

基于BP神经网络的消费者网络 质量安全信息预警实证研究*

余凡 余红伟 许伟

摘要:针对目前产品网络质量安全事件频发的现状,本文以“天涯”论坛中“圣元”奶粉的所有帖子为数据源,构建了一个包括输入层、隐含层和输出层的三层BP神经网络。通过筛选,选取发帖单位量、发帖累计量、发帖单位频度、发帖累计频度、回复单位量、回复累计量、回复单位频度、回复累计频度8个指标作为输入层参数,选取安全、轻警、中警、重警、严重警5个预警等级作为输出层参数,利用隐含层的BP神经网络对输入层的质量信息进行计算,得到预警等级。实验证明,BP神经网络能够有效地进行预警,预警结果略低于期望值,预警敏感度还有待加强。

关键词:BP神经网络;网络质量安全;预警

一、问题的提出

质量安全信息出现后,会产生相应的负面结果。将这些事件背后的因果关系提炼出来,应用到现有的质量安全信息,能够推测出可能性结果。质量安全信息预警是政府有效预测质量安全事件并成功阻止其大规模爆发的重要手段(程虹,2013)。想要得到准确、科学的预警结果,必须要有客观、迅速、全面的质量信息源。面对近年来质量安全事件频发的现状,政府除了采取诸如制定更加严格的标准,加大处罚质量违规人员的力度等措施外,对产品从生产、流通到交易的全过程也进行了严密地监管(李酣,2013)。然而由于信息不对称造成的政府监管部门人员的有限性与需要监管产品种类的无限性之间的矛盾,政府监管技术的滞后性与产品技术发展的先进性之间的矛盾,证明无论从信息的客观性、及时性还是全面性,利用监管获取的信息源都难以进行有效地预警(罗英,2013)。大数据时代的到来,为政府有效预警带来了契机。

网络的便捷性、普遍性促使其能够成为质量预警的数据平台。作为大数据时代信息传播的主要平台,网络已经进入了人们的日常生活中。2013年7月发布的第32次《中国互联网发展状况统计报告》显示,截至2013年6月底,我国网民规模达到5.91亿,较2012年底增加2656万人。互联网普及率为44.1%,较2012年底提升2%^①。平均不到3人就有1人能够上网,加之手机等移动设备与互联网相连,我们已经生活在一个布满网络的世界。与互联网的紧密联系,使得人们更加倾向于在网络上进行交流和沟通。2013年5月发布的《2012年中国人权事业的进展》白皮书指出,截至2012年底,中国网络微博用户规模为3.09亿。据对中国最有影响的10家网站统计,网民每天发表的论坛帖文和新闻评价达300多万条,微博客每天发布和转发的信息超过2亿条^②。频繁的信息交流让人们把信息发布到网上当成一种习惯。当人们习惯于把每天经历、遇到的事情发布到微博和论坛上时,质量安全信息一定会成为人们讨论的重点之一。当消费者使用过有质量问题的产品后,这些经历都有可能出现在网络上成为

* 余凡、余红伟、许伟,武汉大学质量发展战略研究院,电子邮箱:303717695@qq.com、yuhongwei928@163.com、16661021@qq.com。本文为国家社科基金重大项目(11&ZD158)、“十二五”国家科技支撑计划项目(2011BAK06B06)和科技部质检公益性行业科研专项(201210117)的阶段性成果。衷心感谢研究院所有的研究人员对本文提出的宝贵意见。感谢匿名审稿人,文责自负。

①资料来源:中国互联网络信息中心,2013:《第32次中国互联网络发展状况统计报告》。

②资料来源:中国人大网,2013:《2012年中国人权事业的进展》白皮书。

质量安全信息。这些质量安全信息不仅全面反映了种类繁多的产品质量问题,而且能够第一时间在网络上发布,是任何其它平台无法比拟的。网络质量安全信息是能够对产品质量进行预警的非常宝贵的数据。

