线。

从直观上看，丽水时代院线、武汉春晓院线和各自的预测辅助院线均在同一省内，且距离相对较近；莆田新星院线和其预测辅助院线不仅均位于福建省内，而且同处沿海地区。可见预测辅助院线的选择结果符合院线放映特征的地区性规律。

2) 构造时空序列预测放映场次

根据以上预测辅助院线的选择结果，按照 2.2 节方法为每个待预测院线构造时空序列作为 BP 神经网络的学习样本。其中设置 $L=2$，则部分样本数据如表 5 所示。

表 5　部分学习样本数据

| 预测院线 | 学习样本输入 | 样本输出 |
|---|---|---|
| 丽水时代院线 | {46,59,33,23,52,58,77,37,29,60} | {63} |
| | {21,29,15,11,24,26,35,17,13,28} | {32} |
| | …… | …… |
| 莆田新星院线 | {19,8,11,5,9,67,28,20,18,24} | {223} |
| | {14,6,5,10,7,51,22,16,15,19} | {167} |
| | …… | …… |
| 武汉春晓院线 | {18,46,8,21,12,20,11,35} | {31} |
| | {8,18,5,13,7,11,6,12} | {24} |
| | …… | …… |

设置 BP 神经网络的学习速率为 0.4，学习动量为 0.3，学习次数选取 5000 次。将表 5 的学习样本进行规一化处理后对 BP 神经网络进行训练。将训练好的神经网络输入待预测院线和预测辅助院线 2009 至 2010 年 1 至 5 月放映场次的规一化时空序列，计算神经网络的输出，并将输出结果反规一化得到 2011 年 1 至 5 月的预测放映场次。将计算结果与实际放映场次进行比较，结果如表 6 所示。其中，定义：

$$\text{相对误差}=\frac{|\text{实际放映场次}-\text{预测放映场次}|}{\text{实际放映场次}}\times 100\% \quad (5)$$

表 6　院线科教片实际放映场次与预测场次对照表

| 预测院线 | 时间 | 实际场次 | 预测场次 | 相对误差/% |
|---|---|---|---|---|
| 丽水时代院线 | 2011 年 1 月 | 94 | 96 | 2.13 |
| | 2011 年 2 月 | 48 | 49 | 2.08 |
| | 2011 年 3 月 | 69 | 73 | 5.80 |
| | 2011 年 4 月 | 213 | 225 | 5.63 |
| | 2011 年 5 月 | 291 | 302 | 3.78 |
| 莆田新星院线 | 2011 年 1 月 | 729 | 743 | 1.92 |
| | 2011 年 2 月 | 554 | 560 | 1.08 |
| | 2011 年 3 月 | 822 | 836 | 1.70 |
| | 2011 年 4 月 | 621 | 646 | 4.03 |
| | 2011 年 5 月 | 353 | 369 | 5.13 |
| 武汉春晓院线 | 2011 年 1 月 | 80 | 83 | 3.75 |
| | 2011 年 2 月 | 41 | 39 | 4.88 |
| | 2011 年 3 月 | 197 | 206 | 4.59 |
| | 2011 年 4 月 | 355 | 366 | 3.10 |
| | 2011 年 5 月 | 511 | 529 | 3.52 |

## 4　性能分析

本文提出的基于时空相关性的数字电影放映场次预测，将最邻近法与时空序列结合起来：通过最邻近法选择预测辅助院线，通过时空序列的构造将待预测院线和预测辅助院线放映数据的空间属性、时间属性有机地融合，得到更加精确的预测模型。图 2 对比显示了传统的基于人工神经网络的时间序列预测方法与本文预测方法预测全国部分院线 2011 年 1 月放映场次的相对误差。由图 2 可以看到，本文提出的预测方法预测相对误差相比传统方法平均减少了大约 1/2。(图 2 中院线编号 1~13 依次代表：丽水时代院线，金华新农村院线，衢州新农村院线，温州新农村院线，武汉春晓院线，孝感澴川院线，黄冈新农村院线，咸宁新农村院线，莆田新星院线，宁德新农村院线，福州华益院线，泉州闽南院线，厦门新农村院线。)

