基于原料分类的常压塔软测量模型建立方法研究

党添添1，吕文祥1，黄德先1

1. 清华大学 自动化系, 北京 100084

E-mail: lvwx79@tsinghua.edu.cn

摘 要:针对原油蒸馏过程常规软测量模型难以适应原油进料性质变化的问题，提出了利用非监督数据驱动分类方法对原油进行分类，建立基于类别的非线性软测量模型的策略。通过聚类算法对不同加工原油的生产数据进行聚类，赋予训练样本所属类别。利用批标准化（Batch Normalization）和随机失活（Dropout）对神经网络进行优化，建立分类模型，提高了分类的鲁棒性和准确度。对进料原油先分类再对过程数据进行软测量建模，避免了原油性质差异对于软测量模型的影响，提高了软测量模型的回归精度和泛化水平。仿真结果表明，该基于原料分类建模的软测量模型对于原油性质有很好的适应性和预测准确性。

关键词: 原油蒸馏，软测量，无监督分类模型，人工神经网络。

**Research on Soft-sensor Modeling Method of Atmospheric Column Based on Crude Oil Classification**

Tiantian Dang1, Wenxiang Lyu1, Dexian Huang1

1. Department of Automation, Tsinghua University, Beijing 100084, China

E-mail: lvwx79@tsinghua.edu.cn

**Abstract:** Aiming at the problem that the conventional soft sensor model of crude oil distillation process is difficult to adapt to the change of crude oil feed properties, an unsupervised data-driven classification method was used to classify crude oil and establish a class-based nonlinear soft sensor processing model. The crude oil is clustered by the clustering algorithm, and the training samples are assigned to the category. BN and Dropout are used to optimize the neural network and establish the classification model, which improves the robustness and accuracy of classification. Based on the classification model, the process data is classified before modeling, which avoids the influence of the original property difference on the soft sensor model, and improves the regression accuracy and generalization level of the soft sensor model. Simulation results show that the soft sensor model based on classification has good adaptability and prediction accuracy for the original properties.

**Key Words:** Crude Distillation Unit, Soft-Sensor, Unsupervised Data-Driven Classification, Artificial Neural Network

[[1]](#footnote-1)

引言

常减压蒸馏过程是炼油生产过程的重要环节，原油经过常压、减压蒸馏塔，根据沸点高低分馏成多种质量指标不同的石油产品和后续加工装置的原料组份，其各侧线产品的组份合格和稳定对于整个炼油过程的生产安全稳定、产品质量合格继而提高原油资源利用率、节能降耗具有重要作用。如何精确地在生产过程中估计产品的质量指标是长期以来的复杂问题[1]。

在生产过程中往往有一些像常减压产品质量指标这种关键的变量难以直接测量，因而影响到直接质量控制的实现，只能使用对应的温度等变量进行间接控制。因此，需要对这些变量进行软测量(Soft sensor)。软测量的实际使用由来已久，许多测量仪表都是基于物理原理建立的易测过程变量与难以直接测量的待测过程变量的数学关系来计算获取[2,3]。这种意义上的软测量方法在1978年Brosillow提出的推理控制(Inferential control)的基本思想和方法后得到了系统化的描述：采集过程中比较容易测量的二次变量(Secondary Variable,或称辅助变量)，构造推断估计器来估计过程主要变量(Primary Variable)，并克服扰动和测量噪声对过程主要变量的影响[4]。推理控制的出现进一步推动了软测量技术的研究，基于各种模型的软测量方法得到了更为广泛的研究和应用[5]。软测量和预测控制相结合推动了炼油生产过程的直接质量闭环控制和卡边优化，即先进控制的发展和应用[1,3,6] 。

但是，软测量在炼油工业中的应用方面存在着一些制约性难题，原料性质不稳定和生产负荷变动致使在特定原料生产和负荷条件下所建立的模型产生模型失配。为了解决这一问题，吕文祥等将能够表现原油成份变化信息的过程变量加入软测量模型输入中，考虑其原油成份变化对软测量估计值的影响，同时，对直接测得的过程变量进行机理变换以减少加工负荷变化所造成的软测量模型非线性程度，明显提升了软测量模型的推广性能，在常减压装置的先进控制应用中取得了好的应用效果[7,8]。

