

文章编号: 1006-2475(2020)10-0012-05

基于注意力机制和 LSTM 的电力通信设备状态预测

吴海洋 陈 鹏 郭 波 蒋春霞 李霁轩 朱鹏宇

(国网江苏省电力有限公司信息通信分公司, 江苏 南京 210024)

摘要: 随着电力通信网络的快速增长, 网络中通信设备的在线状态预测对于提升运维可靠性具有重要意义。在实际场景中, 设备工作数据来源复杂, 往往存在数据维度高、特征稀疏且模式重复等问题, 导致传统的预测方法性能非常受限。本文提出一种基于注意力机制和 LSTM(长短期记忆) 模块的设备状态预测模型。模型训练分 2 阶段进行, 保证注意力机制能够通过端到端学习对原始特征进行充分降维并提取出最相关的信息进行状态预测。基于电力通信网络真实运维数据进行一系列验证实验, 结果表明所提方法在设备状态预测问题中的有效性。

关键词: 注意力; 长短期记忆; 神经网络; 设备状态预测

中图分类号: TP277; TP18

文献标志码: A

DOI: 10.3969/j.issn.1006-2475.2020.10.003

Attention and LSTM Based State Prediction of Equipment on Electric Power Communication Networks

WU Hai-yang, CHEN Peng, GUO Bo, JIANG Chun-xia, LI Ji-xuan, ZHU Peng-yu

(Information and Communication Branch, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China)

Abstract: With the rapid growth of electric power communication networks, the importance of predicting the working state of on-line equipment is increasing as well. Since the running data of typical communication devices always come from heterogeneous resources, the prediction models have to be learned from features with high dimension, high sparsity as well as repetitive patterns. This problem severely restricts the performance of conventional machine learning approaches. This paper proposes a novel state prediction model based on the integration of attention mechanism and LSTM (Long Short-Term Memory). By a two-stage learning strategy, the attention mechanism can achieve both dimensionality reduction and feature extraction of original input. Meanwhile, most related features are extracted for final prediction from the end-to-end learning. Extensive experimental results on practical running data of electric power communication networks demonstrate that, the proposed method has high performance in the working state prediction problem.

Key words: attention; LSTM; neural networks; state prediction of equipment

0 引 言

近年来, 我国智能电网的覆盖规模迅速扩展, 作为相应配套承载的电力通信网络也迅速发展, 设备数量逐年增多。现代电力通信网络不仅承担了电网管理功能, 还需要对关键设备的运行状态进行记录和维护。因此, 电力通信网络设备的健康运行本身的重要性也与日俱增。传统的电力通信网络设备维护仍然基于周期性巡检或者故障告警等手段^[1], 响应慢且故障代价高。而随时监控设备运行状态, 通过运维记录对可能出现的设备故障进行预警正逐渐成为电力

通信网运维的重要手段^[2-3]。

虽然在工业领域对于设备运行状态的在线预测问题已经有很多相关研究^[4], 然而对于电力通信网络设备的状态预测仍然是一个挑战性问题。这主要是因为此类设备的相关数据来源复杂, 不仅包括现场各类传感器数据, 还包含了各类网络运维记录, 且考虑到相关节点的数据对于目标节点上设备状态的预测同样有辅助作用, 最终所需要考虑的输入数据维度往往很高; 而相对地, 数据数量则不是很多, 且大部分数据模式单一。

传统的基于现场有限数据进行状态预测的一些

收稿日期: 2020-03-07; 修回日期: 2020-03-22

基金项目: 国网江苏省电力有限公司科技项目 (SGJSXT00DDJS1900167)

作者简介: 吴海洋 (1983—), 男, 江苏南京人, 高级工程师, 博士, 研究方向: 电力通信、机器学习, E-mail: wuhaiyang128@163.com; 陈鹏 (1989—), 男, 江苏盐城人, 工程师, 硕士, 研究方向: MU-MIMO, E-mail: topecp@sina.com。

手段^[5],往往依赖于手工的特征设计,需要有很具体的技术经验背景,在不同的场景下用途有限。而另一个思路是在原始数据的基础上首先使用一些数据降维手段进行预处理,然后通过训练支持向量机^[6-7]或者决策树^[8-9]等回归预测模型实现状态预警^[10]。此类方法的局限性在于,一般的降维手段只能作为预处理独立使用,不能和预测模型进行端到端的联合学习,对于有效信息的挖掘效率受到限制,最终导致性能有限。

