CLASSIFYING QUESTIONS

SHRI SHRUTHI SHRIDHAR

SOWMYA S SUNDARAM

PROBLEM STATEMENT

- Identify the type of question being asked by the user with these categories - Who, What, When, Affirmation (Yes / No) or Unknown.

PROBLEM STATEMENT

- Identify the type of question being asked by the user with these categories - Who, What, When, Affirmation (Yes / No) or Unknown.
- Build an interface to display it to the user, and validate your model's response with that of the user's actual intent.

PROBLEM STATEMENT

- Identify the type of question being asked by the user with these categories - Who, What, When, Affirmation (Yes / No) or Unknown.
- Build an interface to display it to the user, and validate your model's response with that of the user's actual intent.
- Show analytics of performance

PROBLEM STATEMENT

- Identify the type of question being asked by the user with these categories - Who, What, When, Affirmation (Yes / No) or Unknown.
- Build an interface to display it to the user, and validate your model's response with that of the user's actual intent.
- Show analytics of performance
- Improve performance dynamically

DATASET ANALYSIS

- No of samples – 1483
- Distribution
- Missing class labels – taken as unknown

Category	Number
What	574
Who	371
When	89
Affirmation	103
Unknown	342

CHALLENGES

- Key challenge is differentiating affirmation and unknown

- If we take conventional word2vec approaches all question words such as who, what, when, why, how etc may have similar values. Hence classifying between them might be difficult.

- Class imbalance may make the model weak for affirmation

APPROACHES ATTEMPTED

- Multiple Classifiers after applying mean of glove
- Using first two words of sentence and building glove
- Applying over-sampling to balance training

APPROACH - 1

- Multiple Classifiers after applying mean of glove
- After splitting dataset into 80-20% for training and test, we evaluated various models

Classifier	Test Accuracy
Logistic Regression	0.68
SVM	0.67
Random forest classifier	0.52
Multi Layer Perceptron	0.64

APPROACH - 2

- Multiple Classifiers after applying mean of glove on first 2 words
- After splitting dataset into 80-20% for training and test, we evaluated various models

Classifier	Test Accuracy
Logistic Regression	0.89
SVM	0.89
Random forest classifier	0.74
Multi Layer Perceptron	0.91

APPROACH - 3

- Sampled all datasets equally to minority class
- Performed badly, all accuracies ~ 0.2

POST MODEL BUILD

- Chose multi layer perceptron with 2 hidden layers – 150 and 10 nodes
- Trained on entire dataset before launching interface

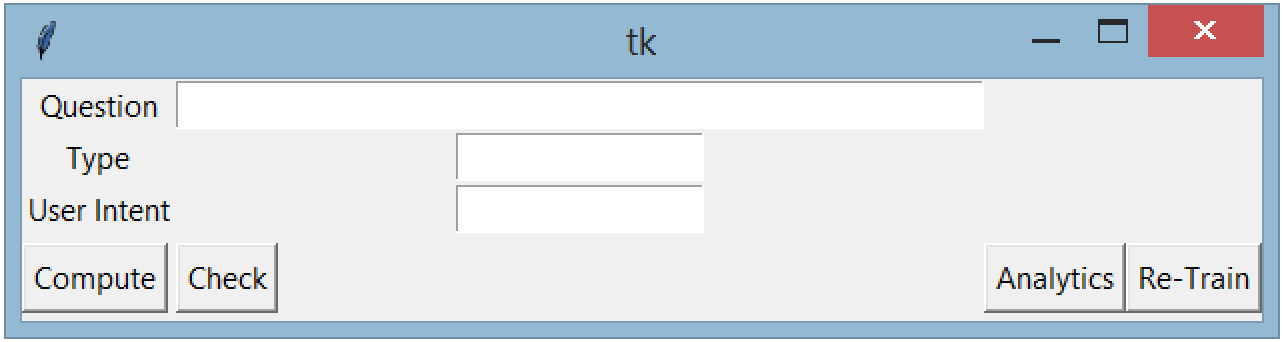
INTERFACE

- Desktop interface
- Trained on entire dataset before launching interface
- Functionalities
  - Query Model
  - Check Model against User Intent
  - See statistics of queries in this session
  - Retrain model

