JAIST-SIIT-NECTEC Collaboration in Audio Processing

Masashi Unoki (JAIST)
Chai Wutiwiwatchai (NECTEC)
Stanislav Makhanov (SIIT)
Pakinee Aimmanee (SIIT)

Audio watermarking
Jessada Karnjana
Speech analysis using empirical mode decomposition
Surasak Boonkla

Combining Different Lexical Units in RNN-based Language Modeling for Open-domain Thai Speech Recognition

Chai Wutiwiwatchai
National Electronics and Computer Technology Center (NECTEC)
Thailand

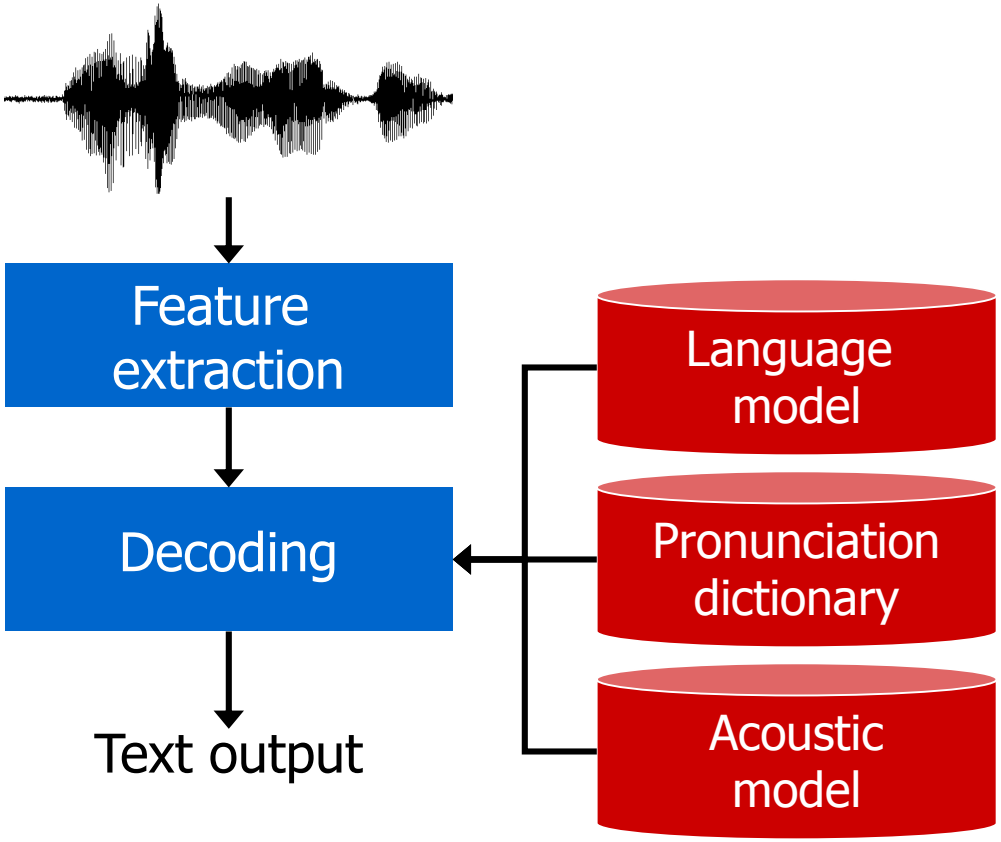
Outline

- Background
- Thai language and lexical units
- Combining different lexical units in
  - WFST-based language modeling
  - RNN-based language modeling
- Evaluation and discussion
- Conclusion and future work

Outline

- Background
- Thai language and lexical units
- Combining different lexical units in
  - WFST-based language modeling
  - RNN-based language modeling
- Evaluation and discussion
- Conclusion and future work

Open-domain LVCSR



- Challenge in reducing an out-of-vocabulary (OOV) rate with optimal system resources

Sub-word Based Language Modeling

世界反兴奋剂组织开展的调查显示，
俄罗斯田径运动员存在大规模服用兴奋剂问题，
或将被禁止参加2016年奥运会。

Wertpapierhandelgesetz

Using sub-word units

- Reducing the OOV rate
- Increasing acoustic confusion

- Y. He et al., "Subword-based modeling for handling oov words in keyword spotting," ICASSP 2014, pp. 7864–7868.
- X. Liu et al., "Syllable language models for mandarin speech recognition: Exploiting character language models," The Journal of the Acoustical Society of America, vol. 133, no. 1, pp. 519–528, 2013.