针对这些问题,本文提出基于注意力机制和长短时记忆(Long Short-Term Memory, LSTM)模型构建端到端的设备状态预测框架,对网络节点上的设备运行状态进行在线预测。作为深度学习领域^[11]常见的特征提取和序列预测手段,注意力机制和 LSTM 模型在自然语言处理^[12]、视频内容检测^[13]等场景中都有非常成功的应用。本文主要解决将此类需要海量数据驱动的复杂模型应用于有限数据场景时的一些具体问题。通过数据生成以及分阶段训练等手段成功地实现端到端的模型训练,提升最终的预测性能。

1 背景资料

1.1 注意力机制

注意力机制^[14]可以充分发掘信号中的自相关性,突出信号中与最终预测任务相关的部分特征,并降低信号混淆程度。由于从现场多种数据来源采集的信号普遍存在信号源之间的相关性以及信号源内部通道间的相关性,因此在传统注意力机制的基础上,本文进一步引入通道注意力模块^[15]和传统的注意力模块构成的混合注意力模块,如图 1 所示。

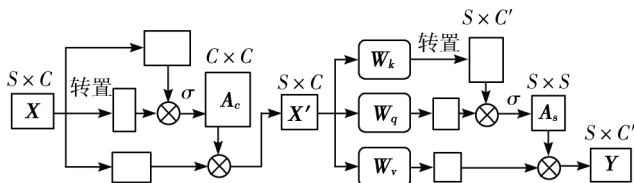


图 1 混合注意力模块

该模块由通道注意力模块和传统注意力模块串联而成,兼具了相关性提升和信号降维的作用。给定输入 $X \in \mathbf{R}^{S \times C}$,其中 S 为信号长度, C 为每组信号的通道数, σ 为 softmax 函数产生的非线性关系。首先在前半部分通过一个通道注意力模块,将特征与其转置矩阵相乘得到不同通道的自相关阵,该矩阵通过一个 softmax 函数后即可作为通道注意力矩阵 $A_c \in \mathbf{R}^{C \times C}$ 对原始信号进行加权。通过该模块后,原始特征通道间的相关性下降,而有用的信息被突出,进一步输入紧随的降维注意力模块。

对于降维注意力模块而言,其输入 $X' \in \mathbf{R}^{S \times C'}$ 被 3 个映射矩阵 W_k 、 W_q 和 W_v 投影到一个低维空间 $\mathbf{R}^{S \times C'}$ 上,通过类似的方法产生自注意力矩阵 $A_s \in \mathbf{R}^{S \times S}$,对输入信号进行加权。降维注意力机制可以简单表达如下:

$$Y = A_s X' W_v = \sigma(X' W_q (X' W_k)^T) X' W_v \quad (1)$$

最终得到降维后的输出 $Y \in \mathbf{R}^{S \times C'}$ 。

在该模块中,仅有 3 个需要学习的 $C \times C'$ 维参数矩阵: W_k 、 W_q 和 W_v 。在第 3 章的实验环节中可以看到,仅采用少量的参数就可以将输入特征的维度大幅降低,实现状态预测。

1.2 LSTM 模型

长短时记忆 LSTM 模型^[16-17]是一种常见的递归神经网络实现,多用于顺序信号的状态预测^[13,18-20]。一个标准的 LSTM 单元由 4 个门函数和一个记忆细胞构成,分别是输入门 i 、输入调制门 g 、遗忘门 s 、输出门 o 和记忆单元 c 。给定 t 时刻的输入 $x_t \in \mathbf{R}^{SC'}$,输出为 $z_t \in \mathbf{R}^N$,传递的隐藏状态为 $h_t \in \mathbf{R}^N$,其中 N 为隐藏单元数;实际运行时,注意力单元的输出 Y 被拉直以后作为 LSTM 的输入使用。如果将内部各门函数和记忆单元上的非线性激活函数统一用 $f^\lambda(\cdot)$, $\lambda \in \{i, g, s, o, c\}$ 表示,而输入输出分别用 x^λ 和 y^λ 表示,则 LSTM 的前向计算可简单表示为:

$$y^\lambda = f^\lambda(W^\lambda x^\lambda), \lambda \in \{i, g, s, o\} \quad (2)$$

$$y^c = f^c(y^s \odot c_{t-1} + y^i \odot y^g) \quad (3)$$

$$z_t = h_t = y^o \odot y^c \quad (4)$$

其中 \odot 表示向量点积。对于 LSTM 而言,需要学习的参数主要为 4 个门函数上的线性加权参数 W^λ ,实际使用中可以通过标准的时序反向传播方法^[21]对模块进行训练。