由于原油种类频繁变化且各原油间性质差距明显，周长等提出了首先进行油品分类，再对不同油品分别建模的方法，并基于bootstrap模型在不同的三类原油上分别取得了对干点的良好预报效果，证实了利用类别结构信息的重要性[2]。

本文在周长等人的方法[9]上进行改进。首先，使用神经网络对不同加工原油生产过程数据进行分类，针对无监督的仿真数据，通过聚类赋予进料原油对应的类别，从而训练分类模型，实现无监督的训练。然后，对于每一类原油基于神经网络训练一个软测量模型。为了提高模型的鲁棒性和预测准确性，对两种模型的神经网络训练增加了BN层[10]和Dropout[11]来。本文将改进后的模型方法与原方法进行了比较，表明将分类模型和软测量模型结合起来，实现预测精度更高的石化产品质量指标的预测。

1. 基于分类的软测量建模
   1. 总体框架和改进措施

基于分类的软测量模型整体结构如图1所示。首先通过利用分类模型根据多种原油进料的比例，将进料原油分为多类（图中示意为三种原油，分成三类），然后根据进料原油所属类别，选择对应的软测量模型。最后根据所选择的软测量模型，对过程数据进行分析，预测产品质量指标。

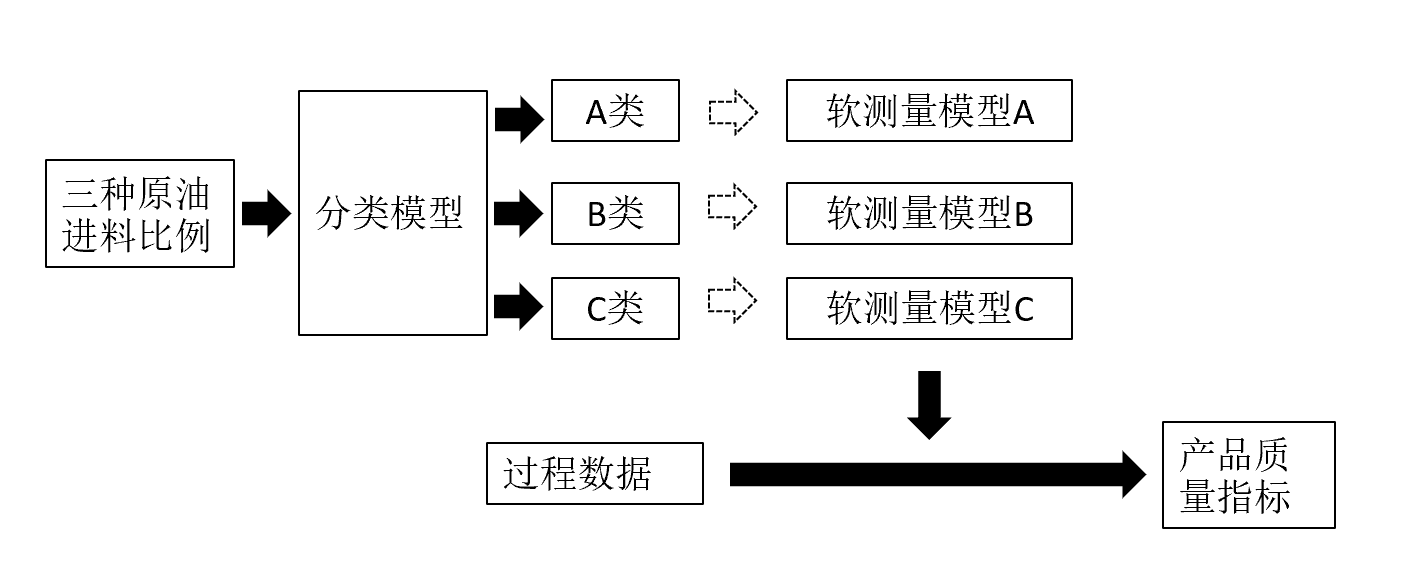


图1 整体结构图

总体框架分为两部分，一是基于进料原油的分类模型，研究分类方法和模型训练方法。

分类方法是根据多种原油进料()的比例，将混合的进料原油进行分类。分类的类别数可以和原油进料的种类数相一致。对于每个样本的所属类别，可以根据产品出料流量进行聚类，给予一个伪标签用于训练过程的监督信号。假设出料流量为()，进料流量和为()。根据两者的比值()进行kmeans聚类。将不同比例的原油进料混合的进料原油聚成A, B, C三类，并将这个聚类的类别作为其对应混合进料原油的伪标签用于分类模型的监督训练。