网络的虚拟性、完全自由性导致消费者对于网络质量信息的真实性产生怀疑。甚至还有专门从事网络造谣的炒作公司每天散播虚假信息。但是,消费者每次网购时会把其他消费者的评价信息作为衡量产品质量好坏的重要标准。评价信息千差万别,但当某件产品销售量非常大时,个别不同观点的评价会被大数据淹没,主流的评价会成为影响消费者购买动机的直接推动力。对点击率、回复率非常高的帖子进行分析,发现它们具有一个共性:提供大量的诸如图片、新闻链接、视频等能够证明信息真实性的材料。以蒙牛关于“出口比内地产品质量更好”的言论发帖为例,由于楼主提供了新闻截图,引发了接近25万人的点击和3000人的回复。如果说点击量只能说明消费者对此事比较关注,那么回复量则能说明消费者群体之间产生了共鸣。大量点击率和回复率为零的帖子没有提供能够证明其真实性的材料,其信息真实性存在问题,消费者不关注说明帖子发布的内容并不重要。从两个例子中不难发现,一方面消费者非常信赖网络上的质量评价信息,一方面在网络上发表言论并且想到得到关注时,一定会引用真实材料做信息支撑。同时大数据的特性帮助消费者过滤掉非主流信息和不值得关注的信息。综合三方面的因素可以推出,通过消费者对于网络质量信息的自由发布和选择,形成网络质量大数据,大数据会将真实的、值得关注的质量信息提供给消费者,将虚假的、不值得关注的质量信息淘汰。

网络信息传播的迅猛性、无边界性使得网络质量信息产生的负面效应不亚于质量事件本身造成的后果,甚至更为严重。网络上存在大量的诸如新浪微博、天涯论坛的虚拟社区。同一位用户可以注册多个虚拟社区,用户在不同的社区拥有不同的网络关系。整个网络由错综复杂的不同关系叠加交织在一起。当用户的网络关系越多时,质量信息一经发布,会在最短的时间内通过转载、关注的方式无限扩散,酝酿成为网络质量事件。2008年9月16日的三聚氰胺事件一经报导,仅圣元奶粉在天涯论坛的发帖数和回复数分别由前一天的3和16个上升到12和306个,可见网络质量信息传播速度之快难以想象。网络使得质量信息不对称最小化,消费者掌握的信息越多,企业面临的风险越大(中国质量观测课题组,2013)。因此网络质量信息的迅速传播对于企业造成的后果是毁灭性的。三聚氰胺事件通过网络平台让消费者了解到国内奶粉存在严重的质量问题后,全国各地的年轻妈妈们通过网络交流奶粉经验,共享使用奶粉过程中出现的各种问题,最后都义无反顾地选择国外奶粉。经过几年的质量建设工作,国内奶粉市场仍然没有太大的起色。

综上所述,消费者在网络上发布的信息能够第一时间全面、真实、迅速地反映不同产品的质量,我们能够利用网络质量信息预警。同时由于网络传播的巨大负面效应也要求我们对其预警。因此,消费者在网络平台发布质量信息之时进行严密监控,掌握信息的传播和发展趋势,提前做好预警工作,避免酿成巨大损失,是十分迫切和必要的。然而,由于网络质量信息纷繁复杂,平台和语言各异,传播迅速,很难找到质量信息的传播发展规律和预警的关键节点,这也是目前关于利用消费者网络质量信息进行预警研究非常少的原因。本文将消费者的网络质量信息为研究数据,通过预警指标的构建,尝试利用BP神经网络形成预警模型,对网络质量信息进行预警,减少其带来的负面影响。

二、文献回顾

按照预警依托的不同平台,可以把预警分为网上预警和网下预警两块。网下预警的出现、发展、应用都早于网上预警,预警主要涉及安全性要求比较高的领域。具体而言,预警主要包括以下几个方面:金融风险预警,研究者运用不同的数学模型对存在的财务危机进行预测。Fitzpatrick(1932)利用财务指标进行单变量预测分析。Altman(1968)利用多变量分析模型对财务危机进行预测。黄福员(2013)提出一种基于粗糙集理论的模糊神经网络模型,用于金融风险预警。环境预警,研究者对土地、水等环境的恶化、污染进行预警。Herrick(2002)对耕地进行监测预警。Katlan(1999)对土地的荒漠化进行预

警研究。罗艳(2013)采用生态风险预警评估法评估遵义东南部地区农业土地的重金属含量。张立辉(2013)基于韦伯-费希纳定律建立水库水环境预警评价模型。危险工作场所预警,研究者对存在重大安全隐患的场所进行预警。Lee(1992)利用神经网络算法对设备的使用情况进行预测,预防安全事件发生。马建(2013)通过多种地压监测方式对某矿综放工作面形成综合监测预警系统。疾病预警,研究者对高发且危险的疾病进行预警。盛静宇(2013)探讨了心率减速力在评价肥厚型心肌病患者自主神经功能方面的应用以及对高危患者进行预警方面的临床价值。食品安全预警,研究者对食品生产到销售各个环节中存在的安全问题进行预警。陈璐(2013)探索建立江苏省昆山市餐饮食品安全风险监测三级网络和预警信息平台。

