arXiv:2512.17935v1 [cs.SD] 11 Dec 2025

chatter：用于将信息论和人工智能/机器学习模型应用于动物交流的 Python 库

Mason Youngblood

美国石溪大学高级计算科学研究所 masonyoungblood@gmail.com

软件

- GitHub
- 文档
- PyPI

已发布日期： 2025年12月10日

概要

动物交流研究通常涉及将交流单元归类为不同类型（例如，鸣禽的音节或座头鲸的音符）。这种方法在许多情况下都很有用，但必然会简化真实交流系统所具有的复杂性和细微差别。 `chatter` 是一款新型 Python 库，利用信息论和现代机器学习技术，在连续潜在空间中分析动物交流。它不受分类体系限制，并已通过鸟类、蝙蝠、鲸类和灵长类的发声进行测试。通过运用多种不同架构，包括变分自编码器和视觉变换器， `chatter` 将发声序列表示为高维潜空间中的轨迹，从而无需手动或自动对单元进行分类。该库提供端到端工作流程——从预处理和分段，到模型训练和特征提取——使研究人员能够量化发声序列的复杂性、可预测性、相似性和新颖性。

需求声明

近年来，动物行为学家开始使用机器学习，将发声投射到高维潜空间中（Goffinet et al., 2021; Merino Recalde, 2023; Sainburg et al., 2020）。然而，除少数例外（Alam et al., 2024）外，这些方法通常用于识别发声类别，而非将发声序列作为连续向量进行分析。 `chatter` 通过提供一个易于使用、模块化且易上手的界面来应用这些技术，填补了这一空白。它补充了现有工具（如 `pykanto` （Merino Recalde, 2023）和 `AVGN` （Sainburg et al., 2020）），这些工具侧重于离散分析，而它则提供了用于连续分析的并行工作流程。如此一来，它降低了研究人员的入门门槛，使那些对计算方法感兴趣、却可能没有时间或专业知识从零开始搭建机器学习流水线的人也能使用这些方法。该库的设计兼顾灵活性，既能满足高级用户的需求，也为新手提供合理的默认设置。

`chatter` 的核心工作流程分为三个步骤，全部通过配置字典进行管理。首先，用户初始化 `Analyzer` 类，并设置最低和最高频率、振幅阈值及去噪参数等。 `Analyzer` 负责处理原始音频，通过降噪、滤波和压缩进行预处理，并根据振幅或其他标准将音频分割为单独的声学单元。它还为处理鸟鸣数据的用户提供使用 BirdNET（Kahl 等，2021）识别目标物种的功能。随后，这些声学单元会转换为声谱图。

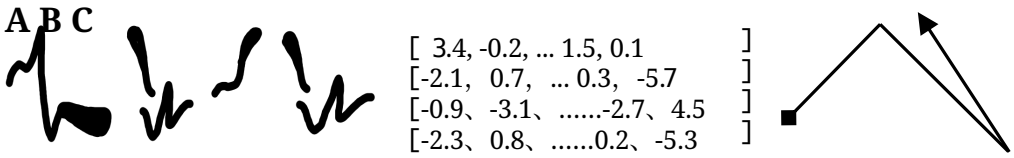


图 1： `chatter` 工作流程的基本示意图，展示了从声谱图到潜在特征，再到二维空间中的可视化结果的过程。请注意，所有信息论分析都在原始潜在空间中进行，而非降维后的二维空间。

接下来，使用 `Trainer` 类训练卷积变分自编码器，以学习发声的高维潜在表示（Goffinet 等，2021）。这种无监督方法能够以相对无损的方式描述复杂发声，使用户能够根据学习到的一组声学特征重建原始声音。随后，会提取所有声学单元的潜在特征。用户还可以选择将 DINOv3 视觉Transformer应用于声谱图，以提取潜在特征；当把序列作为处理单元时（即使用整个序列的声谱图，而非序列内各个单元的声谱图），这种方法效果尤佳（Siméoni 等，2025）。最后，该库提供工具，可使用 PaCMAP（Wang 等，2021）等降维技术将潜在空间可视化。PaCMAP 是 UMAP 的扩展，能更好地捕捉全局结构。图2展示了卡辛氏蚋莺音节的这一输出结果（录音取自 Hedley（2016））。

