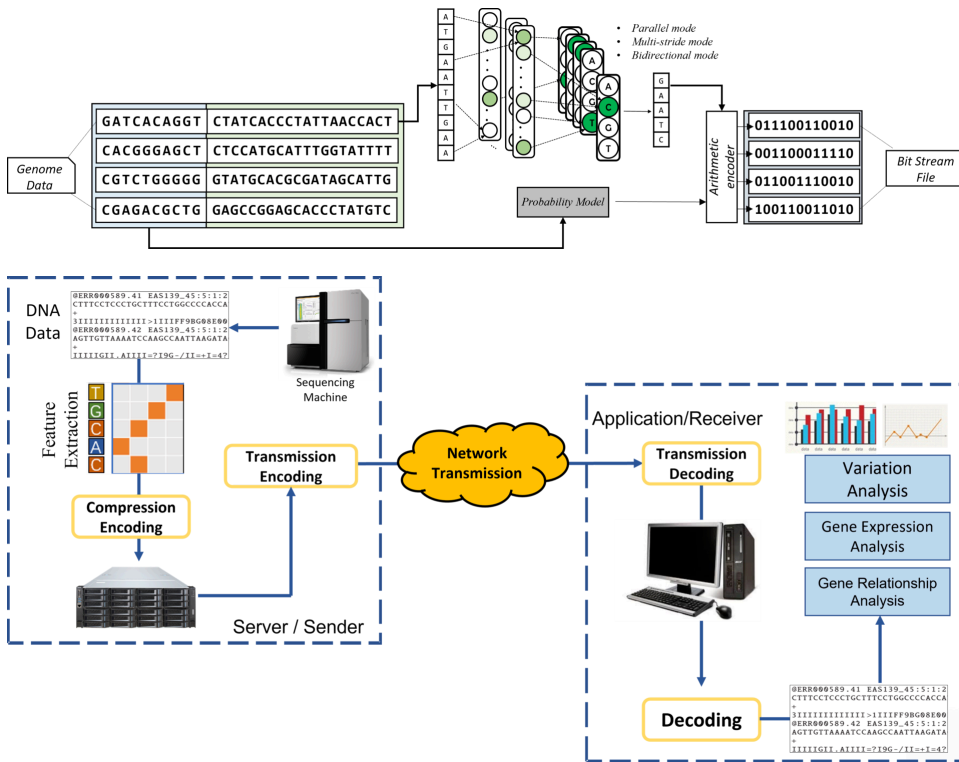


基于学习的基因组编解码器

通信和信息

生物医学与基因工程/化工产品
计算机/人工智能/数据处理和信息技术



Remarks

第 48 届日内瓦国际发明展 (IEIG) (2023) - 银奖

IP 状态

专利已存档



技术成熟度等级 (TRL) ?

6

发明人

邝得互教授
王诗淇教授
王萌
孙圳昊

机会

基因测序技术的快速发展从根本上改变了人类研究生命蓝图的方式，并促进了基因组学、生物信息学、系统生物学、合成生物学等学科的建立和发展。随着基因测序技术的不断完善，测序成本持续下降。然而，测序数据增长速度远远超过了摩尔定律，使得基因组数据压缩和传输成为生物信息产业上游的基础关键技术。依赖于人工智能，本发明专注于通过机器学习和上下文建模实现高效率的基因组压缩。原型软件在测试数据上能够实现超过100倍的压缩比，具有个性化、低复杂度和高压缩效率的特点，有效降低了基因组数据的存储和传输成本。

技术

在本发明中，我们使用机器学习技术从同种生物的大量遗传数据中学习潜在的数据模式。利用这些模型，我们可以高效地压缩同种生物的基因组数据。此外，为了提高压缩速度，我们创新性地引入了三种预测方法，包括并行预测、多步长预测和双向预测。原型软件的实验结果表明，这些预测模型能够在压缩过程中有效降低机器学习模型的计算复杂度。



优势

- **高效**：我们的发明实现了超过100倍的压缩比，远高于传统编码器（例如4倍）。
- **快速**：与没有编解码器优化的方法相比，我们的技术提高了处理速度（50倍加速），为多种应用提供了便利。
- **智能**：通过紧凑的表示，我们可以探索不同物种之间的关系，为生物学研究提供基础支持。

应用

- 遗传云数据库
- 远程医疗
- 遗传诊断

