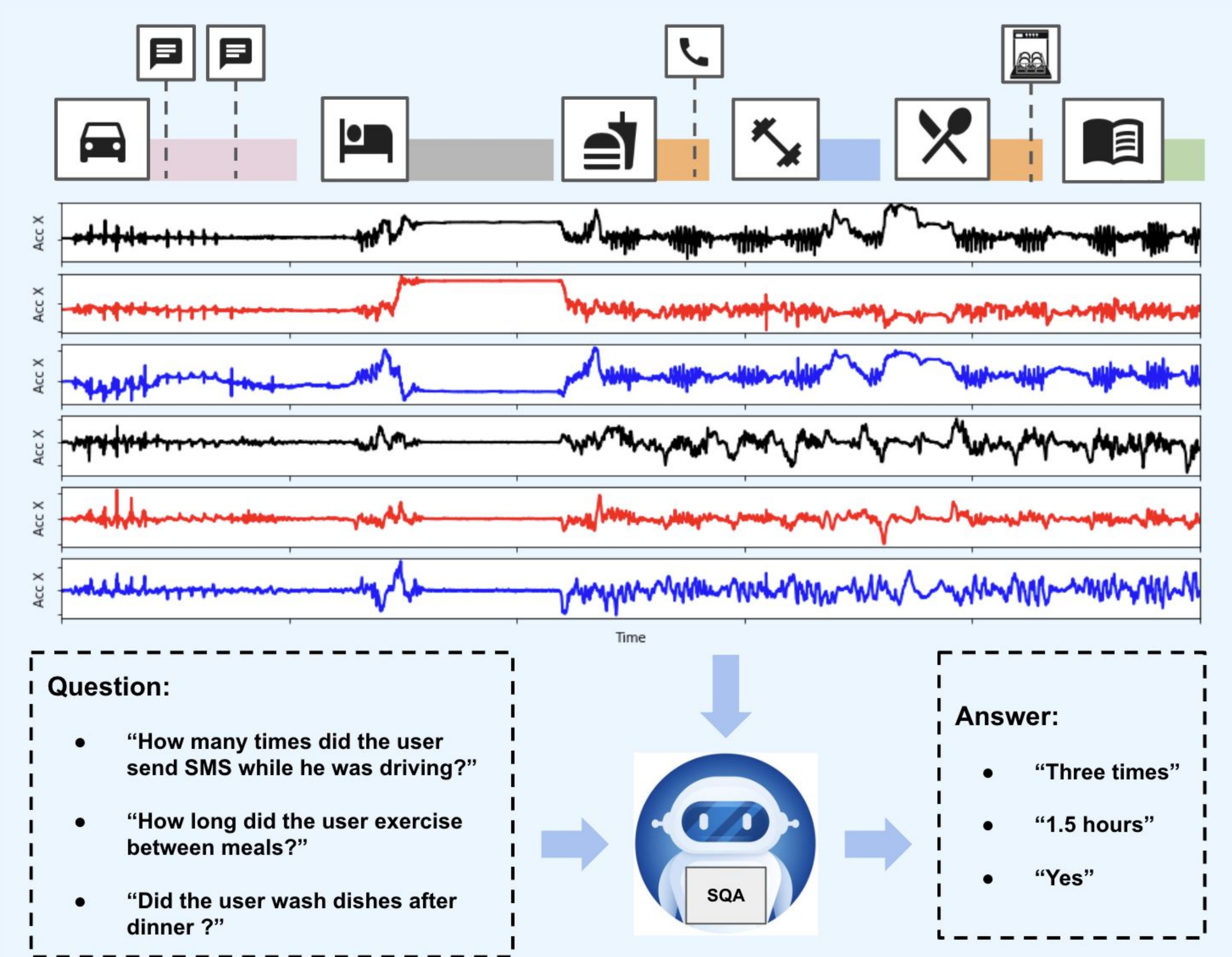
DeepSQA: Understanding Sensor Data via a Deep Learning Approach to Sensory Question

Tianwei Xing, Marc Roig Vilamala, Luis Garcia, Federico Cerutti, Lance Kaplan, Alun Preece and Mani Srivastava

