



基于 GPU 加速随机森林算法的大豆基因型填充研究

李明亮^{1,2}, 李卓², 黄斌², 于军¹, 辛鹏¹, 张继成³, 唐友^{1,2}

(1. 吉林化工学院 信息与控制工程学院, 吉林 吉林 132022; 2. 吉林农业科技学院 电气与信息工程学院, 吉林 吉林 132101; 3. 东北农业大学 电气与信息工程学院, 黑龙江 哈尔滨 150030)

摘要: 基因型填充 (Genotype Imputation, GI) 是一种利用现有的基因型信息来推断未测定或不完整基因型的技术。为了探索在大豆基因组测序中处理不完整数据的高效填充方法, 以提高数据处理速度和效率, 本研究采用真实的大豆参考面板基因型数据, 通过对数据进行 2%、5%、10% 和 25% 的完全随机缺失处理, 利用 GPU 加速的随机森林机器学习算法构建填充模型, 并对不同缺失比例的数据进行填充。同时, 对比分析了不同处理器的准确性和性能。结果显示: 基于 GPU 加速的随机森林算法在大豆基因组中实现了优秀的填充精度。与主流基因填充软件相比, 该方法至少提供了 4 倍以上的运算时间优势。因此, GPU 加速的基因型填充策略可应用于大规模基因型数据处理中, 提高大豆基因型数据处理的速度和效率, 同时减少计算时间和资源消耗。

关键词: 大豆基因填充; 随机森林算法; GPU 加速; 数据处理

Research on Soybean Genotype Imputation Based on GPU-Accelerated Random Forest Algorithm

LI Mingliang^{1,2}, LI Zhuo², HUANG Bin², YU Jun¹, XIN Peng¹, ZHANG Jicheng³, TANG You^{1,2}

(1. School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China; 2. Electrical and Information Engineering College, Jilin Agricultural Science and Technology University, Jilin 132101, China; 3. College of Electronic and Information, Northeast Agricultural University, Harbin 150030, China)

Abstract: Genotype Imputation (GI) is a technique that uses existing genotype information to infer unobserved or incomplete genotypes. This study aims to explore efficient imputation methods for handling incomplete data in soybean genomic sequencing, with the goal of improving data processing speed and efficiency. Real soybean reference panel genotype data was used in the study, and the data was subjected to complete random missingness at rates of 2%, 5%, 10%, and 25%. A GPU-accelerated random forest machine learning algorithm was employed to construct imputation models and fill in the missing data at different missingness rates. Additionally, the accuracy and performance of different processors were compared and analyzed. The research results demonstrate that the GPU-accelerated random forest algorithm achieves excellent imputation accuracy in the soybean genome. Compared to mainstream genotype imputation software, this method provides at least a fourfold computational time advantage. Therefore, the GPU-accelerated genotype imputation strategy can be applied to large-scale genotype data processing, improving the speed and efficiency of soybean genotype data processing while reducing computational time and resource consumption.

Keywords: soybean genotype imputation; random forest algorithm; GPU acceleration; data processing

大豆作为重要的粮食和油料作物, 在全球范围内广泛种植, 具有较高的经济价值。了解大豆的基因型对于改良和优化品种、提高农作物产量和抗性, 以及增强大豆的适应性和营养价值具有重要意义。基因型填充 (Genotype Imputation, GI) 是指通过基因分型技术, 对大豆个体的基因组进行高密度基因型的预测和填充, 以填补基因组中的缺失数据^[1]。通过基因型填充, 我们可以更好地理解大豆的遗传特性^[2], 推动相关研究和应用, 促进大豆品种改良和农业可持续发展。

大豆是一种遗传多样性相对较低的自交系作物, 特别是在双亲本衍生群体中, 相关单倍型的延伸较长, 重复基因型所占比重较大^[3]。总的来说, 大豆基因型的特点包括遗传多样性、自交繁殖、基因连锁不平衡、基因副本和基因表达调控等, 选择合适的算法和模型来解决大豆基因型填充问题是一个挑战。需要考虑算法的准确性、效率和可扩展性, 并结合大豆遗传学和基因组学领域的知识进行优化和改进。大豆基因型填充可以采用多种不同的基于统计学和概率模型的填充方法, 如基于 k 近

