AIDAS: Immersive Interaction within Vehicles

Ian Lane, Yi Ma, Vishwanath Raman, Antoine Raux, Yanjing Zhang

Background

Current driver support systems are unable to leverage **situational context** during interaction: i.e.. location, visible landmarks, status of driver, status of vehicle, external environment

Context key to support situated interaction within vehicles (i.e. context-aware navigation or situated information search). Robust belief-tracking critical to enable immersive interaction
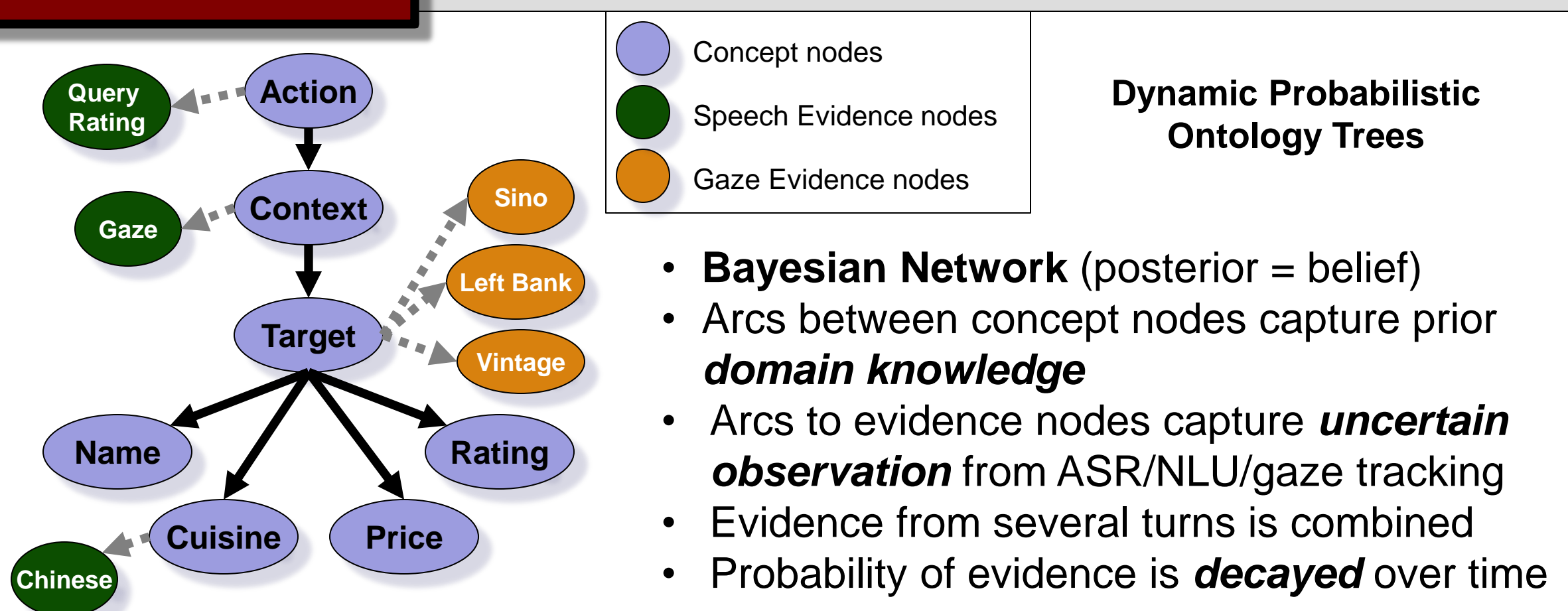
Architecture
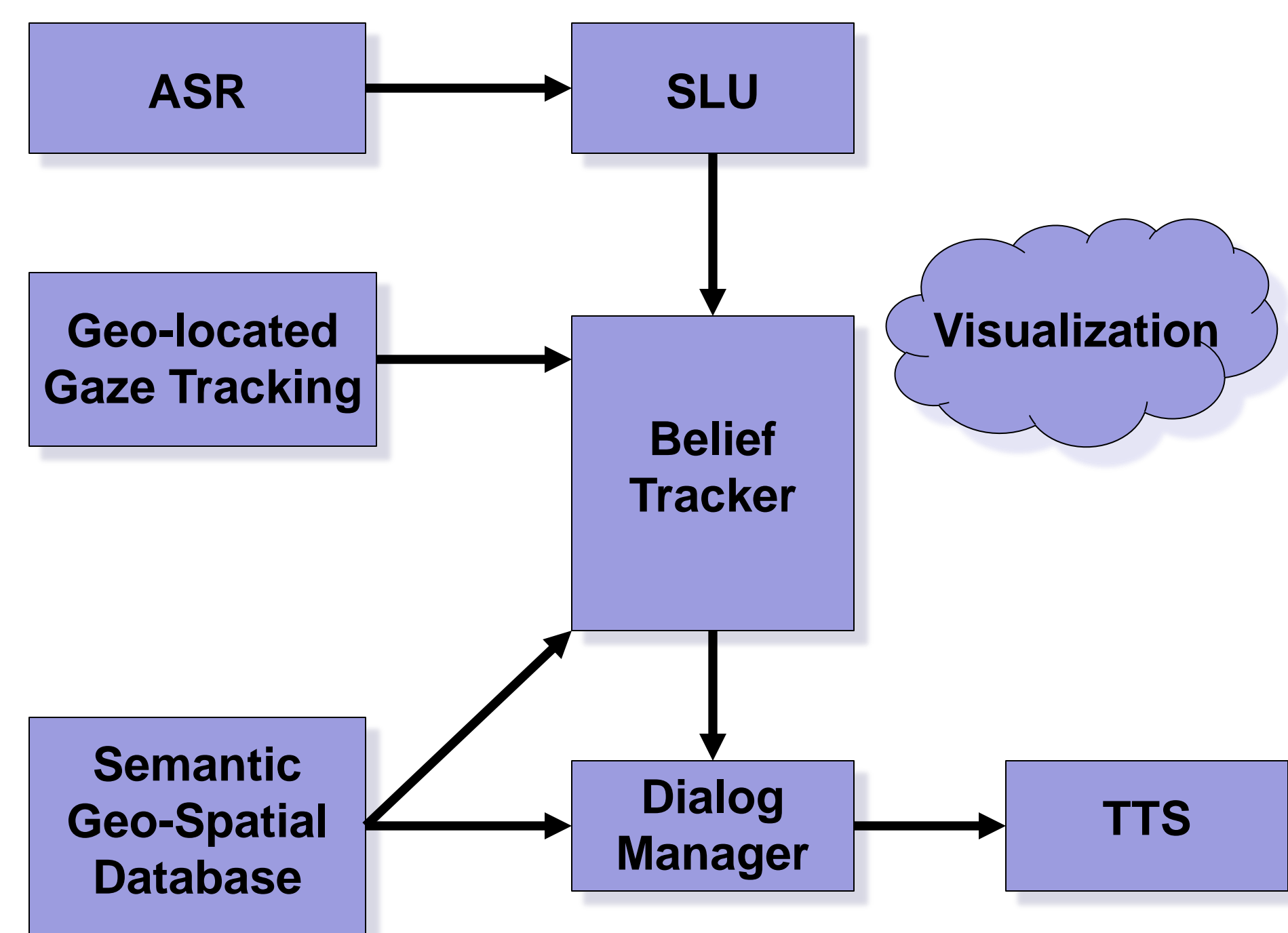
ASR: Automatic Speech Recognition
SLU: Spoken Language Understanding
TTS: Text To Speech Generation

