大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
  出典 Publication
```

背景 Introduction 卷号 Introduction

既存研究 Related Researches

正解の例 Example of True Positive 据検出の例 Example of False Positive 不具合があるクローンの脳検出率

コードクローンの四つの参照 クローンに基づく不具合輸出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徵選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

公頼は・頼似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 輕価方法·平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

結論と会後の課題 Conclusion and Future Work

者容、結論及び今後の課題 妥当性の考察 Threats to Validity

Active Refinement of Clone **Anomaly Reports**

MD 輪講

修士課程1年 楊 嘉晨

大阪大学大学院コンピュータサイエンス専攻楠本研究室

2012年7月4日(火)

2012 年7日4日(火)

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction

卷号 Introduction 既存研究 Related Researches

正解の例 Example of True Positive 据検出の例 Example of False Positive 不具合があるクローンの脳検出率

コードクローンの四つの参照 クローンに基づく不具合輸出

クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ 洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出・五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徵選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分籍法

公頼法・頼似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

經価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

者容、結論及び今後の課題

順番を調整した例 Example of Re-ordering 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

出典 Publication

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

京 Introduction 特景 Introduction 既存研究 Related Researches 正解の例 Example of False Positive 誠検出の例 Example of False Positive 不具合があるクローンの誠検出率 ーードクローンのの全関

コートクローンの四つの家阪 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出: 五つの特徴 特徴抽出: 五つの特徴

前処理: 特徴選択 前処理: Pata Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 頼似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 fine制制中の3 Information Gain 顕番老期準し上例 Example of Re-ordering

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work 出典 Publication

Active Refinement of Clone Anomaly Reports

- ICSE 2012
- Similarity and Classification

Lucia, David Lo, Lingxiao Jiang, and Aditya Budi

Singapore Management University

2012年7月4日(火)

者容、結論及び今後の課題

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

出典 Publication

背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの部検出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴

特徴抽出:五つの特徴の例

前処理: 特徵選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

考察,結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

背景 Introduction

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction 既存研究 Polated Passas

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

従来のクローンレポートの静的洗練 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 書容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012 年 7 日 4 日(火)

2012年7月4日(火

背景

Introduction

コードクローンはソフトウェア保守に対して悪い影響

不具合 (anomaly)があるコードクローンに バグを含む可能性が高い

バグが含む不具合があるクローンを 正解 (True Positive)

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

展集 in Modeling in Manager in Manager in Manager in Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率 フードクローンの男の

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

従来のクローンレポートの静的洗練 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公額法・類似度の計算

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報刊得 Top-3 Information Gain

10p-3 情報利得 10p-3 Information Gain 順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

---- - - - - · - · · ·

2012年7月4日(火)

背景

Introduction

コードクローンはソフトウェア保守に対して悪い影響

不具合 (anomaly)があるコードクローンに バグを含む可能性が高い

バグが含む不具合があるクローンを 正解 (True Positive)

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 電景 Introduction

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 鍼検出の例 Example of False Positive 不具合があるクローンの銅棒出座

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

クローンレポートの動的洗練法 動的洗練法の流れ 洗練エンジン Refinement Engine

先練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine

特徴抽出: 構文木を構築 特徴抽出: 不つの特徴

特徴抽出: 五つの特徴の例前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公額法: 類似度の計算

万知法: 知似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7日4日(火)

2012年7月4日(火)

背景

Introduction

コードクローンはソフトウェア保守に対して悪い影響

不具合 (anomaly)があるコードクローンに バグを含む可能性が高い

バグが含む不具合があるクローンを 正解 (True Positive)

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 電景 Introduction

青景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 減検出の例 Example of False Positive 不具合があるクローンの額検出率

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

背景

Introduction

コードクローンはソフトウェア保守に対して悪い影響

不具合 (anomaly)があるコードクローンに バグを含む可能性が高い

バグが含む不具合があるクローンを 正解 (True Positive)

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

> 背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誠検出の例 Example of False Positive 不具合があるクローンの超棒出座

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ 洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 fin模利制等 10-93 Information Gain 顯番老期報上方例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

既存研究:クローンに基づく不具合検出

Related Researches: Clone-based Anomaly Detection

クローン間の識別子 (Identifier) の不一致 (Juergens et al., 2009)

E. Juergens, F. Deissenboeck, B. Hummel, and S. Wagner, ``Do code clones matter?" in Proceedings of the 31st International Conference on Software Engineering. IEEE Computer Society, 2009, pp. 485--495.

クローン周りのコード片の差異(Jiang et al., 2007a)

L. Jiang, Z. Su, and E. Chiu, ``Context-based detection of clone-related bugs," in ESEC/FSE, vol. 2007, 2007.

大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
 出典 Publication
背景 Introduction
 卷号 Introduction
 既存研究 Related Researches
 正解の例 Example of True Positive
 編検出の例 Example of False Positive
 不具合があるクローンの脳栓出率
 コードクローンの四つの象別
クローンに基づく不具合輸出
 クローンに基づく不具合検出
動的洗練法 Dynamic Refinement
 従来のクローンレポートの静的洗練法
 クローンレポートの動的洗練法
 動的法練法の流れ
洗練エンジン Refinement Engine
 洗練エンジンの流れ Process of Refinement Engine
 特徴抽出: 構文木を構築
 特徴抽出: 五つの特徴
 特徴抽出: 五つの特徴の例
```

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い SPARSE Experiment and Evaluation

実験評価 Experiment and Evaluation 評価方法:平均正解卷見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 fine制制等 10p-3 Information Gain 期番老期整七:F例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

正解の例

Example of True Positive

Linux-2.6.19 から見つけたクローン

fs/sysfs/inode.c

```
struct dentry * dentry = sd—>s dentry;
220
221
    if (dentry) {
222
         /* the following parts are detected as clones */
223
        spin lock(&dcache lock);
224
        spin lock(&dentry->d lock);
225
        if (!( d unhashed(dentry) && dentry -> d inode)) {
226
            dget locked(dentry);
227
            d drop(dentry);
228
            spin unlock(&dentry->d lock):
229
            spin_unlock(&dcache lock):
230
```

drivers/infiniband/hw/ipath/ipath_fs.c

```
456 struct dentry *tmp;
457
458
    tmp = lookup one len(name, parent, strlen(name));
459
460
   spin lock(&dcache lock);
    spin lock(&tmp->d lock);
461
462
    if (!( d unhashed(tmp) && tmp—>d_inode)) {
463
        dget locked(tmp);
464
        d drop(tmp);
465
        spin unlock(&tmp->d lock):
466
        spin_unlock(&dcache lock):
467
```

大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
 出典 Publication
背景 Introduction
 番号 Introduction
 既存研究 Related Researches
 正解の例 Example of True Positive
 編検出の例 Example of False Positive
 不具合があるクローンの脳栓出率
 コードクローンの四つの象別
クローンに基づく不具合輸出
 クローンに基づく不具合検出
動的洗練法 Dynamic Refinement
 従来のクローンレポートの静的洗練法
 クローンレポートの動的洗練法
 動的法練法の流れ
洗練エンジン Refinement Engine
 洗練エンジンの流れ Process of Refinement Engine
 特徴抽出: 構文木を構築
 特徴抽出: 五つの特徴
 特徴抽出: 五つの特徴の例
 前処理:特徵選択
 前処理: Data Re-balancing
 非ネスト汎化された最も近隣分類法
```

分類法: 類似度の計算 具体的な洗練の振舞い 実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering 素容 結論のびる後の課題

妥当性の考察 Threats to Validity

結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

正解の例

Example of True Positive

Linux-2.6.19 から見つけたクローン

