

基于灰关联和灵敏度的 BP 网络隐含层结构优化

张晓明, 王芳, 金玉雪, 刘晓洋

(太原理工大学 信息工程学院, 太原 030024)

摘要: 在优化 BP 神经网络隐含层结构时, 采用灰关联剪枝法是每次删除灰关联度小于灰关联阈值的隐节点, 该方法学习时间短, 但由于灰关联阈值的选取具有一定的主观性, 可能会导致误删节点或不能完全删除冗余节点; 而采用灵敏度剪枝法是每次只删除灵敏度最小的一个隐节点, 故学习时间较长; 因此, 提出一种基于灰关联-灵敏度的 BP 神经网络隐含层结构调整算法; 首先在网络前期学习过程中, 采用灰关联法对隐含层节点进行“粗删”, 直到剩余隐节点的灰关联度都大于动态灰关联阈值, 然后在网络后期学习过程中, 采用灵敏度剪枝法对隐含层节点进行“细删”, 直到删除后的学习误差增大, 则保留该节点, 并结束学习; 文章将结构优化后的神经网络应用于风电功率预测, 仿真结果表明, 该方法在满足学习误差要求的同时, 不仅精简了神经网络结构, 而且避免了灰关联剪枝法中灰关联阈值精确选取困难所带来的问题。

关键词: 隐含层结构优化; 灵敏度剪枝法; 灰关联度; BP 神经网络

Hidden Layer Structure Optimization of BP Network Based on Grey Incidence Degree and Sensitivity Degree

Zhang Xiaoming, Wang Fang, Jin Yuxue, Liu Xiaoyang

(College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: When using grey incidence method to optimize the hidden layer of BP neural network structure, this method takes a short learning time to delete the hidden redundant nodes whose grey incidence degree is less than grey incidence threshold. But the selection of grey incidence threshold has certain subjectivity, which may result in deleting useful nodes by mistake or being unable to delete redundant nodes completely. When using sensitivity pruning method, only the node with the minimum sensitivity will be deleted each time, therefore it takes a long learning time. In view of this, this paper presents a BP structure optimization method based on the grey incidence and the sensitivity degree. At the early stage of the network learning, this paper uses grey incidence method to delete the redundant nodes rapidly, until the grey incidence degrees of the remaining hidden nodes are greater than dynamic grey incidence threshold value. Then in the later learning process of the network, it uses sensitivity pruning method to delete the hidden nodes precisely, until the learning error increases after the deletion of the node. Then keep the node and stop the learning. In this paper, the neural network with the optimized structure is applied to wind power prediction. The simulation results show that this method can meet the requests for forecasting error. And it can not only simplify the structure of neural network, but also solve the problem brought by grey incidence threshold's precise determination of grey incidence method.

Keywords: hidden layer structure optimization; sensitivity pruning method; grey incidence degree; BP neural network

0 引言

研究表明, 一个 BP 神经网络的隐含层节点数目越多, 网络对函数的逼近能力就越好, 但从神经网络的学习速度来考虑, 隐含层的节点数目不宜过多^[1]。因此如何确定神经网络结构一直是 BP 神经网络实际应用时的难点和关键。在实际应用中, 常常参考 $\varphi = \sqrt{b+c} + \zeta$ 来确定隐含层节点数, 式中, ζ 取值范围为 $[1, 10]$; b 和 c 分别是网络的输入和输出节点数^[2]。近几年, 许多学者致力于这方面的研究, 并提出了几种较为有效的方法。其中, 剪枝算法是先构建一个足够大的网络, 在其学习过程中再逐个删去冗余隐节点, 并重新学习, 直到获得能满足精度要求的最小网络为止^[3]。唐万梅^[4]提出了利用灰色关联分析法(灰关联法)优化神经网络隐含层结构, 但

在删除节点时如果灰关联阈值选取的不当, 可能存在误删节点或不能完全删除冗余节点的问题。Mozer-Smolensky^[5]提出的灵敏度剪枝方法, 不仅可以剪除冗余的隐节点, 而且还可以删除冗余的输入节点。但由于它是用一个节点被删除时网络的误差变化的大小来计算节点的灵敏度, 每次只删除灵敏度最小的一个节点, 再重新学习, 导致学习时间长^[6]。因此, 本文将两种方法有机结合, 提出一种基于灰关联-灵敏度的隐含层剪枝法。首先在神经网络前期学习过程中, 采用灰关联法对隐含层节点进行“粗删”, 直到剩余隐节点的灰关联度都大于灰关联阈值, 然后在网络后期学习过程中, 采用灵敏度剪枝法对隐含层节点进行“细删”, 直到删除后的学习误差增大, 则保留该节点, 并结束学习。