2 方法设计与模型训练

通过将上述混合注意力机制和 LSTM 进行串联,本文提出如图 2 所示的电力通信设备状态预测方法。其中从现场各类传感器以及运行记录中所得到的原始数据经过填充到固定长度以后进入预处理模块,通过常见的归一化以及中心化等处理形成系统输入。通过图 1 所示的混合注意力模块后输入 LSTM 单元实现时序预测。最终通过最后一个全连接(Fully Connected, FC)层输出预测结果。

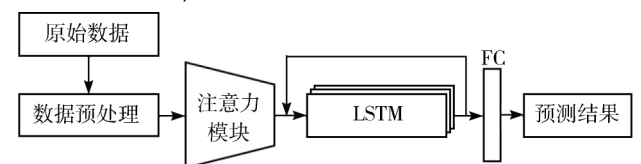


图 2 预测系统架构

对于如图 2 所示的预测模型,如果直接使用常见的反向传播方法进行训练,所面临的问题主要有 2 个:首先,可供训练的样本维数很高,然而数量并不充分;其次,大部分训练样本模式单一,实际运行中通信设备异常状态的出现并不常见。使用这样的数据直接驱动模型学习往往会导致过拟合,预测效果并不好。因此在实际进行模型训练时,本文采用 2 阶段的方式对模型进行逐步学习。

在第 1 阶段,首先通过可变分自编码器(Variational Auto-Encoder, VAE)^[22]方式对混合注意力模块进行预训练,同时得到一个生成模块用于产生同分布的采样样本补充进入训练样本库。VAE 基于传统的自编码器提出,其区别在于在降维后的中间隐层对表示均值和方差的低维向量添加随机噪声,并保证生成模块在随机噪声的影响下仍然能够复原输入样本。本文使用一个如图 1 所示的注意力模块串联一个三层神经网络构成一个 VAE,保证 C 维特征在被降维至 C' 维后再恢复为 C 维生成特征。再与训练时使用生成特征和原始输入的最小均方误差作为损失函数训练 VAE。当训练完成后,一方面可以得到一个预训练完成的注意力模型,另一方面通过对原始样本增加小随机噪声在 VAE 中生成新样本补充进样本库进行下一阶段学习。在第 2 阶段,对如图 2 所示的完整预测模型进行端到端训练。在训练时模型前段的注意力模块使用第 1 阶段结果进行参数初始化,与后段 LSTM 模块以及 FC 层进行联合训练。同时,将第 1 阶段的生成样本补充进入样本库以后,通过优化传统的交叉熵损失^[23-24]完成模型训练。

3 实验与分析

3.1 数据来源及算法实现

实验部分的数据主要由每个电力通信网络节点上的设备履历信息、检修记录、缺陷记录以及传感器传回的实时监测信息构成。同时作为通信网络节点,其运行特征诸如光功率、抖动、漂移、误码率、误码秒以及信噪比等也是判断其状态的重要参数。

本文实验中从每个标准节点提取一组特征向量,将网络拓扑结构上与该节点最接近的 3 个节点对应的特征向量取出同样用于当前节点的状态预测。所有数据经过抽取、拉直并对空位进行填充后每个节点上得到一组 2048 维的特征向量,将 4 组向量输入图 2 所示预测模型进行状态预测。

本文采用 Pytorch 深度学习框架实现上述注意力以及 LSTM 模块。实验所用数据采集于 2 年内的电力通信网络运营数据,其中 2/3 作为模型的训练数

据,其余 1/3 作为测试数据使用。每组实验中同样配置的方法进行 5 次实验,以检查随机初始化带来的模型性能的细微变化。

3.2 注意力模块参数比较

由于注意力模块的主要作用是将每组 2048 维的高维特征映射到一个低维子空间,并突出和预测相关的有效特征,在预训练阶段首先要对降维后子空间的维数做一些测试。过低的子空间维数不足以保持足够的预测特征,而过高的子空间维数则会直接影响后续 LSTM 模块的复杂度。因此实验部分首先基于 VAE 训练方法比较了采用不同子空间维数时在测试集上生成的样本和原始样本的均方误差。实验采用 32 个样本为批次处理大小,在训练集上训练 20 个周期,最终比较结果如表 1 所示。