收稿日期: 2023-05-05

基金项目: 吉林省科技发展规划项目 (YDZJ202201ZYTS692)。

第一作者: 李明亮 (1999—), 男, 硕士研究生, 主要从事计算机大数据研究。E-mail: brilliant_lml@163.com。

通讯作者: 唐友 (1979—), 男, 博士, 教授, 高级工程师, 主要从事生物信息学及农业信息化研究。E-mail: tangyou@jlnku.edu.cn。

邻插值法、隐马尔可夫模型^[4]和卷积神经网络^[5]等。这些方法在大豆基因型填充中都表现出良好的效果,特别是基于随机森林模型的方法,其精度更具优势^[6]。

标准的基因型填充方法使用隐马尔科夫模型(HMM)和大型全基因组测序参考数据集。这些方法通过利用待填充基因组与参考数据集之间共享的遗传变异,使用 HMM 填充每个样本中的缺失基因型。HMM 中的隐藏状态表示与待填充单倍型最相关的参考数据集中的单倍型。HMM 参数估计还依赖于重组率、突变率和基因型错误率,需要使用马尔可夫链蒙特卡洛算法(MCMC)或最大期望化算法(EM)进行拟合。因此,基于 HMM 的填充是一种计算密集型过程,需要同时访问高性能计算环境和大型、涉及隐私的全基因组测序参考数据集。

随机森林算法(Random Forest, RF)是一种基于集成学习的决策树算法,由机器学习领域的统计学家 Leo Breiman 和 Adele Cutler 于 2001 年提出,基于 CART 算法、bagging 策略和随机特征选择的思想进行改良和优化^[7]。随机森林算法的主要特点是随机抽样和随机属性选择,其基本思想是将样本数据随机地分成多个小集合,并在每个小集合上构建一个决策树,然后将这些决策树结合起来进行预测。由于随机森林算法所构建的每一个树结构相对简洁,可以通过观察树的形态和关键的分支节点,来解释预测结果。随机森林模型采用了多个决策树的集成预测,可以避免单个决策树的过拟合和不足,从而提高了模型的准确性^[8]。在决策树的构建过程中,采用基尼指数或信息增益等指标进行节点的分裂,以使决策树的生长具有更好的泛化性能^[9]。随机森林算法的并行度高,可以快速处理大量数据,并且可以利用 GPU 等硬件进行加速^[10]。

GPU (Graphics Processing Unit)加速技术是指利用图形处理器或其他可编程的硬件来加速计算,具有高并行性、高效性、低成本等优势。GPU 最初是为了图形渲染而设计的,拥有数百个甚至数千个并行的处理单元,可以同时处理大量的数据计算^[11]。在计算科学和机器学习等领域中,GPU 也可以用来加速矩阵计算、深度神经网络、布尔运算和图像处理等任务。

在基因型填充和其他大规模数据处理中,GPU 加速技术可以提高数据处理的速度和效率,同时减少计算时间和资源消耗^[12]。GPU 加速技术可以利用现有的显卡硬件来并行处理大量的深度学习计算,同时降低计算成本和系统复杂度。而对于常规 CPU (Central Processing Unit)运算,其最大的优势是可以一次性地快速执行较小的任务。相比之下,GPU 加速要求问题可以分解为更小的子问题,每个子问题都能并行操作^[13]。这使得 GPU 加速非常适合于大规模的计算任务,尤其是涉及到庞大的数据集时。总体来讲,GPU 加速技术能够帮助在处理大规模数据和进行复杂计算的时候提高计算效率,减少计算成本和时间。本研究以真实大豆参考面板基因型数据为基础,设置完全随机缺失处理,利用 GPU 加速随机森林算法进行基因型填充,旨在获得优异的填充策略,提高大豆基因型数据处理的速度和效率。

