

A review of deep learning applications in human genomics using next-generation sequencing data

生信 2001 张子栋 2020317210101

这篇文章主要介绍了深度学习在人类基因组学中的应用，特别是利用下一代测序（NGS）数据的应用。文章介绍了深度学习的基本概念和原理，以及它与其他机器学习方法的区别和联系。还概括了深度学习在不同子领域的人类基因组学中的发展和应用，例如疾病变异、表观基因组学、基因表达和药物基因组学等。深度学习在人类基因组学中的优势和挑战，例如能够自动提取复杂、高维、非线性的基因组数据的特征，但也面临着数据质量、可解释性、计算资源等问题。文章研究了深度学习在不同子领域的人类基因组学中的发展和应用，例如疾病变异、表观基因组学、基因表达和药物基因组学等；深度学习在人类基因组学中的优势和挑战，例如能够自动提取复杂、高维、非线性的基因组数据的特征，但也面临着数据质量、可解释性、计算资源等问题；深度学习在人类基因组学中的最新应用，例如利用深度学习预测基因功能、调控元件、蛋白质结构和相互作用等。

文章的目的是为了让生物信息学或基因组学的读者了解深度学习在人类基因组学中的重要性和前沿性，以及如何使用深度学习方法来分析人类基因组数据。文章也提出了一些未来生物信息学研究方向的启发，例如提高数据标准化、增加模型可解释性、加强跨领域合作等。

此外文章还对近年来发表在人类基因组学领域的深度学习相关的文献进行了系统的梳理和分析，总结了深度学习在人类基因组学中的主要方法和模型，例如卷积神经网络（CNN）、循环神经网络（RNN）、变分自编码器（VAE）、生成对抗网络（GAN）等。深度学习是一种强大而有前景的技术，在人类基因组学中有广泛而重要的应用。文章也指出了深度学习在人类基因组学中还不存在的不足和需要改进的地方，例如提高数据标准化、增加模型可解释性、加强跨领域合作等。

在目前深度学习在基因组学的应用中，仍然存在以下难点：

- 如何处理和分析大规模、异质、噪声、缺失等特点的基因组数据。
- 如何设计和优化适合不同基因组任务和数据类型的深度学习模型和算法。
- 如何评估和验证深度学习模型在人类基因组学中的性能和效果。
- 如何解释和理解深度学习模型在人类基因组学中的工作机制和生物意义。

总体来看，深度学习是一种强大而有前景的技术，在人类基因组学中有广泛而重要的应用，未来生物信息学研究应该充分利用深度学习的优势，开发更多更好的深度学习模型和工具，解决更多更难的基因组问题。同时，深度学习在人类基因组学中还存在一些不足和需要改进的地方，未来生物信息学研究应该关注深度学习的局限性和挑战，提高数据标准化、增加模型可解释性、加强跨领域合作等，提高深度学习在人类基因组学中的可靠性和有效性。深度学习在人类基因组学中有巨大的潜力和影响，利用深度学习预测基因功能、调控元件、蛋白质结构和相互作用等，促进基因组数据的共享和利用，提高基因组医疗的效率和精准度。