```
fs/sysfs/inode.c
                                                              drivers/infiniband/hw/ipath/ipath fs.c
                                                                456 struct dentry *tmp;
     struct dentry * dentry = sd—>s dentry;
 220
221
                                                                457
      if (dentry) {
                                                                458
                                                                    tmp = lookup one len(name, parent, strlen(name));
 222
         /* the following parts are detected as clones */
                                                                459
 223
         spin lock(&dcache lock);
                                                                460
                                                                    spin lock(&dcache lock);
 224
         spin lock(&dentry->d lock);
                                                                    spin lock(&tmp->d lock):
 225
            (!( d unhashed(dentry) && dentry—>d inode)) {
                                                                462
                                                                       (!( d unhashed(tmp) && tmp->d inode)) {
 226
             dget locked(dentry);
                                                                463
                                                                        dget locked(tmp);
 227
                                                                464
              d drop(dentry);
                                                                         d drop(tmp);
 228
             spin unlock(&dentry->d lock):
                                                                465
                                                                        spin unlock(&tmp->d lock):
 229
             spin unlock(&dcache lock):
                                                                466
                                                                        spin unlock(&dcache lock):
 230
                                                                467
                                             Type-2 クローン
```

大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
  出典 Publication
背景 Introduction
  卷号 Introduction
  既存研究 Related Researches
 正解の例 Example of True Positive
  編検出の例 Example of False Positive
  不具合があるクローンの脳検出率
  コードクローンの四つの象別
クローンに基づく不具合輸出
 クローンに基づく不具合検出
動的洗練法 Dynamic Refinement
  従来のクローンレポートの静的洗練法
  クローンレポートの動的洗練法
  動的法練法の流れ
洗練エンジン Refinement Engine
  洗練エンジンの流れ Process of Refinement Engine
  特徴抽出: 構文木を構築
  特徴抽出: 五つの特徴
  特徴抽出: 五つの特徴の例
  前処理:特徵選択
  前処理: Data Re-balancing
  非ネスト汎化された最も近隣分類法
  公頼は、頼似度の計算
  具体的な洗練の振舞い
事験評価 Experiment and Evaluation
  經価方法·平均正解発見率 APPF
  実験の対象と設定 Settings of Empirical Evaluation
  実験の結果 Results of the Empirical Evaluation
  Top-3 情報利得 Top-3 Information Gain
  順番を調整した例 Example of Re-ordering
者容、結論及び今後の課題
```

正解の例

Example of True Positive

Linux-2.6.19 から見つけたクローン

fs/sysfs/inode.c struct dentry * dentry = sd—>s dentry; 220 221 if (dentry) { 222 /* the following parts are detected as clones */ 223 spin lock(&d ache lock): 224 spin lock(&dentry->d lock): 225 if (!(d unhashed(dentry) && dentry -> d inode)) { 226 dget locked(dentry); 227 d drop(dentry): 228 spin_unlock(&dentry->d_lock); 229 spin_unlock(&dcache_lock): 230

drivers/infiniband/hw/ipath/ipath_fs.c

```
456 struct dentry *tmp;
457
458
    tmp = lookup one len(name, parent, strlen(name));
459
460
    spin lock(&dcache lock);
    spin lock(&tmp->d lock);
461
462
    if (!( d unhashed(tmp) && tmp->d inode)) {
463
        dget locked(tmp);
464
        d drop(tmp);
465
        spin unlock(&tmp->d lock):
466
        spin_unlock(&dcache lock):
467
```

片方に Null であるかの判断

妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
出典 Publication
育意 Introduction
育意 Introduction
育意 Introduction
京意 Introduction
正解の例 Example of True Positive
新検出の例 Example of False Positive
不具合かあるクローンの訴検出率
コートワローンの回つの象徴
フローンに基づく不異合検出
カワーンに基づく不異合検出
対策が表達法 Dynamic Refinement
従来のクローンレポートの動物洗練法
クローンレポートの動物洗練法
動物洗練技の変れ
```

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出: 五つの特徴 特徴抽出: 五つの特徴

特徴相由: ユラの特徴の例 前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近陸分籍法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation

実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity

結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

誤検出の例

Example of False Positive

fs/nfsd/nfs3xdr.c

```
423 if (!( p = decode_fh(p, &args—>fh))
424 ||!(p=decode_filename(p,&args—>name,&args—>len))
425 ||!(p=decode_sattr3(p,&args—>attrs)))
426 return 0:
```

drivers/hwmon/lm87.c

fs/nfsd/nfsxdr.c

