

LLMを用いたソフトウェア開発プロジェクト管理における自動データ分析の評価

2024年3月8日(金)

研究コース5 人工知能とソフトウェア品質 コース

LLM によるプロジェクト管理の高度化チーム

研究員：青山 弦太（株式会社ビッグツリーテクノロジー&コンサルティング）

主査：石川 冬樹（国立情報学研究所）

副主査：栗田 太郎（ソニー株式会社）

副主査：徳本 晋（富士通株式会社）

はじめに - LLMの活用度

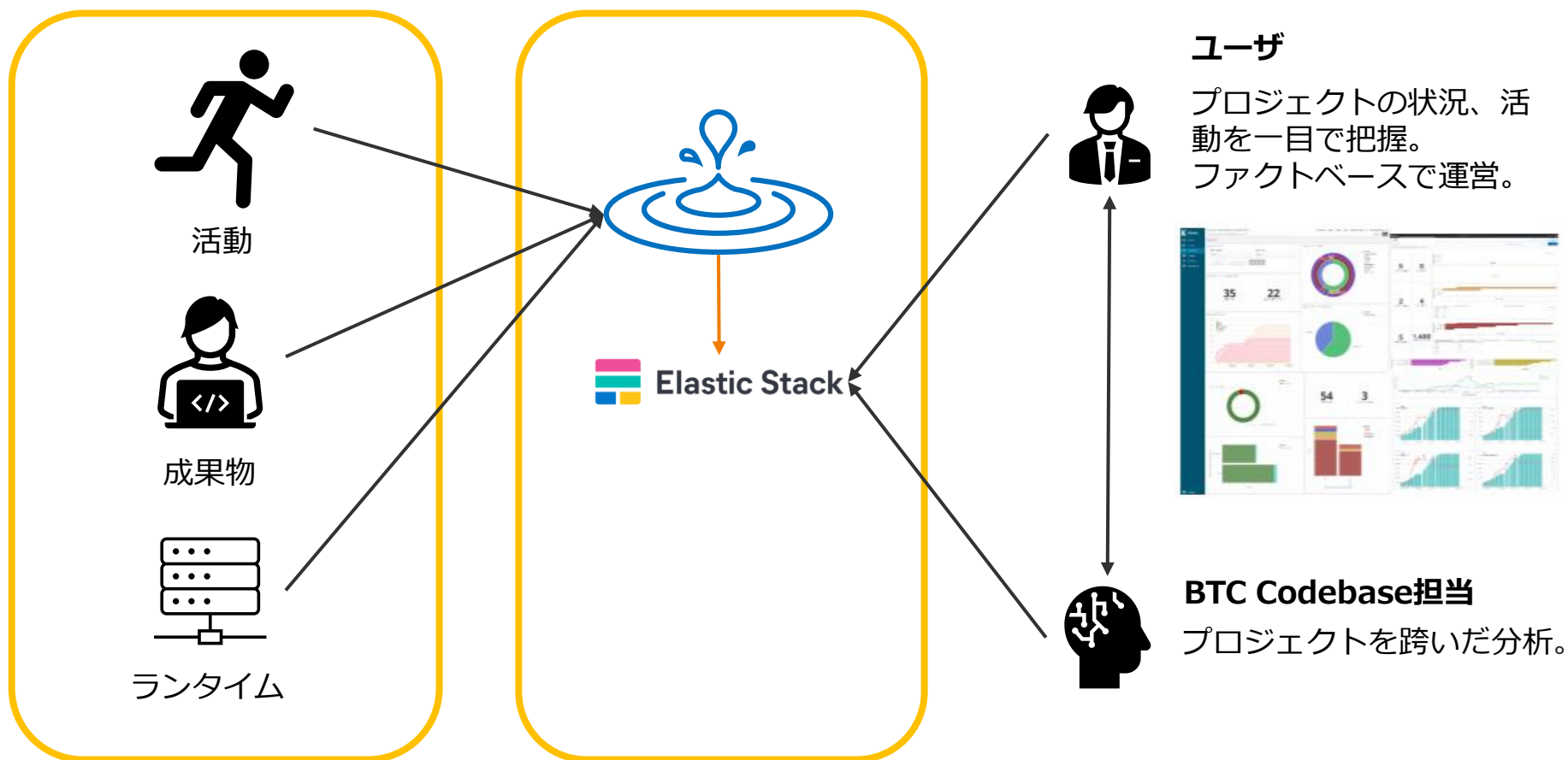
LLMの活用度についておしえてください！

Q1.ChatGPT等、LLM系のサービスを使ったことがありますか？

Q2.ChatGPT等、LLM系のサービスをどのような用途で使っていますか？

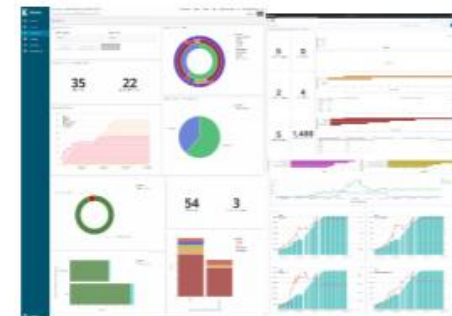
はじめに - 大まかな背景

自組織でソフトウェア開発に関するデータ分析基盤を作っている。



ユーザー

プロジェクトの状況、活動を一目で把握。
ファクトベースで運営。



