

# 神经网络的数据分类算法在物联网中的应用

冯秀芳, 肖文炳

(太原理工大学 计算机科学与技术学院, 山西 太原 030024)

**摘要:**物联网中的基站将世界各地具有传感能力的事物结点连接到一起。它的基站需要负责处理大量数据,所以必须具备能够处理这些数据的计算能力。网络处理器(net processor)正将取代传统的基站处理器,它能在信息处理中表现出更高级别的能力。文中讨论了公共处理器(co-processor)的设计,它能帮助网络处理器分类数据包。这项工作中,基于数据的性质使用神经网络来分类数据包从而克服物联网中高维、海量数据的传输瓶颈。试验中通过对不同结构的前馈反向传播网络(BP)的误差分析,得出用于某特定物联网数据分类的最优BP网络结构。

**关键词:**数据分类;物联网;人工神经网络;公共处理器

**中图分类号:**TP391

**文献标识码:**A

**文章编号:**1673-629X(2012)08-0245-04

## Application of Neural Network Data Classification Algorithm in IOT

FENG Xiu-fang, XIAO Wen-bing

(College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China)

**Abstract:** The internet of things in the base station will be around the world with the capability of sensing nodes connected together. The base station needs to be responsible for the processing of large amounts of data, so must has enough computing ability to deal with these data. Network Processors (NP) are intended to replace conventional processors in the base station and they can bring the new levels of performances in information processing. In this work use neural network to classify packets so as to overcome the networking of high dimension, mass data transmission bottleneck. This classification is first through the different feed forward back-propagation network (BP) structure of the error analysis, so as to get the best results for particular IOT data mode.

**Key words:** data classification; IOT; ANN; co-processor

## 0 引言

人工神经网络(ANN, artificial neural networks)是20世纪80年代迅速兴起的集计算机科学、数学神经学等学科为一体的综合性交叉学科,由大量的称为神经处理单元的自律要素及这些自律要素相互作用形成的网络<sup>[1]</sup>。它是多年来在对神经科学研究的基础上,经过一定的抽象简化与模拟的人工信息处理模型在输出和输入之间建立一个非线性的映射关系。在人工神经网络的实际应用中,80%~90%的人工神经网络模型采用BP网络或它的变化形式,它也是前向网络的核心部分,体现了人工神经网络最精华的部分<sup>[2]</sup>。

物联网的基本概念是将广泛存在在人们周围的各种各样的事物和物体(例如RFID标签、传感器、执行

器、手机等)通过全球唯一的地址模式连接起来,使他们能够互相交流、通讯、协作以达到共同的目的<sup>[3~5]</sup>。

当前,具有巨大市场应用前景的物联网正在掀起第三次信息技术浪潮<sup>[6]</sup>。然而其依托的核心技术之一无线传感器网络具有能量、存储能力、计算能力、通信带宽等资源受限的突出特点,实施数据融合也是其必然选择<sup>[7]</sup>。数据分类作为数据挖掘中一个重要课题,旨在生成一个分类函数或模型,该函数或模型能把数据映射到指定类别中的某一类<sup>[8,9]</sup>。

## 1 物联网中数据分类模型

物联网系统的首要任务是准确、及时地实现各个节点之间的数据传输;其次,为了提高系统的能量效率,物联网系统需要进行数据融合。而这些待处理的数据中大部分是相关联的,所以可以结合已有的数据来实现某种信息处理方法。物联网的主体为点对点网络,主要由很多能够通过某种方式独立上网的物联网节点再加上一个或多个基站所组成,这些基站是功能更强大可以由便携式电脑或服务器所构成的节点,他们连接着世界各地余下的物联网接点。物联网节点数

收稿日期:2011-12-23;修回日期:2012-03-25

基金项目:山西省科技基础条件平台项目(2011091003-0103);山西省回国留学人员科研资助项目(2011-028)

作者简介:冯秀芳(1966-),女,教授,主要研究方向为无线传感器网络、物联网技术、人工智能、虚拟现实;肖文炳(1987-),男,硕士研究生,主要研究方向为物联网、人工神经网络。

