情報システム開発プロジェクトの成否予測の仕組みの提案

- 属性毎に分類したデータの活用 -

Proposal of Project Outcome Prediction System for the IS Development Project

- Utilizing Classified Data according to Attributes -

慶應義塾大学大学院　システムデザイン・マネジメント研究科
Graduate School of System Design and Management, Keio University
○河村 智行　　高野 研一 1)
○Tomoyuki Kawamura　　Kenichi Takano1)

Abstract　Researchers have found that approximately 70% of the Information Systems development projects in Japan have failed, thus increasing the demand for solutions that will raise expected project success rates. It is said that to improve success rates, support should be provided by the organization to which the projects belong. This study aims to identify projects that an organization should preferentially support by predicting a project outcome. In the previous research, the project outcome prediction was performed for "Company A," a Japan-based IT vendor. As a result, a model with a predictive accuracy of 73.9% was obtained. In this study, the authors attempt to predict project outcomes with higher accuracy by utilizing classified project data sets according to attributes. As a consequence, the average of the predictive accuracy of the models using the classified data sets according to Package software attributes was 76.6%. These results are expected to contribute towards identifying the projects that an organization should support preferentially.

1. 背景

情報技術産業は、1950 年代の商用コンピュータの登場以来、飛躍的な成長を遂げてきた。情報技術産業を構成する主要な活動の一つに情報システム開発がある。日本の情報システム開発の投資額は 2014 年に 12 兆円を超え [1]、今後もさらなる成長が期待されている。しかしながら、古くから情報システム開発プロジェクトの失敗が問題となっている。日経 BP [2] によると、2003 年以降、日本の情報システム開発の約 70%が失敗プロジェクトであると言われており、プロジェクトの成功率の向上が望まれている。

情報システム開発のプロセスモデルとして広く利用されている CMMI（能力成熟度モデル統合） [3] は、プロジェクト、およびプロジェクトが所属する組織が実施すべき活動を定義している。CMMI は、リスク管理などの様々なプロセスをプロジェクトが実施する活動として定義している一方で、要員の追加などのプロジェクトのみでは十分に対応できないリスクの対策を組織が支援することで、リスクの影響を軽減し、プロジェクトを成功に導く必要があると述べている。実際に、プロジェクトに対する組織の関与がプロジェクトの成功に良好な影響を与えることが明らかにされている [4]。しかしながら、組織のリソースは有限であるため、組織内の全てのプロジェクトに対し十分に関与できるとは限らない。よって、影響の大きなリスクを抱えるプロジェクト、つまり失敗の可能性の高いプロジェクトを特定し、優先して関与することで、効率的に組織全体のプロジェクトの成功率を向上する必要がある。本研究は、プロジェクトの成否の予測をとおして、組織が優先して関与すべきプロジェクトの特定に寄与することを目的とする。

いくつかの研究がプロジェクトの成否の予測を試みている。Takagi ら [5]、および Mizuno ら [6] は、

慶應義塾大学大学院　システムデザイン・マネジメント研究科
Graduate School of System Design and Management, Keio University
横浜市港北区日吉 4-1-1　　Tel: 045-564-2518　　e-mail: tomoyuki@z5.keio.jp
4-1-1, Hiyoshi, Kohoku-ku, Yokohama Japan

1) 慶應義塾大学大学院　システムデザイン・マネジメント研究科
Graduate School of System Design and Management, Keio University

日本の特定の製造業企業の組込みソフトウェア開発プロジェクトを対象にプロジェクトの混乱発生の予測を試みた。彼らは、プロジェクトメンバへのアンケートによってプロジェクトを評価し、その評価結果を用いて混乱発生の予測を行い、高い精度で予測可能であることを示した。また、Wohlin ら [7] は、NASA のソフトウェア開発プロジェクトを対象に、専門家の判断によってプロジェクトを評価し、その評価結果を用いて早いタイミングでプロジェクトの成否の予測を行い、高い精度で予測が可能であることを示した。これらの研究は、開発に利用する技術が同一、および対象とする顧客が単一、と言った類似した特性を有するプロジェクトを予測対象としている。

