コードクローン変更管理システムの開発と実プロジェクトへの適用

山中 裕樹[1]　崔 恩瀞[1]　吉田 則裕[2]　井上 克郎[1]　佐野 建樹[3]

概要：ソフトウェア保守における大きな問題の一つとしてコードクローンが指摘されている．コードクローンとは，ソースコード中に，互いに一致または類似した部分を持つコード片のことである．コードクローンに対する主な保守作業として，クローンセット（互いにコードクローンとなっているコード片の集合）に含まれる全てのコード片を一貫して編集する同時修正と，クローンセットを１つのサブルーチンにまとめる集約が挙げられる．本研究では，コードクローンに対する保守作業を支援することを目的としたコードクローン変更管理システムの開発を行った．そして，企業で行われているソフトウェア開発に適用することによって，本システムの有用性を確かめることができた．

キーワード：コードクローン，ソフトウェア保守，変更管理

A Development of Clone Change Management System and Its Application to Actual Project

Yuki Yamanaka[1]　Eunjong Choi[1]　Norihiro Yoshida[2]　Katsuro Inoue[1]　Tateki Sano[3]

Abstract: Code clone is one of the major problems for software maintenance. A code clone is a code fragment that has identical or similar portion in source code. In order to manage code clones, software developers should consider consistent modification of clone sets(i.e., a set of code clones identical or similar to each other) and merging clone set into a single function. In this study, we developed a code clone change management system for maintenance to them. Moreover, we confirmed the usefulness of the developed system by applying to industrial software development.

Keywords: Code Clone, Software Maintenance, Change Management

1. まえがき

ソフトウェアの保守工程における大きな問題の一つとしてコードクローンが指摘されている．コードクローンとは，ソースコード中に互いに一致または類似した部分を持つコード片のことであり，主にコピーアンドペーストなどによって発生する [1]．コードクローンに対する主な保守作業として，同時修正と集約が挙げられる．同時修正とは，クローンセット（互いにコードクローンとなっているコード片の集合）に含まれる全てのコード片を一貫して編集することである．例えば，クローンセット中の１つのコード片に欠陥が含まれている場合，そのコード片とコードクローンになっている他のコード片も一貫した修正を行う必要が考えられる．また，集約とはクローンセットを１つのサブルーチンにまとめることである．集約を行うことによって，ソースコード中のコードクローンを削減することが可能である．

上述したコードクローンに対する保守作業を効率良く行うために，コードクローンの変更管理が必要である．例えば，欠陥を含むクローンセット中の一部のコード片が編集された場合，同時修正が行われておらず，修正漏れの可能性がある．また，ソフトウェア開発企業では，システムテ

[1] 大阪大学
Osaka University
[2] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology
[3] 日本電気株式会社
NEC Corporation

ストに大きなコストを要した場合，コードクローンを集約し再度システムテストを行うということは，コスト面の問題から避けられることが多い．従って，システムテスト前に発生したコードクローンを発見・集約することで，大きなコストをかけずコードクローンを削減する必要がある．しかし，検出されたコードクローンの量が膨大となる場合，同時修正が行われていないクローンセットや，新たに発生したクローンセットなどの変更情報を人手で確認することは困難である．

そこで本研究では，コード片の追加，編集，削除といった変更情報に基づいたコードクローンの分類手法を提案する．そして，分類に基づいて，保守作業の対象となるコードクローンの変更情報を定期的に開発者に通知するコードクローン変更管理システムの開発を行った．評価実験として，開発したシステムを企業で行われているソフトウェア開発に約 40 日間適用した．実験の結果，本システムを用いることによって，開発者は保守作業の対象となる 11 個のクローンセットを確認することができた．

本稿の構成は次のとおりである．2 節では，本研究の背景を述べる．3 節では，コードクローンの分類手法について説明する．4 節では，本研究で開発したコードクローン変更管理システムの概要について説明する．5 節では，コードクローン変更管理システムの実プロジェクトに対する適用実験について述べる．6 節では，本研究の関連研究について述べる．7 節では，本研究のまとめと今後の課題について述べる．

2. 背景

本節では，本研究の背景として，コードクローンとその検出ツール CCFinder，および，コードクローンの変更管理の必要性について述べる．

2.1 コードクローン

コードクローン (Code clone) とは，ソースコード中に存在する，互いに一致または類似した部分を持つコード片のことである [1]．一般的に，コードクローンの存在はソフトウェアの保守を困難にすると言われている．例えば，コードクローンとなっているコード片中に欠陥が存在する場合，そのコード片と一致または類似した他のコード片についても同様の欠陥が含まれている可能性があり，確認が必要である．

一般的に，互いに一致または類似したコードクローンの対をクローンペア (Clone pair) と呼び，クローンペアにおいて推移関係が成り立つコードクローンの集合をクローンセット (Clone set) と呼ぶ．

