

見積依頼書および開発実績を用いた 機械学習によるソフトウェア見積支援

キヤノンイメージングシステムズ株式会社

清水 大輔

shimizu.daisuke345@mail.canon

ソフトウェア見積における課題

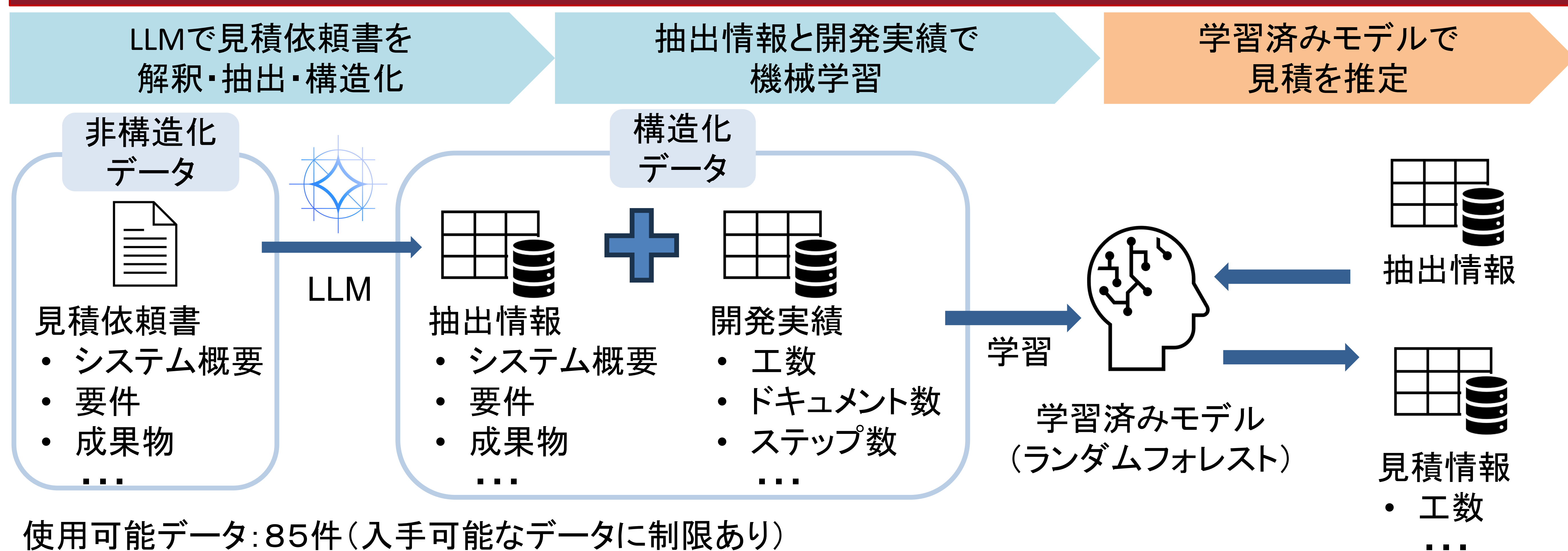
- ・ 見積作成の手法は存在するが、見積者の経験と勘に頼りがち
- ・ 見積もりの経験が浅い人や新規顧客の場合、時間がかかり精度も低くなりやすい
- ・ 顧客の要求が変更になった際に、やり直しのコストがかかる

手法・ツールの適用による解決

機械学習を用いた見積システム構築で解決

- ・ LLMで見積依頼書を解釈・抽出・構造化
- ・ 抽出情報と開発実績で機械学習
- ・ 学習済みモデルで見積を推定
- ・ 機密データを扱うためオンプレ環境で構築

提案手法：機械学習を用いた見積システム



実施内容・結果

- ・ オンプレミス環境で使用可能なLLM12種類、機械学習モデル6種類から精度の高いものを1種類ずつ選定

LLM モデル 候補	Gemma 2	Llama 2	mistral	Phi-3	DeepSeek-R1	LLaVA
	Gemma 3	Llama 3	Qwen2.5	Phi-4	MoonDream 2	Nomic Embed Text
機械学習 モデル候補	Random Forest	Linear	SVM (rbf)	Light GBM	XGBoost	k-NN
Accuracy	0.241	0.206	0.212	0.177	0.218	0.235

- ・ ランダムフォレストの決定木の本数、深さをデータ数に合わせて縮小することで予測精度が向上

- ・ 推定された見積情報と開発実績を比較した結果
Accuracy = 0.241
⇒ Accuracyが低い要因は、学習データの不足およびLLMによる解析誤差が影響

【今後の取り組み】

- ・ 学習データ拡充、LLM解析用プロンプト・学習データ抽出情報見直し、機械学習のパラメータ改善による予測精度向上
- ・ 新たなLLMモデルによる解析を試行、実用検討

