

論文内容の要旨

博士論文題目 欠損値を含むソフトウェア開発データリポジトリを用いた
プロジェクト計画支援

氏 名 角田 雅照

ソフトウェア開発プロジェクトを成功に導くためには、プロジェクトの計画時に、適切なプロジェクト特性(開発期間、プログラミング言語、開発要員数、外部委託比率など)を決定し、さらに、開発工数の予測を精度良く行うことが求められる。本論文では、一般に、多数の未記録の値(欠損値)が含まれる、ソフトウェア開発プロジェクトの特性値を蓄積したリポジトリ(ソフトウェア開発データリポジトリ)を用いて、(1) プロジェクト特性の決定の支援、および、(2) 開発工数予測の支援を行う。具体的な成果は次のとおりである。

(1) プロジェクト特性と生産性との関連の分析

プロジェクト特性の決定に当たっては、各特性(開発期間など)が開発の生産性に与える影響を考慮する必要があるが、従来、日本のソフトウェア開発企業を対象とした生産性に関する分析はほとんど行われていない。本論文では、日本の代表的なエンタープライズ系ソフトウェア開発企業 15 社から得られたソフトウェア開発データリポジトリを分析し、プロジェクト特性と生産性との関連を明らかにした。分析では、まず比例尺度を名義尺度に変換し、全ての特性を名義尺度として同様に扱った。尺度変換により、一元配置分散分析を適用して生産性に対する各特性の寄与率を導出できた。また、クラメールの V を用いて特性間の相互関連を分析することにより、生産性と間接的な関連がある特性を明らかにした。分析の結果、平均開発要員数は生産性と最も関連が強く(寄与率 0.39)、開発規模は平均開発要員数を介して間接的に生産性に関連があることなどが分かった。

(2) 協調フィルタリングに基づく工数予測

従来、ソフトウェア開発データリポジトリを用いた工数予測手法が多数提案されているが、リポジトリに欠損値を含まないことが前提である。本論文では、欠損値の有無や多寡に関わらず工数予測を行うための方法として、協調フィルタリング(CF)に基づく工数予測方法を提案した。CF は、欠損値を大量に含んだデータを入力とした場合でも予測が行えるという特長があるが、ソフトウェア工数予測に適用する方法はこれまで提案されていない。提案方法では、まず、入力となる特性値(メトリクス値)を正規化し、値域を揃える。次に、正規化したメトリクス値を用いて、予測対象(開発中)のプロジェクトと、過去に行われたプロジェクトとの類似度を計算する。最後に、類似度の高い(予測対象プロジェクトと類似した)プロジェクトの工数を類似度で加重平均した値を、予測対象プロジェクトの工数とする。ケーススタディを行った結果、CF に基づく工数予測方法は従来方法(欠損値処理法を用いたステップワイズ重回帰分析)よりも高い精度を示し、相対誤差の平均値(1 プロジェクト当たり)が 22.11 から 0.79 に改善した。

氏 名	角田 雅照
-----	-------

(論文審査結果の要旨)

本論文では、ソフトウェア開発データリポジトリを分析した結果に基づき、プロジェクト特性の決定に関する指針の提示、および、開発工数予測方法の提案を行っている。

まず、プロジェクト特性の決定の支援を目的として、日本のソフトウェア開発企業から得られたソフトウェア開発データリポジトリを分析し、プロジェクト特性と生産性との関連を明らかにしている。分析では、まず比例尺度を名義尺度に変換し、全ての特性を名義尺度として同様に扱い、一元配置分散分析を適用して生産性に対する各特性の寄与率を導出している。また、クramerのVを用いて特性間の相互関連を分析することにより、生産性と間接的な関連がある特性を明らかにしている。分析の結果、生産性に影響を与えている要因のうち、最も有力なものは平均要員数であることを示した。また、FP(開発規模)は生産性に対して弱い関連しか見られなかったが、FP が大きいプロジェクトの大部分は平均要員数も多かったことから、FP は平均開発要員数を介して間接的に生産性に関連があることを示した。従来、日本の複数のソフトウェア開発企業からデータが大規模に収集されたことがなく、日本のソフトウェア開発において、プロジェクト特性と生産性との関連は必ずしも明らかではなかった。本論文の結果は、プロジェクト計画時には平均開発要員数は抑えるべきであるという指針を開発現場に与えた点で大変有用である。

次に、ソフトウェア開発データリポジトリに含まれる欠損値に着目し、欠損値の有無や多寡に関わらず工数予測を行うための方法として、協調フィルタリング(CF)に基づく工数予測方法を提案している。CF は、欠損値を大量に含んだデータを入力とした場合でも予測が行えるという特長があるが、ソフトウェア工数予測に適用する方法はこれまで提案されていない。提案方法では、まず、入力となる特性値(メトリクス値)を正規化し、値域を揃える。次に、正規化したメトリクス値を用いて、予測対象(開発中)のプロジェクトと、過去に行われたプロジェクトとの類似度を計算する。最後に、類似度の高い(予測対象プロジェクトと類似した)プロジェクトの工数を類似度で加重平均した値を、予測対象プロジェクトの工数とする。提案方法を評価するために、従来の予測方法である、欠損値処理法を用いたステップワイズ重回帰分析と、予測精度の比較を行うケーススタディを行い、CF に基づく工数予測方法は従来方法よりも高い精度で予測できることを示している。従来の工数予測手法はリポジトリに欠損値を含まないことが前提であるのに対し、提案方法は、欠損値の有無や多寡に関わらず工数予測を行うことができるという点で新規性、有用性ともに大である。

以上のとおり、本論文は、ソフトウェア開発データリポジトリの分析結果に基づき、プロジェクト計画時におけるプロジェクト特性の決定の指針を提供するとともに、欠損値を含むリポジトリを用いた工数予測方法を提案するものである。ソフトウェアプロジェクトにおいてコストの超過や納期の遅延を防ぐためにはプロジェクト計画が鍵となる。これらの研究成果により、過去の定量的データに基づいた、適切なプロジェクト計画を立案することが可能となる。よって、研究成果はソフトウェア業界の発展に大きく貢献するものであり、本論文は博士(工学)論文として価値あるものと認める。