一种基于神经网络的垃圾邮件过滤方法*)

张鹏鹏　张自力

(西南大学智能软件与软件工程重点实验室　重庆 400715)

摘　要　垃圾邮件问题日益严重,受到研究人员的广泛关注,基于各种技术的垃圾邮件过滤方法应运而生,其中神经网络技术应用广泛。现在主要采用的后向传播(BP)神经网络虽然在垃圾邮件过滤中取得很好的效果,但仍然存在局部极小点、不能适应新样本、学习效率较低等诸多问题。因此,本文将一种有导师、可在线学习的自组织神经网络——预测自适应谐振理论神经网络(ARTMAP),运用于垃圾邮件过滤,提出了一种新的基于 ARTMAP 的垃圾邮件过滤方法。实验表明,基于 ARTMAP 的邮件过滤能够对垃圾邮件进行有效的过滤,在保证正确率的同时,更能适应当前垃圾邮件特征不断变化的环境。

关键词　预测自适应谐振神经网络 ARTMAP,垃圾邮件,过滤

A Neural Network Based Spam Filtering Approach

ZHANG Peng-peng　ZHANG Zi-li

(Key Laboratory of Intelligent Software & Software Engineering, Southwest University, Chongqing 400715, China)

Abstract　The volume of spam in Internet has grown tremendously in the past few years. And this problem has aroused concern from many researchers and they have developed various methods to filter spam. With the development of neural network and artificial intelligence, neural network has been increasingly used in filtering spam, especially BP neural network. It has achieved good results, but there still exist many problems such as easily stacking into local minimum value, inability to meet sample and low learning efficiency, and so on. To this end, we suggest that the theory of predictive adaptive resonance, ARTMAP neural network be used for filtering spam. The experimental results show that this new method based on ARTMAP can not only filter the spam efficiently, but also adapt to the environment in which the characteristics of spam in constant change while the correct rate is ensured.

Keywords　ARTMAP, Neural network, Spam, Filter

1　引言

随着互联网的发展和普及,电子邮件已成为信息交互的重要工具。电子邮件在给人们带来极大便利的同时,其负面影响也日益突出,那就是我们每天收到的邮件中很大一部分是垃圾邮件。据统计,全世界每天的电子邮件中 10%以上是垃圾邮件。所谓垃圾邮件主要有两类,一类是名目繁多的商业广告,另一类是非法团体为其政治、经济等目的而进行的"网络宣传"。因此,垃圾邮件的智能分析、自动过滤已成为目前研究的一个热点。

目前垃圾邮件过滤技术主要可以分为安全认证方法、基于规则的方法和基于概率统计方法三种[1]。其中安全认证方法虽然有效地防止了未经认证的用户发来的邮件,具有很高的安全性,但影响了邮件的易用性。而基于规则的方法,虽然能对邮件头和正文进行有效的过滤,但其规则都是人工指定的,需要用户不断发现、总结和更新,涉及的人为因素比较多,没有经验的用户很难提供有效的规则。基于概率统计的方法往往通过某种计算表达式推出结果,它大部分是利用简单贝叶斯分类算法来过滤邮件,但其过多的简化使得很多对于分类很有用的信息丧失了,使得误判率较大,可能给用户带来更大的损失。

随着神经网络和人工智能的发展,人们正逐步考虑采取智能化的方法过滤垃圾邮件。国内外的一些学者已在尝试将神经网络用于垃圾邮件的过滤[6-11]中,但总的来说,这些研究采用的都是比较低级的有导师监督模型,如 BP 网络。因此,不可避免地存在一定缺陷:如存在局部极小点、收敛速度慢、为适应新的样本,必须使整个网络重新学习。

