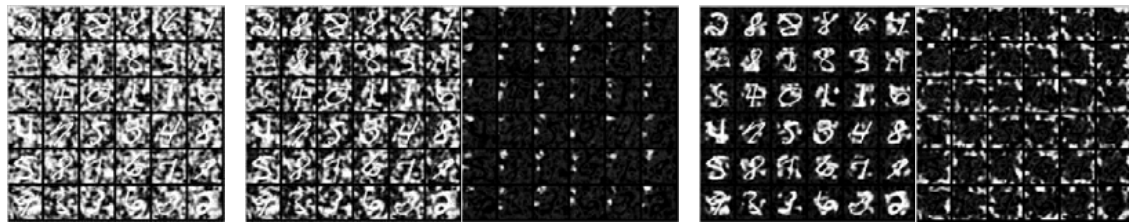


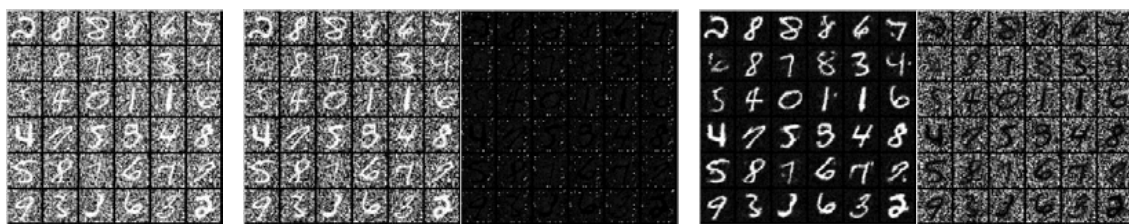
为与分类有关的和与分类无关的两个部分,并且都能在含噪声的数据中自适应找到数据中与分类有关的部分.但是,我们可以从图 5 和图 6 中看出:

- 1) pgRBM 学习得到的与分类无关的图像(也就是噪声)中包含的噪声信息很少,pgRBM 学习得到的与分类有关的图像与原始的含噪声图像相差不是很大.
- 2) pgrncRBM 是在 pgRBM 基础上对其学习到的与分类有关的图像二次去噪,但是 pgrncRBM 可以很好地把图像中与分类有关的信息(也就是手写数字)和噪声信息分别学习出来,特别是 pgncRBM 在 MNIST back-random-f 这个数据集上学习得到的与分类有关的图像基本上没有噪声.
- 3) pgincRBM 是在 pgRBM 基础上对其学习到的与分类有关的图像二次去噪,pgncRBM 可以把图像中与分类有关的信息(也就是手写数字)和图像噪声信息分别学习出来,但是有时也只能学习到部分与分类有关的信息,如第 1 列中的数字 2.



原始图像 pgRBM 学习到的与分类有关/无关的图像 pgrncRBM 学习到的与分类有关/无关的图像

(a) pgRBM 和 pgrncRBM 在 MNIST back-random-a 上学习到的与分类有关/无关图像对比

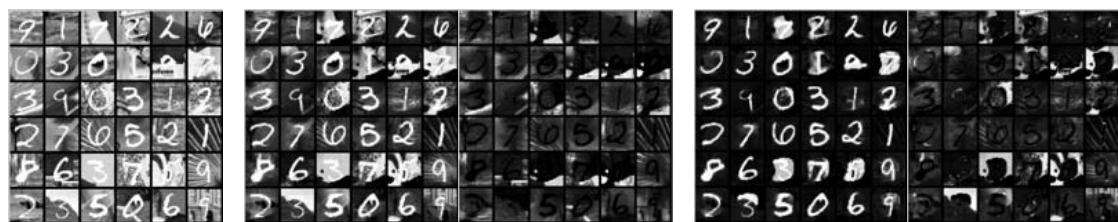


原始图像 pgRBM 学习到的与分类有关/无关的图像 pgrncRBM 学习到的与分类有关/无关的图像

(b) pgRBM 和 pgrncRBM 在 MNIST back-random-f 上学习到的与分类有关/无关图像对比

Fig.5 Learning results of pgRBM and pgrncRBM on MNIST back-random-a/f data sets

图 5 pgRBM 和 pgrncRBM 在 MNIST back-random-a/f 数据集上的学习结果



原始图像 pgRBM 学习到的与分类有关/无关的图像 pgrncRBM 学习到的与分类有关/无关的图像

Fig.6 Learning results of pgRBM and pgincRBM on MNIST back-image data set

图 6 pgRBM 和 pgincRBM 在 MNIST back-image 数据集上的学习结果

表 3 给出了含噪声的原始图像、pgRBM 学习到的与分类有关的图像和 pgincRBM/pgrncRBM 学习到的与分类有关的图像与干净数据的信噪比.我们可以得出,pgrncRBM 和 pgincRBM 可以在传统的 pgRBM 基础上对

