OUKA
The University of Osaka
Institutional Knowledge Archive

| | |
|---|---|
| Title | ソースコードの類似性分析に基づくソフトウェア保守支援に関する研究 |
| Author(s) | 吉田, 則裕 |
| Citation | 大阪大学, 2009, 博士論文 |
| Version Type | VoR |
| URL | https://hdl.handle.net/11094/2431 |
| rights | |
| Note | |

*The University of Osaka Institutional Knowledge Archive : OUKA*

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

The University of Osaka

【10】

| | |
|---|---|
| 氏名 | 吉田則裕（よしだのりひろ） |
| 博士の専攻分野の名称 | 博　士（情報科学） |
| 学位記番号 | 第　２３０５３　号 |
| 学位授与年月日 | 平成21年３月24日 |
| 学位授与の要件 | 学位規則第４条第１項該当<br>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 | **ソースコードの類似性分析に基づくソフトウェア保守支援に関する研究** |
| 論文審査委員 | （主査）<br>教　授　井上　克郎<br>（副査）<br>教　授　増澤　利光　　教　授　楠本　真二 |

## 論文内容の要旨

ソフトウェアに対する保守作業の効率を下げている要因の１つとして、ソースコード中の類似コード片が指摘されている。類似コード片とは、ソースコードの一部分（コード片）のうち、他のコード片と類似性が高いものを指す。保守作業において、ソースコード中のあるコード片を修正すると、その類似コード片を同時に修正する作業が必要になることが多い。

類似コード片の中でも、コードクローン（トークン列や構文木が等価であるコード片が存在するコード片）に着目し、検出を行うツールが数多く開発されている。しかし、これらツールの問題点として、あるクローンセット（コードクローンの同値類）をリファクタリング（単一のメソッドに集約）するためには、他のクローンセットもリファクタリングする必要があるという依存関係を検出しないことや、構文上に差異がある類似コード片を検出できないことが多いことが挙げられる。本研究では、これらの問題点を解決する、類似コード片を対象とした保守支援手法を提案する。

まず、チェーンドクローンセット（同時に集約を検討すべきクローンセットの集合）を検出することで、リファクタリング支援を行う手法を提案する。本手法は、クローンセット中に存在するメソッド呼び出し関係、およびメソッドと変数の利用関係を解析することで、チェーンドクローンセットを検出する。更に、検出したチェーンドクローンセットを含むクラスを要素とする集合の継承関係から、適用可能な集約方法を提示する。適用実験では、複数のソフトウェアから多くのチェーンドクローンセットを検出し、そのうちのいくつかは、提示する集約方法を適用することができた。

次に、識別子の類似性に基づく類似コード片検索手法を提案する。本手法は、クエリとしてコード片を与えると、識別子の類似性に基づいて対象ソースコードから類似関数（クエリとして与えられたコード片の類似コード片を含む関数）を検索する。具体的には、まず自然言語処理の分野で提案されている類義語特定法を用いて、語（識別子を分割・正規化した後の文字列）の類義語を特定する。次に、クエリとして与えられたコード片に含まれる全ての語について、同一もしくは類義語である語を含む関数を検出し、類似関数として提示する。適用実験として、本手法を用いて類似した欠陥を含むコード片の検索を行ったところ、類似した欠陥の多くを提示できることが確認された。

## 論文審査の結果の要旨

近年、長期にわたって運用される大規模ソフトウェアが増加しており、ソフトウェアの保守にかかる費用が大きくなってきている。そのため、ソフトウェア工学において保守作業の効率化が重要な課題として取り上げられるようになった。

保守作業の効率を下げる要因の１つとして、本論文で扱っている類似コード片が挙げられる。類似コード片とは、ソースコードの一部分（コード片）のうち、他のコード片と類似性が高いものを指す。ソースコード中のあるコード片を修正すると、その類似コード片を同時に修正する作業が必要になることが多い。既に、類似コード片の一種であるコードクローン（トークン

列や構文木が等価であるコード片が存在するコード片）を検出するツールが数多く提案されているが、本論文ではこれらのツールの問題点を２つ指摘している。まず、クローンセット（コードクローンの同値類）を単一のメソッドに集約するためには、他のクローンセットも集約する必要があるという依存関係を検出しないことを指摘している。続いて、構文上に差異がある類似コード片を検出できないことが多いことを指摘している。本論文では、これらの問題点を解決する、類似コード片を対象とした保守支援手法を提案している。

最初に、チェーンドクローンセット（同時に集約を検討すべきクローンセットの集合）を検出することにより、集約支援を行う手法を提案している。本手法は、クローンセット間の依存関係を解析することで、チェーンドクローンセットを検出する。更に、検出したチェーンドクローンセットを含むクラス間の継承関係から、適用可能な集約方法を提示する。適用実験において、本手法により提示される集約方法が、従来の集約方法と比較して効率的であることが確認されている。

次に、識別子の類似性に基づく類似コード片検索手法を提案している。本手法は、検索質問としてコード片を与えると、識別子の類似性に基づいて対象ソースコードから類似関数（クエリとして与えられたコード片の類似コード片を含む関数）を検索する。適用実験として、本手法を用いて類似した欠陥を含むコード片の検索を行っている。その結果、類似した欠陥の多くを提示できることが確認されている。

以上のように、本論文の内容は、ソースコードの類似性分析を行うことにより、類似コード片の集約や検索を支援するものであり、効率的な保守作業の実現に大きく寄与している。よって、本論文は博士（情報科学）論文として十分な価値があるものと認める。