一方で、Kawamura ら [8] は、多様な特性を有する日本の IT ベンダの情報システム開発プロジェクトを対象にプロジェクトの成否の予測を試みた。彼らは、専門家による要件確定時のリスク評価を行い、その評価結果を用いて成否の予測を行い、73.9%と一定の精度で予測可能であることを示した。前述の Takagi ら、Mizuno ら、および Wohlin らの研究では、予測精度 80～90%と高い予測力を達成したのに対し、彼らの研究では相対的に低い予測精度となった要因として、彼らの研究対象が多様な特性を有するプロジェクトを遂行する IT ベンダである点が影響している、と述べている。

本研究では、Kawamura らの成果に対して、さらなる成否予測能力の向上を達成したいと考えた。ソフトウェア開発のコスト見積りのモデルを調査した Chen ら [9] は、多様な特性を有するプロジェクトを対象とする場合、類似の特性を有するプロジェクト毎にモデルを作成することで、予測能力の向上が可能である、と指摘している。実際に、Takagi らは、類似の特性を有するプロジェクトを対象としたことで、高い予測精度を達成できた、と述べている。よって、多様な特性を有するプロジェクトを遂行する IT ベンダであっても、類似の特性を有するプロジェクト毎に予測を行うことで、予測能力の向上が可能であると考えた。以上より、本研究では、類似の特性、つまり属性毎に分類したプロジェクトのデータを利用することで、成否予測モデルの予測能力の向上を試みる。

本章に続く第 2 章では、Kawamura らが構築した成否予測の仕組みの概要を説明する。第 3 章では、彼らの研究成果をもとに、属性毎に分類したプロジェクトのデータを利用した成否予測モデルを構築し、その予測精度の変化を確認する。第 4 章では、第 3 章で示した予測精度の変化について考察する。そして、第 5 章では、本研究の成果をまとめる。

2. 先行研究

本章では、Kawamura らが構築した成否予測の仕組みを説明する。

2.1 成否予測の仕組み

日本の IT ベンダである A 社の協力のもと、成否の予測が実施された。A 社は従業員 1,000 名を超える IT ベンダで、日本の主要企業数百社を顧客として、情報システム開発に関連する一連のサービスを提供している。A 社では、プロジェクトの主要マイルストーン時に、プロジェクトの上位マネージャ（SM：Senior Manager）、および組織全体のリスク監査部門（PMO：Project Management Office）がプロジェクトのリスクの特定、および対策の検討を支援し、プロジェクトの遂行可否、および組織的なプロジェクトの関与の程度を決定している。以降、本活動を MS（Milestone）会議と呼ぶ。MS 会議は、プロジェクトの各種成果物の確認、およびプロジェクトメンバへのインタビューを通して実施される。

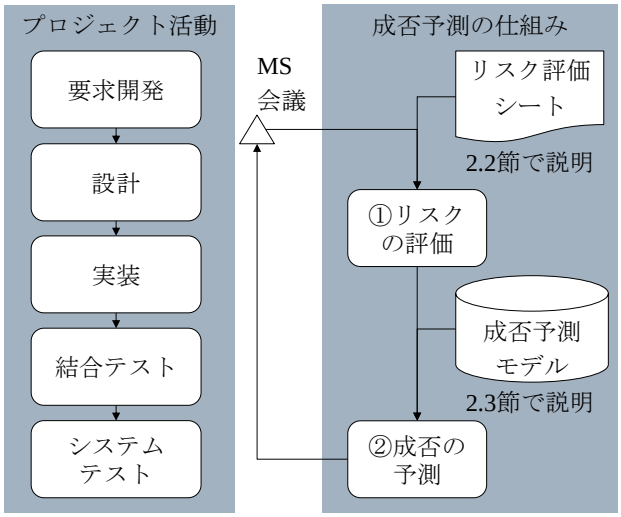

図1　成否予測の仕組み

成否予測の仕組みを、要求確定時の MS 会議に追加して実施することで、組織的なプロジェクトの関与の程度の決定に利用したいと考えた。成否予測の仕組みは図 1 のとおりである。「①リスクの評価」では、要求確定時の MS 会議の内容をもとに、SM、および PMO の合議でリスク

