

利用对抗自编码器生成具有分析物属性信息的近红外光谱样本来增强近红外高斯过程回归模型

赵新宇, 李统乐, 黄光造*

温州大学电气与电子工程学院, 浙江 温州

收稿日期: 2024年3月1日; 录用日期: 2024年6月10日; 发布日期: 2024年6月20日

摘要

通过在近红外光谱校正模型中添加人工生成的光谱样本, 是增强近红外校正模型预测效果的一种有效方法, 尤其对于建模样本数不足的情况。本文提出一种基于对抗自编码器(Adversarial Autoencoder, AAE)的近红外光谱样本生成方法, 并将利用该方法生成的光谱样本补充到近红外高斯过程回归建模集中, 以增强模型的预测效果。该方法基于AAE模型, 利用编码网络将光谱的低维特征分布映射成某种先验分布, 利用解码网络将先验分布中的样本点重构成近红外光谱。因此, 本文利用AAE模型成功的从三个典型的近红外光谱数据集中生成光谱, 然后将这些生成的光谱应用到高斯过程回归模型中。实验结果显示, 所提方法能够有效的生成与建模集同分布的近红外光谱样本, 并提高高斯过程回归的预测性能。由此总结, AAE有潜力生成衍生同源光谱, 并在一些小样本集中扩增光谱数量。

关键词

近红外光谱, 对抗自编码器, 高斯过程回归

The NIR Gaussian Process Regression Model Is Enhanced by Using the Adversarial Autoencoder to Generate NIR Spectral Samples with Analyte Property Information

Xinyu Zhao, Tongle Li, Guangzao Huang*

College of Electrical and Electronic Engineering, Wenzhou University, Wenzhou Zhejiang

*通讯作者。

文章引用: 赵新宇, 李统乐, 黄光造. 利用对抗自编码器生成具有分析物属性信息的近红外光谱样本来增强近红外高斯过程回归模型[J]. 光电子, 2024, 14(2): 35-43. DOI: 10.12677/oe.2024.142005

Abstract

Adding artificially generated spectral samples to near-infrared spectral correction models is an effective method to enhance the prediction effect of near-infrared spectral correction models, especially when the number of modeling samples is insufficient. In this paper, a method of generating near-infrared spectral samples based on Adversarial Autoencoder (AAE) is proposed, and the spectral samples generated by this method are added to the near-infrared Gaussian process regression modeling set to enhance the prediction effect of the model. This method is based on AAE model, using the coding network to map the low-dimensional feature distribution of the spectrum to some prior distribution, using the decoding network to reconstruct the sample points in the prior distribution to form the NIR spectrum. Therefore, this paper successfully generates spectra from three typical NIR spectral datasets using AAE model, and then applies these generated spectra to Gaussian process regression models. Experimental results show that the proposed method can effectively generate NIR spectral samples with the same distribution as the modeling set, and improve the prediction performance of Gaussian process regression. In conclusion, AAE has the potential to generate derived homologous spectra and amplify spectral numbers in some small sample sets.

Keywords

Near Infrared Spectrum, Adversarial Autoencoder, Gaussian Process Regression

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

近红外光谱检测是一种常用的化学分析技术，具有快速、非破坏性、高灵敏度等优点，已经被广泛应用于农业、化工、食品，医药等领域。近红外光谱是一种间接的分析技术，需要建立反应光谱响应与分析物属性间潜在关系的校正模型，才能实现对样品属性的分析。在实际的近红外光谱检测应用中，会遇到样本数不足的情况，导致所构建的校正模型处于过拟合的状态，限制了近红外光谱检测技术的应用发展。目前，近红外光谱校正模型的构建，较多的还是采用基于小样本的传统建模方法，比如偏最小二乘回归模型、最小二乘支持向量机等。随着人工智能的快速发展，一些新的模型校正方法也开始应用于近红外光谱检测领域，并展露出一定的优势，比如基于深度学习的卷积神经网络[1] [2] [3] [4]。这些新的校正模型往往复杂度较高，需要更多的样本数进行训练才能发挥其对复杂关系的拟合效果，建模样本数的不足已经成为制约复杂近红外光谱校正模型构建的一个瓶颈。

在深度学习领域，对于建模样本数不足的情况，一个常用的解决方法是采用人工的方式进行样本生成。样本的人工生成在图像领域中已经得到了较多的研究[5] [6] [7]。相较于肉眼可辨的图像数据，近红外光谱反应的是分析物的物理与化学性质，具有吸收强度较弱，光谱的信噪比低，并且倍频、合频谱带重叠现象严重的特点，比其图像数据更加抽象和难以理解[8] [9]。

