

Application of SOM Neural Network in Operational Environment Data Categorization

Lei Wang¹, Wenjun Ma²

¹PLA 91049 Troops, Qingdao Shandong

²PLA 91033 Troops, Qingdao Shandong

Email: 359219132@qq.com

Received: Feb. 7th, 2019; accepted: Feb. 19th, 2019; published: Feb. 26th, 2019

Abstract

In order to solve the problem that the names of data items in the operational environment were similar but not easy to distinguish in the process of data acquisition, with wide varieties and not easy to aggregate. Four kinds of data items often appeared in the operational environment: environmental documents, environmental data, natural environment and assembly area were taken as examples. This paper used SOM neural network to classify the duplicate names of data items. It was verified by an example. The results show that environmental documents, environmental data and natural environment belong to one category, while the assembly area belongs to another category.

Keywords

SOM Neural Network, Repetition, Operational Environment, Data Item, Classification

SOM神经网络在作战环境数据项归类中的应用研究

王磊¹, 马文君²

¹解放军91049部队, 山东 青岛

²解放军91033部队, 山东 青岛

Email: 359219132@qq.com

收稿日期: 2019年2月7日; 录用日期: 2019年2月19日; 发布日期: 2019年2月26日

摘要

针对数据采集过程中作战环境数据项名称近似而不易区分、种类繁多且不易汇总的问题, 本文以作战环境中经常出现的四类数据项—环境文件、环境数据、自然环境、集结地域为例, 运用SOM神经网络对重复的数据项名称进行归类, 并通过实例进行了验证, 结果证明环境文件、环境数据、自然环境属于一类, 而集结地域属于另一类。

关键词

SOM神经网络, 重复, 作战环境, 数据项, 归类

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

武器系统作战运用数据主要包括各级指挥员对所部(分)队的武器装备在作战中的指挥、控制与使用过程中产生、需要或关联的数据[1]。

武器系统在操作训练或实战演习过程中, 会根据作战流程产生海量的作战运用数据, 而作战运用数据的产生离不开相应的作战环境支撑, 即不同的作战环境会产生不同数值的作战运用数据, 因此在采集武器系统海量作战运用数据的过程中, 必须要注明具体的作战环境, 才能使得作战运用数据具有参考价值。但是目前部(分)队对积累的大量的各种作战环境数还停留在简单的人工记录、查询、统计和备份阶段, 导致作战环境数据项名称容易重复或语义近似等问题, 制约了采集数据的高效应用, 数据利用率不高。解决此类问题首先就是要对大量的作战环境数据项进行系统分析和聚类梳理, 以便为后续利用计算机等自动化设备进行数据处理和运用奠定基础。本文则尝试利用 SOM 神经网络[2]对作战环境数据项进行归类。

2. SOM 神经网络的本质和特点

SOM (SOM, Self Organizing Feature Map)神经网络由一系列神经元组成网络节点, 总共分为两层, 即两者之间形成全连接的输入层和竞争层, 其中, 竞争层为二维平面阵列。SOM 网络具有自组织、自学习的特点[2], 其网络结构如图 1 所示。

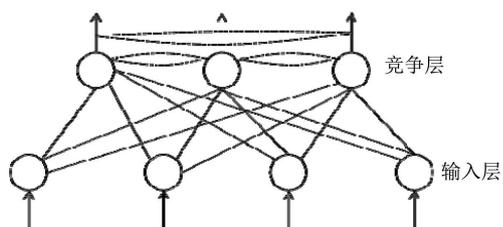


Figure 1. The general structure of SOM neural network

图 1. SOM 神经网络的一般结构组成

SOM 神经网络具有相应的功能模块, 共同完成自组织、自学习的过程[3], 即:

- 1) 处理模块。用于输入信号或某类自变量, 形成完成相应判断功能的“判断函数”。
- 2) 选择比较模块。通过比较, 选择某处理模块, 该处理模块具有最大函数输出值。
- 3) 局部互联功能。通过局部互联, 能同时激励某处理模块及其最邻近的处理模块。
- 4) 修正调整模块。用于修正或调整某处理模块的相应参数, 提高特殊条件下的“容纳性”和“适应性”。

当学习开始后, 竞争层中的神经元会逐一分辨出输入层里的神经元节点(前提是与其相邻近), 同时获胜神经元的权值向量及其邻近神经元的权值向量会按照某种函数进行更新。通过网络的自组织、自学习训练, 权值向量就会按照某种方式变得有规律可循, 即前一次的权值向量更新结果作为下一次更新前的网络输入向量。如果某一节点位置发生变化, 借助权值向量更新的作用, 此变化将影响到该神经元的邻近神经元。一旦学习过程结束, 就可以建立起一种最优的神经元节点布局, 该布局里的每个权值向量都位于聚类中心, 即完成了一次数据聚类任务[3][4]。