```
344 if (!( p = decode_fh(p, &args -> ffh))
345 | |!(p=decode_fh(p,&args -> tfh))
346 | |!(p=decode_filename(p,&args -> tname,&args -> tlen)))
347 return 0;
```

drivers/hwmon/gl520sm.c

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

原 Introduction 背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの総検出度

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ 洗練エンジン Refinement Engine

先練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択 前処理: Data Re-balancing

那処理: Data Re-Data Cing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い 実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

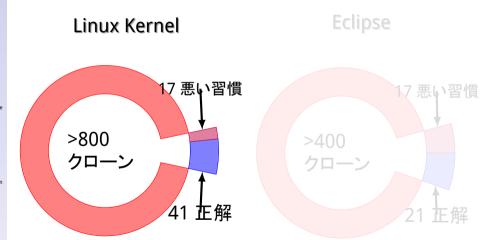
妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2042 5 2 5 4 5 (4)

2012年7月4日(火)

不具合があるクローンの誤検出率

False Positives in Anomaly Clones



L. Jiang, Z. Su, and E. Chiu, ``Context-based detection of clone-related bugs," in ESEC/FSE, vol. 2007, 2007.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率 コードクローンのの発展

コートクローンの四つの家限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ 洗練エンジン Refinement Engine

た Refinement Engine
 洗練エンジンの流れ Process of Refinement Engine
 特徴抽出: 構文木を構築

特徴抽出:五つの特徴
特徴抽出:五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法:類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation

実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

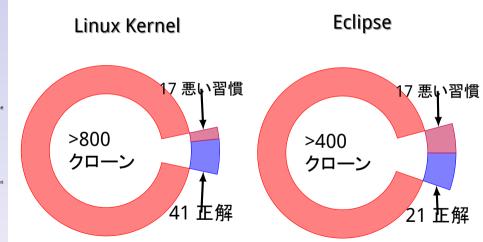
考察, 結論及び今後の課題 妥当性の考察 Threats to Validity

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

不具合があるクローンの誤検出率

False Positives in Anomaly Clones



L. Jiang, Z. Su, and E. Chiu, ``Context-based detection of clone-related bugs," in ESEC/FSE, vol. 2007, 2007.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

動的元禄法の流れ 洗練エンジン Refinement Engine

た練エンシン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い 実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 ff報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 書容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

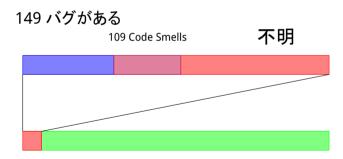
2012 年 7 目 4 日(少)

2012年/月4日(火

不具合があるクローンの誤検出率 II

False Positives in Anomaly Clones II

商用ソフトウェア(Gabel et al., 2010)



500 確かめた 8103 不具合 クローン

M. Gabel, J. Yang, Y. Yu, M. Goldszmidt, and Z. Su, ``Scalable and systematic detection of buggy inconsistencies in source code," in ACM Sigplan Notices, vol. 45, no. 10. ACM, 2010, pp. 175--190.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具会があるクローンの総練出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 認価方法: 平均正解発見率 APPE

評価方法: 平均止解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

条、結構及いう後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

コードクローンの四つの象限

4 Quadrants of Code Clone Group

| | | 一貫性 | |
|-----|---------------|--------------|------------|
| | | Inconsistent | Consistant |
| 可変性 | 厳格 (Rigid) | | |
| | 柔軟 (Flexible) | | |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの超棒出車

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 額似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法: 平均正解発見率 APPF

また面がない。 実験の新集 Results of the Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察,結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7日4日(火)

コードクローンの四つの象限

4 Quadrants of Code Clone Group

| | | | 一貫性 | |
|-------|------------|---------------|--------------|------------|
| | | | Inconsistent | Consistant |
| 可変性 | 厳格 (Rigid) | ✓ | | |
| 一, 久江 | | 柔軟 (Flexible) | ✓ | |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法・類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法: 平均正解発見率 APPF

評価方法: 平均止所発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 考察, 結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

コードクローンの四つの象限

4 Quadrants of Code Clone Group

| | | 一貫性 | |
|-----|---------------|--------------|------------|
| | | Inconsistent | Consistant |
| 可変性 | 厳格 (Rigid) | ✓ | |
| | 柔軟 (Flexible) | ✓ | |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

出典 Publication

背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの認検出率

コードクローンの四つの象限 クローンに基づく不具合検出

プローンに基づく不具合核出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 有又不を情栄 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例

前処理:特徵選択

前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Va

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

クローンに基づく不具 合検出

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法・類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

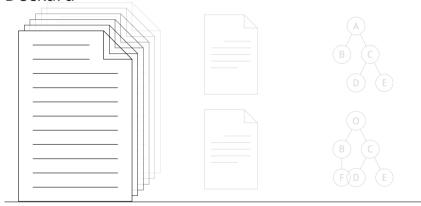
相談とう後の課題 Conclusion and Future Wo

2012年7月4日(火)

クローンに基づく不具合検出

Clone-based Anomaly Detection

Deckard^(Jiang et al., 2007b)



L. Jiang, G. Misherghi, Z. Su, and S. Glondu, ``Deckard: Scalable and accurate tree-based detection of code clones," in Proceedings of the 29th international conference on Software Engineering. IEEE Computer Society, 2007, pp. 96–105.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率 コードクローンの四つの参照

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分額法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

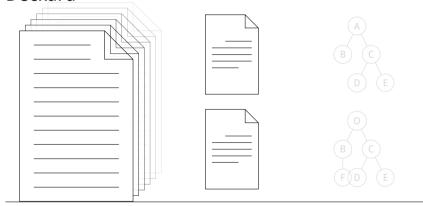
妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

クローンに基づく不具合検出

Clone-based Anomaly Detection

Deckard^(Jiang et al., 2007b)



L. Jiang, G. Misherghi, Z. Su, and S. Glondu, ``Deckard: Scalable and accurate tree-based detection of code clones," in Proceedings of the 29th international conference on Software Engineering. IEEE Computer Society, 2007, pp. 96–105.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率 コードクローンの四つの象限

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分額法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法・平均正解発量率 APPF

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

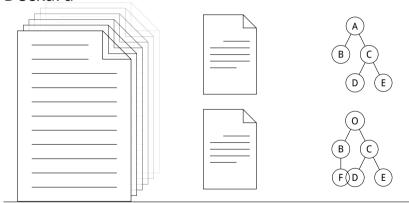
妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

クローンに基づく不具合検出

Clone-based Anomaly Detection

Deckard^(Jiang et al., 2007b)



L. Jiang, G. Misherghi, Z. Su, and S. Glondu, ``Deckard: Scalable and accurate tree-based detection of code clones," in Proceedings of the 29th international conference on Software Engineering. IEEE Computer Society, 2007, pp. 96--105.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分額法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法: 平均正解発息率 APPF

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

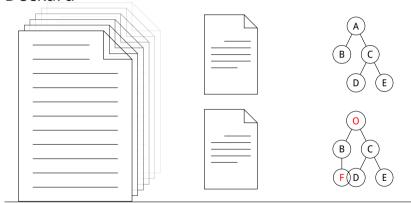
妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

クローンに基づく不具合検出

Clone-based Anomaly Detection

Deckard^(Jiang et al., 2007b)



L. Jiang, G. Misherghi, Z. Su, and S. Glondu, ``Deckard: Scalable and accurate tree-based detection of code clones," in Proceedings of the 29th international conference on Software Engineering. IEEE Computer Society, 2007, pp. 96--105.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

出典 Publication

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具会があるクローンの経緯出塞

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の溶れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例

前処理:特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法・類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

動的洗練法 Dynamic Refinement

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction 卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 据検出の例 Example of False Positive 不具合があるクローンの脳検出率 コードクローンの四つの参照 クローンに基づく不具合輸出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ 洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徵選択 前机理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公頼は・頼似度の計算 具体的な洗練の振舞い 実験評価 Experiment and Evaluation 經価方法·平均正解發見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題 妥当性の考察 Threats to Validity

結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

従来のクローンレポート静的洗練法

Static Refinement of Clone Report in Other Researches

| ID | 内容 | Bug? |
|-----|-----|------|
| 1 | AAA | |
| 2 | BBB | |
| 3 | CCC | |
| 4 | DDD | |
| 5 | EEE | |
| 6 | FFF | |
| 7 | III | |
| ••• | ••• | ••• |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction 卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 据検出の例 Example of False Positive 不具合があるクローンの脳検出率 コードクローンの四つの参照 クローンに基づく不具合輸出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ 洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徵選択 前机理: Data Re-balancing 非ネスト汎化された最も近隣分籍法 公頼は・頼似度の計算 具体的な洗練の振舞い 実験評価 Experiment and Evaluation 經価方法·平均正解發見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題 妥当性の考察 Threats to Validity

結論と会後の課題 Conclusion and Future Work

従来のクローンレポート静的洗練法

Static Refinement of Clone Report in Other Researches

| ID | 内容 | Bug? |
|----|-----|------|
| 1 | AAA | ? |
| 2 | BBB | ? |
| 3 | CCC | Χ |
| 4 | DDD | ? |
| 5 | EEE | ? |
| 6 | FFF | ? |
| 7 | III | Χ |
| | ••• | ••• |

2012 年7日4日(火)

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction 背景 Introduction 既存研究 Related Researches

成行研究 Related Researches
正解の例 Example of True Positive
誤検出の例 Example of False Positive
不具合があるクローンの誤検出率
コードクローンの四つの象限
クローンに基づく不具合検出

クローンに基づく不具合検出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ 洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine

特徴抽出: 構文木を構築 特徴抽出: 五の特徴 特徴抽出: 五のの特徴

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の結果 Results of the Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利博 (pp-3 Information Gain 順番老調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

従来のクローンレポート静的洗練法

Static Refinement of Clone Report in Other Researches

| ID | 内容 | Bug? |
|----|-----|------|
| 1 | AAA | ? |
| 2 | BBB | ? |
| 3 | CCC | Χ |
| 4 | DDD | ? |
| 5 | EEE | ? |
| 6 | FFF | ? |
| 7 | III | Χ |
| | | ••• |

| ID | 内容 | Bug? |
|----|-----|----------|
| 1 | AAA | √ |
| 2 | BBB | Χ |
| | | |
| 4 | DDD | ✓ |
| 5 | EEE | Χ |
| 6 | FFF | Χ |
| | | |
| | ••• | |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 電景 Introduction

r景 Introduction 背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率 コードクローンのの象別

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ 洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7日4日(火)

クローンレポートの動的洗練法

Dynamic Refinement of Clone Report

| ID | 内容 | Bug? |
|----|-----|------|
| 1 | AAA | |
| 2 | BBB | |
| 3 | CCC | |
| 4 | DDD | |
| 5 | EEE | |
| 6 | FFF | |
| 7 | III | |
| | ••• | ••• |

Active Refinement of Clone

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

卷号 Introduction 既存研究 Related Researches

正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳栓出率 コードクローンの四つの参照

クローンに基づく不具合輸出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的法練法の流れ 洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例

前処理:特徵選択 前机理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算 具体的な洗練の振舞い

事験評価 Experiment and Evaluation

輕価方法·平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation

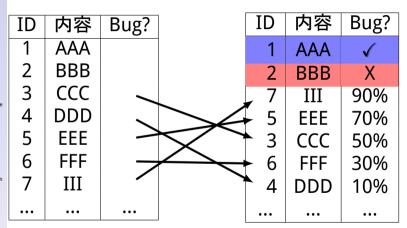
実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

クローンレポートの動的洗練法 **Anomaly Reports**

Dynamic Refinement of Clone Report



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

出典 Publication

背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの部体出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の溶れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出:五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類は・類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation

実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題

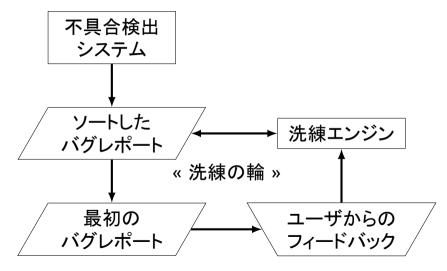
祭, 結画及び今後の課題 妥当性の考察 Threats to Validity

結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

動的洗練法の流れ

Active Refinement Process



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

出典 Publication

背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの語検出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

助的光線法 Uynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の溶れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法・類似度の計算

万規法: 規模度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation

実験の結果 Results of the Empirical Evaluation
Top-3 ff報利得 Top-3 Information Gain

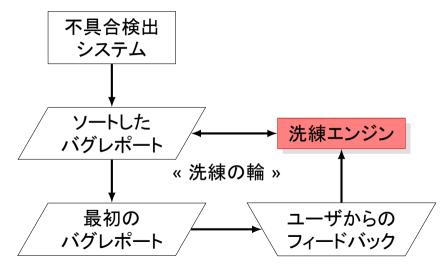
順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

動的洗練法の流れ

Active Refinement Process



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

出典 Publication

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具会があるクローンの経緯出塞

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

洗練エンジン Refinement Engine

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

背景 Introduction 背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出: 不つの特徴

特徴抽出: 五つの特徴の例前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法・平均正解発量率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

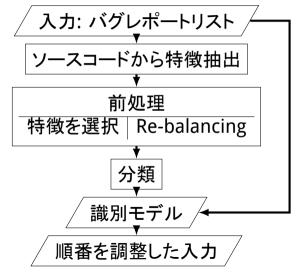
妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Fur

結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

洗練エンジンの流れ

Process of Refinement Engine



大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
出典 Publication
背景 Introduction
育景 Introduction
育芸 Introduction
現子研究 Related Researches
正解の例 Example of True Positive
諸検出の例 Example of Size Positive
不具合があるジローンの類検出率
コージアロー・パロマハの架内
```

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

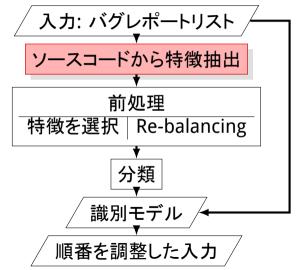
順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

洗練エンジンの流れ

Process of Refinement Engine



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの語検出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

