Dialog State Tracking for Interview Coaching Using Two-Level LSTM

Ming-Hsiang Su , Chung-Hsien Wu, Kun-Yi Huang, Tsung-Hsien Yang, and Tsui-Ching Huang

Department of Computer Science and Information Engineering
National Cheng Kung University, Taiwan

Outline

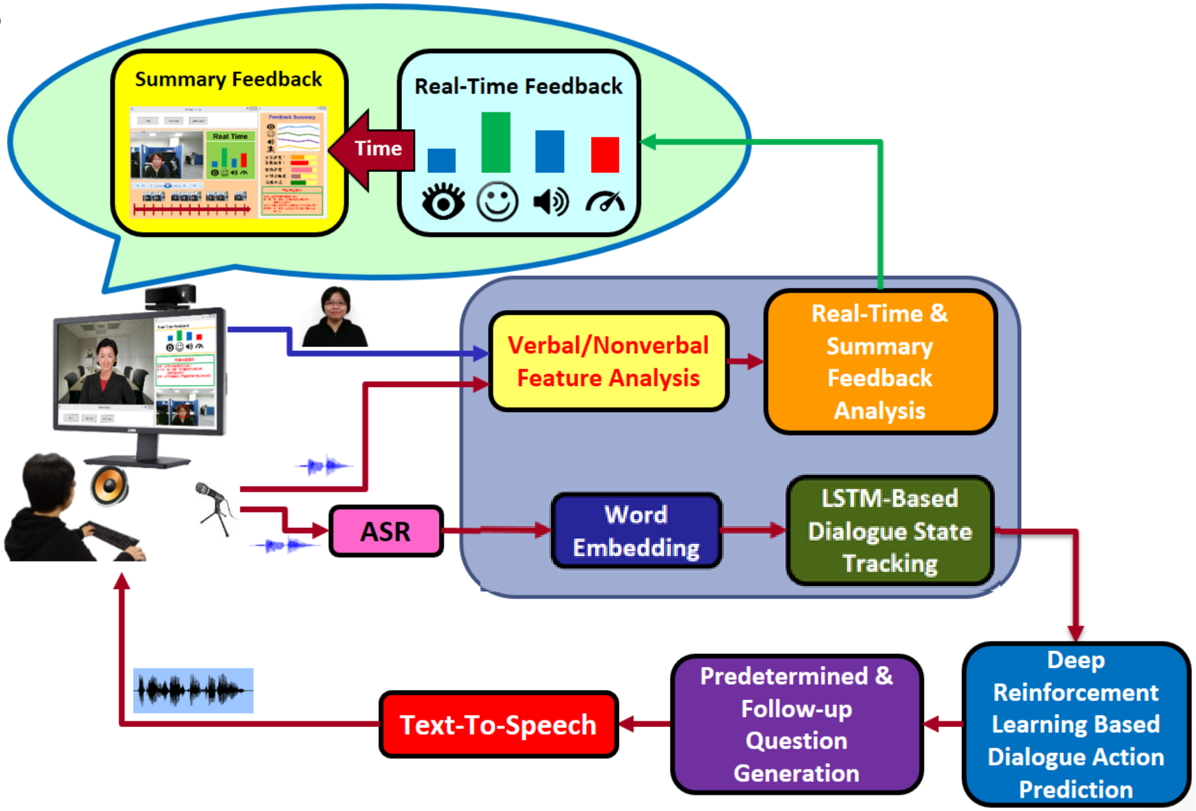
- Introduction
  - Motivation and goal
- Data collection and annotation
- System framework
  - Word embedding
  - LSTM for answer vector extraction
  - ANN for dialog state detection
- Experimental results
- Conclusions

Introduction

- While in school, students are busy in studying for a better future.
  - Students rarely have the opportunity to practice interview for work or study.
- If students are afraid or nervous during the interview,
  - they can not answer questions asked by the interviewers properly.

