

# Modeling Dialogues in Grounded Document

---

Anuja Tayal  
cs521

# Motivation

- Do models have to know everything??
- Feed more data??
- Jeopardy - Just know Trivia Question? Facts? Or Smart?
- Solution - **Grounded Document!!**




# Dataset

- MultiDoc2Dial Dataset Shared Task
- 4 domains- ssa, student aid, va, dmv
- Used dmv subset
- Each dialog turn is annotated with role, dialogue act, human-generated utterance and the grounding span with document information
- user questions, agent follow-up questions and responding turns

documents	149
dialogues	1328
2 Segments	781
> 2 Segments	257
single	290

# Dataset



Home | [About Us](#) | [Contact Your DMV](#) | [Site Map](#)

Auto Insurance Quotes

Start Saving

Home

About Us

DMV Forms

Driver Manuals

Drivers License

Site Map

Home > [Www.ilsos.gov](#) > [Services](#) > **New to Illinois Driving**

## New to Illinois Driving

Services

Welcome to the State of Illinois. Below are links that will give you information on driving in Illinois. If you need further information, please contact the Secretary of State's office by phone at 800-252-8980 (Illinois only) or by completing the [Driver's License contact form](#) or the [Title and Registration contact form](#).

- [Contact Us](#)
- [Driver's License/State ID Card](#)
- [Facility Finder](#)
- [How Do I?](#)
- [Illinois Rules of the Road](#)
- [License Plates Guide](#)
- [Mandatory Insurance](#)

- [Motor Voter Information](#)
- [Organ and Tissue Donation](#)
- [Secretary of State Facility Checklist](#)
- [Vehicle Titles and License Plates for New Illinois Residents](#)
- [Title and Registration](#)
- [Traffic Safety](#)

# Proposed Solution

- 3 components
- Dialog Act Prediction
- Retrieve Model- to retrieve the most important context, given the query
- Question Answering Module - given the context and the query, find the answer span

# Dialog History

```
'You've got it. Another query about DMV. What happens if you seek a hearing with the DMV?[SEP]agent: Yes, you can sign up for MyDMV for all the online transactions needed.||user: Can I do my DMV transactions online?||agent: hi, you have to report any change of address to DMV within 10 days after moving. You should do this both for the address associated with your license and all the addresses associated with all your vehicles.||user: Hello, I forgot to update my address, can you help me with that?'
```

```
{'text': 'the suspension is placed on hold pending the outcome of the hearing. ', 'answer_start': 1140, 'answer_end': 1209}
```

# Dialogue Act- Multi Class Classification

			utterance
da	label	data_type	
query_condition	0	train	3863

index	da	utterance
0	query_condition	Hello, I forgot o update my address, can you help me with that?
1	respond_solution	hi, you have to report any change of address to DMV within 10 days after moving. You should do this both for the address associated with your license and all the address associated with all your vehicles.
2	query_solution	Can I do my DMV transactions online?
3	respond_solution	Yes, you can sign up for MyDMV for all the online transactions needed.
4	query_condition	You've got it. Another query about DMV. What happens if you seek a hearing with the DMV?

respond_solution	1	val	58
		train	4335
response_negative	3	val	765
		train	629
response_positive	5	val	111
		train	633
		val	112

# Dialogue Act

- Train Bert Sequence Classification Model
- Batch\_size 4
- epochs=4
- learning\_rate=1e-5
- Trained through Pytorch transformers