助的光線法 Uynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の溶れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

ディストがにされた版も近 分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報刊得 Top-3 Information Gain

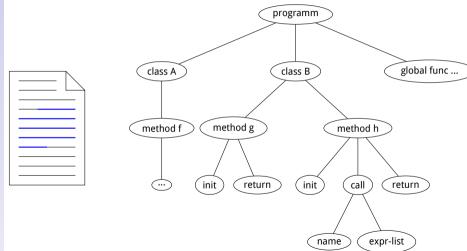
順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

乗、和酬及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

特徴抽出: 構文木を構築

Feature Extraction: Tree Constraction



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの超棒出室

コードクローンの四つの象限 クローンに基づく不具合検出

プローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

映評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation

実験の結果 Results of the Empirical Evaluation Top-3 ff報利得 Top-3 Information Gain

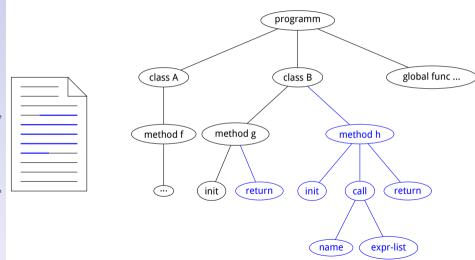
順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

特徴抽出: 構文木を構築

Feature Extraction: Tree Constraction



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 譲検出の例 Example of False Positive 不具合があるクローンの超棒出座

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の溶れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

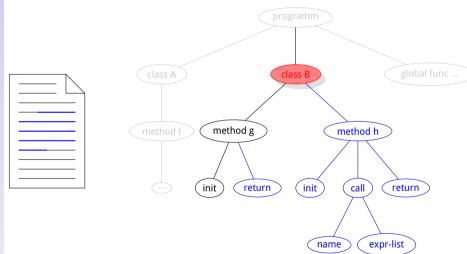
10p-3 Intermation Gain 順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

特徴抽出: 構文木を構築

Feature Extraction: Tree Constraction



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication
出典 Publication
背景 Introduction
背景 Introduction
病子の例 Example of Faise Positive
正解の例 Example of Faise Positive
不具合があるクローンの譲模出第
コードクローンの選つの象別
ウローンに基づて果合検出
クローンに基づて果合検出
複雑のクローンはボーケの影響が接続する
「後春のクローンはボーケの影響が

クローンレポートの動的洗練法 動的洗練法の流れ 洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解竟見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 顕番老期整上仍便 Example of Re-ordering

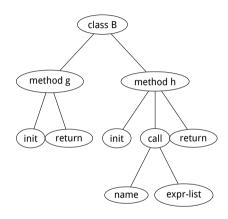
考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

特徴抽出: 五つの特徴

Feature Extraction: 5 features

基本的な特徴 Basic Features 対の特徴 Pair Features 基本的な特徴の割合 Proportional Features—Basic 対の特徴の割合 Proportional Features—Pair 他の特徴 Other Features



大阪大学大学院 CS 専攻 楊 嘉晨

常豊 Introduction 育豊 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 訓検出の例 Example of False Positive 不具合があるシローンの訓検出年 コードワローンの図つの象別 クローンに基づく不具合検出 クローンに基づく不具合検出

出典 Publication
出典 Publication

動的洗練法 Dynamic Refinement 従来のクローンレポートの幹的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine
洗練エンジンの流れ Process of Refinement Engine

洗練エンシンの流れ Process of Refinement Eng 特徴抽出: 構文木を構築 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公額は・類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報判得 Top-3 Information Gain 顕番老期整上近例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

特徴抽出: 五つの特徴

Feature Extraction: 5 features

基本的な特徴 Basic Features

対の特徴

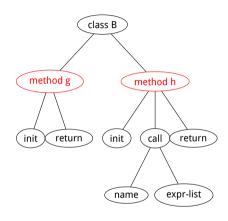
Pair Features

基本的な特徴の割合

Proportional Features―Basic 対の特徴の割合

Proportional Features―Pair 他の特徴

Other Features



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 背景 Introduction 背景 Introduction 背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 損検出の例 Example of False Positive 不具合があるウローンの動検出率 コードクローンの図つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

出典 Publication

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 禁機抽出: 五つの特徴

特徴抽出: 五つの特徴の例前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公額は: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解竟見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 顕番老期整上仍便 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012 年 7 目 4 日(少)

特徴抽出: 五つの特徴

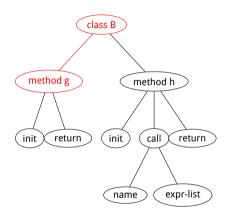
Feature Extraction: 5 features

基本的な特徴 Basic Features 対の特徴

Pair Features

基本的な特徴の割合
Proportional Features—Basic
対の特徴の割合
Proportional Features—Pair
他の特徴

Other Features



大阪大学大学院 CS 専攻

出典 Publication 出典 Publication 背景 Introduction 卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳栓出率 コードクローンの四つの象別 クローンに基づく不具合輸出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ 洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例 前処理:特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 公頼は、頼似度の計算

具体的な洗練の振舞い 実験評価 Experiment and Evaluation 經価方法·平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題

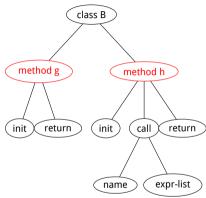
妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

特徴抽出: 五つの特徴

Feature Extraction: 5 features

基本的な特徴 Basic Features 対の特徴 Pair Features 基本的な特徴の割合 Proportional Features—Basic 対の特徴の割合 Proportional Features—Pair 他の特徴 Other Features



大阪大学大学院 CS 専攻

出典 Publication 出典 Publication 背景 Introduction 卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳栓出率 コードクローンの四つの象別 クローンに基づく不具合輸出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ 洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例 前処理:特徵選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公頼は、頼似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 經価方法·平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

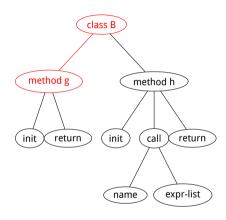
者容、結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

特徴抽出: 五つの特徴

Feature Extraction: 5 features

基本的な特徴 Basic Features 対の特徴 Pair Features 基本的な特徴の割合 Proportional Features—Basic 対の特徴の割合 Proportional Features—Pair 他の特徴 Other Features



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 背景 Introduction 背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳栓出率 コードクローンの四つの参照 クローンに基づく不具合輸出 クローンに基づく不具合検出

出典 Publication

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine

特徴抽出: 構文木を構築 特徴抽出・五つの特徴

特徴抽出: 五つの特徴の例 前処理:特徵選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公頼は、頼似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 經価方法·平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

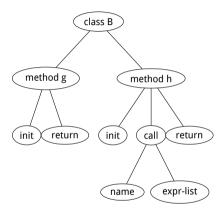
者容、結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012 年7日4日(火)

特徴抽出: 五つの特徴

Feature Extraction: 5 features

基本的な特徴 **Basic Features** 対の特徴 Pair Features 基本的な特徴の割合 Proportional Features—Basic 対の特徴の割合 Proportional Features—Pair 他の特徴 Other Features



クローンの数 $|\mathsf{CG}|$, クローン の平均サイズ $\frac{\sum_{\mathsf{c}\in\mathsf{CG}}|\mathsf{C}|}{|\mathsf{CG}|}$

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction 昨五研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択 前処理: Data Re-balancing

前処理: Data Re-Dalancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 fin模利制 Flory 3 Information Gain 顕番老期報上方例 Example of Re-ordering

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7日4日(火)

者容、結論及び今後の課題

特徴抽出: 五つの特徴の例

Feature Extraction: Example of 5 features

decode_sattr3(p, &args—>attrs)

decode_filename(p, &args ->tname, &args ->tlen)





| 基本 | | 対 | | | | 対割台 | | 他 | |
|-----------|---|---------------------|---|-----------|------|---------------------|------|-----|---|
| タイプ | | タイプ | | タイプ | | タイプ | | タイプ | |
| | 2 | Call/Name | 2 | | | Call/Name | | Num | 2 |
| Name | 2 | Call/ Expr- list | 2 | Name | 100% | Call/ Expr- list | 100% | Avg | |
| Expr-list | 2 | Expr-list/ Expr | 2 | Expr-list | 100% | Expr-list/ Expr | 100% | | |
| Expr | 2 | | | Expr | 100% | | | | |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction 既存研究 Related Researches

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 fin模利制 Flory 3 Information Gain 顕番老期報上方例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

粕調と予恢の課題 Conclusion and Full

2012年7月4日(火)