网上预警的兴起相对较晚,预警主要集中在网络舆情控制方面。丁菊玲(2011)利用BP神经网络进行网络舆情的预警,对网络舆情的发展趋势进行评价和预测。姚福生(2013)认为主题演化是网络舆情演化的重要形式,构建了包括信息捕获与提取模块、分析评估模块和结论通报模块的网络舆情主题演化预警信息工作系统。王铁套(2012)在详细分析网络舆情影响因素和自身属性的基础上,依据模糊综合评价法构建了网络舆情预警模型,并用实例证明了该模型的有效性和准确性。刘毅(2012)利用基于三角模糊数的模糊德尔菲法和模糊层次分析法,对网络舆情指标进行了再次筛选和权重的确定,得到了面向于某一具体公共事务或热点话题的网络舆情预警指标体系。丁菊玲(2011)实现对网络论坛网帖中网民观点极性和观点强度的挖掘,构建基于三粒度挖掘结果的观点树,建立基于观点树的网络舆情危机预警架构,最后通过实验验证其可行性。王青(2011)对现有网络舆情监测指标体系进行整理与归纳,通过E-R模型系统分析主题舆情的属性特征,从舆情热度、舆情强度、舆情倾度、舆情生长度四个维度进行预警。曾润喜(2010)在问卷调查的基础上利用层次分析法构建了警源、警兆、警情三类因素和现象的网络舆情突发事件预警指标体系。

虽然网下、网上预警的侧重点不同,但基本上都采用构建评价指标、计算评价结果、划分预警等级、判断预警结果的步骤进行预警。评价方法已经非常成熟,因此预警的关键在于如何构建科学的评价指标。将网下、网上预警的文献对比,可以发现网下预警指标的领域区分性非常强,比如金融预警的常用指标有净资产收益率、主营业务利润率、资产负债率等,生态环境预警的常用指标有温度、湿度、风速、光照度等。之所以不同领域可以构建适应于该领域的指标,是因为网下预警的数据来源更加多样化。研究者可以通过精密的仪器设备、抽查、调查问卷等方式主动收集评价指标对应的数据。与网下相比,网上预警的数据来源更加单一化,只能通过互联网获取信息。互联网能够提供什么样的信息,我们就只能被动地获取什么样的信息;因此,根据网上的信息设计的评价指标具有很大的局限性,无法主动设计指标,只能被动地接收指标。

另外,网下、网上预警的关注点也是不一样的。网下收集信息的手段丰富,能够准确获取产品相关安全指标的信息,因此网下预警主要对产品进行安全性预警分析。比如金融预警中的净资产收益率就是金融领域一个非常具体的指标,当净资产收益率达到某个值时,能够从一个侧面反映出该企业的财政状况处于何种状态。相反,网上提供的信息与产品的相关性不高,与信息传播的相关性非常高,原因在于信息在网络上传播所带来的负面效应大于信息本身的负面效应,因此网上预警主要对质量事件的传播进行预警分析,更加关注事件的信息传播状态。预警的指标也主要集中在信息传播的指标上,比如:帖子点击数、回复数、参与讨论人数等。

本文主要研究消费者的网上质量信息。根据上文的分析,本文将使用网上预警指标的构建方法构建消费者网络质量信息的预警指标。

三、预警指标

本文选取的预警指标遵循以下几个原则:可获得性、客观性、可用性。

由于网上的预警指标是根据网上信息制定的,而不同的网络虚拟社区提供的信息不同,因此首先需要确定待研究的网络虚拟社区,查看该虚拟社区能够提供哪些信息,然后根据信息确定预警指标。本文

选取国内知名论坛——天涯论坛作为预警的实验数据来源。通过对天涯论坛中帖子相关属性进行分析,发现论坛能够提供的信息有发帖数量、帖子标题、楼主昵称、发布时间、点击数量、回复数量、帖子内容、回复作者昵称、回复时间、回复内容 10 个帖子属性。

在网上预警的文献中,研究者除了使用网上提供的客观信息外,还会在预警指标中加入一些主观信息,比如:帖子真实性、敏感度、主题参与度。这类指标都是用问卷调查、专家打分的方法获取。还有一类诸如主题观点倾向的指标数据会使用情感词匹配的方法从帖子内容、回复内容中获取。笔者认为采用问卷调查、专家打分的方法获取的指标数据可能会受到调查对象、专家个人经验、偏好的影响。对于情感词匹配方法,由于来自于不同地区的消费者在讨论的时候会使用本地的语言,加上中文存在一词多义、多词一义、一词多性的现象,这些因素都会降低匹配的精度,影响最后的预警结果,而采用论坛提供的客观数据能够更加真实地反映质量信息的传播规律;因此本文将从 10 个帖子属性中选择预警指标。