为了解决上述问题,本文采用预测自适应谐振理论——ARTMAP[2]进行垃圾邮件的过滤。ARTMAP 神经网络是自适应谐振(ART)理论系列中的一种,由美国 Boston 大学的 S. Grossberg 于 1991 年提出。ARTMAP 神经网络是一种有导师的自组织神经网络,对于任意的输入序列,能够进行自适应的主动在线识别,具有一定的稳定性和可塑性,其警戒参数可以随着环境的反馈而改变。同时,ARTMAP 神经网络对已经学习过的对象具有稳定的快速识别能力,并且不存在局部极小点,因此不会因样本不当而陷入权重的无休止调整。正因为 ARTMAP 神经网络有上述特点,它已经被应用在文本分类、故障诊断、控制理论中[12-14],并且都取得了不错的效果。然而,在垃圾邮件过滤方面,鲜见与 ARTMAP 神经网络技术结合的报道。

在本文的垃圾邮件过滤系统中,主要是用 ARTMAP 神经网络作为邮件过滤器,即 ARTMAP 过滤器,对已经预处理

*)受到重庆市自然科学基金资助。张鹏鹏　硕士研究生,主要研究方向为人工智能、邮件过滤;张自力　博士,教授,主要研究方向为多代理系统、人工智能、网络安全等。

过的邮件文本进行分类。通过对 BP 和 ARTMAP 的性能比较和仿真实验，可以看到 ARTMAP 过滤器在过滤邮件时，不但收敛速度快，而且比其它神经网络更能适应环境的变化。当新样本出现时，ARTMAP 过滤器能进行“智能化”的自动处理，若该样本特征与已学习过的某种模板特征的差别未超过临界参数值，则将该样本归为此类模板，若差别超过了临界参数值，则产生一种新的邮件类别。

本文后续内容组织如下：第 2 节介绍 ARTMAP 神经网络；第 3 节介绍 ARTMAP 过滤器；第 4 节讨论如何选择 ARTMAP 过滤器中的临界参数；第 5 节给出 BP 和 ARTMAP 神经网络的性能比较；第 6 节介绍实验设计和结果，并对实验结果进行分析；最后给出总结及未来工作。

2 ARTMAP 神经网络

ARTMAP 神经网络是自适应谐振理论 ART 神经网络的一种有导师学习的自组织神经网络，由两个通过 MAP FIELD 域连接的 ART 网络构成，其体系结构如图 1 所示[2]。

ARTMAP 神经网络中的 ART_a 和 ART_b 分别是独立的自组织分类模块，它们由 inter-ART 模块相连接，分别用于读取输入向量 a 和 b。这里的 inter-ART 模块包括一个 MAP FIELD 域，它通过 MAP FIELD 目标控制和 MAP FIELD 面向子系统控制类别学习。其中，MAP FIELD 目标控制是控制从 ART_a 识别的类别到 ART_b 识别的类别的相应映射的学习。MAP FIELD 面向子系统控制 ART_a 临界参数的匹配追踪。如果输入 a 引起的 ART_a 分类和输入 b 引起的 ART_b 分类在 MAP FIELD 中产生了误匹配，就增加 ART_a 的临界值，必要的话，再学习和 ART_b 分类匹配的新的 ART_a 分类。这种 inter-ART 临界参数重置信号是一种信息的后向传播形式，但又不同于 BP 神经网络中的后向传播。例如，inter-ART 重置引起的搜索可以将非直观的特征应用到新的 ART_a 分类中，这就像 S. Grossberg 介绍的用品尝的反馈结果进行绿色香蕉类学习的例子[2]，但 BP 神经网络就不能用后向传播品尝的特征作为香蕉类别区分的直观表示。

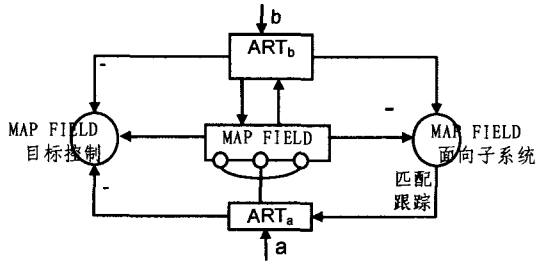

图 1 ARTMAP 体系结构

3 ARTMAP 过滤器

本文提出的 ARTMAP 过滤器，是对已经预处理过的邮件文本进行分类，一类为垃圾邮件类，另一类为合法邮件类，而无法进行分类的文本则放在同一个文件夹中，等待用户的反馈信息，一旦用户对其进行了判断，则提取其特征值并加入到相应类别中，以便下一次过滤时使用。系统框图如图 2 所示。其中，邮件预处理包括对邮件进行有用信息提取和计算特征向量两个部分。一封邮件经过去停用词、提取词干，形成单词序列，然后通过向量空间模型(VSM)的量化转变为预处理过的邮件，以便 ARTMAP 邮件过滤器使用。

