# Dialogue management: discriminative approaches to belief tracking

Milica Gašić

Dialogue Systems Group, Cambridge University Engineering Department

February 4, 2016

Discriminative models for belief tracking

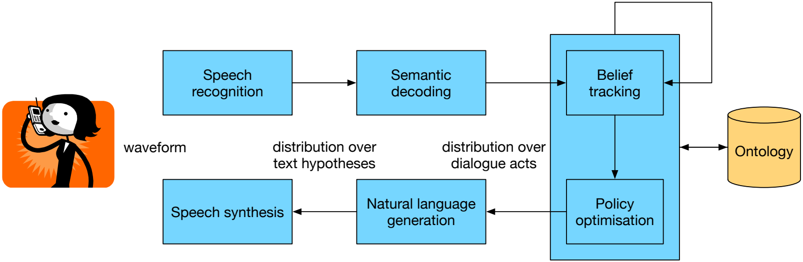
Ranking models

Deep neural network approaches to belief tracking

Recurrent neural network approaches to belief tracking

Integrated approaches to semantic decoding and belief tracking

# Dialogue management



# Generative vs discriminative models in belief tracking

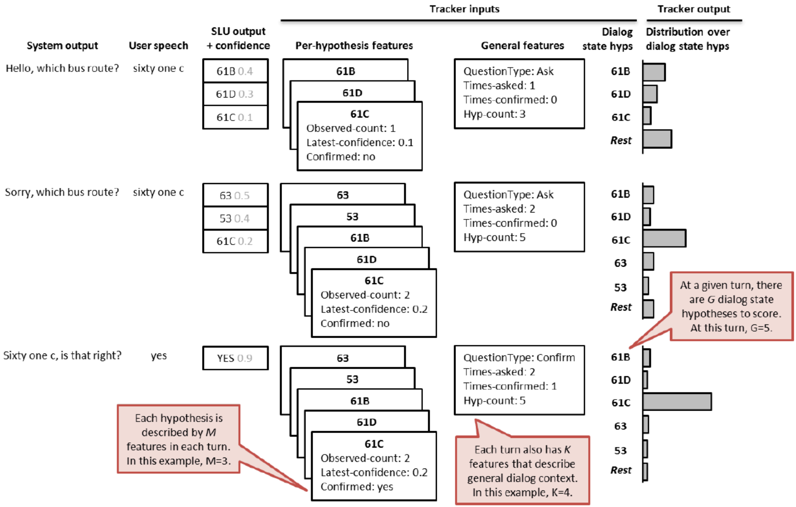
Discriminative models: the state depends on the observation

$$b(s_t) = p(s_t|o_t)$$

Generative models: the state generates the observation

$$b(s_t) = \frac{p(s_t, o_t)}{\sum_{s_t} p(s_t, o_t)} \propto p(o_t|s_t)p(s_t)$$

# Advantage of discriminative belief tracking [Metallinou et al., 2013]



# Problems in generative belief tracking

- Generative models make assumption that observations at each turn are independent
- Discriminative models directly model the dialogue state given arbitrary and possibly correlated input features.

# Dialogue state tracking challenge (DSTC) problem formulation

Common dataset with tools to evaluate the performance of the tracker. The dialogue state consists of three components:

- **goal** for each informable slot, e.g. pricerange=cheap.
- **requested** slots by the user, e.g. phone-number.
- **method** of search for the entities, e.g. *by constraints*, *by alternatives*, *by name*.

The belief state is then the distribution over possible slot-value pairs for goals, the distribution over possible requested slots and the distribution over possible methods.

# Evaluate the quality of the belief state tracker

Accuracy: the fraction of turns where the top dialogue state hypothesis is correct

L2 norm: is squared L2-norm of the hypothesised distribution **p** and the true label

$$L2 = (1 - p_i)^2 + \sum_{j \neq i} p_j^2$$

where $p_i$ is the probability assigned to the true label.

# Focus tracker

The focus tracker accumulates the evidence and changes the focus of attention according to the current observation.

$$b(s_t = s) = o(s) + (1 - \sum_{s' \in S} o(s'))b(s_{t-1} = s)$$

# Class-based approaches to dialogue state tracking

Model the conditional probability distribution of dialogue state given all observations upto that turn in dialogue.

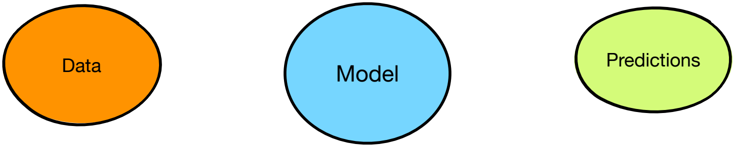
$$b(s_t) = p(s_t | o_0, \cdots, o_t)$$

Features are extracted from $o_0, \cdots, o_t$ and include information about

- latest turn
- dialogue history
- ASR errors

This allows a number of models to be used: maximum entropy linear classifiers, neural networks and ranking models.

