# Multi-Domain Adversarial Learning for Slot Filling in Spoken Language Understanding

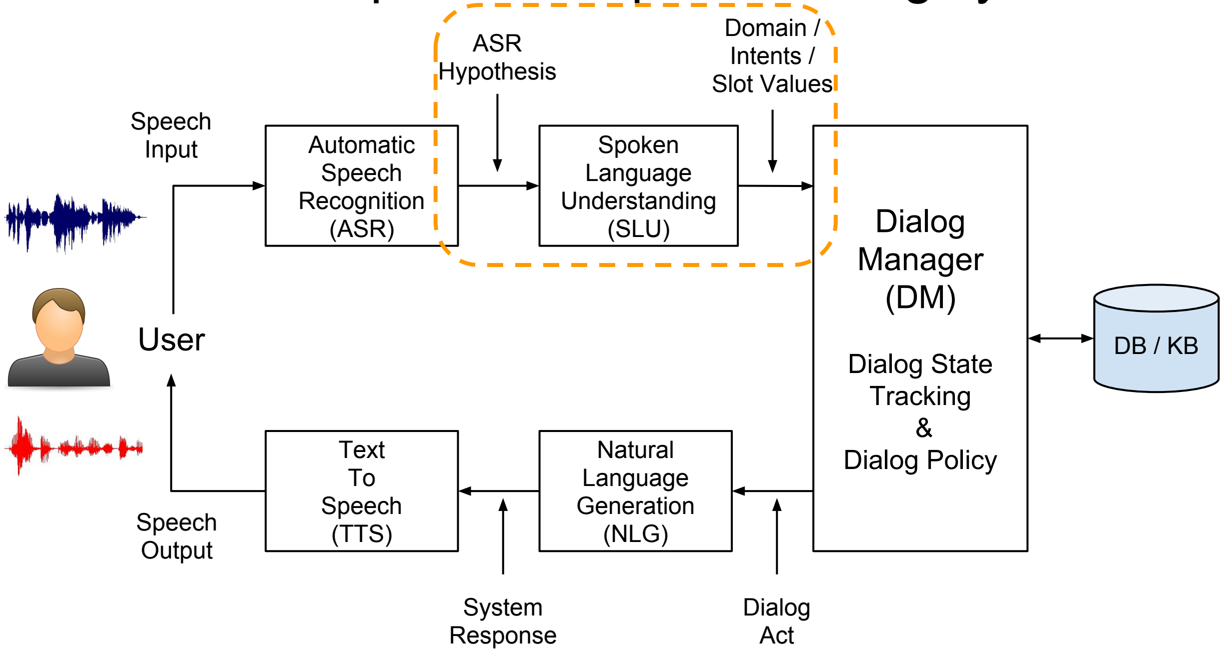
***Bing Liu***, *Ian Lane*

Carnegie Mellon University

liubing@cmu.edu, lane@cmu.edu

# Spoken Language Understanding (SLU)

- SLU is a critical component in spoken dialog systems



# SLU Tasks

1. Domain Classification
2. Intent Determination
3. Slot Filling

User Utterance:

Show me flights from Pittsburgh to Long Beach on Sunday

SLU Outputs:

Domain: Air_Travel
Intent: Show_Flight

Slots:
Departure_City: Pittsburgh
Departure_Date: Sunday
Departure_Time: <to_be_filled>
Destination_City: Long Beach
...

# Slot Filling in SLU

- Slot filling as a sequence labeling problem

| **Utterance** | Flights | from | Pittsburgh | to | Long | Beach | on | Sunday |
|---|---|---|---|---|---|---|---|---|
| **Slot Label** | O | O | B-Dept_city | O | B-Dest_city | I-Dest_city | O | B-Dept_date |

  - Given an utterance $\mathbf{w} = (w_1, w_2, \ldots, w_T)$, find the best sequence of slot labels $\mathbf{y} = (y_1, y_2, \ldots, y_T)$, one for each word in the utterance, such that:

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} P(\mathbf{y}|\mathbf{w}, \theta)$$

  - Popular methods for slot filling: MEMM, CRF, RNNs.

