

一种概率过程神经网络模型及分类算法

许少华^{1,2}, 李盼池¹, 何新贵²

(1. 大庆石油学院 计算机与信息技术学院, 黑龙江 大庆 163318; 2. 北京大学 信息科学技术学院, 北京 100871)

摘要:针对动态信号分类及与先验类别知识融合问题,提出了一种概率过程神经网络模型.模型将贝叶斯概率分类机制与过程神经网络动态信号处理方法相结合,通过在前馈过程神经网络中增加一个模式单元层,以及采用归一化指数类型激励函数,实现基于贝叶斯规则的动态信号分类.分析了概率过程神经网络分类机制与贝叶斯分类规则的等价性,给出了具体的学习算法,实验结果验证了模型和算法的有效性.

关键词:动态信号分类;贝叶斯规则;概率过程神经网络

中图分类号: TP183 **文献标识码:** A **文章编号:** 1673-4785(2009)04-0283-05

Combined probabilistic process neural network and classification algorithm

XU Shao-hua^{1,2}, LI Pan-chi¹, HE Xin-gui²

(1. School of Computer and Information Technology, Daqing Petroleum Institute, Daqing 163318, China; 2. School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

Abstract: A probabilistic process neural network has been proposed in order to provide integration of a priori knowledge with dynamic information classification. In this model, Bayesian classification was combined with the dynamic information processing of process neural networks. Dynamic information classification based on Bayesian rules was realized by adding a pattern neuron layer and a summing neuron layer to a feed forward process neural network and applying the normalized exponential activation function to the hidden layer. Classification equivalence between probabilistic process neural networks and Bayesian rules was analyzed and a concrete learning algorithm presented. Experimental results showed the effectiveness of the proposed model and algorithm.

Keywords: dynamic signal classification; Bayesian rules; probabilistic process neural networks

概率神经网络是由 Specht 提出的^[1],它与统计信号处理的许多概念有着紧密的联系,在结构上类似于反向传播网络,两者的主要区别在于采用以统计方法推导的激励函数代替传统神经网络的 Sigmoid 型激活函数.概率神经网络的理论基础是贝叶斯决策理论,该理论以其独特的不确定性知识表达形式、丰富的概率表达能力、综合了先验知识的增量学习等特性,在当前数据挖掘领域被广泛关注.其基本思想是将已知的类条件概率和先验概率,利用贝叶斯公式估计后验概率,然后根据后验概率的大小进行决策分类^[2].许多研究表明概率神经网络具有

易于训练、收敛速度快、适于实时处理等特性,并可进行任意非线性变换,其判决曲面与最优贝叶斯准则曲面相接近;同时具有很强的容错性,并且各层神经元数目比较固定,因而易于硬件实现等^[3-6].目前,概率神经网络已广泛应用于模式分类^[7]、信号处理^[8]、目标跟踪^[9]等领域中.

在实际工程领域与科学研究中,存在大量动态信号模式识别与分类问题,许多系统的输入是依赖于时间变化的函数.笔者针对动态信号模式分类、故障诊断以及动态系统状态预测等时变信息处理问题,提出和建立了过程神经网络理论和模型^[10-11].过程神经网络可以直接把时变过程作为输入输出信号,是传统人工神经网络在时间域上的一种扩展.对于缺乏先验知识和模型的复杂非线性动态系统的仿真建模、系统辨识、过程模拟以及泛函逼近等问题,过程神经网络表现出明显优势,目

收稿日期:2009-07-16.

基金项目:国家自然科学基金资助项目(60572174);黑龙江省教育厅科学技术研究资助项目(11521013);黑龙江省自然科学基金资助项目(ZA2006-11);黑龙江省科技攻关资助项目(GZ07A103).

通信作者:许少华. E-mail: xush62@163.com.

前已在时变信号的模式识别^[12]、故障诊断^[13]、预测预报^[14]等许多领域获得成功的应用. 文中通过将过程式的时变(函数)信息与文献[1]中的概率神经网络相融合, 提出和建立了一种过程概率神经网络(probabilistic process neural networks, PPNN)模型, 设计了相应的学习算法, 实验结果表明该模型及算法是可行的.