# Class-based approaches to dialogue state tracking

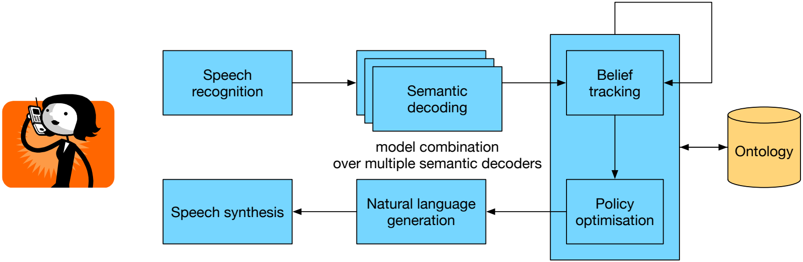



- Observations labelled with dialogue states

- Neural networks
- Ranking models

- Distribution over possible dialogue states – **belief state**

# Dialogue management with multiple semantic decoders



# Ranking approach to dialogue state tracking

Dialogue state tracking of the user goal consists of the following three steps

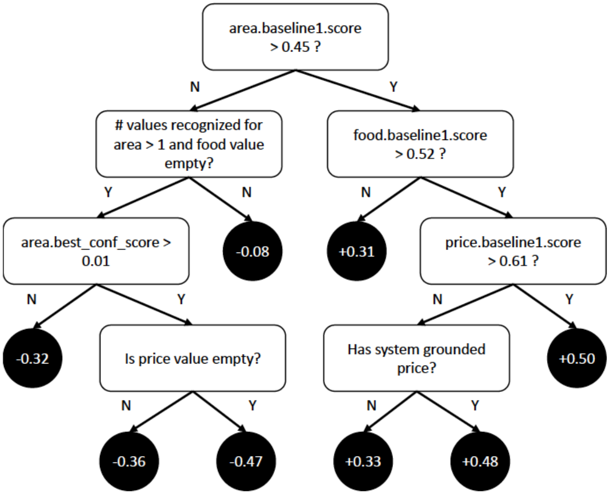
- Enumerate possible dialogue states
- Extract features
- Scoring

Using multiple semantic decoders trained on different datasets can produce a richer set of possible dialogue states.

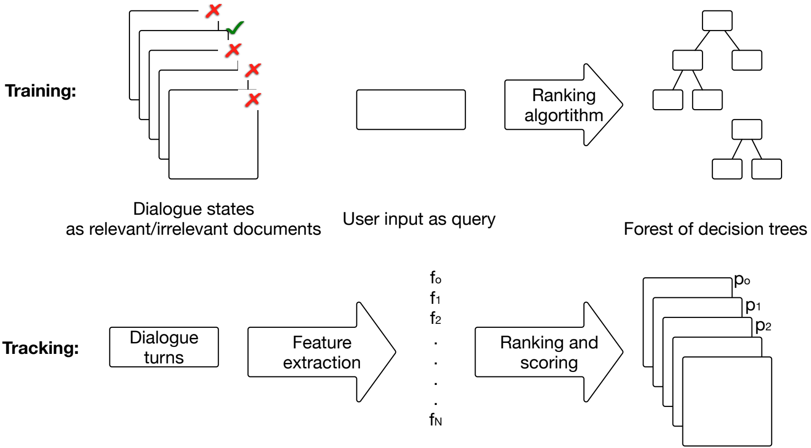
# Theory: Decision trees

- For a set of input data points $\mathbf{x}_1, \cdots, \mathbf{x}_N$ and target values $t_1, \cdots, t_N$ find partitioning of the input space and the set of questions so that the sum-of-squares (in the regression case) or the cross entropy (in the classification case) is minimal.
- Random forests are a way of averaging multiple decision trees trained on different parts of the same training set.