を評価する。リスクの評価は後述するリスク評価シートを利用する。「②成否の予測」は、「①リスクの評価」の結果をもとにプロジェクトの成否を予測する。予測に際しては、後述する成否予測モデルを利用する。MS 会議では、これらの結果を参考に、プロジェクトの遂行可否、および組織的なプロジェクトの関与の程度を決定する。

2.2 評価シートの作成

成否予測の仕組みの構築に当たって、「リスク評価シート」、および「成否評価シート」を作成した。リスク評価シートは、要求確定時のプロジェクトのリスクの評価項目を列挙したものである。一方で、成否評価シートは、プロジェクト完了時の成否結果の評価項目を列挙したものである。成否評価シートは、図 1 に示した成否の予測の実施では不要なものであるが、後述する 2.3 節の予測モデルを構築する際に必要となるため、リスク評価シートと合わせて作成した。

(1) リスク評価シート

リスク評価シートは、Mcleod ら [10] が示した成否に影響を与える要因のフレームワーク、および A 社の過去のナレッジを参照して、A 社の PMO と共同で作成した。また、作成に際しては要求確定時に評価可能な項目であることを考慮した。リスク評価シートは表 1 のとおり 17 のリスク評価項目から構成される。

(2) 成否評価シート

Mcleod ら [10] は、プロジェクトの成否の定義は観点によって異なるため、一律に規定することは困難であると述べている。本研究では、調査対象が IT ベンダであることを考慮し、開発者の視点で比較的良く利用される定義である [11] [12]、品質・コスト・スケジュールの計画値と実績値の差異をもとに成否を評価する。成否評価シートの各項目は以下の 3 設問で構成される。

- 品質　　：想定したとおりの品質（納品後の不具合発生状況）だった
- コスト　　：計画したコストどおりに完了した
- スケジュール：計画した納期どおりに納品した

評価シートの各項目の評価は、設問に回答する形式で行う。また、評価結果を定量的に処理することを考慮し、「1：非常に良く当てはまる」～「6：全く当てはまらない」の 6 段階のリッカート尺度から選択する（「プロジェクト特性」の設問のみ「6：非常に良く当てはまる」～「1：全く当てはまらない」から選択する）。

なお、評価シートには参考情報として各評価項目の詳細な説明、および評価の選択肢の基準を記述しているが、A 社の内部情報が含まれるため本論文では省略している。例えば、評価項目「要求の精度」の評価の選択肢の基準であるが、本研究では、後述するとおりウォーターフォールベースの開発プロセスを採用しているプロジェクトを対象とするため、基本的に「1：非常に良く当てはまる」、または「2：当てはまる」が選択される。しかし、実際には、非常に多くの積み残し課題があり、設計以降に多数の仕様の見直しが想定されるプロジェクトが存在する。それが認識された場合は、「5：当てはまらない」、または「6：全く当てはまらない」を選択することが基準として記されている。

2.3 成否予測モデルの構築

成否予測モデルを以下の 3 つのステップで構築した。

(1) 過去プロジェクトの評価によるデータの収集

前述のリスク評価シート、および成否評価シートを用いて完了したプロジェクトを評価した。評価対象プロジェクトは、A 社内の特定年度に完了した一定規模以上の受託開発プロジェクトである。これらのプロジェクトは顧客・技術と言ったプロジェクト特性が多岐に渡り、ウォーターフォールベースの開発プロセスを採用している。評価は、各プロジェクトの MS 会議に参加していた SM と PMO の合議のもと評価を行った。この結果、88 プロジェクトのデータが収集された。

(2)　プロジェクトの成否の分類

成否評価シートの評価結果をもとに各プロジェクトを成功／失敗に分類した。分類した結果を、予測モデルを作成する際の目的変数として利用する。分類方法は、日経 BP [2]、および The Standish Group [11] と同等の基準を用いて分類する。つまり、成否評価シートの 3 つの評価項目である品質・コスト・スケジュールにおいて、計画値を全て達成したプロジェクトを成功、一つでも計画値を達成できなか

