CausalPrompt: Enhancing LLMs with Weakly Supervised Causal Reasoning for Robust Performance in Non-Language Tasks

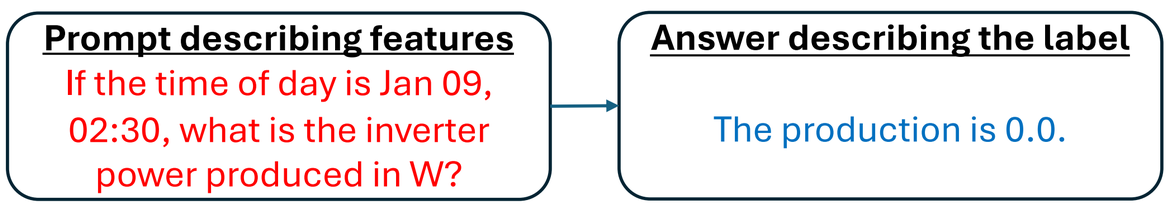
Tung-Wei Lin*, Vanshaj Khattar*, Yuxuan Huang*, Junho Hong, Ruoxi Jia, Chen-Ching Liu, Alberto Sangiovanni-Vincentelli, Ming Jin
UC Berkeley, Virginia Tech, University College London, University of Michigan

Abstract

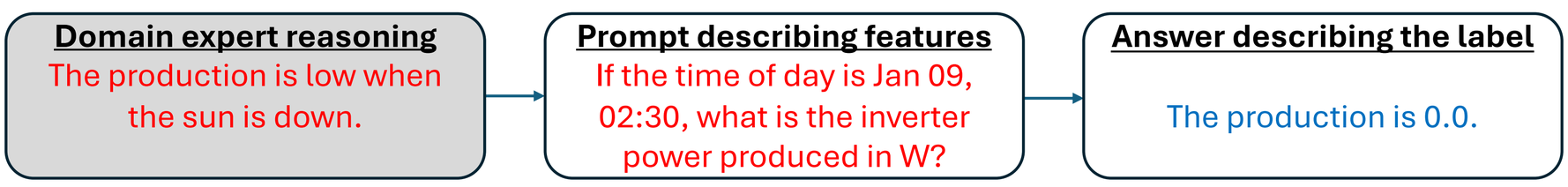
In confronting the pressing issue of climate change, we introduce "CausalPrompt", a prompting strategy that adapts large language models (LLMs) for classification and regression tasks through the application of weakly supervised causal reasoning. We delve into the complexities of data shifts within energy systems, often resulting from the dynamic evolution of sensor networks, leading to discrepancies between training and test data distributions or feature inconsistencies. By embedding domain-specific reasoning in the finetuning process, CausalPrompt significantly bolsters the adaptability and resilience of energy systems to these shifts. We show that CausalPrompt significantly enhances predictions in scenarios characterized by feature shifts, including electricity demand, solar power generation, and cybersecurity within energy infrastructures.

Methodology

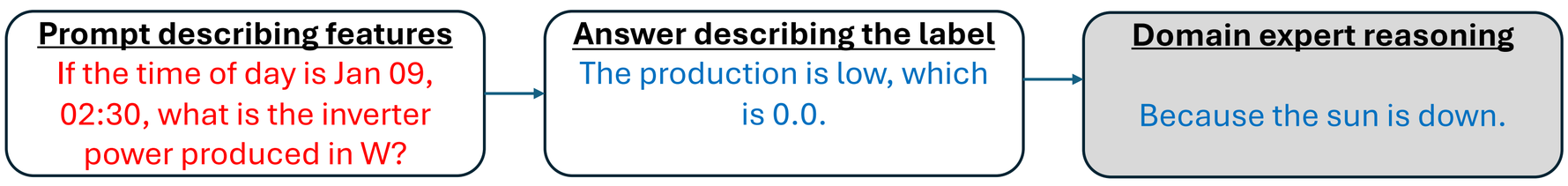
a) LIFT training prompts



b) CausalPrompt1 (CP1)



c) CausalPrompt2 (CP2)



CausalPrompt1 (CP1): CP1 incorporates domain expert reasoning at the beginning of the training prompt. The LLM is fine-tuned to utilize both the query and the domain expert reasoning to generate the corresponding label.

CausalPrompt2 (CP2): In CP2, domain expert reasoning is appended after the label. The LLM is fine-tuned to first produce the label followed by the domain expert reasoning.

Results

Table 1: **RMSEs with and without (w/o) feature shifts on datasets 1 and 2.** (Lower is better.) The best results are in bold, and the second best are underlined.

Dataset	CP1 (ours)	CP2 (ours)	LIFT	GPR	LR	MLP	KNN	DTR
Electric demand dataset (w/o shift)	0.25	0.40	0.41	**0.15**	0.50	<u>0.23</u>	0.17	**0.15**
Electric demand dataset (with shift)	**0.41**	<u>0.60</u>	2.63	0.96	0.92	3.16	0.71	0.72
Solar power prediction (w/o shift)	**83.90**	178.81	182.29	<u>121.48</u>	215.92	147.20	109.07	122.07
Solar power prediction (with shift)	**121.01**	304.57	359.36	280.42	<u>216.52</u>	333.60	287.42	299.36

Table 2: **Accuracy (%) on the cybersecurity dataset with and without (w/o) feature shifts.** (Higher is better.)

Dataset	CP1 (ours)	CP2 (ours)	LIFT	LSTM	GRU	RNN
Substation cyber-attack (w/o shift)	**64.28**	35.72	<u>21.42</u>	35.71	35.71	35.71
Substation cyber-attack (with shift)	**57.14**	<u>28.57</u>	7.15	14.29	<u>28.57</u>	21.43

Table 3: **Percentage (%) drop in the performance after feature shift in the test set.** (* - not applicable). (Lower is better.)

Dataset	CP1	CP2	LIFT	GPR	LR	MLP	KNN	DTR	LSTM	GRU	RNN
Electric demand dataset	<u>64.00</u>	**50.00**	541.46	540.00	80.39	1216.67	343.75	380.00	*	*	*
Solar power prediction	<u>44.23</u>	70.33	97.13	130.83	**0.27**	126.63	163.51	145.23	*	*	*
Substation cyber-attack	**11.10**	<u>20.01</u>	66.61	*	*	*	*	*	59.98	<u>20.01</u>	40.00