的快速增长给构建高速网络带来了巨大的挑战,并且在传输中所产生的噪音和必要管理类型的参数也在增加。虽然这些增加的信息处理都可由基站来完成,但网络处理器(NP)正将取代传统的基站处理器,它能在信息处理中表现出更高一级别的能力<sup>[10,11]</sup>。

对于数据包的分析包括以下操作:框架和验证、分析、分类、加密/解密编辑和调度、统计信息收集和转发。数据包分类是将数据包映射到一些确定的类型的过程,它需要大多数网络处理系统协同工作。数据包分析需要最快速度执行才能避免数据的丢失、额外延时、和额外的数据包缓冲区。

对物联网节点读取的数据分类是最重要的,因为在他们所获得的数据具有大规模和极高的维度(类别很多),可以很容易地超出现有数据处理系统的存储容量和处理能力。另一方面,对于时间、空间及不同传感器输入获得的传感器网络的数据往往是自关联的,这种参数逐渐改变后的结果可以通过现有的数据分析得出。

可见数据包分类模块直接影响网络处理器的整体性能。因此,系统架构师认为专业的共处理器(co-processors)应该通过减少复杂的任务,增加明确定义和计算密集型任务等方式来提高网络处理器(NP)的性能<sup>[12]</sup>,如图 1 所示。

