

## アソシエーション分析に基づく再検査基準の決定手法の提案

- Guidelines by Association Threshold Evaluation(GATE)による再検査基準設定 -

### A Proposal of a Decision Method for Standards of Re-inspection based on

#### Association Analysis

- Setting Reinspection Criteria Using Guidelines by Association Threshold Evaluation (GATE) -

リーダー：片桐 汐駿（アズビル株式会社）

研究員：若林 俊弥（株式会社日立システムズ）

越智 恭平（アズビル株式会社）

荻原 美樹（アズビル株式会社）

主査：喜多 義弘（長崎県立大学）

副主査：秋山 浩一（株式会社日本ウィルテックソリューション）

アドバイザー：西田 尚弘（株式会社日新システムズ）

#### 研究概要

本研究では、ドキュメント検査における再検査基準を定量的に決定する手法として、Guidelines by Association Threshold Evaluation (GATE) を提案する。従来、再検査基準は不良密度に基づき経験則で設定されていたが、しきい値の根拠が曖昧であり、品質良と判定されたにもかかわらず事故が発生する事例があった。GATE は、組織内の過去プロジェクトに共通する欠陥パターンをアソシエーション分析により抽出し、関連性の高い欠陥組を基準として再検査の要否を判断する。実験では、過去の欠陥データを用いてパターン抽出と検証を行い、分析用データと検査用データに共通して、関連性の高い欠陥組があることを確認した。

**Abstract** This study proposes Guidelines by Association Threshold Evaluation (GATE) as a method for quantitatively determining reinspection criteria in document inspection. Traditionally, reinspection criteria have been set empirically based on defect density; however, the rationale for the threshold has been unclear, and there have been cases in which failures occurred despite documents being judged as having acceptable quality. GATE extracts defect patterns common across past projects within an organization using association analysis and determines the necessity of reinspection based on highly related defect combinations. In the experiments, defect patterns were analyzed and validated using historical defect data, and it was confirmed that certain highly related defect combinations were consistently observed in both the training dataset and the validation dataset.

#### 1. はじめに

本研究では、ドキュメント検査の再検査基準を定量的に決定する手法を提案する。

ドキュメント検査とは、設計部門が作成し、設計レビューの終了したドキュメントを検査部門が最終的にチェックする<sup>[1]</sup>プロセスのことである。ドキュメント検査の主な目的は以下の2点である。

- ・ 設計仕様書の内容を第三者が客観的に評価することで、一定の設計品質を確保し、品質保証の観点から予防的な管理を行う。
- ・ ソフトウェア開発において工程ごとの区分が曖昧になりやすいことから、各設計工程の成果物を一つずつ確認し、各工程の完了を客観的に判断する。

## 第41年度（2025年度）ソフトウェアテストコース（Team GATE）

研究員の現場では、ドキュメント検査は基本設計フェーズ・詳細設計フェーズの計2回実施される。それぞれに不良密度を指標値として、十分に検査が実施されたかどうか判定している。ここでいう不良密度とは、ドキュメント100ページ当たりの欠陥数である。この不良密度が5.0未満であれば、品質良と判断している。しかし、このしきい値は算出方法が定められておらず、経験則で決められている。そのため、現在の基準の妥当性を説明できない。

また、独立行政法人情報処理推進機構(IPA)により、業界の過去の不良密度データ<sup>[2]</sup>は公開されている。しかし、このデータが表すのは業界の平均であり、ドキュメント検査の質を評価する数値ではない。よって、IPAのデータをもとにしきい値を決定するのは不適切である。したがって、ドキュメント検査の再検査基準を決定する、新たな方法が求められる。

そこで本研究では、経験則によらない再検査基準を決定するための手法として、Guidelines by Association Threshold Evaluation(GATE)を提案する。具体的には、アソシエーション分析を活用して、再検査基準と再検査箇所を提示することができる手法である。