2.2 コードクローン検出ツール CCFinder

ソフトウェアの規模が大きい場合，ツールを用いた自動的なコードクローン検出が行われる．本研究では，字句解析ベースの検出ツールである CCFinder [2] を用いてコードクローンの検出を行っている．CCFinder は，字句解析でソースコードをトークン列に変換し，変換処理で変数名や関数名などを同一のトークンに変換する．そして，閾値以上の長さの共通トークン列を探索し，全てのコードクローンの対のリストを出力する．CCFinder を用いることによって，空白やタブの有無などのコーディングスタイルを除いて完全に一致するコードクローンだけではなく，ユーザ定義名や一部の予約語のみが異なるコードクローンを検出することが可能である [3]．

2.3 コードクローンの変更管理の必要性

コードクローンに対する主な保守作業として，以下の項目が挙げられる．

同時修正: クローンセット中のコード片を一貫して編集すること．

集約: クローンセット中のコード片に共通するロジックを実装するサブルーチンを作り，各コード片をそのサブルーチンの呼び出し文に置き換えること．

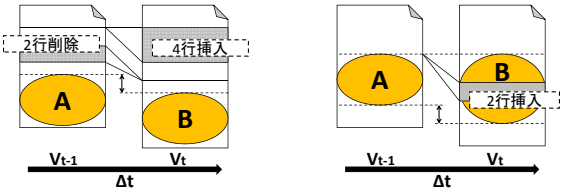
位置情報の記録: ソースコードにコメントとして，コードクローンとなっている他のコード片の位置を書くこと．

上述したコードクローンに対する保守作業を効率良く行うために，コードクローンの変更管理が必要である．例えば，欠陥を含むクローンセットの一部のコード片が修正された場合，そのクローンセットに含まれる他のコード片についても同様の修正を検討する必要がある．また，ソフトウェアのテスト終了後におけるクローンセットの集約は，不具合を生む可能性がありコストが大きい．新たに発生したクローンセットを集約の対象とすることで，バグを生むコストを小さくすることが可能である．また，集約することが困難である場合，新たに発生したコードクローンは位置情報の記録の対象になる可能性がある．

もし検出されたコードクローンの量が膨大となる場合，その中から，同時修正が行われていないクローンセットや新たに発生したクローンセットなど，保守作業が必要となる可能性が高いコードクローンの変更情報を人手で確認することは困難である．コードクローンの変更管理を行うことによって，保守作業の対象となるコードクローンの確認コストを削減することが可能になると考えられる．

3. コードクローンの分類手法

本研究では，コードクローンの変更情報を開発者に提供するため，コード片の追加，編集，削除といった変更に基づいたコードクローンの分類手法を提案する．以下に，本手法の手順を示す．入力は，分析の対象となる 2 バージョンのソースコード V_{t-1}，V_t である．V_t は最新バージョンのソースコード，また，V_{t-1} は 1 つ前のバージョンのソー



(a) ケース 1　　(b) ケース 2

図 1　コードクローンの親子関係
Fig. 1　Parent-child relationship of code clones

スコードを表す．また，V_i のソースコードに含まれる全てのコードクローンの集合を C_i と表す．

手順 1:　V_{t-1}，V_t に CCFinder を適用し，コードクローン集合 C_{t-1}，C_t の検出を行う．

手順 2:　コードクローンの変更情報を分析するために，C_{t-1} に含まれるコードクローンと C_t に含まれるコードクローンの対応関係を求める必要がある．そこで，3.1 節で説明する定義に基づいて，コードクローンの親子関係を求める．

手順 3:　3.2 節で説明する定義に基づいて，2 バージョンのソースコードに含まれる全てのコードクローンを分類する．

手順 4:　3.3 節で説明する定義に基づいて，2 バージョンのソースコードに含まれる全てのクローンセットを分類する．

3.1　コードクローンの親子関係

コードクローンの変更情報を分析するためには，C_{t-1} に含まれるコードクローンと C_t に含まれるコードクローンの対応関係を求める必要がある．本手法では，文献 [4] のコードクローンの履歴分析と同様の手法を用いてコードクローンの親子関係を求める．ここでは，あるコードクローン $A \in C_{t-1}$ に対応するコードクローンが $B \in C_t$ である場合，コードクローン B をコードクローン A の**子クローン**，コードクローン A をコードクローン B の**親クローン**と定義する．

