



大连理工大学

信息检索研究室

*Information Retrieval Laboratory of DUT*

# A Prediction Model Based on Deep Belief Network and Least Squares SVR Applied to Cross-Section Water Quality

2020年12月11日

## ● 开源代码：无

## ● 数据集：

- ✓ 数据来源于河北廊坊数据监测站，三河东桥的水源质量数据
- ✓ 数据每四小时被收集一次
- ✓ 数据包含以下9个特征：水温、pH值、溶解氧、电导率、浊度、高锰酸钾指数、总磷、氨氮
- ✓ 使用前八个特征预测最后一个特征

- 水污染问题严重（人为因素+非人为因素）
- 使用水源质量参数预测、控制水污染
- 利用**粒子群**优化后的**DBN**神经网络，结合最小二乘支持向量回归(**LSSVR**)进行水质预测
- 对比BP、DBN、LSSVR、DBN-LSSVR模型取得了更好的效果

- 由一系列的RBM单元组成，每层神经元之间不存在连接
- 可见层与隐层能量计算方式：

$$E(v, h|\theta) = -\sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i w_{ij} h_j$$

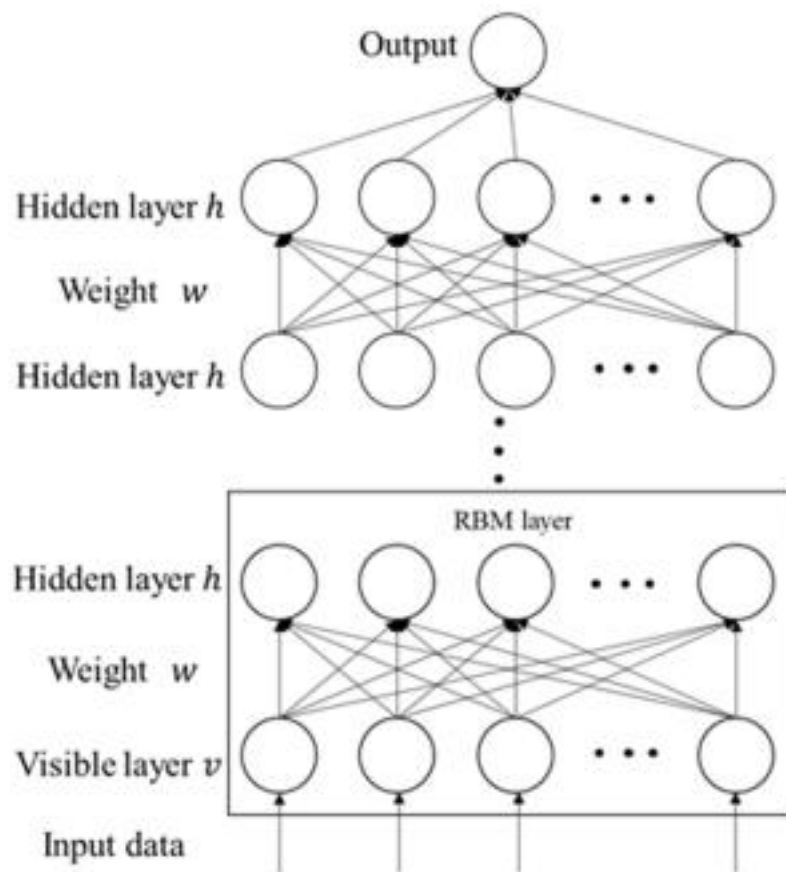


Figure 3. Deep belief network (DBN) model architecture.