1 BP 神经网络结构

BP 神经网络是 D. E. Rumelhart 和 J. L. McClelland 于 1986 年提出的, 是目前使用最广泛的神经网络之一。其基本思想是采用最速梯度下降法调整网络权值, 以使整体网络的实际输出值与期望输出值之间的误差最小。

收稿日期: 2014-04-20; 修回日期: 2014-05-24。

基金项目: 国家自然科学基金(51277127)。

作者简介: 张晓明(1989-), 男, 河北昌黎人, 研究生, 主要从事复杂系统建模与控制方向的研究。

BP 神经网络由输入层、隐含层及输出层组成，网络各层之间的神经元节点由权值连接，同一层之间的神经元相互独立，不存在相互连接关系。常用的网络多为三层结构^[7]，本文所用的 BP 神经网络结构图如图 1 所示。

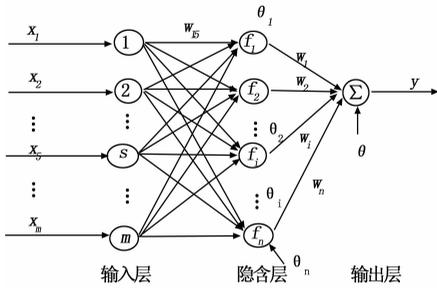


图 1 BP 神经网络结构

图 1 中，神经网络有 m 个输入， n 个隐含层节点，1 个输出。每个神经元节点的激活程度由输入、激活函数以及节点的阈值来决定，并且通过调整相邻节点之间的连接权值来增强或减弱对输出层输出和隐含层输出的作用。

2 BP 神经网络隐含层结构优化

研究表明：在神经网络学习的过程中，一部分隐含层节点对网络的输出影响大，在神经网络中处于比较重要的位置，而另一部分节点对网络的输出影响较小，对网络的作用相对较小，可以认为它们是冗余的，应该将其删除^[4]。本文采用灰关联法和灵敏度剪枝算法相结合的方法来优化 BP 神经网络隐含层结构。

2.1 基于灰关联法的隐含层结构优化

灰关联法是指根据因素之间发展趋势的相似或相异程度（即灰关联度）来衡量因素间相互关联程度的一种方法^[8]。

假设有 N 个训练样本，记 $o_i = (o_i(1), o_i(2), \dots, o_i(j), \dots, o_i(N))$ ($i=1, 2, \dots, n$) 为 BP 神经网络隐含层第 i 个神经元学习 N 个样本数据的输出； $y = (y_1, y_2, \dots, y_j, \dots, y_N)$ 为输出层神经元对 N 个学习样本数据的输出。当网络学习满足误差要求后，把神经网络输出的期望值序列 y 作为参考序列（母序列），第 i 个隐节点的输出序列 o_i 作为比较序列（子序列）。

在第 j 个样本下，隐节点 i 与网络期望输出的灰关联系数定义为：

$$\xi_i(j) = \frac{\min_i \min_j |y_j - o_i(j)| + l \times \max_i \max_j |y_j - o_i(j)|}{|y_j - o_i(j)| + l \times \max_i \max_j |y_j - o_i(j)|} \quad (1)$$

式中， $l \in [0, 1]$ 为分辨系数， l 取值越大，灰关联度值也越大，一般取 l 为 0.5； $\min_i \min_j |y_j - o_i(j)|$ 为所有样本中，网络期望输出与隐含层各节点输出之差的最小值； $\max_i \max_j |y_j - o_i(j)|$ 为所有样本中，网络期望输出与隐含层各节点输出之差的最大值。