1 材料与方法

1.1 材料

本研究采用的大豆基因组数据来源为 SoyBase (<https://soybase.org>),试验数据主要包括大豆的 1 007 个个体和 70 310 个单核苷酸多态性(Single Nucleotide Polymorphism, SNPs)。运用遗传算法和 plink 软件的模拟方法,基于真实数据生成了 1 个包含 1 000 个个体和 10 000 个 SNPs 的模拟基因数据集。

1.2 试验设计

试验硬件环境为两颗 Intel 至强银牌 4110 CPU 中央处理器,基于 X64 架构,2816 最大线程;64 G 内存;NVIDIA GeForce RTX 3060 12G GPU 显卡,3 584 个流处理单元。使用 simFrame 对模拟数据随机缺失,缺失方式为完全随机,缺失比例分别为 2%、5%、10% 和 25%。分别使用 CPU 处理器和 GPU 处理器进行基因数据模型训练和填充。采用随机森林算法使用参考面板数据训练填充模型,再使用训练好的填充模型对不同比例缺失文件进行填充,并比较不同处理器在内存、时间和准确率方面的表现。并将论文提出的 GPU 加速随机森林算法与主流基因填充算法进行效果对比。具体试验流程如图 1 所示。

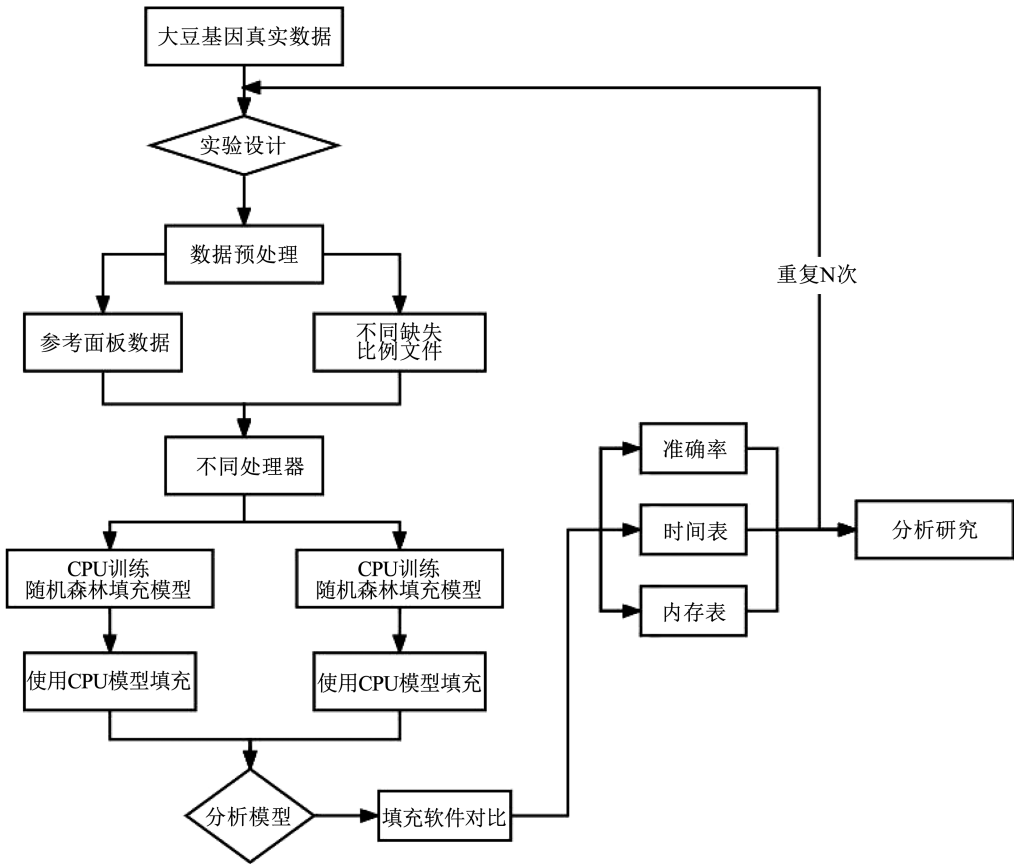


图1 试验设计流程图

Fig. 1 Experimental design flow chart

1.3 方法

1.3.1 将基因型数据转为 012 格式 将 VCF 文件中的基因型转换为 012 规则^[14],是为了便于在基因型数据中进行数值计算和统计分析。具体地说,在 VCF 文件中,010 表示两个等位基因都是参考基因型,将其转换为数字 0 是比较直观的。对于 011 或 110,这种基因型只有 1 个等位基因是替换基因型,因此可以将其转换为数字“1”,表示有 1 个替换基因型。对于 111,这种基因型表示两个等位基因都是替换基因型,因此可以将其转换为数字“2”,表示有两个替换基因型。对于缺失基因型 .1. 用“-1”表示。这样就完成了基因型到数字的转换,将基因型的信息简化为更少的状态,方便进行分析和比较。

1.3.2 随机森林填充算法分析 随机森林算法首先对参考面板文件进行有放回抽样得到 k 个训练子集 $D = \{D_1, D_2, \dots, D_k\}$ 。 k 个训练子集训练得到 k 棵决策树 $DT = \{DT_1, DT_2, \dots, DT_k\}$,训练得到随机森林模型。在每个决策树的构建过程中,采用更加高效的基于信息增益排序的特征选择策略,通过

