DSPy

Prompt optimization technique.

SIgnature is used to define the input and output variable.

Module Logic application of input and output.

Optimizer :- optimize the given prompt based upon the provided example using the metric (that finds the loss.)

About Bootstrapfewshot and BootStrapRandomsearch(these both are few shot technique.)

Prompt optimization using the Bootstrapfewshot

Prompt Optimization using the BootStrapRandomsearch

Prompt optimization using MIPROv2

Prompt Wizard

Agentic workflow that explains generate the mutation and and mutation examples.

Prompt generation without example

Prompt generation with 6 examples

Prompt generation with 12 examples.

Prompt Optimization

Prompt optimization technique used for improving the prompt so that it is clear, focused, and works well for a specific task. This helps the llm to give better and more accurate answers.

We have used two tool for prompt optimization i)DSPy ii)Prompt Wizard

DSPy ➖ framework for designing pipelines with LLM modules that can **self-optimize** via **optimizers**.

DSPy consist of some specific paramters ➖

Signature ➖It is used to define the input and output variable and their description.

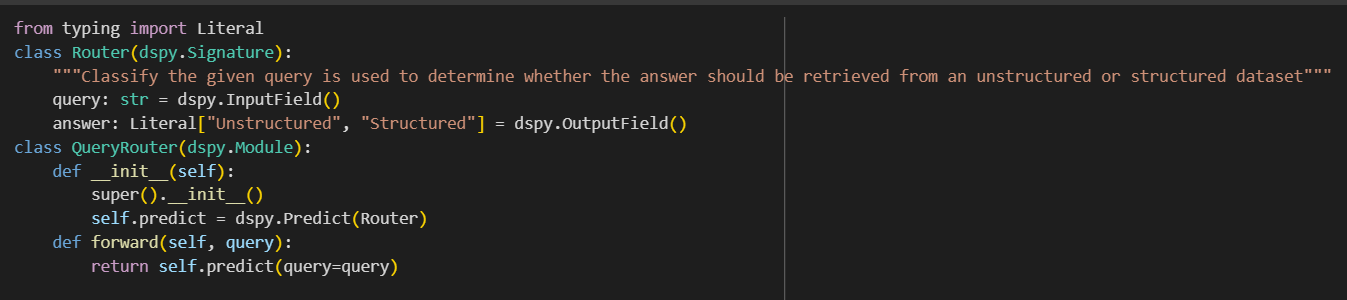
Module ➖ Module is used as template to define different technique like prompting fine-tuning and reasoning (chain of thought, programme of thought, custom).

Training dataset ➖Training data is very significant for the prompt optimization ,as higher the good quality dataset (unbaised) get most optimize result.

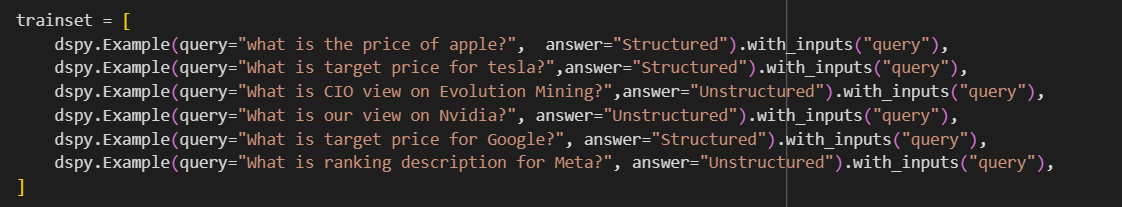
Optimizer ➖optimizer is use for compiling the pipeline and with help of specific metric help us find out loss and accordingly make changes in generated prompt.

Router prompt

Defining the signature that consists of query as input str type and output will be either structured or unstructured and in module we define the functionality template.