特徴抽出: 五つの特徴の例

Feature Extraction: Example of 5 features

decode_sattr3(p, &args->attrs)

decode_filename(p, &args->tlen)





| 基本 | | 対 | | 基本書 | | 対割台 | | 他 | |
|-----------|---|---------------------|---|-----------|------|---------------------|------|-----|---|
| タイプ | | タイプ | | タイプ | | タイプ | | タイプ | |
| | 2 | Call/Name | 2 | | | Call/Name | | Num | 2 |
| Name | 2 | Call/ Expr- list | 2 | Name | 100% | Call/ Expr- list | 100% | Avg | |
| Expr-list | 2 | Expr-list/ Expr | 2 | Expr-list | 100% | Expr-list/ Expr | 100% | | |
| Expr | 2 | | | Expr | 100% | | | | |

大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
出典 Publication
背景 Introduction
背景 Introduction
既存研究 Related Re
```

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの説検出率 コードクローンの四つの参照

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

クローンレポートの動的洗練法動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation

実験では、平均正解発見車 APPF 実験の対象と設定 Settings of Empirical Evaluation 天験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering 来転 結論の 150 金砂 四脚脚

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

特徴抽出: 五つの特徴の例

Feature Extraction: Example of 5 features

decode_sattr3(p, &args->attrs)

decode_filename(p, &args->tname, &args->tlen)





| 基本 | | 対 | | 基本害 | 引合 | 対割合 | ì | 他 | |
|-----------|---|---------------------|---|-----------|------|---------------------|------|-----|-----|
| タイプ | 数 | タイプ | 数 | タイプ | 割合 | タイプ | 割合 | タイプ | 値 |
| Call | 2 | Call/Name | 2 | Call | 100% | Call/Name | 100% | Num | 2 |
| Name | 2 | Call/ Expr- list | 2 | Name | 100% | Call/ Expr- list | 100% | Avg | 5.5 |
| Expr-list | 2 | Expr-list/ Expr | 2 | Expr-list | 100% | Expr-list/ Expr | 100% | | |
| Expr | 2 | - | | Expr | 100% | • | | | |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誠検出の例 Example of False Positive 不具合があるクローンの総検出座

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 結微抽出: 五つの特徴

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法・節似度の計算

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法・平均正解発量率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

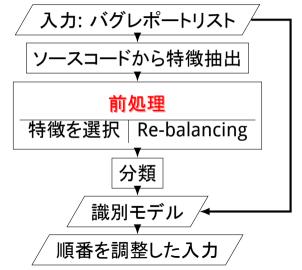
順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

洗練エンジンの流れ (再掲)

Process of Refinement Engine (Again)



大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
出典 Publication
```

原 Introduction 背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誠検出の例 Example of False Positive 不具合があるクローンの誠検出率

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例前処理: 特徴選択

削処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Éxample of Re-ordering 考察、結論及び令後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

前処理: 特徵選択

Preprocessing: Feature Selection

- c はクローンのクラスラベル, f は一つの特徴
 - ・ 正解は ve クラス, 誤検出は -ve クラス

情報利得 (Information Gain) は:

$$IG(c|f) = H(c) - H(c|f)$$
 (1)

$$H(c) = -\sum_{c_i \in \{\pm ve\}} P(c_i) \log P(c_i)$$
 (2)

$$H(c|f) = -\sum_{c_i \in \{\pm ve\}} P(c_i|f) \log P(c_i|f)$$
 (3)

情報利得を基づいて Weka(Holmes et al., 1994) を用いて特徴選択

G. Holmes, A. Donkin, and I. Witten, ``Weka: A machine learning workbench," in Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on. Ieee, 1994, pp. 357–361.

大阪大学大学院 CS 専攻

出典 Publication 出典 Publication

書号 Introduction 卷号 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳検出率

コードクローンの四つの象別 クローンに基づく不具合輸出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的法練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出・五つの特徴

特徴抽出: 五つの特徴の例

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公頼は、頼似度の計算

具体的な洗練の振舞い

影験評価 Experiment and Evaluation

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

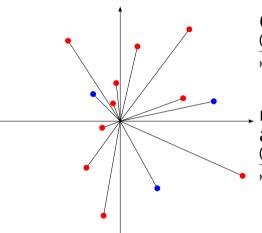
順番を調整した例 Example of Re-ordering 容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

前処理: Data Re-balancing

Preprocessing: Data Re-balancing



Cosine-similarity (Kantardzic, 2011)

M. Kantardzic, Data mining: concepts, models, methods, and algorithms. Wiley-IEEE Press, 2011

nearest neighbor approach (Renieres and Reiss, 2003)

M. Renieres and S. Reiss, ``Fault localization with nearest neighbor queries." in Automated Software Engineering, 2003, Proceedings, 18th IEEE International Conference on, IEEE, 2003. pp. 30--39.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

育景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 調検出の例 Example of False Positive

誤検出の例 Example of False Posit 不具合があるクローンの誤検出率 コードクローンの四つの象限

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の変わ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

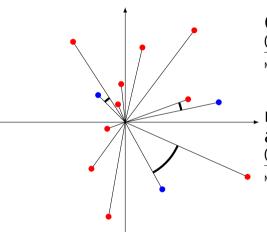
順番を調整した例 Example of Re-ordering 背察, 結論及び今後の課題 妥当性の考察 Threats to Validity

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

前処理: Data Re-balancing

Preprocessing: Data Re-balancing



Cosine-similarity (Kantardzic, 2011)

M. Kantardzic, Data mining: concepts, models, methods, and algorithms. Wiley-IEEE Press, 2011.

nearest neighbor approach (Renieres and Reiss, 2003)

(Renieres and Reiss, 2003)

M. Renieres and S. Reiss, ``Fault localization with nearest neighbor queries," in Automated Software Engineering, 2003. Proceedings. 18th IEEE International Conference on. IEEE, 2003, pp. 30–39.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 旺解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの経緯出室

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ。

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出:五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 整備方法・平均正解発見来 APPE

新加力法: 平均正所来見年 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

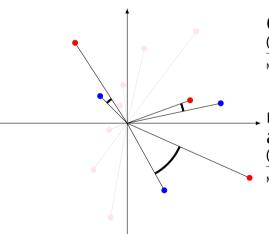
順番を調整した例 Example of Re-ordering

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

前処理: Data Re-balancing

Preprocessing: Data Re-balancing



Cosine-similarity (Kantardzic, 2011)

M. Kantardzic, Data mining: concepts, models, methods, and algorithms. Wiley-IEEE Press, 2011.

nearest neighbor approach (Renieres and Reiss, 2003)

M. Renieres and S. Reiss, ``Fault localization with nearest neighbor queries," in Automated Software Engineering, 2003. Proceedings. 18th IEEE International Conference on. IEEE, 2003, pp. 30–39.

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳検出率

コードクローンの四つの象別 クローンに基づく不具合輸出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徵選択

前机理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 經価方法·平均正解發見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

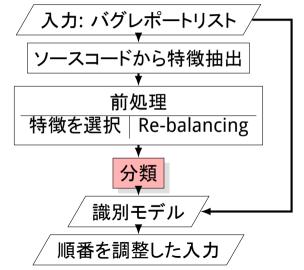
順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012 年7日4日(火)

洗練エンジンの流れ(再掲)

Process of Refinement Engine (Again)



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

計典 Publication

計画 Introduction

対象 Introduction

Intr

クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象を設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報刑得 Top-3 Information Gain 調番を想撃した例 Example of Re-ordering

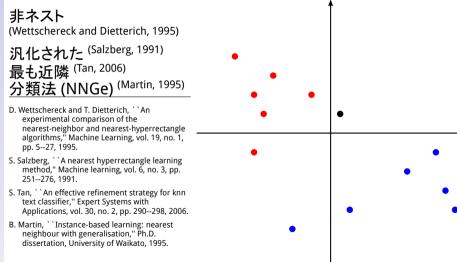
考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2042 5 7 7 4 7 (1)

2012年7月4日(火)

非ネスト汎化された最も近隣分類法

NNGe: Nearest Neighbor Classification with Non-Nested Generalization



大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 計典 Publication 計量 Introduction 行者 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 訓検出の例 Example of False Positive 不具合があるクローンの選択 コードクローンに基づて不具合検出 クローンに基づて不具合検出

従来のクローンレポートの静的洗練法

クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine

特徴抽出: 構文木を構築 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例 前処理: 特徴選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算 具体的な洗練の振舞い

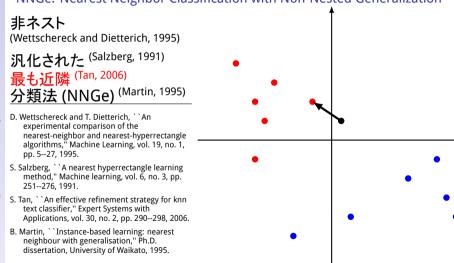
実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象と記定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報刑得 Top-3 Information Gain 調番を認即と元例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7日4日(火)