计算特征在样本分类中的不确定性来减少评估特征的重要性,选择具有区分能力的特征,从而减少特征计算的开销。训练好的随机森林模型文件包含以下内容:决策树的数量 ($n_estimators$) 和每棵树的最大深度 (max_depth) 等超参数信息。决策树的信息 (比如节点、分裂标准、特征的划分点等) 等决策树的结构信息。数据处理的信息 (比如缺失值填充、类别变量的编码等)。所有训练过的特征 (特征名称和特征重要性等)。函数和模块的导入信息 (用于加载模型时导入需要的函数和模块)。再使用模型对不同缺失比例文件进行填充,采用多数投票法进行投票表决,得票最多的分类结果,作为最终分类结果输出 $Y = argmax_i \sum_{i=1}^k f_i(x)$,其中 DT 表示决策树集合, Y 表示输出变量。

1.3.3 GPU 加速技术 采用数据布局优化技术,在 GPU 上进行计算时,数据的访问模式对性能有很大影响,对数据进行内存对齐,使得每个线程块可以高效地访问连续的内存区域,最大限度地减少数

据访问延迟并提高内存带宽利用率。

在构建随机森林时,对每个决策树进行剪枝,以减少决策树的规模和复杂度,在传统的剪枝过程中,需要计算每个节点的指标,如损失函数或不纯度指标。对于每个节点,需要计算其左右子节点的指标,并与合并后的父节点指标进行比较。这些指标的计算可以通过并行化在 GPU 上进行。将每个节点的计算分配给不同的 GPU 核心,可以同时计算多个节点的指标,从而加快剪枝过程。

随机森林预测过程涉及多个决策树的集成,传统的方法是对每个决策树的结果进行简单的投票或平均,但这可能在 GPU 上的并行执行中引入数据依赖。所以采用更加高效的并行投票方法,使用原子操作确保多个线程或核心访问共享变量时的原子性和一致性,以充分利用 GPU 的并行计算能力。

1.3.4 验证填充数据的准确性 通过填充后的基因型数据与完整数据进行比对,利用均方误差 MSE (Mean Square Error)和决定系数 R^2 (R-Squared)判断准确性。MSE 具体表示为: $MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$,其中 y_i 表示准确值也就是真实值, \hat{y}_i 表示预测估计值, m 表示缺失数据的个数。MSE 越小,估计值与真实值之间越接近,算法填充准确率

越高。 R^2 具体表示为: $R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2}$,其

中, y 表示真实值, \hat{y} 表示预测估计值, \bar{y} 表示所有真实值的均值。系数 R^2 的取值范围再 0 到 1 之间, R^2 越高,表示预测偏差越小,说明准确性就越高。