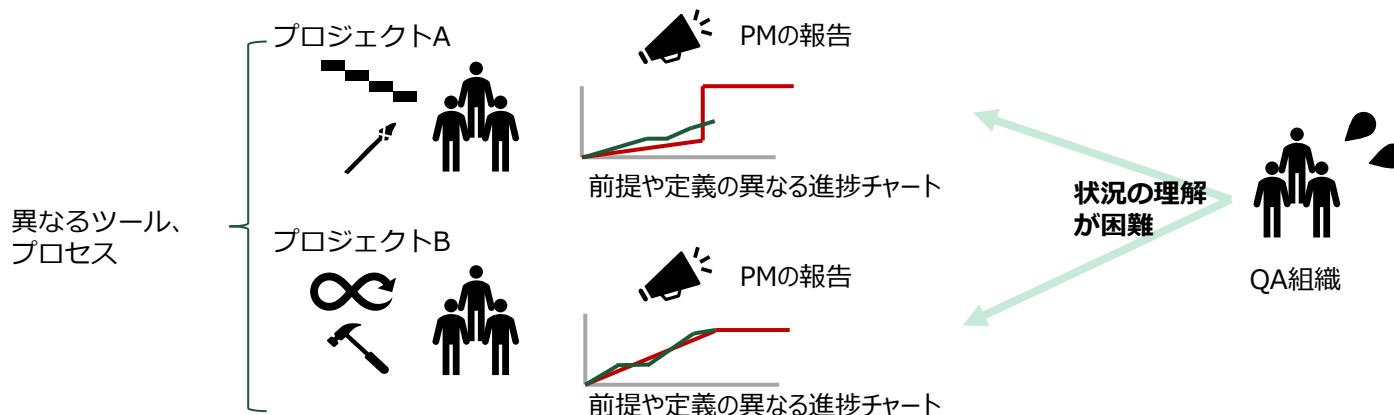
BTC Codebase担当

プロジェクトを跨いだ分析。

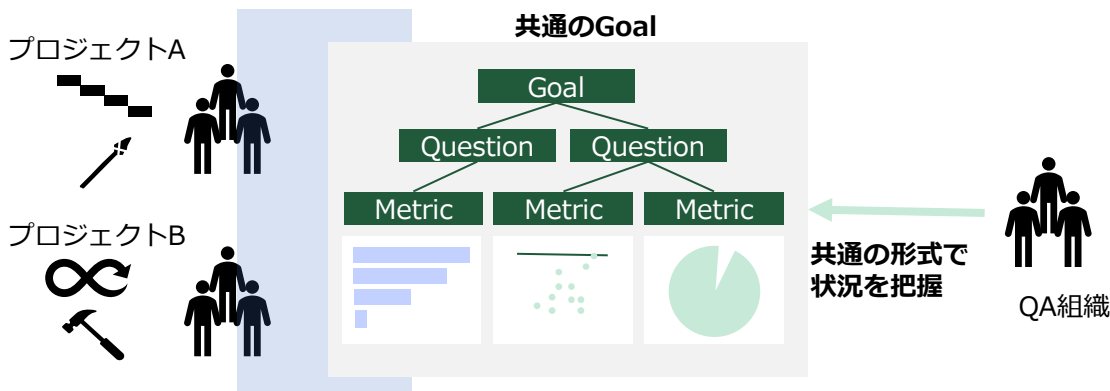
はじめに - 大まかな背景

自組織のQAチームが、多くの社内プロジェクトを共通の形式のダッシュボードで見ることで分析を行っている。

Before : 各プロジェクトの細かい事情を把握していない場合に、状況の理解が困難。



After : 共通の形式で網羅的にプロジェクトの状態が可視化されているため、状況の理解が容易！



はじめに - 大まかな背景

プロジェクトの分析は知識が必要な一方で、数が多く特定の専門知識を持つ人に負担がかかっている。

Goal

Question

Metrics

納期遅延なくプロジェクトを完了したい

計画は立てられている？

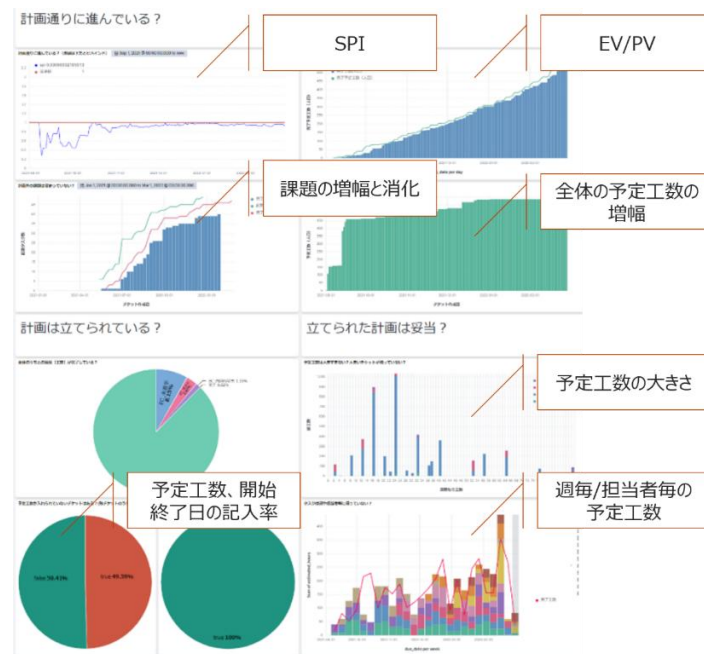
- 予定工数が入力されていないタスク数の割合
- 開始日終了日が入力されていないタスクの工数割合

