

Weekly Report

WU Zihan

August 25, 2023

Catalogue

Introduction

Method

Introduction

- ▶ The main task of this week is to run cocluster on the n -gram matrix.
- ▶ After getting the access to fat-node of HPC from CityU, currently the biggest problem is the running time.
- ▶ Some local running results are shown in the following slides.

Method

- ▶ $A \in \mathbb{R}^{27651263 \times 16094862}$ is the 3-gram matrix, where we have 27651263 paragraphs and 16094862 3-grams.
- ▶ Locally, we select 10000 paragraphs and all 3-grams to run cocluster.

Attempt 1

- ▶ First, all zero columns are removed. Then $A \in \mathbb{R}^{10000 \times 79728}$.
- ▶ Try to do cocluster on A since the linear patterns are emerging.

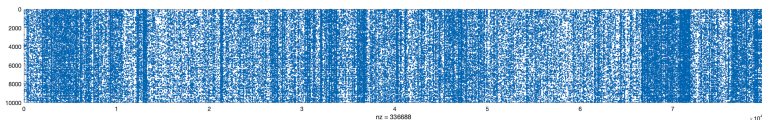


Figure: The 3-gram matrix A