

# F-NLP at VLSP 2021-MRC Shared Task: Joint Learning and Ensemble Method for Vietnamese Machine Reading Comprehension

The 8th International Workshop on Vietnamese Language and Speech Processing (VLSP 2021)

Phan Xuân Phúc<sup>1</sup>, Nguyễn Kỳ Tùng<sup>2</sup>, Nguyễn Hải Dương<sup>2</sup>, Dương Minh Trí<sup>3</sup>  
 {phanxuanphucnd, nguyengkyltungcntt04, nhd.a1.chel, minhtridn1999}@gmail.com

<sup>1</sup> FTECH CO., LTD,

<sup>2</sup> Hanoi University of Science and Technology, Ha Noi, Viet Nam

<sup>3</sup> Danang University of Science and Technology, Da Nang, Viet Nam

## INTRODUCTION

- **Motivation:** Machine Reading Comprehension (MRC) is one of the key problems and the next trend in Nature Language Processing, where the task is to read and comprehend a given text passage, and the answer questions based on it. MRC could be widely applied in many applications, such as search engines, intelligent agents, dialog systems, question answering systems and chatbots.
- **Task Definition:**
  - **Input:** Given a text passage and the question.
  - **Output:** A span of text in the given text passage.

## DATASET

- **Name:** UIT-ViQuAD2.0 (Vietnam Question Answer Dataset 2.0)
- **Size:** 35K+ question on 170+ articles, containing about 23K+ question-answer pairs with about other 12K unanswerable questions.
- **Domain:** Vietnamese Wikipedia.

	Train	Public Test	Private Test	All
Number of articles	138	19	19	176
Number of passages	4,101	557	515	5173
Number of total question	28,457	3,821	3,712	35,990
Number of unanswerable questions	9,217	1,168	1,116	11,501

Table 1: Overview of the UIT-ViQuAD 2.0 dataset.

**Context:** Phạm Văn Đồng (1 tháng 3 năm 1906 – 29 tháng 4 năm 2000) là Thủ tướng đầu tiên của nước Cộng hòa Xã hội chủ nghĩa Việt Nam từ năm 1976 (từ năm 1981 gọi là Chủ tịch Hội đồng Bộ trưởng) cho đến khi nghỉ hưu năm 1987. Trước đó ông từng giữ chức vụ Thủ tướng Chính phủ Việt Nam Dân chủ Cộng hòa từ năm 1955 đến năm 1976. Ông là vị Thủ tướng Việt Nam tại vị lâu nhất (1955–1987). Ông là học trò, cộng sự của Chủ tịch Hồ Chí Minh. Ông có tên gọi thân mật là Tô, đây từng là bí danh của ông. Ông còn có tên gọi là **Lâm Bá Kiệt** khi làm Phó chủ nhiệm cơ quan Biện sự xử tại Quế Lâm (Chủ nhiệm là Hồ Học Lãm).