立てた計画は妥当？

- タスクの予定工数の大きさ
- リソースに対する週ごとの予定工数

計画通りに進んでいる？

- EV/PV/SPI(完了工数/予定工数)
- 予定工数の増加率
- 課題の総数、消化数の推移



※チケット管理システムに登録された以下のデータを基に描画
開始日、期日、完了日、作成日、ステータス、予定工数、担当者

はじめに - 大まかな背景

プロジェクトの分析は知識が必要な一方で、数が多く特定の専門知識を持つ人に負担がかかっている。

Action

Metrics

LLWを用いることで、人の代わりにやってくれないかな？

納を

- 立てた計画は実行されている？
 - タスクの予定工数
 - リソースに対する週ごとの予定
- 計画通りに進んでいる？
 - EV/PV/SPI(完了工数/予定工数)
 - 予定工数の増加率
 - 課題の総数、消化数の推移

計画通りに進んでいる？

SPI

EV/PV

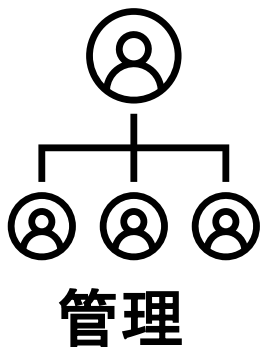
課題の増幅と消化

全体の予定工数の増幅

終了日の記入率

※チケット管理システムに登録された以下のデータを基に描画
 開始日、期日、完了日、作成日、ステータス、予定工数、担当者

はじめに - 管理への適用



調査



開発

- KKDを排除するために定量管理などデータが活用されてきた。
- 一方でLLMはあまり活用されていない。

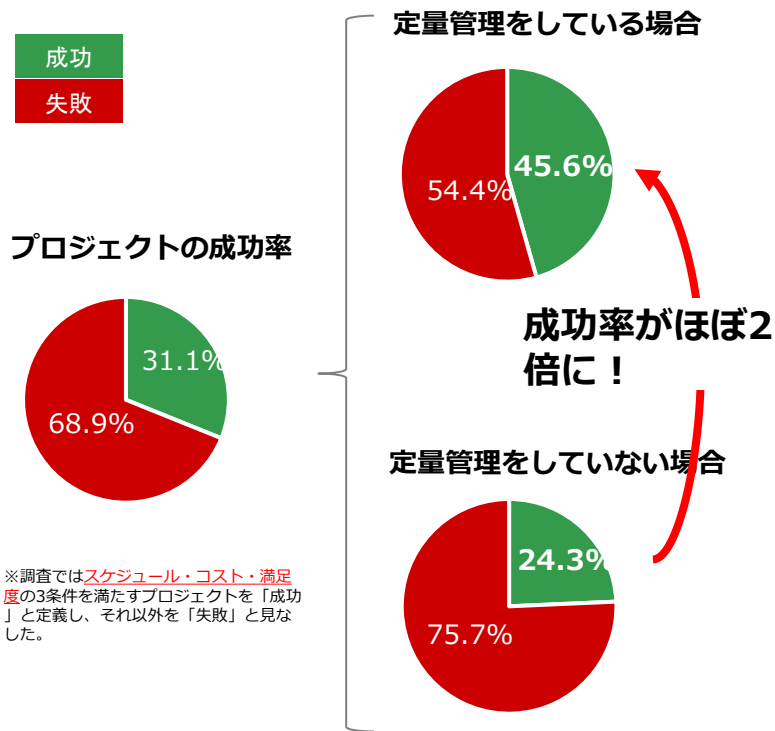
属人性を排除し品質をあげるために、管理にLLMを適用できるのでは？

背景 - プロジェクトの成功と定量管理

プロジェクトの成否はプロジェクト管理に強く依存している。
 その中でも、定量管理の実施有無で成功率に大きな差があるといわれている。

定量管理とプロジェクトの成功率

(測る企業は成功率が2倍に 日経コンピュータ 2015/11/30)



成否を左右する取り組みランキング

(ITプロジェクト実態調査 2018 日経コンピュータ 2018/3/1)

順位	取り組み	区分
1	プロジェクトが予定通りに進捗しない場合の対策などを事前に用意	リスク管理
2	プロジェクトの成果物を定量把握できる管理手法を導入	定量管理
3	プロジェクトに関わる全ての作業を文章化	文書化
4	プロジェクトに関わるコストを定量把握できる管理手法を導入	定量管理
5	ユーザ企業がRFP(提案要求書)を作成してベンダーを選定	文書化
6	プロジェクトのスケジュールの進捗を定量把握できる管理手法を導入	定量管理
...		