## 2. 関連研究

### 2.1 欠陥パターン

細川ら<sup>[3]</sup>は、ソフトウェアの欠陥の組合せ、つまり複数の欠陥が同時に発生する「共起構造」に着目し、潜在的な欠陥を推定する手法を提案している。欠陥パターンとは、この共起構造を特徴とする欠陥の集合を指す。この手法は、潜在している欠陥を効率的に除去し、後工程への欠陥流出抑制につながるという点で有用である。我々は、この手法から着想を得て、欠陥のパターンに着目することで、欠陥検出が十分であるか判定して再検査基準としたり、潜在している欠陥を検出できるように再検査箇所を提示したりすることができるのではないかと考えた。

### 2.2 アソシエーション分析

再検査基準や再検査箇所を明確にするためには、欠陥パターンを構成する欠陥間の共起関係を定量的に扱う必要がある。この共起関係を数値化する手法として、アソシエーション分析に着目した。アソシエーション分析とは、Agrawalらによって提案された、複数の事象間の関連性を抽出する手法である<sup>[4][5][6]</sup>。関連性を抽出したい事象をA、Bとしたとき、アソシエーション分析で多く用いられる評価指標である支持度(support)、信頼度(confidence)、リフト値(lift)の定義を以下に示す。

$$\text{support}(A, B) = P(A \cap B)$$

$$\text{confidence}(A, B) = P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\text{lift}(A, B) = \frac{P(B|A)}{P(B)}$$

ドキュメント検査で指摘された欠陥AとBを例に、アソシエーション分析について述べる。ドキュメント検査で指摘された検査項目の組に対して、全プロジェクトの件数をN、欠陥Aが検出されたプロジェクトの件数を $N_A$ 、欠陥AとBの両方が検出されたプロジェクトの件数を共起回数 $N_{A \cap B}$ としたとき、欠陥Aの発生確率 $P(A)$ および欠陥AとBの同時発生確率 $P(A \cap B)$ は、以下の式で表される。

$$P(A) = \frac{N_A}{N}$$

$$P(A \cap B) = \frac{N_{A \cap B}}{N}$$

## 第41年度（2025年度）ソフトウェアテストコース（Team GATE）

支持度は欠陥 A と B が同時に発生した確率を示し、欠陥 A と B の組合せがドキュメント検査でどの程度頻繁に現れているかを表す。支持度が高い場合は、その組合せが頻繁に現れていることを意味し、分析対象とすべき組合せの抽出に利用できる。

信頼度は、欠陥 A の発生を前提としたときに欠陥 B がどれだけの確率で同時に発生しているかを表す。信頼度を用いることで「欠陥 A が発生したとき欠陥 B も発生する」といった方向性のある関連性に着目した分析が可能となる。

リフト値は 1 をしきい値として関連性の有無を明確に判断できる。リフト値が 1 より大きいときは、 $P(B|A) > P(B)$  であり、「欠陥 A が検出されたプロジェクトにおいて欠陥 B が検出される確率」が、「全てのプロジェクトで欠陥 B が発生する確率」よりも高い。言い換えれば、欠陥 A が見つかった時点で欠陥 B の有無に注目して再検査することが、効率的な欠陥検出につながることを意味する。このとき、欠陥 A と B の間には正の関連があると言える。一方で、リフト値が 0 以上 1 未満のときは欠陥 A と B の間には負の関連があり、リフト値が 1 のときは関連がない。

以上の性質に着目すると、ドキュメント検査で指摘された欠陥の組合せについて、支持度、信頼度、リフト値を評価し、それらが一定値以上であれば、再検査基準と再検査箇所の提示に活用可能な欠陥パターンとなるのではないかと考えた。

### 3. 提案

あるプロジェクトのドキュメント検査の結果が与えられたときに、そのプロジェクト内の未検出の欠陥をアソシエーション分析により推定し、再検査基準と再検査箇所を提示する手法として、我々は GATE を提案する。

#### 3.1 仮説

我々は、欠陥パターンを用いて、ドキュメントの再検査基準と再検査箇所を決定できるのではないかと考えた。また、欠陥パターンの発見には、アソシエーション分析が有効なのではないかと考えた。

以上から、我々の RQ は以下となる。

『RQ：アソシエーション分析で発見した欠陥パターンを用いることで、未検出の欠陥を含むドキュメントの再検査基準と再検査箇所を決定できるのではないか？』

#### 3.2 提案手法

GATE は、組織内で共通の欠陥パターンをもとに再検査基準と再検査箇所を決定する手法である。共通の欠陥パターンを発見する方法には、アソシエーション分析の関連性指標である支持度と信頼度とリフト値を用いる。