1.3.5 填充性能分析 使用不同处理器对不同缺失比例的数据进行运算,并测量其所需的时间和内存消耗。可以利用 Python 中的 time 和 psutil 模块,调用 time.clock() 方法计算程序的运行时间,调用 psutil.process() 方法获取运行程序所占用的内存值。

2 结果与分析

2.1 不同处理器准确率对比

经过 n 次重复填充实验,由均方误差(表 1)和决定系数(表 2)的分析结果显示,在相同缺失比例

情况下,使用 GPU 填充算法精确性略高于使用 CPU 填充,并且随着缺失比例由 2%、5%、10%、25% 逐渐升高,两种处理器填充精准性也均逐步下降。

表 1 不同处理器在不同缺失比例下的均方误差对比

Table 1 Comparison of MSE of different processors at different missing rates

缺失比例 Deletion ratio	CPU	GPU
2%	0.0129	0.0124
5%	0.0163	0.0147
10%	0.0379	0.0340
25%	0.0658	0.0580

表 2 不同处理器在不同缺失比例下的决定系数对比

Table 2 Comparison of R-squared different processors at different missing rates

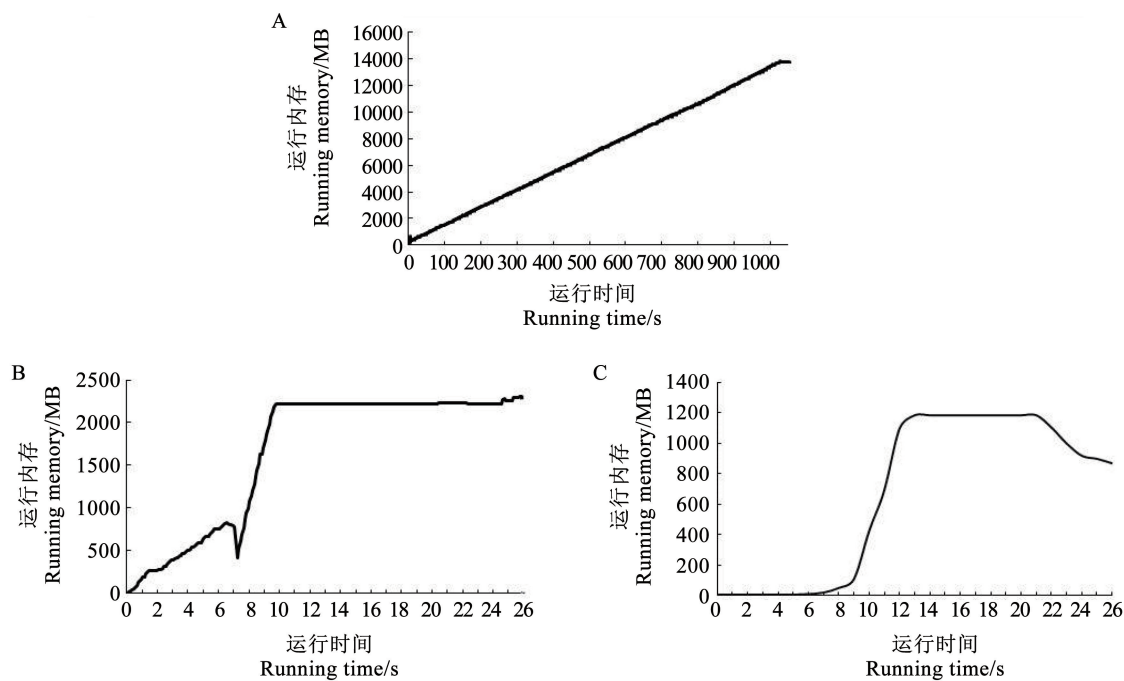
缺失比例 Deletion ratio	CPU	GPU
2%	0.9792	0.9797
5%	0.9674	0.9685
10%	0.9283	0.9292
25%	0.8798	0.8813