传统的近红外光谱样本生成方法主要是模拟在样本进行近红外光谱采集时仪器的基线扰动对光谱响

应的轻微影响,使相同的分析物属性对应多种光谱响应,以达到样本增强的效果。相较于传统的近红外光谱样本生成方法,生成对抗网络(Generative Adversarial Network, GAN)提供了一种新的样本生成策略。GAN 是一种包含生成器(Generator)和判别器(Discriminator)两部分的神经网络。生成器将随机噪声生成与真实样本相似的样本,判别器则用来区分生成样本和真实样本是否一样,二者通过对抗训练,最终达到纳什平衡[10][11][12]。基于生成对抗策略的生成模型已经在许多领域得到应用,尤其在数据增强方面效果显著。Zhun Zhong [13]等人在图像风格同化的任务中,利用 CycleGAN (Cycle-Consistent GAN)生成图像样本来扩展训练数据集。Yu 和 Zhang [14]等人借鉴了强化学习(RL)的策略,提出 SeqGAN (Sequence GAN)来生成文本序列来改善模型对文本的预测性能。Frid-Adar [15]等人通过使用 GAN 生成肝部影像样本来提高肝损伤图像分析的准确性和鲁棒性。Teng [16]等人使用 GAN 扩展了激光诱导击穿光谱(LIBS)的数量,但是没有给出生成光谱的具体应用。Xu [17]等人通过 BEGAN (Boundary Equilibrium GAN)来生成近红外光谱样本,与共识算法(CDPLS)相结合提升了 PLSR 模型的预测精度。

AAE 是一种结合了生成对抗策略的自编码器模型[18],相较于经典的 GAN 网络,AAE 可以连续对比输入数据与生成数据间的差异,因此,AAE 模型的训练过程较传统的 GAN 更稳定,生成的样本也具有更好的连续性。本文探讨 AAE 模型生成模拟小样本近红外光谱数据集对高斯过程回归的增强效果,包括通过一些公认的评价指标来评估增强性能。

2. 理论

2.1. AAE 生成方法

AAE 是一种结合了自编码器(Auto-Encoder, AE)和 GAN 的深度学习模型。如图 1 所示,AAE 通过编码器将高维空间中的样本数据映射成低维空间中的潜变量,然后再通过解码器将潜变量重构为原始样本,由于判别器的存在使得潜在变量空间服从先验分布。通过对抗训练,自编码器和判别器相互竞争,编码器试图生成更符合先验分布特征的潜变量,而判别器则试图准确判别生成的潜变量和先验变量间的差异。这种对抗训练的过程可以迫使自编码器学习到更丰富的先验特征表示,生成更多样化和逼真的样本数据。

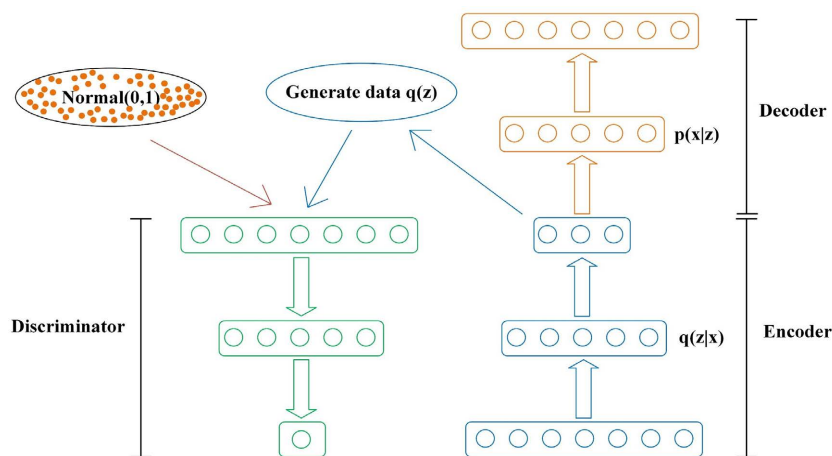


Figure 1. AAE structure diagram
图 1. AAE 结构图

其中 AAE 具体的细节变量如下所述。 X : 输入样本数据向量; X^{\parallel} : 重构样本数据向量; $q(z|x)$: 编码器的分布函数; $p(x|z)$: 解码器的分布函数; Z : 潜在变量空间, 编码出来的潜变量, 服从 $z \sim q(z)$ 分

