玉山銀行 E.SUN BANK

ROCLING 2021

The 33rd Conference on

Computational Linguistics and Speech Processing

October 15-16, 2021 • Online and National Central University, Taiwan

# Integrated Semantic and Phonetic Post-correction for Chinese Speech Recognition

Yi-Chang Chen[1], Chun-Yen Cheng[1], Chien-An Chen[1], Ming-Chieh Sung[1], Yi-Ren Yeh[2]

[1] E.SUN Financial Holding Co., Ltd.

[2] Department of Mathematics, National Kaohsiung Normal University

# Our Team

Yi-Chang Chen (陳宜昌)

Chun-Yen Cheng (鄭俊彥)

Chien-An Chen (陳建安)

Ming-Chieh Sung (宋名傑)

Yi-Ren Yeh (葉倚任)

# Introduction (1/2)

- applications of ASR:
  - voice-activated banking
  - meeting minutes transcription
  - voice content inspection
- HMM-based model  v.s. End-to-end model (requires a huge amount of data)
- one of HMM-based model: kaldi
- Kaldi: acoustic model + language model
- language model of Kaldi: N-gram
  - lack of long-term contextual clues
  - produce many homo-phonic errors (e.g. 你「頭」票了嗎)

# Introduction (2/2)

- Recently, many successful methods have been proposed in natural language processing, such as BERT.
- MLM also could be applied as a post-correction for speech recognition
  - detection model: finetuned BERT (token classification task)

| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 玉 | 山 | 銀 | 行 | 您 | 好 | 很 | 糕 | 新 | 為 | 您 | 服 | 務 |

  - correction model (SIMPLE METHOD): MLM of BERT

| 玉 | 山 | 銀 | 行 | 您 | 好 | 很 | [MASK] | [MASK] | 為 | 您 | 服 | 務 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|

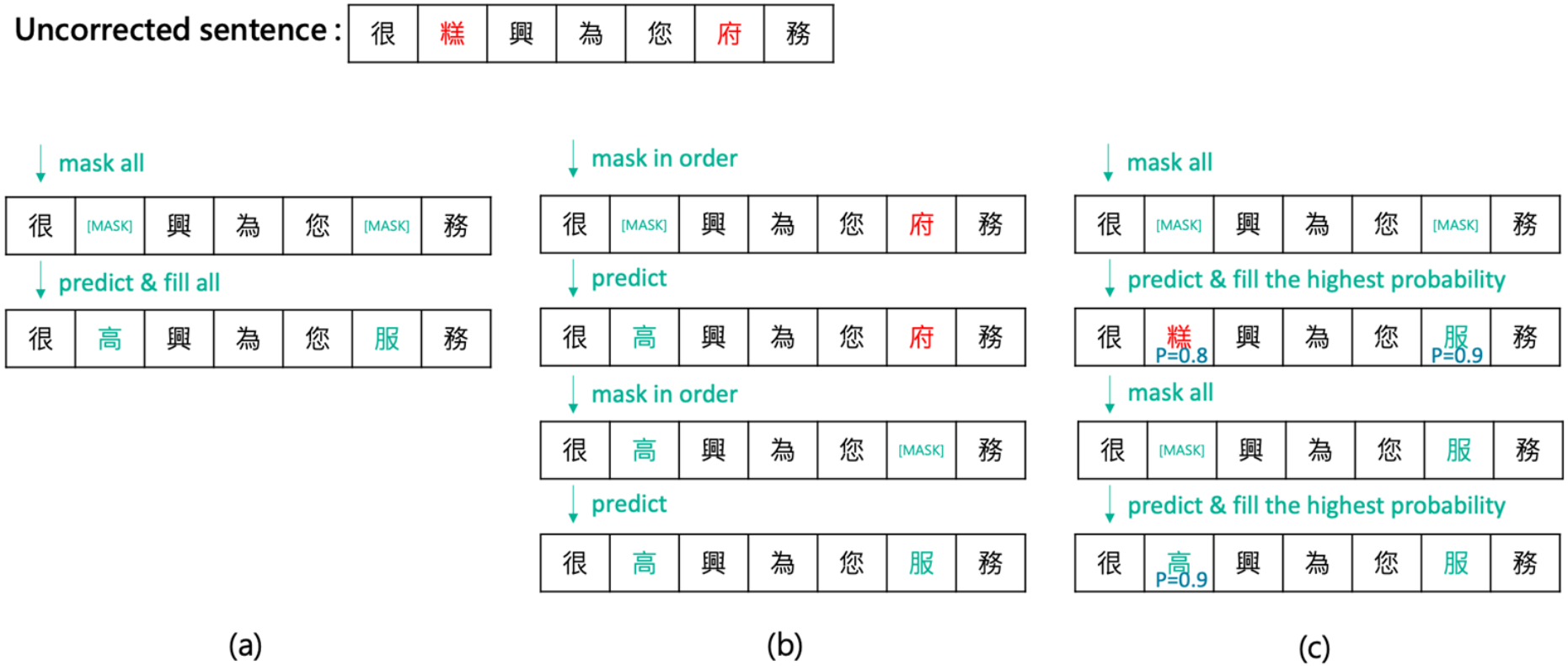
Problem:
only consider semantic info., not phonetic info.