# Current Slot Filling Models

- Current slot filling models are mostly domain-specific
  - Trained and work on individual task domains.
- Hard to transfer knowledge across domains for slot filling
- Costly in annotating semantic tags for each new task domain

# Motivations

*Can we learn common features and representations that can be shared across multiple domains for slot filling in SLU?*

- Benefits:
  - Reduce the amount of annotated data required for developing a new domain
  - Improve slot filling performance with an ensemble of domain-general and domain-specific models

# Learning Shared Representations

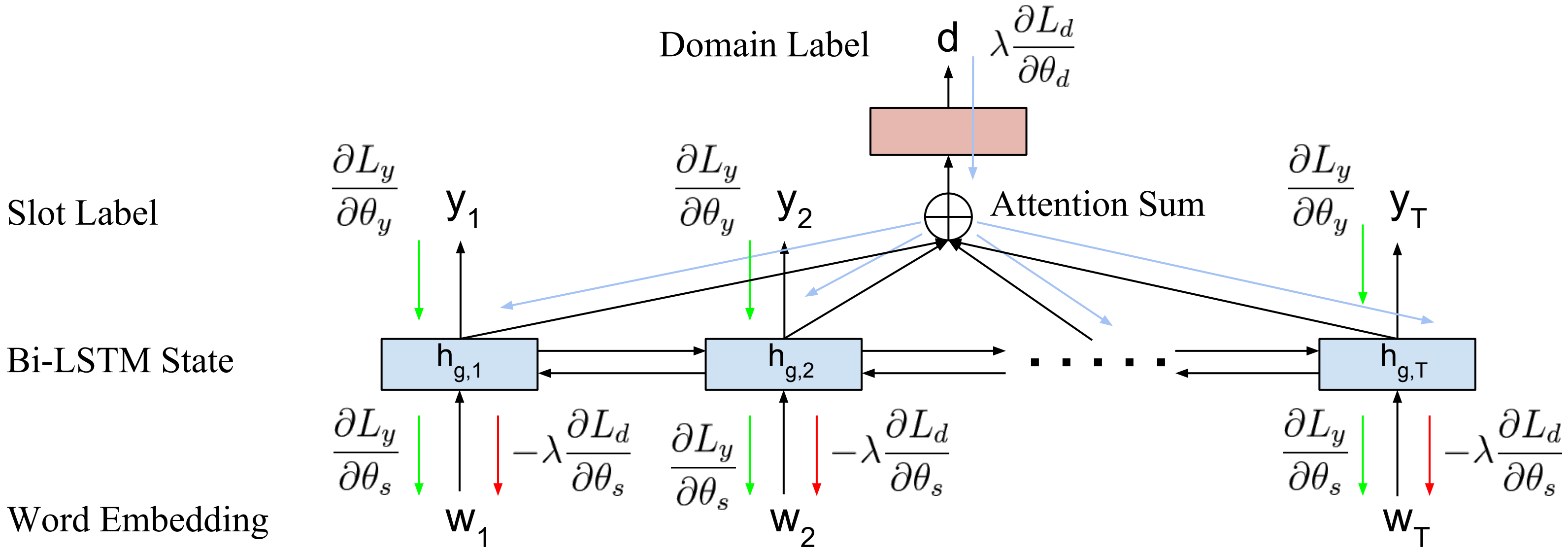
- Naive approach: train a single slot filling model directly on a union of the data from all domains
  - May still learn disjoint domain-specific features due to the very different data distributions in different domains
- Good common representations across domains are the ones based on which a system cannot recover the domain of the original inputs [1]
- Our proposal: Apply domain adversarial learning in training the domain-general slot filling model

[1] Ben-David, Shai, et al. "A theory of learning from different domains." *Machine learning* 79.1 (2010): 151-175.