Motivation and Goal

- For convenient and inexpensive opportunities to practice interview,
  - a coaching system is constructed to improve user's interview skills

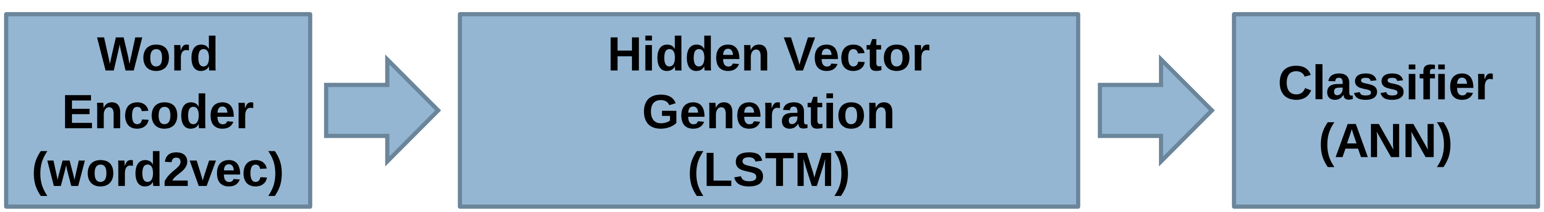


Motivation and Goal

- How to develop a **dialog state tracker (DST)** for user goal inference in an interview coaching system.
  - DST is one of the key sub-tasks of dialog management.
  - DST tasks are crucial because the dialog policy needs the DST to **detect correct dialog state** in order to choose an appropriate action.

Introduction (3)

- In this study, we proposed an **encoder-classifier** model.
  - The word2vec method is used to encode words into word vectors.
  - A two-level LSTM-based method is proposed to encode the answer sentences into answer hidden vectors.
    - One for generating sentences hidden vectors.
    - One for generating answer hidden vectors.
  - The ANN-based method is used to predict dialog states.



Data collection and annotation

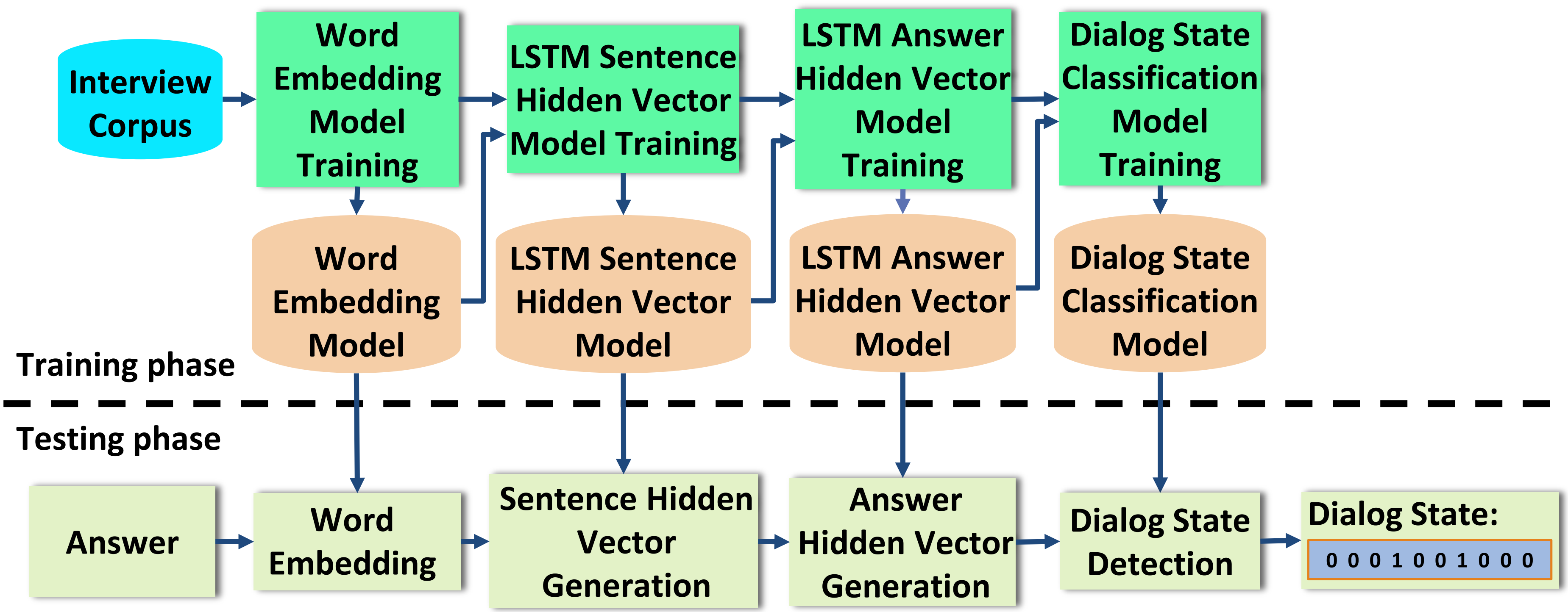
- Interview dialog corpus collection:
  - The domain of the corpus is chosen as the **College Admission Interview**.
  - **12** participants were invited.
  - During corpus collection, two participants completed the interview without prior design questions and answers.
  - A total of **75** dialogs with **540** question and answer pairs were collected.
  - Average number of sentences for each answer is **3.95**.