図 1 はコードクローンの親子関係の例を示している．図 1(a) では，コードクローン $A \in C_{t-1}$ の前で 4 行の挿入と 2 行の削除が行われているため，A に対応するコード片 B の開始行番号と終了行番号は，それぞれ A の開始行番号と終了行番号に 2 行追加した値となる．B が C_t に含まれる場合，コードクローン A の子クローンはコードクローン B となる．また，図 1(b) では，コードクローン A の前で編集操作は行われていないため，対応するコード片 B の開始行番号は A の開始行番号と同じになる．一方 A に 2 行の挿入が行われているために，B の終了行番号は A の終了行番号に 2 行追加した値となる．B が C_t に含まれる場合，コードクローン A の子クローンはコードクローン B となる．このように，あるコードクローン $A \in C_{t-1}$ に対応するコード片をその開始行番号と終了行番号の対応に基づいて求め，そのコード片 B がコードクローンとなっている場合，コードクローン A とコードクローン B の間に親子関係を定義する．

3.2　コードクローンの分類

まず最初に，任意のコードクローン X について以下の 4 つの命題を定める．

- $P(X)$：X の親クローンが存在する
- $C(X)$：X の子クローンが存在する
- $M(X)$：2 バージョン間で X が編集されている
- $CP(X)$：X の親クローンと X がクローンペアである

本手法では，これらの命題を用いて 2 バージョン間のソースコード V_{t-1}，V_t に含まれる全てのコードクローンを以下の 5 項目に分類している．

Stable Clone:　$P(X) \wedge \neg M(X)$ を充足するコードクローン $X \in C_t$ とその親クローンを指す．すなわち，Stable Clone は，2 バージョン間で変更がないコードクローンを意味する．

Modified Clone:　$P(X) \wedge M(X) \wedge CP(X)$ を充足するコードクローン $X \in C_t$ とその親クローンを指す．すなわち，Modified Clone は，変数名の変更などの編集がなされたが，編集後も同じクローンセットに属するコードクローンを意味する．

Moved Clone:　$P(X) \wedge M(X) \wedge \neg CP(X)$ を充足するコードクローン $X \in C_t$ とその親クローンを指す．すなわち，Moved Clone は，文の挿入など大幅な編集がなされたため，異なるクローンセットに属するようになったコードクローンを意味する．

Added Clone:　$\neg P(X)$ を充足するコードクローン $X \in C_t$ を指す．すなわち，Added Clone は，コード片のコピーアンドペーストなどによって V_t で新たに発生したコードクローンを意味する．

Deleted Clone:　$\neg C(X)$ を充足するコードクローン $X \in C_{t-1}$ を指す．すなわち，Deleted Clone は，クローンセットの集約やコード片の削除などによって除去されたコードクローンを意味する．

3.3　クローンセットの分類

本手法では，2 バージョン間のソースコード V_{t-1}，V_t に含まれる全てのクローンセットを以下の 4 項目に分類している（**図 2**）．

Stable Clone Set:　図 2 のクローンセット A のように，V_{t-1}，V_t の 2 バージョンに渡って存在するクローンセットで，属するコードクローンが全て Stable Clone に分類されるクローンセットを指す．

Changed Clone Set:　V_{t-1}，V_t の 2 バージョンに渡って存在するクローンセットで，属するコードクローン

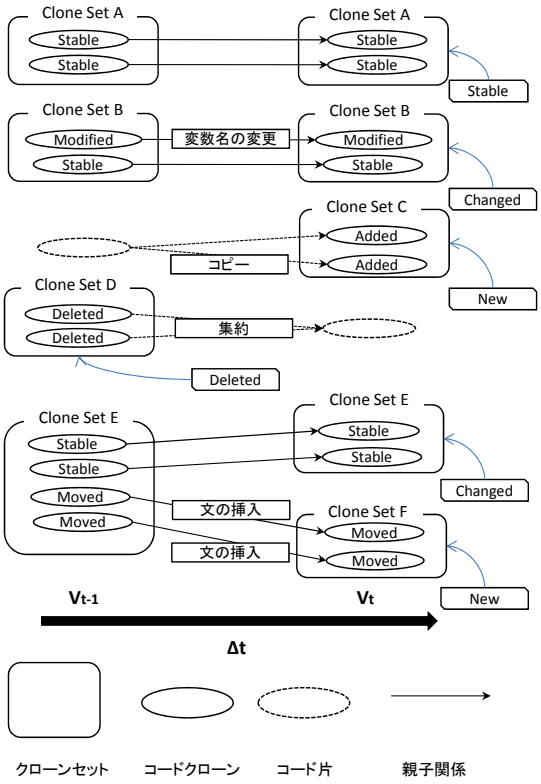


図 2 コードクローン・クローンセットの分類例
Fig. 2 Example of categorization of code clones and clone sets

に 1 つでも Stable Clone 以外のコードクローンが含まれるクローンセットを指す．例えば，図 2 のクローンセット B は一部のコードクローンのみが編集され，Modified Clone に分類されている．また，クローンセット E では，一部のコードクローンに対して文の挿入がなされたため，V_t では異なるクローンセット F を形成している．このように，同時修正がなされておらず，修正漏れの可能性があるクローンセットは Changed Clone Set に分類される．