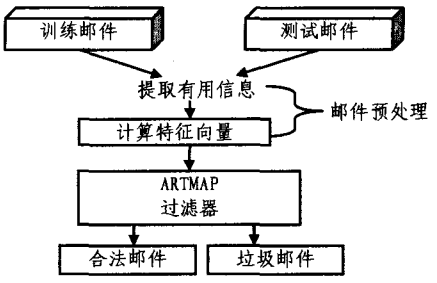

图 2 ARTMAP 过滤器系统框图

在 ARTMAP 过滤器过滤前，首先用进行过邮件预处理的训练邮件对 ARTMAP 过滤器进行训练，训练好的 ARTMAP 过滤器就可以用于邮件过滤了。过滤时，测试邮件首先进行提取有用信息和计算特征向量的邮件预处理过程，处理好的测试邮件表示为一个特征向量，使用这个特征向量作为 ARTMAP 邮件过滤器的输入，对邮件进行分类处理。

在分类处理过程中，预处理好的邮件向量作为 ART_a 端的输入，预测结果向量作为 ART_b 端的输入，两端采用相同的一种三层神经网络结构(见图 1 和图 3)。首先，F0 层接收输入向量并进行完全编码，完全编码后的向量 I 输入到 F1 层，然后在 F2 层对所有神经元进行计算，值最大的神经元获胜。若获胜神经元满足临界条件，则更新原型向量，重新激活 F2 层所有神经元，继续对下一个输入向量进行分类；若不满足临界条件，就将该神经元停用，在 F2 中继续寻找下一个值最大的神经元，如果 F2 中所有神经元都不满足临界条件，则网络将无法学习该输入向量。详细的分类算法流程如图 3 所示。

4 临界参数的选择

ARTMAP 过滤器进行邮件过滤时，临界参数值的选择非常重要，它关系着整个网络的大小。一般来说，MAP FIELD 域的临界参数 ρ 设置为 0.99，ART_a 的临界参数 ρ_a 设置为 0 或 0.7[2]时，在效率相同的情况下，网络最小。在本文的 ARTMAP 过滤器中，当用 200 封相同邮件进行过滤时，可以看到 MAP FIELD 域的临界参数 ρ 为 0.99 时，ART_a 的临界参数 ρ_a 设为 0.7，网络中的神经元个数最少，如图 4 所示。

当训练邮件是合法邮件时，临界参数 ρ_a 的设置对网络大小的影响比较小，更有甚者，无论参数如何设置，网络大小是一样的。而当邮件为垃圾邮件时，网络大小出现了明显的变化，垃圾邮件越多，参数设置为 0.7 的网络中神经元的个数就越少。所以，选择 0.7 作为 ART_a 临界参数 ρ_a，在不影响过滤效果的同时，可以减小网络大小，减少空间的占用。

5 BP 和 ARTMAP 神经网络性能比较

BP 神经网络是应用最为广泛的网络。例如，它曾经被用于文字识别、模式分类、文字到声音的转换、图像压缩、决策支持等。但是，有许多问题困扰着该算法，尤其是收敛速度慢、会陷入局部极小点和不可塑性问题。通过下面的比较实验，可以看到，ARTMAP 神经网络收敛速度快，没有陷入局部极小点的问题，并且具有可塑性。

(1)收敛速度问题

BP 神经网络的训练速度是非常慢的，尤其是当网络的训练达到一定的程度后，其收敛速度可能会下降到令人难以忍受的地步。实验中，采用 BP 和 ARTMAP 神经网络分别进行