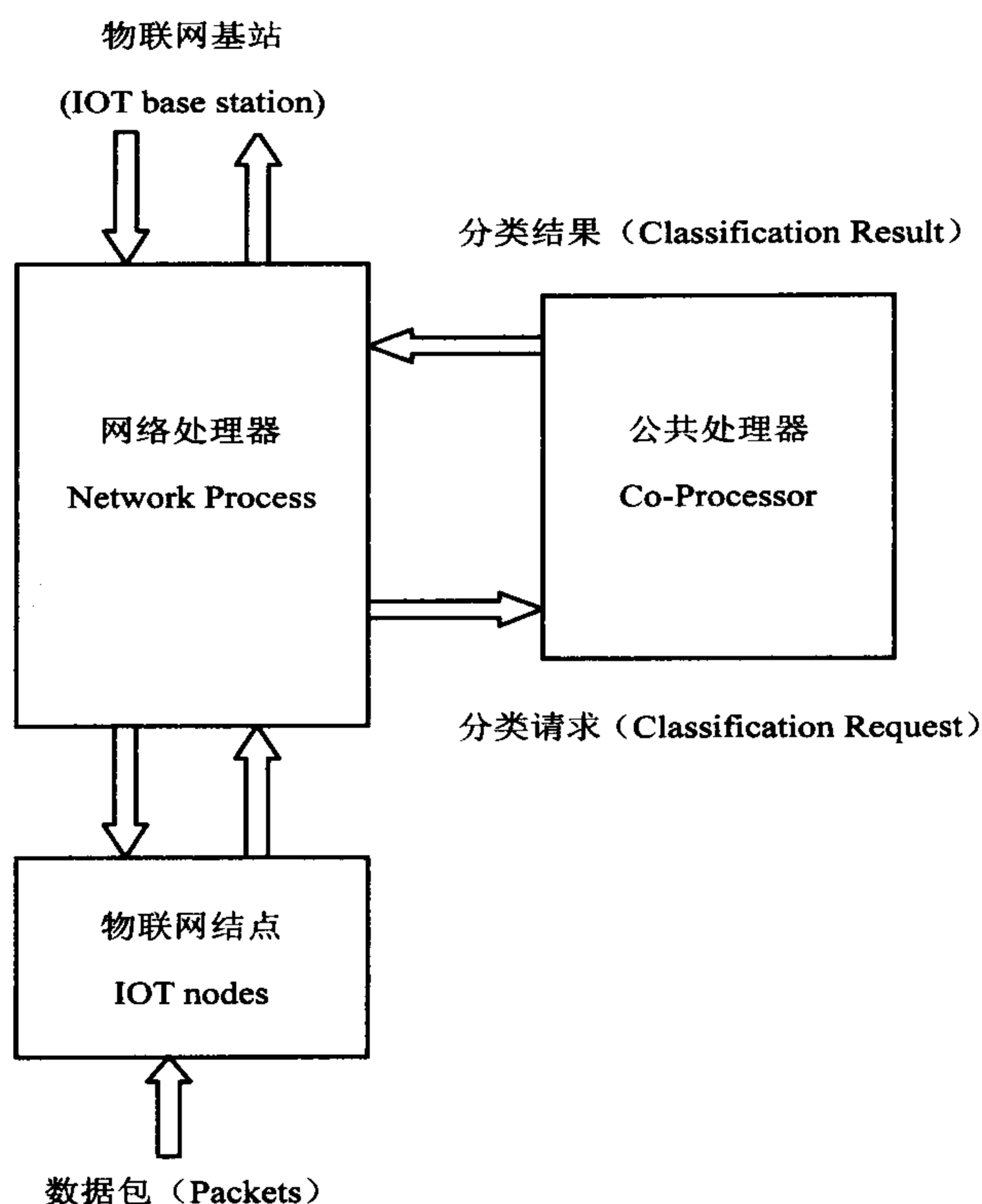


图 1 数据包分类模型

由于即将连接的物体种类很多,而不同物体的通信数据又不尽相同,这样就造成了物联网中传播的数据呈高维状态。虽然高维数据的分类是一个重要而且

具有挑战性的问题,但选取的基于常规的前馈反向传播的人工神经网络(BP)分类方法是一种有效的方法,它具有较强的分类能力和鲁棒性。

## 2 分类实验中的数据模型

物联网主体组成为一种基于事件的系统,它依赖于众多物联网节点在各种环境中细粒度的信息交互。物联网系统设计的一项重要目标是实现有效数据的快速传输,在有效数据传输的同时,用户也可以通过网络接口将他们感兴趣的信息传递给属于整个网络子集的物联网节点,一些特定的信息可以是关于某一现象的触发事件。因此,通过双向的传输功能,物联网节点将有效数据传给用户,用户也可以对自己感兴趣的部分信息进行动态筛选和查询,实现这些功能都离不开对原始数据的分类。

毫无疑问,物联网的主要优势是对日常生活和潜在用户的多个方面行为具有高度影响。在这种情况下,家庭自动化、协助生活、电子保健、智能交通等只是物联网可能的应用环境中的少数几个例子,其中新的模式将在不久的将来发挥主导作用。这预示着,物联网中的处理器将需要处理大量高维度的数据。所以,为了提高数据的处理效率,对这些原始的、高维度的数据进行分类处理是必要的。分类后的数据,能够更好地进行数据融合和数据统计等操作,这样就能提升整体网络的性能。

物联网实验中支持三种类型的事件定义,即关键词的接收、分发、过期。接收事件由一个物联网节点接收消息来产生。每一个定义事件周期的开始和结束都由计时器设置,并定义预设时间之后发生的事件为过期事件。

应用程序层为用户提供了接口(问题查询),物联网节点(查询答复)和传入的答复,可以发现通常查询应该不需要发给特定的节点,所以把基于属性或位置的查询作为实验首选。例如,“节点位置的温度高于 70F”是基于属性的查询。同样,“读取 A 区域节点的温度”,是基于位置查询。

例如,如果温差  $\Delta T$  大于预定义的阈值,这时传感器节点已准备好来传输数据。可表示为:

$$\Delta T > T_a - \alpha T \quad (1)$$

$\Delta T$  - Max temperature difference(最大温度差)

$T_a$  - Actual temperature(实际温度)

$\alpha T$  - 预定义的温度阈值

物联网中的传输可根据参数不同来分类,即源或目标 IP 地址、源端口号等。在多数情况下,连接在物联网上的设备是不需要定期传输数据的。所以,主要研究一个基于事件的方法,节点之间的信息交互由实

现设定的事件来触发。

分类的第一步是确定数据是可以做分类的,在用到的方法中,这些数据是物联网中传感器所监视的参数。要在每个数据包的基础上执行分类是不可能的,因为它的一些属性不允许这样做。具体来讲,来自不同传感器的数据包可能大小相同。所以通过基于事件发生时监控到的参数来分类数据包。

根据确定的字段或数据包的字段和分类规则,为了对数据包分类需要做一些处理操作。数据包中的字段有:

- 时间戳(在几秒钟内)

• 数据(在字节)