New Clone Set: V_t のみに存在するクローンセットを指す．例えば，図 2 のクローンセット C のようにコード片のコピーアンドペーストなどによって新たに発生したクローンセットを意味する．

Deleted Clone Set: V_{t-1} のみに存在するクローンセットを指す．例えば，図 2 のクローンセット D のように集約などによって除去されたクローンセットを意味する．

4. コードクローン変更管理システム

本研究では，3 節で説明したコードクローンの分類に基づいて，保守作業の対象となるコードクローンの変更情報を開発者に提供するコードクローン変更管理システムの開発を行った．分析結果の提供方法として，テキストベースの電子メールによる通知とウェブベースのユーザインタフェースの提供を実現している．

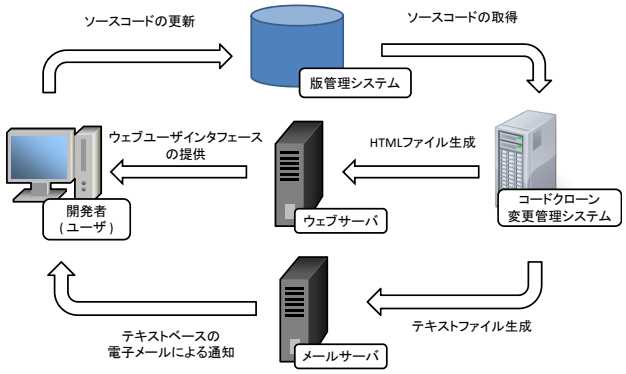


図 3 コードクローン変更管理システムの処理の流れ
Fig. 3 Process of code clone change management system

コードクローン変更管理システムの処理手順を以下に示す（**図 3**）．なお，本システムでは，ソフトウェア開発に CVS[*1]，Subversion[*2] などの版管理システムを用いることを想定している．

手順 1: 最新バージョンのソースコード V_t を版管理システムから取得する．V_{t-1} は過去のバージョンのソースコードであるため，前回の分析時に最新バージョンとして分析したソースコードを V_{t-1} として用いる．

手順 2: 3 節で説明した手法で，V_{t-1} と V_t のソースコードに含まれるコードクローン，および，クローンセットの分類を行う．

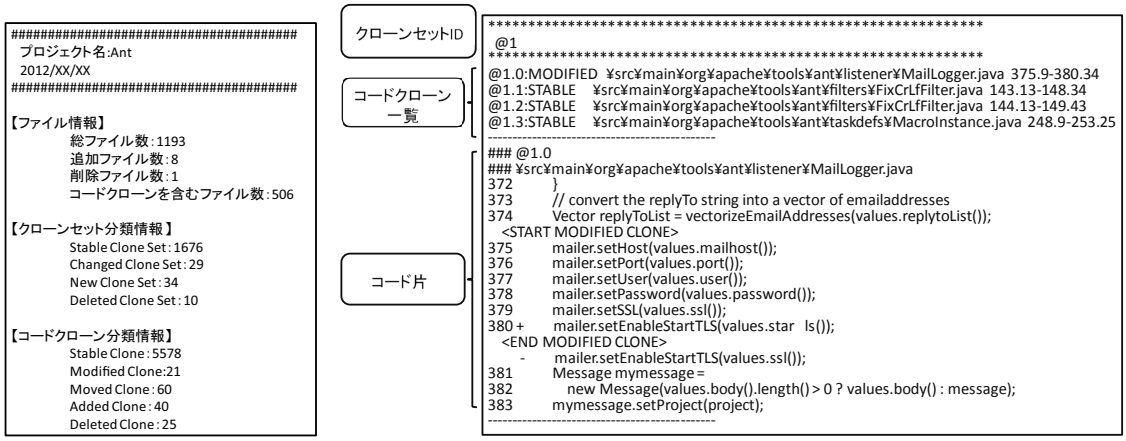
手順 3: 手順 2 の分類結果に基づいて，HTML ファイルと電子メールによる通知のためのテキストファイルの生成を行う．

本システムでは，手順 3 で生成したテキストファイルを添付した電子メールを送信することによって，開発者への通知を行なっている．また，情報提示に必要な HTML ファイルをウェブサーバ上に生成することによって，開発者はウェブブラウザを通して保守作業の対象となるコードクローンの変更情報を詳細に確認することが可能となる．

本システムによって，変更されたコードクローンの分類情報を開発者に提供する利点を以下に示す．

- 新たに発生したクローンセットは，保守作業の対象となる可能性がある．従って，New Clone Set に分類されたクローンセットを確認することで，集約，あるいは，位置情報の記録が必要となるクローンセットを把握することができる．
- Stable Clone と Modified Clone, Moved Clone が同時に含まれる Changed Clone Set を確認することで，同時修正が行われていないクローンセットを把握することができる．