函数逼近，函数为 $y=\sin(x^2)$。BP 神经网络采用 *tansig*()和 *purelin*()分别作为隐层和输出层神经元的传递函数，*traingd* 作为训练函数。达到 0.01 的训练精度时，BP 神经网络经过了 1373 次训练，而 ARTMAP 神经网络只用了 6 次。同样的，在垃圾邮件过滤中，ARTMAP 神经网络的收敛速度也比 BP 神经网络要快。对具有 51 个特征词的邮件文本进行分类时，BP 神经网络用了 579 次，而 ARTMAP 神经网络只用了一次就达到预期精度。其中，BP 使用 MATLAB 进行仿真训练达到收敛的过程如下：

TRAINGD, Epoch 400/1000, MSE 0.0145043/0.01, Gradient 0.0602294/1e-010

TRAINGD, Epoch 579/1000, MSE 0.00999084/0.01, Gradient 0.0416814/1e-010

TRAINGD, Performance goal met.

ARTMAP 的收敛次数显示如下：

The number of epochs needed was 1, Performance goal met.

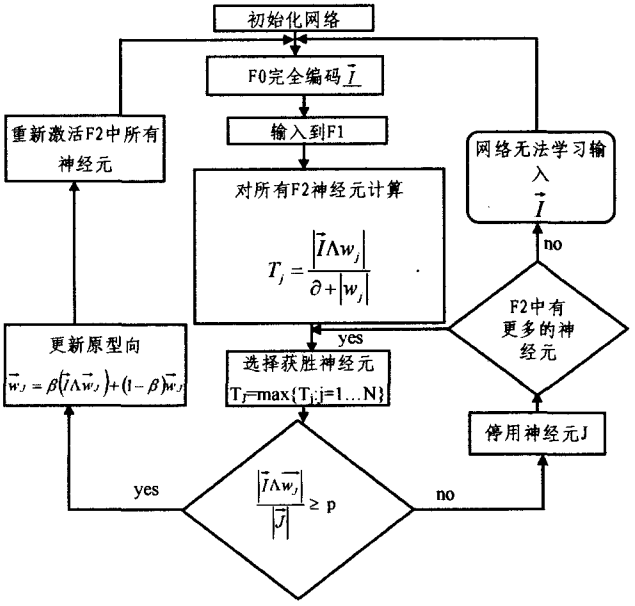

图 3　分类算法流程

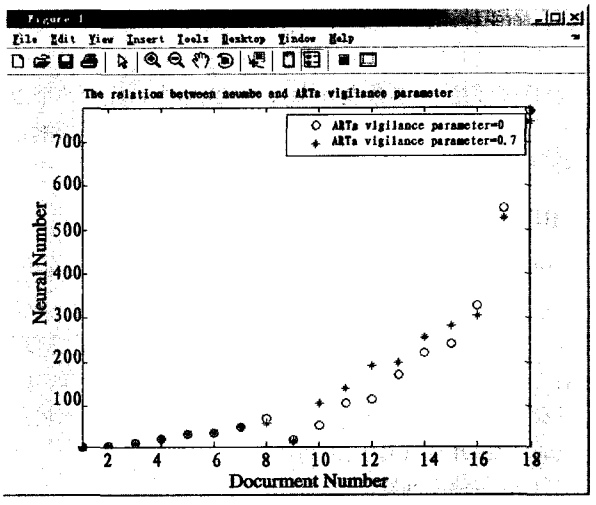

图 4　ART_a 临界参数 ρ_a

(2)局部极小点问题

在使用 BP 神经网络进行函数 $y=\sin(x^2)$ 的函数逼近时，采用 *logsig*()作为隐层和输出层神经元传递函数，*trainlm* 作为训练函数，在训练 6 次之后就陷入了局部极小点，无法逃脱，从而达不到训练目标精度。图 5 和图 6 展示了 BP 神经网络陷入局部极小点的训练过程。