分类模型的输入是原油进料的比例。保持原油进料的总量一定，不同的原油进料比例混合起来作为不同进料原油。分类模型的输出是进料原油的类别。根据类别选择不同的软测量模型。

二是基于原料分类的软测量模型，研究模型训练方法和多个软测量模型的选择方法。软测量模型的输入是经过选择的过程变量等，输出是经过选择的产品质量指标。针对每一种进料原油，设计一种软测量模型。本文将进料原油分为了三类，所以设计了三种软测量模型。

因而，整体基于原料分类的常压塔软测量模型流程如下：根据进料原油的原油进料的比例进行分类，使用对应类别的软测量模型通过分析过程数据，预测产品质量指标。制约模型性能的条件主要有两个：分类方法的准确性和软测量模型的准确性。以往研究[1,3,6,7,8,9]中提到，以神经网络为基础的软测量模型在原油性质变化频繁的情况下的预测精度不高，模型的鲁棒性较差，容易陷入过拟合。近年来，随着数据量的增大和模型深度的增加，Batch Normalization(BN)和Dropout有效地缓解了模型的过拟合，增强了模型的鲁棒性，提高了模型的预测准确性。同时，增加了网络深度，为神经网络提供了更多的可学习参数，为模型提供了更多的训练空间。

本文参考周长等人[9]的方法，对于分类方法，使用BN和Dropout优化分类模型，提高分类准确率。同时，在软测量模型上也用BN和Dropout改进神经网络，减少模型的预测误差。下面简要介绍一下分类模型建模和软测量模型建模。图2，图3分别是分类模型和软测量模型的模型内部结构。在文献[9]的基础上，在神经网络的输入上增加BN和Dropout来增加模型的泛化能力和鲁棒性。

