

Chain of Dictionary(For Machine Translation)

Instructor

Bhaskarjit Sarmah

Vice President, Blackrock



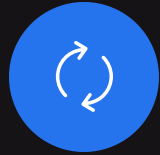
Definition

- **Chain-of-Dictionary** is a machine translation method that enhances LLM's performance for multilingual translation, especially for less-common languages.
-

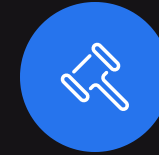
How Chain of Dictionary Works



Chain of Dictionary Prompting (COD) involves using a series of multilingual dictionaries to help translate languages by linking words across different languages in a chain format, which is particularly useful for languages that don't have many resources



This method does not rely on showing the translation program many examples of translations (few-shot demonstrations), which is often not practical for languages that are not widely used or documented[



Instead, COD enhances the translation process by providing a structured path of word meanings from the source language through one or more auxiliary languages before reaching the target language, thereby improving the translation accuracy for low-resource languages

Chain of Density: Example

Shared Translation Prompt

Translate the following text from English into Tamil:

"We have now 4-month-old mice that are non-diabetic that used to be diabetic," he added.

Standard Prompting

###No extra input ###

The standard prompting method uses the shared translation prompt described above only for prompting LLMs. There is no other text included as the prompt.

###No extra input##

Translation Output

"Translation from ChatGPT with lower quality: 'இப்போது நாம் நான்கு மாத வயது தூளான காலங்கள் உள்ளன அதோடு கிடக்கும் மூன்று நிற்பதில்லை தூளங்கள் காலமில்லாமல்', என்று அவர் சொன்னார்."



Chain-of-Dictionary Prompting

"have" means "வேண்டும்" means "haben" means "avoir". "4-month-old" means "4 மாத வயதில்" means "4 Monate alt" means "4 mois". "mice" means "எலிகள்" means "Mäuse" means "souris". "non-diabetic" means "நீரிழிவு நோய் இல்லாத" means "nicht-diabetisch" means "non diabétique". "used" means "பயன்படுத்தப்பட்ட" means "Gebrauch" means "utilisés". "diabetic" means "நீரிழிவு நோய்" means "Diabetiker" means "diabétique". "added." means "சேர்க்கப்பட்டது." means "- und hinzugefügt" means "ajoutée."

Translation Output

"Translation from ChatGPT with higher quality: 'நாங்கள் இப்போது சர்க்கரை நோயில்லா 4 மாத வயது எலிகளை கொண்டிருக்கிறோம், முன்னர் சர்க்கரை நோய் இருந்தவை, அவர் சேர்க்கின்றார்.'"



Chain of Dictionary: Practical Use-case



CoD enhances translation for under-supported languages, broadening accessibility and communication.

Reduces the need for large sets of example translations, which can be hard to find for less common languages.

Ideal for digital platforms seeking to offer multilingual content, enhancing user experience.

Thank You