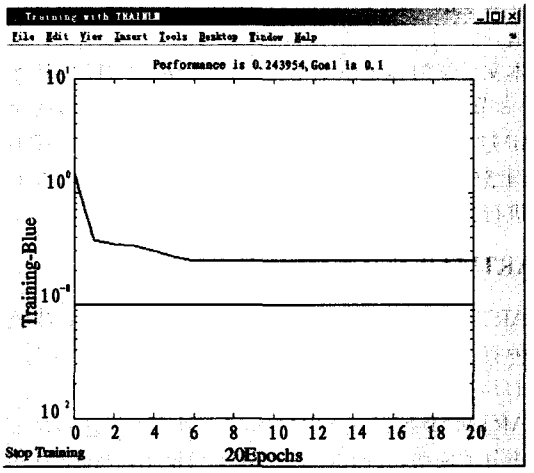

图 5　BP 神经网络训练过程

ARTMAP 神经网络进行同样的函数逼近时并没有陷入局部极小点，训练 6 次后达到精度目标。

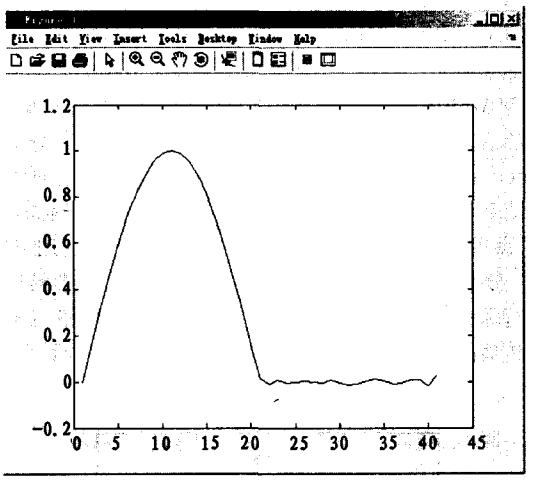

图 6　BP 神经网络陷入局部极小点

(3)可塑性问题

BP 算法必须将整个训练集一次提交给网络，再对它进行联接权的调整，在完成训练后，投入正常的运行。如果在使用的过程中，环境发生了变化，则需要重新构造一个能够表现当前环境的样本集，并用该样本集对网络进行训练。这会破坏掉网络原来已学会的内容，而只记下新的内容。这就是说，网络的长期存储内容只是它最后获得训练时系统所面对的样本集所蕴含的内容，所以，BP 神经网络不具备可塑性。而 ARTMAP 神经网络可以在原有的网络上学习新的样本，从而保证在不破坏原存储内容的基础上将新的内容增加进去。

实验中，首先使用[1,0;0,0]作为训练样本，BP 和 ARTMAP 神经网络都对其进行了训练，训练好的网络如图 7 所示。当环境发生改变，新的样本[1,0;1,1]出现时，BP 神经网络丢掉了原有已训练好的网络，对新的样本重新进行训练，如图 8 所示。而 ARTMAP 神经网络是在原有的基础上，增加了对新样本的训练，从而保证了网络的可塑性，并能适应环境的变化。

