

ベイズ識別器によるソフトウェアプロジェクト の混乱予測システムの開発



若 者

水 野 修*

1. ま え が き

近年のソフトウェア開発プロジェクトでは、短期間で高品質な製品の開発が求められてきている。こうした要求を満たす1つの古典的な解決策は、十分な人と時間をソフトウェアのテスト工程にかけることであった。しかし、テスト工程に必要とされる時間・人的資源まで十分に考慮してソフトウェア開発の計画が行われることは稀であり、その結果として開発途中でのプロジェクトの混乱という現象が発生する。プロジェクトの混乱に対しては、個別に応急処置を行うのではなく、開発プロセス全体の改善を進めていくことが重要であると言われている。

筆者は大阪大学大学院情報科学研究科菊野研究室で助手として6年の間、いくつかの企業のソフトウェア開発プロセスの改善に協力しつつ、研究活動を展開してきている。そうした中で、ある企業からソフトウェア開発の混乱が重要な問題となっており、混乱しそうなプロジェクトの予兆をつかむことができないかという要求が我々に寄せられていた。

本研究はこのような背景の下に、プロジェクトの混乱予測システムを研究・開発した成果をまとめている。本研究ではプロジェクトマネージャへのアンケート調査をプロジェクトの混乱という事象に結びつける手法を提案した。そして、プロジェクトが混乱するか否かを予測するシステムをデータマイニン

グ手法を用いることにより実装した。この分析と評価には協力企業で過去2年間に行われた15個のプロジェクトデータから得られた80回答のアンケートデータを利用している。

2. プロジェクトの混乱状態

開発現場の状況が制御不可能になってしまうプロジェクトを混乱プロジェクトと呼ぶ。しかしこれでは漠然としているので、ここでは混乱プロジェクトを次のように定義する。

計画との相対誤差 混乱状態に陥ったプロジェクトは、工数・開発期間・納期などにおいて計画からのずれが生じる。そこで、これらの値の計画からのずれが基準を超えたものを混乱プロジェクトの候補とする。

開発プロセス 設計・レビュー・テストの各プロセスが組織の内部規約通り実施されていないプロジェクト。

成果物の品質 プロジェクト成果物の品質が著しく悪いプロジェクト。

これらの基準を基に、組織のSEPG(Software Engineering Process Group)により混乱状態に陥ったと認められたものを混乱プロジェクトと定義する。

3. ベイズ識別器によるプロジェクト 混乱予測システム

我々は、ソフトウェアの混乱予測のためにアンケートの収集から混乱の予測までを自動的に行うシステムの開発を目指している(図1)。以降では、混乱予測の流れと、その中で使用するアンケート項目の作成、そしてベイズ識別器の理論について説明する。

3.1 ツールによる混乱予測の流れ

図2は本論文で提案する混乱予測手法の大まかな作業の流れを表している。まず、ステップ1でプロジェクトメンバーからプロジェクトの評価を聞き出



* Osamu MIZUNO
1973年7月生
1999年大阪大学・大学院基礎工学研究科・博士後期課程中退
現在、大阪大学・大学院情報科学研究科・情報システム工学専攻、助手、博士(工学)、ソフトウェア工学
TEL 06-6879-4536
FAX 06-6879-4539
E-Mail o-mizuno@ist.osaka-u.ac.jp

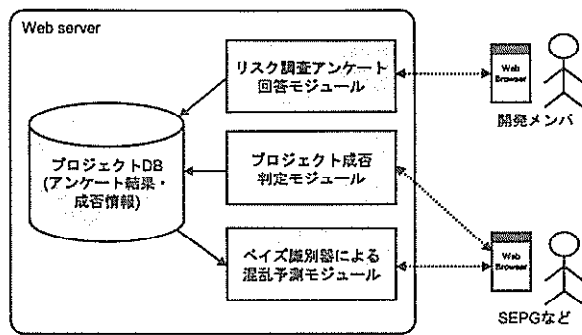


図1 混乱予測システム

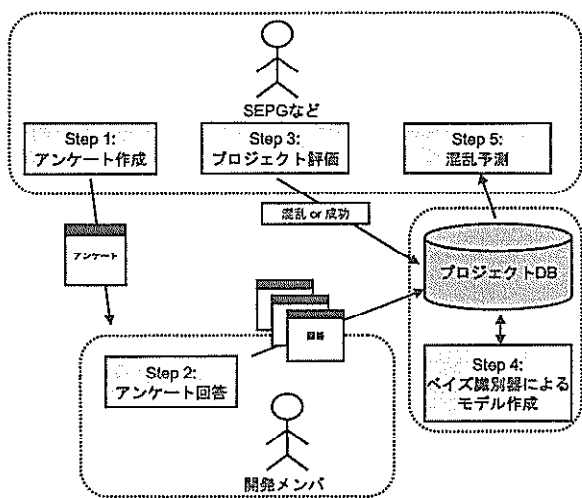


図2 混乱予測の流れ

すためのアンケートを作成する。本研究ではある企業の協力を得て企業のSEPGが中心となって作成した。作成に当たっては、私たちの研究グループが過去に他の企業と協力して作成したものを参考に、企業の内部規約・プロジェクト特性を考慮した。

続いてステップ2では作成したアンケートをプロジェクトメンバーに配布し、アンケートに回答してもらう。本研究では既に終了した15個のプロジェクトを対象としてそのプロジェクトメンバー80人にアンケートを実施した。

提案手法ではプロジェクトの混乱予測を行うにあたり、過去のプロジェクトのデータを利用する。予測を行うモデル作成には混乱の有無が分かっているプロジェクトデータが必要であるため、この混乱の有無の判定をステップ3で行う。