布; $q(z)$: AAE 模型学习得到的分布; $p(z)$: 预先定义的任意先验分布, 通常为标准高斯分布 $Normal(0,1)$ 。AAE 采用的是聚合后验分布:

$$q(z) = \int_x q(z|x) p_d(x) dx, \quad (1)$$

其中 $p_d(x)$ 为真实数据的分布。

AAE 是一种神经网络, 它的训练过程涉及到两个网络: 自编码器和判别器。相关的训练要点总结如下:

- AAE 是通过将聚合后验分布 $q(z)$ 和任意先验分布 $p(z)$ 进行匹配来完成正则化, 让对抗网络不断地使 $q(z)$ 去匹配 $p(z)$;
- 在对抗的过程中, 自编码器和判别器也在不断地最小化目标损失;
- 编码器的目标是确保聚合的后验分布 $q(z)$ 可以欺骗判别器, 使其像是来自真实先验分布 $p(z)$ 。

本文以带 1 维分析物属性的小麦样本生成为例, 进行 AAE 模型构造方法的讨论, 所采用结构如图 2 所示。编码器是由全连接神经网络形成的网络架构, 与解码器相对称, 判别器是通过全连接和卷积神经网络所构成, 其中所拟合分布设置为二维正态分布。初始样本变量在经过编码器 4 个潜在层映射后得到潜变量, 再通过解码器重新映射成原有数据结构, 得到重构损失。判别器则判别自编码器的潜变量是否来自所拟合的分布, 得到判别损失。设置重构损失和判别损失的相对权重为 9:1。在训练初始模型时, 使用 Adam 优化器, 以 0.0004 的初始学习率(learning rate)进行训练, 样本批处理大小设置为 16, 模型训练的世代数为 20,000。对于训练完毕的模型, 对 AAE 模型的解码器输入来自所拟合分布的随机抽样样本, 解码得到相应的近红外光谱生成样本。

本文采用上述的网络结构主要用于 AAE 模型对近红外光谱样本的生成, 虽然本文所采用的结构比较简单, 但是已经可以满足常规近红外光谱样本生成的需要。和其它的深度学习神经网络模型一样, AAE 模型的结构和参数设置可以根据实际问题的复杂程度适当的调节。

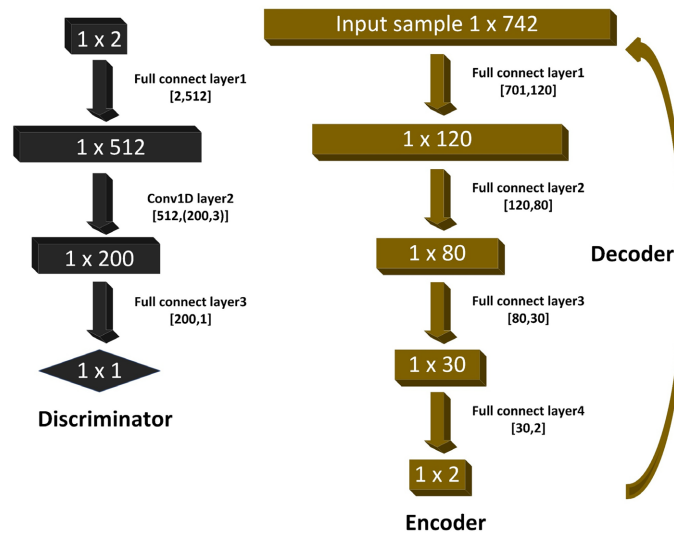


Figure 2. AAE for wheat sample network structure; Encoders and decoders have symmetrical structures

图 2. AAE 对于小麦样本网络结构; 编码器和解码器具有对称结构

2.2. 高斯过程回归

本文采用高斯过程回归(Gaussian Process Regression, GPR)来评价 AAE 生成的光谱样本对高斯过程回

归模拟的增强效果[19]。GPR 是一种非参数的回归分析方法，它基于高斯过程来建立预测模型。GPR 方法假设样本之间的关系可以由一个连续的高斯过程来近似描述，通过计算这个高斯过程的均值和方差，可以获得对未知数据的预测。