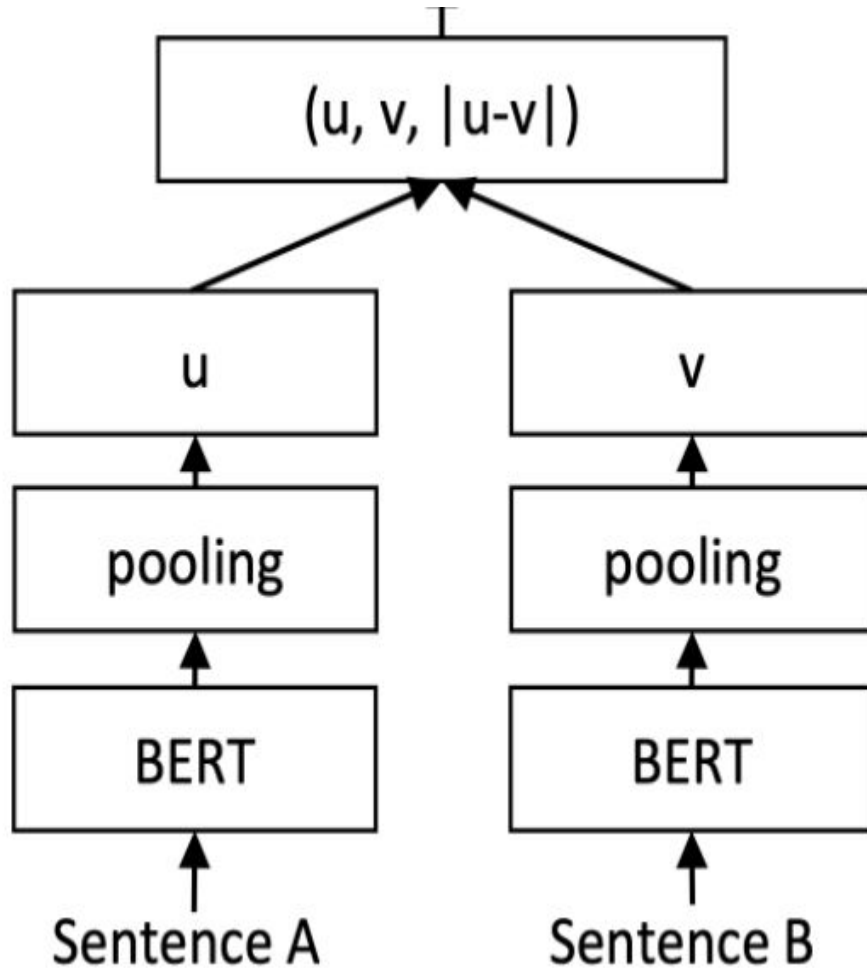


# Retrieval Model

- Retrieve the most important context given the query
- Symmetric Semantic Search- query and context are of same length
- **Asymmetric Semantic Search**- context is huge as compared to query
- **Dot product**- retrieval of long documents
- Cosine Product- retrieval of short documents
- **Msmarco-distilbert-base-dot-prod-v3 model**- Sentence Bert Model
- Used Sentence Transformer Library and FAISS Library

## Related Work

- TF-IDF
- BM-25
- SBert
- Dense Passage Retrievals (DPR)



# Retrieval Model

- **FAISS:** (Facebook AI Similarity Search) is a library that allows to quickly search for embeddings of multimedia documents that are similar to each other.
- Used SBERT to embed contexts
- Indexed the embeddings through FAISS
- For finding the context for given query ,
  - Encode query through sbert
  - Search the embedding with indexed db and retrieve top\_k results

# Retrieval Model

- 127 unique contexts
- 6135 query and context values

True	2788
False	3347
Accuracy	50%

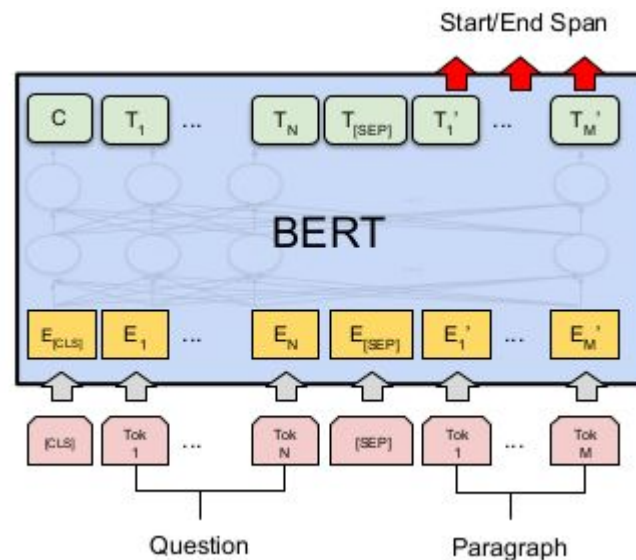
Without history

True	4200
False	1935
Accuracy	70%

With history

# Question Answer Module

- BertForQuestionAnswering
- epochs=3
- Learning rate=5e-5
- Batch size=4
- Evaluation metric=Bleu Score
- Bilingual Evaluation Understudy
- score for comparing a candidate translation of text to one or more reference translations.
- Average Bleu score = .58



# Challenges

- Context length 5000
- Bert max length 512

# Future Work

- Use LongFormer to break context
- Use chained indexing
- Break context into more smaller blocks
- Add personalisations

# References

- MultiDoc2Dial
- Doc2Dial
- Bert For Question Answering
- Faiss
- Bleu
- SBert
- MultiClass Classification

Thank You!!  
Questions??

---