• 来源 IP 地址
- 来源端口数

• 目的地 IP 地址

• 目的端口数

目标是基于传感器的数据包端口号和目标 IP 地址进行分类。该数据被分成两组:训练集和测试集。他们每一个都有 24 个子集,传输的数据包括来自实验传感网络的所有节点所发出的数据包,在考虑到所有可能的数据包之后,来制定训练集。采用的传输数据如表 1 所示。

表 1 实验采用的数据属性

物联网节点捕捉到数据属性	端口号码	数据包数量
温度	60	53
湿度	52	35
气压	46	28
电流	81	46
电压	63	32
速度	45	36
其它	--	26

3 分类原理和分析

对于  $N$  个样本集合  $\{(x_k, y_k) \mid x \in R^m, y \in R^n, k = 1, 2, \dots, N\}$  的离散时间序列, ( $R$  是任意实数,  $R^m$  表示输入层有  $m$  个神经元,  $R^n$  表示输出层有  $n$  个神经元),  $BP$  网络可以完成从输入到输出的高度非线性映射,即可以找到某种映射使  $f: R^m \rightarrow R^n$ 。

定理 1: Kolmogorov 定理, 给定任一连续函数  $f: [0, 1]^n \rightarrow R^m$ ,  $f$  可以精确地用一个 3 层前向神经网络实现, 此网络的第一层即输入层有  $n$  个神经元, 中间层有  $2n + 1$  个神经元, 第三层即输出层有  $m$  个神经元。

定理 2: BP 定理, 给定任意误差  $\varepsilon$  和在  $L_2$  范数下:  $f: [0, 1]^n \rightarrow R^m$ , 存在一个 3 层  $BP$  网络, 它可以在任意  $\varepsilon$  平方误差精度内逼近  $f$ 。

这两个定理说明 BP 神经网络是建立在坚实的数学原理基础上的, 它可以对任一连续函数实现任意精度的逼近。用训练样本经过训练建立映射关系, 然后再用检测样本检验网络能否给出正确的输入-输出关系, 如果可以, 则这个网络具有很好的泛化能力。