考虑到预警主要观察单位时间质量信息传播的数量,帖子标题、楼主昵称、帖子内容、回复作者昵称、回复内容无法使用,而论坛没有提供点击数量随时间变化的量,因此选取发帖数量、发布时间、回复数量、回复时间 4 个指标作为原始指标。对原始指标扩展,最终得到发帖单位量、发帖累计量、发帖单位频度、发帖累计频度、回复单位量、回复累计量、回复单位频度、回复累计频度 8 个指标。其中,单位量和累计量分别指发帖和回复单位时间和累计时间的数量,单位频度和累计频度分别指单位时间和累计时间帖子和回复的变化数量。

四、BP 神经网络预警及结果分析

BP(Back-Propagation)神经网络的前身是误差反向传播。误差反向传播经由 Bryson 的提出和 Werbos、Parker 等研究者的运用,反向传播开始引起广泛的关注,并被逐步引入到人工神经网络。BP 神经网络具有简单的结构、较强的可执行性、较好的自学习性,并且能够有效地解决非线性问题。因此,在众多人工神经网络的模型中,以 BP 神经网络作为研究方法的论文占了最大的比重。

(一) BP 神经网络算法和模型

BP 神经网络包括输入层、隐含层和输出层(王钰,2005)。输入层和输出层分别负责接收输入值和产生输出值。隐含层负责将输入层的值计算得出输出期望值以及反向传播输出值。BP 神经网络算法是一个误差反向传播的循环学习过程。其基本思想为,输入层接受输入值后,传播给隐含层处理,隐含层将处理结果传播给输出层。这个传播过程称为正向传播。如果输出结果与期望值的误差过大,则进入反向传播阶段。误差反向传播阶段,将输出误差反向传播给隐含层,隐含层各神经元都将接收到一部分输出误差,神经元根据输出误差调整权值。通过不断地学习和调整,输出值将无限接近期望值。直到神经网络学习次数达到预定的阈值,神经网络才会停止。

$$y_j = \frac{1}{1 + e^{-x_j}} \quad (1)$$

$$x_j = \sum_i y_i w_{ji} \quad (2)$$

BP 神经网络模型主要由公式 1、2、3、4 构成(李萍,2008)。公式 1 为神经网络的 Sigmoid 激活函数,激活函数用于输入和输出值的非线性映射转换。公式 2 为神经网络的输出值。神经网络中的神经元接收激活函数的值 y_j , 激活公式 2,公式 2 乘以权重,得到神经元的值 x_j , x_j 代入激活函数,激活函数激活下一个神经元,依次循环,直到激活所有神经元,形成神经网络。

$$\Delta W = -l \frac{\partial E}{\partial w} \quad (3)$$

$$\frac{\partial E_j}{\partial w_{ji}} = y_j (1 - y_j) (y_j - d_j) y_i \quad (4)$$

当输出值误差大于期望误差时,神经网络利用反馈机制修正神经元权重。神经元权重的修正模型

如公式 3 所示, $\frac{\partial E}{\partial w}$ 为权重的误差量, l 为学习率, 负号表示权重朝负方向调整。通过公式运算, 权重误差量可以由公式 4 表示, 其中 d_j 为期望输出值。权重的修正是一个迭代的过程, 当权重小于设定阈值, 停止修正。

(二) 基于 BP 神经网络的预警结构

通过以上算法和模型的描述不难发现, 本文的研究可以基于 BP 神经网络进行消费者网络质量信息的预警。将上文提出的 8 个指标作为神经网络的输入层, 预警等级为输出期望, 利用一段时间内的输入和输出值训练并搭建起 BP 神经网络, 最后利用 BP 神经网络对未来的消费者网络质量信息进行预警。预警框架如图 1 所示。预警等级分为安全、轻警、中警、重警、严重警。

