Order-Based Pre-training Strategies for Procedural Text Understanding

NAACL 2024

Abhilash Nandy, Yash Kulkarni, Pawan Goyal, Niloy Ganguly

CNeRG Lab, Indian Institute of Technology Kharagpur, India

Background

- In the realm of procedural text, understanding and processing tasks such as entity tracking are pivotal.
- Conventional Pre-training such as MLM does not capture step-order information
- Small size of the procedural reasoning benchmark datasets

ENTITY TRACKING

		Participants:				
Paragraph (seq. of steps):		water	light	CO2	mixture	sugar
	state0	soil	sun	?	-	-
Roots absorb water from soil						
	state1	roots	sun	?	-	-
The water flows to the leaf.						
	state2	leaf	sun	?	-	-
Light from the sun and CO2 enter the leaf.						
	state3	leaf	leaf	leaf	-	-
The light, water, and CO2 combine into a mixture.						
	state4	-	-	-	leaf	-
Mixture forms sugar.						
	state5	-	-	-	-	leaf

Time

Figure: Example Of Entity Tracking Task

Problem Statement

Coming up with Pre-training tasks that use Order of Steps in a Procedure as a supervision signal to solve procedural reasoning tasks.

- Using different permutation-based and step-similarity-based (based on proximity of steps) objectives (as well as their combinations) to use Order as a Supervision Signal
- Using the fine-tuned model to solve Entity Tracking task that require understanding the sequential nature/order of the steps

Dataset Collection

We collected, and preprocessed a corpus of **more than 2.5 Million** recipes after deduplication was carried out.

- Recipe1M+ dataset (Marin et al., 2018)
- RecipeNLG dataset (Bie n et al., 2020)
- Generating personalized recipes from historical user preferences (Majumder et al., 2019)
- Storyboarding of recipes: Grounded contextual generation. (Chandu et al., 2019)

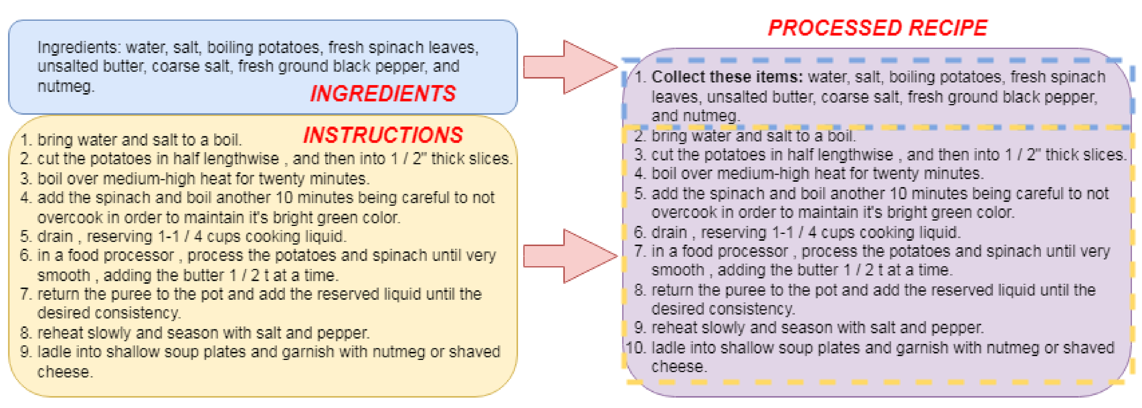

Figure: Modifying a recipe by adding ingredients as a step at the top of the original recipe

Qualitative Analysis

	Ground Truth	Permutation Classification	RoBERTa-BASE
Procedure	flower	flower	flower
(Before the process starts)	-	-	-
1. A seed is planted.	-	-	-
2. It becomes a seedling.	-	-	-
3. It grows into a tree.	-	-	-
4. The tree grows flowers.	tree	tree	tree
5. The flowers become fruit.	-	-	tree
6. The fruits contain seeds for new trees.	-	-	tree

Figure: Analysis of the ground truth and predictions of Permutation Classification vs. a well-performing baseline on a ProPara sample. **Permutation Classification predicts the entity state better when an entity ceases to exist**