TECH DETAILS

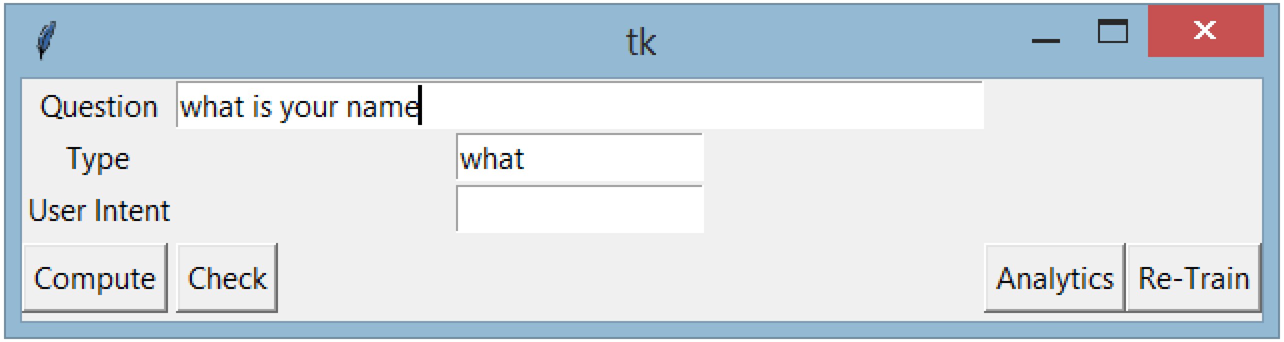
- Python 3
- Sklearn
- NumPy
- Tkinter
- Gensim