2.2 不同处理器性能对比

如图 2 所示,在随机森林填充模型训练部分,使用 CPU 运行总耗时 1 054.1 s,计算机内存为 1 028.6 s 时达到最大值 13 713.77 Mb;使用 GPU 运行总耗时 26.2 s,计算机内存存在 25.8 s 时达到最大值 2 301.96 Mb,显卡内存存在 20.8 s 时达到最大值 1 087.2 Mb。

如图 3 所示,在缺失值填充部分,在时间维度上,随着缺失比例的逐渐增加,CPU 运行时间呈线性增加,GPU 运行时间基本不变。GPU 运行速度相对于 CPU 缺失比例越大提升效果越明显,2% 缺失比例时速度提升约 50 倍,至 25% 缺失比例时速度提升约 500 倍。

如图 4 所示,在内存维度上,随着缺失比例的逐渐增加,CPU 与 GPU 运行内存均基本不变。使用 GPU 填充时计算机内存与显卡内存总和比使用 CPU 填充时计算机内存减少了 70%。



注：A. 使用 CPU 进行模型训练计算机运行内存; B. 使用 GPU 进行模型训练计算机运行内存; C. 使用 GPU 进行模型训练显卡运行内存。

Note: A. Using CPU for model training requires computer memory; B. Using GPU for model training requires computer memory; C. Using GPU for model training requires graphics card memory.

图 2 不同处理器进行模型训练运行内存随时间变化图

Fig.2 Memory usage over time during model training on different processors

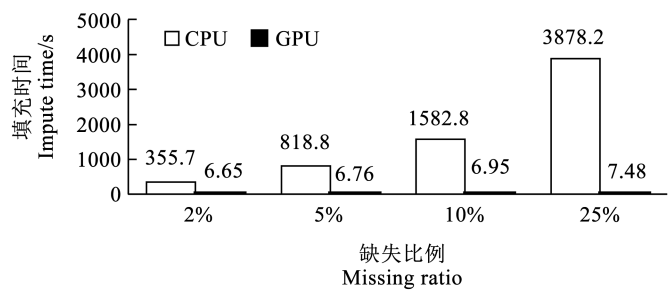


图 3 不同处理器在不同缺失比例下运行时间对比

Fig.3 Time comparison of different processors under different missing ratios

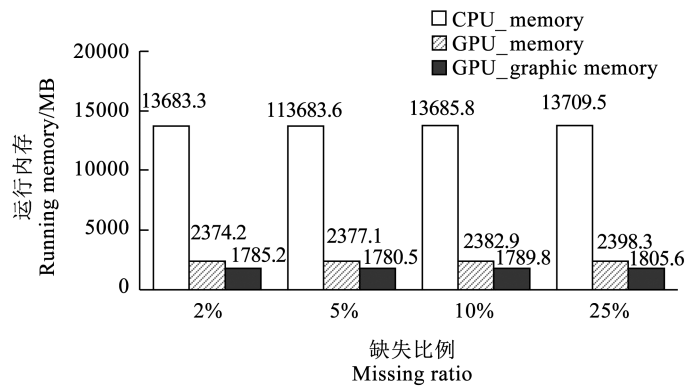


图 4 不同处理器在不同缺失比例下运行内存对比

Fig.4 Memory comparison of different processors under different missing ratios

2.3 填充软件性能对比

通过对多次重复实验取平均值,将 GPU 加速随机森林算法的填充方法与主流填充软件^[15] Beagle5^[16-17] 和 Minimac4^[18] 进行比较实验的结果如表 3 所示。在填充准确率方面的结果显示:当基因数据缺失比例低于 5% 时,算法的填充效果优于 Beagle5 和 Minimac4;当基因数据缺失比例为 10% 时,算法的准确率略优于 Beagle5,但低于 Minimac4;

而当基因数据缺失比例为 25% 时,算法的准确率则低于 Beagle5 和 Minimac4。在填充时间方面的结果显示:算法在模型训练阶段需要 26.2 s 生成填充模型,并将其作为参考模型进行填充操作。对于相同的参考面板文件,填充模型可重复使用,而且算法所需的填充时间均少于 Beagle5 和 Minimac4 算法,特别是在低缺失比例的情况下,效果更加明显。