表 1 注意力模型子空间维数结果比较

实验号	子空间维数	平均均方误差
1	4	207.35 \pm 24.25
2	8	46.78 \pm 4.38
3	16	10.98 \pm 1.23
4	32	8.36 \pm 0.87
5	64	7.57 \pm 0.79

从表 1 的比较中可以发现,子空间维数在 8 维以上时才有可能产生比较有意义的信号重建,也意味着比较有效的特征信息被完整地保留下来。而子空间维数增长到 32 维以上时,平均均方误差的下降则不明显,而带来的复杂度提升则比较严重。因此在端到端训练阶段,主要考虑 8、16 和 32 维这 3 种维度和 LSTM 模块配合使用时的最终性能。

3.3 联合参数选择

在端到端训练阶段,上面筛选出的注意力模型参数和 LSTM 模块参数进行了联合筛选以找出最优的模型超参数。其中 LSTM 模块主要需要筛选的参数有时间步和隐层单元数量 2 个设置。表 2 列出了不同参数对应的最终预测精度,即正确预测数量占总测试样本的比例,以及每种配置需要的训练时间。作为模型效率的参考,此处训练时间仅为第 2 阶段端到端训练时间。其中每次实验同样采用 32 个样本为批次处理大小,在训练集上训练 20 个周期。

从表 2 的比较中可以看出,子空间维度达到 16 维以上, LSTM 隐层单元数达到 8 以上,基本上可以得到性能较好的预测模型。相应地, LSTM 时间步的增加也对最终的预测性能有正向增益。其中子空间维数为 16,时间步和隐层单元数分别为 10 和 16 时,可以得到一个预测性能比较理想的模型。继续增加相关模块的复杂度,所带来的性能提升已不明显,且会带来细微的训练时间增加。

表 2 模型超参数对最终性能和效率影响的结果比较

实验	超参数			性能	
	子空间维度	时间步	隐层单元数	精度	训练时间/h
1	8	5	4	0.633 ± 0.012	2.1
2	8	5	8	0.682 ± 0.008	2.1
3	8	10	4	0.693 ± 0.014	2.2
4	8	10	8	0.726 ± 0.010	2.1
5	16	5	4	0.784 ± 0.009	2.1
6	16	5	8	0.847 ± 0.013	2.2
7	16	5	16	0.849 ± 0.014	2.2
8	16	10	4	0.832 ± 0.010	2.2
9	16	10	8	0.869 ± 0.009	2.3
10	16	10	16	0.874 ± 0.009	2.3
11	32	5	8	0.849 ± 0.011	2.3
12	32	5	16	0.851 ± 0.012	2.3
13	32	10	8	0.873 ± 0.015	2.4
14	32	10	16	0.876 ± 0.011	2.4
15	32	15	8	0.877 ± 0.009	2.4
16	32	15	16	0.877 ± 0.010	2.4

3.4 与相关方法的比较以及消融测试

为进一步验证所提方法的有效性,基于筛选出的最优参数,本文继续将所提模型和一些相关方法进行比较,并就不同的模块组合进行消融测试。参与比较的方法有:PCA + Logistic 回归(P + L)、PCA + LSTM(P + LSTM)、注意力 + Logistic 回归(A + L)、卷积网络(Convolutional Neural Networks, CNN) + LSTM(C + LSTM)以及分别进行1阶段训练和2阶段训练的注意力 + LSTM(A + LSTM1 和 A + LSTM2)。本文基于一维卷积层、池化层以及1个全连接层构建了1个9层CNN,并使用常见的批量标准化和非线性激活层。图3和表3分别比较了不同方法的查准率(Precision)、召回值(Recall)和训练时间。