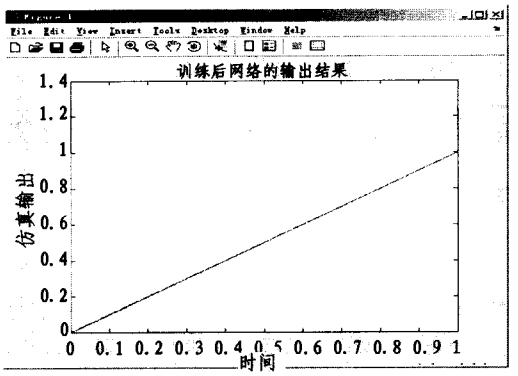

图 7　原样本训练好的网络

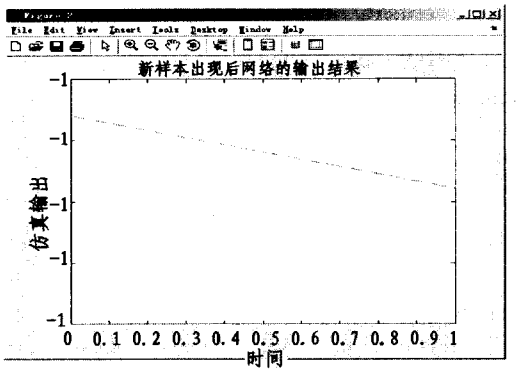

图 8　BP 新样本训练后网络

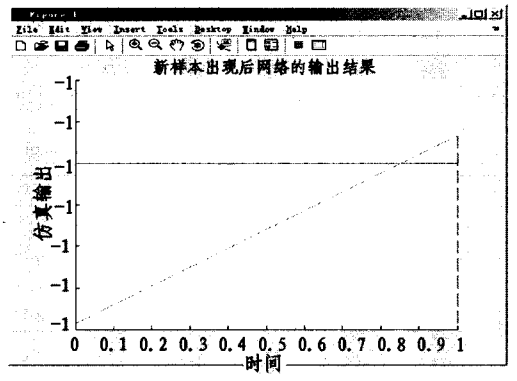

图 9　ARTMAP 新样本训练后网络

6　实验结果及分析

本文采用 Ling-Spam 语料集[3]作为实验数据，VC++语言实现分类算法，实验方法采用 10 次交叉验证。Ling-Spam 语料集是由 Androutsopoulos 等人提供，由提供者收到的垃圾邮件和来自于语言学家列表(Linguist list)的非垃圾邮件组成。首先，对语料集中的邮件文本进行去停用词和提取词干处理，得到单词序列，然后使用向量空间模型(VSM)量化邮件，并且在 Yang[5]介绍的 5 种常用特征提取方法中选取词和类别的互信息熵(Mutual Information Entropy)[3]作为特征选择算法。

这里的向量空间模型是以向量的形式表示文本[4]：

$$((t_1, f_1), (t_2, f_2), \cdots, (t_n, f_n))^T$$

其中 t_i 和 $f_i(i=1,\cdots,n)$ 分别表示项和该项的特征值(或权重)，其中项是邮件分词后的单词，特征值体现了该项在文本中的作用程度，本文使用 TFIDF 公式[2]计算特征值。

在邮件过滤过程中，ARTMAP 神经网络的 MAP FIELD 域临界参数 ρ 设置为 0.99，ART_a 的临界参数 ρ_a 设置为 0.7。

本文将邮件分为合法邮件，垃圾邮件和暂时无法判定邮件。所谓暂时无法判定邮件，是指除了垃圾邮件和合法邮件外所有邮件。因为 ARTMAP 在判定时有一些邮件会暂时无法判定其为垃圾邮件或者合法邮件，故会自动产生除了这两类邮件外的其它类别，这些类别即为暂时无法判定邮件。对这三类邮件进行如表 1 的编码(为了表述方便，这里采用 3 位二进制数进行编码)，然后使用设置好的 ARTMAP 邮件过滤器对样本邮件过滤，其仿真结果如表 2 所示。

仿真结果表明，ARTMAP 具有很好的学习特性，整个网络具有可塑性。当输入第一组样本时，由于此时的神经网络不具备任何知识，所以每输入一个样本，神经网络就产生一个新模板并储存在网络内；当输入第二组样本时，神经网络由于已经具备了这些知识，所以可以快速正确地模式匹配，作出正确的邮件分类；第三组样本包括一些已经学习过的和未学习的样本，对于已学习过的样本，神经网络立即作出了正确的判断。而对未学习的样本，ARTMAP 神经网络则进行了“智能化”的自动处理，若该样本特征与已学习过的某种模板特征的差别未超过临界参数值，则将该样本归为此类模板，若差别超过了临界参数值，则产生一种新的邮件类别，将之称为暂时无法判定邮件。这些新邮件需要经过进一步的处理，才能确定为垃圾邮件或合法邮件。一旦这些暂时无法判定的邮件被具体的确定后，比如为垃圾邮件，ARTMAP 过滤器则把这类邮件的分类模板归为垃圾邮件。

