Automatic Response Generation to Conversational Stimuli

Vishal Raj Dutta 2015115

Sanidhya Singal 2015085

Group 36

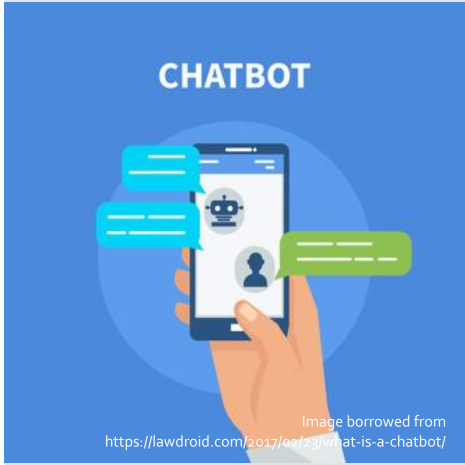


Image borrowed from https://lawdroid.com/2017/02/23/what-is-a-chatbot/

Problem Statement

How to make a computer engage in natural conversation with a person?

Challenges:-

- Previously depended on complex rule-based systems
- Seq2Seq learning - a new approach based on RNNs

  -No rules, learns everything seamlessly from data

Related Work

State-of-the-art for our data:

The Cornell Movie-Dialogue Corpus

Perplexity achieved = 2.74 *

*https://medium.com/botsupply/generative-model-chatbots-e422ab08461e

Dataset & Evaluation

Dataset :

- 20k conversational exchanges from Cornell corpus for training data
- 2k for validation data
- vocabulary of ~1000 most common words
- unknown words - replaced by special token

Evaluation :

- qualitative metric : 'human-ness' score
- quantitative metric : perplexity

Feature Extraction

Trimmed sentences to fixed length, and padded them.

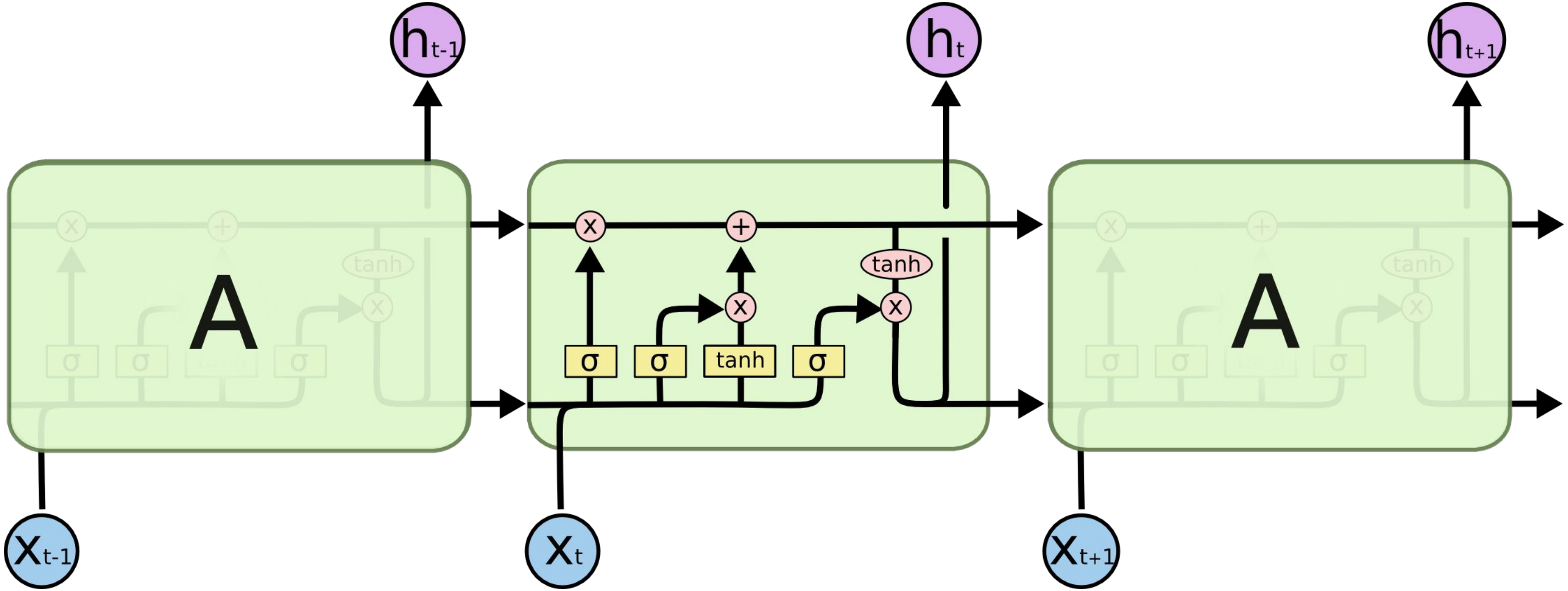
Tokenized into words and used them as features, using:

1. Word2Vec Model (CBOW)
2. 1-Hot vectorization

Strategy: Models Tried

1. Statistical HMM model
   - produced less coherent responses
2. 4-layered LSTM with Word2Vec
   - responses were not good, at all (reasons)
3. 2-layer encoder-decoder based LSTM with 1-hot-vectorization
   - best of the lot!

For comparisons :
Used HMM for statistical model and encoder-decoder for neural model



LSTM

Image borrowed from http://colah.github.io/posts/2015-08-Understanding-LSTMs/

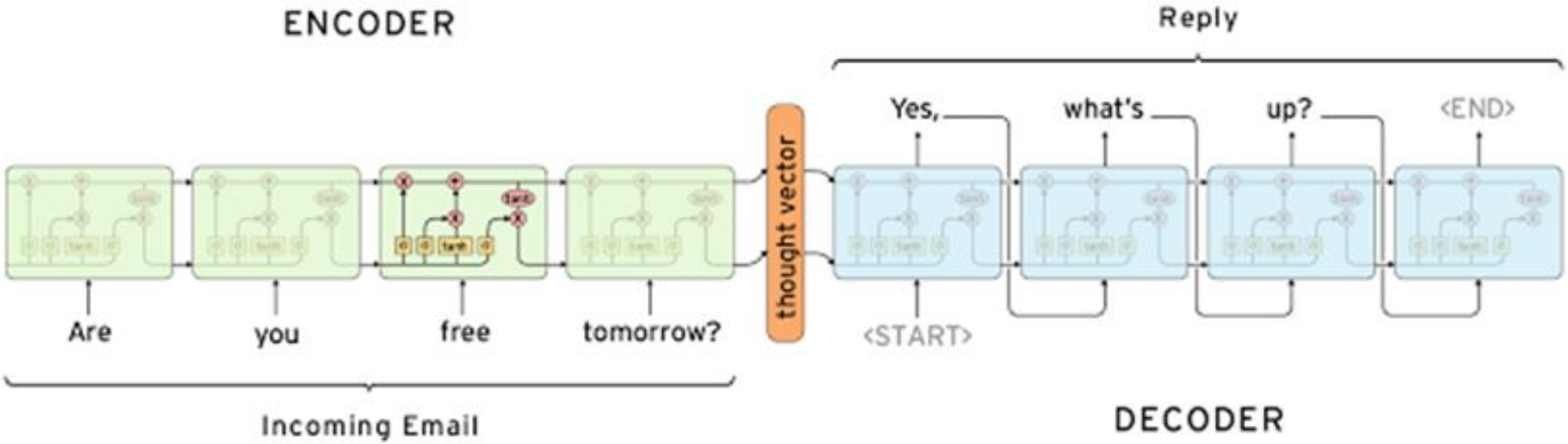


Image borrowed from https://research.googleblog.com/2015/11/computer-respond-to-this-email.html

Analysis & Progress

HMM model :

-used a bigram model

-performed worse than neural model

Encoder-Decoder model :

Analyses carried out -

LSTM v/s GRU

No. of hidden layer dimensions

Batch sizes during training

Encoder-input reversal and non-reversal

Results

Quantitative analysis : Perplexity

SOTA : 2.74

Neural : 8.04

HMM :26.24

Qualitative analysis : Human-ness score*

SOTA : 5.135

Neural : 4.671

HMM : 2.1982

*survey done on IIITD students

Conclusion

Came very close to SOTA!