pgRBM 学习到的数据二次去噪得到蕴含噪声信息较少的图像,从而达到更好的分类效果.pgrncRBM 在 MNIST back-random-f 和 MNIST rotated-random-f 上的错误率分别是 5.03%和 21.49%,而 pgRBM 在这两个数据集的错误率是 pgrncRBM 的 2 倍以上.并且,特别是在文献[18]中,RBM 在 MNIST basic 和 MNIST rotated 的错误率分别是 3.94%和 14.69%,这与 pgrncRBM 在 MNIST back-random-f 和 MNIST rotated-random-f 上的错误率还是比较接近的.

Table 3 Signal-noise ratios of learning results of pgRBM and pgincRBM/pgrncRBM on different valid data sets

表 3 pgRBM 和 pgincRBM/pgrncRBM 学习结果在验证数据集上的信噪比

数据集	原始噪声图像	pgRBM 学习到的 与分类有关的图像	pgincRBM/pgrncRBM 学习到的 与分类有关的图像
MNIST back-random-a	5.567 0	7.070 7	11.935 6
MNIST back-random-b	5.581 3	6.826 4	12.726 2
MNIST back-random-c	5.580 3	6.906 5	13.235 2
MNIST back-random-d	5.555 1	6.400 3	13.661 8
MNIST back-random-e	5.551 1	6.255 9	15.110 0
MNIST back-random-f	5.554 0	6.361 2	18.949 7
MNIST back-image	6.662 1	10.659 7	15.484 3
MNIST rotated -random-a	5.618 2	6.911 6	11.110 2
MNIST rotated -random-b	5.600 5	7.556 0	12.096 0
MNIST rotated -random-c	5.612 4	6.871 5	12.586 5
MNIST rotated -random-d	5.596 6	6.550 9	12.386 4
MNIST rotated -random-e	5.573 0	6.278 9	12.429 1
MNIST rotated -random-f	5.579 6	6.341 6	18.648 2
MNIST rotated-back-image	6.585 5	10.342 8	14.235 5

表 4 给出了 DBN、wDBN、pgDBN、pgwDBN、pgncDBN、pgwncDBN 和 CNN 算法含随机噪声的手写数据集上的错误率.我们从表 4 可以看出,pgwncDBN 在绝大多数数据集上的性能都表现出优于其他深度学习算法的分类效果.通过对比 pgDBN 和 pgncDBN,我们可以看出,pgrncDBN 在所有随机噪声数据集上的分类性能都优于 pgDBN.实验结果表明,pgrncDBN 还是一种有效的神经网络学习算法.我们又可以发现以下结果.

- 1) wDBN 在所有数据集上都要表现出优于 DBN 的分类性能.
- 2) pgwDBN 在所有数据集上都要表现出优于 pgDBN 的分类性能.
- 3) pgwncDBN 仅在 MNIST rotated-back-f 数据集上的错误率略高于 pgrncDBN.
- 4) CNN 仅在 MNIST back-random-a、MNIST back-random-b 和 MNIST rotated-random-a 这 3 个数据集上的错误率低于 pgrncDBN 和 pgwncDBN.

Table 4 Error rates of deep algorithms related to DBN on random noisy data sets (%)

表 4 与 DBN 相关的深度算法在随机噪声数据集上的错误率 (%)

数据集	算法						
	DBN	wDBN	pgDBN	pgwDBN	pgncDBN	pgwncDBN	CNN
MNIST back-random-a	43.21	35.95	28.01	27.73	23.56	23.55	12.32
MNIST back-random-b	40.57	30.40	24.41	23.82	20.31	19.65	15.96
MNIST back-random-c	35.34	26.37	20.41	19.32	15.94	15.86	16.89
MNIST back-random-d	34.78	21.90	16.45	15.81	12.79	12.22	14.41
MNIST back-random-e	32.98	17.67	12.54	11.76	8.60	7.91	12.52
MNIST back-random-f	25.29	12.56	8.54	7.41	4.04	3.88	8.31
MNIST rotated -random-a	72.31	71.13	67.48	67.41	57.61	56.83	46.42
MNIST rotated -random-b	71.04	67.96	62.07	62.24	51.67	51.33	59.20
MNIST rotated -random-c	67.21	63.65	57.42	56.24	45.63	45.44	57.04
MNIST rotated -random-d	67.42	60.33	48.98	48.81	39.23	38.96	55.49
MNIST rotated -random-e	66.25	53.71	42.55	40.39	31.29	31.27	44.40
MNIST rotated -random-f	63.43	43.47	33.02	31.16	15.18	15.45	32.70