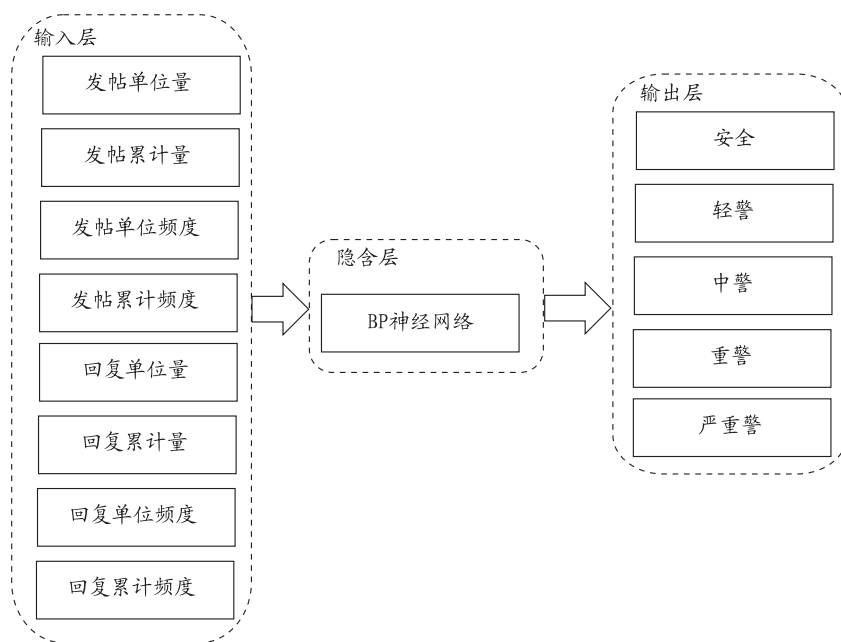


图 1 基于 BP 神经网络的预警结构

(三) 预警流程

本文将使用 Matlab 工具进行预警实验。预警流程细分为 10 个步骤, 实验结果将在对应的步骤中展示。

1. 选定数据源

选定国内知名论坛——“天涯”中与圣元奶粉质量有关的所有帖子作为数据源。数据时间从 2004 年 11 月 29 日到 2013 年 8 月 30 日。

2. 获取输入指标数据和输出预警等级

通过对原始帖子的预处理, 将原始数据以时间为轴统计到 8 个输入指标中。同时, 设定输出预警等级。预警的五个等级从安全和严重分别对应 10000、01000、00100、00010、00001 五个数字, 方便计算。

3. 训练样本输入值归一化

将所有数据分为训练样本和测试样本两个部分。首先利用训练样本构建 BP 神经网络, 然后利用测试样本检验 BP 神经网络, 查看神经网络的预警准确率。为了降低输入值带来的误差, 提高神经网络的精度, 需要对输入值进行归一化处理。本文使用“premnmx”进行归一化处理。归一化前后的输入值如表 1、2 所示。BP 神经网络的核心函数如表 3 所示。

表1 训练样本的部分输入值

时间	x1	x2	x3	x4	x5	x6	x7	x8	期望输出
2008/9/13	29	2	56	2	334	9	493	3	10000
2008/9/14	31	2	56	0	349	15	499	6	01000
2008/9/15	34	3	57	1	365	16	500	1	01000
2008/9/16	46	12	66	9	671	306	790	290	00010
2008/9/17	77	31	85	19	1422	751	1235	445	00001
2008/9/18	95	18	98	13	2196	774	1258	23	00001
2008/9/19	121	26	106	8	3157	961	1445	187	00001
2008/9/20	137	16	116	10	3541	384	2022	577	00010
2008/9/21	149	12	120	4	3831	290	2116	94	00010
2008/9/22	158	9	123	3	4042	211	2195	79	00010
2008/9/23	166	8	124	1	4250	208	2198	3	00010
2008/9/24	173	7	125	1	4424	174	2232	34	00010
2008/9/25	176	3	129	4	4485	61	2345	113	00100
2008/9/26	176	0	132	3	4522	37	2369	24	01000
2008/9/27	179	3	135	3	4620	98	2430	61	00100
2008/9/28	179	0	138	3	4640	20	2508	78	01000
2008/9/29	179	0	138	0	4643	3	2525	17	10000

表2 训练样本的归一化值

时间	x1	x2	x3	x4	x5	x6	x7	x8	期望输出
2008/9/13	-0.8182	-0.8710	-0.6615	-0.7895	-0.8805	-0.9813	-0.7218	-0.9896	10000
2008/9/14	-0.8052	-0.8710	-0.6615	-1.0000	-0.8752	-0.9688	-0.7184	-0.9792	01000
2008/9/15	-0.7857	-0.8065	-0.6554	-0.8947	-0.8694	-0.9667	-0.7178	-0.9965	01000
2008/9/16	-0.7078	-0.2258	-0.6000	-0.0526	-0.7596	-0.3632	-0.5539	0.0052	00010
2008/9/17	-0.5065	1.0000	-0.4831	1.0000	-0.4902	0.5630	-0.3022	0.5425	00001
2008/9/18	-0.3896	0.1613	-0.4031	0.3684	-0.2126	0.6108	-0.2892	-0.9203	00001
2008/9/19	-0.2208	0.6774	-0.3538	-0.1579	0.1322	1.0000	-0.1835	-0.3518	00001
2008/9/20	-0.1169	0.0323	-0.2923	0.0526	0.2700	-0.2008	0.1428	1.0000	00010
2008/9/21	-0.0390	-0.2258	-0.2677	-0.5789	0.3740	-0.3965	0.1959	-0.6742	00010
2008/9/22	0.0195	-0.4194	-0.2492	-0.6842	0.4497	-0.5609	0.2406	-0.7262	00010
2008/9/23	0.0714	-0.4839	-0.2431	-0.8947	0.5243	-0.5671	0.2423	-0.9896	00010
2008/9/24	0.1169	-0.5484	-0.2369	-0.8947	0.5867	-0.6379	0.2615	-0.8821	00010
2008/9/25	0.1364	-0.8065	-0.2123	-0.5789	0.6086	-0.8730	0.3254	-0.6083	00100
2008/9/26	0.1364	-1.0000	-0.1938	-0.6842	0.6219	-0.9230	0.3390	-0.9168	01000
2008/9/27	0.1558	-0.8065	-0.1754	-0.6842	0.6570	-0.7960	0.3735	-0.7886	00100
2008/9/28	0.1558	-1.0000	-0.1569	-0.6842	0.6642	-0.9584	0.4176	-0.7296	01000
2008/9/29	0.1558	-1.0000	-0.1569	-1.0000	0.6653	-0.9938	0.4272	-0.9411	10000

