大规模线性逆问题的分裂-融合-共识算法及在数据价值计算中的应用

综述报告

唐厚钧

2024 年 9 月 5 日

1. 课题背景
2. 研究现状
3. 研究内容
4. 研究思路
5. 参考文献

1 课题背景
2 研究现状
3 研究内容
4 研究思路
5 参考文献

- 逆问题的定义：通过间接观测数据推断未知参数的科学研究方法。
- 重要性：应用于多个领域，如石油勘探、医学成像等。

1 课题背景

2 研究现状

3 研究内容

4 研究思路

5 参考文献

1. 当前研究进展：近年来，稀疏优化和压缩感知等领域的研究取得了显著成果。

1. 当前研究进展：近年来，稀疏优化和压缩感知等领域的研究取得了显著成果。
2. 算法瓶颈：然而，传统算法在面对大规模线性逆问题时，依然存在局限性，如计算效率低下和解决精度不足。

1. 当前研究进展：近年来，稀疏优化和压缩感知等领域的研究取得了显著成果。
2. 算法瓶颈：然而，传统算法在面对大规模线性逆问题时，依然存在局限性，如计算效率低下和解决精度不足。
3. 需求：因此，发展新型算法来应对大规模数据处理的挑战变得尤为重要。

1. 当前研究进展：近年来，稀疏优化和压缩感知等领域的研究取得了显著成果。
2. 算法瓶颈：然而，传统算法在面对大规模线性逆问题时，依然存在局限性，如计算效率低下和解决精度不足。
3. 需求：因此，发展新型算法来应对大规模数据处理的挑战变得尤为重要。
4. 两大基本发展趋势：
    - 传统算法的 AI 化
    - 大规模线性逆问题的分解-融合型算法

1 课题背景

2 研究现状

3 研究内容

4 研究思路

5 参考文献

项目目标

- 设计大规模线性逆问题的新型数值方法，提升计算效率，解决维数灾难和“强困难域”问题

1. 正交子矩阵级联的线性逆问题的分裂-融合算法：

1. 将矩阵分裂成相互正交的子块

$$A = [A_1, A_2, ..., A_j]$$

1. 正交子矩阵级联的线性逆问题的分裂-融合算法:

1. 将矩阵分裂成相互正交的子块

$$A = [A_1, A_2, ..., A_j]$$

2. 目标是充分利用测量矩阵的正交结构，设计高效的求解算法。

1. 正交子矩阵级联的线性逆问题的分裂-融合算法:

1. 将矩阵分裂成相互正交的子块

$$A = [A_1, A_2, ..., A_j]$$

2. 目标是充分利用测量矩阵的正交结构，设计高效的求解算法。
3. 传统算法未能充分利用这种结构，导致潜能未被完全实现。

2. 大规模非结构性线性逆问题的分裂-融合算法:

1. 针对一般非结构的大规模线性逆问题，设计两类算法:
   - 固定分裂策略：在迭代过程中保持矩阵的列分裂方式恒定，子块保持欠定性。

2. 大规模非结构性线性逆问题的分裂-融合算法:

1. 针对一般非结构的大规模线性逆问题，设计两类算法:
   - 固定分裂策略：在迭代过程中保持矩阵的列分裂方式恒定，子块保持欠定性。
   - 自适应分裂策略：根据当前迭代点信息调整矩阵的列分裂，提升收敛速度和稳定性
2. 关键技术问题：如何平衡矩阵分裂个数与算法求解成功率

3. 基于矩阵网格分裂的大规模线性逆问题的分裂-融合算法：

1. 当 m，n 巨大且差距不大时，将矩阵 A 分裂成水平方向上 J 块和纵向方向上 K 层。

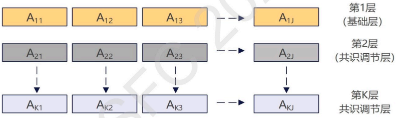

图 2.1：矩阵网格分裂图

3. 基于矩阵网格分裂的大规模线性逆问题的分裂-融合算法:

1. 当 m, n 巨大且差距不大时，将矩阵 A 分裂成水平方向上 J 块和纵向方向上 K 层。

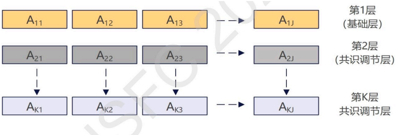

图 2.1：矩阵网格分裂图

2. 子问题解的融合，用于恢复原问题的解；以及行问题之间的共识，用于维持行层子问题解的一致性

3. 基于矩阵网格分裂的大规模线性逆问题的分裂-融合算法:

1. 当 m, n 巨大且差距不大时，将矩阵 A 分裂成水平方向上 J 块和纵向方向上 K 层。

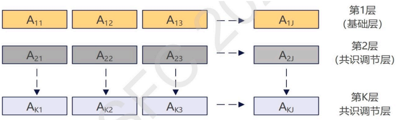

图 2.1：矩阵网格分裂图

2. 子问题解的融合，用于恢复原问题的解；以及行问题之间的共识，用于维持行层子问题解的一致性
3. 根据迭代点提供的信息，以及行层子问题之间的共识程度，调整网格分裂的规则