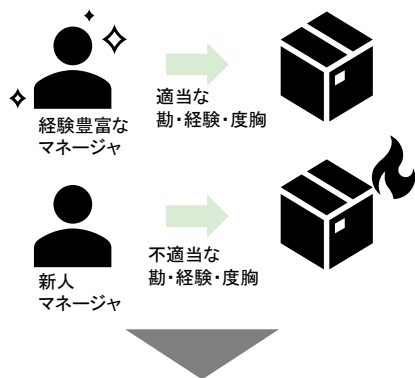
<https://www.ipa.go.jp/files/000072870.pdf>

背景 - データ駆動意思決定の理想と実態

定量管理、データに基づいたマネジメントはプロジェクトの成功に強く寄与する。
一方で、実際に利用されるべきと考えている人は多いが、利用できている実務者は多くないという問題がある。

KKDによるマネジメント

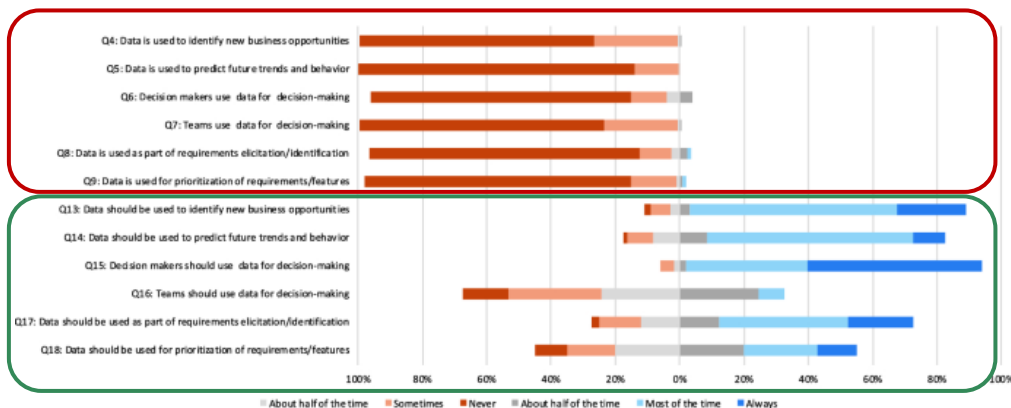
プロジェクトの成否がマネージャに強く依存している。



ASD(※)におけるデータ駆動型意思決定[1]

※ : アジャイル開発 Agile Software Development

Q4~Q9: 意思決定にほとんど**データ**が**利用**されていない。

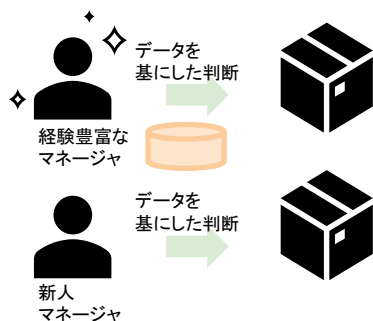


[1] Fig. 5. Use of data as part of decision making. Present (Q4~Q9) and future (Q13~Q18)

Q13~Q18: 意思決定に**データ**が**利用**されるべきと考えられている。

データに基づいたマネジメント

データに基づいて属人性を排除した再現性のある判断ができる。

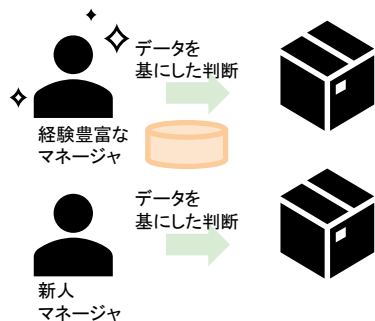


背景 - LLMによるPM分析業務の代替

LLMがデータに基づいたマネジメント実現の一助にならないかを考える。

問題

データに基づいたマネジメントをするべきと考えている人は多い。
一方で、実際にできている人は少ない。



原因(仮説)

リソース不足、分析スキル不足により実施されないことが原因では？

リソース(時間、人的コスト)