1 贝叶斯决策相关理论

1.1 贝叶斯定理

贝叶斯(Bayes Thomas)在《论机会学说问题的求解》一文中, 提出了一种归纳推理的理论, 其中的“贝叶斯定理(或贝叶斯公式)”给出了在已知结果 E 后, 对所有原因 C 计算其条件概率(后验概率)的公式, 可以看做是最早的一种统计推断程序. 以后被一些统计学者发展为一种系统的统计推断方法, 称为贝叶斯方法. 其基本内容是: 通过先验概率 $P(A)$ 与条件概率 $P(B|A)$ 来估计后验概率 $P(A|B)$, 即

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (1)$$

1.2 贝叶斯判定策略

用于模式分类的判定规则或策略的公认标准是: 在某种意义上, 使预期风险最小. 这样的策略称贝叶斯策略. 以2类判别为例, 设模式状态 P 为 P_A 或 P_B , 欲根据 n 维向量 $\mathbf{X} = [x_1 \ x_2 \ \cdots \ x_n]$ 描述的一组测量结果, 判定 $P = P_A$ 或 $P = P_B$, 贝叶斯判定规则变成:

$$d(\mathbf{X}) = \begin{cases} P_A, & h_A l_A f_A(\mathbf{X}) > h_B l_B f_B(\mathbf{X}), \\ P_B, & h_A l_A f_A(\mathbf{X}) < h_B l_B f_B(\mathbf{X}). \end{cases} \quad (2)$$

式中: $f_A(\mathbf{X})$ 和 $f_B(\mathbf{X})$ 分别为 A 和 B 的概率密度函数; l_A 为 $P = P_A$ 时判定 $d(\mathbf{X}) = P_B$ 的损失函数, l_B 为 $P = P_B$ 时判定 $d(\mathbf{X}) = P_A$ 的损失函数(正确判定时的损失等于0); h_A 为 $P = P_A$ 的先验概率, $h_B = 1 - h_A$ 为 $P = P_B$ 的先验概率.

贝叶斯判定规则 $d(\mathbf{X}) = P_A$ 的区域与贝叶斯判定规则 $d(\mathbf{X}) = P_B$ 的区域间的界限可用下式求得

$$f_A(\mathbf{X}) = K f_B(\mathbf{X}). \quad (3)$$

式中: $K = h_B l_B / h_A l_A$.

使用式(3)的关键是根据样本模式估算概率密度函数的能力. 通常先验概率为已知, 或者可以准确地加以估计, 损失函数需要主观估计. 然而若模式的概率密度函数未知, 并且给出的是一组训练模式(训练样本), 得出概率密度函数的唯一线索只有这些样本.

1.3 概率密度函数估计方法

判别边界的准确度取决于概率密度函数估计的

准确度. Parzen 提出了 $f(\mathbf{X})$ 的一簇估值公式:

$$f_n(\mathbf{X}) = \frac{1}{n\lambda} \sum_{i=1}^n \varpi\left(\frac{\mathbf{X} - \mathbf{X}_{Ai}}{\lambda}\right). \quad (4)$$

同时, Parzen 证明了 $\lim_{n \rightarrow \infty} |f_n(\mathbf{X}) - f(\mathbf{X})|^2 = 0$.

Cacoullos 扩展了 Parzen 的结果, 在 Gaussian 核的特殊情况下, 多变量估计可表达为

$$f_A(\mathbf{X}) = \frac{1}{(2\pi)^{p/2} \sigma^p} \cdot \frac{1}{m} \sum_{i=1}^m \exp\left[-\frac{(\mathbf{X} - \mathbf{X}_{Ai})^T (\mathbf{X} - \mathbf{X}_{Ai})}{2\sigma^2}\right]. \quad (5)$$

式中: i 表示样本号, m 表示训练样本总数, \mathbf{X}_{Ai} 表示类别 P_A 的第 i 个样本, σ 表示平滑参数, p 表示度量空间的维数.

2 概率过程神经元网络

在众多时变动态信号处理问题中, 受多种非线性扰动因素、信号间的耦合作用以及噪声的影响, 对于综合评判的结果, 不能给出完全精确的肯定(取值为1)或否定(取值为0)的回答; 而往往存在某种程度上的肯定或否定(取值为0~1之间), 即评判结果呈现一定的概率性或模糊性. 对这类问题的解决, 期待着新模型的出现, 这种新模型应该既体现评价结果的概率性, 即应与贝叶斯决策理论的判决结果相一致, 又能处理各种时变的过程(函数)信号. 针对这一问题, 本节首先提出过程概率神经元的概念, 进而提出一种过程概率神经网络模型.