- 可见层与隐层联合概率分布：

$$P(v, h|\theta) = \frac{e^{-E(v,h|\theta)}}{Z^\theta}$$

$$Z^\theta = \sum_{v,h} e^{-E(v,h|\theta)}$$

- 使用极大似然训练模型参数：

$$L(\theta) = \sum_{n=1}^N \ln(v^n, h)$$

$$\theta = \operatorname{argmax} L(\theta) = \operatorname{argmax} \sum_{n=1}^N \ln(v^n, h)$$

- **PSO算法：简单、易于实现、参数少、速度快**
- **实现步骤：**
  - ◆ （1）初始化粒子种群（随机位置和速度）
  - ◆ （2）评估，寻找Pbest和Gbest
  - ◆ （3）以一定的速度进行更新

- 基于统计理论和结构风险最小化准则的方法
- 求解二次规划=》求解线性方程组
- 提高计算速度和收敛精度
- 训练方法：

$$\min J(\omega, \xi) = \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^l \xi_i^T \xi_i$$

$$s.t. y_i = \omega^T \varphi(x_i) + \mu + \xi_i \quad (i = 1, 2, \dots, l)$$

# 基于PSO优化DBN网络和LSSVR的预测模型



大连理工大学  
信息检索研究室



Information Retrieval Laboratory of DUT

## ● 实现步骤：

- ◆ (1) 训练DBN模型的每个RBM层
- ◆ (2) 使用PSO防止DBN陷入局部最优
- ◆ (3) 决定LSSVR参数
- ◆ (4) 使用顶层LSSVR模型迭代更新每一层的RBM参数，finetune整体的DBN网络



## ● MAE、MPAE、RMSE、R<sup>2</sup>

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$R^2 = \frac{\left( \sum_{i=1}^n (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i) \right)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)^2}$$