# Semantic Post-correction by MLM

Assume all errors were detected...

| | Datasets | |
|---|---|---|
| | AISHELL-3 | Wiki |
| ✦ mask-all-and-replace-all | 11.69 % | 75.14 % |
| mask-one-and-replace-one | 9.89 % | 73.84 % |
| mask-all-and-replace-one | 11.75 % | 75.62 % |

Table 1: The correction accuracies for different masking and replacement strategies.

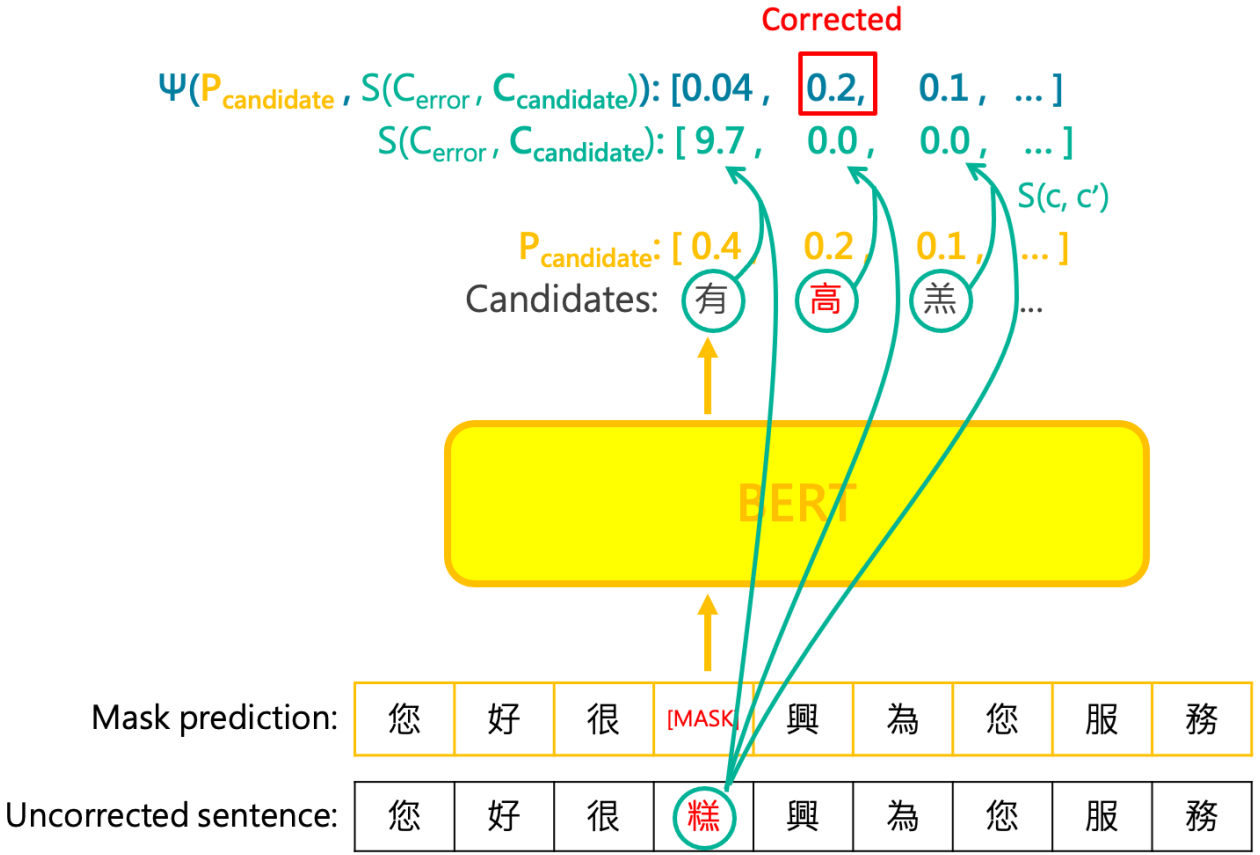


# Phonetic MLM for Post-correction

- DIMSIM: phonetic distance
  - S( c , c′ ) ≥ 0
  - S between two homo-phonic characters is 0
  - S will be larger while the phonic difference is more significant
- Phonetic MLM