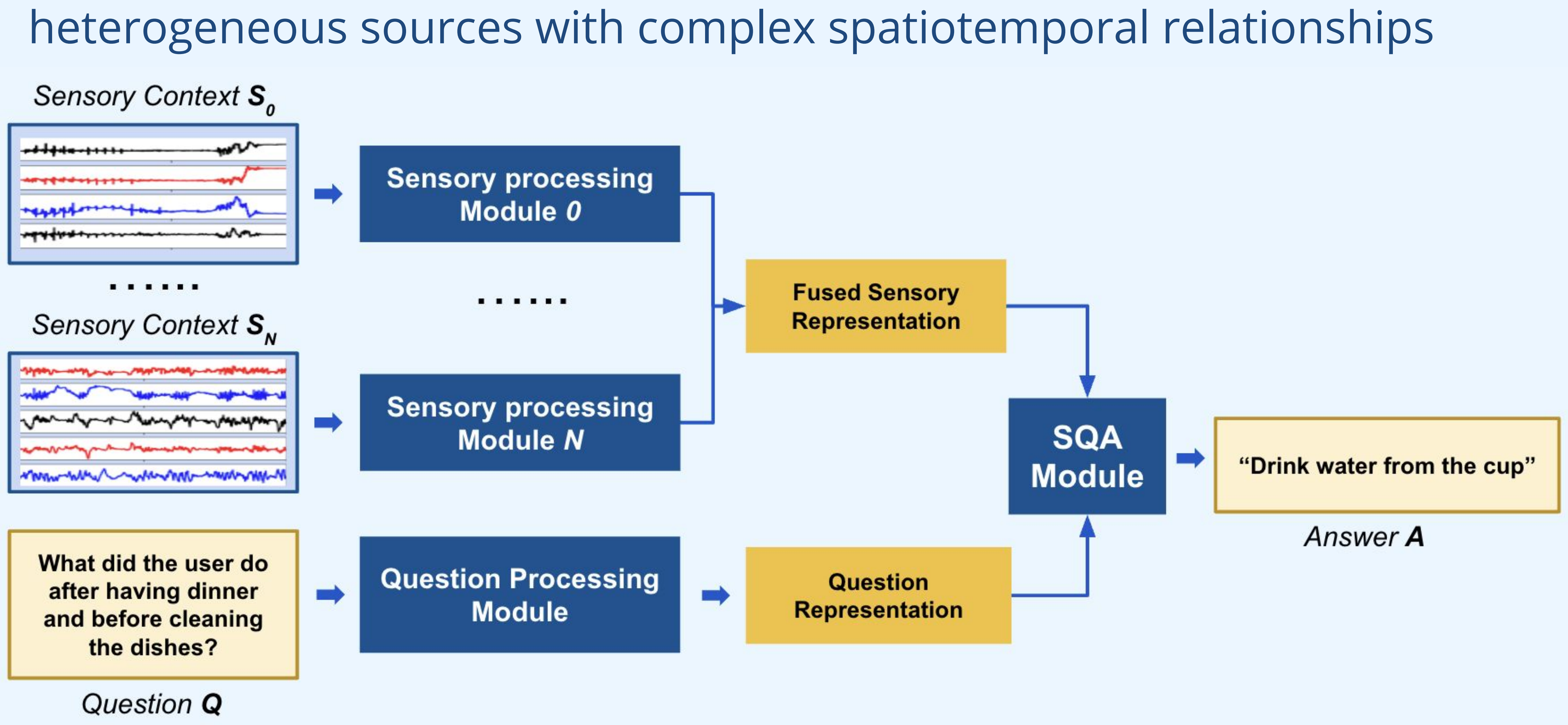
MOTIVATION

- Current machine learning systems can only perform one or a few limited task on sensory data.
- Introducing new tasks requires re-training of the system.
- User cannot get other information except for predefined high-level labels.