ドキュメント検査で指摘された欠陥 A と B を例に GATE の手順を示す。

##### (1) データの収集

過去のプロジェクトのドキュメント検査で検出された欠陥データを集める。

##### (2) 過去の欠陥データの事前処理

過去のプロジェクトのドキュメント検査で検出された欠陥を、検査項目ごとに分類する。分類は、ドキュメント検査のプロセスを定めた文献<sup>[1]</sup>を参考にする。

さらにプロジェクトごとに、各分類の欠陥の数を「検出された」/「検出されていない」という 2 値に変換する。検出された欠陥数が 1 件以上の分類は「検出された」、検出された欠陥数が 0 件の場合は、「検出されていない」と判定する。

##### (3) 支持度と信頼度とリフト値の計算

過去のプロジェクトのドキュメント検査で指摘された全ての検査項目の組合せに対して、支持度と信頼度とリフト値を計算する。

## 第41年度（2025年度）ソフトウェアテストコース（Team GATE）

### (4) 欠陥パターン候補の抽出

(3)で算出したリフト値と支持度に基づいて、欠陥パターン候補を特定する。本手法では、以下の2条件を満たす検査項目の組の一覧を欠陥パターン候補とする。

$$support(A, B) \geq 0.05$$

$$lift(A, B) > 1$$

### (5) 欠陥パターンの特定

(4)で算出した欠陥パターン候補の信頼度に基づいて、欠陥パターンを特定する。本手法では、以下の条件を満たす検査項目の組の一覧を欠陥パターンとする。

$$confidence(A, B) \geq 0.5$$

これにより得られた欠陥パターンを、「A→B」のように表記する。

### (6) 検査対象のドキュメントの評価

(5)で得られた欠陥パターンについて、進行中のプロジェクトのドキュメント検査結果において欠陥Aが検出されているが欠陥Bが検出されていない場合は、再検査とする。その際、再検査箇所として欠陥Bに着目して再検査するように検査担当者に指示する。

## 4. 評価

### 4.1 実験概要

本論文では、アソシエーション分析が欠陥パターンの発見に有効であることを示すための実験を行う。実験に使用する欠陥データは、研究員が所属する組織における過去プロジェクトから収集したものである。欠陥データは9:1の割合でランダムに分割し、分析用データと検査用データとして利用する。なお、ここで用いる過去プロジェクトのデータについては、従来の方法にて欠陥の検出が済んでいるデータとして扱う。この条件下で、アソシエーション分析の有効性は以下の手順により評価する。

- (1) 分析用データにおける全ての検査項目の組合せについて、支持度と信頼度とリフト値を算出する。
- (2) 算出した支持度とリフト値に基づき、欠陥パターン候補を抽出する。
- (3) 算出した信頼度に基づき、欠陥パターンを特定する。
- (4) 特定した欠陥パターンについて、検査用データにおけるリフト値と信頼度を算出する。
- (5) 特定した欠陥パターンのリフト値、信頼度に基づきアソシエーション分析の有効性を評価する。

### 4.2 実験準備

本実験では、研究員が所属する組織で特定の期間行われたドキュメント検査における欠陥のデータを実験対象とする。欠陥の例を表1に示す。1つのプロジェクトに対しては複数の欠陥が記録されている。記録されている欠陥は、ドキュメント検査を実施した担当者が表1に示した事象分類表のうち適切な事象を記録している。本実験データは、120件分のプロジェクトで行われたドキュメント検査の検査記録であり、計415回の欠陥が発生しているデータを利用する。なお、同じプロジェクト内で発生した同じ欠陥は一つとして集計する。120件のプロジェクトに対して9:1の割合で分割するため、108件を分析用データ、12件を検査用データとなるようにランダムに分割した。

