# Bing Liu

Carnegie Mellon University
NASA Research Park, Bldg 23
Moffett Field, CA 94035

liubing@cmu.edu


## 1 Research Interests

My main area of research lies in **Natural Language Understanding (NLU)** and its applications in **Spoken Dialogue Systems (SDS)**. I am particularly interested in **knowledge representation** and integrating knowledge to dialogue systems. My initial research focus is on distributed representation of text and its applications in spoken language understanding. My recent efforts are more towards learning distributed representation of knowledge and context. I am keen on exploring knowledge-enabled end-to-end trainable dialogue systems that can perform tasks with specific goals.

### 1.1 Text Representation and SLU

Spoken language understanding (SLU) system interprets the semantic meanings conveyed by speech signals. Major components in SLU systems include identifying speaker's intent and extracting semantic constituents from the natural language query, two tasks that are often referred to as intent detection and slot filling. Intent detection can be treated as a semantic utterance classification problem, and a number of standard classifiers can be applied. Slot filling can be treated as a sequence labeling problem. These two tasks are usually processed separately by different models. Prior work on intent classification and slot filling includes using uni-bi-trigrams, together with head word, hypernym, and other hand-coded rules in classifiers like SVM.

Motivated by the success of distributed representations of words in many NLP applications (Collobert et al., 2011), we designed an RNN-based SLU model with scheduled training method (Liu and Lane, 2015) that can effectively capture the sequential structural patterns of the text. In a following-up work (Liu and Lane, 2016a), we developed an RNN-based model that can be used to jointly perform intent detection and slot filling in SLU. Such joint models simplify SLU systems, as only one model needs to be trained and deployed. In this work, we also explored the effectiveness of applying neural attention mechanism in the RNN model, and evaluated the attention based RNN model in sequence-to-sequence learning settings.

In many real world applications using speech interface, real time responses from the agent are desired. In speech recognition, instead of receiving the transcribed text at the end of the speech, users typically prefer to see the ongoing transcription while speaking. In SLU, with real time intent identification and semantic constituents extraction, the downstream systems will be able to perform corresponding search or query while the user dictates. In a recent work (Liu and Lane, 2016b), we designed an RNN-based online joint SLU model that keeps tracking the intent variations as word in the transcribed utterance arrives and uses it as contextual features in the joint model. In addition, we modeled the interaction between the SLU and the language modeling tasks and showed that such joint modeling led to better ASR and SLU performance. As a next step, we want to see how such joint model can be further extended for belief tracking in dialogue systems when considering the dialogue history beyond a single utterance.

### 1.2 Knowledge Representation

Introducing knowledge to dialogue and question answering systems is an important step towards building intelligent conversational agents. Knowledge comes in various forms, and how these knowledge can be represented and integrated to the intelligent system in an end-to-end trainable manner is an active research problem. Large scale knowledge bases are good sources for structured knowledge. Such knowledge is entity centric and is typically represented by properties and relations between entities. Embedding methods in representing the entity and relations has been explored in literature (Socher et al., 2013). Key challenges in question answering with knowledge base include the very large search space in knowledge base and query compositionality. Knowledge in unstructured form can come from web documents (e.g. text and images) that are returned by search engine given a query. Knowledge from web document is likely to be more complete and updated comparing to that from knowledge base. The main challenges are the accurate understanding of the query and knowledge extraction from data in multi-model form.

My current research interest lies in extracting and integrating task-specific knowledge to the dialogue manager. To control the search space, we want the knowledge to be dynamically retrieved from knowledge base or web doc-

uments based on the task and detected user intent.

### 1.3 Task-oriented Dialogue Systems

Sequence to sequence models have shown promising improvement in building end-to-end trainable, open domain dialogue systems (Vinyals and Le, 2015). Such models typically require a large amount of data to train, which is often not available for task-oriented applications. To mitigate this problem, popular approaches in closed domain dialogue system design cast the task as a partially observable Markov Decision Process (POMDP) and use reinforcement learning for online policy optimization by interacting with users (Gašić et al., 2013). Recently, a neural network based end-to-end trainable task-oriented dialogue system (Wen et al., 2016) is proposed with a novel crowdsourced data collection framework and showed promising performance. In my research, I would like to whether the problem of insufficient training data in task-specific end-to-end trainable dialogue systems can be mitigated by the effective usage of the introduced knowledge.

## 2 Future of Spoken Dialog Research

In the long run, I think a truly intelligent conversational agent will not only be able to handle comprehensive queries, but also to understand human emotions and express personality. The agent should be able to understand complex questions, provide solutions with updated real world knowledge, and respond in a manner that is adjusted based on the user's characteristics.

In the near future, I believe we will see the growing importance of data-driven approaches in designing dialogue systems, in both open and task-oriented domains. Current spoken dialogue systems contains large amount of handcrafting, and this substantially limits the systems' domain scalability. Data driven approach with customized domain knowledge can be a promising solution. Moreover, I see good potential of applying transfer learning in spoken dialogue systems. It will be interesting to see how a well performing task-specific dialogue model can be efficiently adapted to a new domain that with limited training data.

## 3 Suggestions for Discussion

I would like to suggest the following topics for discussion.

- Integrating knowledge or external memory to dialogue systems.
- Transfer learning in task-specific dialogue systems.
- Evaluation: what can be good metrics in evaluating the responses generated by conversational agent?
- What are the main challenges in deploying real world spoken dialogue systems?

## References


Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537.

M Gašić, Catherine Breslin, Matthew Henderson, Dongho Kim, Martin Szummer, Blaise Thomson, Pirros Tsiakoulis, and Steve Young. 2013. On-line policy optimisation of bayesian spoken dialogue systems via human interaction. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8367–8371. IEEE.

Bing Liu and Ian Lane. 2015. Recurrent neural network structured output prediction for spoken language understanding. In *Proc. NIPS Workshop on Machine Learning for Spoken Language Understanding and Interactions*.

Bing Liu and Ian Lane. 2016a. Attention-based recurrent neural network models for joint intent detection and slot filling. In *INTERSPEECH*.

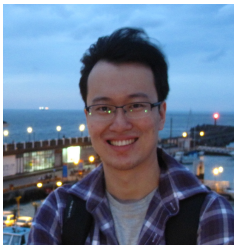
Bing Liu and Ian Lane. 2016b. Joint online spoken language understanding and language modeling with recurrent neural networks. In *SIGDIAL*.

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems*, pages 926–934.

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, David Vandyke, and Steve Young. 2016. A network-based end-to-end trainable task-oriented dialogue system. *arXiv preprint arXiv:1604.04562*.


## Biographical Sketch



Bing Liu is a PhD student in the Department of Electrical and Computer Engineering at Carnegie Mellon University, working under the supervision of Professor Ian Lane. His research interests include using machine learning for natural language processing, spoken language understanding, and dialogue systems. Bing graduated with a Bachelor degree in Electrical and Electronic Engineering from Nanyang Technological University (Singapore).