Building a Natural Language Understanding Module

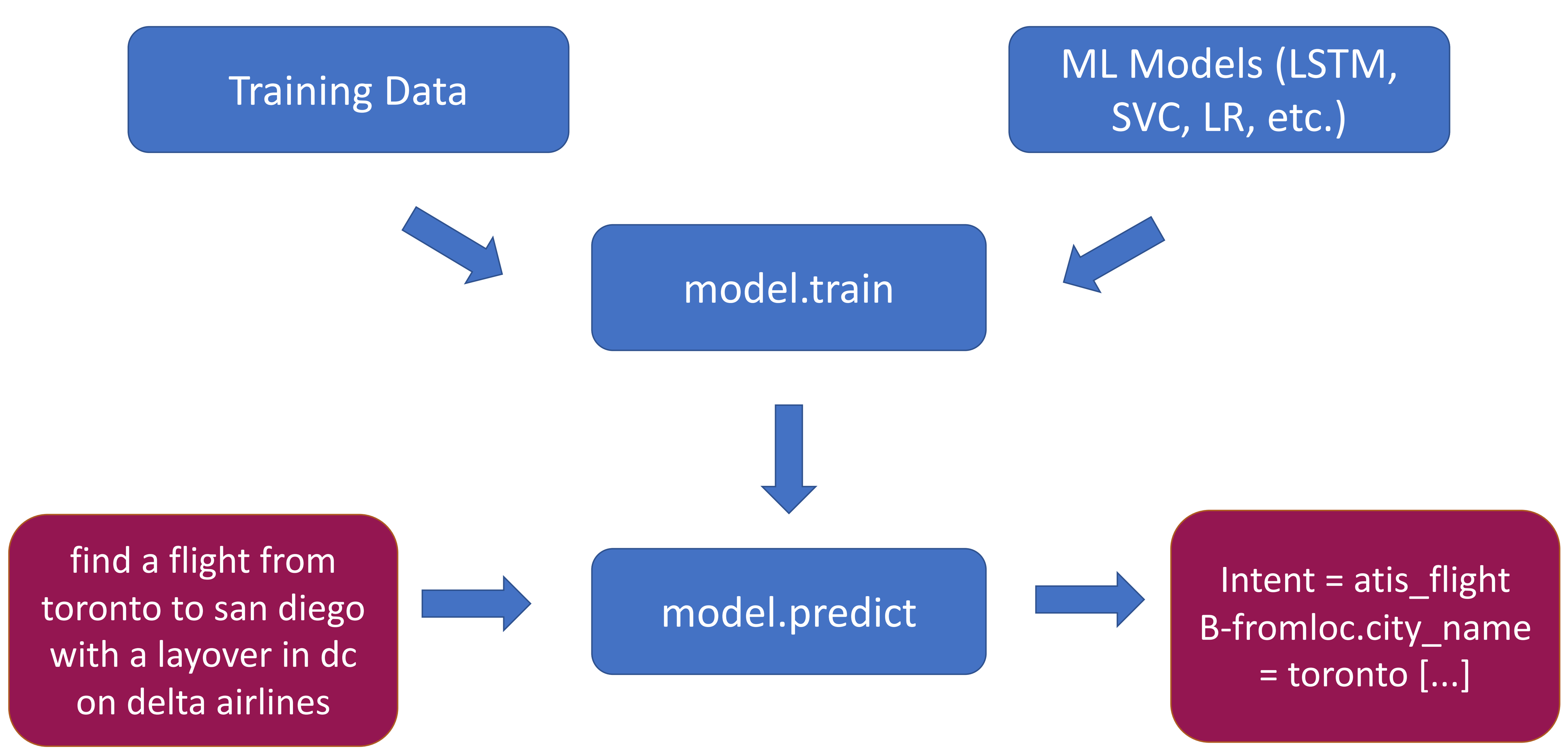
Mckenna Brown (mckennab@andrew.cmu.edu)