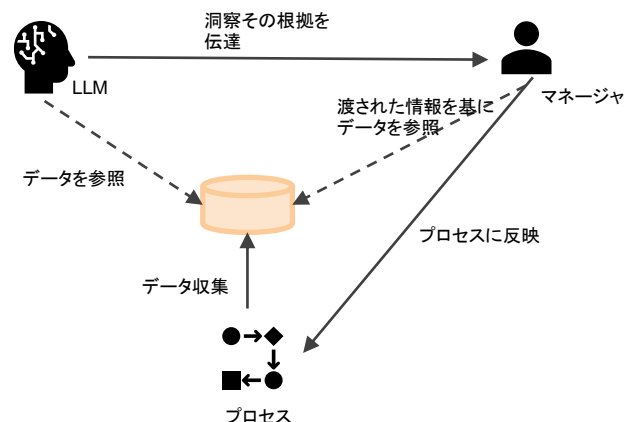
管理工数は十分に確保されないのでは？
忙しいとまず生産に回ってしまうのでは？

分析スキル

管理工数は十分に確保されないのでは？
忙しいとまず生産に回ってしまうのでは？

解決案

LLMに洞察を獲得してもらうことで、スキルやリソースを賄えないか？



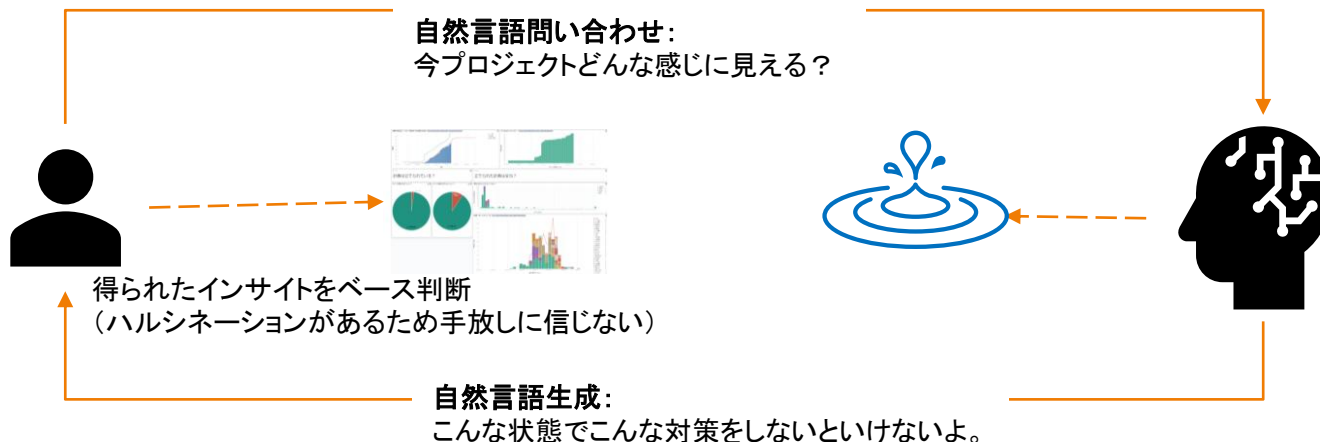
テーマ

PMによる分析の一部をLLMにより代替することで、
意思決定までの時間短縮, 意思決定の品質向上が可能なのではないか？

※意思決定の品質：決定されたことの根拠が強いことや、決定したことがより目的に寄与するなど、総合的な意思決定の期待効果ととらえる。

提案 - LLMによる定型確認

特に本発表では、探索的な分析に用いるのではなく、定型的な確認をLLMに実行させることに絞る。



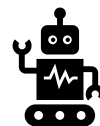
定型レポート生成

特徴

- ・固定されたプロンプトで定期的に自動実行する

メリット

- ・入出力を簡単にエンドユーザがコンピューティング可能



マシン支援型インサイト

<https://yellowfin.co.jp/blog/29-hqblog1-what-is-augmented-analytics>

特徴

- ・データを見て気になったことを自由に質問する。
- ・困っていることをどう解決すればよいか自由に質問する。

メリット

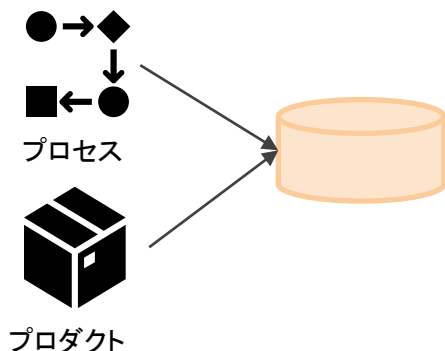
- ・その場タイミングで困っていることを対話的に質問できる。
- ・探索的データ分析が民主化される。

提案 - GQMベースでの定型確認

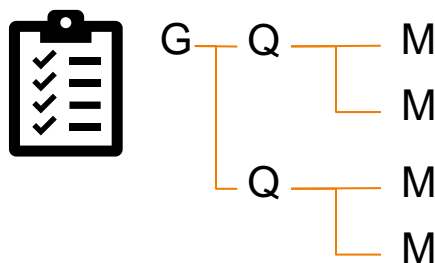
人間がデータを読み取る時は、GQMなどをベースに注視観点を定め分析をする。
 本発表では、LLMにもGQMを基に作られた注視観点を基に意志決定させる。

前提

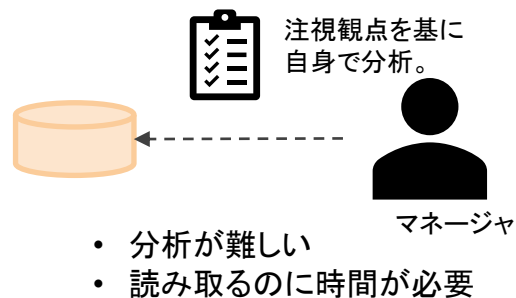
組織的なデータが収集されている



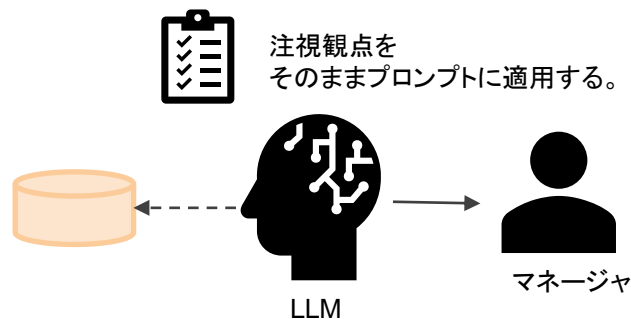
データの標準的な注視観点が存在する



ASIS: (自組織での運用)