表 1　リスク評価シートの項目

リスク評価項目		リスク評価項目の説明	評価時の設問
プロジェクト特性			
	11_顧客の新規性	プロジェクトが顧客（企業・担当者）との開発作業経験が豊富かどうかを評価する。	顧客の新規性は高い
	12_業務の新規性	プロジェクトが開発するシステムの業務分野（決済業務、販売管理業務、など）の知識・経験が豊富かどうかを評価する。	業務の新規性は高い
	13_技術の新規性	プロジェクトが開発に利用する技術（言語、ミドルウェア、など）の知識・経験が豊富かどうかを評価する。	技術の新規性は高い
	14_規模の大きさ	プロジェクトの規模の大きさ（総工数、ピーク時要員数）を評価する。	プロジェクトの規模は大きい
	15_ステークホルダ	プロジェクトが関与する必要のあるステークホルダ（ユーザ、他システム、他ベンダ、など）の数・複雑さを評価する。	ステークホルダは複雑
	16_要求の難易度	開発対象システムの要求（機能要求・非機能要求）の実現難易度が高いかどうかを評価する。	実現する要求は難しい
	17_既存システムの理解	プロジェクトが既存システム（今回開発するシステムと同様の機能を有する稼動中のシステム）に対する理解が高いかどうかを評価する。	既存システムの理解は少ない
開発プロセス			
	21_期間の制約	プロジェクト期間が充足しているかを評価する。また、プロジェクト全体だけでなく、各工程の期間の充足も考慮する。	プロジェクト期間は十分ある
	22_見積りの精度	見積り（金額・工数）を十分に精査しているかを評価する。	見積りを十分に検討している
	23_計画の精度	見積り（金額・工数）を除くプロジェクト計画に関する一連の活動（WBS の計画、レビュー計画、成果物構成、など）を十分に精査しているかを評価する。	プロジェクト計画を十分に検討している
	24_要求の精度	上流工程（要件定義、設計、など）における要求の整理（システム化の範囲の特定、システム化の仕様の明確化、など）を十分に実施しているかを評価する。	要求は明確
人と活動			
	31_顧客のスキル	期待される役割（要求の提示、成果物のレビュー、など）に対して、顧客（ユーザ含む）のスキル・リソースが充足しているかを評価する。	顧客のスキル・リソースは充足している
	32_顧客コミュニケーション	プロジェクトメンバと顧客の意思疎通・調整事項の合意が、円滑に実施できるかを評価する。	プロジェクトメンバと顧客のコミュニケーションは良好
	33_プロジェクトマネージャのスキル	プロジェクトマネージャに期待される役割（プロジェクト計画、進捗管理、など）に対して、プロジェクトマネージャのスキル・リソースが充足しているかを評価する。	プロジェクトマネージャのスキル・リソースは充足している
	34_業務 SE のスキル	業務 SE に期待される役割（業務の要件定義、設計、など）に対して、業務 SE のスキル・リソースが充足しているかを評価する。	業務 SE のスキル・リソースは充足している
	35_技術 SE のスキル	技術 SE に期待される役割（アーキテクチャの設計、インフラの設計、など）に対して、技術 SE のスキル・リソースが充足しているかを評価する。	技術 SE のスキル・リソースは充足している
	36_PG のスキル	実装者に期待される役割（プログラミング、テスト、各種設定、など）に対して、実装者のスキル・リソースが充足しているかを評価する。	実装者のスキル・リソースは充足している

ったプロジェクトを失敗と分類する。

成否評価シートの 3 つの評価項目である品質・コスト・スケジュールのうち、全て「1：非常に良く当てはまる」、または「2：当てはまる」と評価しているプロジェクトを「成功」、それ以外のプロジェクトを「失敗」に分類した。また、分類した結果に対し、各プロジェクトの MS 会議に参加していた SM・PMO が彼らの認識と大きな齟齬がないことを確認した。その結果、43 プロジェクト（49%）が成功、45 プロジェクト（51%）が失敗に分類された。

