

# ツール評価(1)

5K-2

## - 評価方法論 -

菅野 聰美 吉田 ゆき 手島 文彰

(株式会社東芝 システム・ソフトウェア技術研究所)

### 1. はじめに

ソフトウェア開発効率の向上や品質向上を目的としてCASEをはじめとするさまざまな開発支援ツールが導入されはじめている。ツールの導入にはツールの評価が必須だが、従来のツール評価では評価者が評価結果を得るために力をおいていたため、評価項目や評定基準、さらにそれらの決定理由など評価過程のデータが残されていることが少なく、評価結果を共有することを難しくしていた。

そこで、評価結果を共有・蓄積する事を目的として、評価者の経験や主観等の背景に依存しない再利用性の高い評価データを獲得するための作業手順、成果物、ガイドラインからなる評価方法論について提案する。

### 2. ツール評価方法における問題

従来のツール評価における問題点の一つは、『評価者が異なれば評価結果も異なり、評価結果を共有できない』ということである。我々は、この原因は評価項目に対する計測値の解釈といったものが、評価者によって異なり、その相違点がわからないということであると考えた。

そこで、ツール評価が図1に示すプロセスモデル<sup>[2]</sup>に従うものとすると、評価者の評価目的、経験、個性などが影響を及ぼし、上記問題を引き起こす可能性があるのは、次のようなプロセスである。

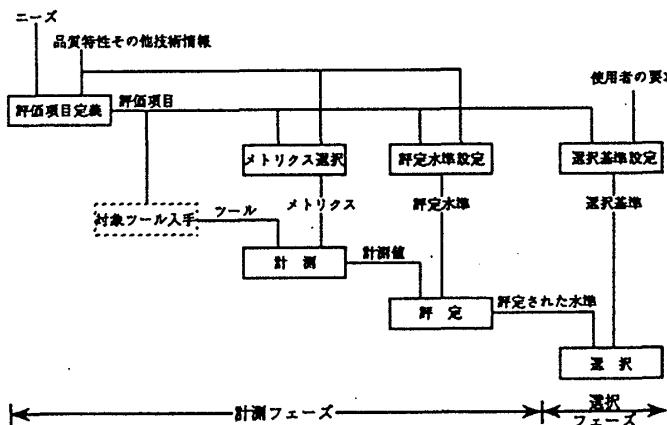


図1. ツール評価プロセスモデル

- ①評価項目定義プロセス
- ②メトリクス選択プロセス
- ③計測プロセス
- ④評定水準設定プロセス
- ⑤評定プロセス

### 3. 提案する評価方法

問題を解決するための一つの方法は、各評価プロセス内の評価過程を追跡することができ、自分と違うところから順次自分の意見を反映させることができるような枠組みを提供することだと考えた。そこで、以下の評価方法を提案する。

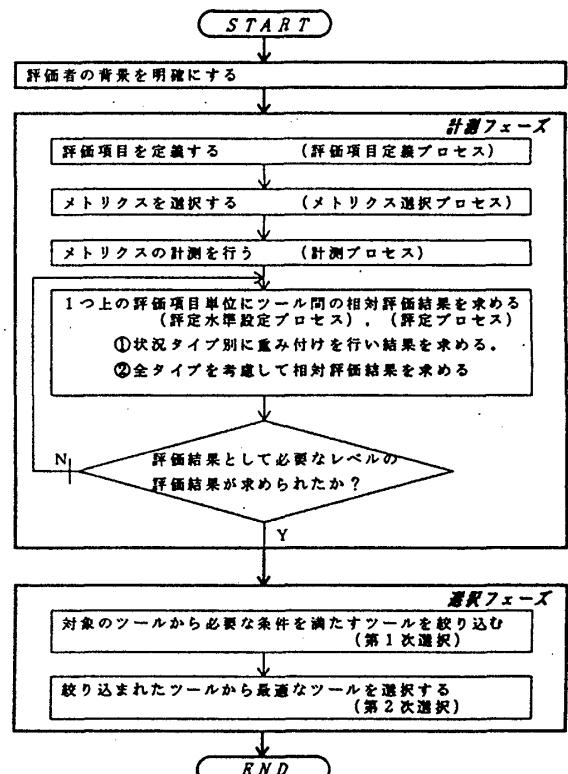


図2. ツール評価方法

また、上記評価方法における留意点をいくつか示す。

- 評価者の目的、経験、想定した精度、観点などによって評価項目の種類、詳細度、メトリクスが異なる。そこで、これらの関係を第三者が理解できるように定型フォーマットを埋めていく。

表1. 数値化手法の問題点

	定量的	定性的
絶対的	計測データ：評価者の主觀・個性 ・経験などの影響が含まれる ①計測(measurement)時はない ②評定(rating)時はある 評価者の影響：小	計測データ：評価者の主觀・個性 ・経験などの影響が含まれる ①計測(measurement)時はある ②評定(rating)時もある 評価者の影響：大
相対的		計測データ：評価者の主觀・個性 ・経験などの影響が含まれる ①計測(measurement)時はある ②評定(rating)時もすこしある 評価者の影響：中