Data collection and annotation

- According to the collected corpus,
  - **10** semantic slots were defined.

Semantic Slot	Number
Association and cadre leadership	50
Performance and achievement	57
Leisure interest	50
Pros. and cons.	45
Motivation	54
Study and future plan	49
Curriculum areas	143
Programming language and professional skills	38
Personality trait	136
Others	12

System Framework

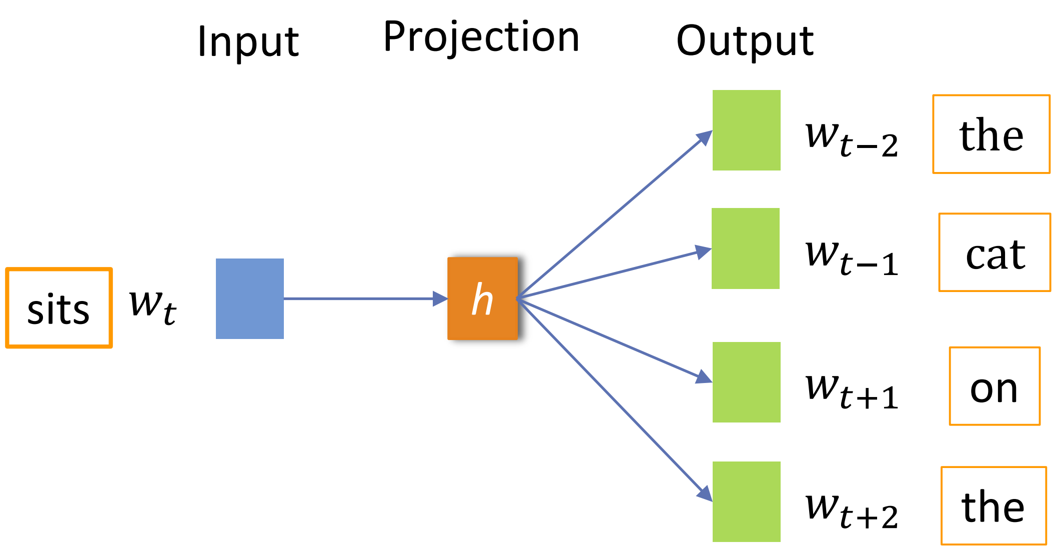


Word embedding model

- Each word is mapped to its corresponding **word vector** $\boldsymbol{w_i}$ by using word2vec.
  - Word2vec creates **vectors** that are distributed numerical representations of word features, such as the context of individual words.
  - The purpose and usefulness of word2vec is **to group the vectors of similar words together in the vector space**.
  - Word2vec encodes each word in a vector and trains words against other words that neighbor them in the input corpus.