(3) ロジスティック回帰分析による成否予測モデルの作成

リスク評価シートにより収集した 17 項目の評価結果を説明変数、成功／失敗の分類結果を目的変数とし、ロジスティック回帰分析を適用して成否の予測モデルを作成した。そして、10-Fold 交差検証法を適用して、予測精度を確認したところ、73.9%の予測精度を有するモデルが得られた。なお、予測精度とは、全予測件数の中で予測結果と実際の結果が一致した件数の割合のことである。例えば、予測の実施 1000 件に対し、739 件が実際の結果と一致した場合、予測精度 73.9%となる。

3. 属性毎に分類したデータを利用した成否予測モデルの構築

本章では、第 2 章で示した先行研究をもとに、属性毎に分類したプロジェクトのデータを利用した成否予測モデルを構築し、予測精度の向上を試みる。予測モデルを以下の 4 つのステップで構築した。

3.1 モデル作成用のデータの追加

属性毎にデータを分類することで各モデルの作成に利用できるデータ数が減少するため、モデル作成の処理が一部実行できなくなる可能性がある。この問題を回避するために、モデル作成用のデータの追加を行う。追加対象のデータは、第 2 章の先行研究で入手した A 社内の特定年度に完了したプロジェクトの次の年度に完了したプロジェクトとする。これらのプロジェクトは、先行研究と同様に、顧客・技術と言ったプロジェクト特性が多岐に渡り、ウォーターフォールベースの開発プロセスを採用している、一定規模以上の受託開発プロジェクトである。

第 2 章と同様に、完了したプロジェクトの MS 会議に参加していた SM・PMO が、リスク評価シート、および成否評価シートを用いて対象のプロジェクトを合議のもと評価することで、データを入手した。この結果、新たに 91 プロジェクトのデータが追加され、合計で 179 プロジェクトのデータが利用可能となった。これら 179 件のデータのリスク評価シート、および成否評価シートの各項目の評価結果の分布は表 2、および表 3 のとおりである。

表2 リスク評価シートの評価結果の分布

リスク評価項目	平均	標準偏差
11_顧客の新規性	3.00	1.76
12_業務の新規性	2.86	1.48
13_技術の新規性	2.77	1.42
14_規模の大きさ	3.04	1.19
15_ステークホルダ	3.68	1.25
16_要求の難易度	3.42	1.16
17_既存システムの理解	2.81	1.57
21_期間の制約	2.80	1.29
22_見積りの精度	2.60	1.00
23_計画の精度	2.75	1.00
24_要求の精度	2.77	1.04
31_顧客のスキル	3.09	1.03
32_顧客コミュニケーション	2.39	0.91
33_プロジェクトマネージャのスキル	2.46	0.77
34_業務SEのスキル	2.58	0.96
35_技術SEのスキル	2.66	0.95
36_PGのスキル	2.60	0.95

3.2 プロジェクトの成否の分類

179 件のデータを成功／失敗に分類する。分類した結果を、予測モデルを作成する際の目的変数として利用する。分類方法は、第 2 章と同様に、成否評価シートの 3 つの評価項目である品質・コスト・スケジュールにおいて、計画値を全て達成したプロジェクトを成功、一つでも計画値を達成できなかったプロジェクトを失敗と分類する。

成否評価シートの 3 つの評価項目である品質・コスト・スケジュールのうち、全て「1：

表3 成否評価シートの評価結果の分布

成否評価項目	平均	標準偏差
品質	2.35	1.09
コスト	2.32	1.24
スケジュール	1.94	0.98

非常に良く当てはまる」、または「2：当てはまる」と評価しているプロジェクトを「成功」、それ以外のプロジェクトを「失敗」に分類した。また、分類した結果に対し、各プロジェクトの MS 会議に参加していた SM・PMO が彼らの認識と大きな齟齬がないことを確認した。その結果、96 プロジェクト（54%）が成功、83 プロジェクト（46%）が失敗に分類された。