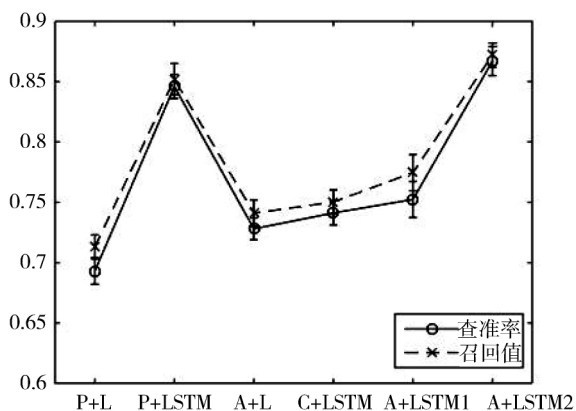


图 3 不同方法的查准率和召回值比较

从图3中可以看出,LSTM模块对于时序信号预测的作用比较明显,直接使用Logistic回归难以进行有效预测,而LSTM也需通过降维手段配合才能有效工作。在该问题中,将降维模块与LSTM串联直接进行1阶段端到端学习得到的模型性能有限,而基于

VAE方法的2阶段训练则显著提高了最终性能。另一方面,直接使用CNN等流行的深度学习模型在此处的效果并不是十分理想。其原因可能是因为通常的卷积核尺寸有限,对于局部特征可以有效捕捉,对于比较稀疏的输入信号而言则无法直接对整个特征进行处理,而增大卷积核尺寸进行结构改进又将使模型复杂度变得过高。

表 3 不同方法的训练时长比较

方法	P + L	P + LSTM	C + LSTM	A + LSTM
训练时长/h	1.6	2.3	2.2	4.5

表3中比较了不同方法训练模型所需要的时长。对于本文所提的注意力 + LSTM模型,由于需要采用2阶段训练方式进行训练,导致训练的总时长较长,然而测试阶段的模型前向效率是基本近似的。

4 结束语

本文提出了一种基于注意力机制和LSTM的预测模型,着重解决了电力通信设备状态的预测问题。模型使用端到端的方式进行训练,将原始特征数据的降维、信息挖掘与最终预测整合成一步完成,对于类似场景中数据维度高、特征稀疏以及训练数据量不足的问题有较好的效果。后续研究中还将进一步优化模型复杂度,以提升方法在线预测的时间效率。

参考文献:

- [1] ANDERSSON L, BRAND K P, BRUNNER C, et al. Reliability investigations for SA communication architectures based on IEC 61850 [C]// Proceedings of the 2005 IEEE Russia Power Tech Conference. 2005, DOI: 10.1109/PTC.2005.4524707.
- [2] ZHAO Z D, LOU Y Y, NI J H, et al. RBF-SVM and its application on reliability evaluation of electric power system communication network [C]// Proceedings of the 2009 International Conference on Machine Learning and Cybernetics. 2009, 2: 1188-1193.
- [3] GAO H S, GUO J. Application of vulnerability analysis in electric power communication network [C]// Proceedings of the 2009 International Conference on Machine Learning and Cybernetics. 2009, 4: 2072-2077.
- [4] 李兆桐, 张卫山, 郭武武. 基于LSTM的工业互联网设备工作状态预测[J]. 计算机与现代化, 2020(1): 1-5.
- [5] LEEMA A A, HEMALATHA M. Proposed prediction algorithms based on hybrid approach to deal with anomalies of RFID data in healthcare [J]. Egyptian Informatics Journal, 2013, 14(2): 135-145.
- [6] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.
- [7] 李元诚, 方廷健, 于尔铿. 短期负荷预测的支持向量机方法研究[J]. 中国电机工程学报, 2003, 23(6): 55-59.
- [8] SAFAVIAN S R, LANDGREBE D. A survey of decision tree classifier methodology [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1991, 21(3): 660-674.

