

VSLIM: A Vietnamese Explicit Slot-Intent Mapping for Joint Multi-Intent Detection and Slot Filling



Phong Chung¹, Kha Le-Minh², Xuan-Bach Le¹ and Tho Quan¹

¹ Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh City, Vietnam. {cdphong.sdh232, lexuanbach, qttho}@hcmut.edu.vn

² University of Information Technology (UIT), VNU-HCM, Ho Chi Minh City, Vietnam. 23520664@gm.uit.edu.vn

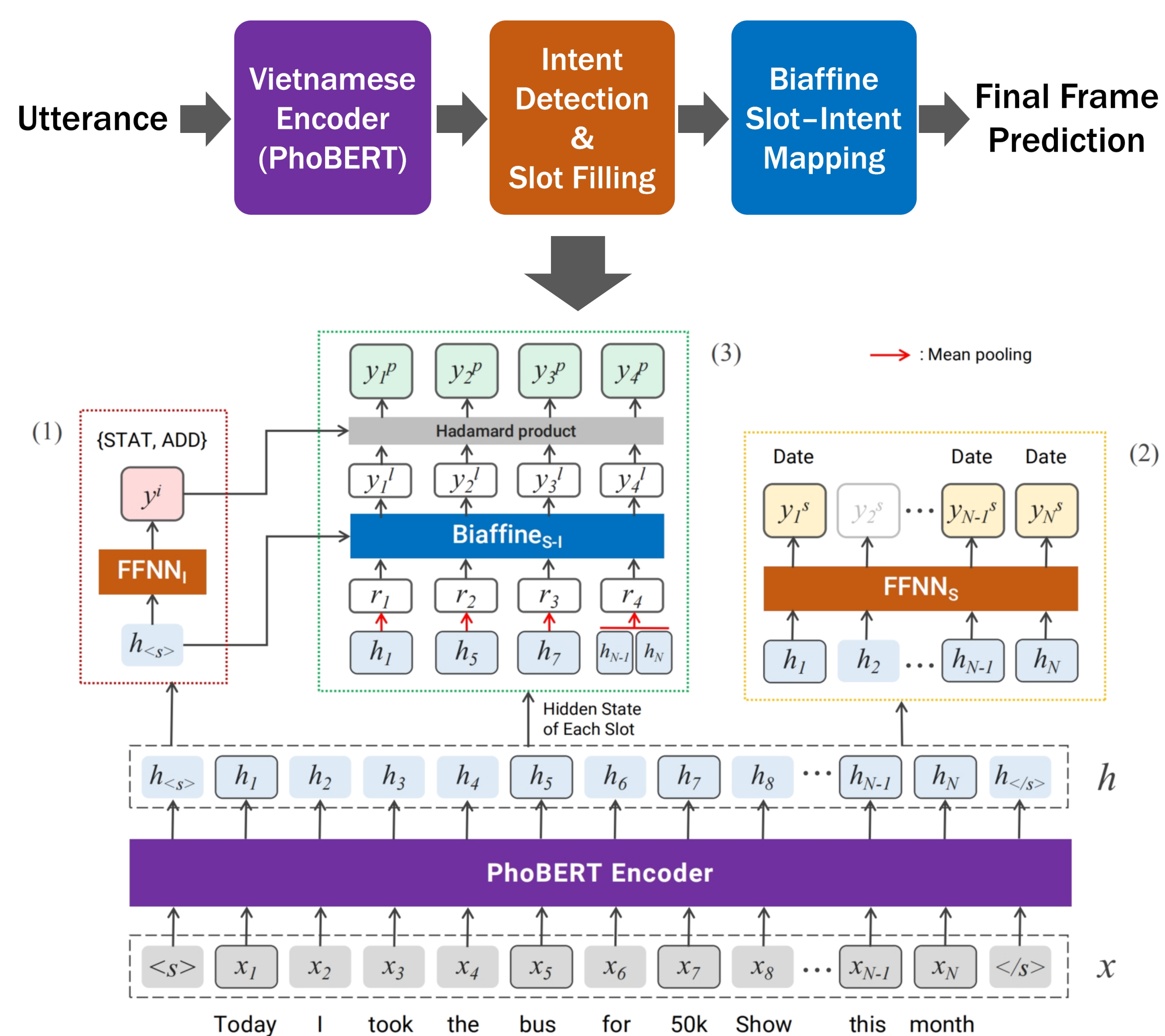


Introduction

- **Problems:**
 - Intent Detection and Slot Filling are crucial tasks
 - Slots must be linked to the correct intent
- **Gaps:**
 - Most models do not explicitly map slots to intents
 - Vietnamese resources are limited (LowRes)
- **Contributions:**
 - VSLIM: explicit slot-intent mapping model
 - VPED: Vietnamese multi-intent dataset