尽管 BP 神经网络和 ARTMAP 神经网络都是有导师的神经网络结构，但 ARTMAP 神经网络不存在局部极小点，并且比 BP 神经网络更能适应环境的变化，特别适合于过滤现在这种特征不断变化的垃圾邮件。本文采用的 Ling-Spam 语料集中的垃圾邮件是语言学家在生活中收集的真实的垃圾邮件，它表现出了特征不断变化的特性。ARTMAP 神经网络在能适应不断变化的垃圾邮件的同时，也能比较正确地对邮件进行类别判定，其正确率可以达到 99.01%，查全率也能达到 99.50%。

表 1　邮件编码

邮件类型	数字编码
合法邮件	101
垃圾邮件	110
暂时无法判定邮件	111
暂时无法判定邮件	011
暂时无法判定邮件	100
暂时无法判定邮件	001
暂时无法判定邮件	010
暂时无法判定邮件	000

结束语　工作垃圾邮件是目前亟待解决的问题，智能型过滤器已成为当前研究的热点。目前，把神经网络，尤其是 BP 神经网络应用于垃圾邮件的过滤较为普遍，但 BP 神经网络对垃圾邮件的误判率比较高，收敛速度慢，为适应新的样本，必须使整个网络重新学习。因此，本文提出了一种基于 ARTMAP 神经网络的垃圾邮件过滤技术。在 ARTMAP 过滤器中，把预处理好的邮件文本向量和预期的分类结果向量分别作为 ARTMAP 神经网络两端的输入，设置临界参数 ρ_a 为 0.7 进行过滤，最后输出分类结果。通过与 BP 神经网络邮件过滤的对比实验，可以看到，在具备一定的正确率的同

(下转第 208 页)

可。由于 Maekawa 算法将节点随机散布在网络中，请求节点需要逐个将消息发送给仲裁集，消息传播的多跳性导致该算法的消息复杂度激增。

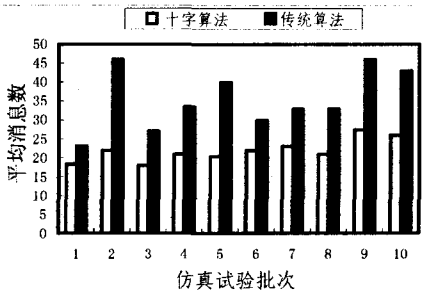

图 2　消息复杂度仿真结果对比

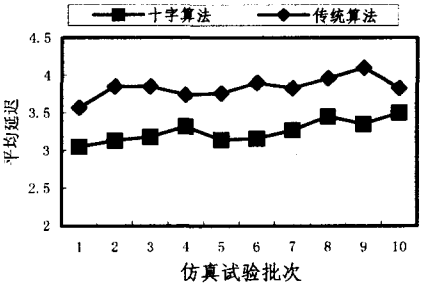

图 3　时间复杂度仿真结果对比

图 3 是 MNCME 算法和 Maekawa 算法的响应延迟对比。尽管传统 Maekawa 算法不需要遍历行列，但仲裁集随机散布在整个网络中，往往导致消息在线路上冲突，且随机扩散的仲裁集间距较大，因而使该算法在网格网络中的响应延迟反而比 MNCME 长。

结束语　本文根据网格网络的特点，提出了新型的基于十字仲裁集的分布式互斥算法 MNCME。该算法基于节点所处的行列生成十字仲裁集，并且用 Lamport 逻辑时戳保证消息的时序性。分析与仿真证明，和传统 Maekawa 算法相比，该算法具有较低的消息复杂度和时间复杂度。

参 考 文 献

[1] 舒继武，等. 大规模问题数据并行性能的分析[J]. 软件学报，2000，11(5)：628-633

[2] 黄铠，徐志伟. 可扩展并行计算技术、结构与编程[M]. 陆鑫达，等译. 北京：机械工业出版社，2000：145-223

[3] 尹俊文，邹鹏，等. 分布式操作系统[M]. 长沙：国防科技大学出版社，2001：68-82

[4] Maekawa M. A logN Algorithm for Mutual Exclusion in Decentralized Systems [J]. ACM Trans Computer Systems, 1985, 3 (2): 145-159