3.3 属性毎のデータの分類

収集した 179 件のデータを属性毎に分類する。適用する属性は「プロジェクトタイプ」、および「業務パッケージの利用」の 2 つである。これらの属性は、情報処理推進機構が収集しているデータ [13] を参考に、A 社が収集している属性である。なお、この 2 つ以外にも A 社はいくつかの属性を収集しているが、本仕組みの予測実施タイミングである要求確定時に判明しており、かつモデルの予測精度の差異に影響を与えると考えられる属性として、この 2 つを採用した。

属性「プロジェクトタイプ」は 3 つの分類「新規開発」、「改良開発」、および「再開発」から構成される。各分類の定義は以下のとおりである。また、分類毎のデータの件数は、「新規開発：75 件」、「改良開発：90 件」、および「再開発：14 件」である。

- 新規開発：既存システムが存在しない、または既存システムを全面的に作り直す場合
- 改良開発：既存システムが存在し、修正・追加をする場合
- 再開発　：既存システムが存在し、機能仕様をほぼ変更することなくプラットフォームを変更する場合

一方で、属性「業務パッケージの利用」は 2 つの分類「利用あり」、および「利用なし」から構成される。分類毎のデータの件数は、「利用あり：103 件」、および「利用なし：76 件」である。

3.4 予測モデルの作成

属性毎に分類された 5 パターンのデータを用いてベイズ識別器を適用し、予測モデルを作成する。さらに、10-Fold 交差検証法を適用して、各予測モデルの未知のデータに対する予測能力を確認する。第 2 章に示した先行研究では、予測モデルの作成にロジスティック回帰分析を利用した。しかし、近年では、欠損値があるデータでも識別が可能という理由でベイズ識別器が利用されるケースが増えてきている。A 社では、将来的に業務パッケージの開発など、顧客の新規性と言った一部の項目を評価できないプロジェクトを予測対象に含めることが想定されるため、ベイズ識別器を採用した。予測モデルの作成、および 10-Fold 交差検証法の実施には、多数の機械学習用のライブラリを有する汎用言語 Python を利用した。

まず、リスク評価シートにより収集した 17 項目の評価結果を説明変数、成功／失敗の分類結果を目的変数に設定し、ベイズ識別器を適用して成否の予測モデルを作成した。そして、10-Fold 交差検証法を適用して、予測精度を確認した。属性毎に分類された 5 パターンのデータを用いて上記手順を実施したところ、表 4 に示す予測精度を有するモデルが得られた。また、予測精度の比較のために、179 件の全データを用いて同様にモデルを作成したところ、75.0%の予測精度を有するモデルが得られた。

4. 考察

4.1 予測精度の比較

本節では、まず先行研究と本研究のモデルの予測精度を比較し、次に本研究で実施した属性毎に分類したデータを利用して構築したモデルの予測精度の比較を行う。

(1) 先行研究と本研究の比較

先行研究のモデルの予測精度、および全データを利用して構築した本研究のモデルの予測精度の違いを比較する。先行研究では、88 件のデータを利用し、ロジスティック回帰分析を適用して

表4　属性毎に分類したデータを用いたモデルの予測精度

属性	分類	N	予測精度	*平均予測精度
プロジェクトタイプ	新規開発	75	74.4%	75.8%
	改良開発	90	77.3%	
	再開発	14	73.1%	
業務パッケージの利用	利用あり	103	70.1%	76.6%
	利用なし	76	85.4%	
全データ		179		75.0%

*平均予測精度は、各分類のNと予測精度から属性毎に平均値を算出したもの。

モデルを構築した結果、予測精度が 73.9%であった。一方で、本研究では、91 件のデータを追加し合計 179 件のデータを利用し、ベイズ識別器を適用してモデルを構築した結果、予測精度が 75.0%に向上した。向上した要因として、適用した手法の違いの影響が考えられる。

