

トピックモデリングを用いたLLMによる要件の自動矛盾検知の手法

オープンソースソフトウェア工学研究室
S2310236 古川陽希

要件定義

システムが備えるべき機能や性能を明確に定義する工程



要件の種類や内容

①**機能要件**：システムを用いて直接行う振る舞いを定義

-検索機能,登録機能…

②**非機能要件**：機能以外の品質や制約事項を定義

-性能,**セキュリティ**…

要件定義の現状

要件同士の間が矛盾が無い必要がある

①通信障害時や電波の届かない店舗であっても決済を行えるよう、**オフライン状態**での決済処理を許可し、完了できなければならない。

②残高の改ざんや二重支払いを確実に防止するため、すべての決済処理において必ず**オンライン通信**を行い、サーバー側の承認を得る必要がある

⇒×

手動による意味論的な分析が必要

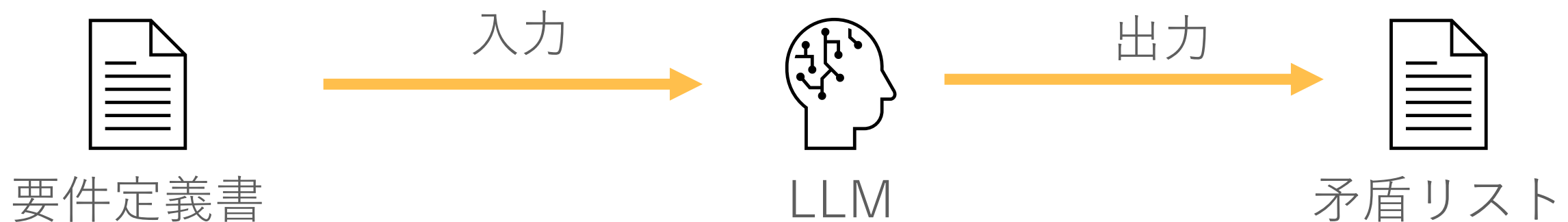
人間の専門家が行っていた

- ・膨大なコスト
- ・非常に難しいタスク

⇒自動矛盾検知によって負担を軽減

自動矛盾検知と課題

要件の総当たり



大規模な要件定義書の処理

- ・ 比較回数が**多**すぎる

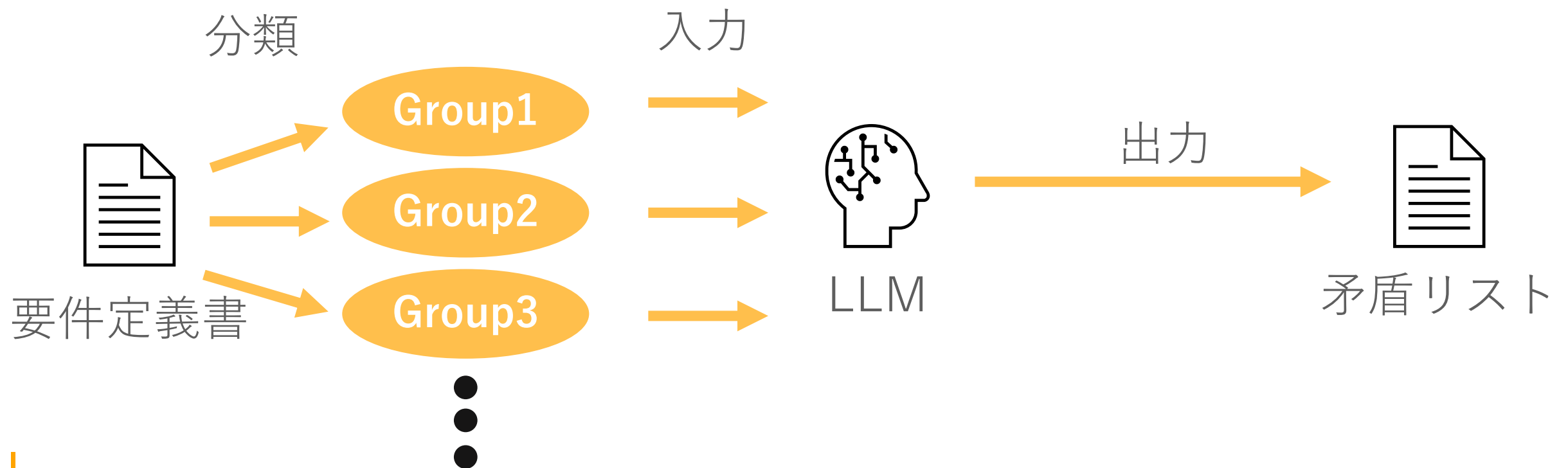
例) 200個の要件の計算量 = $\frac{200(200-1)}{2} = 19900$ 回