可以采用输入节点为  $m$  个、输出节点为  $n$  个、隐节点为  $p$  个的 3 层 BP 神经网络来实现, 如图 2 所示。

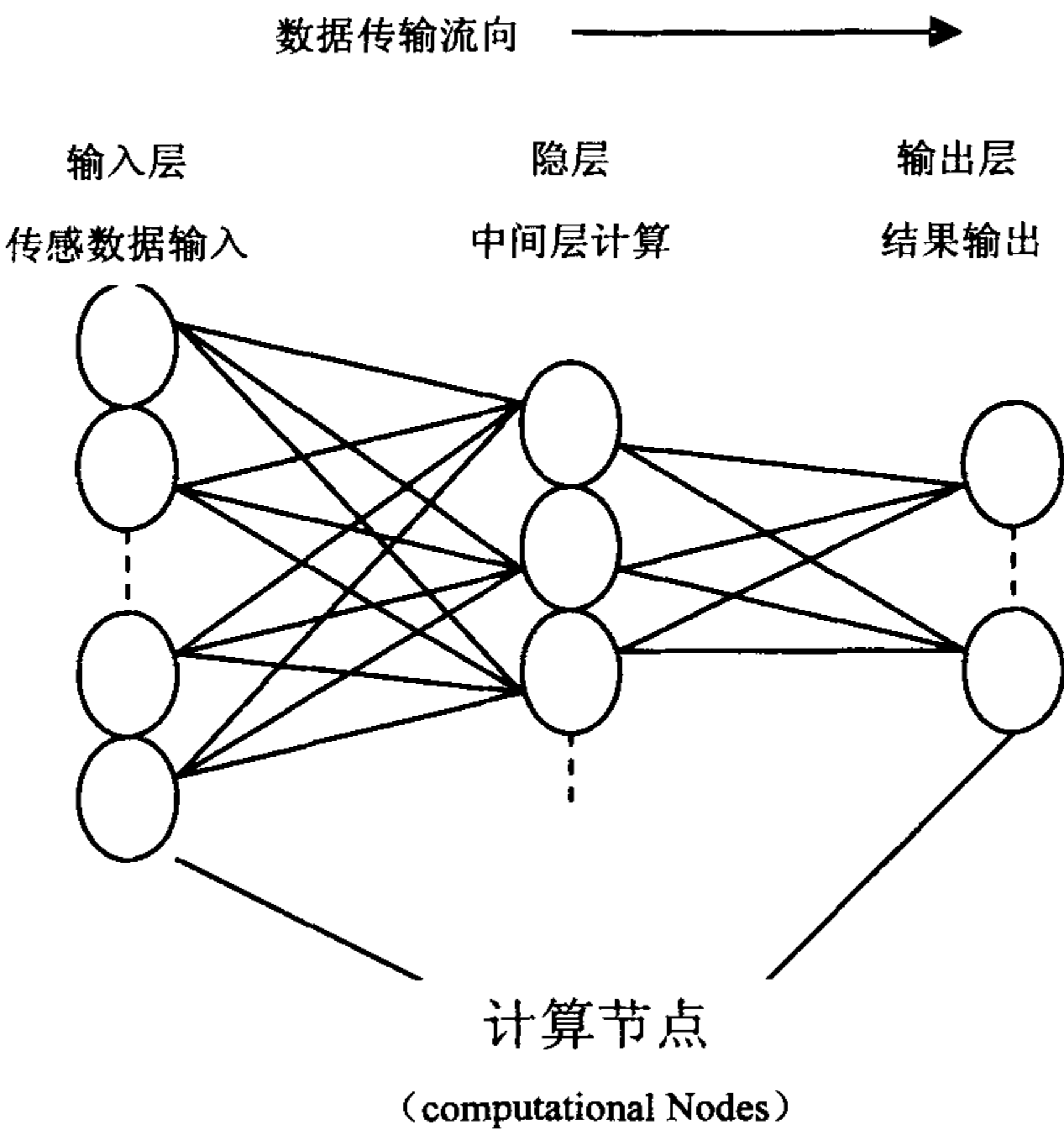


图 2 BPN 反馈网结构模型

采用的 BP 迭代算法是:

$$X_{k+1} = X_k - A_k G_k$$

(2)

$X_k$  - 当前权值和偏差向量

$G_k$  - 当前的梯度

$A_k$  - 当前的学习率.

误差计算法为 MSE

$$MSE = \sum_{p=1}^p \sum_{j=1}^k (l_{p,j})^2$$

(3)

误差  $(l_{p,j})$  的计算结果为:

$$l_{p,j} = o_{p,j} - d_{p,j}$$

(4)

$l_{p,j}$  表示输入为  $X_p$  时, 输出节点为  $j$  的输出误差。

$o_p$  - 输出模式

$d_p$  - 期望模式

输入层有 256 个神经元对应感知接点, 输出层有 6 个神经元, 每一个神经元代表一类。因此, 当网络模拟开始, 理想的输出应在 1 到 6 中, 而其他的 5 个中应该没有输出。

对于隐藏层的神经元数目选择基于以下实验的结果。为了防止其它类型的分类数据通过培训, 使用交叉验证的方法。

4 实验结果

第一次的实验用训练集来培训神经的网络, 用测试集来交叉验证。第二次的实验中, 交换使用的 2 套数据集, 做同样的测试。在每个实验中, 记录下神经网络在隐藏层中有 10 到 100 个节点的结果, 发现这些结果平均迭代次数超过了 1000。这就表示, 在拥有 10 个节点的隐藏层进行 1000 次迭代的方差已被记录, 它



的均值就作为 10 个节点的方差。

在实验中,通过 2 个参数进行研究,训练神经网络训练集的方差(在 MATLAB 下测试)和用于交叉验证所用到的测试集所产生的方差。这些参数由图 3 和图 4 中绘制出。