表1中, x5、x7 的值明显大于其他值, 利用这些原始值构建神经网络, 会影响神经网络的精度。归一化后所有的值均处在 $[-1, 1]$ 之间, 方便神经网络的搭建。

4. 创建神经网络

使用“newff”创建神经网络。本文构建三层神经网络, 隐含层只有一层。隐含层的节点数通常通过反复测试确定。通过反复测试, 确定隐含层的节点数为10。隐含层的激活函数选用双曲正切函数“tansig”。输出层的激活函数选用线性函数“purelin”。训练函数选用梯度下

表3 BP神经网络中的核心函数

函数名称	函数意义
premnmx	输入值归一化
newff	构建BP神经网络
postmnmx	输出值反归一化
tansig	隐含层双曲正切传递函数
purelin	输出层线性传递函数
traingdx	网络训练函数
train	网络训练执行函数
sim	计算相似度

降自适应学习率训练函数“traingdx”。

5.设置网络参数

BP 神经网络的常用配置参数如表 2 所示。通过反复测试,本文选定训练次数显示间隔为 100,网络学习速率为 0.05,最大网络训练次数为 10000,网络允许误差为 0.05。

表 4 BP 神经网络的配置参数

参数名称	参数值	参数意义
net.trainParam.show	100	网络训练次数的显示间隔
net.trainParam.lr	0.05	网络学习的速率
net.trainParam.epochs	10000	最大的网络训练次数
net.trainParam.goal	0.05	网络训练目标的允许误差

6.训练神经网络

使用“train”开始训练神经网络,Matlab 将弹出训练监控窗口,并实时反馈误差信息。当误差小于配置值时,训练过程结束,神经网络形成。

表 5 从输入层到隐含层的神经元权值

x1	x2	x3	x4	x5	x6	x7	x8
0.2342	-1.1993	0.1986	-1.1135	0.7197	-1.7925	-0.3839	0.2719
0.4441	-0.0644	0.0274	-0.0887	0.4839	0.7663	0.8429	-1.2884
-0.3546	0.2618	-0.8304	0.1967	0.8551	-0.7056	-0.5017	-0.6756
0.6455	-0.9422	-1.2858	0.9243	-0.0477	2.5366	-1.6349	0.5499
0.3996	-0.1971	-0.2732	0.3588	-0.3375	0.4781	0.6965	0.7309
-1.4408	2.0666	-1.3870	-0.7014	-1.5409	-1.1248	-0.6796	-0.1017
-0.9464	-1.1321	-1.5559	0.0893	-0.0404	-0.7975	-0.3998	-0.4209
0.5027	-0.5426	-1.1498	0.0030	0.4882	-0.5123	-0.3895	-0.2911
0.1279	-0.4324	-0.8850	-0.3162	0.7717	-0.8348	-0.4627	0.5662
-0.3502	-0.5965	-0.0084	0.3613	0.6287	-2.2507	-0.2463	-1.0556

表 6 从隐含层到输出层的神经元权值

x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
-0.3129	-0.6333	0.6529	-2.5251	0.0065	1.9116	0.1015	0.1772	0.7752	2.8601
0.2443	-0.0128	-0.8141	1.9691	-0.7051	-1.6393	-0.4506	-1.4188	0.1485	-0.5338
0.2013	-0.7243	-0.6732	0.5182	-0.4999	-1.1333	-0.0759	-0.3086	-0.4428	-1.1044
1.0084	1.6910	-1.1758	0.1154	1.2104	1.0562	0.4838	-0.0303	0.1499	-1.1240
-1.1847	0.0828	-0.5633	0.0591	-0.1531	0.0006	-0.0064	0.4448	0.2342	-0.0084