[*1] http://www.cvshome.org/.
[*2] http://subversion.tigris.org/



(a) プロジェクト情報　　(b) クローンセット一覧

図 4　テキストファイルの出力例

Fig. 4　Example of text-based visualization

プロジェクト名：Ant

クローンセット一覧

Changed Clone Set

クローンセット ID:66			
ID	分類	ファイル	位置
79	DELETED	¥src¥main¥org¥apache¥tools¥ant¥filters¥ExpandProperties.java	73.9–87.17
117	STABLE	¥src¥main¥org¥apache¥tools¥ant¥filters¥PrefixLines.java	86.9–100.17
138	STABLE	¥src¥main¥org¥apache¥tools¥ant¥filters¥SuffixLines.java	87.9–101.17

クローンセットID:44			
ID	分類	ファイル	位置
232	MODIFIED	¥src¥main¥org¥apache¥tools¥ant¥listener¥MailLogger.java	375.9–380.34
82	STABLE	¥src¥main¥org¥apache¥tools¥ant¥filters¥FixCrLfFilter.java	143.13–148.34
87	STABLE	¥src¥main¥org¥apache¥tools¥ant¥filters¥FixCrLfFilter.java	144.13–149.43
823	STABLE	¥src¥main¥org¥apache¥tools¥ant¥taskdefs ¥MacroInstance.java	248.9–253.25

(a) クローンセット一覧ページ

```
92        */
          [START ID:78(Deleted Clone) CLONESET:39(Deleted Clone Set)]
93        public int read() throws IOException {
94 +          if (index > EOF) {
95 +              if (buffer == null) {
96 +                  String data = readFully();
-
          [START ID:79(Deleted Clone) CLONESET:66(Deleted Clone Set)]
-             int ch = -1;
-
-             if (queuedData != null && queuedData.length() == 0) {
-                 queuedData = null;
-             }
-
-             if (queuedData != null) {
-                 ch = queuedData.charAt(0);
-                 queuedData = queuedData.substring(1);
-                 if (queuedData.length() == 0) {
-                     queuedData = null;
-                 }
-             } else {
          [END ID:78]
-                 queuedData = readFully();
-                 if (queuedData == null || queuedData.length() == 0) {
          [END ID:79]
-                     ch = -1;
-                 } else {
97                   Project project = getProject();
98 +                 GetProperty getProperty;
99 +                 if (propertySet == null) {
100+                     getProperty = PropertyHelper.getPropertyHelper(project);
101+                 } else {
```

(b) ソースファイルページ

図 5　ウェブユーザインタフェースの例

Fig. 5　Example of web-based user interface

- Deleted Clone Set は，集約された事によってクローンセットを形成しなくなった可能がある．従って Deleted Clone Set を確認することで，集約されたクローンセットを把握することができる．

4.1　電子メールを用いた通知

電子メールに添付されるテキストファイルの例として，Apach Ant プロジェクト[*3]のある 2 バージョン間に適用した結果を**図 4** に示す．テキストファイルには以下の情報が出力される．

- **プロジェクト情報**（図 4(a)）
  - ファイル情報: 総ファイル数，追加ファイル数，削除ファイル数，コードクローンを含むファイル数を示す．
  - コードクローン分類情報: コードクローンの各々の分類数を示す．
  - クローンセット分類情報: クローンセットの各々の分類数を示す．
- **クローンセット一覧** (図 4(b))
  Changed Clone Set, New Clone Set, Deleted Clone Set に分類されたクローンセットの一覧が出力される[*4]．図 4(b) は Changed Clone Set に分類されたクローンセットの一例を示している．各々のクローンセットに関して，ク以下の情報を出力している．
  - クローンセット ID
  - クローンセットに属するコードクローンの一覧
  - コードクローンとなっているコード片

4.2　ウェブユーザインタフェースの提供

ウェブユーザインタフェースの例として，Apach Ant プロジェクトのある 2 バージョン間に適用した結果を**図 5** に示す．本稿では紙面の都合上，クローンセット一覧ページとソースファイルページについて説明する．

- **クローンセット一覧ページ**（図 5(a)）
  2 バージョン間に含まれるクローンセットの一覧が，

[*3] http://ant.apache.org/.
[*4] Stable Clone Set は変更がないクローンセットであるため，ユーザに提供する価値は低くテキストベースの通知では省略している．

