Report (task-1)

Paper title: Deep learning evaluation using deep linguistic processing

Paper link: https://aclanthology.org/W18-1003.pdf

Summary:
The study emphasizes the shortcomings of conventional methods for assessment in tasks such as visual question answering (VQA) and suggests augmenting existing methods with generated data. In the framework of VQA, the emphasis is on visually grounded language understanding. The authors contend that there are flaws in current datasets, such as the well-known VQA Dataset, that enable algorithms to perform competitively even in the absence of true linguistic understanding. Changes to current datasets and the creation of fake VQA datasets are suggested as solutions to this. The ShapeWorld framework, which generates difficult abstract datasets by utilizing deep language processing technologies, is introduced in this research.

Contribution:
1.Bringing out concerns with real-world datasets, especially in VQA, and suggesting the use of synthetic data for focused research.
2.Introduction of new datasets such SHAPES, CLEVR, NLVR, and ShapeWorld as well as changes to the already-existing VQA Dataset.
3.The focus is on how artificial data might serve as an essential reference point for true language comprehension skills, as opposed to merely enhancing current models with data augmentation.
4.An assessment methodology based on language processing resources is presented, utilizing a bidirectional grammar and compositional semantic approach.

Methodology:
1.Utilizing high-precision English Resource Grammar and the Dependency MRS (DMRS) in deep linguistic processing, one may create natural language sentences from abstract DMRS graphs.
2.creation of synthetic VQA data using randomly selected abstract world models that define entities, microworlds, and their characteristics.
3.An endless number of captions with any level of syntactic complexity may be created by using compositional semantic representations.
4.To demonstrate the intricacy of produced captions and the requirement for visual reasoning, concentrate on quantification instances.

Conclusion:
The benefits of employing generated data in difficult test situations, avoiding the Clever Hans effect, offering flexibility and reusability, and facilitating rich assessments of deep neural networks (DNNs) are emphasized in the final section of the paper. The authors stress that rather than depending on monolithic datasets with a single accuracy number, modular testing for certain subtasks is vital.

Limitations:

1st:
In order to solve the scope ambiguity of nested quantifiers, the study admits the need for more tests and extensions.

2nd:
The authors point out that while generated data offers insightful information, performance on more limited datasets could not always translate to better results on larger datasets that include the VQA Dataset.

Synthesis:
The study analyzes the shortcomings of the VQA evaluation procedures now in use and offers a thorough remedy based on artificial intelligence and language processing technologies. As an application of this strategy, the ShapeWorld framework is shown, with a focus on creating abstract datasets for specific assessments. Exemplary instances of quantification and compositional semantic representations are used to highlight the intricacy of language understanding problems and their requirement for visual reasoning. For in-depth analyses of DNNs, the modular testing technique is recommended, guaranteeing a deeper comprehension of their strengths and weaknesses.