非ネスト汎化された最も近隣分類法

NNGe: Nearest Neighbor Classification with Non-Nested Generalization



大阪大学大学院 CS 専攻 楊 嘉晨

出身で助icitation 出身を助icitation 背景 Introduction 現存研究 Related Researches 正解の所 Example of True Positive 諸機性の例 Example of False Positive 不具合があるクローンの回の部機は第 コードクローンの回の部機は第 フローンに基づて具合検出 クローンに基づて具合検出 グローンに基づて具合検出 後来のクローンはボーケの影響が連絡

クローンレポートの動的洗練法 動的洗練法の流れ 洗練エンジン Refinement Engine

た練エンシン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine

特徴抽出: 構文木を構築 特徴抽出: 五つの特徴

特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象を設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報刑得 Top-3 Information Gain 調番を想撃した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7日4日(火)

非ネスト汎化された最も近隣分類法

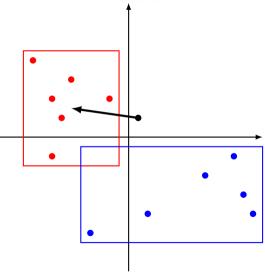
NNGe: Nearest Neighbor Classification with Non-Nested Generalization

(Wettschereck and Dietterich, 1995)

汎化された (Salzberg, 1991)
最も近隣 (Tan, 2006)
分類法 (NNGe) (Martin, 1995)

非ネスト

- D. Wettschereck and T. Dietterich, ``An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms," Machine Learning, vol. 19, no. 1, pp. 5–27, 1995.
- S. Salzberg, ``A nearest hyperrectangle learning method," Machine learning, vol. 6, no. 3, pp. 251--276, 1991.
- S. Tan, ``An effective refinement strategy for knn text classifier," Expert Systems with Applications, vol. 30, no. 2, pp. 290--298, 2006.
- B. Martin, ``Instance-based learning: nearest neighbour with generalisation," Ph.D. dissertation, University of Waikato, 1995.



大阪大学大学院 CS 専攻 楊 嘉晨

出奏 Publication 出奏 Publication 青豊 Introduction 青貴 Introduction 青貴 Introduction 東京 Introduction 既存研究 Related Researches 振特の例 Example of False Positive 不具命があるクローンの週の意報出 コードクローンの週のの意報 フローンに基づ不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 結微抽出: 五つの特徴

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法: 平均正解発見率 APPF 実験の対象を設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報刑得 Top-3 Information Gain 調番を想撃した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7日4日(火)

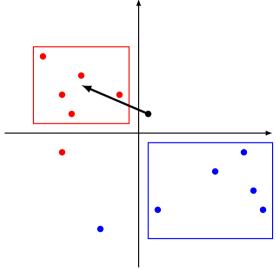
非ネスト汎化された最も近隣分類法

NNGe: Nearest Neighbor Classification with Non-Nested Generalization

非ネスト (Wettschereck and Dietterich, 1995)

汎化された ^(Salzberg, 1991) 最も近隣 ^(Tan, 2006) 分類法 (NNGe) ^(Martin, 1995)

- D. Wettschereck and T. Dietterich, ``An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms," Machine Learning, vol. 19, no. 1, pp. 5–27, 1995.
- S. Salzberg, ``A nearest hyperrectangle learning method," Machine learning, vol. 6, no. 3, pp. 251--276, 1991.
- S. Tan, ``An effective refinement strategy for knn text classifier," Expert Systems with Applications, vol. 30, no. 2, pp. 290--298, 2006.
- B. Martin, ``Instance-based learning: nearest neighbour with generalisation," Ph.D. dissertation, University of Waikato, 1995.



大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
出典 Publication
背景 Introduction
存成 Filter True Positive
近待郊泉 Related Researches
近接他の例 Example of False Positive
将具合があるクローンの路検出率
コードクローンの四つの泉泉
```

クローンに基づく不具合検出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine

特徴抽出: 構文木を構築 特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徴選択 前処理: Data Re-balancing 非ネスト汎化された量も近端分類法

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 整価方法・平均正解発見率 APPE

評価方法: 平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

28/40

分類法: 類似度の計算

Classification: Calculation of Likelihood

$$\begin{split} \text{LH}(\text{dp}) &= 0.5 + \frac{RS(\text{dp})}{2} \\ RS(\text{dp}) &= \frac{|\sum_{\text{d}_T \in D_T} \text{sim}(\text{d}_p, \text{d}_T)|}{D_T} \\ &- \frac{|\sum_{\text{d}_F \in D_F} \text{sim}(\text{d}_p, \text{d}_F)|}{D_F} \\ \text{sim}(\text{d}_p, \text{d}) &= 1 - \text{dist}(\text{d}_p, \text{d}) \\ \text{dist}(\text{d}_p, \text{d}) &\in [0, 1] \end{split}$$

LH(dp) によって Weka の NNGe 分類法を用いて分類

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 譲検出の例 Example of False Positive 不具合があるクローンの誤検出率 コードクローンのの参照

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法・類似度の計算

分類法: 類似度の計算 具体的な洗練の振舞い

実験評価 Experiment and Evaluation 認価方法: 平均正解発見率 APPE

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 顯番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity

妥当性の考察 Inreats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

29/40

具体的な洗練の振舞い

Concrete Refinement Process

| _ |
|-----|
| |
| |
| - 1 |
| |
| |
| _ |
| |
| |
| _ |
| |
| |
| _ |
| |
| |
| _ |
| |
| |
| _ |
| |
| |
| - 1 |
| |
| |
| _ |
| |
| |
| _ |
| |
| |
| _ |
| |
| |
| _ |
| |
| |
| _ |
| |
| _ |
| - 1 |
| |
| |
| - 1 |
| |
| _ |
| - 1 |
| |
| _ |
| _ |
| |
| _ |
| |
| |
| _ |
| _ |
| |
| |
| - 1 |
| |
| |
| - 1 |
| |
| |
| |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

背景 Introduction 背景 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの超検出率

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の溶れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法: 平均正解発見率 APPF

評価方法: 半马止解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity

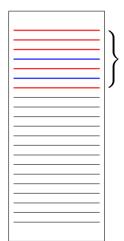
妥当性の考察 Inreats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

2012年/月4日(火

具体的な洗練の振舞い

Concrete Refinement Process



最初の k 個のレポート, 正解を含む

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

卷号 Introduction

既存研究 Related Researches 正解の例 Example of True Positive 据検出の例 Example of False Positive 不具合があるクローンの脳検出率

コードクローンの四つの象限 クローンに基づく不具合輸出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 公頼は・頼似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 經価方法:平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

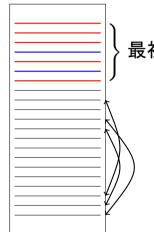
者容、結論及び今後の課題 妥当性の考察 Threats to Validity

結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

具体的な洗練の振舞い

Concrete Refinement Process



最初の k 個のレポート, 正解を含む

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive

不具合があるクローンの脳栓出率 コードクローンの四つの象限 クローンに基づく不具合輸出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的体練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 公頼は・頼似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation

經価方法·平均正解發見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

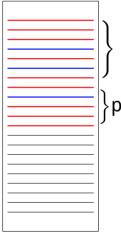
者容、結論及び今後の課題 妥当性の考察 Threats to Validity

結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

具体的な洗練の振舞い

Concrete Refinement Process



最初の k 個のレポート, 正解を含む

р 個のフィードバックプール

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction

卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 据検出の例 Example of False Positive 不具合があるクローンの脳検出率

コードクローンの四つの参照 クローンに基づく不具合輸出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例

前処理: 特徵選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分籍法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation

經価方法·平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

者容、結論及び今後の課題 妥当性の考察 Threats to Validity

結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

具体的な洗練の振舞い

Concrete Refinement Process

| } 最初の k 個のレポート, 正解を含む |
|--------------------------|
| }p 個のフィードバックプール |
| |
| . ! |

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

雪景 Introduction

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの超検出室

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法

動的洗練法の流れ 洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法・類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法: 平均正解発見率 APPF

評価方法: 平均止所発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 書容. 結論及び今後の課題

妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

実験評価 Experiment and Evaluation

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction 卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳検出率 コードクローンの四つの象別 クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的法練法の流れ

洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出・五つの特徴

特徴抽出: 五つの特徴の例

前机理: Data Re-balancing 公頼は、頼似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

者容、結論及び今後の課題 妥当性の考察 Threats to Validity

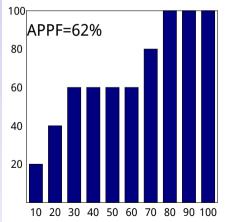
結論と今後の課題 Conclusion and Future Work

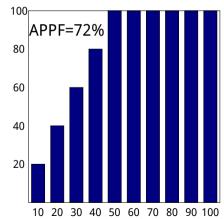
2012 年7日4日(火)

評価方法: 平均正解発見率

Evaluation Criteria: APPF: Average Percentage true Positives Found テストケースの優先順位付け領域から借りた概念.

平均障害発見率 (APFD: Average Percentage Faults Detected).





大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
出典 Publication
電景 Introduction
```