Hybrid Word/Sub-word Language Modeling

- Compromising between the use of word and sub-word language modeling

世界反兴奋剂组织开展的调查显示，
俄罗斯田径运动员存在大规模服用兴奋剂问题，
或将被禁止参加2016年奥运会。

世界反兴奋剂组织开展的调查显示，
俄罗斯田径运动员存在大规模服用兴奋剂问题，
或将被禁止参加2016年奥运会。

- M. A. B. Shaik, A. E.-D. Mousa, R. Schlüter, and H. Ney, "Hybrid language models using mixed types of sub-lexical units for open vocabulary German LVCSR" in INTERSPEECH, 2011, pp. 1441–1444.
- M. Kang, T. Ng, and L. Nguyen, "Mandarin word-character hybrid-input neural network language model." in INTERSPEECH 2011, pp. 625–628.

Outline

- Background
- Thai language and lexical units
- Combining different lexical units in
  - WFST-based language modeling
  - RNN-based language modeling
- Evaluation and discussion
- Conclusion and future work

Thai Scripts

เจ้าหน้าที่คณะกรรมการการเลือกตั้งชูบัตรลงคะแนนเสียงระหว่างการนับคะแนนที่คูหาแห่งหนึ่งในนครย่างกุ้ง|ประธานคณะกรรมการการเลือกตั้งแห่งสหภาพพม่าเผยว่าจะมีการจัดการเลือกตั้งซ่อมในช่วงปลายปีนี้ เนื่องจากยังมีที่นั่งในสภาว่างอยู่อีก 28 ที่นั่ง|

- Spelling with a finite no. of characters
- No explicit sentence nor word boundary
- A word segmentation tool is required for further language processing

Thai Sub-word Unit: Pseudo-Morpheme (PM)

- A smallest textual unit that can make all its contextual units and itself correctly pronounceable

Word	PM Segmentation
Exclamation	Ex | cla | ma | tion
Presentation	Pre | senta | tion

- As the clear definition and a finite no. of possible patterns, automatic PM segmentation is highly accurate

- W. Aroonmanakun, "Collocation and thai word segmentation," in SNLP and Oriental COCOSDA 2002, pp. 68–75.

Outline

- Background
- Thai language and lexical units
- Combining different lexical units in
  - WFST-based language modeling
  - RNN-based language modeling
- Evaluation and discussion
- Conclusion and future work

WFST-based Hybrid Language Modeling

- Maintain the most frequently used word units
- Segment the rest into PM units
- Train a language model using text with the mixed-type units

เจ้าหน้าที่คณะกรรมการการเลือกตั้งชูบัตรลงคะแนนเสียงระหว่างการนับคะแนนที่คูหาแห่งหนึ่งในนครย่างกุ้ง ประธานคณะกรรมการการเลือกตั้งแห่งสหภาพพม่าเผยว่าจะมีการจัดการเลือกตั้งซ่อมในช่วงปลายปีนี้ เนื่องจากยังมีที่นั่งในสภาว่างอยู่อีก 28 ที่นั่ง

- K. Thangthai, A. Chotimongkol, and C. Wutiwiwatchai, "A hybrid language model for open-vocabulary Thai LVCSR" in INTERSPEECH 2013, pp. 2207–2211.

WFST-based Hybrid Language Modeling

- **Test sets**: 5,145 utterances from various domains e.g. broadcast news, law, social media, blog, speech translation app, etc.