**Question 1:** Tên gọi nào được Phạm Văn Đồng sử dụng khi làm Phó chủ nhiệm cơ quan Biện sự xử tại Quế Lâm?

**Answer:** Lâm Bá Kiệt // The answer is extracted from the context at start index 507.

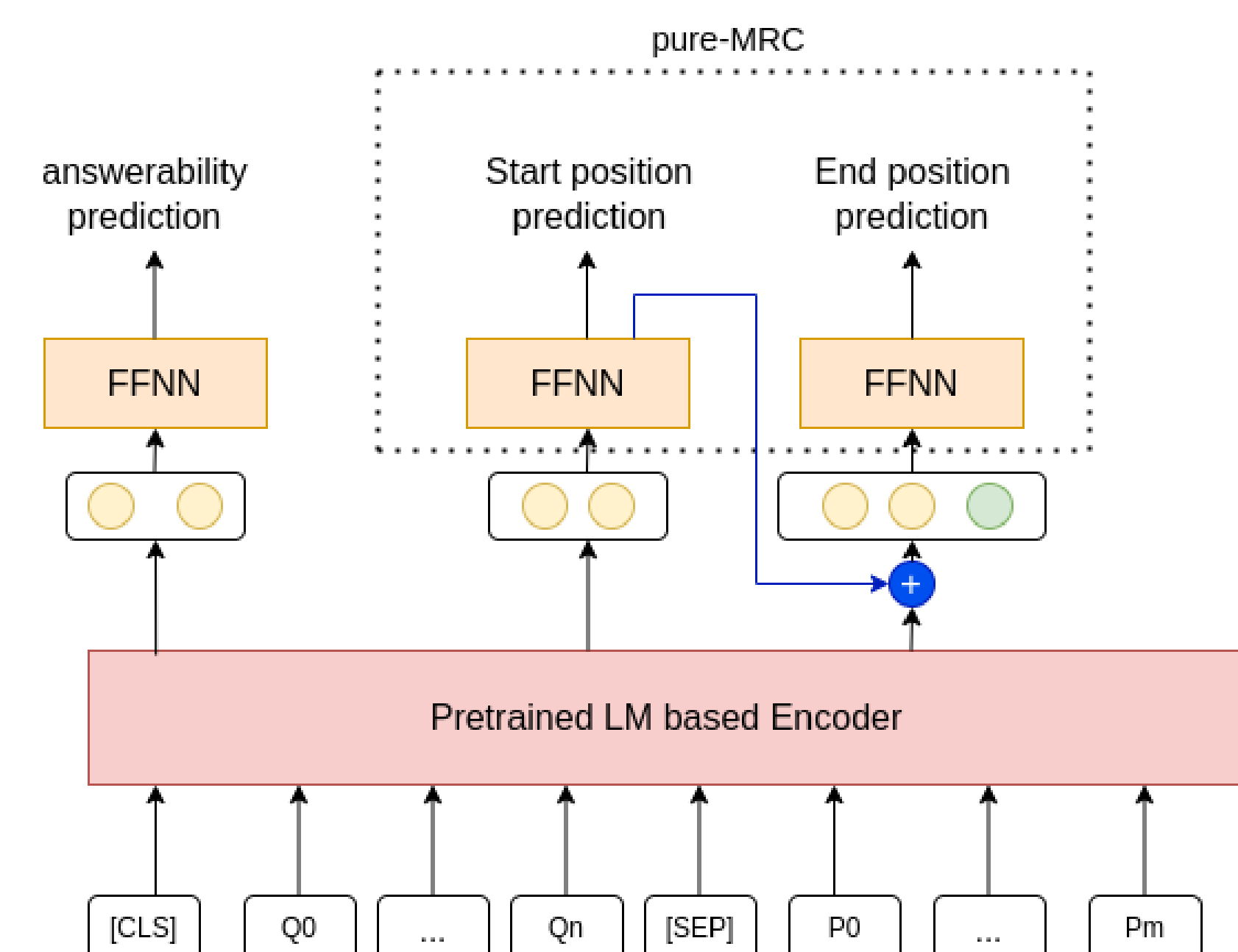
Table 2: An example of the UIT-ViQuAD 2.0 dataset.

## CONCLUSION AND FUTURE WORK

- **Conclusion:**
  - Proposed two types of joint models and ensemble method for Vietnamese Machine Reading Comprehension.
  - Top 3 of the VLSP 2021-MRC Shared Task competition.
- **Future Work:**
  - Experiment with some pre-trained language models with large models such as GPT-2, mT5, ...
  - Research some other verification strategies for ensemble method to improve model performance.

## EXPERIMENTS

- **Data Pre-processing**
  - Build  $M$  ( $M=5$ ) different datasets, each dataset comprises a train and dev set with a corresponding ratio of 0.9:0.1;
  - Randomly divided at the paragraph-level (article-level).
- **Our Models**



- Using the XLM-RoBERTa pre-trained LM as Encoder;
- Two types of joint models for answerability prediction and pure-MRC prediction (start-end position predictions) with/ without dependency mechanism;
- A dependency mechanism to learn the correlation between the start and end positions;
- Ensemble models a way that gives the top K answers for each model and a verification strategy to choose the best answer.
- **Evaluation Metrics**
  - Extract Match (abbreviated as EM)
  - F1-score
- **Experimental Results**

Method	Public Test		Private Test	
	F1	EM	F1	EM
Our model	<b>80.578</b>	<b>70.662</b>	<b>76.456</b>	<b>64.655</b>
Baseline	63.031	53.546	60.338	49.353

Table 3: The results on the public and private test sets of UIT-ViQuAD 2.0 dataset, evaluated on EM and F1-scores.

Method	Public Test	
	F1	EM
With Ensemble	<b>80.578</b>	<b>70.662</b>
Without Ensemble	76.657	65.768

Table 4: The results on the public test set of the UIT-ViQuAD 2.0 dataset, comparisons between with/ without ensemble method.