

テキスト構造に着目した細分化モデルの可読性評価指標

Readability Metrics for Sliced Models Focused on Text Structure

結城 翔†
Sho Yuki

織田 健†
Takeshi Oda

1 はじめに

ソフトウェアの大規模化や複雑化に伴う開発コストの増大に対処するため、形式手法が注目されている。我々は形式手法の一つである B Method を用いた、既存ソフトウェアの部品再利用による形式的ソフトウェア合成手法 (MSSS 手法) を提案している [1]。MSSS 手法において必要な部品を再利用できず不足した部品は人が記述するが、文字列一致による細分化モデルの検索を可能にするために字面統一が施された式は人には難解な表現になる。そこで、不足部品の式の可読性向上手法が提案されたが [2]、この提案は式の可読性評価指標が厳密に定まっていなかったなどの課題があった。よって本稿では従来手法でも注目していた式の行数や演算子の数などの特徴に加え、識別子の数やその長さ、使用する演算子の種類や数式の構文木の深さといったテキスト特徴のセットを提案する。さらに、提案した各テキスト特徴が式の可読性に与える影響の大きさを考慮し、総合的に式の可読性を評価する計算式を提案する。

2 背景と目的

2.1 B Method と MSSS 手法

B Method [3] は集合論と一階述語論理に基づく形式手法の一種で、無矛盾なモデルを記述し、それを段階的に詳細化することにより最終的な実装の整合性を保証する。MSSS 手法 [1] は B Method を用いた、ソフトウェア部品の再利用の自動化を行う手法である。この手法ではモデルとして与えられた要求を細分化し、それらをキーとして字面一致する部品を検索する。得られた候補から、結合可能な部品を選択・結合することで要求モデルを満たす実装を出力する。該当する部品が無い場合は不足部品を人が記述する。MSSS 手法におけるモデル細分化は、部品の粒度を細かくし、部品の再利用性の向上と数学的に等価なまま文字列での部品検索を可能とする目的がある。しかし、部品検索を可能するために字面統一が施された式は、人には難解な表現になる。

2.2 不足部品の可読性向上手法

不足部品は字面統一が施され難解な表現で書かれている為、不足部品の可読性向上手法を提案してきた [2]。この手法では可能な限り要求モデルの式を再現するために、字面統一後の変数に注目した制約条件抽出などを行った。図 1 はこの手法の大まかな流れである。

字面統一後の変数に注目した制約条件抽出

字面統一後の式に含まれる変数に注目し、その変数のみを使用している元のモデルの制約条件を細分化前の式から抽出する。それらに対し字面統一を施す。

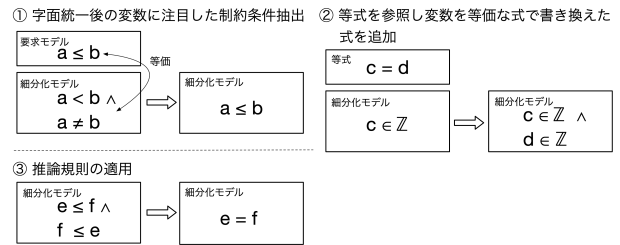


図 1: 従来手法の流れ

等式による式の書き換え

要求モデルと細分化モデルに含まれる等式を抽出し、その等式を構成する定数・変数・集合 (以降これらを項と呼ぶ) を含む細分化モデルの式に対して、等価な項により式を書き換え細分化モデルに追記する。

推論規則の適用による書き換え

B Method の開発環境である Atelier B の定理証明器で用いられる 2208 個の規則のうち、常に書き換えが成立するような 444 個の推論規則を用いて可読性向上のための推論規則を整備する。その後細分化モデルの式群に対して整備した推論規則を適用する。

2.3 先行研究における課題

従来手法における可読性評価指標を以下に示す。

- なるべく元のモデルの式を使用している。
- 式の行数が少ない。
- 使用される項の数が少ない。
- 使用される項の重要度が高い。

ここで「項の重要度」とは、以下の仮定に基づいて計算される。

- 等式の右辺より左辺のほうが項の重要度が高い。
- 項のうちリテラルの重要度は低い。

しかし、右辺より左辺の項の重要度が高いという指標は、根拠に乏しく説得力に欠けるものであった。また、式の行数が少ないほど可読性が高いという指標も、構文木のサイズが小さい 2 行の式とそれを構文木のサイズが大きい 1 行の式に書き換えた場合でどちらが読みやすいかの議論の余地があった。

2.4 Scallabrino らによる可読性向上手法

ソフトウェアの可読性向上に関して、テキスト構造に着目した Scallabrino らによる研究 [4] が存在する。この研究は 5,000 人以上の人々によって 600 の指標に基づき可読性を評価させる実験を行っており、実験を通じて整理されたテキスト構造とそれらが可読性に与える関係のセットは信憑性の高いものであると言える。

2.5 研究目的

2.2 節の先行研究における可読性評価指標は不完全で、厳密に定まっていなかったという課題があった。そこで本研究では、従来手法において整備されていた式の行数や演

†電気通信大学大学院情報理工学専攻

表 1: 可読性評価に用いるテキスト特徴

特徴 (変数名)	影響
行の長さ (len), 識別子の数 (id_N), 括弧の数 (par), 行数 ($line$), 構文木の深さ (dep)*	大
識別子の長さ (id_L)**, 数値の数 (lit), 演算子の数 (op)	小

* 式の構文木の深さの平均. ** 全識別子の文字数の平均.

