

氏名(本籍・生年月日) 上田健之(神奈川県 昭和39年7月24日)  
学位の種類 博士(工学)  
学位記番号 乙第229号  
学位授与の日付 平成26年3月20日  
学位授与の要件 信州大学学位規程 第5条第2項該当  
学位論文題目 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  
論文審査委員 主査 教授 海尻賢二 准教授 海谷治彦  
教授 山本博章 助教 小形真平  
教授 樫山淳雄(東京学芸大学)

## 論文内容の要旨

ソフトウェア保守はそのソフトウェアの価値を維持し、高めるための重要な工程であるが、生産性が低く不具合の混入しやすい困難な工程であることも知られている。この困難さの原因の一つがソフトウェアアーティファクト間のトレーサビリティリンクの喪失である。喪失したリンクを自動的に回復させる手法として近年、情報検索手法の利用が注目されている。

情報検索手法によるトレーサビリティリンク回復は単語の共起性を元にドキュメント間の類似度を求め、類似度の高いものをリンクと判断して回復を行う手法であるが、単語の処理や作成した単語ベクトルの処理などにさまざまな手法のオプションがあり、それらは対象ドキュメントとの相性があることが従来から言われている。又、ソフトウェアアーティファクトは要求仕様、設計文書、ソースコード、テスト仕様などの様々なドキュメントから構成される上、その特徴はプロジェクト毎に多彩である。

情報検索手法に基づくトレーサビリティリンクの回復を現場で行う場合、ドキュメントの特徴に対して適切な手法を選択する必要がある。理想的な状況は、ドキュメントの特徴量から正しい手法が選択されることだが、そのような研究はまだない。

そこで、本研究は現場でのリンク回復の実践において正しい手法を選択するシステムティックな方法として手法オプションマイニングを提案する。これは参照データセットに対して種々の手法の組み合わせを元にリンク回復の予備実験を総当たりで行い、その精度と特徴量と手法との関係のマイニングを行う事で、ある特徴量をもつドキュメント対についての適切な手法を予測するというものである。予測の精度が十分なものであれば、精度の悪い手法によるリンク回復の実施を避けることができる。

さらに、リンクを回復する対象となるドキュメントの構造には一方が他方を包含するなどの非対称性があるものもあるのだが従来の類似度計算にはその非対称性を考慮するものはなかった。そこで本研究では情報検索手法のオプションの一つとして非対称類似度を提案した。

本研究ではこの提案手法について下記の5つの仮説を用意し、それを3つの実験で検証する。

仮説1：すべてのドキュメント対について良いリンク回復精度を与える一意な手法は存在しない。

仮説2：手法オプションの適否はドキュメントの特徴量に依存する。

仮説3：リンクの種別によっては非対称類似度はリンク回復精度の向上に有効である。

仮説4：本提案手法によって、妥当なリンク回復精度を与える手法オプションを有意な識別精度で識別することができる。

仮説5：目標とするリンク回復精度と識別精度には関連がある。

上記の仮説を検証するために参照データ集合に基づいて下記の実験を行った。

実験1：各種ドキュメントに対して各種手法の選択の組み合わせを用いてリンク回復の実験を行い、正解値との比較により文書、手法のオプション毎の精度を求める。この実験により仮説1及び仮説3を検証するとともに、この実験結果を元に実験2、実験3を行う。

実験2：実験1の結果を使って構成した訓練データ集合の交差検証を行い仮説4を検証するとともに作成した識別器の決定木の条件が文書対の特徴量になっている事並びにその識別器によって有意な精度で適切な手法を選択できることを確認することで仮説2を確認する。

実験3：実験1で作成した訓練データ集合を、プロジェクト毎に、自らの文書対によるデータだけを除き、他の全てのデータを訓練データとして作成した識別器の精度を自らの文書対に適用して調べる。これにより、識別精度に対するプロジェクトの影響を評価し、プロジェクトを跨る条件下での仮説4の検証を行う。さらに目標とするリンク回復精度を変化させて識別精度の変化を確認することで仮説5の検証を行う。

これら仮説検証により提案手法の有効性を実験によって検証した。