表 3 GPU 加速随机森林算法与主流填充软件对比

Table 3 Comparison of GPU accelerated random forest algorithm and mainstream imputation software

项目 Item	缺失比例 Deletion ratio	GPU	Beagle5	Minimac4
R^2	2%	0.9797	0.9471	0.9685
	5%	0.9685	0.9407	0.9673
	10%	0.9292	0.9253	0.9630
	25%	0.8813	0.8906	0.9502
时间 Time/s	2%	6.65	24.10	14.65
	5%	6.76	27.98	17.15
	10%	6.95	29.77	22.42
	25%	7.48	31.42	39.17

3 讨论

基于人工智能的数据挖掘技术正在革新生物信息学和分析领域,本研究展示了 GPU 加速随机森林算法在生物遗传学中执行基因型填充这一基本分析任务的潜力,在计算效率高且可移植的框架中产生优越结果。经过训练的随机森林模型可以轻松转移,并快速执行其功能,即使在较小的计算环境中,也不需要将私有基因型数据转移到外部填充服务器中。该模型和算法可以基于大豆遗传基因的特点,结合单倍型和基因频率等信息,提高基因型填充的精确度和效率。

随机森林填充方法对填充大豆基因组数据准确性较高^[19],然而,不同的大豆参考群体可能会导致模型训练参数的调整问题。在使用随机森林算法进行基因型数据分析时,需要首先训练一个模型,在训练模型时需要设定一些参数,如树的数量、每棵树的最大深度、节点分裂的方式^[20]等。而这些参数的设定通常依赖于基因型数据的特征,并且不同参考面板群体可能具有不同的特征。所以,选择合适的参考面板群体对于随机森林算法是至关重

要的,它直接影响到模型训练的效果和结果的可信度。在选择参考面板群体时,需要综合考虑多个因素,包括参考面板群体基因数据的质量、对研究对象的适用性以及模型训练参数的设定等问题。这种方法在训练过程中需要大量的计算资源,但一旦完成训练,它能够以快速且准确的进行填充,并且无需进一步访问参考面板数据。

GPU 加速基因型填充技术采用了 GPU 计算技术进行并行计算,可以加速基因型数据填充的过程。GPU 在大规模并行计算中具有很高的性能和效率,可以大大缩短数据处理时间。因此,GPU 加速基因型填充技术可以有效地提高大豆基因组学数据分析的效率和质量,并且可以应用于大规模基因型数据处理中,提高基因组关联研究的能力,使基因数据分析的覆盖范围更加完整,并通过精细映射改善对大豆病因变异的鉴定。

4 结论

本研究以大豆基因型数据作为研究对象,利用 GPU 加速技术实现了随机森林算法在模型训练和缺失数据填充方面的加速。研究结果表明,GPU 加

速技术显著提高了数据处理的速度和效率。此外,与生物信息学领域主流的基因型填充软件相比,在保证一定填充质量的前提下,本方法能够提高填充速度。一旦训练完成,基于本方法的填充过程快速而准确,不需要进一步访问参考面板数据的特点,因此有望成为现有方法的一种可行的替代方案。本研究提出的基于 GPU 加速的大豆基因型填充技术为加速大豆育种进程、研究大豆基因表达调控以及推动大豆基因组学研究的进展提供了更好的基础,并对其他农业基因型数据的处理和分析具有一定的参考和借鉴价值。

参考文献

[1] LI Y, WILLER C, SANNA S, et al. Genotype imputation[J]. Annual Review of Genomics and Human Genetics, 2009, 10: 387-406.

[2] MARCHINI J, HOWIE B. Genotype imputation for genome-wide association studies[J]. Nature Reviews Genetics, 2010,11(7): 499-511.

[3] CHEN L F, YANG S P, ARAYA S, et al. Genotype imputation for soybean nested association mapping population to improve precision of QTL detection[J]. Theoretical and Applied Genetics, 2022, 135(5):1797-1810.