2.1 概率过程神经元模型

与普通过程神经元模型相似, 笔者提出的概率过程神经元(probabilistic process neuron, PPN)由时变信号输入、时空加权聚合以及激励输出等运算构成, 模型如图1所示.

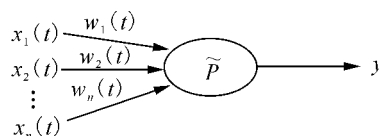


图1 概率过程神经元模型

Fig. 1 Probabilistic process neuron model

图1中, $x_1(t), x_2(t), \dots, x_n(t)$ 为时间 $[0, T]$ 上的过程式输入, $w_1(t), w_2(t), \dots, w_n(t)$ 分别为各维输入的加权函数, y 为输出. 与普通过程神经元所采用的 Sigmoid 型激励函数不同, 该模型采用具有概率统计特性的指数型激励函数:

$$g(x) = \exp\left[\frac{x-1}{\sigma^2}\right]. \quad (6)$$

式中: σ 为平滑参数, $\sigma = 0.4$ 时, $g(x)$ 的分布如图2

所示.

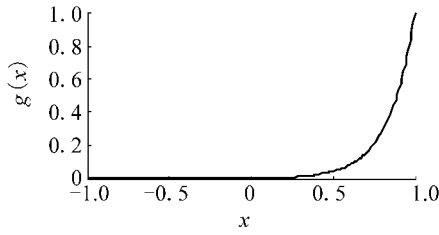


图2 激励函数 $g(x)$ 的图像

Fig. 2 Image of activation function $g(x)$

由图2可知,当 $x \in [-1, 1]$ 时, $g(x) \in (0, 1]$, 因此,对自变量 x 具有计算概率的意义. 概率过程神经元的输入输出关系为

$$y = \exp \left[\frac{\int_0^T \mathbf{W}(t) \mathbf{X}(t) dt - 1}{\sigma^2} \right] = \exp \left[\frac{\int_0^T \sum_{i=1}^n w_i(t) x_i(t) dt - 1}{\sigma^2} \right]. \quad (7)$$

2.2 概率过程神经网络模型

概率过程神经网络模型由4层组成,第1层为时变信号输入层,每个神经元均为单输入单输出,其传递函数为线性函数;第2层隐层由概率过程神经元组成,与输入层之间为全互连接,该层权函数矩阵由训练样本的各个聚类中心样本函数向量确定;第3层为模式层,根据隐层权函数向量的类别属性对第2层输出进行选择性地求和;第4层为输出层,最终完成非线性映射,如分类、函数逼近和预测等. PPNN 模型如图3所示.

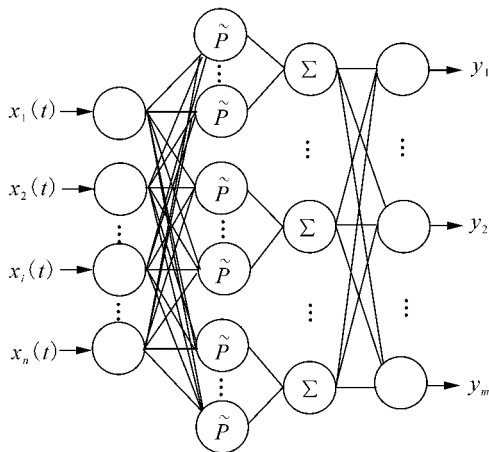


图3 概率过程神经网络模型

Fig. 3 Probabilistic process neural network model

对于图3所示模型,从结构上来看,只有模式层与输出层间的连接权及隐层激励函数中的平滑参数需要训练,而且输出层每个节点的输出是模式层输出的线性叠加,因此,PPNN 的训练速度将远远快于

普通前馈神经网络. 同时,对 PPNN 来说,由于模式层将隐层的输出按类别求和,即使某一训练样本与所有隐中心矢量的距离都很远,即对应于该样本的隐层所有神经元的输出都很小;但由于模式层求和的作用,使得模式层所对应的输出都有较大的值,并保证相互间能够明显区分,这样就不会出现在输出端“拒识”该样本的情况,从而保证了 PPNN 具有较强的分辨能力.