算子の数、数式の構文木の深さなどのテキスト特徴に加え、Scallabrino らによる研究を参考に識別子の長さなどを追加したテキスト特徴のセットを考え、それらを用いた可読性評価の計算式を提案する。

3 テキスト構造に着目した細分化モデルの可読性評価指標

3.1 提案手法の概要

本稿では Scallabrino らによる研究 [4] で整備されていたテキスト構造の一部と従来手法 [2] で用いていた可読性評価指標を組み合わせたテキスト特徴、及びそれらを使用した可読性評価計算式を定義し、可読性が異なるとされる複数のモデルに対し計算を適用する実験を行う。

3.2 可読性評価に用いるテキスト特徴

表 1 は可読性評価に用いるテキスト特徴と、各特徴が可読性に与える影響の大きさである。可読性に与える影響の大きさは、Scallabrino らによる研究において「大」と分類されているものと従来手法 [2] で用いているものを大、それ以外を小とした。

Scallabrino らによる研究は一般のプログラムソースが可読性向上の対象であった一方、本手法では B Method で書かれたモデルを可読性向上の対象としている。そのため Scallabrino らによる研究で参照されていたテキスト特徴のうち、B Method のモデルには存在しない「ループの数」などいくつかの特徴は本手法の対象としなかった。一方で「行の長さ」や「識別子の数」といった二つの研究対象に共通する特徴の他に、従来手法でも参照していた「数式の構文木の深さ」と、Scallabrino らによる研究で参照されており B Method で書かれたモデルの可読性評価計算に用いることができる「識別子の長さ」を可読性評価の計算に用いることとした。

3.3 可読性評価計算式

表 1 をもとに、可読性評価計算式 (1) を定義した。ここで、 R の値が大きいほど可読性が高いことを表す。

$$R = \frac{1}{2(len + id_N + par + line + dep)} + \frac{1}{id_L + lit + op} \quad (1)$$

ここで、表 1 において影響「大」としている特徴は係数を 2 としている。この係数は今後より具体的に決定する。式 1 により、図 2 と図 3 に示す等価なモデルの可読性を評価し計算した。計算の結果、モデル 1 は $R = 0.0808$ 、モデル 2 は $R = 0.0648$ となり、モデル 1 の方が可読性が高いことが示された。

```
MODEL ManageStudent
VARIABLES
  alpha1Class, alpha2Class, alpha3Class
OPERATIONS
  ret <- select_excellent_student =
    PRE
      alpha1Class V alpha2Class V alpha3Class != {}
    THEN
      ret := alpha1Class V alpha2Class V alpha3Class
    END
END
```

図 2: モデル 1

```
MODEL Mch
ABSTRACT_VARIABLES
  av001, av002, av003
OPERATIONS
  op001 <- operation =
    PRE
      not(av001 : POW({})) or
      not(av002 : POW({})) or
      not(av003 : POW({}))
    THEN
      op001 := av001 V av002 V av003
    END
END
```

図 3: モデル 2

4 考察

本手法では可読性評価計算式において影響「大」となるテキスト特徴の係数を決めうちで 2 としたが、今後はこの係数の妥当性や各テキスト特徴の影響度の分類の妥当性を検証する必要がある。また、極端に省略された名称の識別子と通常の単語からなる識別子の読みやすさの差など、識別子の長さが短いほど可読性が向上するという指標には議論の余地がある。今後は識別子が辞書に登録されている単語を含むかなども、可読性評価指標を決定する上で考慮する必要がある。さらに、論理積だけではなく論理和で区切られた部分も行端とするなど、細分化モデルの行数の定義を厳密に行う必要がある。

5 終わりに

本稿では細分化モデル中で使用する識別子の数や長さ、演算子の種類や構文木の深さといったテキスト特徴を示し、それらを用いて総合的に式の可読性を評価する計算式を提案した。今後は計算式や細分化モデルの行数の定義をより厳密に行い、実験規模を拡大する必要がある。

参考文献

- [1] 中村 丈洋. B Method における部品再利用によるソフトウェア合成と高信頼ソフトウェア部品の整備. 電気通信大学 電気通信学研究科 博士 (工学) 学位論文, 2014
- [2] 結城 翔, 織田 健. 形式的ソフトウェア合成手法における細分化モデルの可読性向上手法. 情報処理学会第 85 回全国大会講演論文集, vol.1 pp.275-276, (2023.03)
- [3] 来間 啓伸. B メソッドにおける形式仕様記述. 近代科学社. 2007
- [4] Simone Scallabrino, Mario Linares-Vasques, Denys Poshyvanyk, and Rocco Oliveto. *Improving Code Readability Models with Textual Features*. University of Molise Pesche Italy, The College of William and Mary Williamsburg Virginia USA, 2016.