⇒ LLMの精度が低下

- ・ 出力をセーブしてしまう

クラスタリングによる自動矛盾検知

似たことが記述された要件同士で矛盾がある仮定



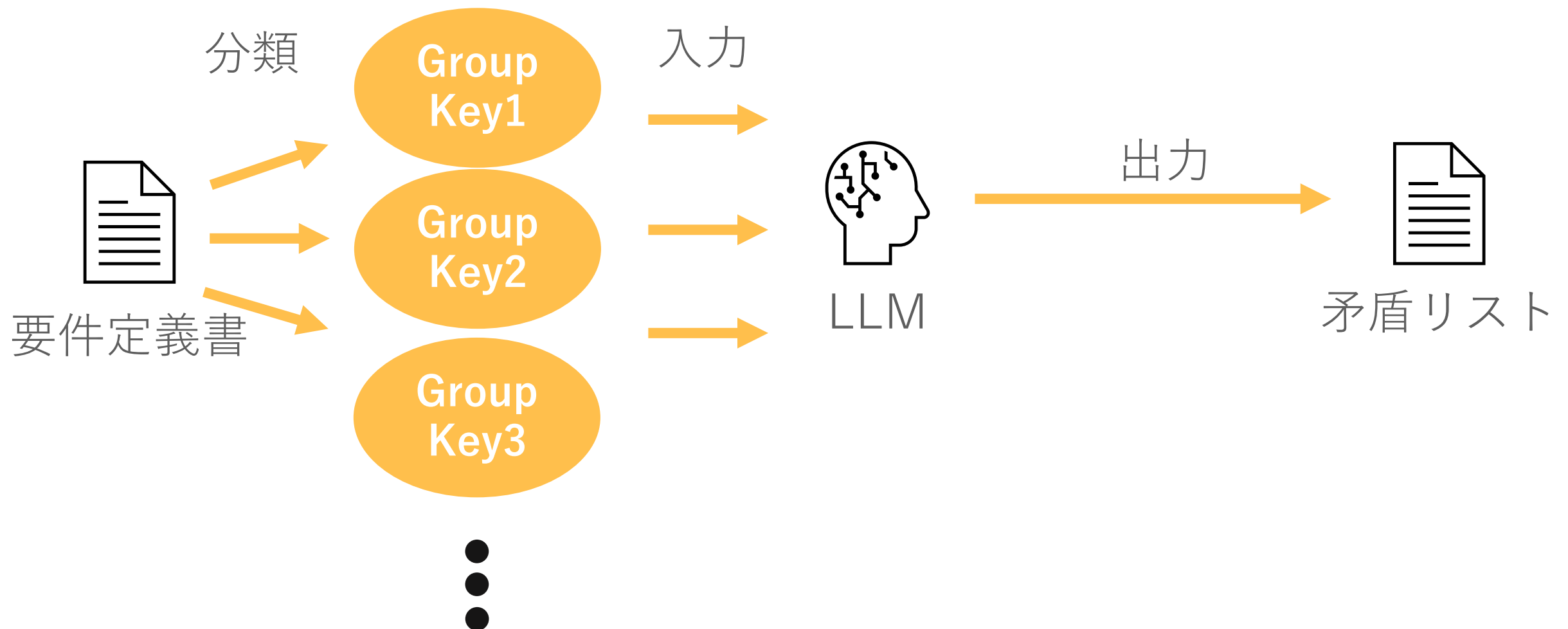
グループごとにLLMが矛盾を検証

- ・ 一回の比較回数が劇的に減少

先行研究

ルールベースによるクラスタリング

- ① キーワードをあらかじめ人が決定
- ② キーワードを含む要件をグループに分類する
- ③ グループごとにLLMが矛盾検証

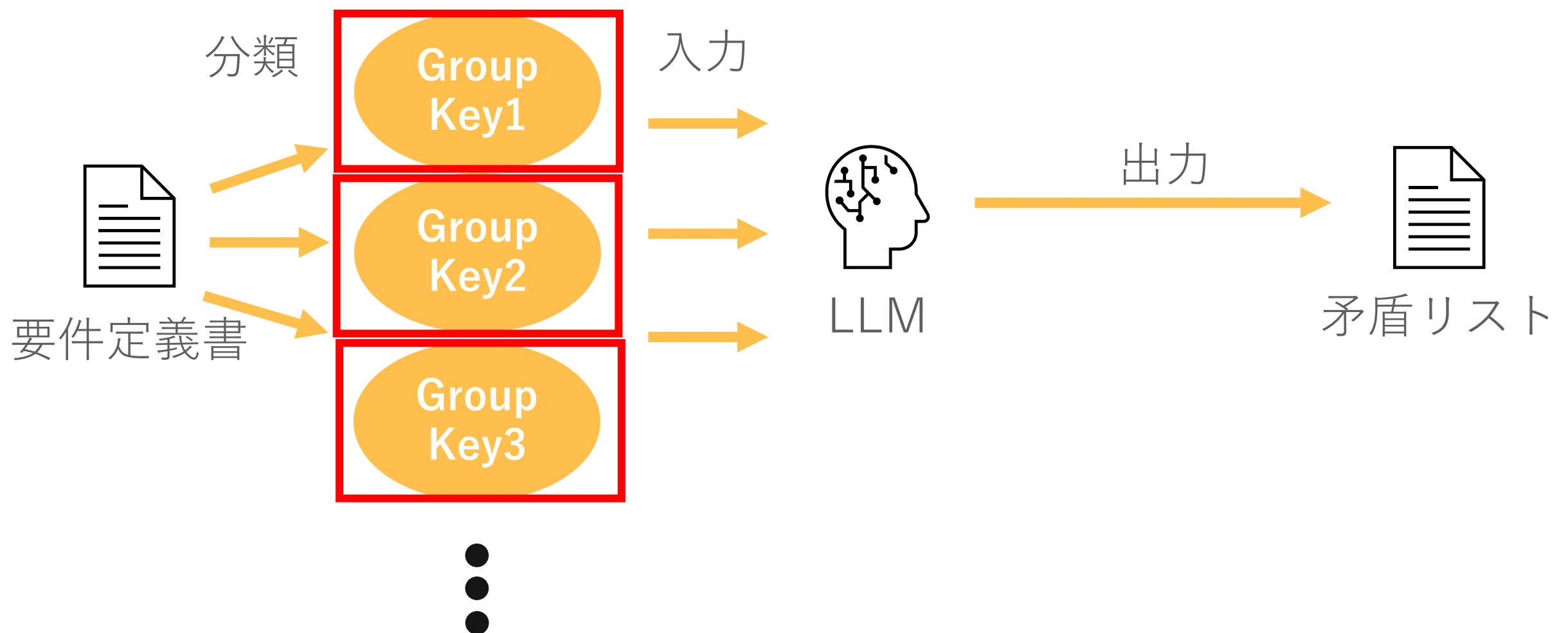


先行研究の課題

ルールベースによるクラスタリング

×それぞれの要件定義書で人間がキーワードを決める必要がある

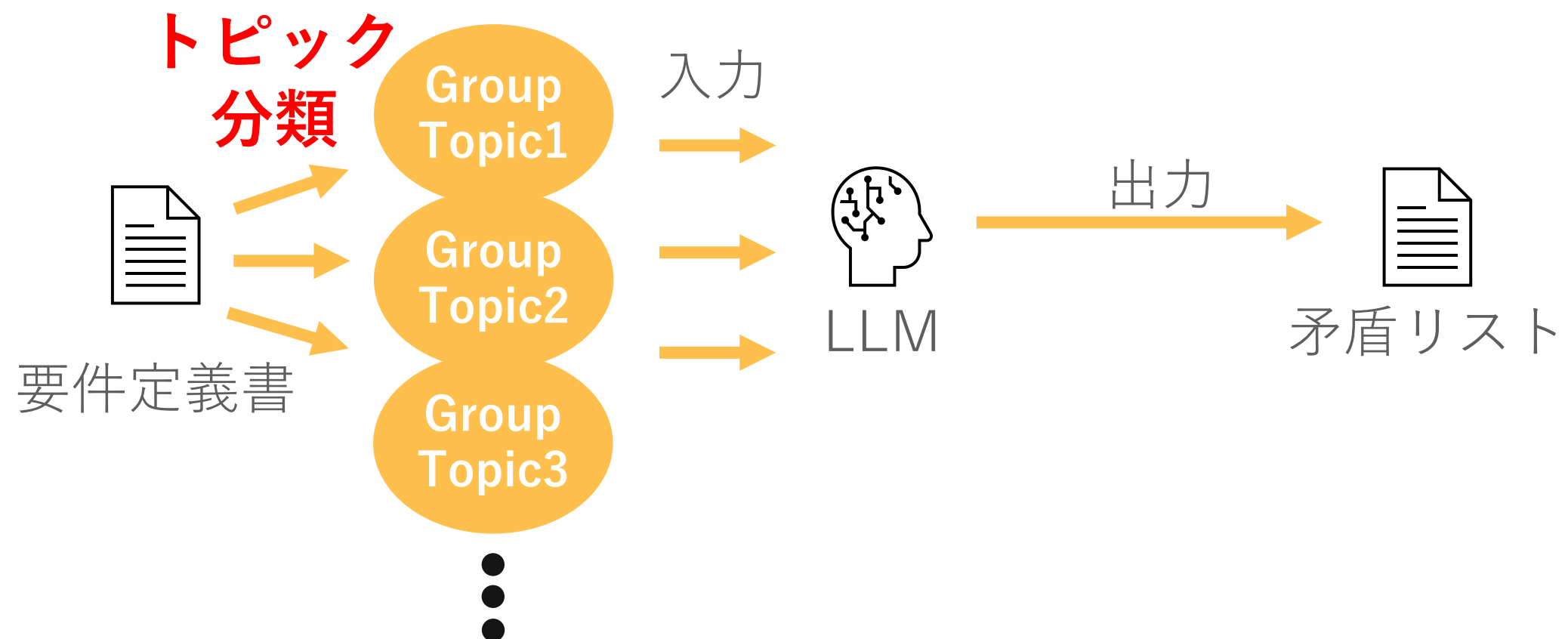