- 評価項目を定義するためのガイドラインとして、水平的・階層的・時間的・空間的観点を利用する。
- 数値化に関する問題点（表1）を考慮し、できるだけ定量的な評価を行う。また、評定時には相対的な数値化が望ましい。このために、意志決定支援法のAHP (Analytic Hierarchy Process)<sup>[3]</sup>を用いる。
- メトリクスに対する重要度は評価者によって異なる。このため、計測データを複数の異なるユーザ間で共有するためには、いろいろな評価者（これを以下状況と呼ぶ）を想定した重要度を設定する。
- 全部の状況タイプを考慮して最終的な評価結果を求める。さまざまな状況を考慮したことにより客観的で汎用的な評価結果を求めることができる。このときも、ツールの評価結果をAHP等を使って相対化することにより、評価結果の一般性を向上させる。

#### 4. 評価結果の再利用の考え方

評価結果を再利用する際の基本的な考え方について具体例を用いて説明する。

本評価方法による評価結果として蓄積されるのは、図3に示すものである。

選択フェーズに渡されるデータは、最終的な『評価項目の得点』であり、従来の評価結果に相当する。この結果はさまざまな状況を考慮して求められたものなので、そのままでも汎用性はある。しかし、もし得られた評価結果に対して疑問がある場合は、評価結果を図3の流れを逆にたどることで追跡することができる。そして、相違点を見つかったときにはその補足欄（所見）に変更理由などを記録し、図中の網かけ部分を操作することで評価

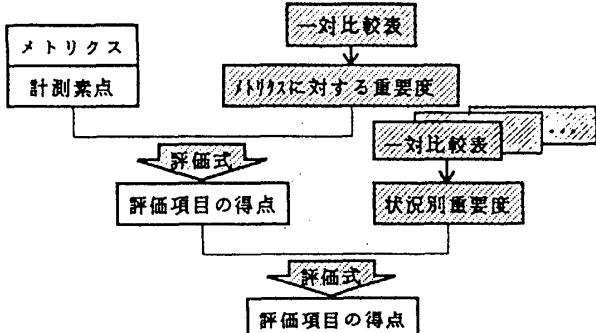


図3. 蓄積される評価データおよび評価結果

データを修正する。

これにより、ユーザは評価データを再利用できるとともに、その得点が得られるまでの経緯を理解することができる。そして、自分の目的に合致するようなツールを選択することが可能となる。

たとえば、図4の2通り（機能性重視、操作性重視）の状況別の評価結果から評価結果が求められているとする。しかし、ユーザは「ツールBはツールFと同程度である」という評価結果に疑問を持ったのであれば、その理由を記録し、評価結果を追跡する。そして、状況別重要度で性能重視の状況が抜けていることを指摘し、その場合の評価得点を求め、一对比較を再度行うことで評価結果を「ツールBはツールFよりは良い」というように修正することができる。

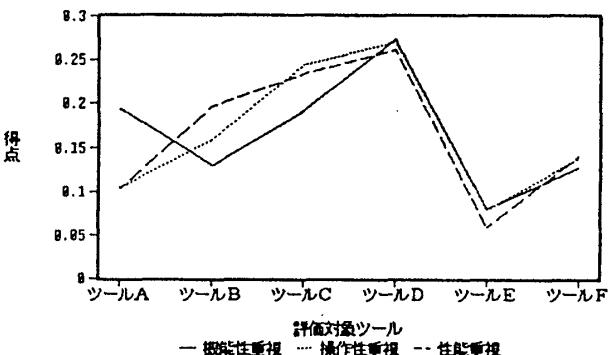


図4. 評定の際の状況別グラフの例

#### 5. まとめ

本稿では、従来のツール評価における問題点の一つを明確にし、それを解決するために手順も含めたツール評価方法論について提案した。従来の評価方法との最大の違いは、『評価結果の再利用ができるように評価のフェーズを切り分け、評価過程の追跡性を保証した』ことである。

また実際に本方法論をCASEツール評価に適用し<sup>[1]</sup>、方法論のツール支援も検討中である。具体的には、評定プロセスを関数化するとともに、選択時に目的を指定すれば評価項目の重要度を自動的に調整してくれるというようなものである。

今後は、より多くのツールに適用し、評価方法論とその結果の妥当性を検討していく。

#### 参考文献

- [1]吉田他、「ツール評価（2）－CASEツール評価－」情報処理学会、第44回全国大会 5K-3
- [2]SOFTWARE QUALITY CHARACTERISTICS, I S O DP9126, 1987
- [3]刀根、「ゲーム感覚意志決定法：AHP入門」日科技連出版社、1986