[5] 刘丹，刘心松. 基于读写特征的分布式互斥算法 [J]. 电子学报，2004，32(2)：326-329

[6] Fu A W. Delay-optimal Quorum Consensus for Distributed Systems[J]. IEEE Transactions on Parallel and Distributed Systems, 1997, 8(1): 59-69

(上接第 193 页)

时，ARTMAP 神经网络更符合当前垃圾邮件特征不断变化的情况，更能适应环境的变化。

表 2　仿真结果

组别	输入	匹配模板	产生新模板
1	101		合法邮件
	110		垃圾邮件
	110		无法判定
	100		无法判定
2	101	合法邮件	
	110	垃圾邮件	
	100	无法判定	
	110	无法判定	
3	111	无法判定	
	110	无法判定	
	100	无法判定	
	110	合法邮件	

虽然 ARTMAP 神经网络过滤器能有效减少用户损失，但是不同的用户对同一封邮件有不同的偏好，如一封广告邮件，对需要它的人来说就是合法邮件，对不需要它的人，它就是一封垃圾邮件，这在 ARTMAP 过滤器中并没有体现出来。因此，如何实现在过滤过程中根据用户不同偏好，实现自动化、个性化的邮件过滤，将是下一步进行研究的重点。

参 考 文 献

[1] 刘洋，杜孝平，黄星华，等. 垃圾邮件的智能过滤系统设计探讨. 微机发展，2003，13(4)：1-3

[2] Gail C, Stephen G, John R. ARTMAP: Supervised Real-Time Learning and Classificaion of Nonstationary Data by a Self-Organizing Neural Network. Neural Networks, 1991(4): 565-588

[3] 潘文峰. 基于内容的垃圾邮件过滤研究. 学位论文. 中国科学院计算技术研究所，2004

[4] 庞剑锋，卜东波，白硕，等. 基于向量空间模型的文本自动分类系统的研究与实现. 计算机应用研究，2001(9)：23-26

[5] Yang Yi Ming, Jan P. A Comparative Study on Feature Selection on Text Categorization// International Conference on Machine Learning (ICML). 1997: 412-420

[6] Zhan Chuan, Lu Xianliang, Hou Mengshu, et al. A LVQ-based neural network anti-spam email approach. ACM SIGOPS Operating Systems Review, 2004: 35-39

[7] 吴跃，王佳. 基于 NB 的双级分类模型在邮件过滤中的研究. 计算机科学，2006，33(5)：110-112

[8] 赵治国，谭敏生，李志敏. 基于改进贝叶斯的垃圾邮件过滤算法综述. 南华大学学报(自然科学版)，2006，20(1)：33-38

[9] 成宝国，冯宏伟. 一个基于 Naive Bayesian 垃圾邮件过滤器的改进. 计算机技术与发展，2006，16(2)：98-99

[10] 刘震，周明天. 基于有监督 Bayesian 网络的垃圾邮件过滤. 计算机应用，2006，6(3)：558-561

[11] James C , Irena K , Josiah P. A Neural Network Based Approach to Automated E-mail Classification // Proceedings of the 2003 IEEE/WIC/ACM international Conference on Web Intelligence(WI'03). 2003: 702-705

[12] Cheepeng L, Jennhwai L, Kuan MeiMing. A Hybrid Neural Network System for Pattern Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005, 27(4): 648-653

[13] Dimitrios C, Michael G, Takis K. Classification of Noisy Signals Using Fuzzy ARTMAP Neural Networks. IEEE Transaction on Neural Networks, 2001, 12(5): 1023-1036

[14] Kiong L. Accurate and Reliable Diagnosis and Classification Using Probabilistic Ensemble Simplified Fuzzy ARTMAP. IEEE Transactions on Knowledge and Data Engineering, 2005, 17 (11): 1589-1593