

フィルター活用可能な 大規模言語モデル構築

総合情報学科 知能情報システム研究室

永井ゼミ 4年

板橋 賢志

目次

- 研究の動機と目的
- 関連(既存)研究
- 研究環境
- モデル構築案
- 検証
- 研究計画

動機

- レビューやSNSフィルタリングを行う必要があるがそのフィルタリングを回避するユーザーが現れる。これはパターンマッチングの限界である。そこで大規模言語処理モデルを活用したフィルターの開発を模索する。

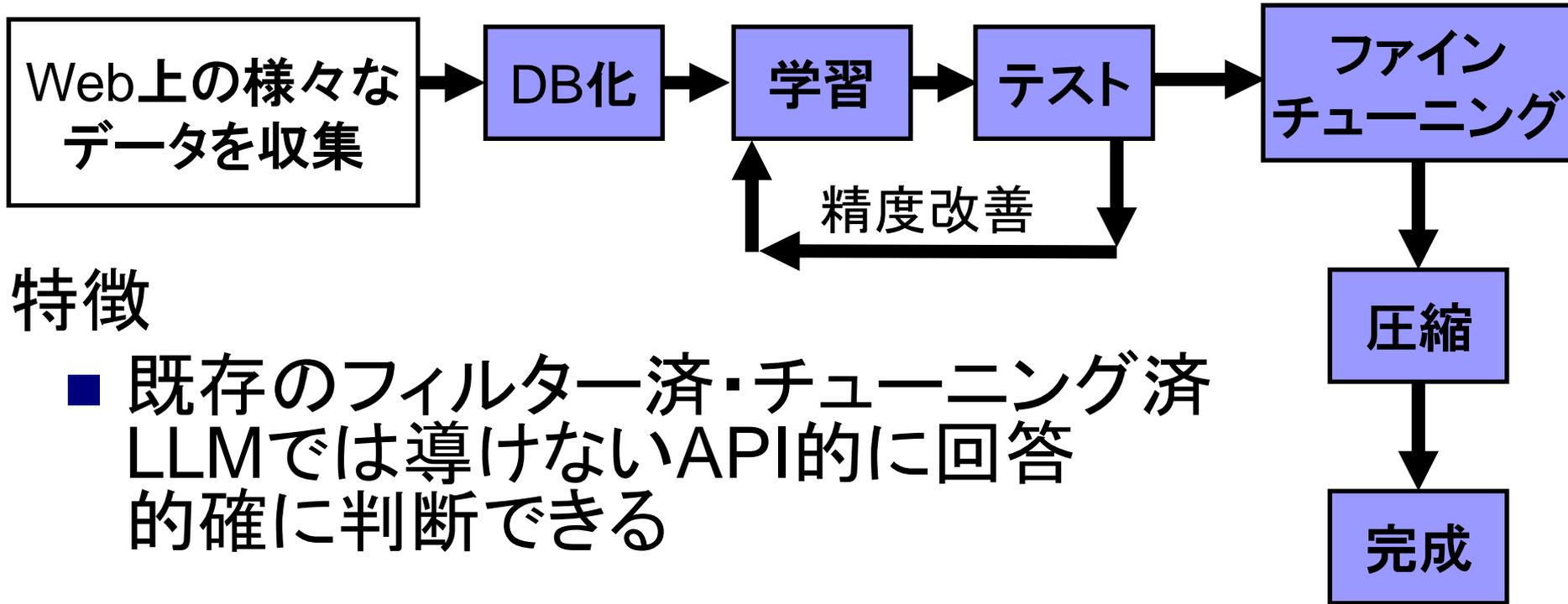
目的

- 大規模言語処理モデルを活用したフィルターの開発。

関連研究(既存研究)

- **Language Models are Few-Shot Learners** 2022-11-21 閲覧
- **Improving Language Understanding by Generative Pre-Training** 2023-7-5 閲覧

モデル構築案



特徴

- 既存のフィルター済・チューニング済 LLMでは導けないAPI的に回答的確に判断できる

→この点が研究たる所以と
事前学習から行う必要性

具体例

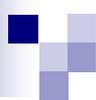
- Transformer Decoder系統（GPTを踏襲）
- Filter GPT（仮）をトレーニング
- ファインチューニングを行う
- Few-Shot One-Shotで検証
- Loraで圧縮を試みる

検証

- ネットスラングや同音異義語など判定が難しいワードを含む適当なワードを作成
- NGと答えることができるかを判定
- NGに見えるけどよく見るとNGでもなんでもない適当なワードを作成
- OKと答えることができるかを判定

現状終わっている事

- LLaMaやGPT X・AlpacaなどLLMに関連する技術調査
- LLM構築ツールの調査(Transformers)
- LLM構築実験用データの用意Wikipedia(JP)
- LLM構築リソースの確保



研究環境

- Transformers
- Google Colaboratory
- Python3

研究計画

1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
---	---	---	---	---	---	---	---	---	----	----	----	---	---	---

言語の学習



ラベル付け



プログラム作成



改善案検討



論文作成