Advisors: Maxine Eskenazi, Shikib Mehri

Summary

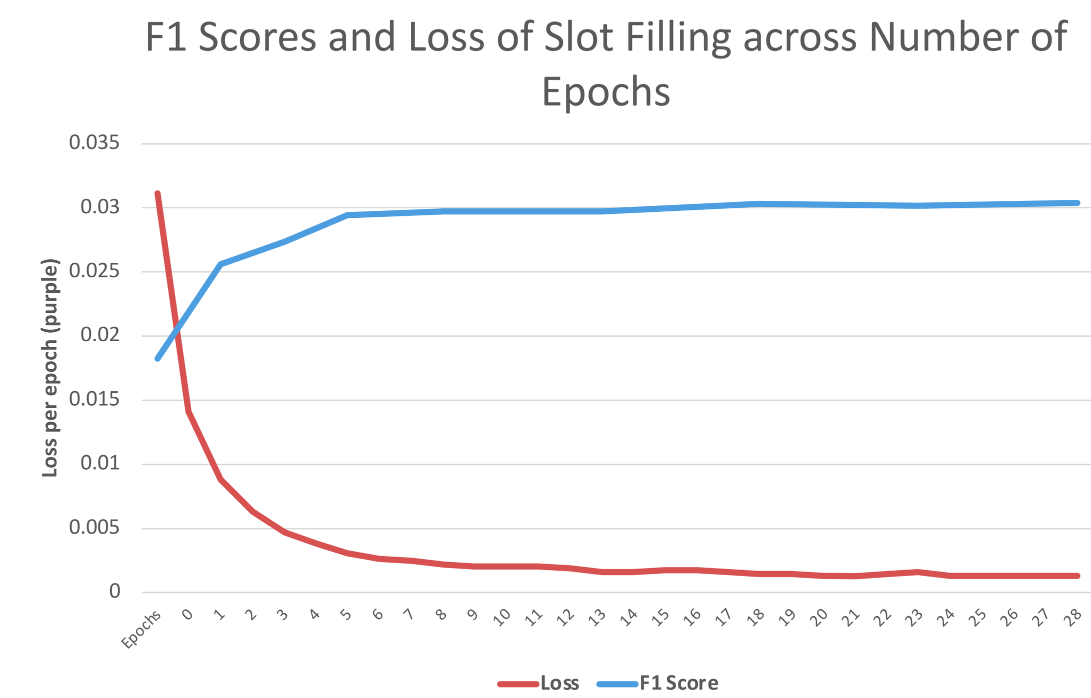
- Creates a general and extendable **Natural Language Understanding** (NLU) module
- Task-oriented NLP use cases are in increasing demand
- Takes a transcribed sentence and predicts:
  - Intent: the general intent behind the sentence
  - Slots: the key details provided in the sentence
- Uses various methods including:
  - Support Vector Machines (SVMs)
  - Logistic Regression
  - Long Short Term Memory Networks (**LSTMs**), a type of **Recurrent Neural Network**
- Will be made open source for use in the Dialog research community

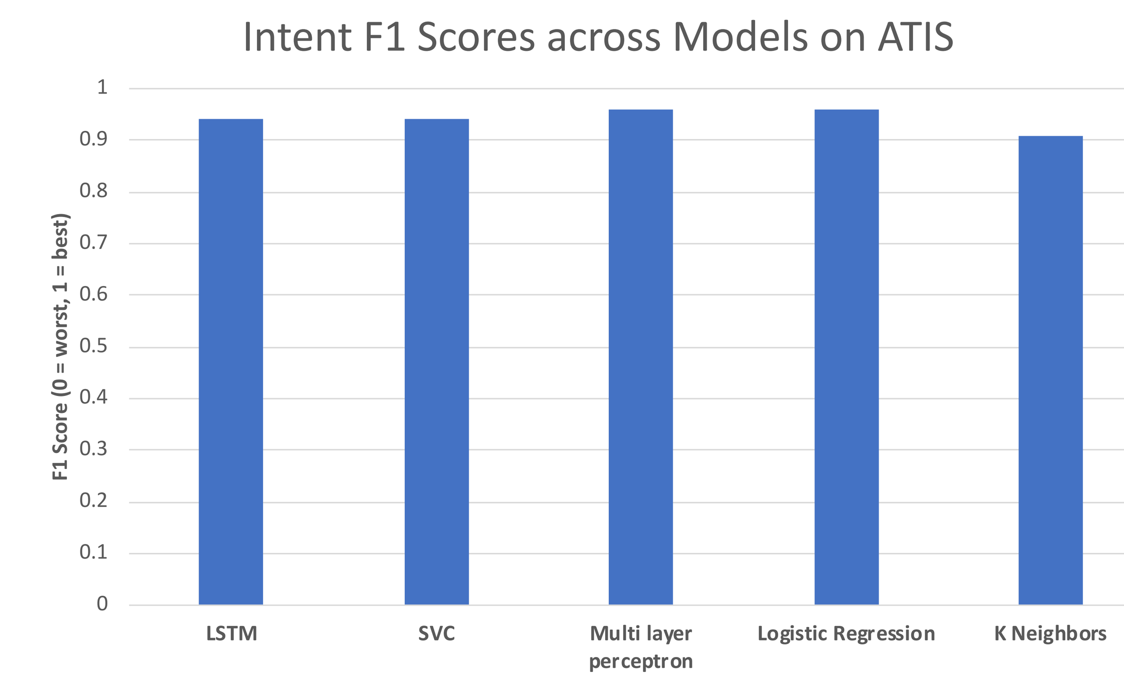
Methods:



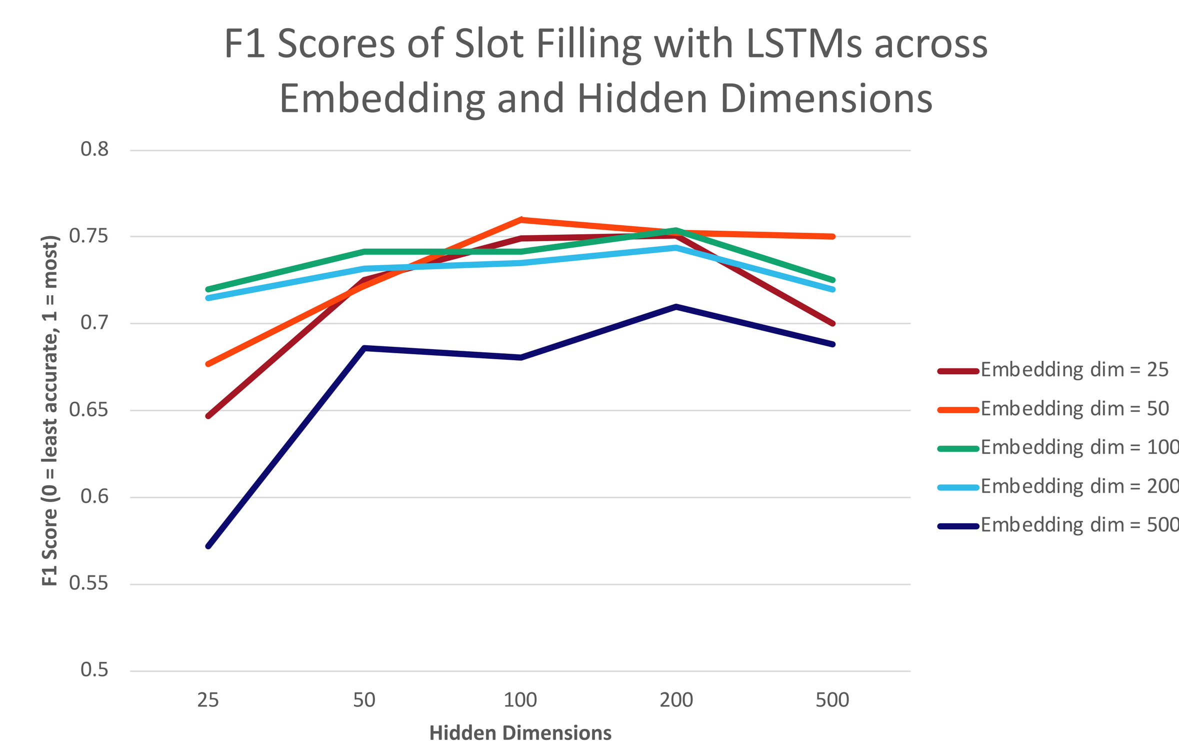
- Separate models for Intent Classification and Slot Filling
- Any ML model with a train and predict feature can be used
- Many common ML models were analyzed, including LSTMs which dynamically include past context

Results/Discussion





Across various ML models Intent Classification performed similarly well, with F1 scores above 0.9



Data for both above graphs was gathered with Loss = 0.01, batch size = 10 for LSTM slot filling in ATIS. An embedding dimension of 50 and hidden dimension of 100 maximized accuracy, and 10 epochs were chosen for efficiency.

- On ATIS intent classification achieved high accuracy (F1 = .94 - .96)
- In a more complex dataset (MultiWOZ), Intent F1 of .755 was achieved
- Slot Filling reached an F1 score of .760 on ATIS
- Top challenge with Slot Filling on ATIS involved correcting for skewed dataset
- Including pre-trained word embeddings from Google News lowered accuracy for both intents and slots

Conclusions

- Handling similar but unseen words (such as “November” vs. “December”) was not assisted by decreasing vocab size or including word embeddings
- In skewed datasets, similar majority class terms were used instead of the minority versions
  - A minority class utterance might slot “Monday” as “B-depart_date.day_name” while it should have predicted “B-day_name”
- Combining intent classification and slot filling into a single model could be used to help account for slot filling’s failures on minority classes
- Incorrect/Correct scoring on results can be enhanced to distinguish between slightly wrong and completely wrong predictions
  - With IOB tagging, mistakes such as “B-depart_time.period_of_day” instead of “I-depart_time.period_of_day” are less harmful and could be fixed with simple rules
- Extending the analysis over several datasets could enhance database-independent models

DialRC

Carnegie Mellon University Language Technologies Institute

Carnegie Mellon University School of Computer Science