Multimodal Situated Interaction

Sino: *"Chinese fusion restaurant. Good rating. Expensive prices."*

Left Bank: *"Popular American restaurant. Average Rating. Medium prices."*

Is that Chinese place over there any good?

Semantic Geo-Spatial Database

Speech Input

GPS and gaze information:

Geo-located Gaze Tracking

Agent Beliefs
Context: Gaze
Action: Query Rating
Cuisine: Chinese
Restaurants in Gaze:
Sino
Left Bank
Vintage Wine Bar

Target: Sino

Dialog Understanding via Belief Tracking

Reference Resolution

Is it expensive?

Dialog History

Agent Beliefs
Context: Dialog
Action: Price Rating
Last Restaurant Mentioned:
Sino

Target: Sino

Concept nodes
Speech Evidence nodes
Gaze Evidence nodes

Dynamic Probabilistic Ontology Trees

- **Bayesian Network** (posterior = belief)
- Arcs between concept nodes capture prior *domain knowledge*
- Arcs to evidence nodes capture *uncertain observation* from ASR/NLU/gaze tracking
- Evidence from several turns is combined
- Probability of evidence is *decayed* over time

Challenges and Future Work

- **Situated understanding of speech**
  - Is the target within driver's field of view?
- **Semantic grounding in the real world**
  - "What's that place with the yellow umbrellas?"
  - "You mean right after that red brick building?"

- **Multimodal interaction that is always "ON"**
  - Is the driver speaking to AIDAS?
  - When should AIDAS interrupt?
  - What is the driver doing?