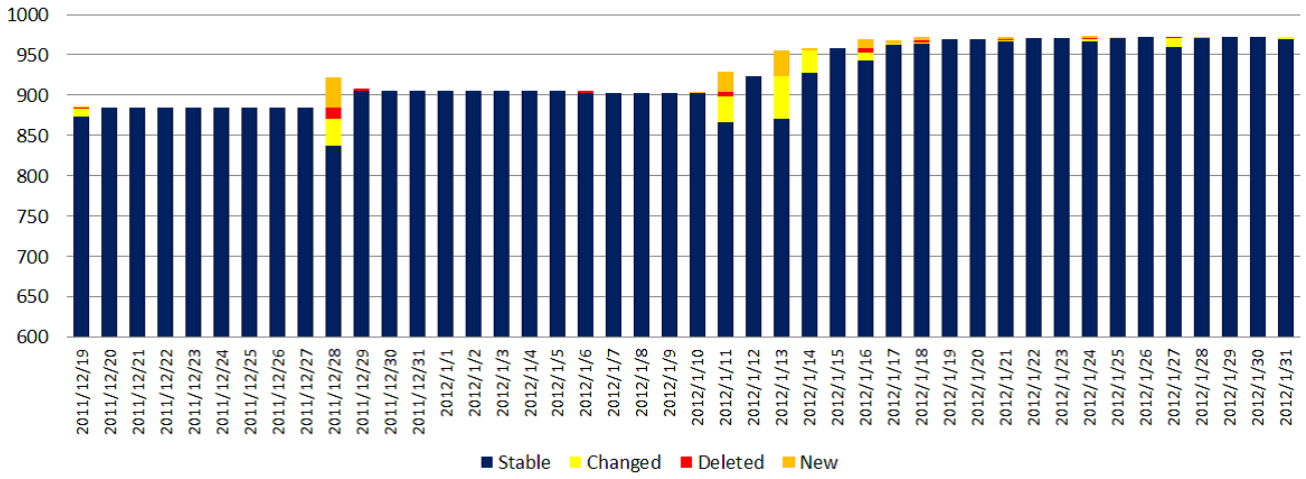


図 6 クローンセットの分類
Fig. 6 Categorization of clone sets

Changed Clone Set，New Clone Set，Deleted Clone Set，Stable Clone Set の順で表示される．また，各々のクローンセットに関して，属するコードクローンの一覧が表示される．各々のコードクローンは，ID，分類，コードクローンが含まれるソースファイル名，ソースファイル中のコードクローンの位置が表示される．Modified Clone，Moved Clone，Added Clone，Deleted Clone の 2 バージョン間で変更されたコードクローンはハイライトされ，見やすくなっている．コードクローン ID をクリックすることによって，そのコードクローンが含まれるソースファイルページへ移動する．

- **ソースファイルページ**（図 5(b)）
  コードクローンが含まれるソースファイルが表示される．コードクローンとなっているコード片は黄色の背景色が付いている．“+”は追加行を，“-”は削除行を示し，削除行は灰色の背景色がついている．

5. 適用実験

実験として，企業で行われているソフトウェア開発に対してコードクローン変更管理システムの適用を行った．適用実験の目的は以下の通りである．

- 開発者がコードクローン変更管理システムを用いて保守作業が必要となるコードクローンを発見することができたか，本システムの有用性を評価する．
- 保守作業の対象となったコードクローンの特徴を調査する．

5.1 実プロジェクトに対する評価実験

5.1.1 実験内容

本研究では，日本電気株式会社のウェブアプリケーションソフト開発を対象に評価実験を行った．対象としたソフトウェアの実装言語は Java であり，ファイル数は約 350，行数は約 12 万である．コードクローン変更管理システムを 2011/12/18 から 2012/01/31 の約 40 日間適用し，1 日毎に，電子メールによる開発者への通知とウェブユーザインタフェースの更新を行った．なお，本実験では CCFinder のトークンの閾値をデフォルトである 30 に設定した．そして，コードクローン変更管理システムの有用性を評価するために，アンケートを実施した．アンケートの対象者は，Java 言語の経験が 10 年以上のプロジェクトマネージャーであり，コードクローンの管理を行っている．主なアンケート内容を以下に示す．

Q1 保守作業の対象となったコードクローンの存在を既に知っていたか
Q2 どのような保守作業を行ったか

- 集約
- 同時修正
- 位置情報の記録
- その他

5.1.2 結果と考察

アンケートの結果，開発者はコードクローン変更管理システムを用いて保守作業の対象となる 11 個のクローンセットを発見することができたことがわかった．**表 1** 左部は，保守作業の対象となった各々のクローンセットに対するアンケートの回答を示している．保守作業の対象となったのは，主に機能追加によって発生したコードクローンである．Q1 に対する回答は全て “NO” であり，これらのクローンセットは本システムを用いて新たに発見することができたものである．この結果から，本システムの有用性が評価できる．

図 6 は，1 日毎のクローンセット数の遷移と，それぞれのクローンセットの分類数を示している．この結果，実際に変更があったクローンセットは最大でも全体の 8.9%であ