人間がキーワードを決定する



本研究について

要件に対して**トピックモデリング（BERTopic）**を行って分類

- ・ **教師無し学習**：潜在的なトピックを自動抽出
- ⇒ トピックを人間が考える必要がない



研究目的

トピックを考えなくても分類し矛盾を検知できる

- ・ 教師無し学習：潜在的なトピックを自動抽出

⇒ トピックを人間が考える必要がない

Prompt2（ゼロショット）

	ChatGPT	Claude	Gemini
適合率	0.464	0.515	0.483
再現率	0.367	0.533	0.400
F1スコア	0.374	0.433	0.294

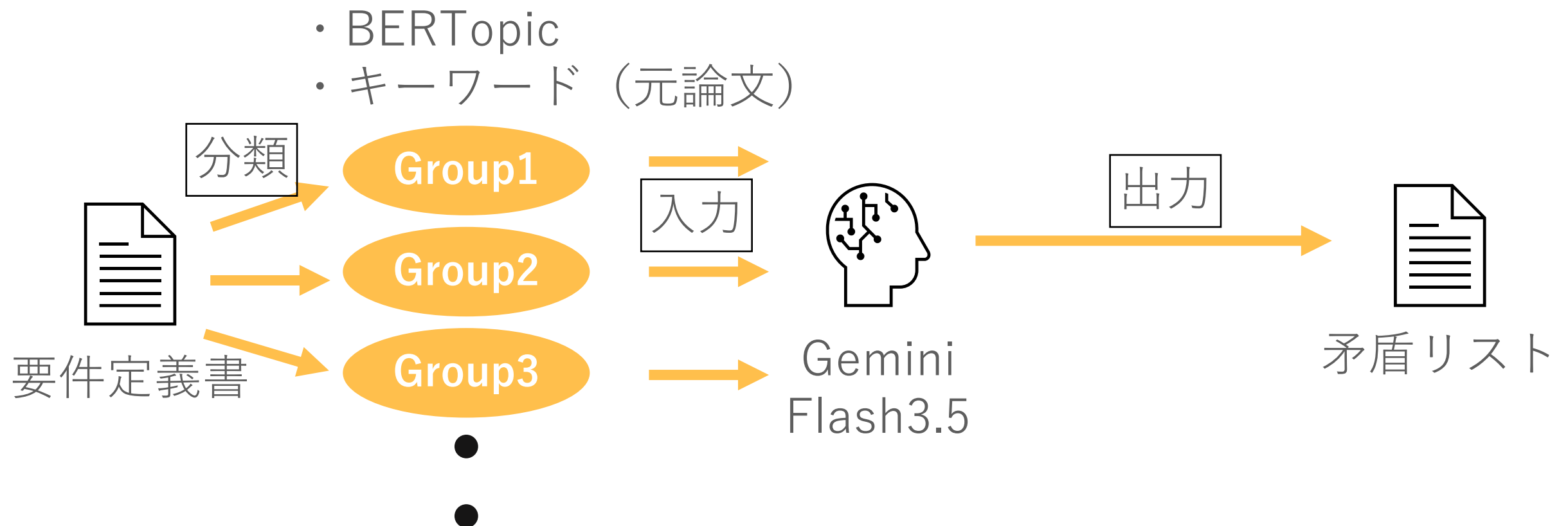
Prompt3(ルールベースのクラスタリング)

	ChatGPT	Claude	Gemini
適合率	0.689	0.734	0.700
再現率	0.633	0.700	0.400
F1スコア	0.660	0.717	0.509

実験

実験手順

- ①元論文で使用されたバグを仕込んだデータを用意
- ②BERTopicを用いてトピックモデリングを行う
- ③分類されたデータのみでGemini3.5Flashで矛盾検知を行う
(元論文の手法も実施)



正解ラベル付け

正解

◎意図的に仕込んだ矛盾

○仕込んだ矛盾によって発生した意図しない矛盾

○元々存在していた未発見だった矛盾

×矛盾していない

今後

| 元論文で言及されていない矛盾のペアの評価

今回の実験で選ばれた要件のペアが矛盾しているかしていないかを複数名に評価してもらう。