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Posit

取行研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率 コードクローンの四つの象限

クローンに基づく不具合検出 クローンに基づく不具合検出

助的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

従来のクローンレポートの静的洗練 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 経価方法: 平均正解発見率 APPF

計画力法: 平均正辨明見率 APPP 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

Top-3 情報利得 Top-3 Information Gain 順番を調整した例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

実験の対象と設定

Settings of Empirical Evaluation

| 名前 | バージョン | 不具合数 | 正解数 | k | p |
|---------|----------|------|-----|----|---|
| Linux | 2.6.19 | >800 | 57 | 50 | 1 |
| Eclipse | 20070108 | >400 | 38 | 50 | 1 |
| ArgoUML | | >50 | 15 | 10 | 1 |

大阪大学大学院 CS 専攻 楊 嘉晨

```
出典 Publication
  出典 Publication
背景 Introduction
```

卷号 Introduction 既存研究 Related Researches

正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳検出率 コードクローンの四つの象限

クローンに基づく不具合輸出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出・五つの特徴

特徴抽出: 五つの特徴の例 前処理:特徵選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公頼は・頼似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 經価方法·平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題

妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012 年7日4日(火)

実験の結果

Results of the Empirical Evaluation

| 名前 | APPF | Top-5 順番の調整 |
|---------|------|---|
| Linux | 11% | $694 \mapsto 18$, $672 \mapsto 64$, $760 \mapsto$ |
| | | $131,770 \mapsto 179,792 \mapsto 206$ |
| Eclipse | 87% | $373 \mapsto 4$, $348 \mapsto 11$, $394 \mapsto$ |
| • | | $29,388 \mapsto 43,370 \mapsto 49$ |
| ArgoUML | 86% | $40 \mapsto 12, 35 \mapsto 15, 34 \mapsto 11,$ |
| - | | $29 \mapsto 9$, $23 \mapsto 8$ |

Linux の結果は低いのは、この手法では改名に関する バグが取れない.

大阪大学大学院 CS 専攻 楊 嘉晨

| 出典 Publication |
|--|
| 出典 Publication |
| 音 Introduction |
| 背景 Introduction |
| 既存研究 Related Researches |
| 正解の例 Example of True Positive |
| 誤検出の例 Example of False Positive |
| 不具合があるクローンの誤検出率 |
| コードクローンの四つの象限 |
| ローンに基づく不具合検出 |
| クローンに基づく不具合検出 |
| 的洗練法 Dynamic Refinement |
| 従来のクローンレポートの静的洗練法 |
| クローンレポートの動的洗練法 |
| 動的洗練法の流れ |
| た練エンジン Refinement Engine |
| 洗練エンジンの流れ Process of Refinement Eng |
| 特徴抽出: 構文木を構築 |
| 特徴抽出: 五つの特徴 |
| 特徴抽出: 五つの特徴の例 前処理: 特徴選択 |
| 前処理: 特殊抵抗 前処理: Data Re-balancing |
| 前処理: Data Re-Dalancing 非ネスト汎化された最も近隣分類法 |
| 分類法: 類似度の計算 |
| 具体的な洗練の振舞い |
| 医験評価 Experiment and Evaluation |
| 評価方法: 平均正解発見率 APPF |
| 実験の対象と設定 Settings of Empirical Evaluat |
| 実験の結果 Results of the Empirical Evaluation |
| Top-3 情報利得 Top-3 Information Gain |
| 順番を調整した例 Example of Re-ordering |
| 背察, 結論及び今後の課題 |
| 妥当性の考察 Threats to Validity |
| 結論と今後の課題 Conclusion and Future Work |

Top-3 Information Gain

Top-3 情報利得

| Тор | 特徴 | 情報利得 | | | | |
|-----|---|----------|--|--|--|--|
| | Linux kernel | | | | | |
| 1 | extern_ definition ^P | 0.015941 | | | | |
| 2 | extern_ definition _1 ^P | 0.015941 | | | | |
| 3 | program##extern_ definitions ^P | 0.015941 | | | | |
| | Eclipse | | | | | |
| 1 | BOOL_OR_TK ^P | 0.01898 | | | | |
| 2 | conditional_ or_ expression ## conditional_ or_ ex- | 0.01898 | | | | |
| | pression ^P | | | | | |
| 3 | BOOL_OR_TK ^B | 0.01898 | | | | |
| | ArgoUML | | | | | |
| 1 | local_variable_declaration_statement ^B | 0.145772 | | | | |
| 2 | variable_ initializer ^B | 0.145772 | | | | |
| 3 | block_ statement ## local_ variable_ declaration_ | 0.145772 | | | | |
| | statement ^B | | | | | |

2012年7月4日(火)

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

出典 Publication 背景 Introduction 卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive 不具合があるクローンの脳栓出率 コードクローンの四つの象限 クローンに基づく不具合輸出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的体練法の流れ 洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築 特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例 前処理:特徵選択 前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 公頼は・頼似度の計算 具体的な洗練の振舞い 実験評価 Experiment and Evaluation 經価方法·平均正解發見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering

順番を調整した例: Linux 696 → 18

Example of Re-ordering: Linux $694 \mapsto 18$

```
drivers/net/wireless/bcm43xx/bcm43xx wx.c
drivers/net/wireless/hcm43xx/
                                                      615 struct bcm43xx private *bcm = bcm43xx priv(net dev):
bcm43xx sysfs.c
                                                      616
     struct bcm43xx private *bcm = dev to bcm(dev);
     mutex lock(&(bcm)—>mutex);
     switch (bcm43xx current phy(bcm)—>type) {
                                                      618 mutex lock(&bcm->mutex):
 350
         case BCM43xx PHYTYPE A:
                                                          mode = bcm43xx current radio(bcm)—>interfmode:
 351
                                                      620
                                                          mutex_unlock(&bcm -> mutex):
                                                      621
                                                          switch (mode) {
                                                      622
                                                              case BCM43xx RADIO INTERFMODE NONE:
                                                      623
 362
         default:
 363
            assert(0):
 364
 365
     mutex unlock(&(bcm)->mutex):
                                                      632
                                                              default:
                                                                 assert(0):
```

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction 背景 Introduction 既左研究 Related I

育景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの誤検出率 コードクローの全層

クローンに基づく不具合検出 クローンに基づく不具合検出

動的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法

従来のクローンレポートの静的洗練: クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 fin模利制 Flory 3 Information Gain 顕番老期報上方例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

2012年7月4日(火)

順番を調整した例: Eclipse 373 → 4

Example of Re-ordering: Eclipse $373 \mapsto 4$

 ${\tt debug/internal/core/LaunchConfiguration.java}$

debug/internal/core/

 $Launch {\tt ConfigurationWorkingCopy.java}$

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication

```
出典 Publication
書号 Introduction
  番号 Introduction
  既存研究 Related Researches
  正解の例 Example of True Positive
  編検出の例 Example of False Positive
  不具合があるクローンの脳検出率
  コードクローンの四つの象限
クローンに基づく不具合輸出
 クローンに基づく不具合検出
動的洗練法 Dynamic Refinement
  従来のクローンレポートの静的洗練法
  クローンレポートの動的洗練法
  動的法練法の流れ
洗練エンジン Refinement Engine
  洗練エンジンの流れ Process of Refinement Engine
  特徴抽出: 構文木を構築
  特徴抽出・五つの特徴
  特徴抽出: 五つの特徴の例
  前処理:特徵選択
  前処理: Data Re-balancing
  非ネスト汎化された最も近隣分類法
  分類法: 類似度の計算
  具体的な洗練の振舞い
影験評価 Experiment and Evaluation
  經価方法·平均正解発見率 APPF
  実験の対象と設定 Settings of Empirical Evaluation
  実験の結果 Results of the Empirical Evaluation
  Top-3 情報利得 Top-3 Information Gain
  順番を調整した例 Example of Re-ordering
5容. 結論及び今後の課題
  妥当性の考察 Threats to Validity
```

