



Title	ソフトウェア再利用時におけるライセンス違反検出技術に関する研究
Author(s)	真鍋, 雄貴
Citation	大阪大学, 2011, 博士論文
Version Type	VoR
URL	<a href="https://hdl.handle.net/11094/2268">https://hdl.handle.net/11094/2268</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名	ま 真 鍋 雄 貴
博士の専攻分野の名称	博士（情報科学）
学 位 記 番 号	第 2 4 9 2 0 号
学 位 授 与 年 月 日	平成 23 年 9 月 20 日
学 位 授 与 の 要 件	学位規則第4条第1項該当
	情報科学研究科コンピュータサイエンス専攻
学 位 論 文 名	ソフトウェア再利用におけるライセンス違反検出技術に関する研究
論 文 審 査 委 員	(主査) 教 授 井上 克郎  (副査) 教 授 増澤 利光 教 授 楠本 真二

## 論 文 内 容 の 要 旨

ソフトウェア開発コスト削減のために、ソフトウェアの再利用が行われている。ソフトウェアの再利用とは、既存のソフトウェアに含まれるソフトウェア部品を同一、もしくは異なるソフトウェアの開発に利用することである。一方、オープンソースソフトウェア（以下、OSS）の再利用によるソフトウェアライセンス違反が問題となっている。ソフトウェアライセンス（以下、ライセンス）とは、著作物を利用するための許諾と、その許諾を得るために義務である。あるソースファイルをあるソフトウェアに再利用した際、ソースファイルのライセンスとソフトウェアのライセンスが矛盾した場合、ライセンス違反となり、ソースコードの公開や、販売停止に追い込まれる可能性がある。

本研究では、これらのライセンス違反を予防するためにライセンス違反を検出することを目的とし、ライセンス特定とソフトウェア再利用検出に関して以下を行った。

初めに、ソースファイル中のコメントでのライセンスを指定する記述（ライセンス記述）の表記ゆれのため、ライセンスの特定は容易でないという問題を解決するため、オープンソースソフトウェアから抽出した知識に基づいたライセンス特定手法を提案する。提案手法では、複数の大規模OSSから抽出した知識を用い、ソースファイルのライセンスを特定する。実験により、提案手法は既存手法より精度と実行時間の点で優れていることを示した。

次に、OSSにおけるライセンスの分布を調査し、どの程度ライセンスが変更され、その変更には傾向があるか調べた。調査結果から、調べた対象のOSSには様々なライセンスのファイルが含まれているだけでなく、その比率がソフトウェアの進化に沿って大規模に変化することが分かった。また、オペレーティングシステムと非オペレーティングシステムでは進化の傾向が異なっていることが分かった。

最後に、あるソフトウェア間で再利用が行われていると判断するための明確な基準が存在しないという問題を解決するため、コードクローンメトリクスに基づくソースコード再利用判定閾値の決定手法を提案する。提案手法では、手作業により再利用の有無が判別されたソフトウェアの組を正解集合として用い、閾値と適合率の関係から、再利用がある場合の閾値と再利用がない場合の閾値を求める。実験により、提案手法によりコードクローンメトリクスに対して、閾値を求めることができ、その閾値が正解集合の大半をカバーすることを示した。

## 論 文 審 査 の 結 果 の 要 旨

ソフトウェア開発コスト削減のために、ソフトウェアの再利用が行われている。ソフトウェアの再利用とは、既存のソフトウェアに含まれるソフトウェア部品を同一、もしくは異なるソフトウェアの開発に利用することである。一方、オープンソースソフトウェア（以下、OSS）の再利用によるソフトウェアライセンス違反が問題となっている。ソフトウェアライセンス（以下、ライセンス）とは、著作物を利用するための許諾と、その許諾を得るために義務である。あるソースファイルをあるソフトウェアに再利用した際、ソースファイルのライセンスとソフトウェアのライセンスが矛盾した場合、ライセンス違反となり、ソースコードの公開や、販売停止に追い込まれる可能性がある。

本学位論文では、これらのライセンス違反を予防するためにライセンス違反を検出することを目的とし、ライセンス特定とソフトウェア再利用検出に関して以下の研究を行ったことが報告されている。

第2章では、ソースファイル中のコメントでのライセンスを指定する記述（ライセンス記述）の表記ゆれのため、ライセンスの特定は容易でないという問題を解決するため、オープンソースソフトウェアから抽出した知識に基づいたライセンス特定手法を提案している。提案手法では、複数の大規模OSSから抽出した知識を用い、ソースファイルのライセンスを特定する。実験により、提案手法は既存手法より精度と実行時間の点で優れていることが示されている。

第3章では、OSSにおけるライセンスの分布を調査し、どの程度ライセンスが変更され、その変更には傾向があるか調査した結果が報告されている。調査結果から、調べた対象のOSSには様々なライセンスのファイルが含まれているだけでなく、その比率がソフトウェアの進化に沿って大規模に変化することが示されている。また、オペレーティングシステムと非オペレーティングシステムでは進化の傾向が異なっていることを示している。

第4章では、あるソフトウェア間で再利用が行われていると判断するための明確な基準が存在しないという問題を解決することを目的とし、コードクローンメトリクスに基づくソースコード再利用判定閾値の決定手法が提案されている。提案手法では、手作業により再利用の有無が判別されたソフトウェアの組を正解集合として用い、閾値と適合率の関係から、再利用がある場合の閾値と再利用がない場合の閾値を求める。実験により、提案手法によりコードクローンメトリクスに対して、閾値を求めることができ、その閾値が正解集合の大半をカバーすることが示されている。

以上の通り、本学位論文ではソフトウェアライセンス違反検出という問題に対し、ソフトウェアライセンスの特定とソフトウェア間での再利用検出という2つの方向から問題解決に取り組み、顕著な成果を上げたことが報告されている。また、これらは世界でも先駆的な研究であり、ソフトウェア工学分野における新たな研究分野の開拓において成功したという多大な貢献を行っているといえる。よって、博士（情報科学）の学位論文として価値のあるものと認める。