TOBE:



- 自動で分析
- 結果だけを見て判断する。

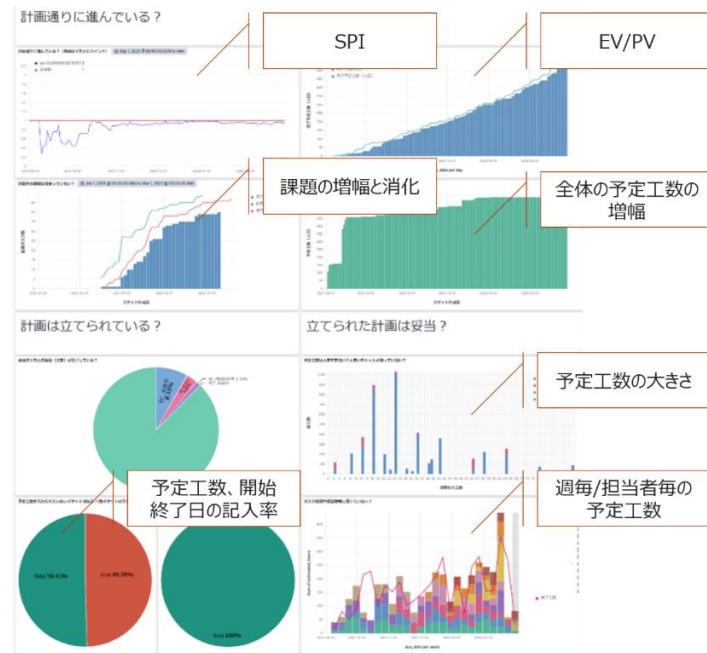
提案 - 自組織での適用例

自組織内で定めている、進捗に関するGQMとデータに対して、人間による分析をLLMによって代替する。



納期遅延なくプロジェクトを完了したい

- 計画は立てられている？
 - 予定工数が入力されていないタスク数の割合
 - 開始日終了日が入力されていないタスクの工数割合
- 立てた計画は妥当？
 - タスクの予定工数の大きさ
 - リソースに対する週ごとの予定工数
- 計画通りに進んでいる？
 - EV/PV/SPI(完了工数/予定工数)
 - 予定工数の増加率
 - 課題の総数、消化数の推移



※チケット管理システムに登録された以下のデータを基に描画
 開始日、期日、完了日、作成日、ステータス、予定工数、担当者

提案 - 分析のためのプロンプト

対象データを読み込ませ、分析するためのプロンプト例。

添付ファイルのデータ(以下、WBSとする)を分析して欲しいです。
新人プロジェクトのマネージャの管理しているプロジェクトのWBSです。

WBSには、以下の前提条件があります。
 ・プロジェクトメンバーの休暇は日本の休祝日に準拠する。土日も休日とする。
 ・それぞれのメンバーは正社員で8月から10月までは専任でこのプロジェクトでの業務を行う
 ・C山は経験が浅い。
 ・本日は8/21とする。
 ・プロジェクト期間を2023/3/10-2023/10/20とする。

以下のあなたは上位のプロジェクトマネージャとして、
以下の3個の要求にこたえてください。

要求1:計画が破綻していないか、以下の観点でレビューしてください。
 ・計画を立てるために必要な情報(開始予定、期日、予定工数、担当者)が入力されているか。
 ・同じ担当者が実施するタスクに期間の重複はないか。(オーバーラップ)
 ・予定工数に対して十分な期間が設定されているか。(実働時間と期日の差分)
 ・タスクは十分な大きさに分解されているか。
 ・タスクの依存関係は加味されているか。

要求2:上記に対して、EVM(EV/PV/SPI)を用いた管理を行いたい。以下の質問に答えてください。
 Q-1.EVM(EV/PV/SPI)を用いた管理を行えるデータの入力状況であるかを教えてください。
 Q-2.Q-1を満たしている場合に、現在のEV/PV/SPIを求めて、計画は順調に進んでいるかを教えてください。
 PVは、完了予定日に予定工数がPVとして計上されるよう計算してください。
 Q-3.Q-2の結果を基に、プロジェクトマネージャとしてどのようなアクションを起こすべきかを教えてください。

要求1,2を通して、最終的なプロジェクトの状況、必要なアクションを以下の形式で教えてください。
 必要なアクションについては、要求1,2の情報を基にアクションを提案してください。
 また、すべてのポイントがアクションに含まれていることを確認してください。

【進捗の良しあし】
 SPI: ★SPIが書かれる
 プロジェクト期間進行度: ★プロジェクトの全期間に対する経過期間の割合が書かれる
 プロジェクト完了進行度: ★プロジェクト全体のPVIに対する現在のEVの割合が書かれる
 【計画の良しあし】
 ★計画の良しあしが書かれる
 【必要なアクション】
 ・★必要なアクションが箇条書きで書かれる
 ・★必要なアクションが箇条書きで書かれる
 ・★必要なアクションが箇条書きで書かれる

