Evaluating Lottery Tickets Under Distributional Shifts

Shrey Desai\textsuperscript{1} \quad Hongyuan Zhan\textsuperscript{2} \quad Ahmed Aly\textsuperscript{2}

\textsuperscript{1}The University of Texas at Austin \quad \textsuperscript{2}Facebook Assistant
Lottery Tickets

There exist **small, sparse subnetworks** within a **large, overparameterized model** that, if trained in isolation, will achieve commensurate test performance.

Frankle and Carbin (2019)
Lottery Tickets

Initialize

\[ f(x; \theta_{\text{orig}}) \]

Frankle and Carbin (2019)
Lottery Tickets

\[ f(x; \theta_{\text{orig}}) \]

\[ \text{train}[f(x; \theta_{\text{orig}})] \]
Lottery Tickets

\[ f(x; \theta_{\text{orig}}) \]

\[ \text{train}[f(x; \theta_{\text{orig}})] \]

\[ m \in \{0, 1\}^{|\theta|} \]

Frankle and Carbin (2019)
Lottery Tickets

\[ f(x; \theta_{\text{orig}}) \]

\[ \text{train}[f(x; \theta_{\text{orig}})] \]

\[ m \in \{0, 1\}^{|\theta|} \]

\[ \text{train}[f(x; \theta_{\text{orig}} \odot m)] \]

Frankle and Carbin (2019)
Lottery tickets exist! Resetting network weights to their initial values is crucial.
Lottery tickets exist!Resetting network weights to their initial values is crucial.

Actually, initial values are questionable. Randomly initialized tickets perform well if the LR is tuned.

Frankle and Carbin (2019)
Lottery Tickets

Frankle and Carbin (2019) - Lottery tickets exist! Resetting network weights to their initial values is crucial.

Actually, initial values are questionable. Randomly initialized tickets perform well if the LR is tuned. - Liu et al. (2019)

Morcos et al. (2019) - Lottery tickets obtained on ImageNet offer inductive bias. Initial values are important?
Lottery Tickets

Q1: How well do lottery tickets generalize?
Lottery Tickets

Q1: How well do lottery tickets generalize?

Q2: If so, what is required for generalization?
Cross-Domain Setup

- Amazon Reviews Dataset (20K train, 10K dev, 10K test)

Books  Electronics  CDs  Movies  Home

- Binary sentiment analysis task

He and McAuley (2016)
Testing Generalizability

Source Domain Samples → Lottery Ticket Algorithm → $f^{\text{src}}(x; \theta_{\text{orig}} \odot m)$
Testing Generalizability

Source Domain Samples

Lottery Ticket Algorithm

\[ f^{src}(x; \theta_{\text{orig}} \odot m) \]

Target Domain Samples

Lottery Ticket Algorithm

\[ f^{trg}(x; \theta_{\text{orig}} \odot m) \]
Testing Generalizability

Source Domain

Samples

\( f^{\text{src}}(x; \theta_{\text{orig}} \odot m) \)

Lottery Ticket

Algorithm

Target Domain

Samples

\( f^{\text{trg}}(x; \theta_{\text{orig}} \odot m) \)
Testing Generalizability

Q1: Do LTs transfer?

\[ f_{\text{src}}(x; \theta_{\text{orig}} \odot m) \]

\[ f_{\text{trg}}(x; \theta_{\text{orig}} \odot m) \]
Testing Generalizability

Target Domain Samples

\[ f^{src}(x; \theta_{\text{orig}} \odot m) \]

\[ f^{src}(x; \theta_{\text{rand}} \odot m) \]

\[ f^{trg}(x; \theta_{\text{orig}} \odot m) \]

Q2: What transfers?
Results

\[ f_{\text{src}}(x; \theta_{\text{orig}} \circ m) \]
\[ f_{\text{src}}(x; \theta_{\text{rand}} \circ m) \]
\[ f_{\text{trg}}(x; \theta_{\text{orig}} \circ m) \]

Electronics $\rightarrow$ CDs

Electronics $\rightarrow$ Movies

Source tickets maintain reasonable performance
Performance begins to drop off in high sparsity regime.
Results

Phase transition point; initialization begins to matter
Results

Source tickets have similar performance as target ticket:

- $f^{src}(x; \theta_{orig} \odot m)$
- $f^{src}(x; \theta_{rand} \odot m)$
- $f^{trg}(x; \theta_{orig} \odot m)$
However, once again, performance drops off later.
Results

Phase transition point around 98%+ sparsity levels
Results

Sometimes, source tickets perform better than target tickets.

\[ f^{\text{src}}(x; \theta_{\text{orig}} \odot m) \]

\[ f^{\text{src}}(x; \theta_{\text{rand}} \odot m) \]

\[ f^{\text{trg}}(x; \theta_{\text{orig}} \odot m) \]
Conclusion

1. Lottery tickets exist in natural language processing
Conclusion

1. Lottery tickets exist in natural language processing

2. Cross-domain transfer possible with lottery tickets
Conclusion

1. Lottery tickets exist in natural language processing

2. Cross-domain transfer possible with lottery tickets

3. Phase transition points show the importance of initialization
Thank You!

@shreydesai