隐节点 i 的灰关联度定义为：

$$r_i = \frac{1}{N} \sum_{j=1}^N \xi_i(j) \quad (2)$$

式中， $i = 1, 2, \dots, n'$ (n' 为当前隐节点个数)。

灰关联度的大小表明隐节点 i 的输出随网络输出值变化的

能力。灰关联度越小，表明隐节点 i 对网络输出的影响小，那么认为该隐节点是冗余的，可以删除。

删除的条件为：

$$r_i < \epsilon_{\text{iter}} = \eta^{iter} r^- \quad (3)$$

式中， r^- 为灰关联度的均值； $\eta \in (0.8, 1)$ ； $iter$ 为灰关联结构优化的次数。

本文在神经网络前期学习的过程中，首先利用灰关联法对隐含层节点进行“粗删”，直到剩余节点的灰关联度都大于灰关联阈值，然后再用灵敏度剪枝法对隐含层结构进行进一步优化。

2.2 基于灵敏度剪枝法的隐含层结构优化

灵敏度计算方法是指在网络学习后，计算隐节点对网络误差的贡献（灵敏度），每次删除贡献最小的节点，重新学习，直到删除节点后的学习误差大于未删除该节点时的学习误差，返回未删除前的结构，停止剪枝，并确定网络结构。

本文选取的网络误差 E 为：

$$E = \frac{1}{2} \sum_{j=1}^N (d^{(j)} - y^{(j)})^2 \quad (4)$$

式中， N 为学习样本数； $d^{(j)}$ 为网络对第 j 个样本的期望输出； $y^{(j)}$ 为网络第 j 个样本的实际输出。

节点灵敏度为：

$$\rho_i = E_{\text{WithoutUnit-}i} - E_{\text{WithUnit-}i} \quad (5)$$

式中， $E_{\text{WithoutUnit-}i}$ 为删除第 i 个节点后的网络误差； $E_{\text{WithUnit-}i}$ 为未删除第 i 个节点时的网络误差。

由于按式 (5) 计算灵敏度需很大的计算量，因此为逼近灵敏度 ρ_i ，对每个隐节点都引入一个重视度系数 a_i ^[9]，则对应的网络输出可表示为 $y = f(\sum_i \omega_i a_i o_i - \theta)$ ，其中 o_i 为第 i 个隐节点的输出； ω_i 为第 i 个隐节点到输出的连接权值； $f(\cdot)$ 为激活函数，本文选择 Sigmoid 函数。若 $a_i = 0$ ，则隐节点 i 对网络输出没有影响，相当于删除了该节点；若 $a_i = 1$ ，则保留隐节点 i ，于是隐节点 i 的灵敏度可表示为 $\rho_i = E_{a_i=0} - E_{a_i=1}$ ，

利用误差函数 E 对 a_i 的导数信息为：

$$\frac{\partial E}{\partial a_i} \Big|_{a_i=1} = \lim_{r \rightarrow 1} \frac{E_{a_i=r} - E_{a_i=1}}{r - 1} \quad (6)$$

对上式作工程近似，即当 $r = 0$ 时上式依然成立，则 ρ_i 为：

$$\rho_i = - \frac{\partial E}{\partial a_i} \Big|_{a_i=1} \quad (7)$$

将 (4) 式代入可得隐节点 i 的灵敏度为：

$$\rho_i = \sum_{j=1}^N (d^{(j)} - y^{(j)}) \omega_i o_i^{(j)} y^{(j)} (1 - y^{(j)}) \quad (8)$$

式中， $o_i^{(j)}$ 为第 i 个隐节点对第 j 个样本的输出； ω_i 为第 i 个隐节点的输出权值。

为了减小数据间的差异，令当前时刻隐含层内各节点灵敏度的绝对和为 $S = \sum_i^{n^*} |\rho_i|$ (n^* 为剩余隐节点个数)，则归一化后的灵敏度为：

$$\hat{\rho}_i = \frac{\rho_i}{S} \quad (9)$$

结构优化过程中，每次删除灵敏度最小的节点并重新学习

神经网络, 直到删除节点后的学习误差大于未删除该节点时的学习误差, 返回未删除前的结构, 停止剪枝。

2.3 BP 神经网络结构优化的算法步骤

基于灰关联—灵敏度的 BP 神经网络结构优化算法的具体步骤如下。

步骤 1: 根据实际问题选定网络输入和输出节点, 初始隐含层节点数选取为经验公式 $\varphi = \sqrt{b+c} + \zeta$ 的两倍以上;

步骤 2: 选定学习率、最大学习次数, 初始化神经网络的学习权值和阈值等参数;