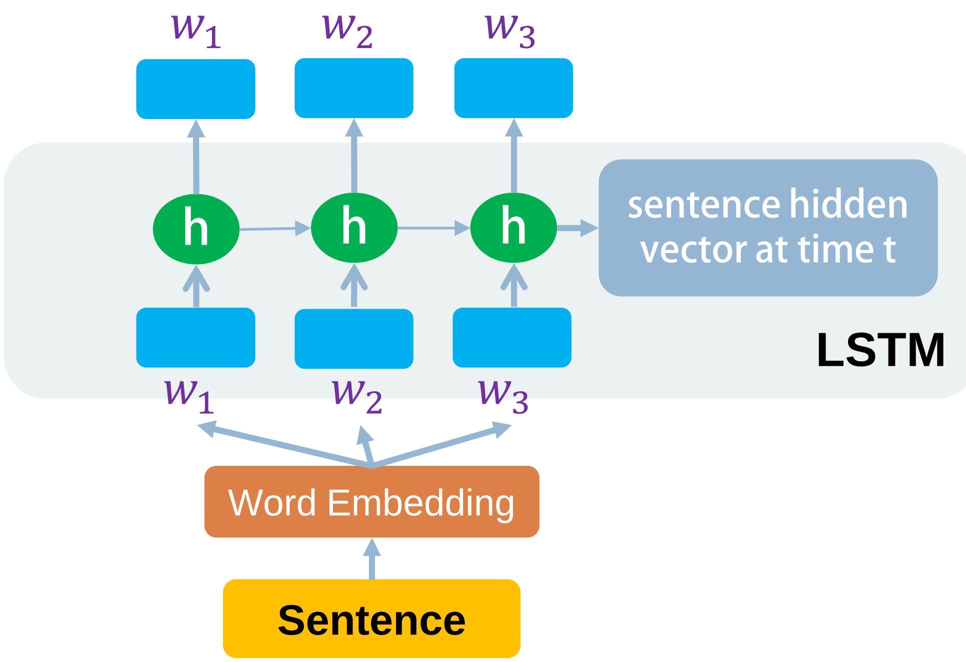
Word embedding model

- This work uses the Skip-gram model.
  - The word vector is connected to the vector representation of the sentence.
  - We use Chinese Gigaword corpus to train word2vector model.
  - Totally, 42619 words were obtained.