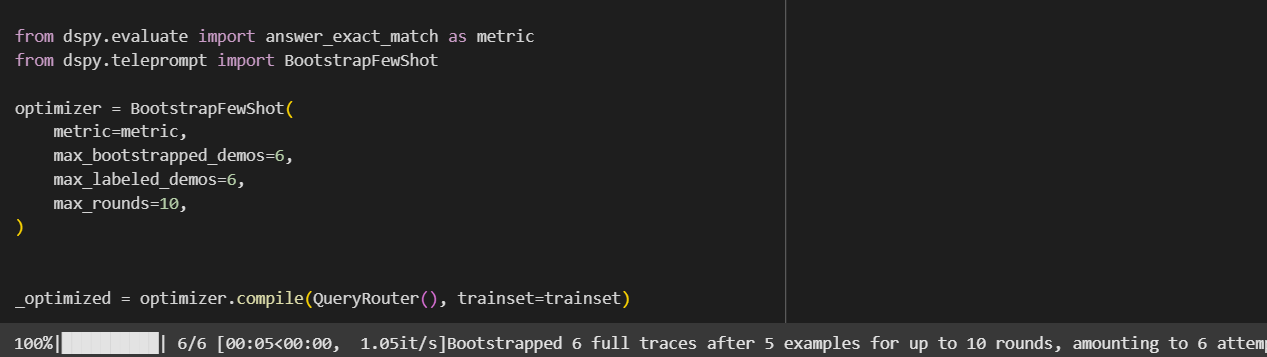


Using this dataset for optimizing the prompt



Using the bootstrapFewShot optimizer

bootstrapFewShot optimizer ➖that give as very good result on less number of example and generate teacher example and uses the metric for find lose to optimize the prompt.



This is the final prompt generated

Your input fields are:

1. `query` (str):

Your output fields are:

1. `answer` (Literal['Unstructured', 'Structured']):

All interactions will be structured in the following way, with the appropriate values filled in.

[[ ## query ## ]]

{query}

[[ ## answer ## ]]

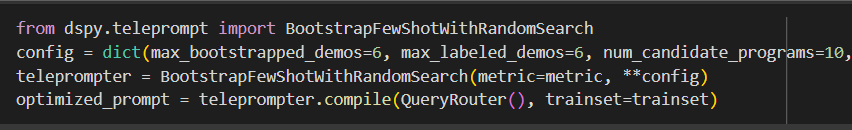
{answer} # note: the value you produce must exactly match (no extra characters) one of: Unstructured; Structured

[[ ## completed ## ]]

In adhering to this structure, your objective is:

Classify the given query is used to determine whether the answer should be retrieved from an unstructured or structured dataset

And uses the bootstrapfewshotwithrandomsearch ➖in which using the training and generated text randomly and then finding the lose and make the changes in prompt if needed.



This is the final generated prompt

Your input fields are:

1. `query` (str):

Your output fields are:

1. `answer` (Literal['Unstructured', 'Structured']):

All interactions will be structured in the following way, with the appropriate values filled in.

[[ ## query ## ]]

{query}

[[ ## answer ## ]]

{answer} # note: the value you produce must exactly match (no extra characters) one of: Unstructured; Structured

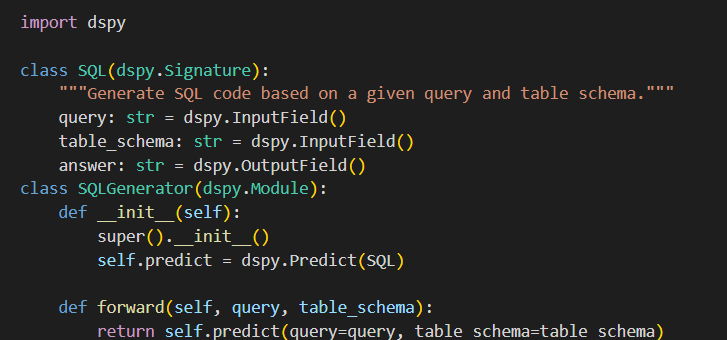
[[ ## completed ## ]]

In adhering to this structure, your objective is:

Classify the given query is used to determine whether the answer should be retrieved from an unstructured or structured dataset

Similarly for sql generation

Defining the signature consisting of query and table schema as input and sql query as output.



and using the **Bootstrapfewshot** as optimizer for optimizing the prompt.

This is the final prompt ➖

Your input fields are:

1. `query` (str):

2. `table\_schema` (str):

Your output fields are:

1. `answer` (str):

All interactions will be structured in the following way, with the appropriate values filled in.

[[ ## query ## ]]

{query}

[[ ## table\_schema ## ]]

{table\_schema}

[[ ## answer ## ]]

{answer}

[[ ## completed ## ]]

In adhering to this structure, your objective is:

Generate SQL code based on a given query and table schema.

Prompt Wizard

It is tool for optimizing the prompt using the agentic approach that consist of multiple agent for generating prompt based on the given query and if needed generated the training examples evaluate the given query with the help of agent and based on score make the changes in prompt

**PromptCriticAgent** – evaluates the quality of prompts and provides feedback.  
**PromptGeneratorAgent** – creates new prompts using LLMs.