CONTEXT & FORMULATION

Background: QA in other domains (text QA, visual QA) use various techniques (bilinear pooling, attention, etc.) to fuse information and make inference. However, none of them works on sensory data from heterogeneous sources with complex spatiotemporal relationships



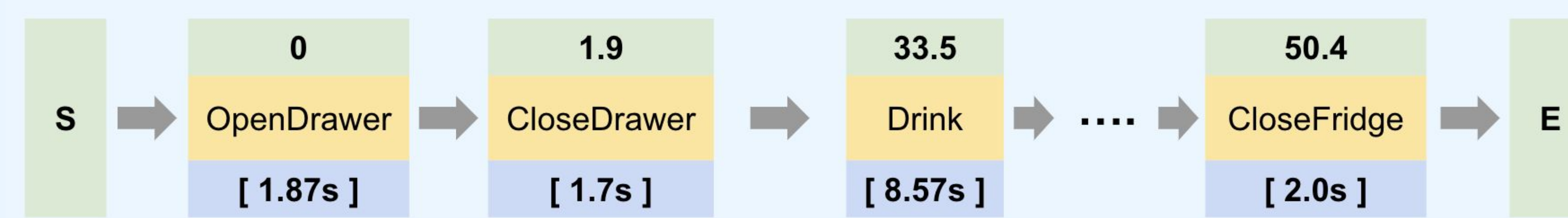
DeepSQA framework:

- Sensory data are first processed and fused by deep learning modules.
- Questions are transformed into embedding sequences by learned word embedding matrix, and then processed using bi-directional LSTM.
- A recursive attention network is used to extract answers iteratively from sensory data based on questions.

SQA DATASET GENERATION

- SQA dataset are generated automatically using a template-based approach. Each element in SQA has its own semantic representation.

Semantic Representation: Scene List $S = \{A_1, A_2, \ldots A_n\}$, $A_i = \{t_i, d_i, s_i\}$



Semantic Representation: Functional Program

What did the user do **[before]** **<opening the fridge>** **and** **[after]** **<closing the drawer>**?

query_action_type(AND(relate(before, open the fridge), relate(after, close the drawer)))

Answer obtained from question engine: Apply the functional program above on the scene list

PRELIMINARY RESULT

- **Evaluate SQA models on OppQA dataset:**
  - **Source dataset:** OPPORTUNITY, ~8h, 4 users, 24 runs
  - **Question family**: 16, (query, existence, counting, comparison...) 110+ templates,
  - **Avg length:** 21 words per question
  - **Answers**: 28 unique answers for classification task

Question Types	Prior	Prior-Q	Neuro-Symbolic	CNN	LSTM	CRNN	SAN	DeepSQA
Overall	44.16%	56.50%	45.61%	44.23%	63.81%	72.09%	57.16%	**74.77%**
Existence	60.08%	41.69%	42.62%	59.88%	62.77%	**73.11%**	62.13%	72.02%
Counting	0%	55.89%	28.40%	0%	57.30%	65.07%	56.99%	**67.56%**
Query	0%	5.67%	0%	0%	40.25%	46.13%	39.49%	**52.77%**
Compare	59.67%	75.27%	69.43%	59.81%	72.15%	76.34%	57.97%	**79.75%**
Compare-act	45.45%	45.45%	0%	46.17%	56.10%	**80.00%**	53.18%	75.00%

- **Adopt a representative subset of methods from VQA task:**
  - **Prior and Prior-Q**: predict answer based on training answer statistics
  - **LSTM/ CNN** : baselines using only question/ sensory data as input
  - **CRNN**: LSTM + CNN with easy information fusion before classification
  - **SAN**: Stacked attention network
  - **DeepSQA**: proposed method using compositional attention model.
- **DeepSQA model shows better performance than other baselines**

FUTURE WORK

- Study the SQA model robustness to linguistic variations, and train robust SQA model using the cycle-consistency framework.
- Investigate the neural-symbolic approach for sensory QA, explicitly parse the input question and construct task-dependent model during the runtime.

RELATED WORK

[1] Xing, T., Vilamala, M., Garcia, L., Cerutti, F., Kaplan, L., Preece, A. and Srivastava, M., 2019, June. "DeepSQA: Understanding Sensor Data via a Deep Learning Approach to Sensory Question Answering" Submitted 2020.

UCLA

Project Repo: https://github.com/nesl/DeepSQA

Contact: twxing@ucla.edu

NESL Networked & Embedded Systems Laboratory