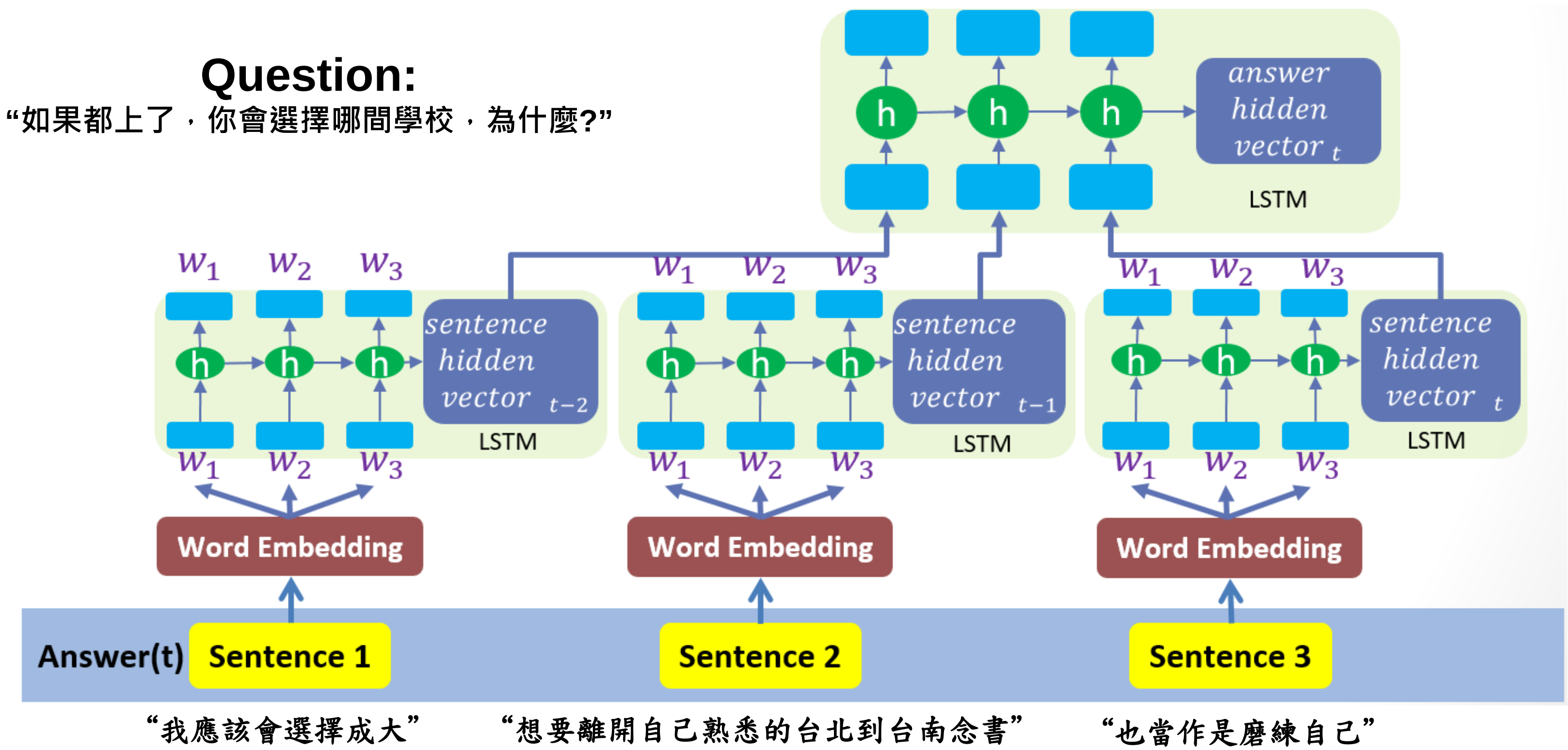
Dialogue State Tracking

- The **semantic relationship** between consecutive sentences in an answer is important for state detection,
  - the fist LSTM is employed to combine word vectors of a sentence to obtain the hidden vector of the sentence.