表 1. 欠陥データ例

プロジェクト ID	ドキュメント検査項目
REQ-101	機能の整合性
REQ-101	ドキュメント間の整合性
REQ-101	形式誤り, 記述不統一, 用語不統一
REQ-102	ドキュメント間の整合性
REQ-102	形式誤り, 記述不統一, 用語不統一
REQ-103	形式誤り, 記述不統一, 用語不統一
REQ-104	誤字, 脱字, 文法誤り
:	:
REQ-220	誤字, 脱字, 文法誤り

### 4.3 実験結果

図 1 は、各検査項目の組合せにおける共起回数を示したものであり、共起が発生したすべての検査項目の組合せについて、共起回数をマトリックス形式で表している。図 1 より、共起回数が最も多い検査項目の組合せは、「記述誤り」と「誤字, 脱字, 文法誤り」であり、108 件中 34 件（31%）のプロジェクトにおいて同時に発生していることが分かる。

次に、共起が発生している検査項目の組合せに対してリフト値を算出した結果を図 3 に示す。図 2 は、共起が確認された欠陥パターンのみについてリフト値を算出し、関連性の強さを示したマトリックスである。図 3 より、リフト値が最も高い検査項目の組合せは、「機能の整合性」と「形式誤り, 記述不統一, 用語不統一」であり、リフト値は 2.27 であることが分かる。

分析用データにおいて、支持度が 0.05 以上(本実験においては共起回数が 6 回以上)かつリフト値が 1 より大きい、欠陥パターン候補を表 2 に示す。発見した欠陥パターン候補のうち、信頼度が 0.5 以上の欠陥パターンを表 3 に示す。本実験では、6 組の欠陥パターンを発見した。

検査用データにおいて、発見した欠陥パターンに対して、支持度、リフト値、信頼度を算出した結果を表 4 に示す。

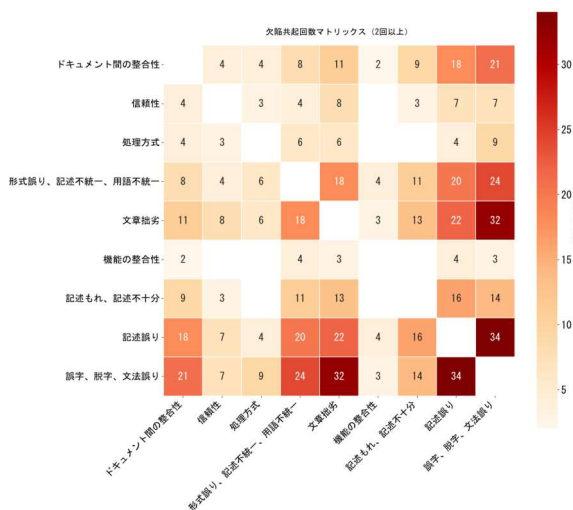


図 1. 共起回数マトリックス(分析)

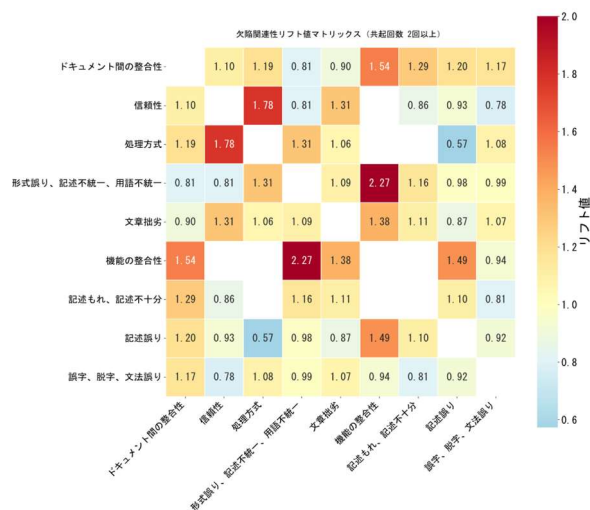


図 2. リフト値マトリックス(分析)