プロジェクト変数部

- WBSには、以下の前提条件があります。
- ・プロジェクトメンバーの休暇は日本の休祝日に準拠する。土日も休日とする。
 - ・それぞれのメンバーは正社員で8月から10月までは専任でこのプロジェクトでの業務を行う
 - ・C山は経験が浅い。
 - ・本日は8/21とする。
 - ・プロジェクト期間を2023/3/10-2023/10/20とする。

プロジェクトに関する定量情報や、定性情報を自然言語で記載する。

GQM分析部

- 要求1:計画が破綻していないか、以下の観点でレビューしてください。
- ・計画を立てるために必要な情報(開始予定、期日、予定工数、担当者)が入力されているか。
 - ・同じ担当者が実施するタスクに期間の重複はないか。(オーバーラップ)
 - ・予定工数に対して十分な期間が設定されているか。(実働時間と期日の差分)
 - ・タスクは十分な大きさに分解されているか。
 - ・タスクの依存関係は加味されているか。

- 要求2:上記に対して、EVM(EV/PV/SPI)を用いた管理を行いたい。以下の質問に答えてください。
 Q-1.EVM(EV/PV/SPI)を用いた管理を行えるデータの入力状況であるかを教えてください。
 Q-2.Q-1を満たしている場合に、現在のEV/PV/SPIを求めて、計画は順調に進んでいるかを教えてください。
 PVは、完了予定日に予定工数がPVとして計上されるよう計算してください。
 Q-3.Q-2の結果を基に、プロジェクトマネージャ

GQMのマトリクスの計算、Questionへの回答を行う。
 マトリクス間に何らかの依存関係がある場合は、ステップを分けて記載する。

出力調整部

- 要求3:
 要求1,2を通して、最終的なプロジェクトの状況、必要なアクションを以下の形式で教えてください。
 必要なアクションについては、要求1,2の情報を基にアクションを提案してください。
 また、すべてのポイントがアクションに含まれていることを確認してください。
 【進捗の良しあし】
 SPI: ★SPIが書かれる
 プロジェクト期間進行度: ★プロジェクトの全期間に対する経過期間の割合が書かれる
 プロジェクト完了進行度: ★プロジェクト全体のPVIに対する現在のEVの割合が書かれる
 【計画の良しあし】
 ★計画の良しあしが書かれる
 【必要なアクション】
 ・★必要なアクションが箇条書きで書かれる
 ・★必要なアクションが箇条書きで書かれる
 ・★必要なアクションが箇条書きで書かれる

最終的に人間が欲しい情報のみを、人間が見やすい形式で出力する。

提案 - プロンプトの適用例

該当プロンプトでプロジェクトのデータを分析した際のアウプットサンプル。

調子のよいプロジェクト

要求3: 最終的なプロジェクトの状況と必要なアクション

進捗の良しあし

- SPI : 1.42
- プロジェクト期間進捗度 : 約28.57%
- プロジェクト完了進捗度 : 約37.96%

計画の良しあし

- プロジェクトは予定より進んでおり、計画は良好です。

必要なアクション

- **進捗の継続的な監視:** 予定より進んでいるものの、継続的な監視と進捗状況の評価が必要です。
- **リソースの効率的な管理:** 進捗に合わせてリソースの再配分や優先順位の調整を行います。
- **ステークホルダーとのコミュニケーション:** 進捗状況をステークホルダーに定期的に報告し、必要に応じてフィードバックを得ます。

これらの分析とアクションは、2023年3月21日時点のプロジェクト状況と利用可能なデータに基づいて提案されています。定期的なレビューと適切な対応策の実施が重要です。[-]

調子の悪いプロジェクト

要求3: プロジェクトの状況と必要なアクション

進捗の良し悪し

- SPI : 0.49
- プロジェクト期間進捗度 : 約70.21%
- プロジェクト完了進捗度 : 約50.39%

計画の良し悪し

- 計画には改善の余地があり、特に顧客レビューの遅れが進捗に影響しています。

必要なアクション

- 顧客レビューの遅れに対処するための対策を講じる。
- タスクの期間やリソース割り当ての見直しを行う。
- 定期的な進捗レビューと計画の見直しを行う。

これらの分析とアクションは、現在のプロジェクト状況と利用可能なデータに基づいて提案されています。定期的なレビューと適切な対応策の実施が重要です。[-]

実験 - 実験のコンディション

ツール: ChatGPT(GPT4)

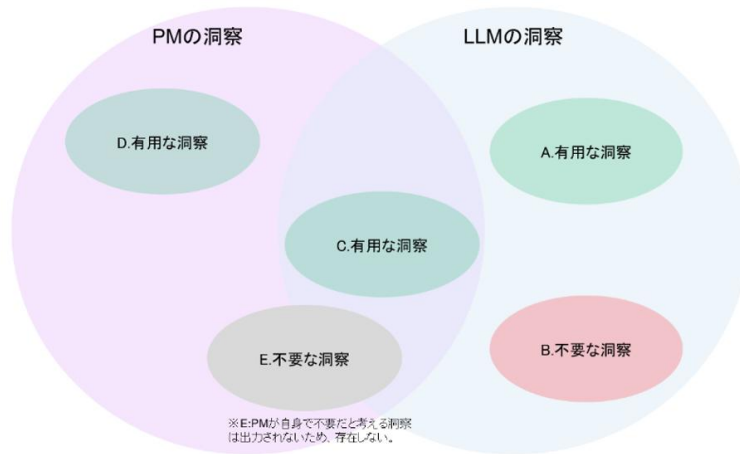
プロンプト: 前項に示した要領で作成

データ: 前項に示したものを使用

実験方法:

