

## Hadoop による分散データマイニング-植生指標モデリングへの適用- Spatio-temporal modelling of Vegetation Index using MapReduce and Hadoop

西前 光<sup>1\*</sup>, 本田 理恵<sup>1</sup>

NISHIMAE, Kou<sup>1\*</sup>, HONDA, Rie<sup>1</sup>

<sup>1</sup> 高知大学理学部応用理学科, <sup>2</sup> 本田理恵

<sup>1</sup>Department of Applied Science, Kochi University, <sup>2</sup>Rie Honda

クラウドやグリッドなどの分散データ処理のフレームワークとして注目されている Hadoop, MapReduce を時空間データマイニングの並列分散化に適用した。具体的な問題として 25 年分の南米の植生指標 (GIMMS) の時空間変動のモデリングを検討した。画像集合からの時系列抽出とこの時系列データの最尤法によるロジスティック変数でのモデリングを MapReduce 向けにコーディングして iMac6 台, Xserve 1 台 (マルチコア) の 2 種類の環境を構築して実験を行った。この際両環境での総コア数に対する速度向上比を調べ、Hadoop, MapReduce による時空間データマイニングにおける分散処理の特徴や有用性を検討した。

キーワード: データマイニング, Hadoop, MapReduce, GIMMS

Keywords: data mining, Hadoop, MapReduce, GIMMS