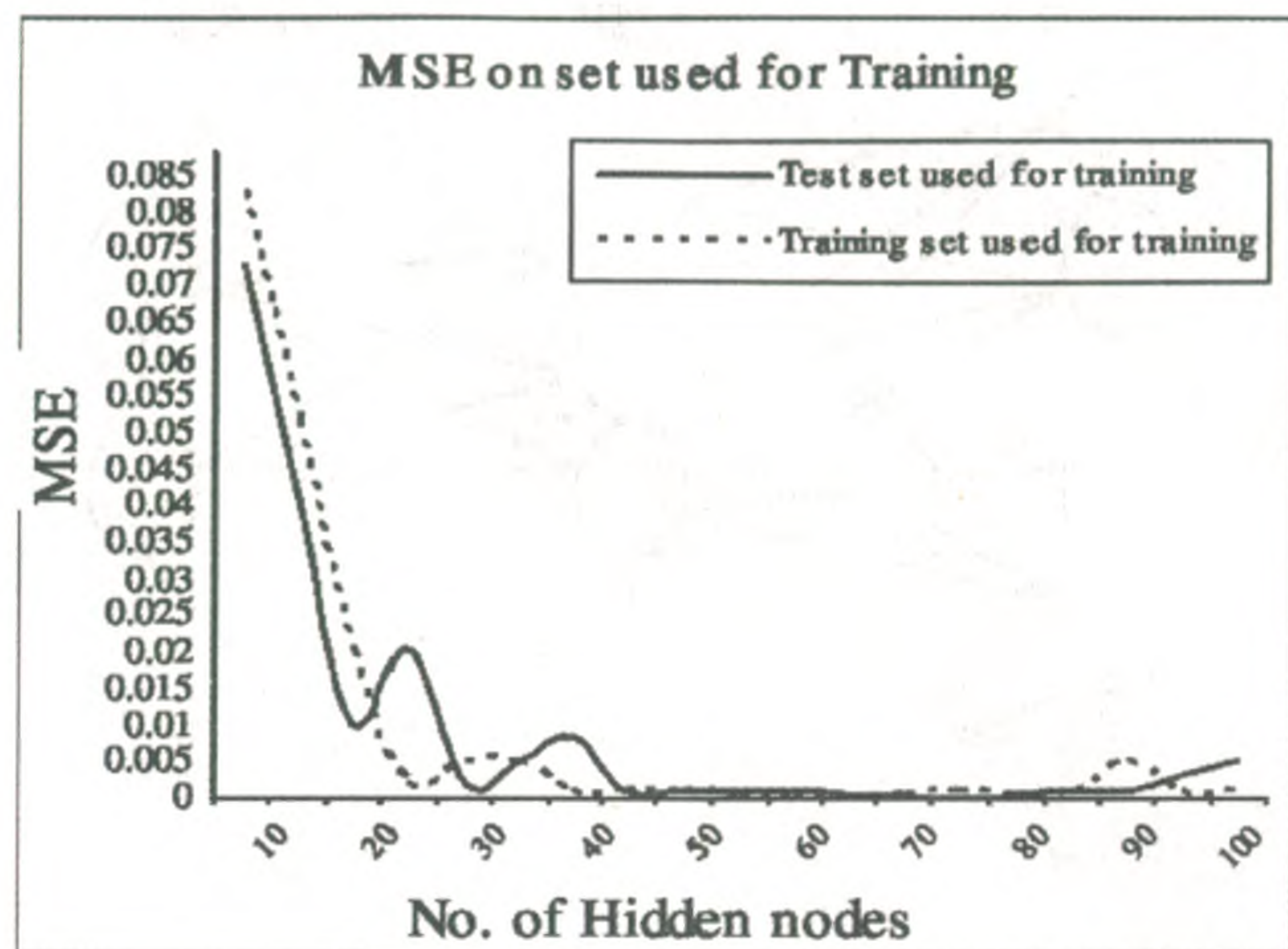


图 3 隐藏层节点数与方差关系图

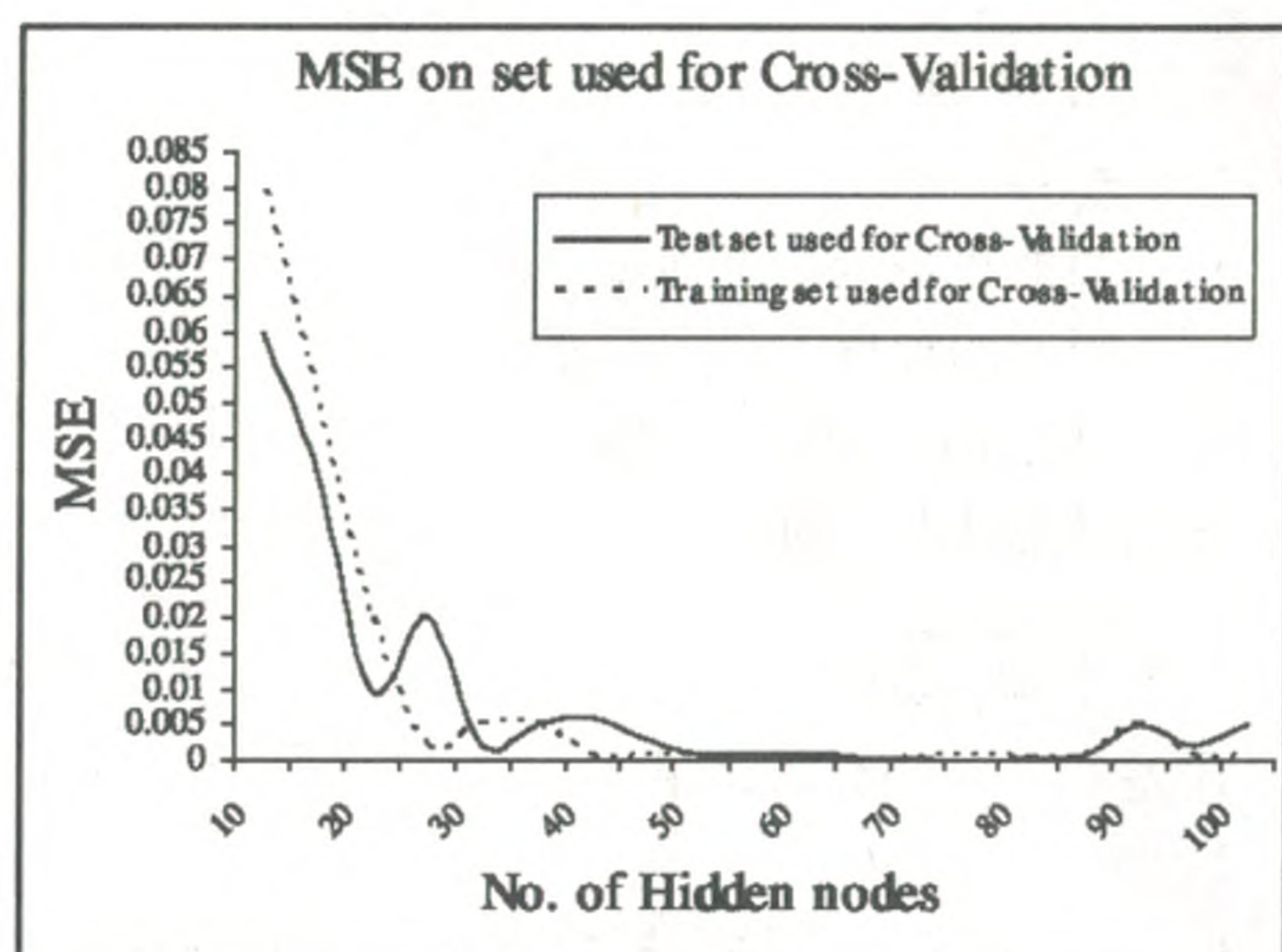


图 4 交换数据后隐藏层节点数与方差关系图

根据方差关系图可以看出隐藏节点超过 72 之后并没有带来显著性能改善。因此,用于分类的神经网络体系结构的隐藏节点数被选为 72。最终,网络结构中有 256 个输入神经元节点,72 个隐藏层神经元节点,和 6 个输出层神经元节点,作为最理想结构被决定采用。

注意到分类的结果,神经网络培训使用培训集培训,然后记录下对测试集分类的结果。输出测试中,每个输入是一个 6 维的向量,其中每个元素都来自于 6 个分类中的一种。

分类结果在列表 2 中。可以看到一共有 3 个分类出错,也就是说分类准确率为 98.82%。

## 5 结束语

对物联网中的数据分类是非常重要,因为数据传输的规模和维度都是空前的。在网络处理器的公共处

表 2 经过交叉训练后的分类结果

数据属性	原始数据	分类结果	出错个数
温度	53	53	0
湿度	35	34	1
气压	28	29	1
电流	46	46	0
电压	32	33	1
速度	36	36	0
其它	26	25	0
合计	256	256	3

