Progressive Continual Learning for Spoken Keyword Spotting

Yizheng Huang*, Nana Hou^, Nancy F. Chen*

*Institute for Infocomm Research, A*STAR, Singapore
^Nanyang Technological University, Singapore
Intelligent voice assistants are:

- everywhere around us.
- awakened by specific speech **keywords (KWS)**.
Background: Voice Assistants/Word Spotting (KWS)

Intelligent voice assistants are,
- everywhere around us.
- awakened by specific speech **keywords**.
- mostly deployed in **edge/mobile** devices.

[1] https://voicebot.ai/2021/03/29/where-custom-voice-assistants-are-deployed-in-2021/
Motivation

Limited vocabularies in KWS models.

Cannot deal with unknown words without a large pre-trained model.

Limited Computational Resource

(small memory, slow training speed, etc).

[1] https://voicebot.ai/2021/03/29/where-custom-voice-assistants-are-deployed-in-2021/
Related Works: Continual Learning

Continual Learning Methods

Regularization-based Methods
- Elastic Weight Consolidation (EWC) [3]
- Synaptic Intelligence (SI) [4]

Replay Methods
- Naive Rehearsal
- Gradient Episodic Memory (GEM) [5]

[3] https://arxiv.org/abs/1612.00796 (James Kirkpatrick et al, Overcoming catastrophic forgetting in neural networks; PNAS'17)
[4] https://arxiv.org/abs/1703.04200 (Friedemann Zenke et al, Continual Learning Through Synaptic Intelligence; ICML'17)
[5] https://arxiv.org/abs/1706.08840 (David Lopez-Paz et al, Gradient Episodic Memory for Continual Learning; NeurIPS'17)
Related Works: Fine-tuning

Fine-tuning the KWS model on unknown keywords

[2] https://arxiv.org/abs/2106.02443 (Awasthi et al, Teaching keyword spotters to spot new keywords with limited examples; InterSpeech’21)

Learning new keywords sequentially

**forgetting issue!**
(*even with a large speech pre-trained model*)
Related Works: Gaps

Fine-tuning the KWS model on unknown keywords

limited the learning ability for more words.

forgetting issue!
(*even with a large speech pre-trained model)

higher memory footprint.

[2] https://arxiv.org/abs/2106.02443 (Awasthi et al, Teaching keyword spotters to spot new keywords with limited examples; InterSpeech'21)
Progressive Continual Learning for KWS (PCL-KWS)

Progressive model expanding

Structure of PCL-KWS
Progressive Continual Learning for KWS (PCL-KWS)

Progressive model expanding

Structure of PCL-KWS

classification layers dedicated for each task (one or more keywords).
Progressive Continual Learning for KWS (PCL-KWS)

Progressive model expanding

Structure of PCL-KWS

store the learned features of previously learned tasks.
Progressive model expanding

create a sub-network according to the new keyword learning task.

frozen while learning new keywords.

Structure of PCL-KWS
Progressive Continual Learning for KWS (PCL-KWS)

Keyword-aware Network Scaling Mechanism

Structure of Sub-network

\[
\{16, 24, 32, 48\} \times \alpha_t
\]
dynamic width multiplier
Progressive Continual Learning for KWS (PCL-KWS)

Keyword-aware Network Scaling Mechanism

\[
\{16, 24, 32, 48\} \times \alpha_t
\]

dynamic width multiplier (factor)

\[
\alpha_t = \mu \frac{C_t}{C_0}, (\mu > 0)
\]
determined by the new keyword numbers, and the pre-trained keywords.

Structure of Sub-network
Evaluation & Insights

Comapre with Continual Learning Baselines

The overall accuracy (%) with the number of learned tasks (each has 3 keywords from Google Speech Commands Dataset)

**Stand-alone:** separate model for each task.

**Fine-tune:** without continual learning.

**from PCL-KWS:**
1. near upper-bound performance.
2. better than all CL baselines.
## Evaluation & Insights

### Compare with Continual Learning Baselines

| Method                        | Accuracy (average of all tasks) | Speed (per-epoch training time) | Memory (extra parameters + buffer size) |
|-------------------------------|---------------------------------|----------------------------------|----------------------------------------|
| Fine-tune (Lower-bound)       | 0.39                            | 109.2s                           | N.A                                    |
| Regularization-based (EWC, SI)| 0.45                            | 133.5s                           | 67.69K                                 |
| Replay-based (NR, GEM)        | 0.73                            | 506.9s                           | 132.4M                                 |
| PCL-KWS (Ours)                | 0.91                            | 97.4s                            | 25.5K                                  |
| Stand-alone (Upper-bound)     | 0.94                            | 123.3s                           | 617.8K                                 |

### Regularization-based:
- **High** training speed.
- **Low** memory footprint.
- **Poor** accuracy.

### Replay-based:
- **Low** training speed.
- **High** memory footprint.
- **Good** accuracy.

### PCL-KWS:
- **High** training speed.
- **Low** memory footprint.
- **Good** accuracy.
Evaluation & Insights

Parameter Growth Rate of PCL-KWS

![Graph a](image1)

![Graph b](image2)
Summary

- Apply various continual learning methods for spoken keyword spotting incremental learning.

- Proposed PCL-KWS, a novel continual learning strategy designed for small-footprint KWS.
  - Compare with regularization-based methods, PCL-KWS has better CL performance.
  - Compare with replay-based methods, PCL-KWS has better system efficiency.

- Introduced a keyword-aware network scaling mechanism to reduce the parameter growth rate.
Thanks