Neural framework performed way better than the HMM base-line.

Further improvements:

more layers?

more training?

Demo

hi

are you the king in the north?

you are funny

are you okay

what is your name? (then)Wouldn't tell me ?

you should go home

love me do

would you like AI ?

do you know of her ?

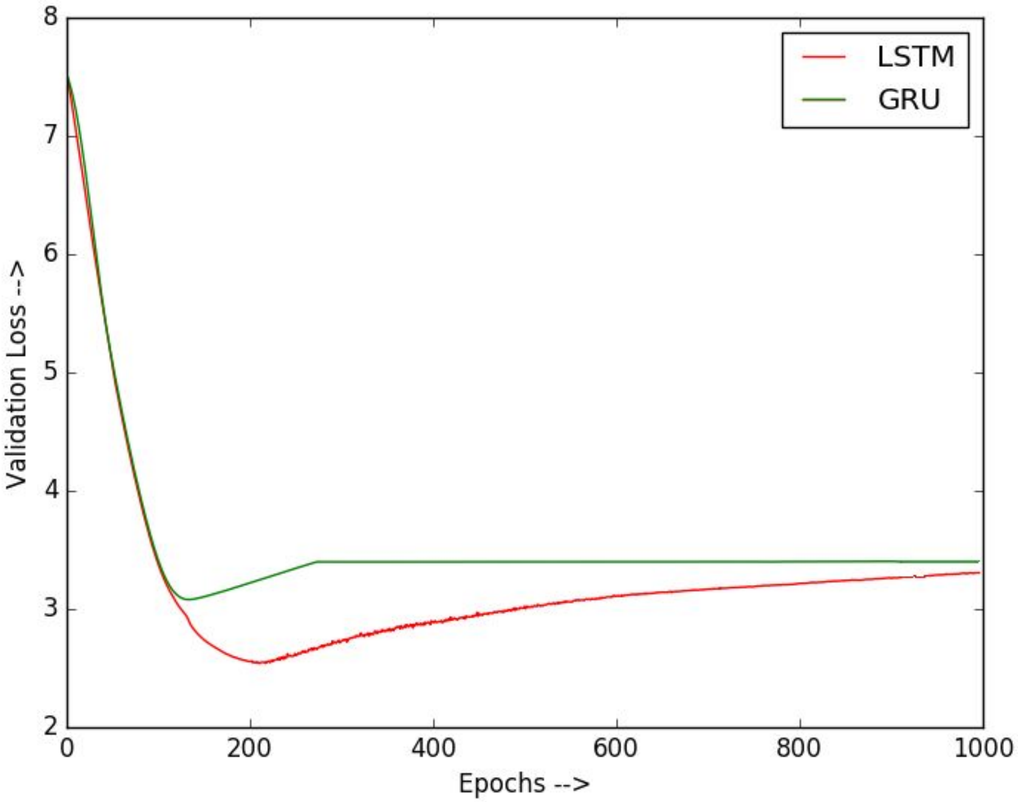


Figure 1. Comparative Study of LSTM and GRU based on Perplexity values
(Batch Size = 64, Latent Dimensions = 512)

