



Deep learning evaluation using deep linguistic processing

Author : Alexander Kuhnle , Ann Copestake
Presented by : Nazia Ahmed Nijhum



Introduction

- *Highlight the increasing performance of deep neural networks (DNNs) in various tasks.*
- *Emphasize the shift towards artificial data for evaluation in fields like reinforcement learning*



Issues with Current Evaluation Practices

- Discuss *problems with standard approaches to evaluating tasks like visual question answering (VQA).*
- Point out the *inadequacies of popular datasets, such as the VQA Dataset, in truly assessing language understanding abilities.*



Artificial Data for Evaluation

- *Introduce the concept of artificial data as a complement to current evaluation practices.*
- *Mention the creation of challenging abstract datasets using deep linguistic processing technology.*



Modified and Artificial VQA Datasets

- *Highlight modifications to existing VQA Dataset and introduction of artificial datasets like SHAPES, CLEVR, NLVR, and ShapeWorld.*
- *Emphasize the importance of these datasets in providing clear and challenging evaluation for VQA systems.*



Advantages of Artificial Data

- *Argue that artificial data is essential for targeted investigations of specific model capabilities.*
- *Stress the capacity of artificial data to reveal fundamental shortcomings of current VQA models.*



Proposed Evaluation Methodology

- *Provide a brief review of problems with current standard evaluation practices.*
- *Propose an evaluation methodology based on linguistic processing resources, emphasizing compositional semantic representations.*



Automatic Generation of Language Data

- *Describe the approach for automatic generation of artificial VQA data using deep linguistic processing technology.*
- *Highlight the use of abstract microworlds and syntactically rich language generation.*



Quantification Example

- *Present an example image with various captions focusing on quantification.*
- *Emphasize the complexity of quantifiers and the need for multimodal reasoning.*



Conclusion: Why Use Artificial Data?

- *Summarize the benefits of using artificial data for evaluation.*
- *Highlight challenging test data, avoidance of the Clever Hans effect, flexibility, reusability, and the potential for rich evaluations.*