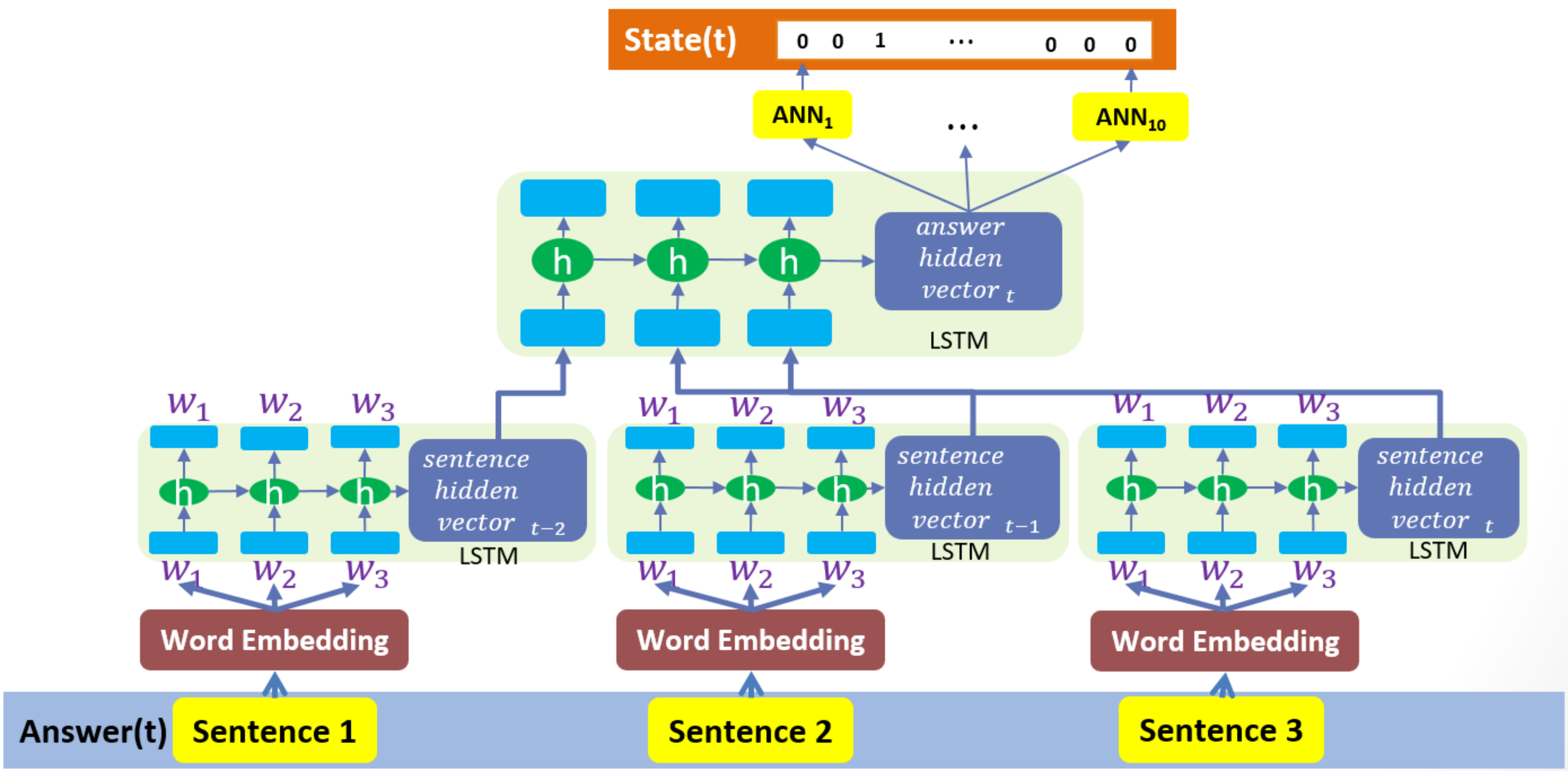
Dialogue State Tracking

- The second LSTM is employed to combine sentence hidden vectors to obtain the hidden vector of the answer.