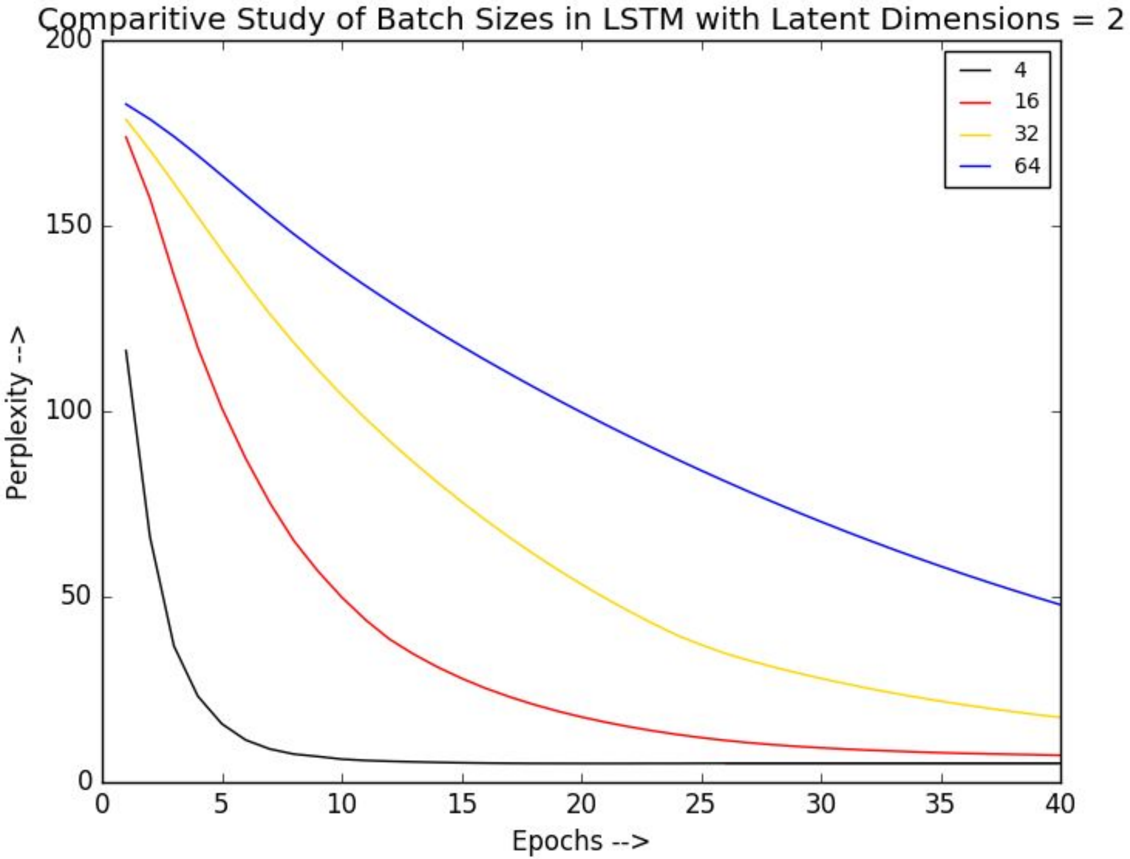


Figure 2. Comparative Study of Batch Sizes in LSTM with Latent Dimensions = 2

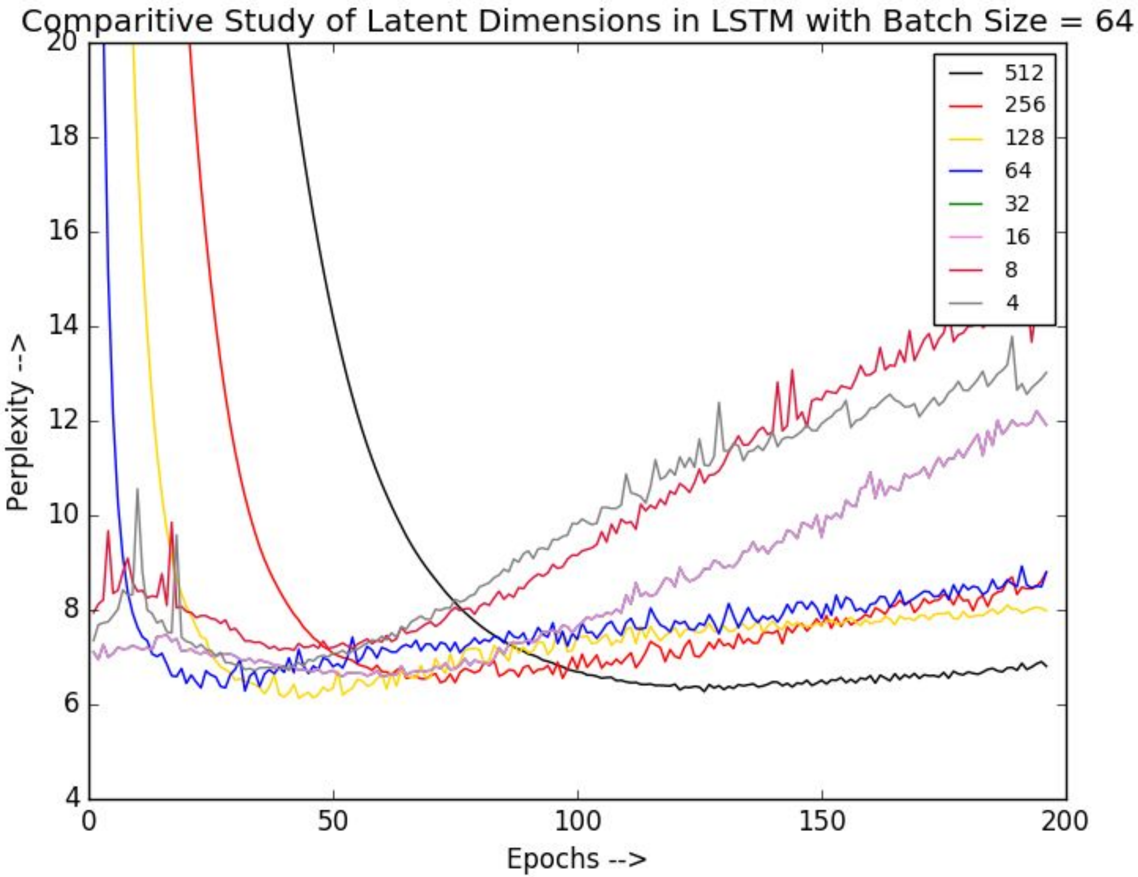


Figure 3. Comparative Study of Latent Dimensions in LSTM with Batch Size = 64

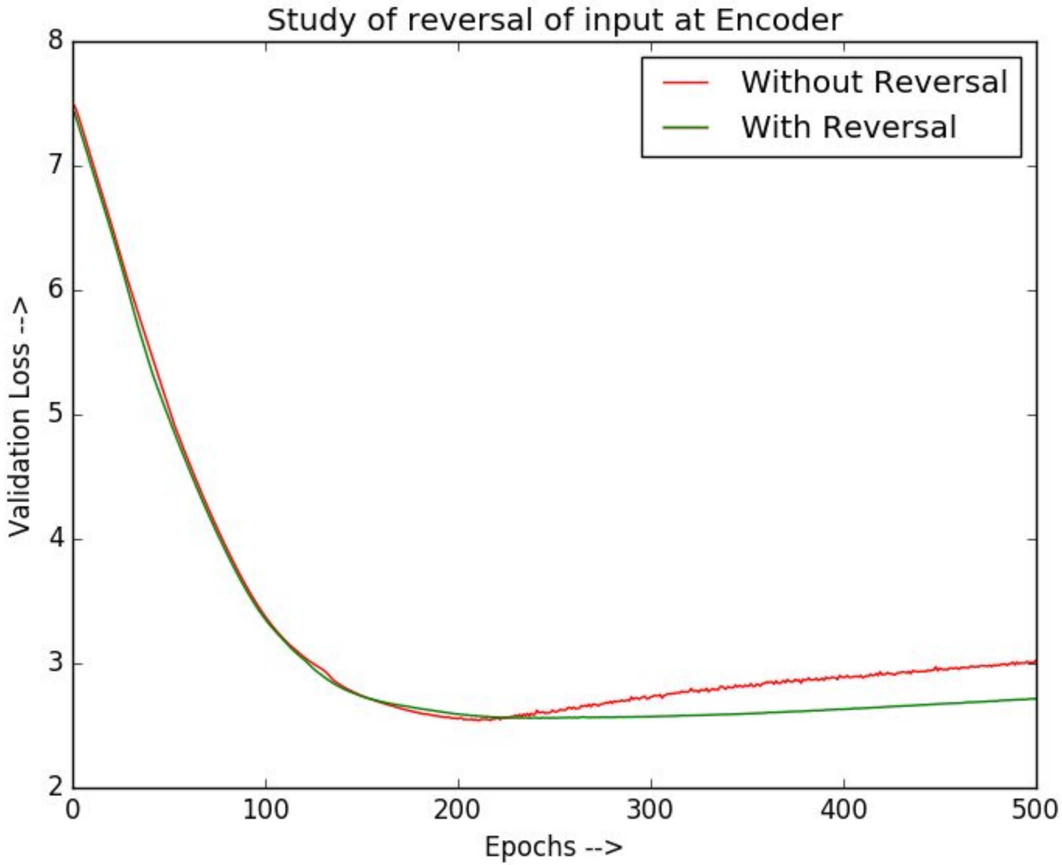


Figure 4. Study of Reversal of Input at the Encoder