4. Shapley Value (SV) 数据价值计算中的分解-融合算法：

1. Shapley Value 的精确计算公式如下：

$$S_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot (Z(S \cup \{i\}) - Z(S))$$

4. Shapley Value (SV) 数据价值计算中的分解-融合算法:

1. Shapley Value 的精确计算公式如下:

$$S_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot (Z(S \cup \{i\}) - Z(S))$$

2. 将 SV 计算建模为大规模线性逆问题，通过统计模型进行近似计算。

4. Shapley Value (SV) 数据价值计算中的分解-融合算法：

1. Shapley Value 的精确计算公式如下：

$$S_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot (Z(S \cup \{i\}) - Z(S))$$

2. 将 SV 计算建模为大规模线性逆问题，通过统计模型进行近似计算。
3. 采用分解-融合算法解决大规模线性逆问题。

4. Shapley Value (SV) 数据价值计算中的分解-融合算法：

1. Shapley Value 的精确计算公式如下：

$$S_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot (Z(S \cup \{i\}) - Z(S))$$

2. 将 SV 计算建模为大规模线性逆问题，通过统计模型进行近似计算。
3. 采用分解-融合算法解决大规模线性逆问题。
4. 拓展至 Shapley Value (SV) 计算，提升计算效率。

4. Shapley Value (SV) 数据价值计算中的分解-融合算法：

1. Shapley Value 的精确计算公式如下：

$$S_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot (Z(S \cup \{i\}) - Z(S))$$

2. 将 SV 计算建模为大规模线性逆问题，通过统计模型进行近似计算。
3. 采用分解-融合算法解决大规模线性逆问题。
4. 拓展至 Shapley Value (SV) 计算，提升计算效率。
5. 研究目标：
   - 建立 SV 计算的线性逆问题模型。

4. Shapley Value (SV) 数据价值计算中的分解-融合算法：

1. Shapley Value 的精确计算公式如下：

$$S_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot (Z(S \cup \{i\}) - Z(S))$$

2. 将 SV 计算建模为大规模线性逆问题，通过统计模型进行近似计算。
3. 采用分解-融合算法解决大规模线性逆问题。
4. 拓展至 Shapley Value (SV) 计算，提升计算效率。
5. 研究目标：
   - 建立 SV 计算的线性逆问题模型。
   - 研发低复杂度的 SV 计算方法。

4. Shapley Value (SV) 数据价值计算中的分解-融合算法：

1. Shapley Value 的精确计算公式如下：

$$S_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot (Z(S \cup \{i\}) - Z(S))$$

2. 将 SV 计算建模为大规模线性逆问题，通过统计模型进行近似计算。
3. 采用分解-融合算法解决大规模线性逆问题。
4. 拓展至 Shapley Value (SV) 计算，提升计算效率。
5. 研究目标：
   - 建立 SV 计算的线性逆问题模型。
   - 研发低复杂度的 SV 计算方法。
6. 关键问题：解决 SV 计算中的“维数灾”，设计高效算法。

1. 课题背景
2. 研究现状
3. 研究内容
4. 研究思路
5. 参考文献

1. 带正交级联矩阵问题

利用正交性，设计基于分裂的快速迭代算法，重点在于降低计算复杂度。
通过矩阵的结构特性，将问题分解为可处理的小规模子问题

2. 大规模非线性结构性逆问题

扩展第一阶段的算法，结合稀疏优化技术，应对问题的规模性与复杂度。
采用列分裂与融合策略，确保在较少计算资源下高效求解

3. 基于矩阵网格分裂的大规模线性逆问题

引入网格分裂-融合-共识算法，通过行列分裂减小子问题规模，利用共识机制保持全局解的一致性。
动态调整网格分裂策略，基于共识误差实现自适应优化。

4. 数据价值计算方法

将 SV 计算问题建模成大规模线性逆问题
基于分裂-融合-共识算法，可以将 SV 计算问题分解为小规模子问题，有效解决维度灾难问题。

1 课题背景

2 研究现状

3 研究内容

4 研究思路

5 参考文献

S. Schoch, et al., CS-SHAPLEY: Class-wise shapley values for data valuation in classification, NeurIPS, 2022.
Y.-B. Zhao and D. Li, Reweighted L1-minimization for sparse solutions to underdetermined linear systems, SIAM J. Optim., 22, 1065-1088, 2012.
K.Bojan, et al., Data debugging with Shapley importance over end-to-end machine learning pipelines, arXiv:2204.11131v2, 2022.
L. Chen, P. Koutris and A. Kumar, Model-based pricing for machine learning in a data marketplace, SIGMOD ’19, June 30-July 5, 2019.

Thank you!