步骤 3: 用最速下降法训练 BP 神经网络, 若满足误差要求则转至步骤 4, 否则如果超过最大学习次数则转至步骤 1;

步骤 4: 根据式 (1)、式 (2) 计算隐节点和输出节点之间的关联度, 计算灰关联阈值 ϵ_{iter} , 删除小于灰关联阈值 ϵ_{iter} 的隐节点后, 神经网络重新学习;

步骤 5: 重复步骤 4, 直到剩余节点的灰关联度都大于灰关联阈值 ϵ_{iter} , 转至步骤 6;

步骤 6: 根据式 (8) 和式 (9) 计算剩余隐节点的灵敏度 $\hat{\rho}_i$, 删除灵敏度最小的隐节点后, 神经网络重新学习;

步骤 7: 重复步骤 6, 直到删除节点后的学习误差大于未删除该节点时的学习误差, 即认为误删了此节点, 则保留此节点, 并结束学习;

步骤 8: 输出结构优化后的 BP 神经网络的各个参数。

3 基于神经网络结构优化的短期风电功率预测

近年来, 风电以其巨大的潜力被广泛应用于电网, 但由于风速具有波动性、间歇性等特点^[10], 导致风电在并入大电网时可能会造成电网电压波动、谐波污染、低压穿越等问题, 影响电网的安全运行。因此为了保证电网的安全运行和更好的电力调度, 准确的风电功率预测是非常必要的^[11]。鉴于此, 本文采用某风电场 1 200 组风速、风向、温度数据作为样本, 数据间隔一小时, 预测未来 1 小时的风电功率。本文将风速、风向和温度作为输入变量, 选取其中前 1 100 组数据作为训练数据, 后 100 组数据作为检验数据, 将样本数据归一化到区间 [0, 1] 内。初始结构选取为 3—24—1。BP 神经网络的初始权值为 [-0.1, 0.1] 内均匀分布的随机值, 最大学习次数 10 000, 学习率 0.05, 网络均方误差 MSE 的目标值小于 0.01。

首先通过灰关联分析法对神经网络结构进行“粗删”, 在网络学习后, 隐节点与输出节点间的灰关联度如表 1 所示。

表 1 隐节点的灰关联度

序号	1	2	3	4	5	6	7
灰关联度	0.400 5	0.476 4	0.763 7	0.579 5	0.519 4	0.591 2	0.530 9
序号	8	9	10	11	12	13	14
灰关联度	0.565 4	0.406 3	0.473 9	0.746 7	0.396 4	0.394 1	0.418 0
序号	15	16	17	18	19	20	21
灰关联度	0.427 3	0.448 5	0.800 1	0.557 2	0.579 8	0.742 6	0.665 9
序号	22	23	24				
灰关联度	0.390 8	0.652 7	0.501 9				

为了避免灰关联阈值选取带来的问题, 经过大量仿真, 本文取 η 为 0.87, 由式 (3) 和表 1 可算得灰色关联阈值 $\epsilon_1 = \bar{\epsilon} \times 0.87^1 = 0.472 3$ 。在删除第 1、9、12、13、14、15、16、22 号隐节点后, 重新学习并计算剩余隐节点的灰关联度, 如表 2

所示。

表 2 隐节点的灰关联度

序号	2	3	4	5	6	7	8
灰关联度	0.740 8	0.389 6	0.748 9	0.454 9	0.387 9	0.427 2	0.734 3
序号	10	11	17	18	19	20	21
灰关联度	0.395 2	0.537 3	0.409 8	0.739 1	0.669 8	0.547 8	0.813 8
序号	23	24					
灰关联度	0.781 7	0.818 5					

由式 (3) 和表 2 可算得灰色关联阈值 $\epsilon_2 = \bar{\epsilon} \times 0.87^2 = 0.454 0$, 第 3、6、7、10、17 号隐节点的关联度小于灰色关联阈值, 将其删除后, 再一次学习并计算剩余隐节点的灰关联度, 如表 3 所示。

表 3 隐节点的灰关联度

序号	2	4	5	8	11	18	19
灰关联度	0.811 2	0.821 5	0.414 3	0.803 3	0.495 8	0.809 2	0.663 7
序号	20	21	23	24			
灰关联度	0.514 6	0.885 1	0.836 6	0.874 6			