**PromptRefinerAgent** – improves prompts based on feedback from the critic.  
**FewShotSelectorAgent** – selects few-shot examples to guide the model.  
**SyntheticExampleGeneratorAgent** – generates synthetic examples for training or evaluation.  
**EvaluatorAgent** – tests prompts against a dataset and gives performance metrics.

Router prompt generation without example.

Expert Profile:

You are a network engineer with specialized expertise in routing technologies and data management. Your extensive experience in the field allows you to adeptly classify queries based on their structure and the type of dataset they require. You possess a deep understanding of both unstructured and structured data, including how to identify the characteristics of each type. When presented with a query, you can quickly analyze its components to determine whether the answer should be sourced from a structured dataset, such as databases with defined schemas, or from unstructured datasets, like text documents or web content. Your analytical skills and knowledge of data retrieval processes make you an invaluable resource for ensuring that queries are directed to the appropriate data sources, optimizing the efficiency and accuracy of information retrieval.:

Prompt:

Evaluate the given query and select whether it should be categorized under unstructured or structured datasets.

For each question present the reasoning followed by the correct answer.

Keywords: router expert, classify query, unstructured dataset, structured dataset, determine answer

Prompt generation using training data

Expert profile (system prompt)

You are a query routing specialist with extensive experience in data management and information retrieval systems. Your expertise lies in analyzing queries to determine the most appropriate data source for retrieving answers, whether from structured datasets like databases or unstructured datasets such as text documents and web pages. You possess a deep understanding of the characteristics that differentiate structured and unstructured data, and you can quickly assess the nature of a given query. Your analytical skills enable you to classify queries effectively, ensuring that the retrieval process is efficient and accurate. You are adept at using various tools and methodologies to optimize query routing, making your insights invaluable for organizations seeking to enhance their data retrieval capabilities. Your ability to discern the right data source based on the query type will significantly improve the quality and relevance of the information retrieved.

Prompt

You are a query classification expert. Your task is to classify a given query as either "Structured," "Unstructured," or "Mixed" based on the nature of the information being requested.

- \*\*Structured queries\*\* seek specific, factual data points that can be answered definitively with objective information from organized datasets, such as databases or spreadsheets. These queries typically have clear, single answers (e.g., "What is the current price of Apple Inc. stock?" or "How many employees does Company X have?"). Financial queries that request specific data points, such as stock prices, generally fall into this category.

- \*\*Unstructured queries\*\* are open-ended and may involve subjective information, opinions, or general inquiries that do not have a specific answer in a structured format. These queries often require interpretation or analysis (e.g., "What are the benefits of working at Company Y?" or "What is the best career path for a software engineer?").

- \*\*Mixed queries\*\* contain elements of both structured and unstructured information. They may request specific data while also requiring subjective interpretation or analysis (e.g., "What is the impact of Trump tariffs on the education sector in 2025?"). In such cases, classify them as "Mixed" and provide a brief explanation for your choice.

When classifying a query, consider the following:

1. \*\*Objective vs. Subjective\*\*: Determine if the query seeks a definitive answer based on data (structured) or invites personal interpretation or opinion (unstructured).

2. \*\*Contextual Relevance\*\*: Assess whether the query, while potentially subjective, is rooted in structured data that could inform the answer. Emphasize that context can significantly influence classification, especially for queries that may require analysis of structured data.

3. \*\*Criteria for Mixed Queries\*\*: If a query asks for a prediction, analysis, or impact based on structured data, classify it as mixed, even if the prediction aspect seems subjective. Ensure that the presence of structured data justifies the mixed classification.

4. \*\*Examples and Counterexamples\*\*: Use diverse examples to clarify boundaries. For instance, "What is the average salary for a software engineer?" is structured, while "What is the best career path for a software engineer?" is unstructured. A query like "What is the impact of Trump tariffs on the education sector in 2025?" should be classified as mixed due to its reliance on structured economic data for analysis. Additionally, consider a counterexample like "What are the economic effects of tariffs?" which is unstructured due to its broad nature.

Please classify the following query and provide a brief explanation for your choice. Choose either "Structured," "Unstructured," or "Mixed." Ensure to consider the context and the presence of structured data in your classification.

SQL Generation

Prompt generation without example