

機械学習システムの要求

～なぜ機械学習プロジェクトはPoCで終わるのか～

荒川 純也, 石谷 規彦, 坂本 竜太, 定塚 和久, 長柄 昌浩

機械学習プロジェクトの問題点

機械学習プロジェクトは下記に示す特性から、一般的に従来のソフトウェア開発と比べて難易度が高いと云われている。

- ・従来プロジェクトに比べて不確実性が高い
- ・データ品質による影響度が高い
- ・顧客側の期待値が高い&理解度が低い

問題解決に向けたアプローチ

文献や論文より収集した機械学習プロジェクトの事例から、プロジェクトの結果に影響を与える要因を抽出・整理した。
更に機械学習に従事する有識者を対象としたアンケートを実施することにより、PoC段階でこれら要因の確認有無とプロジェクト結果を調査し、成功に導く重要な要件を分析した。

アンケート結果から見えてきた重要な要件と成功に向けた考察

重要な要件	次フェーズに繋がらなかった理由(自由記述抜粋)	ソフトウェア工学		機械学習工学	
		適用可能な手法	有効性	必要性	必要な手法
ビジネス 課題／ゴール	技術的可能性の検証が先行し、 <u>ビジネススキームへの展開方向が明確でなかった</u> 。	ゴール指向要求分析 (GQM+Strategies)	◎	×	-
	<u>ビジネスユースケースの価値の検討が不十分</u> な状態で始めてしまい、最終的に実証したいことにズレが生じてしまった。	ビジネスモデル キャンパス	○	×	-
運用後の コスト／体制	PoCを実施することで <u>開発費用がかさみ</u> 、提供段階で顧客の想定を超える金額となった。	見積手法	○	△	機械学習の不確実性を考慮した見積手法
	PoCを実施した結果、システム実装時の <u>リソース確保が難しい</u> ことが判明した。	ユースケース分析 見積手法 (デルファイ法等)	◎	×	-
データ収集 コスト／品質	<u>データが別部署にある</u> などクライアント側の制約が原因でPoC自体の課題設定が小さくなってしまった。	ステークホルダ分析	△	○	データ観点到特化した ステークホルダ分析
	PDCAサイクルを回す業務体制になっておらず、 <u>データ収集に対して理解が得られなかった</u> 。		△	○	Machine Learning Canvasの利用
	正解データを準備するなど <u>顧客が作業負担を大きく感じた</u> 。	見積手法	△	○	データ収集負荷を 考慮した見積手法

アンケート作成ポイント

アンケート作成時に考慮した
分析ポイントと質問内容

<分析ポイント>

1. PoC実施状況
2. ロール別, 立場別
3. 要件の重要度
4. 次に繋がらなかった要因
5. 要件が明確になった時期

質問項目	質問内容
回答者の経験	機械学習プロジェクト経験
	PoC実施経験
	PoC実施結果
PoCについて	PoC実施期間
	PoC実施体制
	回答者の役割・立場
	PoCの目的
要件が明確になった時期	ビジネス観点
	機械学習観点
PoCの結果	結果に影響を与えた要素
	次へ繋がる重要な要素

活動成果

- ・機械学習プロジェクトの実施事例収集と調査
- ・プロジェクト結果に影響を与える要因を抽出
- ・PoCに関するアンケートを実施することで、文献や事例には載っていない“生の声”を収集
- ・文献調査結果とアンケート結果を比較し、調査結果の妥当性を検証
- ・機械学習プロジェクトを成功に導く3つの“重要な要件”を抽出し、適用可能な手法を考察