表 5、表 6 为形成三层神经网络后,神经元的权值。表 5 为输入层到隐含层的神经元权值。输入层 8 个神经元,隐含层 10 个神经元,因此得到一个 10 行 8 列的权值矩阵。表 6 为隐含层到输出层的神经元权值。隐含层 10 个神经元,输出层 5 个神经元,因此得到一个 5 行 10 列的权值矩阵。

7.测试样本输入值归一化

将测试样本同样用“premnmx”进行归一化处理。

8.神经网络仿真

将归一化的值输入到神经网络,检验神经网络的准确率。仿真函数为“sim”。

9.仿真结果反归一化

仿真后的结果是归一化后的值,无法与期望值比较,因此使用“postmnmx”函数对结果进行反归一化处理。

10.结果与期望值比较

将反归一化后的结果与期望值比较,查看神经网络的效果。比较结果如表 7 所示。

表7 测试样本输出值

时间	期望输出	实际输出	实际输出	实际输出	实际输出	实际输出	总输出
2010/8/6	01000	0.7947	0.0484	0.0797	0.0708	0.0008	10000
2010/8/7	00100	0.6749	0.0699	0.1241	0.1208	0.0008	10000
2010/8/8	00010	0.1232	0.1396	0.3117	0.3739	0.0168	00010
2010/8/9	00001	0.0194	0.1583	0.3434	0.4597	0.0009	00010
2010/8/10	00001	0.0213	0.1730	0.3421	0.3762	0.0509	00010
2010/8/11	00001	0.0531	0.1681	0.3350	0.4179	0.0021	00010
2010/8/12	00001	0.0200	0.1945	0.3369	0.3453	0.0641	00010
2010/8/13	00001	-0.0528	0.2680	0.3304	0.4280	0.0249	00010
2010/8/14	00001	-0.1224	0.1950	0.3194	0.5819	-0.0068	00010
2010/8/15	00001	-0.7889	0.9602	0.5196	-0.0202	0.9852	00001
2010/8/16	00001	-0.5340	0.5795	0.0112	0.8297	0.3570	00010
2010/8/17	00001	-2.0584	1.7758	1.1192	0.0451	-0.0328	-
2010/8/18	00001	-1.3194	1.2859	0.8965	0.0538	-0.0212	-
2010/8/19	00010	0.1324	0.2344	0.4172	0.1315	-0.0255	00100
2010/8/20	00010	-0.1814	0.7966	0.2473	0.1374	-0.0031	01000
2010/8/21	00010	0.4892	0.1908	0.1790	0.1372	0.0058	10000
2010/8/22	00100	0.6209	0.1294	0.1405	0.1080	0.0055	10000
2010/8/23	00100	0.6885	0.1047	0.1151	0.0896	0.0058	10000
2010/8/24	00100	0.7114	0.0969	0.1116	0.0752	0.0054	10000
2010/8/25	00100	0.7710	0.0958	0.0685	0.0696	0.0063	10000
2010/8/26	00100	0.6793	0.1584	0.1134	0.0326	0.0006	10000
2010/8/27	01000	0.8029	0.1352	0.0531	0.0135	0.0036	10000
2010/8/28	01000	0.9143	0.0781	0.0225	0.0020	0.0072	10000

从测试的结果来看, BP神经网络具有比较好的预警作用, 预警级别略低于期望。2010年8月8日和2010年8月15日的预警结果与期望值是完全一样的。从2010年8月9日到14日, BP神经网络的预警结果均为重警, 期望值是严重警, 测试结果比期望值低了一个等级。8月6日、8月16日、8月19日、8月27日和8月28日的预警结果比期望值低一个等级。剩下的预警结果均比期望值低两个等级。总的来看, 当出现质量危机时, 基于BP神经网络的预警模型能够预测到危机, 但低于期望结果, 预警敏感度有待加强。

五、研究结论、建议与展望

大数据时代, 基于互联网的产品质量安全事件频发, 利用产品网络质量安全信息进行预警能够有效避免大规模质量安全事件的爆发, 最大程度地减少消费者、企业的损失。本文尝试利用BP神经网络对网络消费者产品质量信息进行预警。通过对圣元奶粉质量安全信息预警的实证测试, 发现BP神经网络具有比较好的预警作用, 但预警结果低于期望值, 预警敏感度有待加强。本文仅是对一个品牌的奶粉进行BP神经网络预警, 今后的工作将集中在利用BP神经网络进行多产品的预警, 将BP神经网络应用到产品质量大数据中。

