

Project proposal



Customer support
ChatBot



SCAN ME

October 2024

Generative Ai ONL1_AIS2_S9e

Table of Contents



1

Project team

Page 1

2

Project Idea

Page 2

3

Project Steps

Page 3

4

The used pre-trained model

Pages 4-5

5

Data used

Page 6



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.



Kha

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.

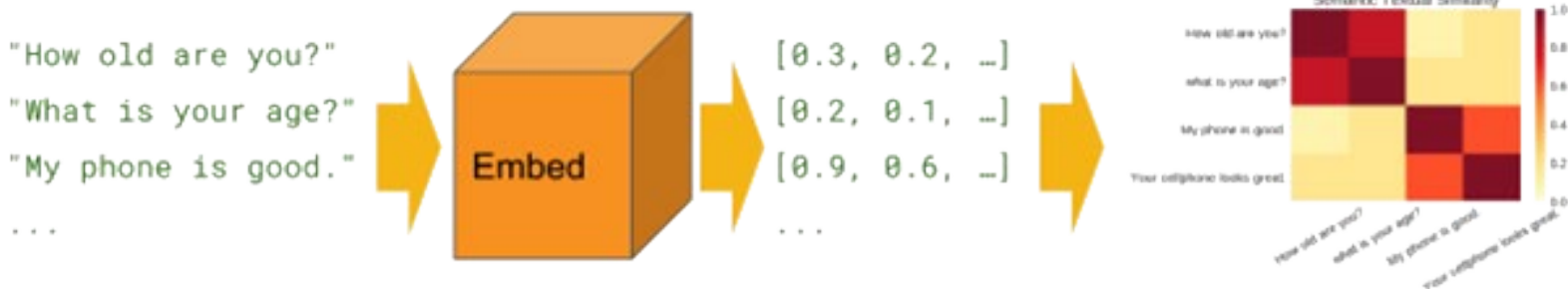
The encoder differs from word level embedding models in that it is trained on a number of natural language prediction tasks that require modeling the meaning of word sequences rather than just individual words.

The pre-trained model



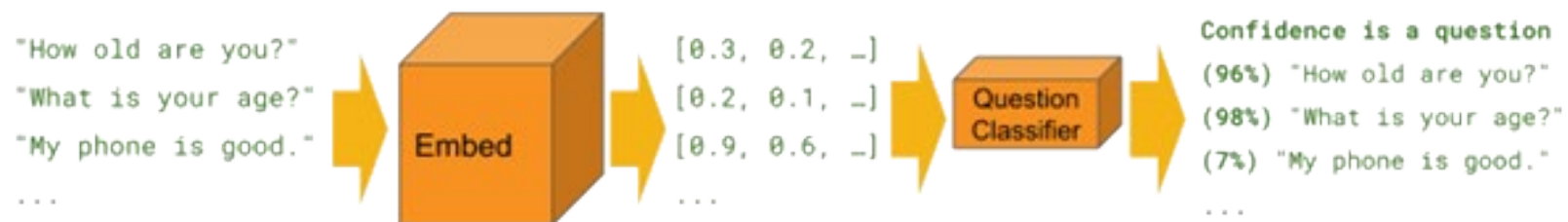
2/2

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.

```
(2022-03-25 18:30:45) Visitor 73930186: for data scientist /Data analyst role
(2022-03-25 18:31:07) Anjali: we can schedule a quick demo session for 10-15 mins . It
will be on google meet and will provide you more understanding about the platform
(2022-03-25 18:31:45) Visitor 73930186: Sure lets do that on monday - 28
(2022-03-25 18:32:01) Anjali: ok sure
(2022-03-25 18:32:16) Anjali: what time would you prefer
(2022-03-25 18:32:43) Visitor 73930186: 11:00 am Est
(2022-03-25 18:32:49) Anjali: okm sure
(2022-03-25 18:32:54) Visitor 73930186: thankyou
(2022-03-25 18:33:02) Anjali: your welcome ...
(2022-03-25 18:34:08) Anjali: we have discount offer goooing on now ... by Monday it will
get expired
(2022-03-25 18:34:23) Anjali: can we do the demo session today or tomm ?
```

```
(2022-03-25 15:58:07) Visitor 50457400: it's fine
(2022-03-25 15:58:11) Visitor 50457400: let's go
(2022-03-25 15:59:01) Visitor 50457400: and grow the kowleage with courses and practice
in project side
(2022-03-25 15:59:03) Visitor 50457400: thank
(2022-03-25 15:59:05) Visitor 50457400: s
(2022-03-25 15:59:30) Anjali: ok'
(2022-03-25 15:59:46) Anjali: may i have your email address
(2022-03-25 15:59:55) Visitor 50457400: Anjali thanks again for your help
(2022-03-25 16:00:12) Visitor 50457400: what kind of email personnal or pro or academic ?
(2022-03-25 16:27:26) Visitor 50457400: here ?
```

```
(2022-03-25 15:58:07) Visitor 50457400: it's fine
(2022-03-25 15:58:11) Visitor 50457400: let's go
(2022-03-25 15:59:01) Visitor 50457400: and grow the kowleage with courses and practice
in project side
(2022-03-25 15:59:03) Visitor 50457400: thank
(2022-03-25 15:59:05) Visitor 50457400: s
(2022-03-25 15:59:30) Anjali: ok'
(2022-03-25 15:59:46) Anjali: may i have your email address
(2022-03-25 15:59:55) Visitor 50457400: Anjali thanks again for your help
(2022-03-25 16:00:12) Visitor 50457400: what kind of email personnal or pro or academic ?
(2022-03-25 16:27:26) Visitor 50457400: here ?
```