假设 $\{f(x), x \in H\}$ 是一个随机变量集，其中 H 是指示集。如果对于任意 n , $x_1, x_2, x_3, \dots, x_n \in H$ 是服从混合高斯分布的变量，然后 $\{f(x), x \in H\}$ 称为由均值函数 $m(x)$ 和协方差函数 $k(x_i, x_j)$ 所决定的高斯过程记为：

$$f(x) \sim GP\{m(x), k(x_i, x_j)\}, \quad (2)$$

$$m(x) = E[f(x)], \quad (3)$$

$$k(x_i, x_j) = E[(f(x_i) - m(x_i)) \cdot (f(x_j) - m(x_j))], \quad (4)$$

通过对原始数据进行归一化，可以假设均值函数 $m(x)$ 为零。

考虑训练集为 $D = \{(x_i, y_i) | i = 1, 2, 3, \dots, n\}$ ，其中样本数量为 n , $x_i \in R^d$ 和 $y_i \in R$ 分别是相应的输入和输出。设 $X = [x_1, x_2, x_3, \dots, x_n]$ 表示输入 $n \times d$ 矩阵， Y 表示 $n \times 1$ 输出向量，那么训练集可以表示为 $D = (X, Y)$ 。对于新的输入 x_* ，GPR 的任务是计算出输入 x_* 所对应输出 y_* 值的条件分布 $P(y_* | x_*, X, Y)$ 。认为 GPR 为：

$$y = f(x) + N(0, \sigma_n^2), \quad (5)$$

其中 $f(x) \sim GP(m(x), k(x_i, x_j))$ 并且 σ_n^2 是噪声强度。 Y 的先验分布为：

$$Y \sim N(0, K + \sigma_n^2 I), \quad (6)$$

其中 $K = k(X, X)$ 是一个 $n \times n$ 的对称正定协方差矩阵，矩阵中的 $K_{ij} = k(x_i, x_j)$ 表示 $f(x_i)$ 和 $f(x_j)$ 之间的相关性。当一对样本 (x_*, y_*) 被添加时，根据 GPR 的定义，随机向量 $(f(x_1), f(x_2), f(x_3), \dots, f(x_n), f(x_*))$ 符合 $n+1$ 维混合高斯随机分布，记为：

$$\begin{bmatrix} Y \\ f(x_*) \end{bmatrix} \sim N\left(0, \begin{bmatrix} K + \sigma_n^2 I & K_*^T \\ K_* & K_{**} \end{bmatrix}\right), \quad (7)$$

其中 $K_* = [k(x_*, x_1), k(x_*, x_2), k(x_*, x_3), \dots, k(x_*, x_n)]$, $K_{**} = k(x_*, x_*)$ 。这些展开得：

$$P(y_* | x_*, X, Y) = N\left(K_* (K + \sigma_n^2 I)^{-1} y, K_{**} - K_* (K + \sigma_n^2 I)^{-1} K_*^T\right). \quad (8)$$

这个分布的均值可以作为 x_* 对 y_* 的预测值，并且分布的方差可以作为置信度。如图 3 所示，置信区间的宽度取决于模型的不确定性，即置信区间越窄，说明预测值 y_* 分布的方差越小，意味着 GPR 模型对于该点的预测非常自信，反之，置信区间越宽，说明 GPR 模型对于预测不太自信，可能需要谨慎对待这些预测结果。因此，本文在构建 GPR 模型的过程中，将预测值方差大的样本通过 AAE 生成填充到建模集中，使输出预测值 y_* 具有更高的预测置信度，从而提升模型预测精度。

3. 实验设计

3.1. 数据

在本研究中使用了三个公共数据集，旨在探讨 AAE 应用在不同样品光谱数据的可行性。第一个数据集包含了 215 个猪肉样本数据，使用 Infratec Tecator 仪器采集光谱，波长范围在 850~1050 nm，其中样本属性包括水分、脂肪和蛋白质含量[20]。第二个数据集包含 248 个小麦数据，使用制造厂商 A 仪器