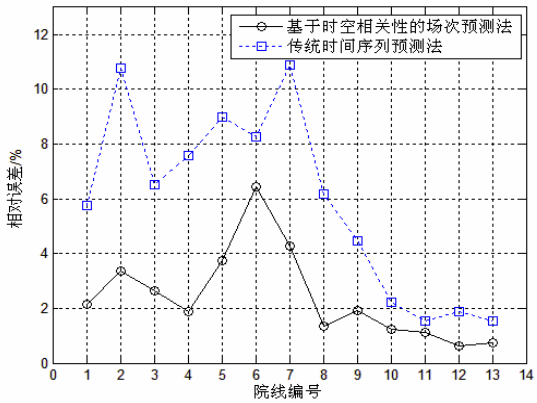


图 2　2011 年 1 月各院线放映场次预测相对误差对比图



© 中国科学院软件研究所 http://www.c-s-a.org.cn

## 5 结语

本文基于数字电影放映场次的时空相关性预测不同院线不同时期的放映场次。通过预测辅助院线的选取，为预测模型的建立增加空间相关数据；通过构造时空序列的方式将相关数据的时间属性、空间属性有机地结合起来，并借助 BP 神经网络得到预测模型。选取预测辅助院线时，本文采用改进的最邻近法，通过聚类预处理，在保证选择对象的时空相关性和选择完整性的前提下，避免了传统最邻近法中盲目搜索的问题。实验表明，本文方法的预测相对误差平均比传统时间序列预测法减少了大约 1/2，提高了预测精度。

## 参考文献


1 师瑞峰,周一民.基于数据挖掘的人口数据预测模型综述.计算机工程与应用,2008,44(9):1–6.

2 王婷婷,钱晓东.时间序列的非线性趋势预测及应用综述.计算机工程与设计,2010,31(7):1545–1549.

3 汪成亮,张硕果.通过确定邻近区域改进 KNN 文本分类.计算机系统应用,2009,18(11):56–58.

4 Eamonn JK, Michael JP. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. Proc. of the 4th International Conference on Knowledge Discovery and Data, 1998: 239–241.

5 马骊溟,徐毅,李泽湘.基于动态网络划分的散乱点 k 邻近快速搜索算法.计算机工程,2008,34(8):10–11.

6 Wang W, Li X, Wang C. River water level forecast based on spatio–temporal series model and RBF neural network. 2nd International Conference on Information Science and Engineering. 2010: 6891.

7 王建军,徐宗本.多元多项式函数的三层前向神经网络逼近方法.计算机学报,2009,32(12):2482–2488.

8 Hornik K, Stinchcombe M, White H. Universal approximation using feedforward networks with non-sigmoid hidden layer activation functions. International Joint Conference on Neural Networks (IJCNN), 1989:613.

9 刘卫宁,王鹏,孙棣华等.基于改进 BP 神经网络的道路交通事故预测.计算机系统应用,2010,19(10):177–181.


（上接第 236 页）

表 1 2 种算法提取的细节点、伪细节点和遗漏细节点比较

| | 文献[4]算法 | | | 本文改进算法 | | |
|---|---|---|---|---|---|---|
| | 细节点总数(个) | 遗漏细节点数(个) | 伪细节点数(个) | 细节点总数(个) | 遗漏细节点数(个) | 伪细节点数(个) |
| 图像 1 | 39 | 6 | 4 | 42 | 1 | 1 |
| 图像 2 | 25 | 4 | 2 | 27 | 0 | 0 |
| 图像 3 | 54 | 10 | 7 | 56 | 3 | 2 |
| 图像 4 | 46 | 9 | 5 | 49 | 2 | 1 |
| 图像 5 | 36 | 6 | 4 | 39 | 0 | 1 |

## 6 总结

本文针对文献[4]算法提取细节点后存在遗漏细节点和伪细节点的不足，在原有二值化图像提取算法的基础上提出了一种改进的提取算法。该算法通过采用形式分割图段、结合纹线方向提取、合并图段等措施，减少了伪细节点和遗漏细节点的出现。实验证明，该算法能显著地减少二值化指纹图像提取细节点过程中出现伪细节点和遗漏细节点的情况，从而提高特征提取的准确性，是一种较为实用的二值化指纹图像细节点提取算法。

## 参考文献


1 田捷,陈新建,张阳阳，等.指纹识别技术的新进展.自然科学进展,2006,16(4):400–408.

2 Jain A, Hong L, Pankanti S, Bolle R. An identity-authentication system using fingerrprint. Proc. of IEEE, 1997,85(9).

3 Jain A. On-Line fingerprint verification. IEEE Trans. PAMI, 1997,19(4):302–324.

4 傅景广,许刚,王裕国.基于二值图像的指纹细节点提取.计算机研究与发展,2004,41(4):720–727.

5 Jiang XD, Yau WY. Detecting the Fingerprint minutiae by adaptive tracting the gray-level ridge. Pattern Recognition, 2001,34: 999–1023.

6 Maio D, Maltoni D. Direct gray-scale minutiae detection in the fingerprints. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997,19(1).




© 中国科学院软件研究所 http://www.c-s-a.org.cn