3. 作战环境数据项归类

作战环境类数据资源是海量的, 因此在梳理分类过程中, 可能遇到不同单位, 或者虽然

同一单位, 但由于不同指挥员在不同时间和地点或者因为记录风格不同造成数据项名称重复或近似, 导致种类繁多而不易归类, 不利于数据资源的高效利用[5]-[10]。

比如, 环境文件、环境数据、自然环境是属于同一类, 而集结地域属于其它类, 因此在实际梳理分类的基础上, 环境文件、环境数据、自然环境应该合并为一类。这里利用 SOM 神经网络算法实现聚类。

3.1. 建立模型[11][12][13][14]

本案例给出了一个含有 4 个待聚类样本的数据集。每个样本有 4 个特征, 分别为每个汉字四小部分的笔画数, 如图 2 所示, 归一化后的数据样本为:

$$T_1, p_1 - 0.2; T_1, p_2 - 0.3; T_1, p_3 - 0.2; T_1, p_4 - 0.3; T_2, p_1 - 0.14; T_2, p_2 - 0.07; T_2, p_3 - 0.36; T_2, p_4 - 0.43; \\ T_3, p_1 - 0.3; T_3, p_2 - 0.2; T_3, p_3 - 0.3; T_3, p_4 - 0.2; T_4, p_1 - 0.11; T_4, p_2 - 0.22; T_4, p_3 - 0.33; T_4, p_4 - 0.33.$$



Figure 2. Word segmentation
图 2. 文字切分

应用 SOM 神经网络聚类的步骤[15][16][17][18]:

1) 选取其中 m 个输入神经元, 并将该 m 个输入神经元到输出神经元的连接赋予较小的随机数权值。选取输出神经元 n 个“邻近神经元”集合 S_n 。其中, $S_n(t)$ 表示时刻 t 的神经元 n 的“邻近神经元”的集合。随着学习的深入, 集合会不断缩小。结合本例, 即为选取数据样本。

2) 将 $X = (x_1, x_2, x_3, \dots, x_m)^T$ 作为输入向量输入给输入层。即对数据样本进行学习, 学习结束后, 对具有最大输出的神经元标以该数据项名称的记号, 并将待检样本作为新一轮输入向量输入到 SOM 神经网络中。

3) 计算映射层的权值向量和输入向量的欧式距离。通过计算得到一个胜出神经元(胜出神经元具有最小距离), 记为 n^* , 与此同时也确定了其邻近神经元。

4) 调整胜出神经元 n^* 及其邻近神经元的权值。

$$\Delta \omega_{qn} = \omega_{qn}(t+1) - \omega_{qn}(t) = \eta(t)(x_q(t) - \omega_{qn}(t))$$

式中 $0 < \eta < 1$, 随着时间变化逐渐下降到 0。

$$\eta(t) = 0.2 \left(1 - \frac{t}{10000} \right)$$

5) 计算 o_k , $o_k = f\left(\min_n \|X - W_n\|\right)$, 式中 $f(*)$ 一般为非线性函数。

6) 如达到聚类要求则学习过程结束; 否则从步骤(2)开始, 继续网络学习。

3.2. 结果分析

本文使用的 SOM 神经网络共有 36 个节点, 即 $6 * 6$ 结构。经过算法的运行和学习过程的模拟, 我们得到了不同神经元之间的欧式距离情况, 如图 3 所示:

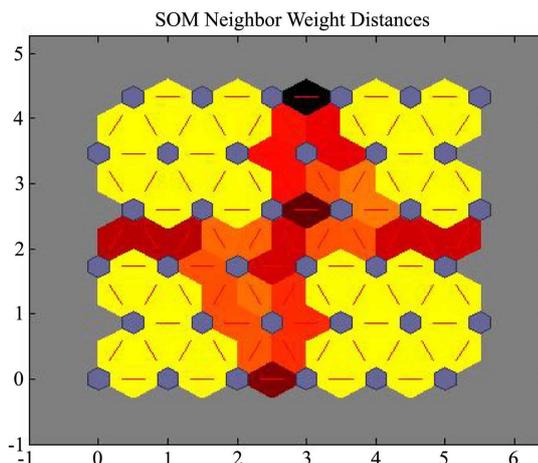


Figure 3. Euclidean distance between different neurons
图 3. 不同神经元间的欧式距离

其中 36 个蓝色节点代表 SOM 网络神经元, 神经元之间连线如果标红, 说明神经元直接连接; 由 36 个节点相互构成的具有不同颜色的菱形, 不同的颜色表示神经元之间距离的远近程度: 当颜色原来越深, 说明神经元之间的距离越来越远。比如神经元之间的距离较近时, 用黄色表示; 当变黑时, 说明神经元之间的距离最远。图 4 中标蓝的神经元为学习过后, 竞争胜出的神经元。

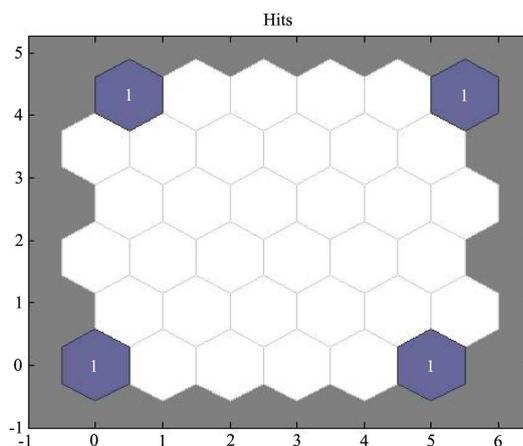


Figure 4. Competition winning neurons (blue)
图 4. 竞争胜出神经元(蓝色)