Methodology

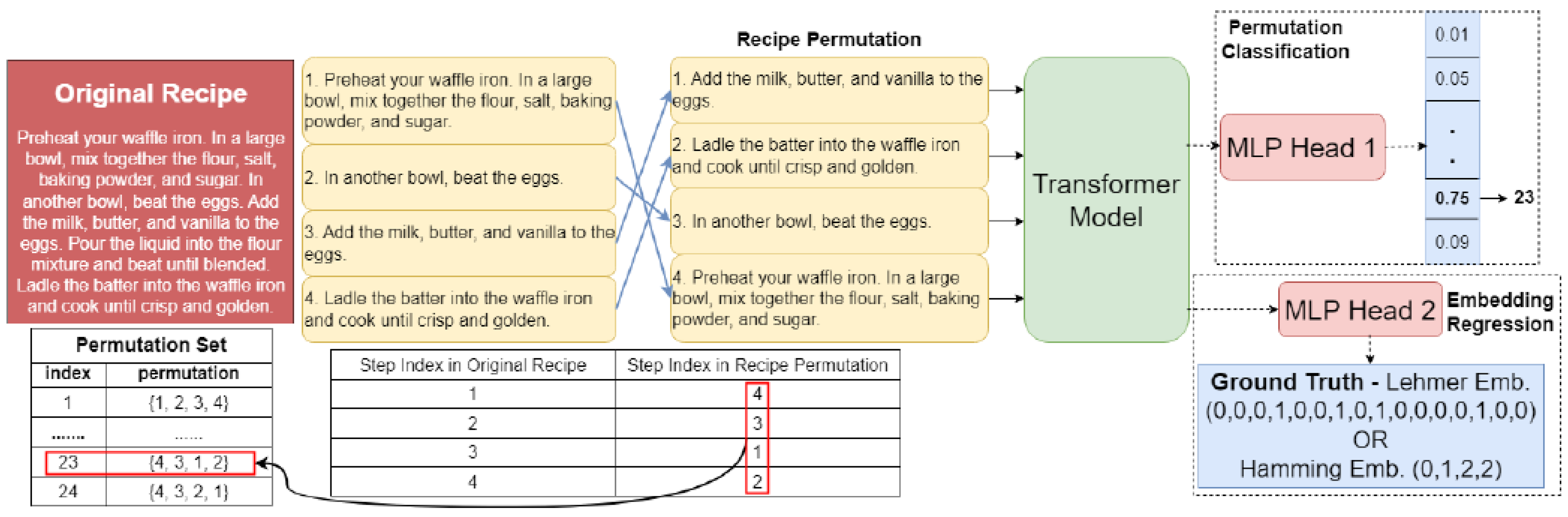

Figure: ***Permutation Classification*** and ***Embedding Regression*** for a 4-step recipe. The Permutation Classification Task is to predict the index of the chosen permutation which in this case is 23, and Embedding Regression Task is to predict the corresponding Lehmer/Hamming Embedding.

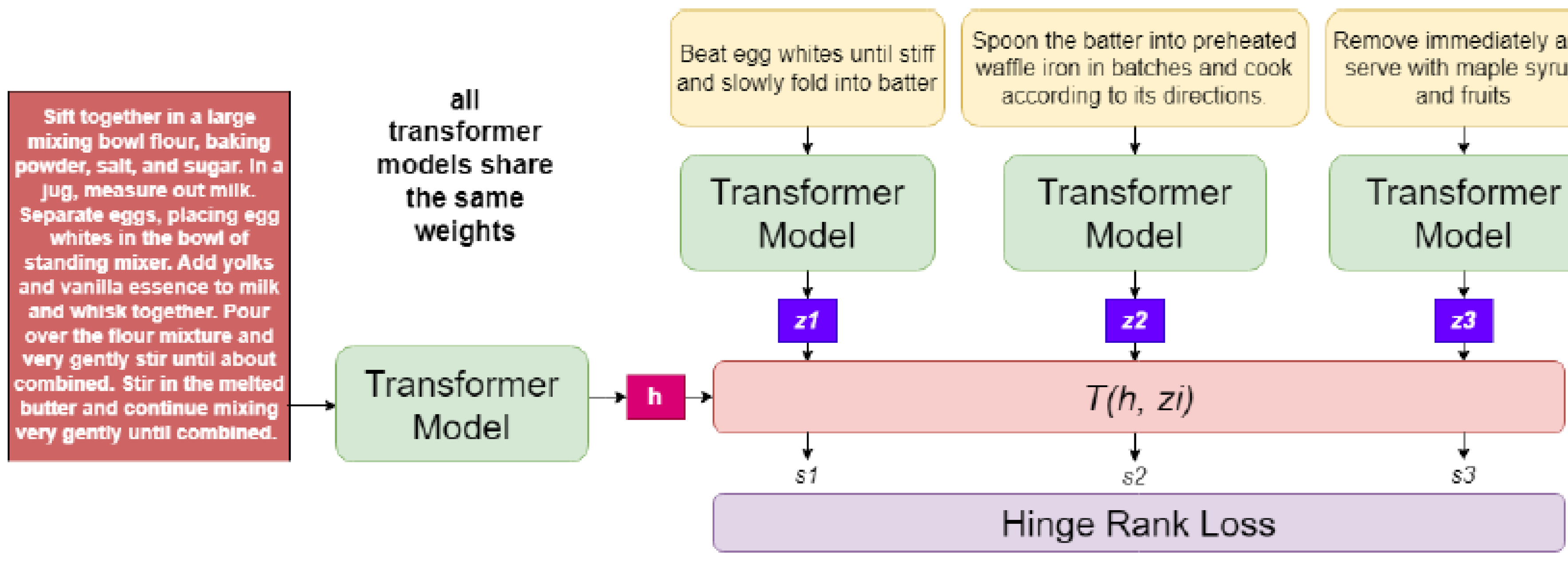

Figure: ***Skip-Clip*** model with a 6-step context and 3 target steps. The task is to rank the target steps based on scores obtained from a scoring function and their order in the recipe using hinge rank loss.