理器进行数据包分类过程中,运用神经网络能提高其性能。在实验中,对有特性(可分类)数据包,使用神经网络对其进行分类。这种分类基于普通的前馈反向传播网络原理,结果证明使用该方法可以实现数据的分类。此外通过误差的分析得出最佳的神经网络结构,使分析准确率达到了 98.82%。这种人工神经网络模型能通过使用硬件描述语言(HDL),将其转换到公共处理器中(Co-Processor)。

## 参考文献:

- [1] 蒋宗礼. 人工神经网络导论[M]. 北京:高等教育出版社, 2001.
- [2] 阴江宁,肖克炎,李楠,等. BP神经网络在化探数据分类中的应用[J]. 地质通报, 2010(10): 1564-1571.
- [3] Atzori L, Iera A, Morabito G. The Internet of Things: A Survey [J]. Computer Networks, 2010, 54(15): 2787-2805.
- [4] 王保云. 物联网技术研究综述[J]. 电子测量与仪器学报, 2009(12): 1-7.
- [5] 李园园,毕晓冬,张永胜,等. 物联网框架安全威胁及相应策略研究[J]. 计算机技术与发展, 2011, 21(11): 148-151.
- [6] 郭苑,张顺颐,孙雁飞. 物联网关键技术及有待解决的问题研究[J]. 计算机技术与发展, 2010, 20(11): 180-183.
- [7] 胡向东,魏琴芳,唐慧. 物联网中数据融合的信誉度模型与仿真[J]. 仪器仪表学报, 2010(11): 2636-2640.
- [8] 汪加才,陈奇,俞瑞钊. 面向分类数据的自组织神经网络[J]. 计算机工程与应用, 2003(5): 96-101.
- [9] Zhang G P. Neural networks for classification: a survey [J]. IEEE Trans. on System, Man, and Cybernetics-Part C: Application and Reviews, 2000, 30(4): 451-462.
- [10] Zhang H, Arora A. Gs3: scalable self-configuration and self healing in wireless sensor networks [J]. Computer Networks, 2000, 43(4): 459-480.
- [11] Gupta P, McKeown N. Algorithms for Packet Classification [J]. IEEE Network Magazine, 2001, 15(2): 24-32.
- [12] Mhatre V, Rosenberg C. Homogeneous vs. heterogeneous clustered sensor networks: a comparative study [C]//IEEE International Conference on Communications (ICC 2004). Paris, France: [s. n.], 2004: 33-40.



参考文献(12条)

1. 蒋宗礼 [人工神经网络导论](#) 2001
2. 阴江宁;肖克炎;李楠 [BP神经网络在化探数据分类中的应用](#) [期刊论文]•[地质通报](#) 2010(10)
3. Atzori L;Iera A;Morabito G [The Internet of Things:A Survey](#) 2010(15)
4. 王保云 [物联网技术研究综述](#) [期刊论文]•[电子测量与仪器学报](#) 2009(12)
5. 李园园;毕晓冬;张永胜 [物联网框架安全威胁及相应策略研究](#) 2011(11)
6. 郭彪;张顺顺;孙雁飞 [物联网关键技术及有待解决的问题研究](#) [期刊论文]•[计算机技术与发展](#) 2010(11)
7. 胡向东;魏琴芳;唐慧 [物联网中数据融合的信度模型与仿真](#) [期刊论文]•[仪器仪表学报](#) 2010(11)
8. 汪加才;陈奇;俞瑞钊 [面向分类数据的自组织神经网络](#) [期刊论文]•[计算机工程与应用](#) 2003(05)
9. Zhang G P [Neural networks for classification:a survey](#) 2000(04)
10. Zhang H;Arora A [Gn3:scalable self-configuration and selfhealing in wireless sensor networks](#) 2000(04)
11. Gupta P;McKeown N [Algorithms for Packet Classification](#) 2001(02)
12. Mhatre V;Rosenberg C [Homogeneous vs.heterogeneous clustered sensor networks:a comparative study](#) 2004

本文链接: [http://d.g.wanfangdata.com.cn/Periodical\\_wjfr201208063.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjfr201208063.aspx)