# Example decision tree for belief tracking [Williams, 2014]



# Web-style ranking [Williams, 2014]



# Theory: Deep neural networks

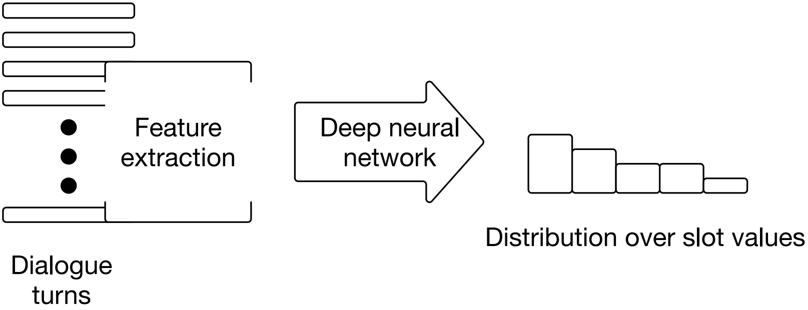
$$\mathbf{h}_0 = g_0(W_0\mathbf{x}^\mathsf{T} + b_0)$$
$$\mathbf{h}_i = g_i(W_i\mathbf{h}_{i-1}^\mathsf{T} + b_i), 0 < i < m$$
$$\mathbf{y} = \text{softmax}(W_m\mathbf{h}_{m-1}^\mathsf{T} + b_m)$$
$$\text{softmax}(\mathbf{h})_i = \exp(h_i)/(\sum_j \exp(h_j))$$

where

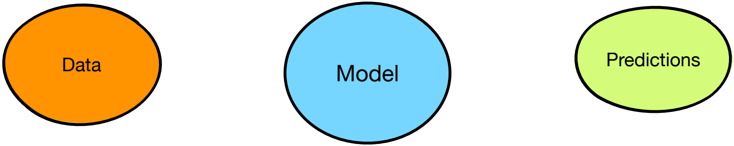
- $g_i$ (differentiable) activation functions hyperbolic tangent tanh or sigmoid $\sigma$
- $W_i, b_i$ parameters to be estimated

# Deep neural networks for belief tracking [Henderson et al., 2013]

- Outputs a sequence of probability distributions over an arbitrary number of possible values
- Learns tied weights using a single neural network
- Uses a form of sliding window for feature extraction



# Sequence-to-sequence approaches to dialogue state tracking



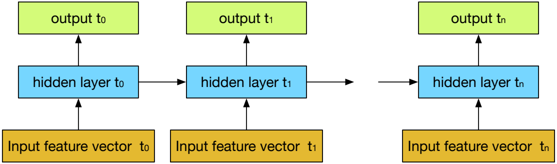

- Sequence of observations labelled with dialogue states

- Recurrent neural networks

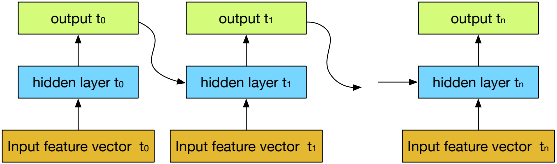
- Distribution over possible dialogue states – **belief state**

# Theory: Recurrent neural networks

## Elman-type



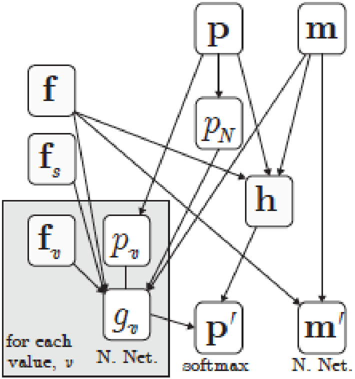

## Jordan-type



# Recurrent neural network based belief tracking [Henderson, 2015]

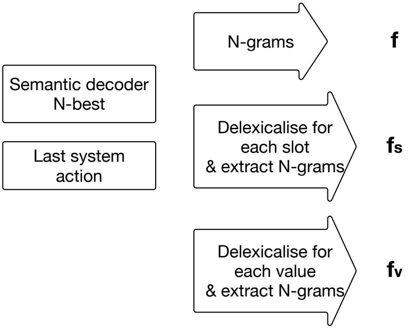
- Contains internal memory which represents dialogue context
- Structurally a combination of Elman and Jordan types
- Takes the most recent dialogue turn and last machine dialogue act as input, updates its internal memory and calculates distribution over slot values.

# RNN structure




- **f** slot independent features, $\mathbf{f}_s$ are slot dependent features and $\mathbf{f}_v$ are value dependent features
- **m** is the internal memory from the previous time step and **m**′ is the memory in the next step
- **p** is the distribution over slot value pairs from the previous time step and **p**′ is the estimated distribution
- **h** and $\mathbf{g}_v$ are estimated with Neural network with one hidden layer and sigmoid activation function

# Feature engineering



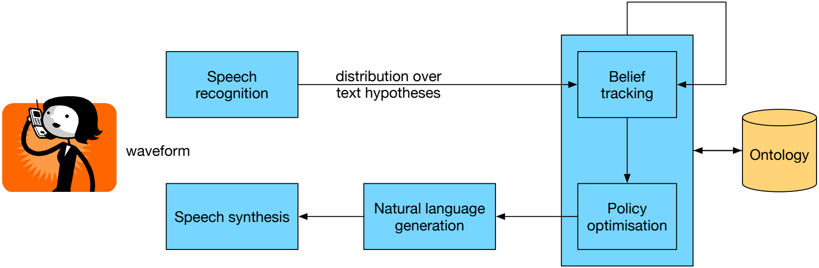

- For the same input feature vectors will be different for different slots and values
- These inputs then query different recurrent neural networks to produce distribution over slot value pairs

