Research Statement

Bhiman Kumar Baghel
CS PhD

1 Previous Research Experience

I have research experience in the domain of Conversational AI like Bixby, Alexa, Google Assistant and Sir. They generally take an audio input (known as an utterance), convert it into text and then extract information from it in order to perform task stated in the utterance. My research focused on extracting information from the text. This information has two major components - 1) Intent and 2) Slot. Intent signifies the action the user want the Conversational AI to perform. Whereas, slots are the object on which the action is performed. For utterance ”Turn on TV”, the intent is ”TurnOn” and ”TV:Device” is a slot with ”Device” as slot type and ”TV” as slot value. Due to their high correlation they are jointly modeled ([3]) in multi-tasking fashion with intent predicted using classification and slot extracted using sequence tagging. This approach worked in scenarios where the utterance contains single intent. However, it can not handle multi-intent scenarios. For utterance ”Turn on TV and turn off Light”, it is able to find ”TurnOn+TurnOff” intent but contains no mechanism to relate slots ”TV:Device” to ”TurnOn” and ”Light:Device” to ”TurnOff”. All my research finding were applied in smart home domain, so my examples would be following that theme.

1.1 Multi Intent and slot Extraction

Multi-intent scenarios come with its own challenges. The challenges identified and resolved through my research are as follows:

1. **Intent and Slot Relation:** Intents should be related only to valid slot and their values. For example in the utterance - ”Turn on TV and turn off the light”, there are two intents *TurnOn* and *TurnOff* with the correct slot assignment as *Device:TV* and *Device:light* respectively. Swapping this assignment would result in the wrong outcome. So in a multi-intent scenario maintaining the correct intent-slot relationship is important, which is ignored in a single-intent scenario as there is only one intent to relate to.

2. **Ranking:** The order of intent execution also becomes important in the multi-intent scenario. For example in the utterance - ”Set volume of TV to 20 after turning the TV on”, the TV should be turned on first then its volume should be set to level 20. It’s logically not possible to perform actions in reverse order.

3. **Scaling:** A sound multi-intent system should not be limited to a pair or triplet of intents. It should be theoretically scalable to relate an infinite number of intents and corresponding slots and rank them.

4. **Partial Utterances:** User don’t want to use short utterances because of its convenience. This is where multi-intent utterances come in handy. User would prefer ”Turn on Tv and set volume 20” instead of ”Turn on TV and set volume of TV to 20”. In the partial utterance it can be observed that different intents ”TurnOn” and ”SetVolume” are related to same slot ”TV:Device”.

My work, STACK [5], proposed modeling intent prediction as sequence tagging for each token, similar to slot tagging, as shown in Fig. 1 d). This created a 1-to-1 mapping between intents and their slots. Hence, resolving the issue of relating intents and slots. However, observing the model design reviles that it is not explicitly modeling intents and slot relation. This was the reasons for this works reliance on stacking ensemble of different strong models for good performance.

Works like Slim [2] explicitly modeled intents and slot relation, as shown in Fig 1 c). This lead to good performance. However, the system was complex. The motivation to develop a simpler system with same advantage lead to my work, SPARCS. SPARCS simply predicted all the information as a single sequence maintaining the 1-to-1 relation, as shown in Fig 1 b).

All the above works can easily accommodate ranking and scaling. However, they fail to handle partial commands because they can not allocate multiple intents to one slot i.e. one-to-many mapping. This is were my most recent work come in picture. It predict a sequence of intents and a sequence of slots for each intent which helps in creating the one-to-many mapping as same slot token can be used in for different intents.

