

学位論文の審査結果の要旨

本論文では、ソフトウェアドキュメント間の関連付け（トレーサビリティリンク）を自動的に行う手法の一つである情報検索手法について、最適な手法を見つけるためにデータマイニング手法を適用するというアイデアを提案し、その有効性を実証している。

ソフトウェアはその開発において種々のドキュメントが作られる（要求文書、設計文書、テスト文書、ソースコード、ユーザマニュアル等）。そしてこれらのドキュメントは種々の観点から同じ内容を表現しているため、相互に関連を持つ。ソフトウェアの保守においては、構成するドキュメント間の関連付けに基づく処理が多く行われる。たとえば要求を変更しようとする場合、対応する設計文書やソースコードを見つける必要がある。また、個々の要求項目が正しく実装されているかどうかを調べるためにその間の関連の認識が不可欠となる。このようにトレーサビリティリンクはソフトウェア保守においてきわめて重要であるが、一般にこのような関連性は正しくは維持・管理されずに、必要に応じて再構築されることが多い。これを自動化しようというのがリンクの再確立であり、再確立手法の主たる手法が情報検索手法である。さらに情報検索手法にも種々のvariationがあり、対象ドキュメントの特性によりその適否、すなわち再確立の精度が異なる。従来これら手法オプションの適否は研究されているが、ドキュメント毎に適切なオプションを見つけようという研究は行われていなかった。

本論文ではドキュメントの種々の特徴がこの適否に関係すると考え、ドキュメントの特徴量とオプションの関係を、リンク再確立の精度と関連付けて知識ベース化し、その知識ベースを使って最適なオプションを見つける手法を提案し、参照データセットを使って有効性を評価し、実証している。

具体的には以下のresearch questionを立てて、その検証を行っている。

- ①最適な手法オプションは、適用するソフトウェアプロジェクトによって異なる。
- ②最適な手法オプションの識別には、適用するソフトウェアプロジェクトの特徴量が関係する。

- ③知識ベースを使った最適な手法オプションの識別は、有意な精度で可能である。

情報検索手法では対象とするソフトウェアドキュメントを単語－文書行列で表現し、この個々の文書に対応する単語ベクトル間の類似度でリンクの有無を判断する。そこでオプションとしては以下のようなものを採用した：単語－文書行列の要素の値の種別、単語の選択法としてのstop wordの使用、同じくstemmingの使用、類似度の判定法としてのVSM、LSI手法等。またドキュメントの特徴量としては以

下のような量を利用した：ドキュメントの種別、サイズ、単語数、平均リンク数。結論としては、1) オプションと特微量の組み合わせによりリンク再確立の精度が異なる、2) ソフトウェアプロジェクトによるが概ねオプションのマイニングは可能である、などが示された。但し、回復精度が低いプロジェクトや、マイニングの精度の低いプロジェクトがあり、これらの改善が今後の課題として指摘されている。

本研究に関しては査読付きの国際会議で発表しており、またその内容を発展させて、電子情報通信学会の和文論文誌に掲載されている。また国際会議論文は査読付きであることを査読報告で確認した。これにより数理情報システム学講座の課程博士の審査基準を満たすことを確認した（上田君は博士課程修了1年以内の申請であり、課程博士の基準での審査が認められている）

以上、本研究はソフトウェア保守において重要な役割をはたすtraceability linkの再確立に寄与するものであり、審査委員全員一致で、博士（工学）論文に値するものと判断した。

公表主要論文名

1 : 上田 健之、小形 真平、海谷 治彦、海尻 賢二：情報検索手法に基づくトレーサビリティリンク回復のための手法オプションについてのマイニングの提案と評価、電子情報通信学会 和文論文誌D Vol. J97-D, No. 3 2014、pp. 414-426

2 : Takeyuki Ueda, Shinpei Ogata, Haruhiko Kaiya, Kenji Kaijiri: IR Based Traceability Link Recovery Method Mining, International Conference on Software Engineering Advances, 2013.10、 pp.278-284