$$\Psi(P_{candidate}, S(c_{error}, c_{candidate})) = P_{candidate} \times exp(-\alpha \times S(c_{error}, c_{candidate})),$$

  - S = 0 ⇒ exp(-α × S) = 1
  - S > 0 ⇒ exp(-α × S) < 1

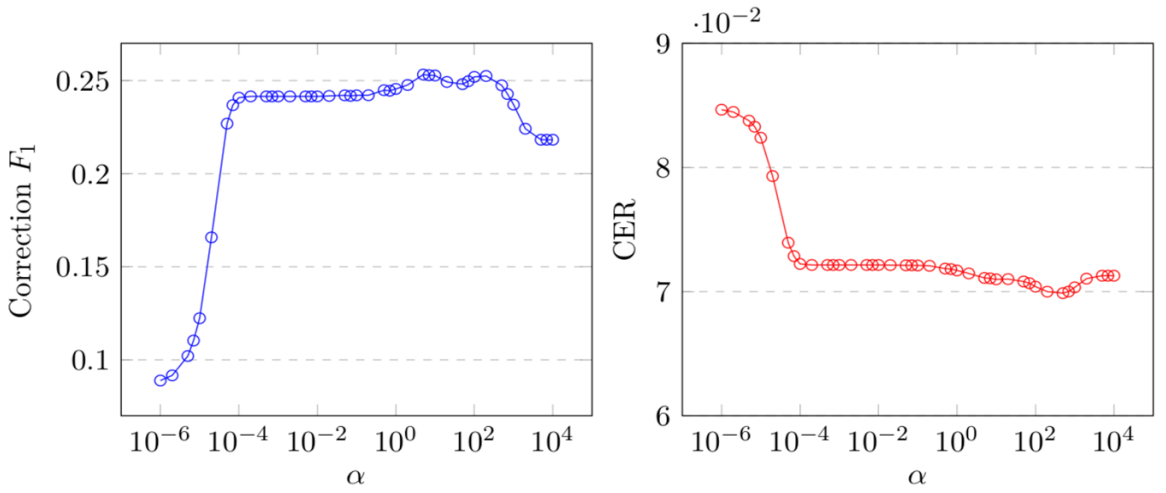


# Phonetic MLM for Post-correction: Result

- Data: AISHELL-3
- Grid search on α

$$\Psi(P_{candidate}, S(c_{error}, c_{candidate})) = P_{candidate} \times exp(-\alpha \times S(c_{error}, c_{candidate})), \quad (2)$$

best: α = 500




- Result

CER of ASR = 9.1%

| | Correction | | | |
|---|---|---|---|---|
| | Pre. | Rec. | $F_1$ | CER |
| MLM | 0.099 | 0.061 | 0.075 | 10% |
| Ours ($\alpha = 500$) | **0.404** | **0.179** | **0.248** | **8.3%** |

# Some Examples

| | |
|---|---|
| ASR: | 也山茶 |
| MLM: | 高山茶 |
| Phonetic MLM: | 野山茶 |

| | |
|---|---|
| ASR: | 首領還並著 |
| MLM: | 首領還活著 |
| Phonetic MLM: | 首領還病著 |

| | |
|---|---|
| ASR: | 直任喜歡女生 |
| MLM: | 直接喜歡女生 |
| Phonetic MLM: | 直認喜歡女生 |

| | |
|---|---|
| ASR: | 我本來可以比了他的 |
| MLM: | 我本來可以殺了他的 |
| Phonetic MLM: | 我本來可以斃了他的 |

| | |
|---|---|
| ASR: | 想發展中國家提供廉價衛星 |
| MLM: | 為發展中國家提供廉價衛星 |
| Phonetic MLM: | 向發展中國家提供廉價衛星 |

| | |
|---|---|
| ASR: | 勸業積極備戰融資融券轉常規 |
| MLM: | 企業積極備戰融資融券轉常規 |
| Phonetic MLM: | 券業積極備戰融資融券轉常規 |