# Results from dialogue state tracking challenge

Taking into account only semantic decoding features:

| | Goals | | Method | | Requested | |
|---|---|---|---|---|---|---|
| | Acc. | L2 | Acc. | L2 | Acc. | L2 |
| Focus | 0.719 | 0.464 | 0.867 | 0.210 | 0.879 | 0.206 |
| RNN | 0.742 | 0.387 | 0.922 | 0.124 | 0.957 | 0.069 |
| Web-style ranking | 0.775 | 0.758 | 0.944 | 0.092 | 0.954 | 0.073 |

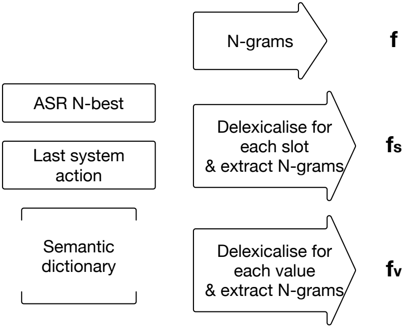
# Alternative dialogue system architecture



# Integrated approaches to semantic decoding and belief tracking [Henderson et al., 2014]

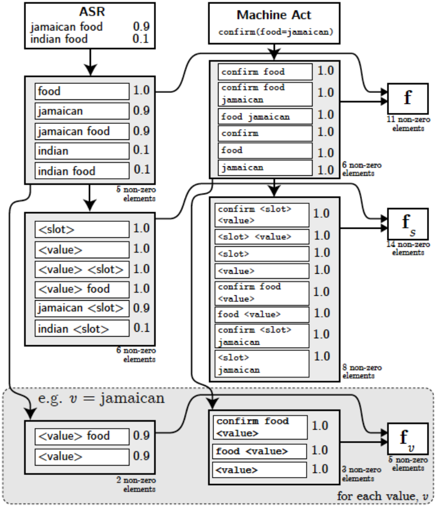
- Instead of extracting features from semantic decoding hypotheses extract features from ASR hypotheses
- Apply the same neural network structure
- Avoids information loss resulting from compact semantic representation of traditional approach
- Output: distribution over slot-value pairs

# Feature extraction from ASR hypotheses




- For limited vocabulary dialogue system possible to extract N-gram features from ASR
- In order to deal with data sparsity need to delexicalise input
- Unlike for semantic decoding output, here it is not obvious which word corresponds to which slot and value
- Semantic dictionary is therefore needed to define possible values

# Example input features



# Results from dialogue state tracking challenge

Taking into account only semantic decoding features:

| | Goals | | Method | | Requested | |
|---|---|---|---|---|---|---|
| | Acc. | L2 | Acc. | L2 | Acc. | L2 |
| Focus | 0.719 | 0.464 | 0.867 | 0.210 | 0.879 | 0.206 |
| RNN | 0.742 | 0.387 | 0.922 | 0.124 | 0.957 | 0.069 |
| Web-style ranking | 0.775 | 0.758 | 0.944 | 0.092 | 0.954 | 0.073 |

Taking into account only ASR features:

| | Goals | | Method | | Requested | |
|---|---|---|---|---|---|---|
| | Acc. | L2 | Acc. | L2 | Acc. | L2 |
| RNN | 0.768 | 0.346 | 0.940 | 0.095 | 0.978 | 0.035 |

# Delexicalisation - elephant in the room

- Most of the performance gain comes from delexicalised features
- This requires a separate semantic dictionary which for all values from ontology defines their possible realisations, for example expensive $\rightarrow$ luxurious, upmarket, pricey
- In real systems this poses a major problem

# References I


Henderson, M. (2015).
*Discriminative Methods for Statistical Spoken Dialogue Systems*.
PhD thesis, University of Cambridge.

Henderson, M., Thomson, B., and Young, S. J. (2013).
Deep Neural Network Approach for the Dialog State Tracking Challenge.
In *Proceedings of SIGdial*.

Henderson, M., Thomson, B., and Young, S. J. (2014).
Word-based Dialog State Tracking with Recurrent Neural Networks.
In *Proceedings of SIGdial*.

# References II


Metallinou, A., Bohus, D., and Williams, J. D. (2013).
Discriminative state tracking for spoken dialog systems.
In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL), Sofia, Bulgaria*. Association for Computational Linguistics.

Williams, J. D. (2014).
Web-style ranking and slu combination for dialog state tracking.
In *Proceedings of SIGDIAL*. ACL Association for Computational Linguistics.