

Optimizing Coordinate Descent over Normalized Data in Column Stores

ABSTRACT

1. INTRODUCTION

Example

2. PRELIMINARIES AND BACKGROUND

Generalized Linear Models (GLMs)

Coordinate Descent (CD)

Column Stores

Problem Definition

3. SIMPLE APPROACHES

3.1 CD After a Join: Materialize

3.2 CD Over a Join: Stream

4. FACTORIZING CD

4.1 Main Algorithm

4.2 Extensions

4.2.1 Scalability

4.2.2 Shared-nothing Parallelism

4.2.3 Multi-table Joins

4.3 Factorizing Block CD

5. EXPERIMENTS

Simulation Setup

Real Datasets

Experimental Setup

5.1 High-level Performance and Quality

5.2 Performance Drill-Down

5.3 Evaluation of Extensions

6. RELATED WORK

7. CONCLUSION AND FUTURE WORK

8. REFERENCES