# Recoverable Ability of Phonetic Distance

An error word of interest is unrecoverable if there exists a candidate that satisfies the following two conditions:

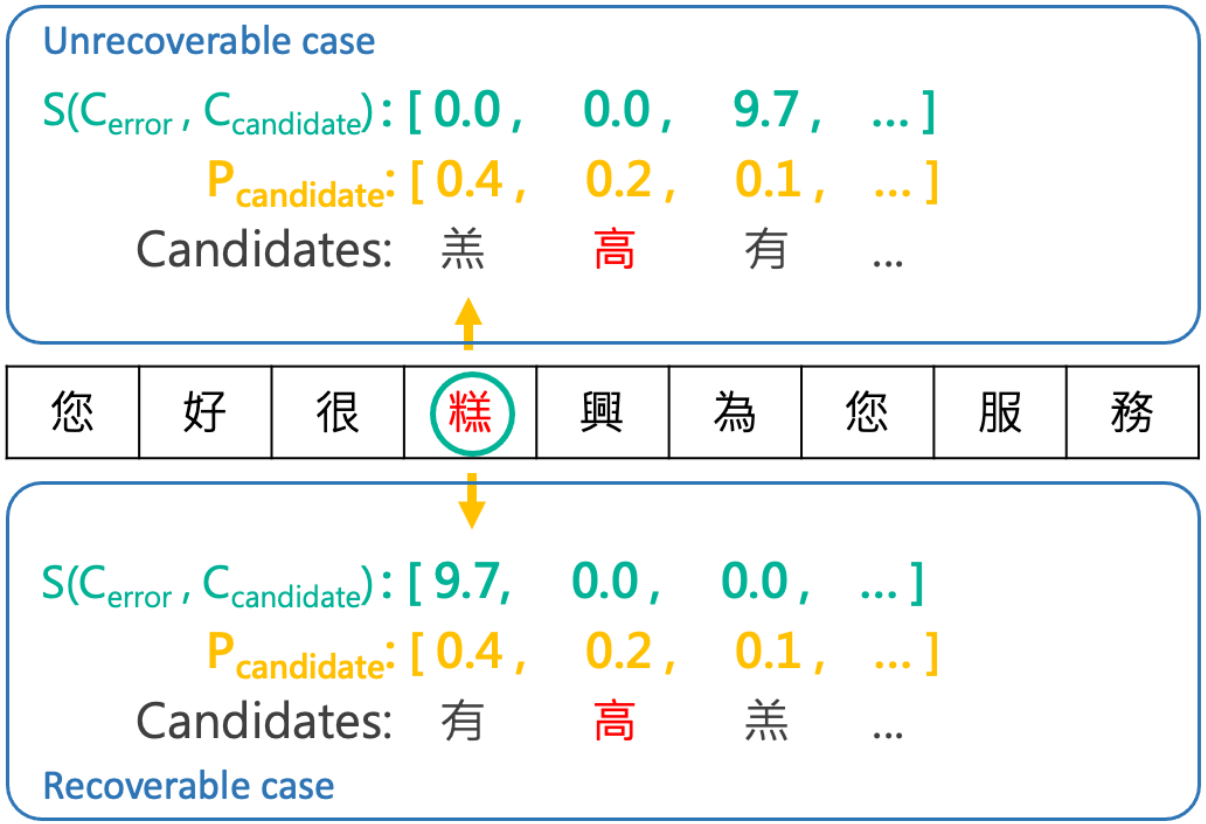
$$P_{error\ candidate} \geq P_{correct\ candidate}$$

and

$$S(C_{error}, C_{error\ candidate}) \leq S(C_{error}, C_{correct\ candidate}),$$

MLM: we have 6,483 recoverable characters (~29.7 %)

Our method can refine 4,671 characters (~72.1%) correctly



# Conclusion

- We observe: Kaldi produces many homo-phonic errors
- A novel approach for post-correction: Phonetic MLM
  - semantic info. + phonetic info.
  - semantic info.: MLM
  - phonetic info.: DIMSIM phonic distance
  - formula:

$$\Psi(P_{candidate}, S(c_{error}, c_{candidate})) = P_{candidate} \times exp(-\alpha \times S(c_{error}, c_{candidate})),$$