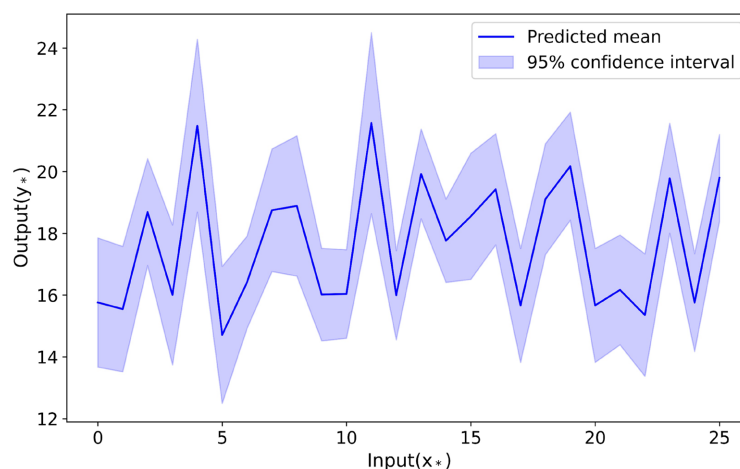


Figure 3. Gaussian process regression
图 3. 高斯过程回归

采集光谱, 波长范围在 730~1100 nm, 其中样本属性为蛋白质含量(available from https://www.cnirs.org/content.aspx?page_id=22&club_id=409746&module_id=239453)。第三个数据集包含了 462 个由配方 R862 制备的三聚氰胺 - 甲醛样本数据, 使用 FT-NIR 光谱仪采集光谱, 其中样本属性为样本缩合过程中聚合的程度[21]。

3.2. 实验

实验由三个部分组成: 首先, 将猪肉、小麦、三聚氰胺 - 甲醛样本数据都按 3:7 的比率划分训练集和测试集, 使训练集处于不充分状态, 通过训练集来模拟构建小样本 GPR 模型。其次, 根据不同原始数据训练集样本来确定 AAE 模型的超参数, 训练构建好的模型并实时监测 AAE 在不同训练阶段生成的光谱样本, 以及自编码器的重构误差和判别器的准确率, 以确保训练进程。最后, 采用决定系数(R^2)和均方根误差(RMSEP)来评价通过 AAE 样本增强对小样本 GPR 模型的提升效果。评估指标 R^2 和 RMSEP, 定义如下:

$$RMSEP(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (9)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (10)$$

其中 n 是样本个数, y_i 是样本的真实值, \hat{y}_i 是第 i 个样本的预测值, \bar{y} 是测试集的平均值。

所有程序都在台式电脑上进行了测试, 其特定的软件配置和版本为: Intel Core i7-7700 HQ CPU, 8GB RAM, NVIDIA GeForce GTX 1080 Ti, Python 3.9, Tensorflow-GPU 2.6.0。

4. 结果与讨论

4.1. 对比生成样本与真实样本

AAE 的主要目的是生成尽可能与真实样本分布相匹配的样本数据, 可以通过观察样本的概率密度曲线来检测生成的样本分布是否与真实样本分布相匹配。

在 AAE 模型保持稳定状态后, 将模型保存下来, 停止训练。以小麦数据为例, 利用保存下来的解码器, 通过二维随机高斯噪声($N \times 2$)生成一系列样本数据, 取 $N = 148$ 来生成与小麦训练集等量的生成样本。如图 4 所示, 粉色柱状图和曲线表示真实数据分布, 紫色表示通过 AAE 生成的数据分布, 可以观察到两者的分布基本相似, 这说明通过 AAE 生成的样本数据具有与真实数据相似分布特征。

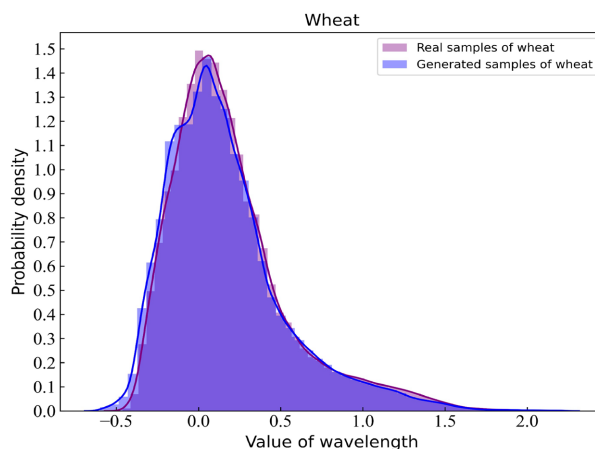


Figure 4. Probability density curve of real sample and generated sample
图 4. 真实样本与生成样本的概率密度曲线

