Exploring deep learning for multimodal understanding

Lao, M.

Propositions

pertaining to the thesis

Exploring Deep Learning for Multimodal Understanding

by Mingrui Lao

1. Significant accuracy in answer prediction may not mean significant VQA model ability in open-world applications. [Chapter 2]

2. Learning a fine-grained multimodal fusion feature establishes the relationships between visual and textual modalities in VQA task. [Chapter 3]

3. Debiasing strategies unavoidably need to make a trade-off between in-distribution and out-of-distribution performance. [Chapter 4 & 5]

4. Shortcut biases are typically more severe and challenging when the multimodal QA systems involve more modalities. [Chapter 6]

5. The crucial factor to accomplish a multi-domain lifelong learning machine is to extract informative knowledge from previously learned domains.

6. In federated learning under severe data heterogeneity, client models inevitably forget generic knowledge aggregated by central server during the local training.

7. The generalization and continual learning abilities are key to achieve human-level AI.

8. The working of the human brain and its reasoning behaviour are an important inspiration to design better neural networks.

9. The success of multimodal understanding models in the future will be determined by how interpretable they are.

10. There are no shortcuts to any place worth going.