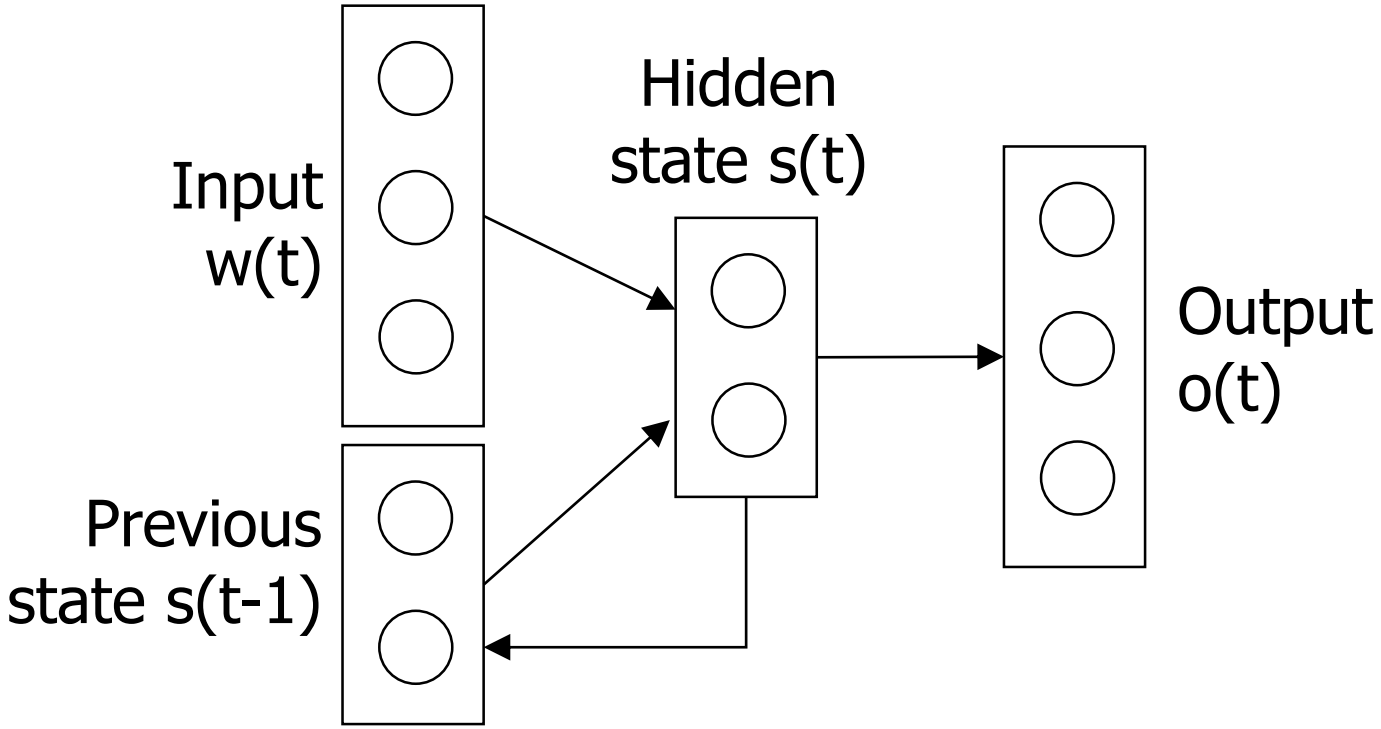
#Words	35K	20K	10K
#PMs	0K	15K	25K
Perplexity	164.3	151.2	140.8
%OOV	2.66	→ 0.57	0.40
%PER	39.64	→ **39.32**	39.50

Outline

- Background
- Thai language and lexical units
- Combining different lexical units in
  - WFST-based language modeling
  - RNN-based language modeling
- Evaluation and discussion
- Conclusion and future work