4.2. 利用 AAE 生成样本增强 GPR 模型

对于不同的样本数据都使用 AAE 生成 n ($n \geq 1500$) 个样本, 以保证生成样本的多样性, 再分别利用训练集样本构建 GPR 模型, 通过输入测试集样本得到 m (m 等于样本测试集) 个分布方差, 分布方差表示模型对预测结果的不确定性, 方差越大意味着模型对目标值预测越不准, 也说明原始训练集中缺少目标值所对应的样本变量, 可以通过数据增强的方式来弥补。通过将 m 个分布方差对应的预测均值作为筛选生成样本属性标签的基准, 利用生成的 n 个样本的属性标签来选择距离基准最近的 m 个样本, 并将这 m 个样本作为增强样本。然后, 将增强样本补充到原始样本训练集中构成增强训练集, 使用增强的训练集构建 GPR 模型, 其在测试集上的预测性能反应了数据增强的效果, 实验结果如表 1。

Table 1. Experimental results of wheat and R862 data augmentation in regression models. Δ represents the increment of training set samples, m represents augmented samples

表 1. 小麦和 R862 数据在回归模型中增强的实验结果。 Δ 表示训练集样本的增量, m 表示增广样本

Dataset	Δ	GPR	
		R^2	RMSEP
Wheat	0	0.8495	0.8262
	$\frac{m}{2}$	0.8573	0.8046
	m	0.8708	0.7655
R862	0	0.8094	7.1547
	$\frac{m}{2}$	0.8664	5.9894
	m	0.8943	5.3277

由于猪肉数据集含有三种类别的属性标签, 可以针对每一种标签来分别进行 AAE 样本生成, 从而实现不同分析物属性的样本增强, 其回归模型增强效果如表 2。

Table 2. Experimental results of meat data augmentation in regression models. Δ represents the increment of training set samples, m represents augmented samples

表 2. 猪肉数据在回归模型中增强的实验结果。 Δ 表示训练集样本的增量, m 表示增广样本

Analyte	Δ	GPR	
		R^2	RMSEP
moisture	0	0.8199	4.229
	$\frac{m}{2}$	0.8813	3.4333
	m	0.9044	3.3681
fat	0	0.8638	4.7094
	$\frac{m}{2}$	0.8798	4.4251
	m	0.8991	4.0549
protein	0	0.7787	1.4337
	$\frac{m}{2}$	0.8133	1.3169
	m	0.8240	1.2783

通过 AAE 样本生成在三个数据集上都表现出增强效果, 并且在每个增强数据集上的 GPR 预测性能都有所提升。因此, AAE 具有非常强大的学习能力, 能够完美的处理光谱响应与分析物属性间的潜在对应关系, 从而对小样本近红外数据集进行人工高质量扩充。

5. 结论

本文研究基于 AAE 方法的近红外光谱样本增强, 通过对三个典型的近红外光谱数据进行增强来验证增强样本在 GPR 上的性能表现。实验结果显示, AAE 能够有效地学习到近红外光谱样本数据的分布信息, 从而使生成的样本数据服从原有数据分布, 这得益于生成对抗网络的强大学习能力。此外, AAE 生成的光谱具有更大的内部多样性和与真实光谱的相似度, 并且对原始训练集具有更好的互补效果。因此, AAE 可以根据学习到的数据分布知识, 合理地创建这些区域的光谱样本。更具体地说, AAE 可以生成高质量的浓度值在训练集之外的光谱样本, 特别是那些理论上存在但在实践中难以收集的光谱样本。

综上所述, AAE 生成具有浓度值的光谱样品的能力。因此, AAE 可以作为一种新的光谱样本生成方法来增强光谱数据集, 从而提升光谱 GPR 性能。

致 谢

作者感谢研究生培养经费的资助。

参考文献

- [1] Le, B.T. (2020) Application of Deep Learning and near Infrared Spectroscopy in Cereal Analysis. *Vibrational Spec-*