第41年度（2025年度）ソフトウェアテストコース（Team GATE）

表 2. 欠陥パターン候補

No	欠陥 $\alpha$	欠陥 $\beta$	$support(\alpha, \beta)$	$lift(\alpha, \beta)$	$confidence(\alpha, \beta)$	$confidence(\beta, \alpha)$
1	信頼性	文章拙劣	0.07	1.31	<u>0.57</u>	0.17
2	処理方式	形式誤り, 記述不統一, 用語不統一	0.05	1.31	0.46	0.16
3	ドキュメント間の整合性	記述もれ, 記述不十分	0.08	1.29	0.32	0.33
4	ドキュメント間の整合性	記述誤り	0.15	1.20	<u>0.64</u>	0.31
5	ドキュメント間の整合性	誤字, 脱字, 文法誤り	0.18	1.17	<u>0.75</u>	0.30
6	形式誤り, 記述不統一, 用語不統一	記述もれ, 記述不十分	0.09	1.16	0.29	0.41
7	文章拙劣	記述もれ, 記述不十分	0.11	1.11	0.28	0.48
8	記述もれ, 記述不十分	記述誤り	0.13	1.10	<u>0.59</u>	0.28
9	形式誤り, 記述不統一, 用語不統一	文章拙劣	0.15	1.09	0.47	0.38
10	処理方式	誤字, 脱字, 文法誤り	0.08	1.08	<u>0.69</u>	0.13
11	文章拙劣	誤字, 脱字, 文法誤り	0.27	1.07	<u>0.68</u>	0.46
12	処理方式	文章拙劣	0.05	1.06	0.46	0.13

表 3. 欠陥パターン

No	欠陥パターン
1	「信頼性」 → 「文章拙劣」
2	「ドキュメント間の整合性」 → 「記述誤り」
3	「ドキュメント間の整合性」 → 「誤字, 脱字, 文法誤り」
4	「記述もれ, 記述不十分」 → 「記述誤り」
5	「処理方式」 → 「誤字, 脱字, 文法誤り」
6	「文章拙劣」 → 「誤字, 脱字, 文法誤り」

表 4. 検査用データの欠陥パターンに対するアソシエーション分析結果

No	欠陥 $\alpha$	欠陥 $\beta$	$support(\alpha, \beta)$	$lift(\alpha, \beta)$	$confidence(\alpha, \beta)$
1	信頼性	文章拙劣	0.01	4.00	1.00
4	ドキュメント間の整合性	記述誤り	0.01	2.00	1.00
5	ドキュメント間の整合性	誤字, 脱字, 文法誤り	0.01	1.50	1.00
8	記述もれ, 記述不十分	記述誤り	0.04	1.25	0.63
10	処理方式	誤字, 脱字, 文法誤り	0.03	1.50	1.00
11	文章拙劣	誤字, 脱字, 文法誤り	0.02	1.00	0.67

### 5. 考察

#### 5.1 実験結果の考察

本研究の実験結果に対して、考察を次に示す。

表4に示す欠陥パターンのうち、『「文章拙劣」→「誤字，脱字，文法誤り」』を除く5組の欠陥パターンにおいて、1よりも高いリフト値となることを確認した。『「文章拙劣」→「誤字，脱字，文法誤り」』でリフト値が1以下になった原因は、分析用データで算出したリフト値が1.07と、極めて1に近い値であったことだと考えられる。リフト値が極めて1に近い場合、関連が極めて弱いといえるため、欠陥パターンとして不適切である可能性がある。仮に欠陥パターン候補を導出する際の分析用データのリフト値のしきい値を1.1以上とした場合、特定した全ての欠陥パターンにおいて、検査用データでも1より高いリフト値となることが確認できた。したがって、本実験では、リフト値が1より高いことを欠陥パターン抽出の条件にしたが、実際の現場で適用する場合には、リフト値のしきい値を高い値に設定することで、より精度の高い欠陥パターンの抽出が可能となると考えられる。

表4に示す欠陥パターンのうち、全ての欠陥パターンの信頼度について0.5より大きいことを確認した。つまり、GATEを適用して再検査基準と再検査箇所を提示したとき、その提示が正しい確率が50%より大きいことを意味する。