- [9] 苗夺谦,王珏. 基于粗糙集的多变量决策树构造方法[J]. 软件学报,1997 8(6):425-431.
- [10] 吴海洋,缪巍巍,郭波,等. 基于遗传算法的 BP 神经网络蓄电池寿命预测研究[J]. 计算机与数字工程,2019 47(5):1275-1278.
- [11] KRIZHEVSKY A, HINTON G. Learning Multiple Layers of Features from Tiny Images[R]. University of Toronto, 2009.
- [12] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019: 4171-4186.
- [13] SUN Y Q, WU Z X, WANG X, et al. Exploiting objects with LSTMs for video categorization[C]// Proceedings of the 24th ACM International Conference on Multimedia. 2016: 142-146.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.
- [15] CHEN T L, DING S J, XIE J Y, et al. ABD-Net: Attentive but diverse person re-identification[J]. arXiv preprint arXiv:1908.01114, 2019.
- [16] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997 9(8):1735-1780.
- [17] LYU Q, ZHU J. Revisit long short-term memory: An optimization perspective[C]// Proceedings of the 2014 Workshop on Deep Learning and Representation Learning. 2014.
- [18] LI Z Y, GAVVES E, JAIN M, et al. VideoLSTM convolves, attends and flows for action recognition[J]. arXiv preprint arXiv:1607.01794, 2016.
- [19] WUH Y, LU X, MIAO W W, et al. Dynamic routing programming for power communication networks by recurrent neural networks based reliability prediction and particle swarm optimization[C]// Proceedings of the 5th IEEE International Conference on Advances in Electrical and Electronics, Information, Communication and Bio-Informatics (AEEICB2019). 2019: 1217-1221.
- [20] 章鑫锋,张彩友,冯毅萍,等. 基于多粒度 LSTM 模型的换流站设备分析研究[C]// 第 30 届中国过程控制会议 (CPCC 2019) 论文集. 2019.
- [21] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2625-2634.
- [22] KINGMA D P, WELING M. Auto-encoding variational Bayes[J]. arXiv preprint arXiv:1312.6114, 2013.
- [23] ZHANG Z L, SABUNCU M R. Generalized cross entropy loss for training deep neural networks with noisy labels[C]// Proceedings of the 32nd Conference on Neural Information Processing Systems. 2018: 8778-8788.
- [24] FARAHNAK-GHAZANI F, BAGHSHAH M S. Multi-label classification with feature-aware implicit encoding and generalized cross-entropy loss[C]// Proceedings of the 24th Iranian Conference on Electrical Engineering (ICEE). 2016: 1574-1579.

(上接第 11 页)

- [10] WANG H, WANG N Y, YEUNG D-Y. Collaborative deep learning for recommender systems[C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015: 1235-1244.
- [11] KIM D H, PARK C, OH J, et al. Convolutional matrix factorization for document context-aware recommendation[C]// Proceedings of the 10th ACM Conference on Recommender Systems. 2016: 233-240.
- [12] WU C-Y, AHMED A, BEUTEL A, et al. Joint training of ratings and reviews with recurrent recommender networks[C]// Proceedings of the 5th International Conference on Learning Representations. 2017.
- [13] 曾安,聂文俊. 基于深度双向 LSTM 的股票推荐系统[J]. 计算机科学,2019 46(10):84-89.
- [14] 李南星,盛益强,倪宏. 用于个性化推荐的条件卷积隐因子模型[J/OL]. 计算机工程,(2019-05-28) [2020-02-16]. <https://doi.org/10.19678/j.issn.1000-3428.0054209>.
- [15] KIM Y. Convolutional neural networks for sentence classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014: 1746-1751.
- [16] WU Y, DUBOIS C, ZHENG A X, et al. Collaborative denoising auto-encoders for top-n recommender systems[C]// Proceedings of the 9th ACM International Conference on Web Search and Data Mining. 2016: 153-162.
- [17] STRUB F, MARY J. Collaborative filtering with stacked denoising autoencoders and sparse inputs[C]// NIPS Workshop on Machine Learning for eCommerce. 2015.
- [18] SEDHAIN S, MENON A K, SANNER S, et al. AutoRec: Autoencoders meet collaborative filtering[C]// Proceedings of the 24th International Conference on World Wide Web. 2015: 111-112.
- [19] SHIBATA H, TAKAMA Y. Behavior analysis of RBM for estimating latent factor vectors from rating matrix[C]// Proceedings of the 2017 6th International Conference on Informatics, Electronics and Vision, 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMT). 2017.
- [20] SARWAR B, KARYPIS G, KONSTAN J, et al. Incremental singular value decomposition algorithms for highly scalable recommender systems[C]// Proceedings of the 2002 International Conference on Computer and Information Science. 2002: 27-28.
- [21] ZHAO L L, LU Z Q, PAN S J L, et al. Matrix factorization + for movie recommendation[C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence. 2016: 3945-3951.
- [22] HE X N, CHUA T-S. Neural factorization machines for sparse predictive analytics[C]// Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2017: 355-364.
- [23] SALAKHUTDINOV R, MNIH A, HINTON G. Restricted Boltzmann machines for collaborative filtering[C]// Proceedings of the 24th International Conference on Machine Learning. 2007: 791-798.