经过学习, 得到的聚类结果如表 1 所示:

Table 1. The clustering results of SOM network
表 1. SOM 网络聚类结果

训练步数	聚类结果			
	T_1	T_2	T_3	T_4
10	3	2	4	1
30	1	36	3	5
50	1	36	7	24
100	25	36	1	30
200	31	36	1	6
500	6	31	1	36
1000	31	36	1	6

当 SOM 网络的学习步数为 10 步时, 样本集 T_1, T_2, T_3, T_4 中的每一个样本自成一类; 当 SOM 网络的学习步数为 30、50、100、200、500、1000 步时, 样本集 T_1, T_2, T_3, T_4 中的每一个样本同样自成一类, 但是种类不一样了, 可以看出随着 SOM 网络的学习步数逐渐增加, 样本集中的每一个样本的归类越来越精确。

4. 结论与说明

一是对于本例, 若输出神经元在输出层的位置与某数据项名称样本的位置相同, 说明待检样本就是某数据项; 若输出神经元在输出层的位置介于很多数据项名称之间, 说明这几种数据项名称都有可能, 具体取决于欧氏距离的大小。

二是 SOM 神经网络的算法学习步数影响网络的最终归类性能, 本文通过 10 次学习过程就可以将样本集中的每一个样本完全分开, 因此就没有必要增加学习步数了。

三是经过 SOM 神经网络对作战环境数据的分析归类, 该方法比传统人工查询、记录、统计的分类速度更快。后续可以以本文的研究成果为基础, 进一步对武器系统其它作战运用数据进行归类、分析和整合, 以提高采集数据的利用率 and 应用效果, 为数据采集应用平台的建立和运行奠定坚实基础。

参考文献

- [1] 高刚, 蔡译锋, 甘艺. 面向作战应用的数据工程建设方法初探[J]. 电子对抗, 2016(1): 10-14.
- [2] 自组织映射 SOM 型神经网络的实现[EB/OL].
<https://wenku.baidu.com/view/313c42f29b89680203d825bf.html>
- [3] 李娟娟, 孟国营, 谢广明, 贾一凡. 基于小波包与 SOM 神经网络的传感器故障诊断[J]. 传感技术学报, 2017, 30(7): 1035-1039.
- [4] 胡桂莉, 喻昕, 黄秋婵. 基于 SOM 神经网络的移动客户细分研究[J]. 微型机与应用, 2015, 34(23): 51-53.
- [5] 高刚, 蔡译锋, 甘艺, 金晶. 面向作战应用的数据工程建设方法初探[J]. 电子对抗, 2016(1): 10-14.
- [6] 翟连政. XML 技术在军事信息系统数据交换和集成中的应用研究[J]. 软件导刊, 2013, 12(2): 106-107.
- [7] 李旭, 程雄, 欧中红. 军事信息系统系统集成技术研究[J]. 计算机与数字工程, 2015, 43(7): 1266-1268.
- [8] 数据字典及其使用[EB/OL]. <https://yq.aliyun.com/ziliao/445060>, 2018-09-06
- [9] 数据、数据元素、数据项、数据对象的理解[EB/OL].
https://blog.csdn.net/qq_31442743/article/details/79868242, 2018-04-09.
- [10] 数据类型[EB/OL].
<https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E7%B1%BB%E5%9E%8B/10997964?fr=aladdin>
- [11] 飞思科技产品研发中心. 神经网络理论与 MATLAB7 实现[M]. 北京: 电子工业出版社, 2005: 231-255.

- [12] 张德丰. MATLAB 神经网络应用设计[M]. 北京: 机械工业出版社, 2008: 50-77.
- [13] Kroenke, D.M. 数据库原理[M]. 清华大学出版社, 2008: 55-60.
- [14] 赵志高, 魏巍, 蔡佳明, 何飞飞. 自组织竞争神经网络 SOM 在轴承故障诊断应用[EB/OL]. <http://www.doc88.com/p-3197646681022.html>
- [15] 高峰, 刘江, 李艳, 杨新刚. 基于 Kohonen 自组织竞争网络的机床温度测点辨识研究[J]. 中国机械工程, 2014, 25(7): 862-866.
- [16] 王晓燕. K-均值算法与自组织神经网络算法的改进研究及应用[D]: [博士学位论文]. 太原: 中北大学, 2017: 14-20.
- [17] 侯淼, 刘陈帅, 王保荣. 基于灰色关联分析的 SOM 神经网络在葡萄分类中的应用[J]. 智能计算机与应用, 2017, 7(6): 42-46.
- [18] 黄紫晖, 许学军. 基于 matlab 实现人工神经网络在股票评级中的应用研究[J]. 中国林业经济, 2018(4): 60-62.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org