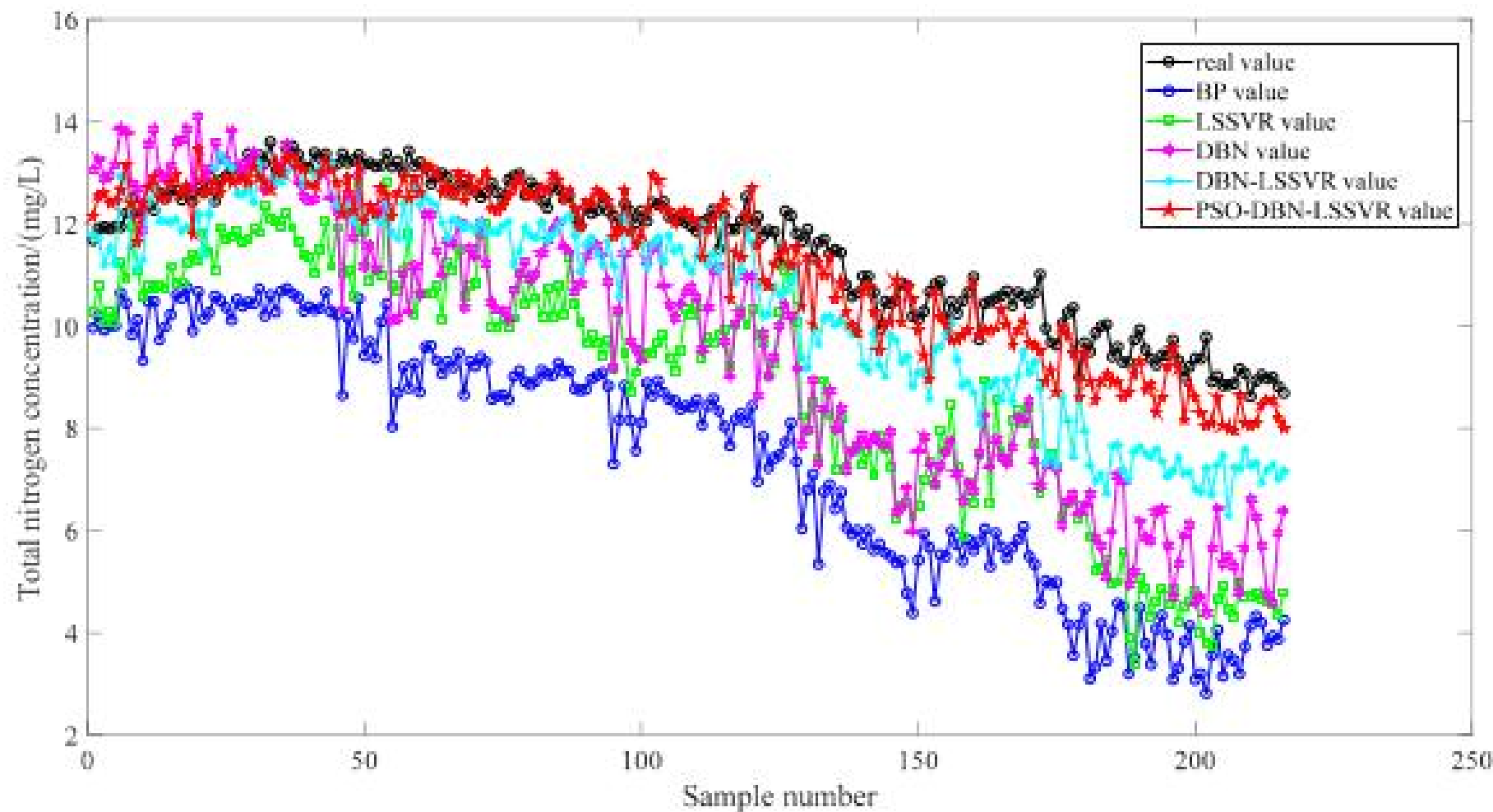
# 实验结果—真实结果对比



大连理工大学  
信息检索研究室



Information Retrieval Laboratory of DLUT



## ● 每四小时实验结果对比

Time	Original Value (mg/L)	BP Value (mg/L)	LSSVR Value (mg/L)	DBN Value (mg/L)	DBN-LSSVR Value (mg/L)	PSO-DBN-LSSVR Value (mg/L)
2019-02-20 00:00	13.22	8.6979	10.6828	10.1474	11.7031	12.5979
2019-02-20 04:00	13	9.1988	11.0839	11.0232	12.3903	12.9361
2019-02-20 08:00	13.43	8.7839	11.7840	10.2788	11.6919	12.5109
2019-02-20 12:00	13.14	9.2670	10.2441	11.1805	12.7838	12.9132
2019-02-20 16:00	13.2	8.7083	10.6196	10.6156	12.3741	12.6128
2019-02-20 20:00	13.08	9.5872	10.6417	12.1926	12.5264	13.1486
2019-02-21 00:00	12.78	9.6111	10.6462	12.1894	12.4182	13.1224
2019-02-21 04:00	12.92	9.3237	10.8038	11.4544	12.3650	12.9969
2019-02-21 08:00	12.99	9.0675	10.1418	10.9945	12.0069	12.6309
2019-02-21 12:00	12.9	9.3087	11.2319	11.6579	11.7297	12.7271
2019-02-21 16:00	12.59	9.1945	11.0802	11.6121	12.0948	12.6160
2019-02-21 20:00	12.74	9.4751	11.5087	12.0508	12.0435	13.0189
2019-02-22 00:00	12.84	8.6493	10.6011	10.3545	11.924	12.4661

# 实验结果—评价指标



Model	MAE	MAPE (%)	RMSE	$R^2$
BP	4.0943	36.99	4.2746	0.2871
LSSVR	2.8406	19.86	2.4957	0.6142
DBN	2.6679	24.54	2.9354	0.6454
DBN-LSSVR	1.1290	10.48	1.3306	0.8714
PSO-DBN-LSSVR	0.4765	4.32	0.4877	0.9327

- 提出一种基于DBN的水质参数深度学习方法  
预测断面水质
- 克服了随机初始化的DBN网络参数会影响模型的预测性能的缺点
- 下一步研究方向：
  - ◆ 收集重金属参数、流量等水质数据
  - ◆ 调整深度学习模型参数

# 谢谢！ 请多提意见！



大连理工大学  
信息检索研究室



Information Retrieval Laboratory of DLUT



大连理工大学

## 信息检索研究室

搜人搜物搜信息 重情重义重认知

<http://ir.dlut.edu.cn>