参考文献:

- [1] 陈璐、朱振华, 2013:《昆山市餐饮食品安全风险监测三级网络和预警信息平台建立及应用效果评价》,《中国食物与营养》第11期。
- [2] 程虹, 2013:《2012年中国质量状况——消费者感知与模型构建》,《宏观质量研究》第1期。
- [3] 丁菊玲、勒中坚, 2011:《基于观点树的网络舆情危机预警方法》,《计算机应用研究》第9期。

- [4] 丁菊玲、勒中坚、薛圈圈,2011:《定量网络舆情危机预警模型构建》,《图书情报工作》第20期。
- [5] 黄福员,2013:《金融风险预警的RST-FNN模型研究》,《电脑知识与技术》第9期。
- [6] 李酣,2013:《从市场失灵到政府失灵——政府质量安全规制的国外研究综述》,《宏观质量研究》第2期。
- [7] 李萍,2008:《基于MATLAB的BP神经网络预测系统的设计》,《计算机应用与软件》第4期。
- [8] 刘毅,2012:《基于三角模糊数的网络舆情预警指标体系构建》,《统计与决策》第2期。
- [9] 罗英,2013:《共享与善治:质量公共服务对质量监管效果的影响——基于我国宏观质量观测数据的实证分析》,《宏观质量研究》,第1期。
- [10] 罗艳、何锦林、许锡娟,等,2013:《遵义东南部地区农业土壤重金属污染生态风险预警研究》,《贵州科学》第6期。
- [11] 马建、王震,2013:《煤矿工作面动力灾害综合预警方式》,《科技创新与应用》第35期。
- [12] 盛静宇等,2013:《心率减速力预警高危肥厚型心肌梗死患者的临床价值》,《江苏实用心电学杂志》第6期。
- [13] 王青、成颖、巢乃鹏,2011:《网络舆情监测及预警指标体系构建研究》,《图书情报工作》第8期。
- [14] 王铁套、王国营、陈越,2012:《基于模糊综合评价法的网络舆情预警模型》,《情报杂志》第6期。
- [15] 王钰,2005:《基于改进BP神经网络的预测模型及其应用》,《计算机测量与控制》第1期。
- [16] 武汉大学质量发展战略研究院中国质量观测课题组,2013:《2012年中国质量发展报告》,《宏观质量研究》第1期。
- [17] 姚福生、钱芳、苏畅,2013:《大学生网络舆情主题演化探究及其预警指标设计》,《学术论坛》第2期。
- [18] 曾润喜,2010:《网络舆情突发事件预警指标体系构建》,《情报理论与实践》第1期。
- [19] 张立辉,2013:《基于韦伯-费希纳定律的浅水水库水环境预警评价》,《现代农业科技》第22期。
- [20] Altman, E. I., 1968, "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy", *The Journal of Finance*.
- [21] Fitzpatrick, P. J., 1932, "A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms", *The Accountants Publishing Company*.
- [22] Herrick, J. E., Brown, J. R., Trugel, A. J., et al., 2002, "Application of Soil Quality to Monitoring and Management Paradigms from Range Land Ecology", *Agronomy Journal*.
- [23] Katlan, B., Sayyed, M. A., 1999, *Regionnal Study on Use of Geograohical Information System and Early Warning in Desertification Control and Movement of Schistocerca Gergana*, MAAR Press.
- [24] Lee, J., Kramer B. M., 1992, "Analysis of Machine Degradation using a Neural Networks Based Pattern Discrimination Model", *Manufacturing Systems*.

An Empirical Study Early-warning for Network Quality Safety of Consumer Based on BP Neutral Network Information

Yu Fan, Yu Hongwei and Xu Wei

(Institute of Quality Development Strategy, Wuhan University)

Abstract: To Cope with frequently occurred issues of network quality safety, this paper aims to study information early-warning for network quality safety by using BP neutral network. Data source is all posts about Shengyuan milk powder downloaded from Tianya forum. BP neutral network includes three layers: input layer, hidden layer, and output layer. Eight indicators includes posts per unit, cumulative posts, posting frequency per unit, cumulative posting frequency, reply per unit, cumulative reply, reply frequency per unit, cumulative reply frequency are constructed, so as to five early-warning levels including safety and mild warning, middle warning, serious warning, very serious warning. The empirical study finds out that taking advantage of BP neutral network has an efficient impact on early-warning. However, the early-warning results are lower than expected. And the early-warning sensitivity is to be strengthened.

Key Words: BP Neutral Network; Network Quality Safety; Information Early-warning

■ 责任编辑 李酣