由式 (3) 和表 3 可算得灰色关联阈值 $\epsilon_3 = \bar{\epsilon} \times 0.87^3 = 0.474 7$, 在删除第 5 号隐节点后, 重新学习并计算剩余隐节点的灰关联度, 如表 4 所示。

表 4 隐节点的灰关联度

序号	2	4	8	11	18	19	20
灰关联度	0.831 1	0.800 0	0.716 3	0.530 0	0.800 2	0.546 4	0.494 0
序号	21	23	24				
灰关联度	0.784 4	0.811 7	0.739 8				

由式 (3) 和表 4 可算得灰色关联阈值 $\epsilon_4 = \bar{\epsilon} \times 0.87^4 = 0.404 1$ 。剩余节点的灰关联度都大于灰关联阈值, 神经网络的“粗删”结束。

然后再根据式 (8)、(9) 计算剩余隐节点的灵敏度, 删除灵敏度最小的节点, 再重新学习, 直到删除节点后会引神经网络学习误差增大为止, 则停止剪枝, 并保留该节点。仿真结果为, 当保留的隐节点数为 9 个时, 神经网络的学习误差 SSE 为 0.002 6, 并且此时剩余隐节点的灵敏度如表 5 所示, 当删除隐节点 18 时, 神经网络的学习误差相比上一次误差增大, 因此停止剪枝, 并且隐节点数最终确定为 9 个。

表 5 剩余隐节点的灵敏度

序号	2	4	11	18	19	20
灵敏度	0.001 6	0.146 4	-0.089 6	-0.304 5	-0.002 3	-0.106 7
序号	21	23	24			
灵敏度	0.068 7	-0.059 4	-0.220 7			

为了验证剪枝后的 BP 神经网络模型的有效性, 本文分别应用未经过结构优化、灰关联法 (阈值较大) 结构优化、灰关联法 (阈值较小) 结构优化、灵敏度法结构优化和本文方法结构优化的 BP 神经网络对 100 组检验数据进行预测, 预测误差结果如图 2、图 3 所示, 并计算均方根误差 (RMSE)、平均绝

5 结论

本文主要研究了基于 TEDS 接口的智能传感网络系统，整个系统的设计以 IEEE1451 标准为指导思想，实现了对温度和倾角传感器实时测量和步进电机控制。通过实验证明，本系统可以在变送器前端拓展传感器/执行器，最多可达 255 个。在系统总线上也可以拓展从网络适配器，有很高的实用价值，为各种复杂的传感网络系统设计提供参考。

参考文献:

[1] 卞亦文, 吴仲城, 等. 基于 IEEE 1451.2 即插即用网络化传感器

(上接第 3057 页)

对百分比误差 (MAPE) 和绝对值平均误差 (MAE), 结果如表 6 所示。

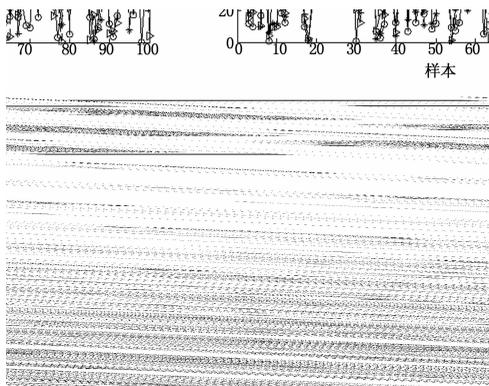


图 2 不同方法优化神经网络结构后的预测误差 (1)

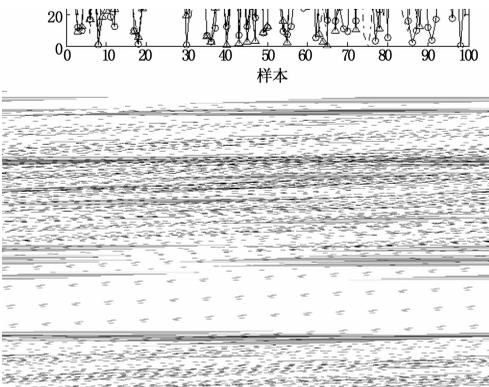


图 3 不同方法优化神经网络结构后的预测误差 (2)

表 6 不同方法优化神经网络结构后的预测误差结果

模型	RMSE(kW)	MAPE(%)	MAE(kW)
未结构优化的神经网络	60.514 4	8.44	49.645 0
灰关联法(阈值较大)优化后的神经网络	60.842 2	8.67	50.120 0
灰关联法(阈值较小)优化后的神经网络	58.146 4	7.95	46.829 0
灵敏度法优化后的神经网络	58.237 8	8.14	47.452 0
本文方法优化后的神经网络	52.298 4	6.56	39.991 0

由图 2、图 3 和表 6 可知，本文优化后的神经网络的风电

的研制 [J]. 传感技术学报, 2003, (1): 50-53.

