



学科・専攻	電子・情報工学	学籍番号	1515028	氏名	杉山 桃香
題目	spark 上で様々な GA の実行				

報告日までの取り組み

PDCA サイクル	設定目標 (P)	A-GA のサンプルプログラムを spark シェル上で実行する。 B-GA のサンプルプログラムを並列分散処理で実行する。 C-参考論文を読み要点をまとめる
	取組内容 (D)	A-python で GA を実行しているサイトを参考に spark シェル上でいくつかの GA を実行した。 【参考サイト】 sample1 遺伝的アルゴリズム-python と機械学習 (http://darden.hatenablog.com/entry/2017/03/29/213948) sample2 DEAP-python と機械学習 (http://darden.hatenablog.com/entry/2017/04/18/225459) NSGA-II 多目的最適化-python と機械学習 (http://darden.hatenablog.com/entry/2017/05/26/234845) B-Spark と Hadoop YARN による並列分散処理のサンプルプログラムを実行した。 C-参考論文の要点をまとめ、Beamer を用いてスライドを作った。
	課題整理 (C)	A-完了 B-完了 【疑問】ローカルで実行したときより並列実行したときの方が出力までに時間がかかった。
	改善方策 (A)	A-多目的遺伝的アルゴリズムの問題を本研究の問題に近づけていく必要がある。 B-【原因】扱ったデータファイルが小さいためローカルで実行する場合と並列で実行する場合とに時間の差はほぼないと考えられる。そのため、並列処理のためにデータを受け渡す時間の分、並列実行の方が時間がかかってしまったと考えられる。よって、もっと処理に時間がかかる問題なら並列処理の良さが感じられると思われる。

報告日

やりたいたいことよりやるべきことを	コメント (出席者)	
	備忘録 (自分)	