表 5 给出了 DBN、wDBN、pgDBN、pgwDBN、pgincDBN、pgwincDBN 和 CNN 算法在含图像噪声的手写数据集上的错误率.除了在 MNIST back-image 上的错误率略高于 CNN 以外,pgwincDBN 在所有图像噪声数据集上都表现出优于深度浅层学习算法的分类效果.通过对比表 5 中的算法,我们还可以发现:

- 1) wDBN 在所有图像噪声数据集上的分类性能略优于 DBN.

- 2) pgwDBN 在所有图像数据集上的错误率略高于 pgDBN.
- 3) pgwincRBM 在所有图像噪声数据集上都表现出优于 pgncRBM 的分类性能.
- 4) CNN 仅仅在 MNIST back-image 数据集上的错误率略低于 pgincDBN 和 pgwincDBN.

Table 5 Error rates of deep algorithms related to DBN on image noisy data sets (%)**表 5** 与 DBN 相关的深度算法在图像噪声数据集上的错误率 (%)

数据集	算法						CNN
	DBN	wDBN	pgDBN	pgwDBN	pgincDBN	pgwincDBN	
MNIST back-image	21.65	16.97	13.17	13.49	13.44	12.98	12.50
MNIST rotated-back-image	62.44	46.42	39.03	39.52	37.37	37.09	40.60

从表 4 和表 5 中我们可以得出:1) 权值不确定性方法有效地解决了 pgrncDBN 和 pgincDBN 两种深度网络中出现的过拟合问题;2) pgrncDBN/pgwrcDBN 和 pgincDBN/pgwincDBN 可以在绝大多数数据集上的性能超过 CNN 的主要原因是 pgrncRBM/pgwrcRBM 和 pgincRBM/pgwincRBM 可以学习得到更为“干净”的数据.通过对比 RBM、DBN 和 wDBN,我们可以得出,DBN 在处理含噪声数据时出现过拟合现象,并没有展现出比 RBM 更好的学习能力,并且权值不确定性方法有效地解决了 DBN 深度网络中出现的过拟合问题.通过对比 pgRBM 与 pgDBN、pgrncRBM 与 pgrncDBN、pgincRBM 与 pgincDBN,我们又可以发现,堆叠 pgRBM/pgrncRBM/pgincRBM 和 RBM 构造出的深度网络(pgDBN、pgrncDBN 和 pgincDBN)展现出优于浅层网络(pgRBM、pgrncRBM 和 pgincRBM)的学习能力.

5 结束语

Point-wise Gated 受限玻尔兹曼机是一种针对噪声数据的浅层学习算法.本文在此基础上利用干净数据提升其在噪声数据上的学习能力,提出了两种基于噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机 (pgrncRBM 和 pgincRBM)方法,然后将它们学习到的与分类有关的数据子集用到深度置信网中,提出了两种基于噪声数据与干净数据的 Point-wise Gated 深度置信网(pgrncDBN 和 pgincDBN).pgrncRBM 和 pgincRBM 在绝大多数手写数据集上都表现出优于 pgRBM 的学习能力.同样,以 pgrncRBM 和 pgincRBM 为基石的 pgrncDBN 和 pgincDBN 一般都优于 pgDBN.然后,本文将权值不确定性方法用在所提出的 4 种算法中,并将这几种算法称为基于噪声数据与干净数据的 Point-wise Gated 权值不确定性受限玻尔兹曼机(pgwrncRBM 和 pgwincRBM)以及基于噪声数据与干净数据的 Point-wise Gated 权值不确定性深度置信网(pgwrncDBN 和 pgwincDBN).实验结果表明,权值不确定性方法能够有效地解决 pgrncRBM、pgincRBM、pgrncDBN 和 pgincDBN 在处理含噪声数据时出现的过拟合问题.我们同时发现,pgRBM、pgrncRBM 和 pgincRBM 并不能在部分噪声数据集(数据中与分类有关的部分被噪声损坏时)上取得理想的结果.如何把它们应用到更多的噪声数据上,也是我们下一步的研究方向.