# Proposed Method

- Model slot filling with bidirectional LSTM (bi-LSTM)
- Train domain-general model using a union of data from all domains, with an additional domain adversarial loss
  - Enforce the bi-LSTM model to learn common representations across domains to “fool” a domain classifier
- Train domain-specific models using individual domain data
- Combine domain-general and domain-specific models at output layer via a multi-layer perceptron (MLP) for slot filling

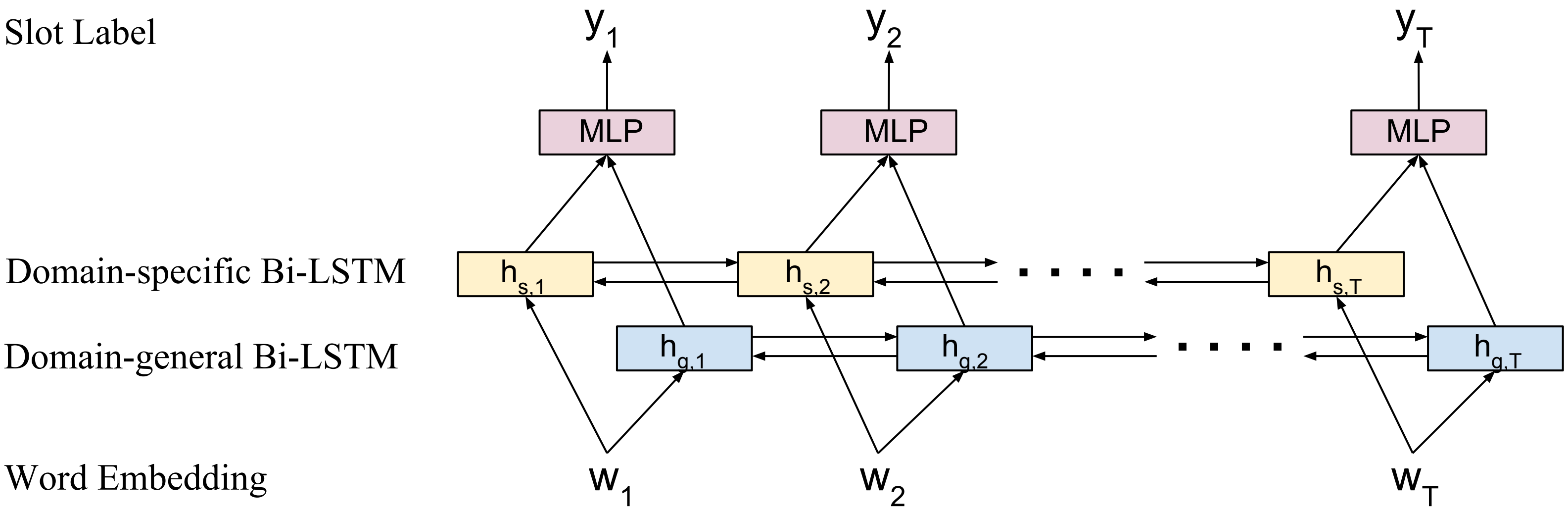
# Domain Adversarial Training



# Domain Adversarial Training

- Model Parameters:
  - Slot filling output MLP: $\theta_y$
  - Domain classification output MLP: $\theta_d$
  - Word embedding & bi-LSTM: $\theta_s$
- Losses:
  - Slot Filling Loss: $$\min_{\theta_s, \theta_y} L_y = \min_{\theta_s, \theta_y} -\frac{1}{T} \sum_{t=0}^{T} \log P(y_t^* | \mathbf{w}; \theta_s, \theta_y)$$
  - Domain Adversarial Loss: $$\max_{\theta_s} \min_{\theta_d} L_d = \max_{\theta_s} \min_{\theta_d} - \log P(d^* | \mathbf{w}; \theta_s, \theta_d)$$
  - Total Loss: $$L = L_y + \lambda L_d$$