Method

- **VSLIM** extends SLIM with biaffine slot-intent mapping and evaluation on a Vietnamese multi-intent dataset.
- Overview of work flow:



- We also propose a new multi-intent dataset - Vietnamese Personal Expense Dataset (**VPED**):
 - 50%+ utterances contain multiple intents
 - Designed for personal expense management tasks

Table. Distribution of utterances in VPED in comparison with PhoATIS'

Dataset	% of utterances				
	1-intent	2-intent	3-intent	4-intent	5-intent
PhoATIS	98.88	1.09	0.03	-	-
VSLIM	46.08	25.54	21.96	5.53	0.90

Results

- **Baselines:** We compare with strong baselines from single and multi intent approaches, such as AGIF, JointIDSF, Co-Guiding...
- **Benchmarks:** We evaluate on 2 Vietnamese benchmarks: PhoATIS (single-intent) and our proposed VPED (multi-intent)

Table 3. Results on PhoATIS dataset. ^{mul} indicates multi-intent model. Results with ^r are taken from previous papers. **Bold** numbers are the best results in each column, while the second best is underlined. VSLIM with PhoBERTv1 shows competitive result, even without slot-intent labels.

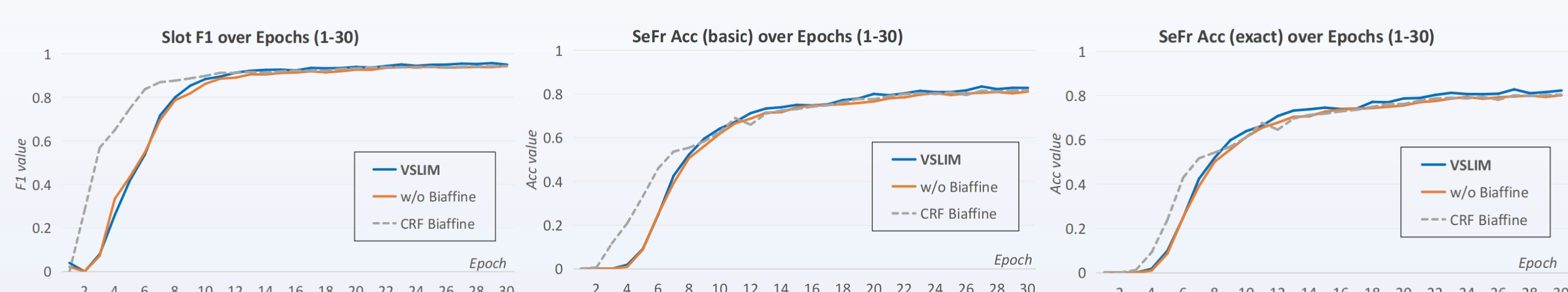
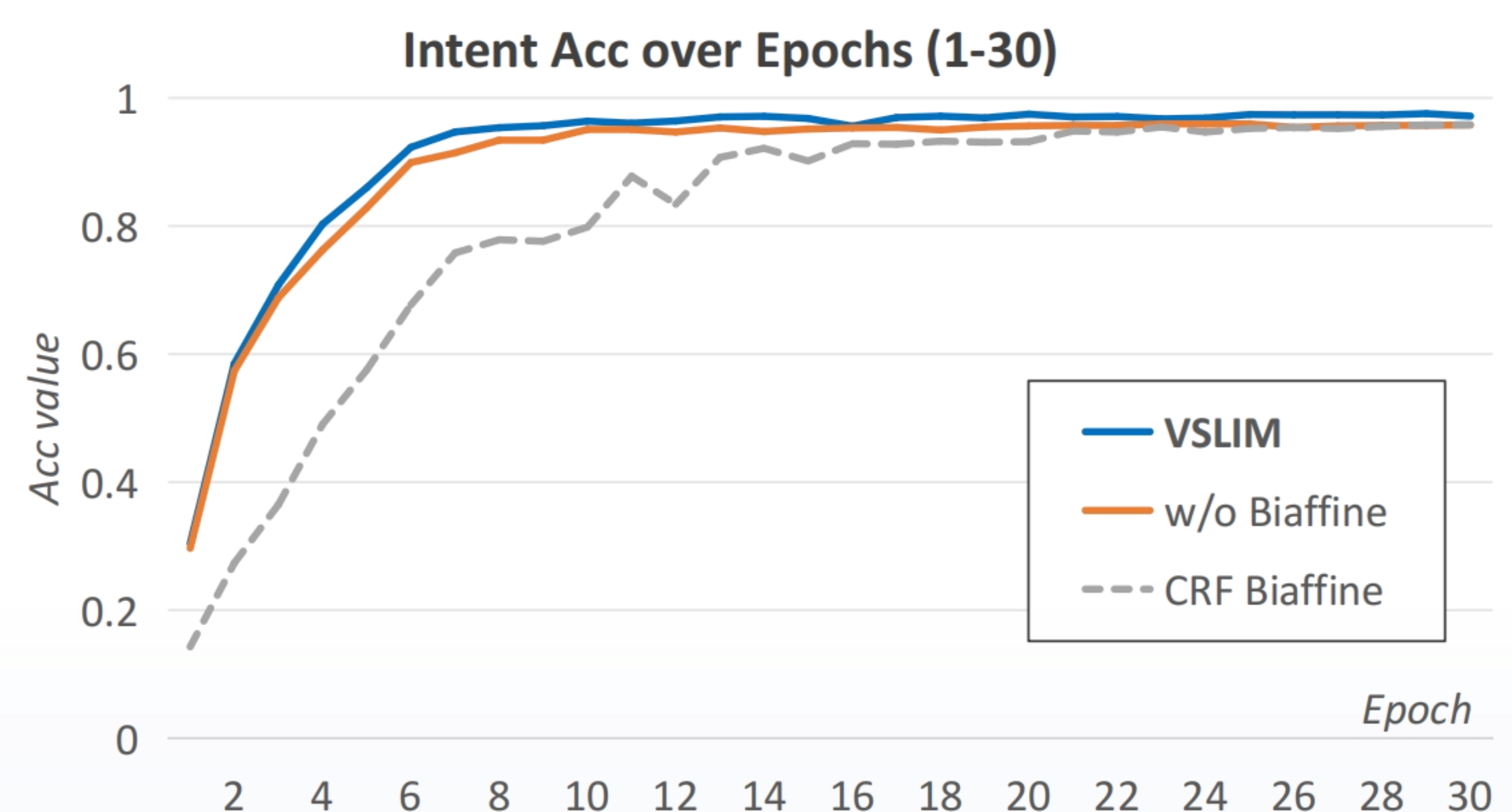
Model	PhoATIS		
	Intent Acc	Slot F1	SeFr Acc (basic)
AGIF ^{mul}	95.63	92.42	78.42
JointBERT+CRF + PhoBERTv1	97.40 ^r	<u>94.75^r</u>	85.55 ^r
JointIDSF + PhoBERTv1	97.62 ^r	94.98^r	86.25^r
MISCA ^{mul} + PhoBERTv1	97.06	94.60	84.71
Co-Guiding ^{mul}	95.14	93.22	80.44
Our VSLIM ^{mul} + PhoBERTv1	97.80	<u>94.75</u>	<u>86.09</u>

