Ravestate: Multimodal Contextual Dialog State Tracking As Bayesian Specific Signal Transduction

Birkner, Joseph
joseph.birkner@tum.de

Karimi, Negin
negin.karimi@tum.de

Hostettler, Rafael
rh@gi.ai

Dolp, Andreas
andreas.dolp@tum.de

Airiian, Wagram
wagram.airiian@tum.de

Kharchenko, Alona
unicorn@roboy.org

1 Abstract

A major challenge in the development of Natural Language Dialogue Systems is to determine the intent of a user utterance, and to map the intent of an utterance U to a certain dialogue application state T. While recent work in this area focuses on embedding these variables as Neural-Network generated latent representations, we hypothesize that a symbolic approach to Dialogue State tracking might deliver higher utility with reduced development effort: By observing dialogue system behaviour as words out of a formal language over signal spikes in the application context, with application states acting as contextual nonterminals, we set up a basic formal framework for dialogue state propagation. Furthermore, we propose the notion of constraint-based Bayesian state specificity as a measure of utility to resolve conflicts between overlapping application states. We implement our system in the open-source library RAVESTATE. Experiments with the implemented system both in text- and speech based scenarios with additional video input show very robust contextual behaviour, while operating fully causally explainable and transparently.

2 Introduction

3 Related Work

4 Signal Transduction

4.1 Formalising Causal Intuitions

Any dialogue system $D : X, H_t \rightarrow Y, H_{t+1}$ is a function mapping input variables X and a context H_t at timestep t to output variables Y and a context H_{t+1}. For example, inputs may be a textual user utterance or visual stimuli, and outputs may be a textual response or a gesture. RAVESTATE models these variables as so-called **Properties**:

Definition 1 (Property)**.** *A property $p^i \in P$ is a pair $\langle L^i, v^i \rangle$ containing a synchronisation primitive L^i and a value v^i. It supports the operations* $\text{READ}(P^i) = v_i$ *and* $\text{WRITE}(P^i, x) \rightarrow \text{READ}(P^i) = x$.

An intuitive approach towards modelling a dialogue system is to articulate it's behaviour as a specifically conditioned reaction to certain events or *signals* which are derived from it's inputs.

Definition 2 (Signal)**.** *A signal $c^i \in C = \langle id^i, age^i_{min} \in \mathbb{R}, age^i_{max} \in \mathbb{R} \rangle$ is a description of a specific event, which may be fulfilled by event instances in the form of spikes. It consists of a name id^i, a minimum spike age age^i_{min}, and a maximum spike age age^i_{max}.*

Definition 3 (Spike)**.** *A spike $s^i \in S$ is a triplet $\langle id^i, age^i \in \mathbb{R}, cause \subset S \rangle$ which corresponds to a real-world instance of any signal $c^{sig} \in C$ where* $\text{EVAL}(s^i, c^{sig}) = \text{TRUE}$. *The function* $\text{EVAL} : S, C \rightarrow \mathbb{B}$ *is defined as follows:* $\text{EVAL}(s^i, c^{sig}) := (id^i = id^{sig}) \wedge (age^{sig}_{min} \leq age^i \leq age^{sig}_{max})$

Any specific reaction to a set of signal spikes is called a state. In cognitive systems theory, a condition-state pair is called a production. For example, an application state for answering personal questions about the agent may be seen as a reaction to a combination of two events/signals (see figure 1):

1. QUESTION The input is a question.
2. ABOUT_AGENT The input sentence's subject is the agent itself ("you").

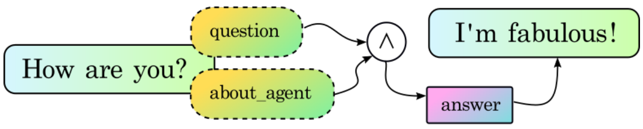

Figure 1: Example of a causal signal-state production relationship.

Such combinations of signals are modelled as **Conditions** in RAVESTATE:

Definition 4 (Condition)**.** *The condition set* $\text{COND}(C)$ *is a boolean algebra over signals C: Let* $\text{COND}(C) := \text{COND}(C \cup \bigcup_{(c_x, c_y) \in C \times C} \{c_x \wedge c_y, c_x \vee c_y\})$. *The definition of the* EVAL *function for conditions is extended as follows:*

In RAVESTATE, productions are directly adapted into a generic state machine:

Definition 5 (Property-Changed-Signal)**.**

Definition 6 (State)**.** *A state T^i is a six-tuple $\langle P_R^i \subset P, P_W^i \subset P, \mathrm{ON}^i \in \mathrm{COND}, f^i : P_R, P_W \rightarrow \mathrm{RESULT}, C_{emit}^i \subset C \rangle$: It can write to a set of properties P_W^i, read from a set of properties P_R^i, execute it's state function f^i in reaction to the condition ON^i, and emit spikes for a subset of signals C_{emit}^i.*

The input question *How are you* is held by a **Property** $p_1 \in P$. For the given example, the QUESTION and ABOUT_AGENT events are **Signals** $\{c_1, c_2\} \subset C$ out of a superset of signals C. In order to close the gap between the property p_1 and it's derivative signals, an intermediate signal is introduced: As p_1 adapts a new value, it emits a CHANGED signal $c_0 \in C$. Furthermore, note the following definition of a **State** in RAVESTATE:

The ANSWER state is a **State** $t_1 \in T$.

The combination of States T, Activations A, Signals C, Spikes $\hat{S}$ and Properties P is called a **Context** $H = \{T, A, C, \hat{S}, P\}$.

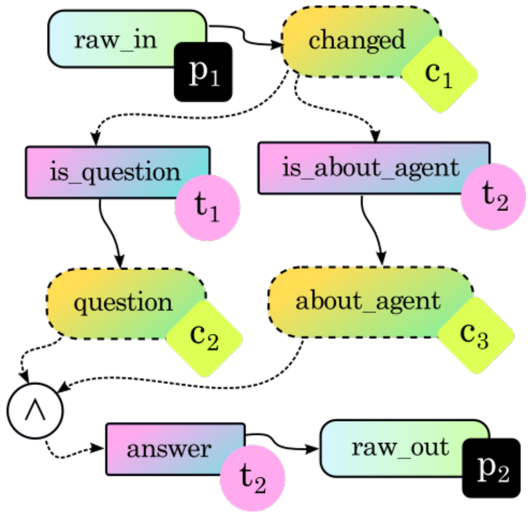

Figure 2: Signal-Flow Diagram for previous example.

4.2 Signal Flow

States $t \in T$ (Processes, Transition Functions, Transducers, Non-Terminals)

Properties $p \in P$ (Data, Channels)

Signals $c \in C$ (Constraints, Chunks)

Spikes $\hat{s}_c \in \hat{S}$

Activations $\hat{a}_t \in \hat{A}$

CausalGroup spike equivalence classes $[\hat{s}_c]$.

4.3 The Transduction Operation

5 Conflict Resolution

5.1 Bayesian Specificity As State Utility

5.2 Causal Groups and Constraint Completion

5.3 Primary and Secondary Signals

6 Experiments

7 Conclusion

8 Future Work

9 References