- troscopy*, **106**, Article 103009. <https://doi.org/10.1016/j.vibspec.2019.103009>
- [2] Qu, J., Liu, D., Cheng, J., Sun, D., Ma, J., Pu, H., *et al.* (2015) Applications of Near-Infrared Spectroscopy in Food Safety Evaluation and Control: A Review of Recent Research Advances. *Critical Reviews in Food Science and Nutrition*, **55**, 1939-1954. <https://doi.org/10.1080/10408398.2013.871693>
- [3] Li, R., Liu, Y., Xia, Z., Wang, Q., Liu, X. and Gong, Z. (2023) Discriminating Geographical Origins and Determining Active Substances of Water Caltrop Shells through Near-Infrared Spectroscopy and Chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **303**, Article 123198. <https://doi.org/10.1016/j.saa.2023.123198>
- [4] Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L. and Engelsen, S.B. (2000) Interval Partial Least-Squares Regression (ipls): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy*, **54**, 413-419. <https://doi.org/10.1366/0003702001949500>
- [5] Pumarola, A., Agudo, A., Martinez, A.M., *et al.* (2018) Ganimation: Anatomically-Aware Facial Animation from a Single Image. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 818-833. https://doi.org/10.1007/978-3-030-01249-6_50
- [6] Chen, R.T., Rubanova, Y., Bettencourt, J., *et al.* (2018) Neural Ordinary Differential Equations. <https://doi.org/10.48550/arXiv.1806.07366>
- [7] Brock, A., Donahue, J. and Simonyan, K. (2018) Large Scale GAN Training for High Fidelity Natural Image Synthesis. arXiv preprint arXiv: 1809.11096. <https://doi.org/10.48550/arXiv.1809.11096>
- [8] Salzer, R. (2008) Practical Guide to Interpretive Near-Infrared Spectroscopy. Wiley Online Library. <https://doi.org/10.1201/9781420018318>
- [9] Xiang, B.R., Li, R., Wu, Y.J., *et al.* (1999) Applications of Near-Infrared Spectroscopy in Pharmaceutical Analysis. *Computers and Applied Chemistry*, **16**, 327-328.
- [10] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. and Bharath, A.A. (2018) Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, **35**, 53-65. <https://doi.org/10.1109/msp.2017.2765202>
- [11] Xie, Q., Dai, Z., Hovy, E., *et al.* (2020) Unsupervised Data Augmentation for Consistency Training. *Advances in Neural Information Processing Systems*, **33**, 6256-6268.
- [12] Arjovsky, M., Chintala, S. and Bottou, L. (2017) Wasserstein Generative Adversarial Networks. *PMLR*, **70**, 214-223.
- [13] Zhong, Z., Zheng, L., Zheng, Z., Li, S. and Yang, Y. (2018) Camera Style Adaptation for Person Re-identification. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5157-5166. <https://doi.org/10.1109/cvpr.2018.00541>
- [14] Yu, L., Zhang, W., Wang, J. and Yu, Y. (2017) Seqgan: Sequence Generative Adversarial Nets with Policy Gradient. *Proceedings of the AAAI Conference on Artificial Intelligence*, **31**. <https://doi.org/10.1609/aaai.v31i1.10804>
- [15] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H. (2018) Synthetic Data Augmentation Using GAN for Improved Liver Lesion Classification. 2018 *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, 4-7 April 2018, 289-293. <https://doi.org/10.1109/isbi.2018.8363576>
- [16] Teng, G.E., Wang, Q.Q., Kong, J.L., Dong, L.Q., Cui, X.T., Liu, W.W., *et al.* (2019) Extending the Spectral Database of Laser-Induced Breakdown Spectroscopy with Generative Adversarial Nets. *Optics Express*, **27**, Article 6958. <https://doi.org/10.1364/oe.27.006958>
- [17] Zhu, D., Xu, L., Chen, X., Yuan, L., Huang, G., Li, L., *et al.* (2020) Synthetic Spectra Generated by Boundary Equilibrium Generative Adversarial Networks and Their Applications with Consensus Algorithms. *Optics Express*, **28**, Article 17196. <https://doi.org/10.1364/oe.390070>
- [18] Makhzani, A., Shlens, J., Jaitly, N., *et al.* (2015) Adversarial Autoencoders. arXiv preprint arXiv: 1511.05644. <https://doi.org/10.48550/arXiv.1511.05644>
- [19] SEEGER, M. (2004) Gaussian Processes For Machine Learning. *International Journal of Neural Systems*, **14**, 69-106. <https://doi.org/10.1142/s0129065704001899>
- [20] Huang, G., Ruan, X., Chen, X., Lin, D. and Liu, W. (2014) A Segmented PLS Method Based on Genetic Algorithm. *Analytical Methods*, **6**, Article 2900. <https://doi.org/10.1039/c3ay41765d>
- [21] Nikzad-Langerodi, R., Zellinger, W., Lughofer, E. and Saminger-Platz, S. (2018) Domain-Invariant Partial-Least-Squares Regression. *Analytical Chemistry*, **90**, 6693-6701. <https://doi.org/10.1021/acs.analchem.8b00498>