結論と会後の課題 Conclusion and Future Work

順番を調整した例: ArgoUML 40 → 12

Example of Re-ordering: ArgoUML $40 \mapsto 12$

```
argouml/uml/diagram/UMLMutableGraphSupport.java

331
if (edge instanceof CommentEdge) {
...
332
...
333
} else if (Model.getFacade().isARelationship(edge)
334
|| Model.getFacade().isATransition(edge)
|| Model.getFacade().isAAssociationEnd(edge))
{
return Model.getUmlHelper().getDestination(edge)

;
} else if (Model.getFacade().isALink(edge)) {
...
338
...
339
}
```

argoum 1/um 1/diagram/UML Mutable Graph Support.java

```
if (edge instanceof CommentEdge) {
362
     else if (Model.getFacade().isAAssociation(edge)) {
363
        List conns = new
364
        ArrayList (Model.getFacade().getConnections(edge)
365
        return conns.get(1):
366
    } else if (Model.getFacade().isARelationship(edge)
367
               Model.getFacade().isATransition(edge)
368
               Model.getFacade().isAAssociationEnd(edge))
369
        return Model.getUmlHelper().getDestination(edge)
    } else if (Model.getFacade().isALink(edge)) {
```

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication

背景 Introduction

背景 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 誤検出の例 Example of False Positive 不具合があるクローンの超棒出車

コードクローンの四つの象限 クローンに基づく不具合検出

クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の溶れ

洗練エンジン Refinement Engine

洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例

特徴抽出: 五つの特徴の伊 前処理: 特徴選択

前処理: 符飯選択 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation

評価方法: 平均正解発見率 APPF

実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 考察、結論及び今後の課題

条、結構及いう後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

考察,結論及び今後 の課題

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication 出典 Publication 背景 Introduction 卷号 Introduction 既存研究 Related Researches 正解の例 Example of True Positive 編検出の例 Example of False Positive

不具合があるクローンの脳検出率

コードクローンの四つの象別 クローンに基づく不具合輸出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的体練法の流れ

洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出・五つの特徴 特徴抽出: 五つの特徴の例 前処理: Data Re-balancing

非ネスト汎化された最も近隣分類法 公頼は・頼似度の計算 具体的な洗練の振舞い

事験評価 Experiment and Evaluation 經価方法·平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 情報利得 Top-3 Information Gain

順番を調整した例 Example of Re-ordering 者容、結論及び今後の課題 妥当性の考察 Threats to Validity 結論と会後の課題 Conclusion and Future Work

妥当性の考察(と質問)

Threats to Validity (and Questions)

- APPF は APFD から借りた概念
 - 他に使われていない
- 本当に Bug であるかとうかは検証していない
 - 人間で目で見て判断した
 - 後のバージョンに修正したかを検証していない (質問)
- 3つのシステムしかない
 - Cと Java しかない,他の言語に対応できるか?
 - 古いバージョンを使った原因は?(質問)
 - Linux と Eclipse の実験の一部は 2007 に発表した研究。 今回追加したのは ArgoUML.
 - Long-term-service ではないから、バグが多いかも

大阪大学大学院 CS 専攻 楊 嘉晨

出典 Publication
出典 Publication
背景 Introduction
背景 Introduction
育秀 Introduction
既存研究 Related Researches
正解の例 Example of True Positive
誘致出の例 Example of False Positive
不思点があるフローンの競争用塞

コードクローンの四つの象限 クローンに基づく不具合検出 クローンに基づく不具合検出 動的洗練法 Dynamic Refinement

助的洗練法 Dynamic Refinement 従来のクローンレポートの静的洗練法 クローンレポートの動的洗練法 動的洗練法の流れ

洗練エンジン Refinement Engine 洗練エンジンの流れ Process of Refinement Engine 特徴抽出: 構文木を構築

特徴抽出: 五つの特徴 特徴抽出: 五つの特徴の例 前処理: 特徴選択

前処理: Data Re-balancing 非ネスト汎化された最も近隣分類法 分類法: 類似度の計算

具体的な洗練の振舞い

実験評価 Experiment and Evaluation 評価方法:平均正解発見率 APPF 実験の対象と設定 Settings of Empirical Evaluation 実験の結果 Results of the Empirical Evaluation Top-3 fin模利制 Flory 3 Information Gain 顕番老期報上方例 Example of Re-ordering

考察, 結論及び今後の課題 妥当性の考察 Threats to Validity 結論と今後の課題 Conclusion and Future Work

2012年7月4日(火)

結論と今後の課題

Conclusion and Future Work

結論

- ・クローンバグレポートの動的な洗練法を提案した
- ・この方法について,評価方法 APPF を採用された
- Linux, Eclipse, ArgoUML の3つのシステムで実験を行った。
 - ・ 結果はそれぞれ 11%, 87%, 86%

今後の課題

- 他のソフトウェアに対応してみたい
- バグレポート以外の不具合レポートに対応してみたい

大阪大学大学院 CS 専攻 楊 嘉晨

2012年7月4日(火)

41/40

私の研究との比較

Comparing with My Research

| | この手法 | 私の研究 |
|------|---------------|-------------------|
| 着目点 | バグを不具合ク | 人によって判断の |
| | ローンから見つけ | 差異 |
| クローン | 構文木 Type-3 | トークン列 Type-2 |
| 学習特徴 | 木の Node, Edge | トークンの型の tf- |
| | の数と割合 | idf |
| 分類法 | 非ネスト汎化された | Cosine-Similarity |
| | 最も近隣 NNGe | の重み付き平均数 |
| 評価方法 | 平均正解発見率 | 訓練集合 -正確率 |
| | APPF | の図 |

大阪大学大学院 CS 専攻 楊 嘉晨

2012年7月4日(火)

42/40

特徴抽出: 五つの特徴の定義

Feature Extraction: Defination of 5 features

定義: 基本的な特徴 (Basic Features)

(t, |CS|), 若し $CS = \{c \in CG | cl ct タイプが存在する\} \land |CS| > 0$

定義: 対の特徴 (Pair Features)

定義: 基本的な特徴の割合 (Proportional Features―Basic)

 $(t, \frac{|CS|}{|CG|})$, 若し $CS = \{c \in CG | clctタイプが存在する\} \land |CS| > 0$

定義: 対の特徴の割合 (Proportional Features―Pair)

 $(t_1,t_2,\frac{|\mathsf{CS}|}{|\mathsf{CG}|})$, 若し $\mathsf{CS}=\{\mathsf{c}\in\mathsf{CG}|\exists_{\mathsf{n}_1,\mathsf{n}_2\in\mathsf{c}}\;\mathsf{n}_1\;\mathsf{t}\;\mathsf{n}_2\;\mathsf{t}$ は連結されている \land n_1

のタイプは $\mathbf{t}_1 \wedge \mathbf{n}_2$ のタイプは $\mathbf{t}_2 \} \wedge |\mathsf{CS}| > 0$

定義: 他の特徴

クローンの数 |CG|, クローンの平均サイズ $\frac{\sum_{c \in G} |C|}{|CG|}$

大阪大学大学院 CS 専攻 楊 嘉晨

2012年7月4日(火)

43/40

参考文献I

Reference

- E. Juergens, F. Deissenboeck, B. Hummel, and S. Wagner, ``Do code clones matter?" in Proceedings of the 31st International Conference on Software Engineering. IEEE Computer Society, 2009, pp. 485--495.
- L. Jiang, Z. Su, and E. Chiu, ``Context-based detection of clone-related bugs," in ESEC/FSE, vol. 2007, 2007.
- M. Gabel, J. Yang, Y. Yu, M. Goldszmidt, and Z. Su, ``Scalable and systematic detection of buggy inconsistencies in source code," in ACM Sigplan Notices, vol. 45, no. 10. ACM, 2010, pp. 175--190.
- L. Jiang, G. Misherghi, Z. Su, and S. Glondu, ``Deckard: Scalable and accurate tree-based detection of code clones," in Proceedings of the 29th international conference on Software Engineering. IEEE Computer Society, 2007, pp. 96--105.
- G. Holmes, A. Donkin, and I. Witten, ``Weka: A machine learning workbench," in Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on. Ieee, 1994, pp. 357--361.
- M. Kantardzic, Data mining: concepts, models, methods, and algorithms. Wiley-IEEE Press, 2011.
- M. Renieres and S. Reiss, ``Fault localization with nearest neighbor queries," in Automated Software Engineering, 2003. Proceedings. 18th IEEE International Conference on. IEEE, 2003, pp. 30--39.



大阪大学大学院 CS 専攻 楊 嘉晨

2012年7月4日(火)

44/40

参考文献 II

Reference

- D. Wettschereck and T. Dietterich, ``An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms," Machine Learning, vol. 19, no. 1, pp. 5--27, 1995.
- S. Salzberg, ``A nearest hyperrectangle learning method," Machine learning, vol. 6, no. 3, pp. 251--276, 1991.
- S. Tan, ``An effective refinement strategy for knn text classifier," Expert Systems with Applications, vol. 30, no. 2, pp. 290--298, 2006.
- B. Martin, ``Instance-based learning: nearest neighbour with generalisation," Ph.D. dissertation, University of Waikato, 1995.