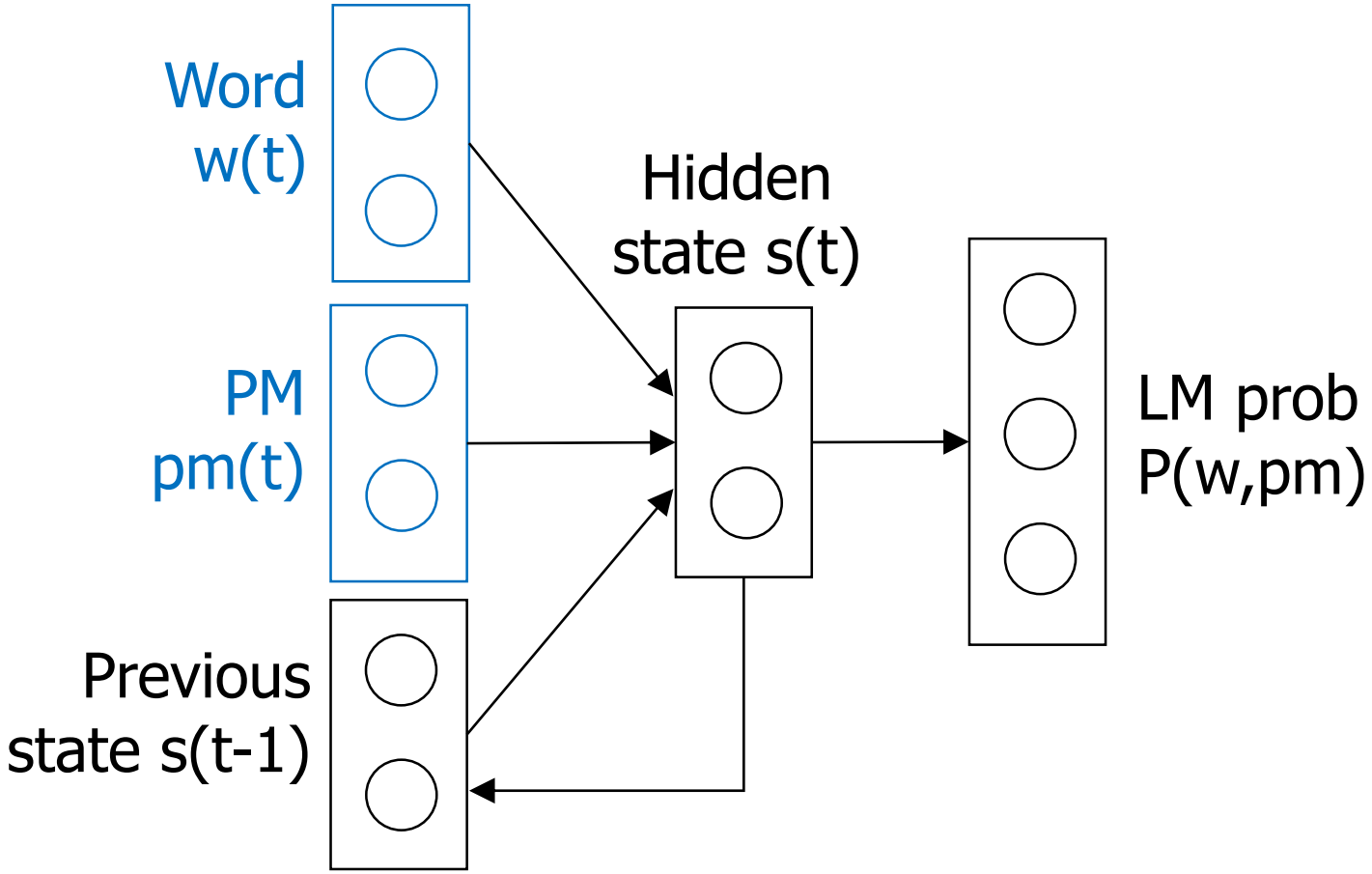
RNN-based Language Modeling (RNNLM)

- Increasingly popular for language modeling
  - Efficiently capture long history
  - Naturally accept multiple input types



RNN-based Hybrid Language Modeling

- RNN can naturally accept different input units, i.e. word and sub-word PM units