日本の特定の製造業企業の組込みソフトウェア開発プロジェクトを対象にプロジェクトの混乱発生の予測を試みた水野ら [14] によると、ロジスティック回帰分析に比べて、ベイズ識別器を適用して構築したモデルの方が高い予測能力を示したと述べている。本研究は、情報システム開発のプロジェクトの成否予測という水野らと類似の予測を実施している点から、同様の傾向が見られたと想定される。実際に、179 件のデータを利用して、ロジスティック回帰分析を適用してモデルを構築し、10-Fold 交差検証法を適用して予測精度を確認したところ、予測精度は 74.1%である。これより、ベイズ識別器を適用することで予測精度が向上していることがわかる。

(2) 属性毎の比較

属性毎に分類したデータを利用して構築したモデルの予測精度の違いを比較する。

まず、2 つの属性の平均予測精度を比較する。表 4 より 179 件の全データを利用して構築したモデルの予測精度が 75.0%であるのに対し、属性「プロジェクトタイプ」でデータを分割して構築したモデルの平均予測精度は 75.8%、属性「業務パッケージの利用」は 76.6%であった。これより、属性毎にデータを分割してモデルを構築することで、全体的な予測精度の改善が行えることが示唆された。

次に、属性毎にデータを分割して構築した 5 つのモデルの予測精度を比較する。表 4 より属性「プロジェクトタイプ」でデータを分割して構築したモデルの予測精度は、「改良開発」が 77.3%と相対的に良好な予測精度を示し、「新規開発」、および「再開発」が、それぞれ 74.4%、および 73.1%と低い予測精度を示した。その一方で、属性「業務パッケージの利用」では、「利用あり」が 70.1%であるのに対し、「利用なし」は 85.4%と顕著な違いを示した。

属性「業務パッケージの利用」において予測精度の顕著な違いを示した点から、現状の仕組みは業務パッケージを利用しないプロジェクトにおいて良好な精度の予測が可能であること、そして業務パッケージを利用するプロジェクトに対しては仕組みの改善の余地があることが示唆されたと言える。

4.2 本研究の成果と今後の課題

4.1 節の結果をもとに、A 社の関係者に、本成否予測の仕組みが優先して関与すべきプロジェクトの特定に有効であるかを確認した。その結果、多くの関係者が属性「業務パッケージの利用」毎にデータを分割して構築したモデルを利用すべきであり、本仕組みは実務において役に立つと述べていた。

属性「業務パッケージの利用」毎にデータを分割して構築したモデルを利用すべきと述べられた理由は、平均予測精度が 76.6%と最も高い点、そして改善課題が明らかであるため今後予測精度の向上が期待できる点を述べていた。また、実務において役に立つと述べられた主な理由は、本仕組みを適用する要求確定時がプロジェクトの初期のタイミングのためである。要求確定時は、プロジェクト全期間の 15%程度を経過したタイミングと言われている [13] （工数では 8%程度を消化したタイミング）。このタイミングで SM、および PMO の経験をもとにプロジェクトの最終的な成否を予想し、組織の関与の程度を決定するのは困難であり、70%を超える予測精度であれば、参考情報として十分に役立つ、という意見であった。さらに、SM、および PMO がプロジェクトメンバなどの関係者に組織の関与の程度を説明する際、A 社の過去実績にもとづいた説明が可能となるため合意を得やすい、という利点も述べていた。この結果受け、A 社では本仕組みを実務に適用中である。

以上より、本成否予測の仕組みは、属性「業務パッケージの利用」毎にデータを分割して構築したモデルを利用することで予測能力の向上が可能であり、そして組織が優先して関与すべきプロジェクトの特定に寄与できると考える。

今後の課題として、業務パッケージを利用したプロジェクトの予測精度の向上に注力したいと考える。業務パッケージを利用したプロジェクトの成否の要因を調査した先行研究 [15] [16] の成果を考慮すると、本研究のリスク評価シートの項目には見直しの余地があると考えられる。具体的には項目「パッケージの品質」、「パッケージの適合性」、「パッケージベンダのサポート」、および「上位マネジメントの関与」の追加を検討可能である。今後、これらの項目をリスク評価シートに追加し、A 社におい