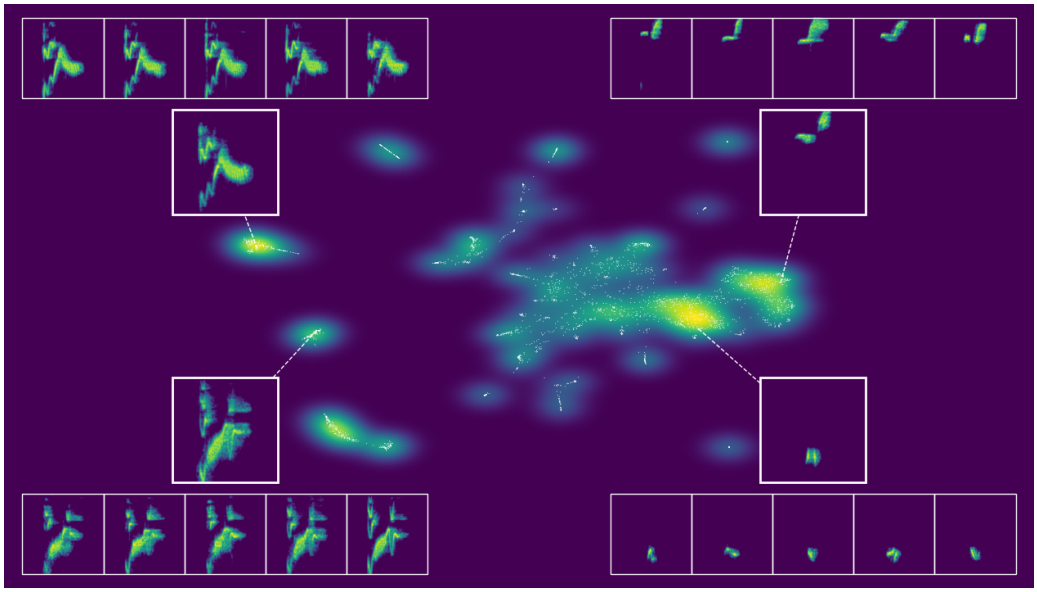


图2： 卡辛氏蚋莺音节的潜在空间。该图展示了这些音节在二维潜在空间中的分布；该潜在空间是将 PaCMAP 应用于变分自编码器生成的潜在特征而得到的，并叠加了具有代表性的声谱图。

最后，`FeatureProcessor` 类提供了多种用于下游分析的函数。复杂度通过计算潜在空间中连续声学单元之间的路径长度（余弦距离）并按时长归一化来量化，从而连续衡量声学多样性随时间的变化（Alam et al., 2024）；这与先前基于声谱图互相关的方法类似（Sawant et al., 2022）。为评估可预测性，`chatter` 会基于潜在特征序列训练向量自回归模型（Sims, 1980）。该模型能够学习时间依赖关系，并量化未来状态的概率。

chatter 还使用 denmarf估计声学单元的常见程度和稀有程度。 denmarf是一个密度估计库，利用掩码自回归流计算高维空间中的密度（Papamakarios 等，2017）。这能够衡量给定发声相对于学习到的分布有多常见或多罕见。最后，通过对潜在向量序列应用动态时间规整，计算整个单元序列之间的相似度（即图 1 中的 B 面板）。总之，这些指标能够相当复杂且细致地描述动物发声序列。

参考文献

Alam, D., Zia, F., & Roberts, T. F. (2024). 雄性斑胸草雀求偶鸣唱中隐藏的适应度。《自然》，627，851–857。https://doi.org/10.1038/s41586-024-07207-4 Goffinet, J., Brudner, S., Mooney, R., & Pearson, J. (2021). 低维学习到的特征空间量化发声曲目中的个体与群体差异。eLife，10，e67855。https://doi.org/10.7554/eLife.67855 Hedley, R. W. (2016). 卡辛氏蚋莺鸣唱中句法的复杂性、可预测性与时间齐次性。PLOS ONE，11(3)，e0150822。https://doi.org/10.1371/journal. pone.0150822 Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET：用于鸟类多样性监测的深度学习解决方案。《生态信息学》，61，101236。https://doi.org/10. 1016/j.ecoinf.2021.101236 Merino Recalde, N. (2023). Pykanto：加速野生鸟类鸣唱研究的 Python 库。《生态学与进化方法》，14(6)，1639–1648。https://doi.org/10.1111/ 2041-210X.14155 Papamakarios, G., Pavlakou, T., & Murray, I. (2017). 用于密度估计的掩码自回归流。载于 I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fer- gus, S. Vishwanathan, & R. Garnett（编），《神经信息处理系统进展》（第30卷）。https://proceedings.neurips.cc/paper_ files/paper/2017/file/ 6c1da886822c67822bcf3679d04369fa-Paper.pdf Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). 在多样化的动物发声曲目中发现、可视化并量化潜在结构。PLOS 计算生物学，16(10)，e1008228。https://doi.org/10.1371/journal.pcbi.1008228 Sawant, S., Arvind, C., Joshi, V., & Robin, V. V. (2022). 声谱图互相关可用于衡量鸟类发声的复杂性。《生态学与进化方法》，13(2)，459–472。https://doi.org/10.1111/2041-210X.13765 Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., ... Bojanowski, P. (2025). DINOv3。https://arxiv.org/abs/2508.10104 Sims, C. A. (1980). 宏观经济学与现实。《计量经济学》，48(1)，1–48。https: //doi.org/10.2307/1912017 Wang, Y., Huang, H., Rudin, C., & Shaposhnik, Y. (2021). 理解降维工具的工作原理：一种解读数据可视化中 t-SNE、UMAP、TriMap 和 PaCMAP 的实证方法。《机器学习研究杂志》，22(201)，1–73。http://jmlr.org/papers/v22/20-1061.html