表 1 保守作業の対象となったクローンセット
Table 1 Clone sets that required additional maintenance

検出日	Q1	Q2	RAD(S)	LEN(S)	RNR(S)	NIF(S)	POP(S)	DFL(S)
’11/12/28	NO	集約	1	36	83	2	2	36
’11/12/28	NO	集約	0	141	78	1	2	141
’12/01/13	NO	集約	1	48	95	7	7	288
’12/01/13	NO	位置情報の記録	3	54	90	2	2	54
’12/01/13	NO	集約	0	52	61	1	2	52
’12/01/13	NO	集約	0	57	94	1	2	57
’12/01/13	NO	集約	0	32	93	1	2	32
’12/01/16	NO	集約	0	37	83	1	2	37
’12/01/16	NO	集約	2	32	84	2	3	64
’12/01/18	NO	集約	0	53	90	1	2	53
’12/01/24	NO	集約	0	72	86	1	3	144

表 2 マンホイットニー U 検定の結果
Table 2 Result of Mann-Whitney U test

	RAD(S)	LEN(S)	RNR(S)	NIF(S)	POP(S)	DFL(S)
p 値	0.5106	0.2419	0.0368	0.5955	0.2793	0.2283
帰無仮説	採択	採択	棄却	採択	採択	採択

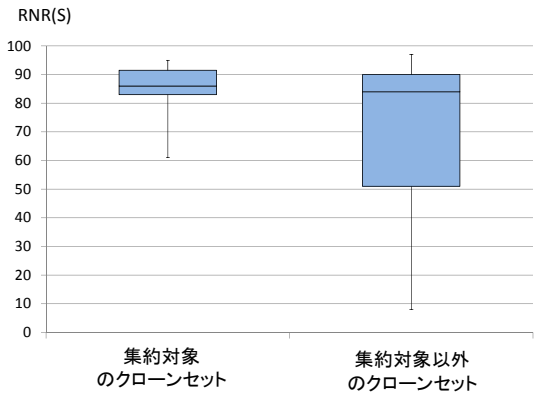


図 7 RNR(S) メトリックの比較
Fig. 7 Comparison of RNR(S) metric

ることがわかった．従って，全クローンセット数は約 850 から 950 であり，その全ての中から保守作業の対象となるコードクローンの変更情報を人手で発見することは困難であると考えられる．従って，本システムを用いることによって保守作業の対象となるクローンセットの変更情報を確認するコストを削減できたと考えられる．

5.2 保守作業の対象となったコードクローンの調査

5.2.1 調査方法

本実験では，Gemini [5], [6] を用いてクローンセットメトリクスの抽出を行った．あるクローンセット S が与えられたとき，Gemini で分析することができるクローンセットメトリクスを以下に示す．

RAD(S): クローンセット S 中のコード片が含まれるファイル集合が，ファイルシステムの中でディレクトリ構造的にどれだけ分散しているかを表す．

LEN(S): クローンセット S 中のコード片のトークン数の平均値を表す．

RNR(S): クローンセット S 中のコード片がどの程度非繰り返しであるかを表す．f をクローンセット S 中のコード片，$TOC(f)$ をコード片 f を構成している字句の数，$TOCrepeated(f)$ をコード片 f を構成している字句のうち，繰り返し要素の字句の数を表すとする．このとき RNR(S) は以下の式で表される．

$$RNR(S) = 1 - \frac{\sum_{f \in S} TOCrepeated(f)}{\sum_{f \in S} TOC(f)}$$

RNR(S) が低い場合，“ソフトウェア開発・保守を行う視点でコードクローン情報を扱う場合に特に対象とする必要が無いもの”である可能性が高いことがわかっている [7]．

NIF(S): クローンセット S 中のコードクローンを所有するファイルの数を表す．

POP(S): クローンセット S 中のコード片単位の要素数を表す．

DFL(S): クローンセット S 中の全コード片を集約した場合，減少が予測されるトークン数を表す．

本実験では，集約対象のクローンセットと保守作業の対象以外のクローンセットの各々のメトリクス値の分布に対して違いがあるか否か，マンホイットニー U 検定を用いて調査した．なお，同時修正が必要であったクローンセットは確認できなかったため，集約の対象となったクローンセットについてのみ分析を行う．ここでは，“集約対象のク

ローンセットと集約対象以外のクローンセットの間でメトリクス値による違いがない”という帰無仮説が棄却されるか否か，有意水準 0.05 で片側検定を行った．

5.2.2 調査結果

表 1 の右部に保守作業の対象となったクローンセットの各々のメトリクス値を示す．また，**表 2** に各々のメトリクス値におけるマンホイットニー U 検定の結果を示す．

