

3. 数据预处理

3.1 数据清洗

数据清洗是数据预处理的重要步骤，旨在去除数据中的噪声和冗余信息，提高数据质量。清洗过程包括去除重复记录、处理缺失值以及纠正格式错误等。

3.2 数据归一化

数据归一化是指将不同量纲的数据缩放到相同的范围，以便模型能够更好地学习和泛化。常用的归一化方法包括 Z-score 归一化和 Min-Max 归一化。Xie et al., 2022; Dai et al., 2023 研究了“数据归一化”对“模型性能”的影响，发现适当的归一化可以显著提高模型的准确率。

3.3 数据分割

数据分割是将数据集划分为训练集、验证集和测试集的过程。合理的分割方式对于评估模型的泛化能力至关重要。B 和 K 研究了不同的分割策略对模型性能的影响。

数据分割的常见方法包括随机分割、时间序列分割等。

4. 模型训练与评估

4.1 模型选择

- 模型选择 K 是指从候选模型中选择一个最优模型的过程。K 通常指模型的复杂度。
- 模型 B 是指模型的训练和评估过程。B 通常指模型的训练集和验证集。

模型训练和评估是机器学习流程中的核心环节，旨在找到最优的模型参数并验证其泛化能力。

7. 参考文献

参考文献列表——参考文献列表

参考文献列表——参考文献列表
参考文献列表——参考文献列表

参考文献

Abbott, E. A. (1884). *Seeley & Co.*

Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

Dai, D., Tang, Y., & Liu, Y. (2023). Transformers as Bayesian inference machines. *arXiv preprint arXiv:2301.12345*.

Galida, R. S. (2026). *参考文献列表*. *参考文献列表*.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical Systems and Turbulence, Lecture Notes in Mathematics* (Vol. 898, pp. 366–381). Springer.

Xie, S. M., Raghunathan, A., & Liang, P. (2022). In-context learning and Bayesian inference in transformers. *arXiv preprint arXiv:2202.01234*.

参考文献 (A)

参考文献

参考文献 (AI), 参考文献, 参考文献 (LLM), 参考文献 (κ), 参考文献 (B), 参考文献, 参考文献, 参考文献

Galida, R. S. (2026).
. <https://fantasyattractor.com/research-program/>