若令输入为 $\mathbf{X}(t) = [x_1(t) \ x_2(t) \ \cdots \ x_n(t)]$, 则隐层输出 $\mathbf{H} = [h_1 \ h_2 \ \cdots \ h_K]$ 的计算式为

$$h_j = \exp \left[\frac{\int_0^T \sum_{i=1}^n w_{ij}(t) x_i(t) dt - 1}{\sigma_j^2} \right]. \quad (8)$$

式中: σ_j 为第 j 个神经元的平滑参数,初值可取为

$$\sigma = \frac{d_{\max}}{\sqrt{H}} \quad (9)$$

式中: d_{\max} 是训练样本各个聚类中心向量之间的最大欧氏距离, H 是中心的数目. 模式层输出为

$$p_k = \sum_{j \in \Omega_k} h_j \quad (k = 1, 2, \cdots, m) \quad (10)$$

式中: m 为样本实际的类属模式数, Ω_k 为第 k 个模式包含的隐层节点序号集合.

网络的最终输出为

$$y_k = \sum_{j=1}^m v_{jk} p_k = \sum_{j=1}^m v_{jk} \sum_{s \in \Omega_j} \exp \left[\frac{\int_0^T \sum_{i=1}^n w_{is}(t) x_i(t) dt - 1}{\sigma_s^2} \right], \quad k = 1, 2, \cdots, m. \quad (11)$$

式中: v_{jk} 为输出层连接权.

概率过程神经元采用式(1)定义的指数函数为激励函数,网络结构与一般概率神经网络类似,因此,该网络分类机制与贝叶斯决策理论是一致的.

事实上,模型的非线性映射过程主要在隐层完成,若令 $\mathbf{W}_j(t)$ 等于训练集中的某个 $\mathbf{X}(t)$, 且 $\mathbf{W}_j(t)$ 和 $\mathbf{X}(t)$ 均已规格化成单位长度,即

$$\int_0^T (\mathbf{W}_j(t))^2 dt = \sum_{i=1}^n \int_0^T (w_{ij}(t))^2 dt = 1, \quad (12)$$

$$\int_0^T (\mathbf{X}(t))^2 dt = \sum_{i=1}^n \int_0^T (x_i(t))^2 dt = 1. \quad (13)$$

故有

$$\sum_{i=1}^n \int_0^T w_{ij}(t) x_i(t) dt \leq \frac{1}{2} \left(\sum_{i=1}^n \int_0^T (w_{ij}(t))^2 dt + \sum_{i=1}^n \int_0^T (x_i(t))^2 dt \right) = 1, \quad (14)$$

$$\sum_{i=1}^n \int_0^T w_{ij}(t) x_i(t) dt \geq -\frac{1}{2} \left(\sum_{i=1}^n \int_0^T (w_{ij}(t))^2 dt + \sum_{i=1}^n \int_0^T (x_i(t))^2 dt \right) = -1. \quad (15)$$

因此

$$-1 \leq Z_j = \sum_{i=1}^n \int_0^T w_{ij}(t) x_i(t) dt \leq 1. \quad (16)$$

根据激励函数 $g(Z_j) = \exp\left[\frac{Z_j - 1}{\sigma^2}\right]$ 可知, 隐层

输出 $0 \leq h_j = g(Z_j) \leq 1$, 求和单元简单地把隐层输出相累加, 然后经加权聚合, 在输出层给出 $[0, 1]$ 区间的实数值, 以确定最终的模式类属, 显然输出结果呈现出一定的概率性。