ステップ4では開発した混乱予測ツールを利用し

て、学習対象としたアンケート回答とプロジェクトの結果からベイズ識別器によるモデルを作成する。このとき、アンケートの回答結果は部分的に未回答の項目があるものも問題なく扱える。引き続き、ステップ5では新規プロジェクトについて混乱予測を行う。混乱予測は新規プロジェクトのアンケート結果をステップ4で作成したモデルに当てはめることで行う。ステップ5の作業も開発したツールを用いることで自動的に行える。

3.2 リスク調査アンケート

本節では、混乱予測を行うためにプロジェクトメンバーに対して行ったリスク調査アンケートについて説明する。

このアンケートの目的は、プロジェクトを混乱状態へと陥らせるような要因の有無・程度をプロジェクトメンバーから聞き出すことである。

アンケートの作成に当たっては、私たちの研究グループにおいて過去に作成したアンケート^[3]を基に、企業の内部規約・プロジェクト特性などを考慮して企業のSEPGが中心となって作成した。作成したアンケートは次の9つの主要な要因に整理される：要求仕様の問題点、見積りの問題点、プロジェクト体制の問題点、工程計画の問題点、進捗管理の問題点、開発技術の問題点、開発メンバーの問題点、プロジェクト環境の問題点、その他の外的要因に関する問題点。

アンケートでは、プロジェクトメンバーが各質問項目に対する評価を行う。評価は各項目ごとに回答選択肢を用意し、その何れかを選ぶことにより行う。

3.3 ベイズ識別器

ベイズ識別器はベイズの定理を利用してカテゴリカルデータをいくつかのクラスに高い精度で分類する手法として良く知られている。ベイズ識別器は各属性が互いに独立であるという仮定を置いているが、経験的に仮定が破られたデータセットであっても極めて精度の高い予測が可能であることが知られている^[1]。

3.4 混乱予測モジュールの実装

混乱予測モジュールはWeb上で動作するアプリケーションとして実装した。使用した言語はPerlとRで、開発規模はほぼ1000行であった。開発したシステムを用いることで、アンケート結果から混乱す

るかどうかの予測を非常に簡単に行える。なお、開発しているツールは当研究室のWebページ(<http://www-ise4.ist.osaka-u.ac.jp/>)にて公開予定である。

4. 適用実験

本章では、3.2節で作成したアンケートを実際のソフトウェア開発プロジェクトに対して適用し、アンケートの回答結果からベイズ識別器による混乱予測を行った結果を示す。

対象としたプロジェクトはある企業において2001年から2002年の間に実施された15個のプロジェクトで、アンケートは1プロジェクトにつき1人から9人の計80人のメンバーに対して行った。このメンバーには、プロジェクトの評価を行ったメンバーなど、一部分のみに携わったメンバーも含まれる。

本実験における対象プロジェクトは既に終了しているため、各プロジェクトの評価をSEPGが行った。プロジェクトの評価は、2節で示したように、プロジェクトの計画・プロセス、およびプロジェクト成果物の品質を基にSEPGが総合的に判断した。

4.1 混乱予測結果

得られたアンケート結果及びプロジェクト評価を利用して、ベイズ識別器により混乱予測を行った。

本研究では、80個のアンケート回答をプロジェクト単位に分割したジャックナイフ法^[2]で実験を行っ

た。すなわち、80回答から1プロジェクトに相当する個数の回答データを取り出してテストデータとし、残りの14プロジェクト分の回答データを学習データとしたモデルを作成し、最後にこのモデルにテストデータを適用することで精度の計算を行っている。なお、1つのプロジェクトには複数の回答者がいるため、得られた予測結果の多数決を取ることで、そのプロジェクトの成功・失敗の予測を行っている。そして、得られた予測結果とSEPGによって評価された実際の結果を比較した。結果を表1に示す。

表1に示すように、15個のプロジェクト中12個について正しく予測することができ、予測精度は80%となった。このことから提案手法によりプロジェクトの混乱予測が可能であることが確認できた。また、この結果にフィッシャーの正確確率検定を行ったところ、 $p < 0.01$ となり統計的にもこの手法が有用であることが分かった。

5. まとめと今後の課題

本研究では、ソフトウェア開発プロジェクトの混乱予測を行う手法として、プロジェクトメンバーへのアンケートとベイズ識別器を用いる手法を提案し、そのためのツールの実装を行った。また実際のプロジェクトデータを用いて適用実験をおこなった。その結果、非常に高い精度でプロジェクトの混乱予測が可能であることがわかった。また、現在進行中のプロジェクトに対しても同様のアンケートを行っており、そのプロジェクトの成否についても今後、プロジェクトの終了時に評価する予定である。

参考文献

- [1] P. Domingos and M. J. Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3): 103-130, 1997.
- [2] R. O. Dura, P. E. Hart, David G., 尾上守夫監訳. パターン識別. 新技術コミュニケーションズ, 2001.
- [3] O. Mizuno, T. Kikuno, Y. Takagi, and K. Sakamoto. Characterization of risky projects based on project managers' evaluation. In *Proc. of 22nd International Conference on Software Engineering*, pp.387-395, 2000.

表1 プロジェクト成否結果

プロジェクト	実際の結果	予測結果
A	混乱	混乱
B	成功	混乱
C	成功	成功
D	成功	成功
E	混乱	混乱
F	混乱	成功
G	成功	混乱
H	混乱	混乱
I	混乱	混乱
J	混乱	混乱
K	混乱	混乱
L	混乱	混乱
M	混乱	混乱
N	成功	成功
O	成功	成功