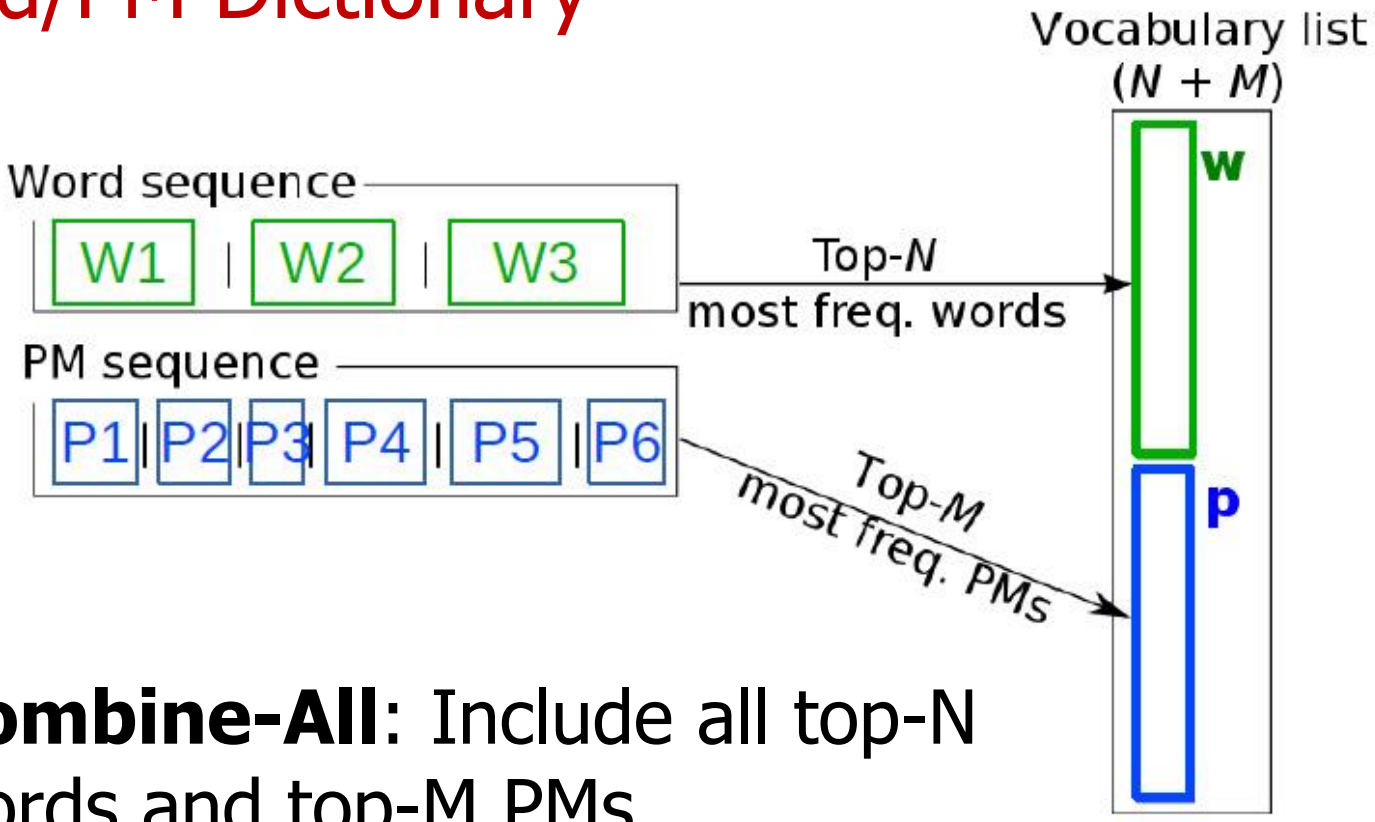


RNN-based Hybrid Language Modeling

- RNNLM as the second-pass re-scoring to reduce the overall system complexity

- X. Liu, Y. Wang, X. Chen, M. J. F. Gales, and P. C. Woodland, "Efficient lattice rescoring using recurrent neural network language models," in ICASSP'14, 2014, pp. 4908–4912.

Word/PM Dictionary



- **Combine-All**: Include all top-N words and top-M PMs
- **Combine-Select**: Select only top-N words then segment the rest words to PMs before select only top-M PMs

Word-PM N-best Lattice: **Combine-All**

- Adding PM paths to the word N-best lattice produced from the first pass decoding

Word-PM N-best Lattice: **Combine-Select**

- Split words not included in the selected top-N words into PMs

Outline

- Background
- Thai language and lexical units
- Combining different lexical units in
  - WFST-based language modeling
  - RNN-based language modeling
- Evaluation and discussion
- Conclusion and future work

Experimental Resources

- **Acoustic model**: GMM with MPE discriminative training by 224 hours of speech
- **Language model**: Open-domain 9.4M words with lexical words limited to 35,000
- **Test set**: 5K utterances from broadcast news and a speech translation mobile app
- **Measure**: Word error rate (WER) and PM error rate (PER)

Evaluation Results

Type	First-pass (3gr)	Second-pass			
		4gr	RNNLM		
			W	PM	Combine-All
PER	20.50	19.78	19.31	19.56	18.99
WER	23.41	22.67	22.25	22.39	22.02

Measures	Combine-All	Combine-Select
PER	18.99	19.01
WER	22.02	22.06
Training time	75h50m	28h02m
Decoding time	15m35s	07m43s

24 cores 2.67 GHz CPU, 98 GB memory

Outline

- Background
- Thai language and lexical units
- Combining different lexical units in
  - WFST-based language modeling
  - RNN-based language modeling
- Evaluation and discussion
- Conclusion and future work

Conclusion and Future Work

- Combining word and sub-word (pseudo-morpheme, PM) units helps improving the performance of open-domain Thai LVCSR both based on WFST and RNNLM

- Including all frequently used words and PMs in the system dictionary produces the lowest recognition errors, but requires much more processing time than that including only most frequently used words and PMs segmented from the rest words

Conclusion and Future Work

- Future work includes
  - Extending RNNLM and hybrid lexical units also to the first pass decoding
  - Analyzing the relationship between the optimal RNN topology and the lexical unit type