Table 4. Results on VPED dataset. ^{mul} indicates multi-intent model. **Bold** numbers are the best results in each column, with statistically significant improvement ($p < 0.05$ under t-test), while the second best is underlined.

Model	VPED			
	Intent Acc	Slot F1	SeFr Acc basic	SeFr Acc exact
AGIF ^{mul}	75.22	62.32	28.31	-
JointIDSF+PhoBERTv2	<u>89.74</u>	85.02	54.08	-
MISCA ^{mul} + PhoBERTv2	89.06	<u>85.39</u>	<u>54.44</u>	-
Co-Guiding ^{mul}	75.59	67.05	29.66	-
Our VSLIM ^{mul} +PhoBERTv2	91.30	87.23	58.95	56.72

- **Ablation (key points):**

- Biaffine classifier shows benefits, while the CRF adds early training difficulty in Intent Detection task.
- Short, variable, limited-contextual-cue utterances likely cause early learning challenges when using a CRF layer.



Discussion & Conclusion

- Biaffine Slot-Intent Mapping significantly improves overall model performance.
- Sequential CRF constraints provide limited benefit in specific multi-intent settings.
- VSLIM demonstrates fast convergence and stable training dynamics.
- The framework can be easily adapted to other languages by replacing the encoder with different variants.

Key References

- Cai et al., 2022 – SLIM: Explicit Slot-Intent Mapping with BERT
- Dao et al., 2021 – Intent Detection and Slot Filling for Vietnamese
- Nguyen & Nguyen, 2020 – PhoBERT: Pre-trained Language Models for Vietnamese

Acknowledgement We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.