Experiments and Results

	Cat1 Acc.	Cat2 Acc.	Cat3 Acc.	Avg. Cat Acc.
Falcon-7B-Instruct (1-shot)	50.42	5.42	0.38	18.74
Falcon-7B-Instruct (3-shot)	48.44	3.15	1.94	17.84
Llama 2-7B-Chat (1-shot)	47.88	9.74	6.44	21.35
Llama 2-7B-Chat (3-shot)	51.27	13.98	11.97	25.74
GPT-3.5 (1-shot)	53.25	24.66	11.37	29.76
GPT-3.5 (3-shot)	62.43	34.66	15.81	37.63
Permutation Classfn.	**73.72**	**43.16**	32.72	**49.87**
$Emb_{Hamming}$	68.5	30.43	**36.16**	45.03
Emb_{Lehmer}	<u>73.3</u>	<u>38.82</u>	<u>35.05</u>	<u>49.06</u>
Skip-Clip	66.94	34.46	32.73	44.71

Table 8: Results on the ProPara Dataset - LLMs vs. proposed permutation-based methods

Model	Location Acc.	Status Acc.	Cat1 Acc.	Cat2 Acc.	Cat3 Acc.	Avg. Cat. Acc.
Rule-based	-	-	57.14	20.33	2.4	26.62
Feature-based	-	-	58.64	20.82	9.66	29.71
ProLocal	-	-	62.7	30.5	10.4	34.53
ProGlobal	-	-	63	36.4	35.9	45.1
EntNet	-	-	51.6	18.8	7.8	26.07
QRN	-	-	52.4	15.5	10.9	26.27
RoBERTa-BASE	56.27	65.71	71.33	31.78	34.05	45.72
Permutation Classfn.	57.71	**70.57**	**73.72**	**43.16**	32.72	**49.87**
$Emb_{Hamming}$	53.05	61.36	68.5	30.43	**36.16**	45.03
Emb_{Lehmer}	**60.61**	<u>68.11</u>	<u>73.3</u>	<u>38.82</u>	<u>35.05</u>	<u>49.06</u>
Skip-Clip	<u>58.19</u>	63.4	66.94	34.46	32.73	44.71

Table 2: Results on ProPara Dataset. Numbers in bold and underlined are the highest and the second-highest scores respectively.

Model	Dev Acc	Test Acc
NPN-Model	-	51.3
KG-MRC	-	51.6
DYNAPRO	-	62.9
RoBERTa-BASE	<u>65.07</u>	64.28
Permutation Classfn.	**65.48**	64.75
$Emb_{Hamming}$	65.03	<u>64.92</u>
Emb_{Lehmer}	63.96	64.29
Skip-Clip	63.87	**65.33**

Table 1: Results on NPN-Cooking Dataset. Numbers in bold and underlined are the highest and the second-highest scores, respectively.

Results on Entity Tracking Tasks

Model	Test Acc
Permutation Classfn.	64.75
$Emb_{Hamming}$	64.92
Emb_{Lehmer}	64.29
Skip-Clip	**65.33**
$Emb_{Hamming}$ + *Permutation Classfn.*	61.88
Emb_{Lehmer} + *Permutation Classfn.*	62.66
Skip-Clip + *Permutation Classfn.*	0.01

Table 3: Results of sequential combination of different pre-training strategies on NPN-Cooking Dataset.

Analysis of Results

- ***Permutation Classification*** gives the **best dev set result** and ***Skip-Clip*** gives **best test accuracy** on the **NPN-Cooking Dataset**.
- Most proposed methods beat baselines on **Propara Dataset**, suggesting that **predicting the next step from a given context helps in Entity Tracking in Recipe Domain**.

Key Insights

- We introduce 3 pre-training tasks - ***Permutation Classification***, ***Embedding Regression*** and ***Skip-Clip*** to learn the sequential nature of procedures.
- ***Skip-Clip*** performs best on the **NPN-Cooking Task**, while ***Permutation Classification*** and ***Embedding Regression*** perform best on the open-domain **ProPara Task**.
- We believe such methods could be extended to procedures in E-Manuals, manufacturing guides, etc.

Paper Link

Code Base Link

For Further Information

Arxiv: `https://arxiv.org/abs/2404.04676`