3 概率过程神经元的学习算法

对于图2所示PPNN模型, 可调参数为输出层权值 v_{jk} 及隐层平滑参数 σ_j , 定义误差函数为

$$E = \frac{1}{2} \sum_{s=1}^M \sum_{k=1}^m (d_{sk} - y_{sk})^2. \quad (17)$$

采用梯度下降算法, v_{jk} 及 σ_j 的调整量计算式为

$$-\frac{\partial E}{\partial v_{jk}} = \sum_{s=1}^M (d_{sk} - y_{sk}), \quad (18)$$

$$-\frac{\partial E}{\partial \sigma_j} = -2 \sum_{s=1}^M \sum_{k=1}^m (d_{sk} - y_{sk}) v_{jk} h_j \times \frac{\int_0^T \sum_{i=1}^n w_{ij}(t) x_i(t) dt - 1}{\sigma_j^3}. \quad (19)$$

v_{jk} 及 σ 的调整计算式为

$$v_{jk} = v_{jk} - \alpha \frac{\partial E}{\partial v_{jk}}, \quad (20)$$

$$\sigma_j = \sigma_j - \beta \frac{\partial E}{\partial \sigma_j}. \quad (21)$$

4 实验分析

采用如下数据集: $C_1: y_1 = x + \varepsilon$; $C_2: y_2 = -x + \varepsilon$; $-10 \leq x \leq 10$. 其中数据噪声 $\varepsilon \sim N(0, \sigma^2)$, $\sigma = 0.1$; 样本数据采样间隔为 $\Delta x = 0.20$. 由 C_1, C_2 各产生 101 个数据点作为训练样本, 如图4所示, 样本数据为 2 维. 为增强样本自身特征, 给每类样本增加第 3 维: $z = \text{sgn}(xy) (\text{abs}(xy))'$, 实验中 $r = 0.01$. 由于 C_1 和 C_2 在 $x=0$ 附近交叉, 故当 $x \in [0.4, 0.6]$ 时, 分类结果都视为正确. 网络参数见表1, 训练结果对比见表2.

实验结果表明, 本文建立的 PPNN 模型与普通 PNN 模型相比, 分类的正确数及正确率均有不同程度

的提高. 这是由于 PPNN 模型中, 隐层激励函数采用了具有概率特性的指数函数, 同时模式层对隐层输出结果起到了较好的分流作用. PPNN 所具有的概率统计特性提高了对样本模式特征的概括及提取能力.

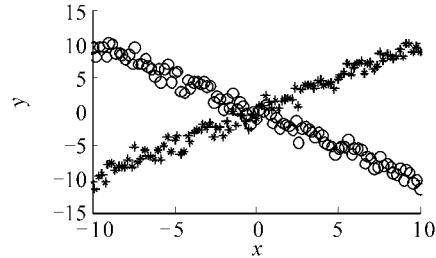


图4 线性交叉数据集聚类样本

Fig. 4 Cluster samples of linear cross dataset

表1 PPNN 参数设置

Table 1 Parameter setting of PPNN

输入节点	隐层节点	PPNN			PPN		
		迭代步数	v_{jk} 学习速率	σ 学习速率	迭代步数	隐层学习速率	输出层学习速率
3	10	100	0.8	0.5	100	0.6	0.7

表2 PPNN 训练结果

Table 2 Training results of PPNN

算法	第1类		第2类	
	正确数	正确率/%	正确数	正确率/%
PPNN	99	98.02	96	95.05
PNN	91	90.09	95	94.06

5 结束语

本文提出了一种概率过程神经网络模型及分类算法. 从神经网络的逼近能力看, 该算法可以认为是一种确定型算法; 但该模型隐层采用具有概率意义的指数激励函数, 使得模型同时具备了随机型算法的某些特征, 整个模型的推理结构与贝叶斯决策理论相一致. 概率过程神经网络采用过程式输入, 有效拓宽了普通概率神经网络的适用范围. 由于模型可调参数少, 收敛速度快, 因而适合信息的实时处理. 实验结果表明, 该模型及算法在模式分类方面具有一定潜力.

参考文献:

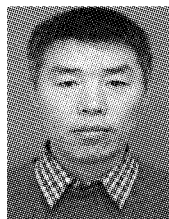
- [1] SPECHT D F. Probabilistic neural networks for classification mapping, or associative memory [C] // Proceedings of IEEE International Conference on Neural Networks. San Diego, CA, 1988: 525-532.