[2] 钱开国, 李海雁, 罗金龙. 面向环境监测的传感器网络系统的设计与实现 [J]. 微型机与应用, 2011, 30 (17): 57-59.
 [3] 肖欣, 张和生, 杨军, 等. 一种交通信息采集传感器网络任务分配方法 [J]. 仪器仪表学报, 2012, 33 (11): 2529-2534.
 [4] 李泽明, 李锦明, 杨燕娇. 基于 IEEE1451.2 标准的温度传感器智能化研究 [J]. 计算机测量与控制, 2013, 21 (10): 2891-2894.
 [5] Stan Woods. IEEE-P 1451.2 Smart Transducer Interface Module [A]. Proceedings of Sensors Conference [C]. Philadelphia, 1996. Helmers Publishing Inc, 1996: 25-38.

功率预测值要优于未结构优化、灰关联法(阈值较大)结构优化、灰关联法(阈值较小)结构优化和灵敏度法结构优化的神经网络的预测值, 验证了本文方法的有效性和可行性。

4 结论

本文对 BP 神经网络的隐含层结构优化算法进行了研究, 针对灰关联法和灵敏度剪枝方法在对神经网络结构优化时, 分别存在不能完全删除冗余节点、误删节点和学习时间长的问题, 本文提出了基于灰关联-灵敏度的 BP 隐含层结构优化算法。在 BP 神经网络学习过程, 先利用灰关联法对隐含层进行剪枝, 然后再利用灵敏度剪枝法对隐含层进行剪枝, 最终确定神经网络结构。仿真实验结果表明, 本文提出的 BP 隐含层结构优化方法, 不仅精简了神经网络结构, 而且避免了灰关联法存在的关联度阈值选取困难带来的问题。

参考文献:

[1] 赵远东, 胡为尧. 神经网络泛化性能改进 [J]. 南京信息工程大学学报, 2011, 3 (2): 164-167.
 [2] 杨淑霞, 韩奇, 徐琳茜, 等. 基于鱼群算法优化 BP 神经网络的电力客户满意度综合评价方法 [J]. 电网技术, 2011, 5 (5): 146-151.
 [3] 乔俊飞, 韩红桂. 神经网络结构动态优化设计的分析与展望 [J]. 控制理论与应用, 2010, 3 (3): 350-357.
 [4] 唐万梅. BP 神经网络结构优化问题的研究 [J]. 系统工程理论与实践, 2005, 10 (10): 95-100.
 [5] Mozer M. C, Smolensky. P. Skeletonization: A Technique for Trimming the Fat from a Network via Relevance Assessment [M]. Advances in Neural Information Processing Systems, Morgan Kaufmann, 1989: 107-115.
 [6] 魏海坤. 神经网络结构设计的理论与方法 [M]. 北京: 国防工业出版社, 2005.
 [7] Peng H W, Liu F R, Yang X F. A hybrid strategy of short term wind power prediction [J]. Renewable Energy, 2013, (50): 590-595.
 [8] 扬帆, 张玉杰, 张彩丽. 发动机运行状态的主分量灰色诊断模型研究 [J]. 计算机测量与控制, 2009, 17 (1): 9-13.
 [9] 郝敏. 基于模糊聚类算法的自适应模糊神经网络研究 [D]. 哈尔滨: 哈尔滨理工大学, 2007.
 [10] Li L, Liu Y Q, Yang Y P. A physical approach of the short-term wind power prediction based on CFD pre-calculated flow fields [J]. Journal of Hydrodynamics, 2013, 25 (1): 56-61.
 [11] 叶晨. 风电功率组合预测研究 [D]. 北京: 华北电力大学, 2011.