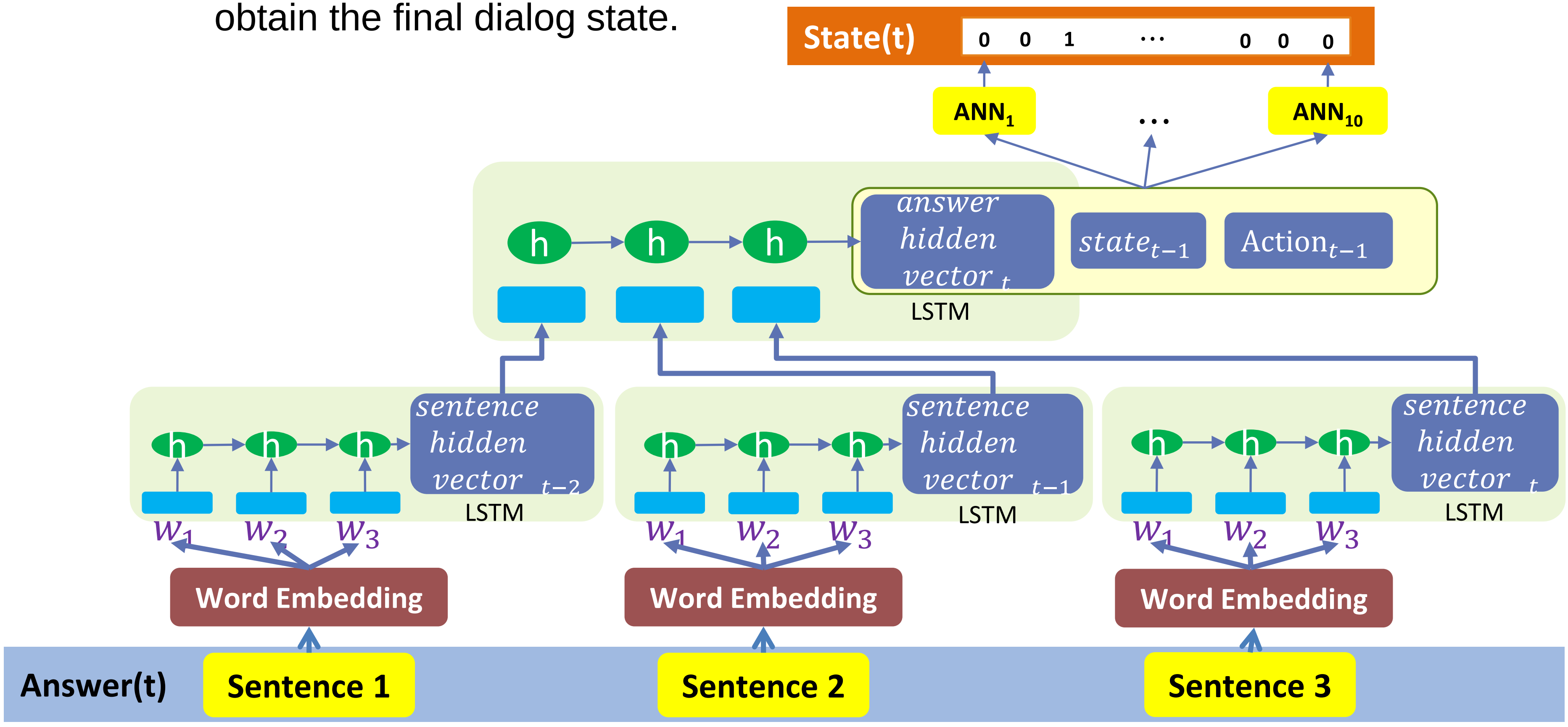
Dialogue State Tracking

- Finally, the answer hidden vector is fed into an ANN to detect the **dialog states** for dialog state representation.