SCREENSHOTS



SCREENSHOTS

Compute



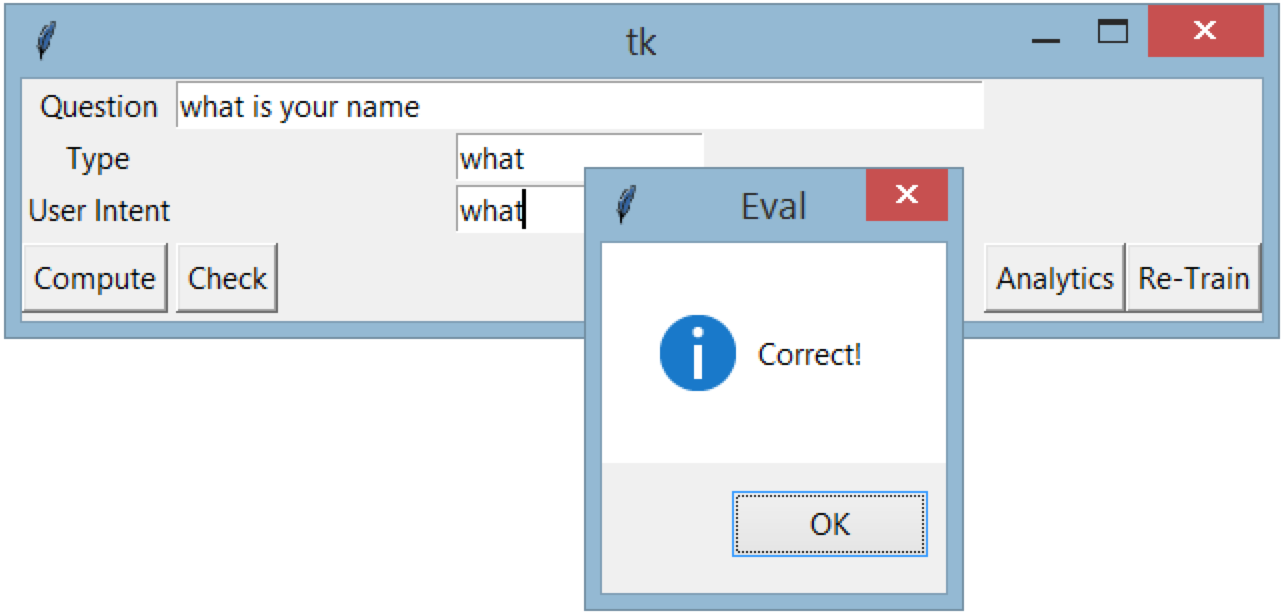
Question what is your name

Type what

User Intent

SCREENSHOTS

Check



Question what is your name

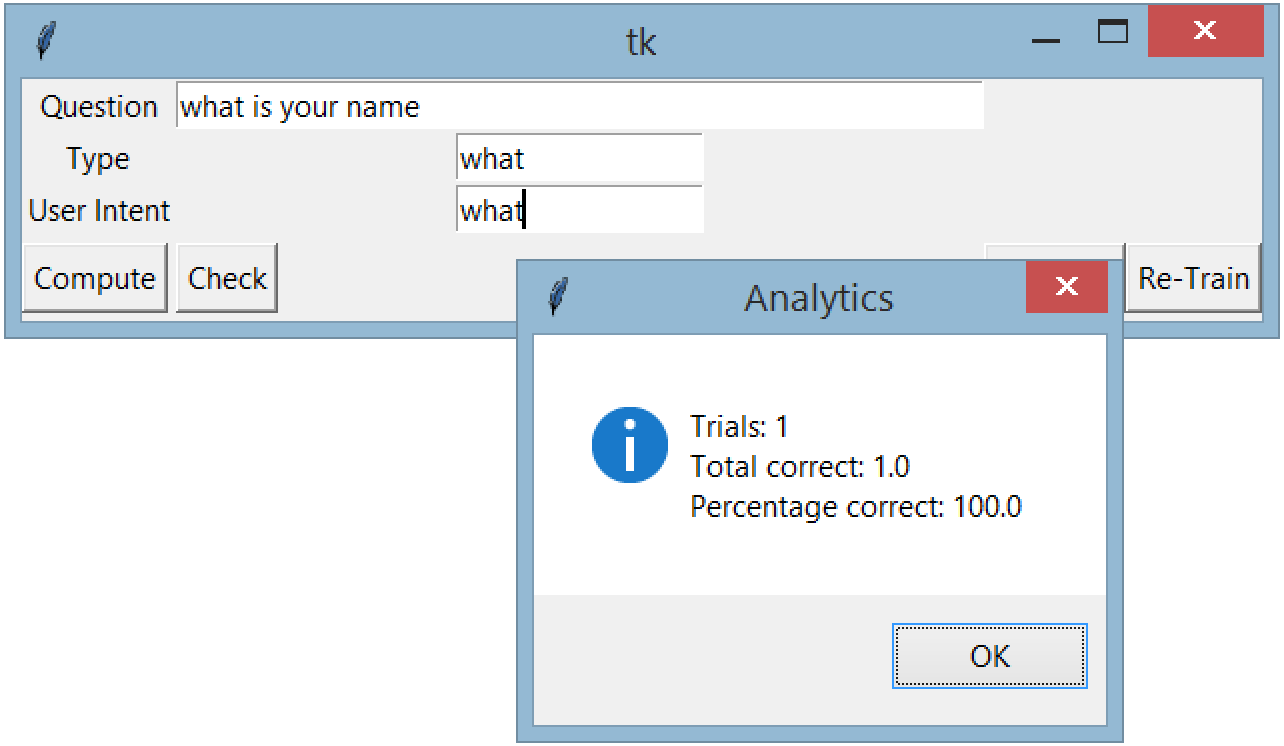
Type what

User Intent what

Correct!

SCREENSHOTS

Analytics



Question: what is your name

Type: what

User Intent: what

Trials: 1
Total correct: 1.0
Percentage correct: 100.0

IMPROVEMENTS

- Experiments with doc2vec
- Partially rule based and partially trained
- Associate a control logic for re-training

APPLICATIONS

- Chatbot pre-processing
- Automated feedback design