VIVEK SURESH RAJ

RESEARCH STATEMENT:

I'm a recent university graduate with interest in contributing to conversational dialogue generation in natural language processing. Artificial Intelligence has successfully solved real world problems in natural language processing and dialogue generation. I'm highly motivated in working through the quality of embodied agents than can perform multi-turn conversations or interactions for language grounding, dialogue understanding and task completion.

PREVIOUS EXPERIMENTS IN THE AREA OF CONTROLLED TEXT GENERATION:

I've previously worked on controlled text generation with seq2seq learning architecture which learns via Conditional probability [P(y | x)]. The objectives of the research were to propose suitable metrics upon which the generated text at the decoder output could be observed and evaluated, to present high-level mechanics of control methods in text generation and to analyse the effects of bottleneck vector/context vector in the model. The experiment was conducted on encoder-decoder based learning architecture with homologous GRU layers.

However, variation at the decoder output was "controlled" with respect to Weighted Decoding (WD) algorithms including greedy search, beam search and top k-sampling methods. The biggest challenge was to introduce "randomness" to overcome generic and repetitive responses at the decoder output. So, each timestep of the decoder was updated with the probability of each word in vocabulary in proportion to its randomness. A suitable weighted parameter was introduced called SoftMax "temperature" (τ) as a hyperparameter, which could be manually altered by the user. Now, that the SoftMax distribution could be scaled by the parameter (τ) which makes the probability distribution, y-hat ($\hat{y}$) as uniformly distributed labels at each timestep at the decoder output. The size of the weight parameter would be encouraged to generate more generic and specific texts at decoder. Observations were made at the generated texts and corresponding evaluations were reported.

FUTURE RESEARCH DIRECTION– AIM AND OBJECTIVE:

I'd like to address on the question that is currently an on-going problem among the conversational AI in the world.

- How best to improvise embodied intelligent agents for language understanding to communicate in natural language and follow natural language instructions for task completion?

In future I intend to collaborate natural language processing with Deep learning and Deep Reinforcement learning, Embodied learning to solve the existing problem via interactive learning.

POSSIBILE APPROACHES TO THE PROBLEM:

In future I intent to proceed my research in general direction of solving harder NLP-Robotics problems by learning to learn agents to communicate in natural language by interacting with dynamic environment, people and language grounding.

Exploitation of embodiment. On a practical note, language grounding is known as an association or alignment between linguistic element to perception and action. For example, I'd like to start arguing with a simple search engine, that sees a noun-phrase like a "dog" or "cat" and brings back the image related to the noun-phrase through a classifier. Thus, it looks simpler with an SVM or neural networks, but I'm more interested to deal with verb-phrases like "Dancing above the table" in which it deals with **w**(*dancing*), **w**(*above*) and **w**(*table*) which requires individual attention. (Sun, March 2020)

Thus, in future I aim to model on a dynamic environment as described above and align it with natural language tokens. I wish to center my efforts on importance of embodied agents than can perform multi-turn conversations or interactions for language learning.

As a first step, I would like to work on multi-modal inputs such as agents understanding of its actuators while it's performing actions as, it will in verbally describing the action as well as to know about the action-object generalization. This is the reason why Embodied language learning is implemented in conversational systems involved in navigation tasks or in tasks where it needs to perform an action on the object to answer the questions. Like dialogue grounded in vision, such interactive scenarios extend language learning into multiple modalities. So, I intend to work on natural language description or explanation of the scene using Deep reinforcement learning neural architectures for processing of multiple modalities in embodied language learning in a dynamic world.

Goal-driven embodied agents for conversational learning. As a second step, I'd like to continue my study further on natural language interactions (like a trainer model) to people in understanding and executing the task (follower model that could interact with environment) by using conversations and *solving for ambiguity, recovering from mistakes*. Finally, I'd like to work on evaluating the models' ability in dialogue understanding, language grounding and task-execution.

On the theoretical side, some questions I want to explore are: Goal discovery vs policy discovery? What if a transformer-based language model is employed in embodied language-conditioned task completion systems?

[I'm also interested in working out the problem on different dimension through language grounding RL + Imitation learning and Self-supervised learning].

REFERENCES:

Sun, Z. L. (March 2020). *Representation Learning for Natural Language.* Beijing: Springer imprint.