てデータの収集、および予測モデルの作成を実施し、業務パッケージを利用したプロジェクトの予測精度を改善可能であるとことを確認したいと考える。

5. まとめ

本研究は組織が優先して関与すべきプロジェクトの特定に寄与することを目的に、属性毎に分類したプロジェクトのデータを活用することで、成否予測モデルの予測精度の向上を試みた。先行研究では、日本の IT ベンダ A 社にて 88 件のプロジェクトのデータを収集し、ロジスティック回帰分析を適用して予測モデルを構築したところ、73.9%の予測精度を有するモデルが得られた。本研究は、新たに追加したデータを含む 179 件のプロジェクトのデータを属性毎に分割し、ベイズ識別器を適用して予測モデルを構築し、予測能力の向上を試みた。その結果、属性「業務パッケージの利用」の 2 つの分類「利用あり」、および「利用なし」にデータを分割してモデルを構築することで、平均予測精度が 76.6%に向上することが示された。これらの結果を活用することで、本研究の目的に寄与できると考える。

今後の課題として、業務パッケージを利用しないプロジェクトのデータを用いたモデルの予測精度が 85.4%と良好な結果を示したのに対し、業務パッケージを利用するプロジェクトでは 70.1%である点を考慮し、業務パッケージを利用するプロジェクトの予測精度の改善に注力する予定である。

参考文献

[1] 経済産業省, 平成 26 年特定サービス産業実態調査(確報), 入手先〈http://www.meti.go.jp/statistics/tyo/tokusabizi/result-2/h26.html〉, 2015.

[2] 株式会社 日経 BP, プロジェクト実態調査 800 社, 日経コンピュータ, 2008 年 12 月 1 日号, pp.36-49, 2008.

[3] Carnegie Mellon University, CMMI for development Version 1.3, 2010.

[4] 河村智行, 高野研一, 情報システム開発の成否に影響を与える組織文化の要因の研究, 情報処理学会論文誌, Vol.53, No.12, pp.2854-2864, 2012.

[5] Takagi, Y., Mizuno, O. and Kikuno, T., An Empirical Approach to Characterizing Risky Software Projects Based on Logistic Regression Analysis, Empirical Software Engineering, Vol.10, No.4, pp.495-515, 2005.

[6] Mizuno, O., Kikuno, T., Takagi, Y. and Sakamoto, K., Characterization of risky projects based on project managers' evaluation, Proceedings of 22nd International Conference on Software Engineering, pp.387-395, 2000.

[7] Wohlin, C. and Andrews, A. A., Prioritizing and Assessing Software Project Success Factors and Project Characteristics using Subjective Data, Empirical Software Engineering, Vol. 8, pp.285-303, 2003.

[8] Kawamura, T. and Takano, K., Development of Project Outcome Prediction System for an IT Vendor, Proceedings of 2015 Asia-Pacific Software Engineering Conference, pp.338-345, 2015.

[9] Chen, Z., Menzies, T., Port, D. and Boehm, B., Finding the right data for software cost modeling, IEEE Software, Vol.22, No.6, pp.38-46, 2005.

[10] McLeod, L. and MacDonell, G. S., Factors that affect software systems development project outcomes: A survey of research, ACM Computing Surveys, Vol.43, No.4, Article24, 2011.

[11] The Standish Group International, CHAOS Summary 2009, The Standish Group International, 2009.

[12] Wixom, B. H. and Watson, H. J., An empirical investigation of the factors affecting data warehousing success, MIS Quarterly, Vol.25, No.1, pp.17-41, 2001.

[13] 独立行政法人 情報処理推進機構, ソフトウェア開発データ白書 2014-2015, 日経 BP 社, 2014.

[14] 水野 修, 安部 誠也, 菊野 亨, プロジェクト混乱予測システムのベイズ識別器を利用した開発--ソフトウェア開発現場への本格導入を目指して, SEC journal, Vol.1, No.4, pp.24-35, 2005.

[15] Mabert, V. A., Soni, A., and Venkataramanan, M. A., Enterprise resource planning: managing the implementation process, European Journal of Operational Research, Vol.146, No.2, pp.302-314, 2003.

[16] Sumner, M., Risk factors in enterprise-wide/ERP projects, Journal of Information Technology, Vol.15, No.4, pp.317-327, 2000.