この結果，RNR(S) のみが帰無仮説が棄却され，集約対象のクローンセットと集約対象以外のクローンセットの間で違いが大きいことがわかった．**図 7** では，集約対象となったクローンセットの RNR(S) メトリック値とそれ以外の RNR(S) メトリック値の分布を比較している．この図から，RNR(S) のメトリック値が比較的大きいクローンセットが集約の対象として選ばれていることがわかる．開発者からも，変数宣言の羅列などのコードクローンが多く検出されているとフィードバックがあった．これらの結果から，RNR(S) メトリック値によるフィルタリングを行うことによって，集約の対象となる可能性が高いクローンセットを開発者に提示することが可能となると考えられる．

6. 関連研究

本研究と同様にコードクローンの履歴を調査する研究として，文献 [8] がある．この研究では，クローンセットの履歴に対してモデルを定義し，長期間に渡って分析することによってコードクローンの存在する期間とその特徴の関係について調査を行っている．文献 [8] では，分析の対象となっているのはクローンセットのみであるが，本研究ではコードクローンも詳細に分類を行っている．また，本研究はソフトウェア開発者への保守作業の対象となるコードクローンの変更情報の提供を目的としている．そのため，コードクローン変更管理システムの開発を行い，企業で行われているソフトウェア開発に適用することによって，その有用性を評価している．

また，本研究と同様に，企業のソフトウェア開発に適用した研究として文献 [7], [9] が挙げられる．文献 [7] では，産業界への適用と意見交換に基づいて，本研究でも利用した Gemini の改良を行っている．本研究では，ユーザである日本電気株式会社の開発者からのフィードバックに基づいて，コードクローン変更管理システムの有用性の評価，および，保守作業の対象となったコードクローンの特徴の調査を行っている．

7. まとめと今後の課題

本研究では 2 バージョン間のコード片の変更に基づいたコードクローンの分類手法を提案した．そして，分類に基づき，保守作業の対象となる可能性があるコードクローンの変更情報を開発者に提供することを目的としたコードクローン変更管理システムの開発を行った．さらに，実際に企業で行われているソフトウェア開発に適用し，その有用性を確かめることができた．

今後の課題として，以下が考えられる．

- 長期に渡る適用と，様々なプロジェクトに対して本システムの有用性を評価する必要がある．また，コードクローン，クローンセットの分類の妥当性も評価する必要がある．
- 今回のアンケート対象者はプロジェクトマネージャ一人であったが，他の開発者からのフィードバックも調査する必要がある．
- 保守作業の対象となる可能性が高いコードクローンの情報を開発者に提示できるように，システムを改善する必要がある．具体的には，RNR(S) メトリック値に基づいたフィルタリングが考えられる．

謝辞 本研究において様々な御協力を頂いた日本電気株式会社 三橋二彩子 氏，岩崎新一 氏に深く感謝する．また，本研究は日本学術振興会 科学研究費補助金 基盤研究（A）（課題番号:21240002），基盤研究（C）（課題番号:22500026）の助成を得た．

参考文献

[1] 肥後芳樹，楠本真二，井上克郎：コードクローン検出とその関連技術，電子情報通信学会論文誌，Vol. J91-D, No. 6, pp. 1465–1481 (2008).

[2] Kamiya, T., Kusumoto, S. and Inoue, K.: CCFinder: A multilinguistic token-based code clone detection system for large scale source code, *IEEE Transactions on Software Engineering*, Vol. 28, No. 1, pp. 654–670 (2002).

[3] Bellon, S., Koschke, R., Antoniol, G., Krinke, J. and Merlo, E.: Comparison and evaluation of clone detection tools, *IEEE Transactions on Software Engineering*, Vol. 31, No. 10, pp. 804–818 (2007).

[4] 川口真司，松下誠，井上克郎：版管理システムを用いたクローン履歴分析手法の提案，電子情報通信学会論文誌，Vol. J89-D, No. 10, pp. 2279–2287 (2006).

[5] Ueda, Y., Kamiya, T., Kusumoto, S. and Inoue, K.: Gemini: Maintenance support environment based on code clone analysis, *Proc. of METRICS '02*, pp. 67–76 (2002).

[6] Higo, Y., Kamiya, T., Kusumoto, S. and Inoue, K.: Method and Implementation for Investigating Code, *Information and Software Technology*, Vol. 49, No. 9-10, pp. 95–98 (2007).

[7] 肥後芳樹，吉田則裕，楠本真二，井上克郎：産学連携に基づいたコードクローン可視化手法の改良と実装，情報処理学会論文誌，Vol. 48, No. 2, pp. 811–822 (2007).

[8] Kim, M., Sazawal, V., Notkin, D. and Murphy, G. C.: An empirical study of code clone genealogies, *Proc. of SIGSOFT/FSE '05*, pp. 187–196 (2005).

[9] 吉村健太郎，ガネサンダルマリンガム，ムーティックディルク：プロダクトライン導入に向けたレガシーソフトウェアの共通性・可変性分析法，情報処理学会論文誌，Vol. 48, No. 8, pp. 2482–2491 (2007).