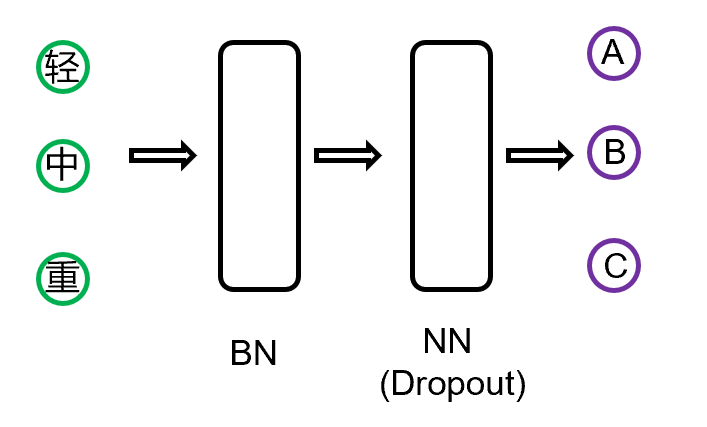


图2 融入BN和Dropout的分类模型网络结构

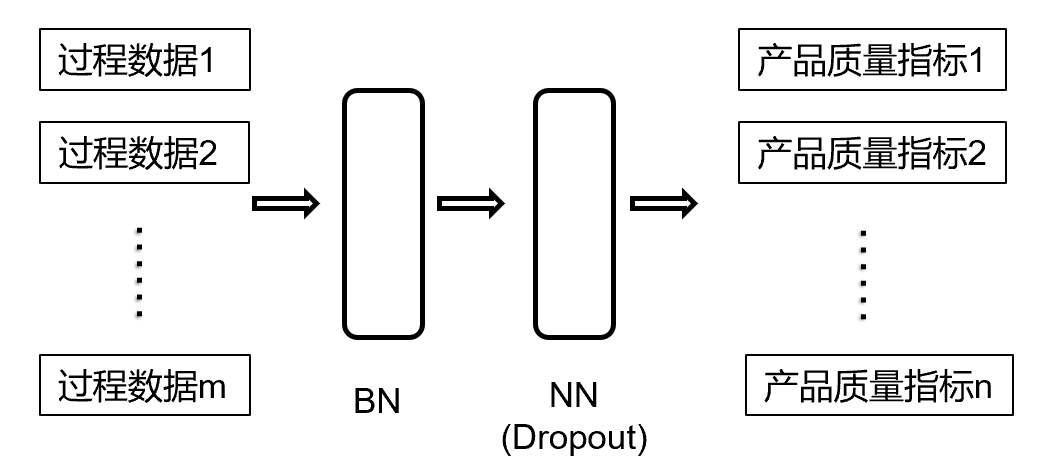


图3 融入BN和Dropout的软测量模型网络结构

* 1. Batch Normalization

BN是对统一批量（batch）的数据进行正态正则化后再进行尺度变换和位移变换的模型结构。具体算法如图4所示。首先，BN对于输入的数据进行预处理，让每一维的均值为0，标准差为1。去除了特征之间的相关性，使得所有特征具有相同的均值和方差。这样处理后的数据的分布会更加集中，离群值的影响大大减小。由于神经网络内部存在内部协变量转移的问题，即各层的输入和输出的分布是不同的，所以多层的神经网络在训练过程中非常容易陷入局部最优，训练的速度比较慢。而正则归一化恰恰能够解决这一个问题。

但是这种强硬的归一化处理，破坏了数据之间本身的特征分布。为了保证数据在归一化之后仍然有相应的特征分布，算法中引入了可学习的参数γ和β用于产生数据偏移，从而恢复数据之间的特征分布。归一化和数据偏移构成了BN的主要内容。使用BN会为神经网络提供了以下几个特点：

（1）加快收敛速度。神经网络中，如果每一层的数据分布都不一样的话，网络的收敛速度是比较慢的。正则化可以让减小数据之间分布的不同，加快收敛。

（2）防止梯度消失和爆炸。如果网络的激活输出（以sigmoid为例）过大，其对应的梯度会比较小，这样网络的学习速度会很慢，甚至难以收敛。正则化保证了数据激活输出不会过大，保证梯度的数值不会过大和过小。

（3）防止过拟合。由于BN层的训练过程中会出现整体的均值和方差，所以模型的输出不仅仅取决于样本本身，还取决于batch中其他样本的情况。从而让模型避免过于关注某个样本，减少过拟合的情况。

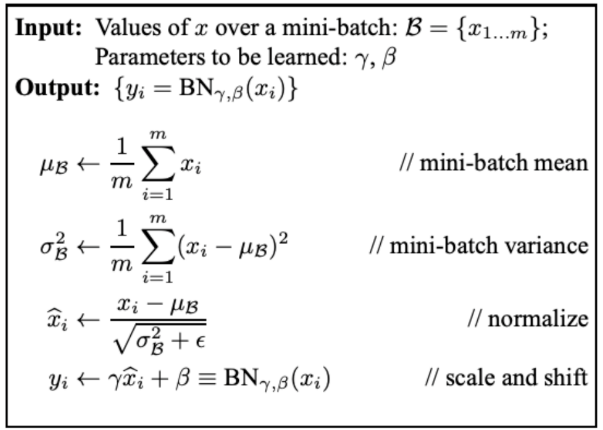


图4 BN算法流程，x是一个mini-batch

假设训练集和测试集是相同分布的。BN层的归一化，降低了数据特征分布之间的不平衡性，使得神经网络在每一层的输入上都是同分布的。这样保证了每一层的训练效果是稳定的，大大减少了优化参数的选择难度。从而保证了训练的效率和效果。