以上から、分析用データで得られた欠陥パターンが、検査用データにおいても高いリフト値、信頼度となっていることから、ドキュメント検査での欠陥パターンの抽出において、アソシエーション分析が有効であることを示すことができた。アソシエーション分析で見出した欠陥パターンをもとに、3.2節で述べた評価を実施することで、経験則によらない再検査基準の決定に繋がると考えられる。具体例として研究員の現場を想定すると、従来の基準では、検査項目「処理方式」において1件の不良が抽出され、不良密度が5.0未満である場合に品質良と判断されていた。これに対し、GATEを利用することで、後工程において手戻りや認識の曖昧さに起因して不良が発生する可能性のある「形式誤り」「記述不統一」「用語不統一」等について、抽出漏れの有無を過去のプロジェクトに基づいて説明可能となる。

#### 5.2 今後の展望

本研究では、データを研究員が手作業で集計したため、利用可能なデータ数が少なかった。そのため統計的仮説検定を行うことができず、実験結果が統計的に有意な結果であることは示せなかった。GATEの有効性を統計的に示すためには、実験データを増やす必要がある。また、アソシエーション分析の精度を高めるためにも、データ数は多い方が望ましい。集計作業の負荷をかけずに大量のデータを扱うために、例えば欠陥指摘を記録するシステムを構築して、構造化されたデータとして管理する方法も検討したい。

また本研究ではGATEの有効性を検証するために過去データのみを用いて数値解析を行ったが、今後は実際の開発現場にGATEを導入した場合の効果を確認する必要がある。そのため、進行中のプロジェクトにGATEを適用し、ドキュメント検査における再検査基準として妥当であるかを検証することが課題である。

また、本研究は再検査を実施するか否かの判断基準を提示することを目的としており、再検査を通じてドキュメントの品質をどのように向上させるかについては検討できていない。今後は、再検査時に重点的に確認すべき検査項目の絞り込みにGATEを活用するなど、再検査によるドキュメント品質向上の手法について検討していきたい。さらに、参考文献では欠陥の要因に着目しているのに対し、本研究ではプロジェクト間に共通する欠陥パターンのみに着目しており、欠陥の背景にある要因については考慮していなかった。今後は本手法を発展させるとともに、要因ごとに共通する欠陥パターンを分析し、その結果について考察を行う必要がある。

### 6. おわりに

本研究では、従来経験則に依存していたドキュメント検査の再検査基準を、客観的かつ定量的に導出する手法として Guidelines by Association Threshold Evaluation (GATE) を提案した。GATEは、過去プロジェクトにおける欠陥の共起パターンをアソシエーション分析によって抽出し、支持度と信頼度とリフト値を用いて欠陥間の関連性を評価することで、再検査の要否を判断する基準を提供するものである。

実験では、組織内で収集された実データを用いてGATEの有効性を検証した。その結果、分析用データと検査用データにおいて、リフト値が1を超える検査項目の組合せが高い割合で一致し、共通の欠陥パターンが組織内に存在する可能性を確認した。

#### 第41年度（2025年度）ソフトウェアテストコース（Team GATE）

一方で、本研究は欠陥パターンの抽出と再検査判定に特化しており、再検査実施後の品質向上プロセスや欠陥要因の分析までは扱っていない。また、実際のプロジェクト運用における効果測定も今後の課題である。

以上より、GATE は組織におけるドキュメント検査の品質保証プロセスを強化し、未検出欠陥の流出リスクを低減するための有効な手法であると考えられる。今後は、現在進行中のプロジェクトへの適用やデータ収集のシステム化により、さらに精度の高い再検査基準の確立を目指す。

#### 参考文献

- [1] 保田 勝通, ソフトウェア品質保証の考え方と実際, 70 - 83, 1995
- [2] 独立行政法人情報処理推進機構, ソフトウェア開発 分析データ集 2022, 2022
- [3] 細川 宣啓, ソフトウェア欠陥の共起性を用いた欠陥推定手法の提案, 2016年度 SQiP 研究会第7分科会
- [4] Rakesh Agrawal et al., Mining association rules between sets of items in large databases, ACM SIGMOD Record, Volume 22, Issue 2, Pages 207 - 216, 1993
- [5] Monden, A., Keung et al., A heuristic rule reduction approach to software fault-proneness prediction, In Proceedings of the 19th Asia-Pacific Software Engineering Conference (APSEC2012), Pages 17 - 24, 2012
- [6] 伊藤 晃ら, アソシエーション分析における可視化を用いた 興味深いルールの探索, 6th Fuzzy System Symposium, 2010