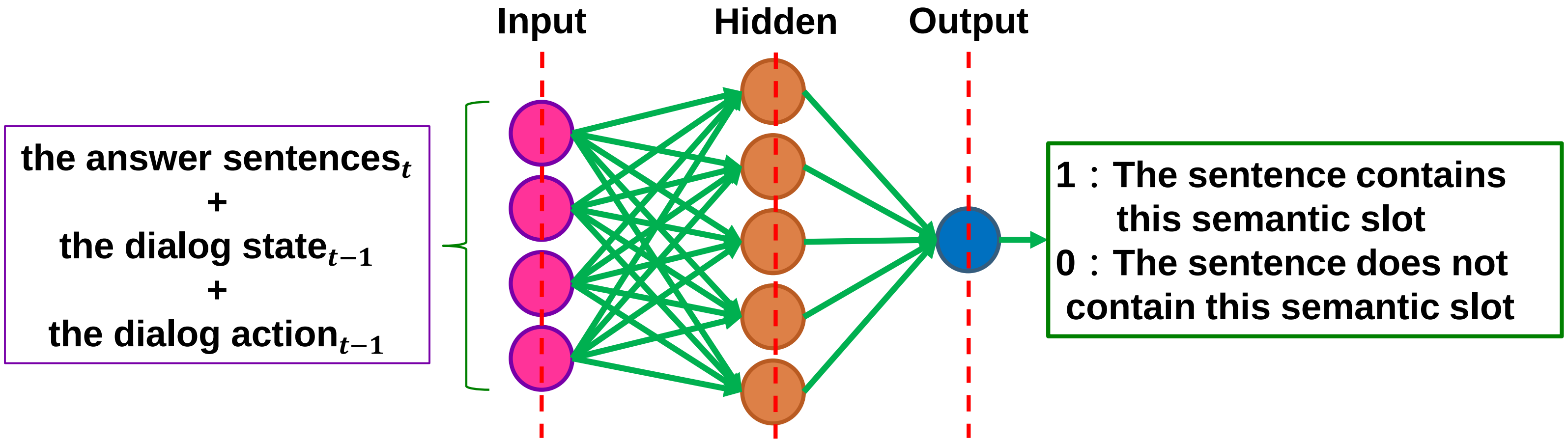
Dialogue State Tracking

- Because the **historical information** could affect the user's current answer,
  - we combine current dialogue hidden vectors and the historical information to obtain the final dialog state.



Dialogue State Tracking

- Dialog state detection:
  - each dialog state consists 10 semantic slot.
  - we use **10** ANNs to detect semantic slots.
  - we feed the **answer hidden vector** into the ANN model to obtain the dialog state.
    - the history of dialog state and dialog action in previous time were combined with the answer hidden vector .



Experimental setup

- The proposed method was evaluated using **5-fold cross validation**.
- Each word vector dimension is set to **30**.
- **10**-dimensional semantic slot output.

$Slot_1$	$Slot_2$	$Slot_3$	$Slot_4$	$Slot_5$	$Slot_6$	$Slot_7$	$Slot_8$	$Slot_9$	$\mathbf{Slot_{10}}$
1	0	0	0	1	0	0	0	0	0

The slot is “**Association and cadre leadership**”

The slot is “**Motivation**”

Experimental results (1)

- When we considered the accuracy of each ANN's output,
  - parameter tuning was conducted to obtain the best performance of the LSTM and ANN models.
  - the number of hidden nodes in the LSTM-based sentence model and answer model were set to **32**.

Experimental results (2)

- This result shows that **the information about the dialog state and dialog action at time *t* - 1** provided a best positive contribution in this experiment.

	Accuracy		
ANN Hidden Nodes	**AHV_t**	**AHV_t+DS_{t-1}**	**AHV_t+DS_{t-1}+ DA_{t-1}**
8	87.44%	**89.45%**	89.69%
16	84.98%	89.08%	**89.93%**
32	84.37%	88.70%	89.64%
64	**88.78%**	88.53%	88.72%
128	87.66%	88.55%	89.05%
256	87.53%	88.17%	88.82%

1. **AHV_t**: Input of ANN contains the answer hidden vectors at time t.
2. **AHV_t+DS_{t-1}**: Input of ANN contains the answer hidden vectors at time t and the dialogue states at time t-1.
3. **AHV_t+DS_{t-1}+ DA_{t-1}**: Input of ANN contains the answer hidden vectors at time t, and the dialog states along with the dialog action at time t-1.

Conclusions

- We propose an approach to DST detection in an interview coaching system.
  - The word2vec model is employed to encode the words to embedding vector for word distributed representation.
  - The LSTM-based model is used for sentence and answer vector representation.
  - The ANN-based model is applied to detect the final dialog state for further policy control.
- In the future, a richer interview corpus and a robust DST model are helpful to improve system performance.

Questions?