* 1. Dropout

Dropout是仅仅用在训练过程中的一种算法。它的具体操作是在每一个训练批次中，通过忽略一般的特征检测器，让一部分隐层节点的值为0，从而减少过拟合现象。如图5所示。将Dropout应用到神经网络中，可以发现它让整个网络变得稀疏，从而减少训练的拟合程度，提高模型的泛化能力。

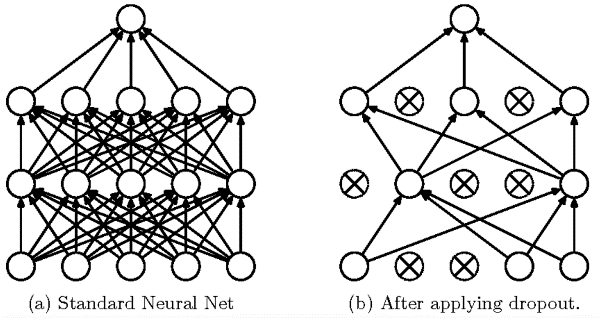


图5 融入dropout的网络结构

Dropout的具体流程如下：在前向过程中，以概率p让一部分神经元的停止工作，即让这部分激活函数的输出为0。在反向过程中，让梯度不流经这部分神经元，不更新这部分神经元的参数，直到参数全部更新完毕之后，再把这部分神经元恢复。重复这一流程。这样便可以在更新全部参数的同时，保证模型的泛化能力，并解决过拟合的问题。

1. 仿真验证
   1. 仿真数据准备

本文采用Unisim Design流程模拟软件模拟常减压分馏过程，并通过调整轻、中、重三种不同性质原油的配比，获得相关的可测量过程数据和质量指标数据。

如图6所示，基于Unisim Design流程模拟软件，搭建了虚拟常减压蒸馏塔仿真平台，其中常压塔侧线产品包括了汽油、煤油和柴油三种石油产品。

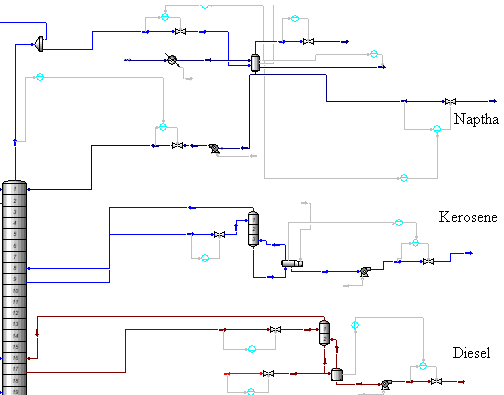


图6 常减压塔局部流程图

使用Unisim Design流程模拟软件模拟了三种原油进料，设为轻油、中油、重油，其实沸点曲线如图7所示。通过流量控制回路控制三种原油的进料量，并保持其总流量稳定。



图7 原油实沸点曲线

仿真数据的生成分为两个部分，第一部分是生成用于分类方法研究的仿真数据。第二部分是用于基于原料分类的软测量模型建模的仿真数据。上文已经提到，仿真的输入是三种原油进料的比例，输出是经过选择的产品质量指标。为了保证我们得到的产品质量指标不会随着进料原油的性质变化而变化，本文通过自动控制来调节仿真软件中的过程变量使产品质量指标基本保持不变。经过挑选的产品质量指标共有五个，分别是：常一线初馏点、常一线终馏点、常一线闪点、常二线终馏点、常三线终馏点。保持这五个质量指标的变化不超过10℃，获得产品出料流量。