- [2] 史忠植. 神经网络[M]. 北京:高等教育出版社,2009: 203-208.
- [3] SPECHT D F. Probabilistic neural networks[J]. Neural Networks, 1990, 3(1): 109-118.
- [4] SPECHT D F, SHAPIRO P D. Generalization accuracy of probabilistic neural networks compared with back propagation networks[C]//International Joint Conference on Neural Networks(IJCNN-91). Singapore:1991:887-892.
- [5] 蔡曲林. 一种新的概率神经网络有监督学习算法[J]. 模糊系统与数学,2006,20(6):83-87.
CAI Qulin. New supervised learning algorithm for probabilistic neural network[J]. Fuzzy Systems and Mathematics, 2006, 20(6): 83-87.
- [6] 邢杰, 萧德云. 基于PCA的概率神经网络结构优化[J]. 清华大学学报:自然科学版,2008,48(1):141-144.
XING Jie, XIAO Deyun. PCA-based probability neural network structure optimization[J]. Journal of Tsinghua University:Science and Technology, 2008,48(1):141-144.
- [7] 杜吉祥, 汪增福. 基于径向基概率神经网络的植物叶片自动识别方法[J]. 模式识别与人工智能,2008,21(2): 206-213.
DU Jixiang, WANG Zengfu. Plant leaf identification based on radial basis probabilistic neural network[J]. Pattern Recognition and Artificial Intelligence, 2008,21(2):206-213.
- [8] 吴婷, 颜国正, 杨帮华, 等. 基于有监督学习的概率神经网络的脑电信号分类方法[J]. 上海交通大学学报, 2008,42(5):803-806.
WU Ting, YAN Guozheng, YANG Banghua, et al. Electroencephalography classification based on probabilistic neural network with supervised learning in brain computer interface[J]. Journal of Shanghai Jiaotong University, 2008,42(5):803-806.
- [9] 王昊, 张波, 田蔚风. 一种基于概率神经网络多信息融合的移动目标跟踪算法[J]. 上海交通大学学报, 2007,41(5):792-796.
WANG Hao, ZHANG Bo, TIAN Weifeng. A multi-cue fused moving object tracker based on probabilistic neural networks[J]. Journal of Shanghai Jiaotong University, 2007,41(5):792-796.
- [10] 何新贵, 梁久祯. 过程神经网络的若干理论问题[J]. 中国工程科学,2000,2(12):40-44.
HE Xingui, LIANG Jiuzhen. Some theoretical issues on procedure neural networks[J]. Engineering Science, 2000,2(12):40-44.
- [11] 何新贵, 梁久祯, 许少华. 过程神经网络的训练及其应用[J]. 中国工程科学,2001,3(4):31-35.
HE Xingui, LIANG Jiuzhen, XU Shaohua. Learning and applications of procedure neural networks[J]. Engineering Science, 2001,3(4):31-35.
- [12] 许少华, 刘扬, 何新贵. 基于过程神经网络的水淹层自动识别系统[J]. 石油学报,2004,25(4):54-57.
XU Shaohua, LIU Yang, HE Xingui. Automatic identification of water-flooded formation based on process neural network[J]. Acta Petrolei Sinica, 2004,25(4):54-57.
- [13] 何新贵, 许少华. 一类反馈过程神经网络模型及其学习算法[J]. 自动化学报,2004,30(6):801-806.
HE Xingui, XU Shaohua. A feedback process neuron network model and its learning algorithm[J]. Acta Automatica Sinica, 2004,30(6):801-806.
- [14] 丁刚, 钟诗胜. 基于时变阈值过程神经网络的太阳黑子数预测[J]. 物理学报,2007,56(2):1224-1230.
DING Gang, ZHONG Shisheng. Sunspot number prediction based on process neural network with time-varying threshold functions[J]. Acta Physica Sinica, 2007,56(2):1224-1230.

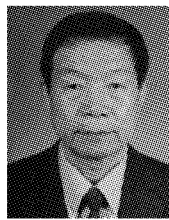
作者简介:



许少华,男,1962年生,博士,教授,博士生导师。主要研究方向为模式识别、神经网络、智能信息处理。在国内外学术期刊发表学术论文50余篇,其中被SCI、EI检索20余篇。



李盼池,男,1969年生,博士,副教授,主要研究方向为量子计算、智能优化算法及其在智能控制、智能信息处理等方面的应用。在国内外学术期刊发表学术论文30余篇,其中被SCI、EI检索10余篇。



何新贵,男,1938年生,教授,博士生导师,中国工程院院士,北京计算机学会理事长,《计算机学报》副主编。主要研究方向为模糊逻辑、神经网络、进化计算、数据库理论,发表学术论文140余篇,其中多篇被SCI、EI检索。