[4] XAVIER A, MUIR W M, RAINEY K M. Impact of imputation methods on the amount of genetic variation captured by a single-nucleotide polymorphism panel in soybeans [J]. BMC Bioinformatics, 2016, 17(1): 1-9.

[5] LIU Y, WANG D L, HE F, et al. Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean[J]. Frontiers in Genetics, 2019, 10: 1091.

[6] JARQUÍN D, KOCAK K, POSADAS L, et al. Genotyping by sequencing for genomic prediction in a soybean breeding population[J]. BMC Genomics, 2014, 15: 740.

[7] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[8] 张晓琴,程誉莹. 基于随机森林模型的成分数据缺失值填补法[J]. 应用概率统计, 2017, 33(1): 102-110. (ZHANG X Q, CHENG Y Y. Imputation of missing values for compositional data based on random forest [J]. Chinese Journal of Applied Probability and Statistics, 2017, 33(1): 102-110.)

[9] 孟杰,李春林. 基于随机森林模型的分类数据缺失值插补[J]. 统计与信息论坛, 2014, 29(9): 86-90. (MENG J, LI C L. Missing data imputation for categorical data based on random forest model[J]. Statistics & Information Forum, 2014, 29(9):

86-90.)

[10] SENAGI K, JOUANDEAU N. Parallel construction of random forest on GPU[J]. The Journal of Supercomputing, 2022, 78: 10480-10500.

[11] 张军阳,王慧丽,郭阳,等. 深度学习相关研究综述[J]. 计算机应用研究, 2018, 35(7): 1921-1928, 1936. (ZHANG J Y, WANG H L, GUO Y, et al. Comprehensive review of deep learning research[J]. Computer Application Research, 2018, 35(7): 1921-1928, 1936.)

[12] 毛海宇,舒继武,李飞,等. 内存计算研究进展[J]. 中国科学: 信息科学, 2021, 51(2): 173-205. (MAO H Y, SHU J W, LI F, et al. Research progress of in-memory computing[J]. Scientia Sinica Informationis, 2021, 51(2): 173-205.)

[13] BRODTKORB A R, HAGEN T R, SÆTRA M L. Graphics processing unit (GPU) programming strategies and trends in GPU computing[J]. Journal of Parallel and Distributed Computing, 2013, 73(1): 4-13.

[14] 唐友. 基于全基因组测序的表型预测方法研究及其体系构建 [D]. 哈尔滨: 东北农业大学, 2017. (TANG Y. Research on phenotype prediction method based on whole genome sequencing and its system construction [D]. Harbin: Northeast Agricultural University, 2017.)

[15] ADRIANO M D, AMR A M, MADHUCHANDA B, et al. A comparative analysis of current phasing and imputation software [J]. PLoS One, 2022, 17(10): E0260177.

[16] BROWNING B L, TIAN X W, ZHOU Y, et al. Fast two-stage phasing of large-scale sequence data[J]. American Journal of Human Genetics, 2021, 108(10): 1880-1890.

[17] BROWNING B L, ZHOU Y, BROWNING S R. A one-penny imputed genome from next generation reference panels [J]. American Journal of Human Genetics, 2018, 103(3): 338-348.

[18] DAS S, FORER L, SCHÖNHERR S, et al. Next-generation genotype imputation service and methods[J]. Nature Genetics, 2016, 48(10): 1284-1287.

[19] 于合龙,刘雨帆,唐友,等. 基于多种机器学习方法填补大豆基因组缺失的比较研究[J]. 大豆科学, 2021, 40(1): 122-129. (YU H L, LIU Y F, ZHANG J C, et al. Comparative study on filling missing soybean genome based on multiple machine learning methods [J]. Soybean Science, 2021, 40(1): 122-129.)

[20] 温博文,董文瀚,解武杰,等. 基于改进网格搜索算法的随机森林参数优化[J]. 计算机工程与应用, 2018, 54(10): 154-157. (WEN B W, DONG W H, XIE W J, et al. Random forest parameter optimization based on improved grid search algorithm [J]. Computer Engineering and Applications, 2018, 54(10): 154-157.)