References:

- [1] Ravanbakhsh S, Póczos B, Schneider J. Stochastic neural networks with monotonic activation functions, In: Gretton A, Robert CC, eds. Proc. of the 19th Int'l Conf. on Artificial Intelligence and Statistics. Cadiz: MIT Press, 2016. 809–818.
- [2] Salakhutdinov R, Hinton GE. An efficient learning procedure for deep Boltzmann machines, Neural Computation, 2012,24(8): 1967–2006. [doi: 10.1162/NECO_a_00311]
- [3] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006,313(5786):504–507. [doi: 10.1126/science.1127647]
- [4] Zhao S, Muraoka Y, Fujimaki R, Carin L. Scalable model selection for belief networks. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. Proc. of the Advances in Neural Information Processing Systems 30. Long Beach: MIT Press, 2017. 4612–4622.
- [5] Zhang N, Ding SF, Zhang J, Xue Y. An overview on RESTRICTED Boltzmann machines. Neurocomputing, 2018,275:1186–1199. [doi: 10.1016/j.neucom.2017.09.065]

- [6] Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 2003,3:1157–1182.
- [7] Sohn BK, Zhou G, Lee C, Lee H. Learning and selecting features jointly with point-wise gated Boltzmann machines. In: *Proc. of the 30th Int'l Conf. on Machine Learning*. Atlanta: MIT Press, 2013. 217–225. <https://icml.cc/2013/>
- [8] Zhang N, Ding SF, Zhang J, Xue Y. Research on point-wise gated deep networks. *Applied Soft Computing*, 2017,52:1210–1221. [doi: 10.1016/j.asoc.2016.08.056]
- [9] Ranzato M, Hinton GE. Modeling pixel means and covariances using factorized third-order Boltzmann machines. In: *Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition*. San Francisco: IEEE Press, 2010. 2551–2558. <http://www.cvpapers.com/cvpr2010.html> [doi: 10.1109/CVPR.2010.5539962]
- [10] Courville A, Desjardins G, Bergstra J, Bengio Y. The spike-and-slab RBM and extensions to discrete and sparse data distributions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2014,36(9):1874–1887. [doi: 10.1109/TPAMI.2013.238]
- [11] Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov RR. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014,15:1929–1958. [doi: 10.1214/12-AOS1000]
- [12] Wan L, Zeiler M, Zhang S, Le Cun Y, Fergus R. Regularization of neural networks using dropconnect. In: *Proc. of the 30th Int'l Conf. on Machine Learning*. Atlanta: MIT Press, 2013. 1058–1066. <https://icml.cc/2013/>
- [13] Iosifidis A, Tefas A, Pitas I. DropELM: Fast neural network regularization with dropout and DropConnect. *Neurocomputing*, 2015, 162:57–66. [doi: 10.1016/j.neucom.2015.04.006]
- [14] Blundell C, Cornebise J, Kavukcuoglu K. Weight uncertainty in neural networks. In: Bach FR, Blei DM, eds. *Proc. of the 32nd Int'l Conf. on Machine Learning*. Lille: MIT Press, 2015. 1613–1622.
- [15] Zhang J, Ding SF, Zhang N, Xue Y. Weight uncertainty in Boltzmann machine. *Cognitive Computation*, 2016,8(6):1064–1073. [doi: 10.1007/s12559-016-9429-1]
- [16] Ding SF, Zhang N, Shi ZZ. Laplacian multi layer extreme learning machine. *Ruan Jian Xue Bao/Journal of Software*, 2017,28(10): 2599–2610 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5128.htm> [doi: 10.13328/j.cnki.jos.005128]
- [17] Hinton GE. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 2002,14(8):1711–1800. [doi: 10.1162/089976602760128018]
- [18] Larochelle H, Erhan D, Courville A, Bergstra J, Bengio Y. An empirical evaluation of deep architectures on problems with many factors of variation. In: Ghahramani Z, ed. *Proc. of the 24th Int'l Conf. on Machine Learning*. New York: MIT Press, 2007. 473–480. [doi: 10.1145/1273496.1273556]

附中文参考文献:

- [16] 丁世飞,张楠,史忠植.拉普拉斯多层极速学习机. *软件学报*,2017,28(10):2599–2610. <http://www.jos.org.cn/1000-9825/5128.htm> [doi: 10.13328/j.cnki.jos.005128]



张楠(1991—),男,安徽萧县人,博士,主要研究领域为机器学习,玻尔兹曼机.



张健(1990—),男,博士生,主要研究领域为深度学习,玻尔兹曼机.



丁世飞(1963—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为人工智能,模式识别,机器学习,数据挖掘.



赵星宇(1994—),男,硕士,主要研究领域为机器学习.