組織内の特定のプロジェクトに対して、設計フェーズの特定のタイミング(20%/50%/70%の期間が経過したタイミング)の進捗に関するデータを用意。

データに対して、PMとLLMそれぞれに洞察を出力させ、その内容を比較する。



洞察分類	特性
A.LLM独自の有用な洞察	人間が気づけない洞察がでているため、付加価値が想像できている。
B.LLM独自の不要な洞察	不要な洞察を無駄に上げているため、ノイズである。
C.LLM,PM共有の有用な洞察	人間と同等の洞察を人的コストなく導出できており、信頼性の補強にもなる。
D.PM独自の有用な洞察	人間が気づく洞察を、漏らしている。
E.PM独自の不要な洞察	存在しない:人間が不要と考える洞察を人間は見つけないため。

※出現した洞察に対する有用、不要は人間が判断する。

実験 - 実験結果

対象プロジェクト規模：10名程度
開発工程にスプリント開発を適用。

洞察分類	件数	内容
C.LLM,PM共通の有用な洞察： 共通	8	<p>計画はきちんと立てられている * 3 開始日期日予定工数の入力率が95%以上</p> <p>計画は妥当である * 2 タスクは十分に分割されている。 週毎、担当者毎に適切に割り当てられている</p> <p>計画通りには進んでいる。 * 1 計画通りにタスクが完了している。</p> <p>計画通りには進んでいない。 * 2 原因は顧客のレビュー待ちが多発している 計画通りにタスクが完了している。</p>
A.LLM独自の有用な洞察： 本手法独自	1	<p>計画の妥当性に一部問題がある 担当者が並列で複数のタスクに対応するスケジュールが常態的に立てられている。</p>
D.PM独自の有用な洞察： LLMの漏れ	1	<p>計画の妥当性に一部問題がある 対応量が多い週と少ない週があり計画の一部が妥当ではない。</p>
B.LLM独自の不要な洞察： 別に要らない	0	

C.LLM,PM共通の有用な洞察 が多い。

共通の有用な洞察が多く、以下の可能性を示せたと考えている。

- PMの手間削減による**意思決定の時間短縮**。
- PMの意思決定を支える根拠となり**意思決定の信頼性向上**。

A.LLM独自の有用な洞察 が一部存在する。

計画の妥当性に一部問題がある

担当者が並列で複数のタスクに対応するスケジュールが常態的に立てられている。

チケットをミクロに参照することで気づく洞察である。

人間では手間がかかるような、多くのチケットを細かく見ないと発見できない洞察は、プログラムやLLMの得意とするところであると考えている。

D.PM独自の有用な洞察 が一部存在する。

計画の妥当性に一部問題がある

対応量が多い週と少ない週があり計画の一部が妥当ではない。

プロンプトや注視観点として明言されていないが、PMが探索的に気付いた洞察であり、LLMが気づかないのは妥当な洞察であった。

プロンプトに明言することでLLMでも洞察を得ることが可能になるが、トークンが多く制限にかかることや、健忘してしまう可能性がある。以下の2点を行うことで、根本的な解決につながるのではと考えている。

- 探索的に気付く洞察は人間,もしくは別のプロンプトでみつけるなど,プロンプトと人間の役割範囲を明確にする。
- RAG (Retrieval-augmented Generation) を導入し,組織のコンテキストや知識をLLMに参照させることで,組織のコンテキストから探索的に気付く洞察を見つけさせる。

まとめ

仮説

PMによる**分析の一部をLLMにより代替**することで、**意思決定までの時間短縮**, **意思決定の品質向上**※が可能なのではないか？

結論

PMの**分析の一部LLMにより代替**でき、以下に寄与する可能性を示せた。

- **意思決定までの時間短縮**
- **意思決定の品質向上**

(特に**意思決定の根拠を補強**することによる**信頼性の向上**)

※意思決定の品質：決定されたことの根拠が強いことや、決定したことがより目的に寄与するなど、総合的な意思決定の期待効果ととらえる。

今後の展望

今後は、実務に即した性能(*)の評価,及び,向上のため、以下の4点に取り組む。

A.より正確な性能評価

実験対象プロジェクト数を増やす。

インプロセスのプロジェクトへの適用。

B.性能向上

RAGの導入：

組織のコンテキストや知識をLLMに参照させる。

より広い範囲の柔軟な気づきが増え人間でしか見つけられない気づきを減らす
/LLM独自の気づきを増やす ことを期待する。

Human in the loopの意識：

実務でどのような運用を行うのかを定義し、それに即した出力を行うようプロンプトを最適化する。**PMが行う作業をより効果的に代替でき、意思決定の時間短縮**につながることを期待する。

※「意思決定までの時間短縮、意思決定の品質向上」への寄与度を性能とする。