设计多种原油进料比例，每改变一次原油进料的比例，我们就需要调整一次仿真软件的过程变量，确保产品质量指标保持基本不变。

软测量模型变量的选择上，根据与作为输出变量的产品指标的相关性来选择所需要的过程变量。上述质量指标软测量模型的输入变量是运行过程中的可测数据，包括顶循回流量、一中回流量、二中回流量，顶循回流采出温度和返塔温度，一中回流采出温度和返塔温度，二中回流采出温度和返塔温度，塔顶压强和温度等等。输出变量包括常一线初馏点、常一线终馏点、常一线闪点、常二线终馏点、常三线终馏点和常顶终馏点。将软件模拟流程设计为动态流程，每次调整仿真对象输入数据，包括原油进料比例和仿真中的某些过程变量等，待仿真环境达到稳态后，进行产品质量指标数据的收集。

为了保证仿真的数据更符合真实的流程数据，我们对于仿真数据增加了高斯噪声。然后我们将样本按照训练集、验证集、测试集8:1:1的比例进行分割。利用BN和Dropout改进之后的人工神经网络进行模型的训练和选择。

* 1. 仿真结果分析

影响产品质量指标的因素有很多，包括原油进料性质，过程温度，过程回流量等等。在分析机理的基础上，选择25个过程变量以及进料变量作为软测量模型建模的输入。同时，为了减少进料量绝对值的影响，将与流量有关的量都换算成对应进料流量和的比率，从而提高模型的鲁棒性和泛化能力。

分类模型和软测量模型使用的神经网络都是融入了BN和Dropout的MLP。在实验中，使用Adam优化算法进行梯度下降。每个神经元的激活函数都是Relu函数，有效的避免了梯度爆炸和梯度消失，同时保证了非线性映射。具体的神经网络实现细节按照[9]实现。数据初始化按照xvaier初始化[12]的方法，一定程度上减少了随机误差。

表1是分类模型精度的对比结果。Aggre-gate network是周长等人[9]的分类模型在测试集上的准确率。本文方法是在他们所用的神经网络上增加BN和Dropout，测试改进方法在测试集上的准确率。表1说明改进的方法相较于前人的方法在准确率上有很大提升，说明改进的分类方法有更好的效果。

表1 分类方法在测试集上的准确率对比

|  |  |  |
| --- | --- | --- |
| Model | Aggregated network[9] | Adjusted NN |
| accuracy | 96.53% | 99.87% |

将进料原油根据机理分为三类，分为轻中重三类。比较分类时的不同类别的分类准确率，结果如表2所示。

表2 不同类别的分类准确率

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | All | Light | Medium | Heavy |
| Aggregated | 96.53% | 100% | 90.96% | 98.63% |
| Modified method | 99.81% | 100% | 99.72% | 99.7% |

仿真实验对比了分类后进行软测量建模和不分类直接进行软测量建模的精度。首先，验证了用BN和Dropout对于软测量模型建模的影响，如表3所示。用未加BN和Dropout的神经网络作为基准，改进的方法是增加了BN和Dropout的神经网络。为了体现BN和Dropout的通用性，在未分类的训练数据上进行软测量建模，即只用一个软测量模型，后续实验分析了分类之后的应用BN和Dropout的软测量模型的效果。从表3中看到，增加了BN和Dropout的神经网络的RMSE要小于未加BN和Dropout的神经网络。这说明增加BN和Dropout之后的神经网络精度更高。同时不难发现，未加BN和Dropout的训练尽管能够收敛，但是不能在测试集上取得更好的效果。这部分实验验证了BN和Dropout对于软测量模型的作用。

表3测试集上平均RMSE

|  |  |  |
| --- | --- | --- |
| Model | Original NN | Modified method |
| RMSE | 3.0 | 2.8 |

下面分析分类方法对于软测量模型建模的作用。表4是改进的方法对比比较周长等人的实验结果。为了保证实验的可靠性，使用的是现有的仿真数据，而不是前人论文中的原本数据，所以测试的Aggregated network的结果和原本论文中的结果有所不同。

