TALK project work at UCAM

Matt Stuttle

August 2005

Cambridge University Engineering Department

Web:  http://mi.eng.cam.ac.uk/ mns25
Email: mns25@eng.cam.ac.uk
Overview

- **TALK project**
  - Goals
  - Baseline system

- **the Hidden Information State (tHIS) system**
  - Proposed system architecture
  - The Speech Understanding (SU) Agent
  - Confidence scores
  - HVS parser

- **Current and future work**
**TALK project:**
Tools for Ambient Linguistic Knowledge

**Project Goals:**

- Implement Information State Update (ISU) approaches
- Develop in-car and in-home multimodal dialog systems
- Learning strategies for the ISU approach
- Integrate multimodality and multilinguality
- Partners include Edinburgh, Gothenburg, DFKI, Saarbrucken, Seville, Bosch, Linguamatics and BMW
**TALK project:**

UCAM Project Workpackages:

- **WP1:** Unifying Multimodality and Multilinguality
  - New Speech Understanding components
  - Developing notion of multimodal dialog act
  - Closely coupling parser, recogniser and dialog manager

- **WP4:** Adaptivity and Learning
  - Reinforcement learning approaches
  - (Partially Observable) Markov Decision Process dialog managers
  - User modelling approaches
Baseline System:
Current baseline system built with Edinburgh:

- Based on SACTI Tourist Information domain

- Built using:
  - DIPPER dialog manager
  - Open Agent Architecture
  - ATK speech recognition using n-gram language models
  - Word spotting parser

- Currently features:
  - Implicit and explicit confirmation
  - Fragmentary clarifications
  - Simple fixed dialog plan (slot filling)
  - Overanswering
  - Multimodal output
Proposed HVS System Overview

System Architecture

Four agents:
- Speech Understanding
- Dialogue manager
- Turn Manager
- Speech Generation
Proposed HVS System Overview

Hidden Information State model

- Operates as a MDP over Belief State $B = P(S_d, S_u)$ where $S_u =$ user state and $S_d =$ dialog state

- Similar advantages to POMDP for modelling uncertainty but like an MDP can be optimised using on-line methods

- Belief state is discrete approximation to continuous state space

- Approximation is a *forest* of most likely states

- User state expands according to *application specific* rules

- User state captures domain ontology

- Actions $A_s$ are based on a linear function approximation of $Q(B, A_s)$
Some of UCAM’s proposed work focuses on developing the SU agent.

- Uses ATK and the Hidden Vector State (HVS) Parser.
- Outputs an n-best list of user dialog acts with scores.
- Operates asynchronously, forwards dialog acts to the Turn Manager agent.
Speech Understanding (SU) Agent

Motivation

- Asynchronous, so does not impose a turn-taking model - will allow barge-in.
- Will output the n-best dialog acts
- Use of HVS and ATK will allow statistical confidence measures to be implemented
- HVS model can be incrementally trained on-line with system data.
- Multimodal inputs can be investigated later (WP1 work)
Confidence Scores

Current implementation:

- ATK currently outputs confidence scores based on recognition scores relative to a “background” model
  - Current scores do not contain much information about correctness of hypothesis

- Recent work has enabled use of n-best lists and lattice generation in ATK
  - Need to generate very long n-best lists to get alternative parse hypotheses.
Confidence Scores
Proposed alternative:

- Generate word posteriors from lattice for each word position
  - Will generate more hypotheses from parser
  - Deletions/insertions handled by “NULL” label

- Can generate alternatives for each word in the best hypothesis
  - Word posteriors have been shown to be more informative
  - Can develop strategies based on fragmentary clarification more easily

- Currently working on generating information on-the-fly
  - Calculating word posteriors requires forward and backward pass through the (complete) lattice
ATK Outputs

I see that’s nearer -21.2
is that nearer -46.1
i see that’s near here -55.4
is the theatre -59.1
i need the theatre -61.6

N-Best

Confidence Network
The Hidden Vector State (HVS) model for Semantic Parsing

Overview

- Trade-off between flat-concept and hierarchical models

- Flat-concept HMM model extended by expanding each state

- States encode stack of push-down automata

- Model is right branching
  - encodes hierarchical context
  - avoids tractability issues of a PCFG approach
Parse tree of semantic concepts is right branching

Tree is represented in vector state sequence below
The Hidden Vector State (HVS) model for Semantic Parsing

Motivation

- HVS is a statistical semantic parser
  - Fits well into UCAM’s plan for a statistical dialogue modelling approach
  - Gives n-best acts together with probability/confidence scores

- HVS model is data-driven
  - allows on-line/incremental adaptation/training
  - requires only lightly supervised data

- HVS can be adapted to work on-the-fly
  - will allow barge-in
  - can use more complicated turn-manager
Training and Parsing with the Hidden Vector State Model

- Domain specific lexical classes
- Abstract semantic annotation for each utterance

HVS-Model

I want to return to Dallas on Thursday

HVS-Parser

Diagram of HVS-Parser with lexical classes and semantic annotations.
Current work

- Training HVS and word-spotting model on SACTI data for subset of tasks.
- Encapsulate HVS parser as an ATK component
- Make speech understanding (ATK+HVS) OAA agent
- Modify ATK to produce word-posterior networks with associated confidence scores
Future plans (SU) Agent

- Combine ATK and HVS probabilities for a parser confidence score
- Adapt HVS approach to work with this data format
- Current parser results using HVS and word-spotting techniques are relatively poor
  - SACTI data is not ideal: will perform a new data collection to bootstrap parser and language models
  - Collection will feature abstract prompts of all features desired in final system
  - May also do some collections with baseline system
  - Plan to use a real town (Cambridge?) for new data collection with a much larger database extracted from Google maps/Yell.com