# Joint Model Training



# Experiments

## ❖ Data sets

- ➢ ATIS (Airline Travel Information Systems): Air travel query
- ➢ MIT Restaurant Corpus[1]: Restaurant query and search
- ➢ MIT Movie Corpus (eng & trivia10k13): Movie query and search

| Datasets | ATIS | MIT Rest. | MIT Mov. eng | MIT Mov. trivia10k13 | Combined |
|---|---|---|---|---|---|
| **Train set size** | 4978 | 7660 | 9775 | 7816 | 30229 |
| **Test set size** | 893 | 1521 | 2443 | 1953 | 6810 |
| **Vocab size** | 572 | 4166 | 7481 | 12145 | 16049 |
| **Slot label size** | 127 | 17 | 25 | 25 | 191 |

[1]The MIT SLU corpora can be downloaded from: https://groups.csail.mit.edu/sls/downloads

# Experiments

- Training Settings
  - LSTM state and output size: 128
  - Output layer MLP size: 200
  - Word embedding size: 128
    - Randomly initialized and fine tuned
  - Dropout: p = 0.5
  - Optimizer: Adam (initial learning rate = 1e-03)
- Evaluation Metrics
  - Slot filling F1 score (harmonic mean of precision and recall)

# Experiment Results

- Domain-specific and domain-general model performance, comparing to published results

| Model | ATIS | MIT Rest. | MIT Mov. eng | MIT Mov. trivia10k13 | Comb. |
|---|---|---|---|---|---|
| Deep LSTM [16] | 95.08 | - | - | - | - |
| RNN-EM [17] | 95.25 | - | - | - | - |
| Encoder-labeler LSTM [18] | 95.40 | - | - | - | 74.41 |
| Attention Bi-LSTM [6] | 95.75 | - | - | - | - |
| BLSTM-LSTM (focus) [19] | **95.79** | - | - | - | - |
| Dom-Spec | 95.55 | 72.42 | **83.43** | **63.64** | - |
| Dom-Gen | 94.09 | **74.25** | 82.95 | 63.34 | 76.03 |

* Dom-Spec: domain-specific Bi-LSTM

* Dom-Gen: domain-general Bi-LSTM (without adversarial learning)

# Experiment Results

- Joint domain-specific and domain-general model performance

| Model | ATIS | MIT Rest. | MIT Mov. eng | MIT Mov. trivia10k13 | Comb. |
|---|---|---|---|---|---|
| Dom-Spec | 95.55 | 72.42 | 83.43 | 63.64 | - |
| Dom-Gen | 94.09 | **74.25** | 82.95 | 63.34 | 76.03 |
| Dom-Gen-Adv ($\lambda$=0.01) | **94.51** | 73.87 | **83.03** | **63.51** | **76.55** |
| Dom-Gen-Adv ($\lambda$=0.1) | 93.88 | 73.98 | 82.31 | 62.83 | 76.01 |
| Dom-Gen-Adv ($\lambda$=1.0) | 84.65 | 62.47 | 75.05 | 52.82 | 66.66 |
| Joint Dom Spec & Gen | 95.62 | **74.47** | 84.87 | 65.16 | - |
| Joint Dom Spec & Gen-Adv ($\lambda$=0.01) | **95.63** | 74.23 | **85.33** | **65.33** | - |
| Joint Dom Spec & Gen-Adv ($\lambda$=0.1) | 95.52 | 74.36 | 85.32 | 64.95 | - |
| Joint Dom Spec & Gen-Adv ($\lambda$=1.0) | 95.52 | 73.57 | 84.26 | 64.38 | - |

* Dom-Spec: domain-specific bi-LSTM

* Dom-Gen: domain-general bi-LSTM

* Dom-Gen-Adv: domain-general bi-LSTM with adversarial learning

# Conclusions

- ❖ We propose applying domain adversarial training in learning cross-domain common features and representations for slot filling task in SLU
- ❖ We show the benefits of applying domain adversarial learning in achieving advanced slot filling F1 scores.
- ❖ Future directions
  - ➢ Perform adversarial learning with sequence level optimization on slot labels (e.g. by adding a CRF layer on top)
  - ➢ Extend SLU model to end-to-end dialogue modeling (poster #7)

# Thanks!