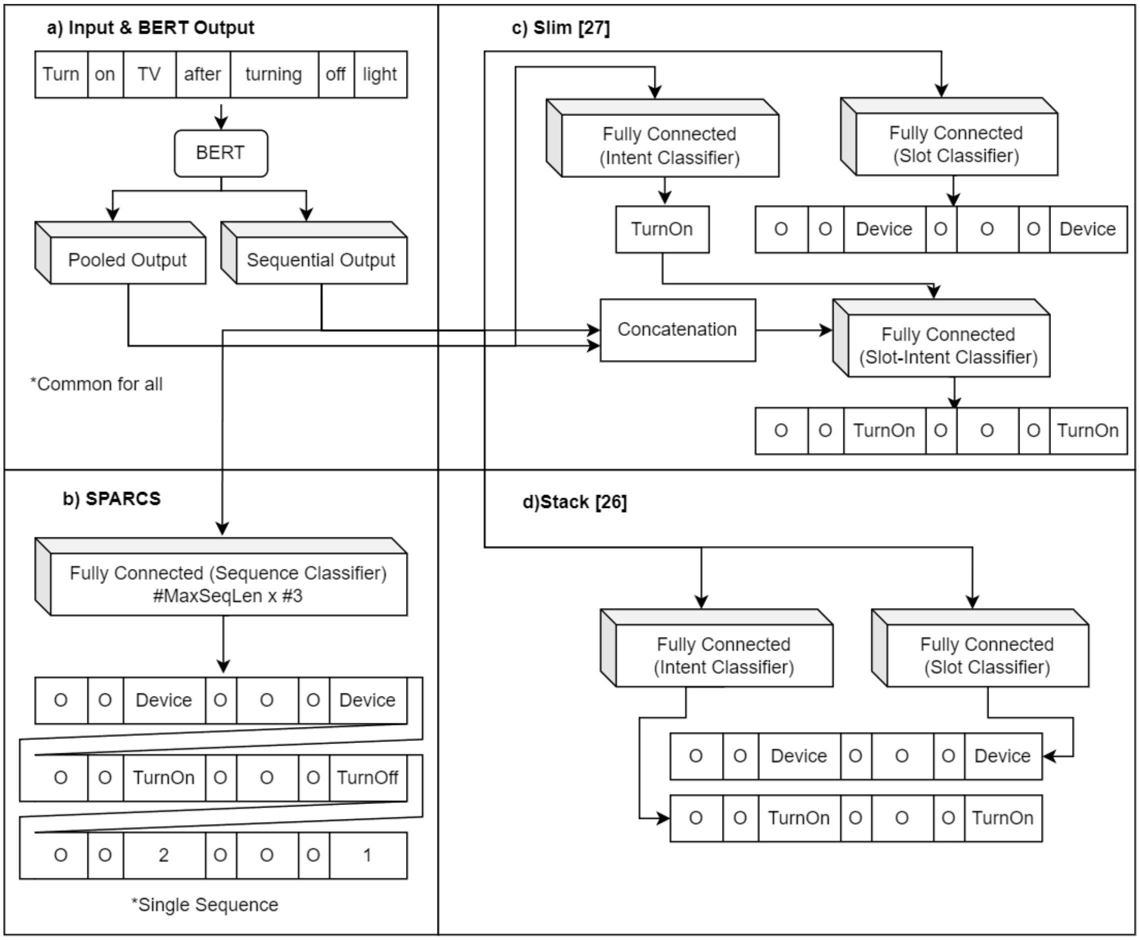

Figure 1: Previous Research

1.2 Handling Missing Slot

What if user hasn't provided necessary slots, like "Set TV 20". Does the user want volume to be set to 20 or brightness? This situation raise the problem of missing slots. Without this slot, the system can not function further. To handle such situation my work (under review at NLE) proposed a dataset creation, annotation and modeling strategy.

1.3 Code-Mixed Language Understanding

I also worked on research to extract intent and slots from code-mixed language. SPARCS is an example of one such work. SPARCS focused on Hinglish which is combination of Hindi and English. However its findings are application of any code-mixed language like Hinglish.

2 Current Research Interest

Now when I look back, I realize that the Conversational AI systems which I have worked with don't have common sense and reasoning capabilities. There were many error scenarios where model predicted the generic output but the expected was as per the user's dynamic context of the devices he has, their states and many other environmental factors. This one of the many observations which leads to my current research interests:

2.1 Grounded Conversational AI

Current Conversational AI's are just atomic task following bots. They are not grounded to the user's environment. Due to this they can't related to user's context and produce personalized results. I developed a PoC (proof of concept) to give user's context as input to BERT in order to provide personalized outputs in smart home environment. However, it never got change to be tested with full data and develop further.

Till now I discussed about Conversational AI performing atomic actions. However, users can provide a way abstract task which needs conversational AI to perform devise and perform a sequence of atomic actions. There are woks like [4], [1] that ground LLMs to work with embodies agents. Current smart homes are only IoT devices. But future will be combination of both. I believe a research direction which enables conversational AIs to ground and reason in order to utilized these devices/agents with full potential is worth exploring.

2.2 Conversational AI as Role players

I believe Conversational AI will integrate into our life playing critical roles as educator, tutor, counselor, law and financial advisor and many more. In order to play this roles they need to have the deeper understand to the user intent and the world which they currently lack.

2.3 Fair & Safe Conversational

Also, such role playing conversational AIs would have high impact on our decision making. So, they should be free from biases and should be reasonable to not share harmful information.

These are the current directions which I would like to explore.

References

[1] Michael Ahn et al. *Do As I Can, Not As I Say: Grounding Language in Robotic Affordances.* 2022. arXiv: `2204.01691 [cs.RO]`.

[2] Fengyu Cai et al. "Slim: Explicit Slot-Intent Mapping with Bert for Joint Multi-Intent Detection and Slot Filling". In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* 2022, pp. 7607–7611. DOI: `10.1109/ICASSP43922.2022.9747477`.

[3] Qian Chen, Zhu Zhuo, and Wen Wang. *BERT for Joint Intent Classification and Slot Filling.* 2019. arXiv: `1902.10909 [cs.CL]`.

[4] Wenlong Huang et al. *Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents.* 2022. arXiv: `2201.07207 [cs.LG]`.

[5] Niraj Kumar and Bhiman Kumar Baghel. "Smart Stacking of Deep Learning Models for Granular Joint Intent-Slot Extraction for Multi-Intent SLU". In: *IEEE Access* 9 (2021), pp. 97582–97590. DOI: `10.1109/ACCESS.2021.3095416`.