表4 分类建模后软测量模型RMSE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | All | Light | Medium | Heavy |
| Aggregated | 10.71 | 4.43 | 5.52 | 7.28 |
| Modified method | 3.52 | 4.24 | 2.03 | 3.32 |

表4的实验结果说明两个问题。第一，相对于之前的工作，改进的软测量模型建模的精度更高，预测能力更强，无论是在分类之后的还是未经分类的。第二，分类之后在进行软测量建模的精度往往比不分类直接进行软测量建模的精度要高。这说明分类方法是行之有效的，对于软测量模型的精度提升是很明显的。不过注意到，分类之后的轻油的软测量建模的精度不如不分类的软测量模型的建模精度。这里可能的原因是数据量过少。由于仿真数据的生成比较复杂且不够精确，当数据过少时，划分的训练集和测试集的分布可能相差比较大。这样的分布状况会导致训练集上的拟合效果可能很难直接移植到测试集上，模型的泛化能力受到阻碍，所以精确度不高。倘如生成更精确的仿真数据，这部分误差也会得到修正。总之，上述实验印证了经过精确分类之后进行的软测量建模的精度是可以得到保障的。

通过综合利用BN和Dropout，在周长等人[9]的方法上进行改进，即对于研究中所用到的神经网络，在每一层的输入上增加BN层，在每一层的输出上增加Dropout，极大的提高了模型的训练速度和泛化能力。通过这些改进，在分类方法上，分类的准确率得到提高，分类模型的收敛也得到了保障；在软测量建模上，软测量模型回归精度提高。在这两个模型基础上，分析先分类再进行软测量建模的精度变化，对比分析不分类直接进行软测量模型建模的精度，结果证明本文的方法可以大大减小软测量模型建模的误差。

1. 结论

本文主要介绍了基于分类的对于进料原油性质变化的原油蒸馏过程非线性软测量建模。通过仿真实验验证可以发现，通过聚类+分类方法可以实现无监督分类，用BN和Dropout对神经网络进行优化可以使分类模型和软测量模型有更好的准确率和鲁棒性，经过分类之后进行的软测量模型有更好的回归精度。

**参考文献**

1. 黄德先; 江永亨; 金以慧. 炼油工业过程控制的研究现状, 问题与展望. 自动化学报, 2017, 43.6: 902-916.
2. Van Den bos A, Eykhoff P. Model building and parameter estimation as means for intelligent measurement. Measurement, 1988, 6(1): 25-32
3. 袁璞,炼油过程先进控制技术的发展与应用, 石油炼制与化工, 1994, 25(10): 28-33)
4. Brosilow C, Tong M. Inferential control of processes: 2. structure and dynamics of inferential control systems. AIChE Journal, 1978, 24(3): 492-500 [4] Souza, Francisco AA, Rui Araújo, and Jérôme Mendes. "Review of soft sensor methods for regression applications." Chemometrics and Intelligent Laboratory Systems 152 (2016): 69-79
5. Souza, Francisco AA, Rui Araújo, and Jérôme Mendes. "Review of soft sensor methods for regression applications." Chemometrics and Intelligent Laboratory Systems 152 (2016): 69-79.
6. 黄德先, 叶心宇, 竺建敏, 李秀改. 化工过程先进控制. 北京: 化学工业出版社, 2006
7. 吕文祥, 摆亮, 黄德先, 江永亨, 刘先广, 朱朝宾, ... & 张代欣. (2009). 针对原油性质变化的常减压先进控制研究与应用. 化工学报, (10), 2522-2528.
8. 吕文祥，黄德先，等. 常压蒸馏产品质量软测量改进方法及应用．控制工程，2004，11（4）：296-298.
9. Zhou, Chang, et al. "Inferential estimation of kerosene dry point in refineries with varying crudes." Journal of Process Control 22.6 (2012): 1122-1126.
10. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. pmlr, 2015: 448-456.
11. Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv preprint arXiv:1207.0580, 2012.
12. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010: 249-256.

1. [↑](#footnote-ref-1)