

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	ソフトウェア工学 (松本 健一 (教授))		
学籍番号	2011017	提出日	令和 4年 1月 20日
学生氏名	池上 綾乃		
論文題目	データ分析プログラム実装におけるKaggle 熟練度とソースコード再利用方法の探索的分析		
要旨			
<p>データ分析技術は情報社会における重要な技術のひとつである。世界規模のデータ分析コミュニティである Kaggle では、データ分析技術を競うコンペティションの実施や、使用したプログラムの公開、共有が行われている。データ分析プログラムには、プログラムの再利用がコピー&ペーストで行われる特徴がある。一方で、類似するプログラム片を複数箇所に記述すると、保守性の低下を招くと言われており、これを防ぐには関数の定義やライブラリの活用が必要である。しかし、データ分析プログラムの場合、プログラムの再利用は容易だが、保守作業が困難であるという報告があり、保守性を保ちつつプログラムを再利用するのは難しい。本研究では、データ分析の熟練者はプログラムを適切に再利用するという仮説に基づき、Kaggle で定義された熟練度の異なる作者が実装したプログラムから、類似するプログラム片の割合やライブラリの利用方法を比較する分析を行った。その結果、熟練度が高い作者ほど関数を多く定義し、類似するプログラム片の割合が少ない傾向が見られた。一方で、ライブラリの種類には大きな差異は見られなかった。これらの結果から、プログラムの再利用の観点でデータ分析の初学者が熟練者に近づくには、多くのライブラリを学習するより、類似する処理がある場合に自作関数を定義するような工夫が重要であると考えられる。</p>			