Project proposal

Customer support ChatBot

SCAN ME

October 2024

Generative Ai ONL1_AIS2_S9e

Table of Contents

Project team

Name	Initials
Khaled El Sherif	K.E.
Mohamed Yousef	M.Y.
Abdoh Elgazar	A.E.
Farrah Tharwat	F.T.

Project Idea

Companies depend on large and round-the-clock support staff to maintain customer engagement. This can be costly and impractical. Chatbots can effectively boost efficiency and lower the cost of business. It resolves unlimited queries with the least human interaction. Creating robust chatbot training data is costly in terms of human time and effort. Additionally, it increases the product design to deployment time. Companies with only a minimal amount of training data may not be able to produce a robust enough model to use for chatbot interactions.

In this project, we will use internal customer support data to create a robust chatbot using the Sequential model. It also demonstrates how to run the chatbot by using the developed model.

Project Steps

1 - Preprocess semi-structured data

2 - Perform exploratory data analysis

3 - Unsupervised labelling

4 - Training data preparation

5 - Hyperparameter tuning

6 - Train the deep learning (DL)

7 - Evaluating the DL model

8 - Use model to predict responses

The pre-trained model

1/2

universal-sentence-encoder

https://www.kaggle.com/models/google/universal-sentence-encoder/tensorFlow2/universal-sentence-encoder/2?tfhub-redirect=true

The Universal Sentence Encoder encodes text into high-dimensional vectors that can be used for text classification, semantic similarity, clustering and other natural language tasks.

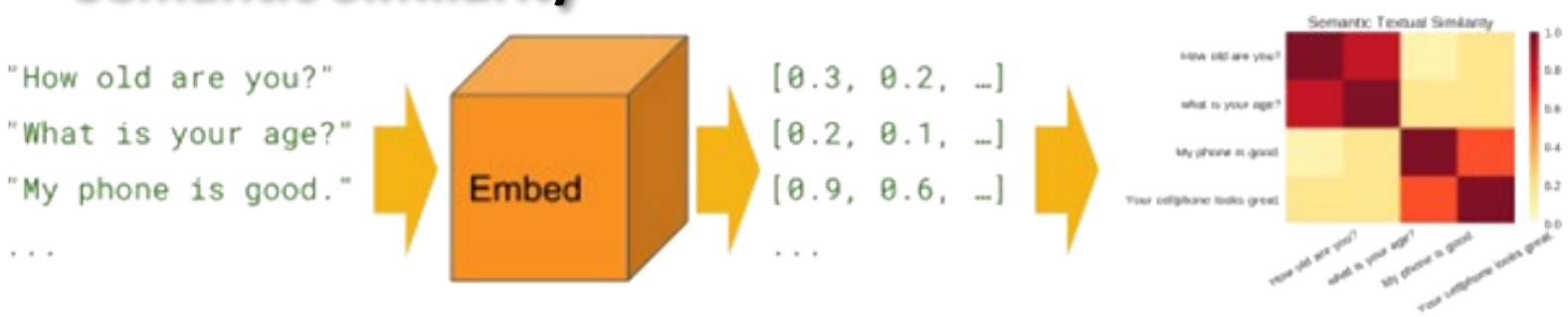
The model is trained and optimized for greater-than-word length text, such as sentences, phrases or short paragraphs. It is trained on a variety of data sources and a variety of tasks with the aim of dynamically accommodating a wide variety of natural language understanding tasks. The input is variable length English text and the output is a 512-dimensional vector. The universal-sentence-encoder model is trained with a deep averaging network (DAN) encoder.

The encoder differs from word level embedding models in that it is trained on a number of natural language prediction tasks that require modeling the meaning of word sequences rather than just individual words.

The pre-trained model

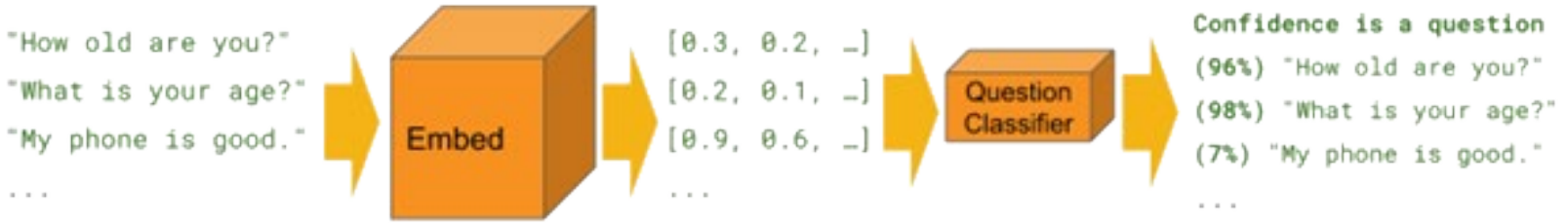
2/2

Semantic Similarity



Semantic similarity is a measure of the degree to which two pieces of text carry the same meaning. This is broadly useful in obtaining good coverage over the numerous ways that a thought can be expressed using language without needing to manually enumerate them.

Classification



The Universal Sentence Encoder was partially trained with custom text classification tasks in mind. These kinds of classifiers can be trained to perform a wide variety of classification tasks often with a very small amount of labeled examples.

Data used

The dataset is an unstructured assortment of customer service inquiry chat logs.

The chat logs consist of timestamped dialogue between a human customer agent and a visitor to the company website.

The dialogue consists predominately of queries about services, prices, location, and signup information.