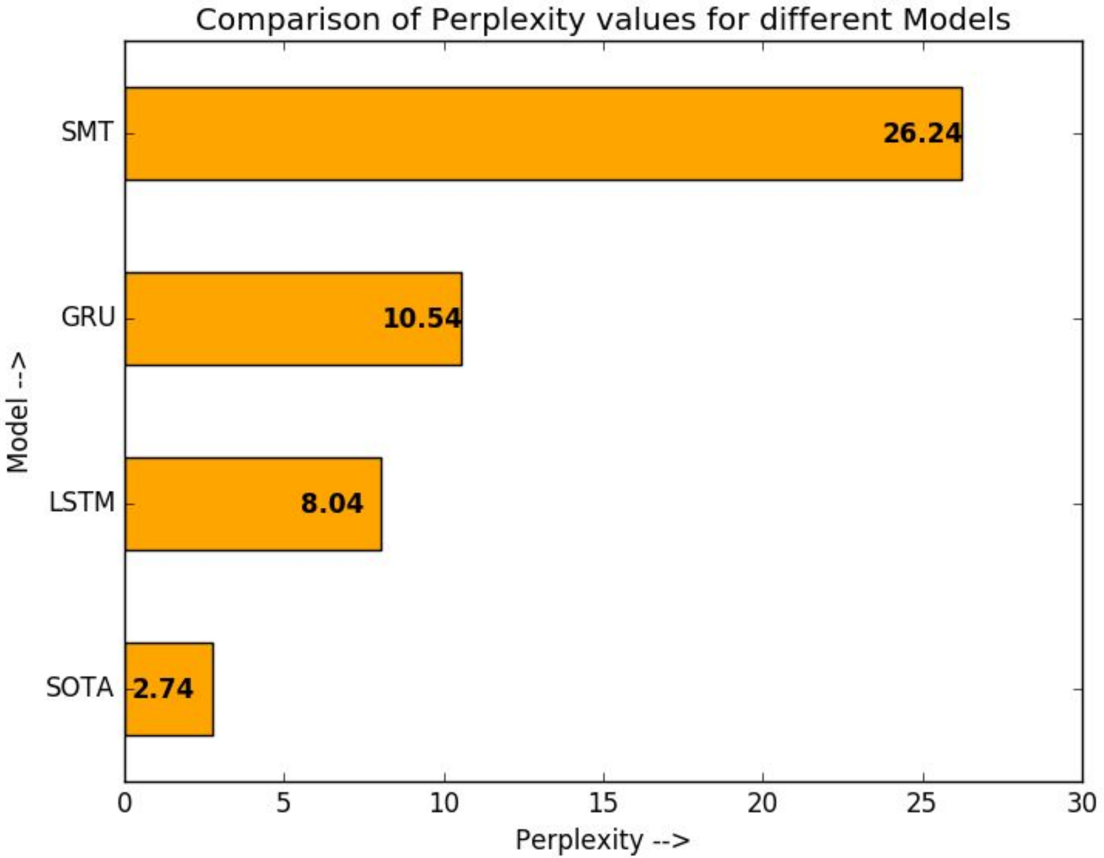


Figure 5. Comparative Study of Different Models

Input Sentence	*seq2seq Model*	*State-of-the-art Model*	*SMT Model*
are you a computer?	what me that	Certainly, Doctor	you re you re
get lost!	no	You got it!	you re
are you a human?	what me that	No, not real	you re you re
who is the president?	no one is it?	Nice,me	you . i m
you are not making sense	i don't know	yeah, I know	. you . i m not
are you drunk?	what me that now please would	I ' m him .	you re not
Hi!	really	Hi!	s
am i a doctor?	was in vsunk work i'm at work u	sure but a can always be a.	you . i m
when will the world end?	wouldn't me	You mean last night would you?	you . i m not
can you teach me something?	this back	what do I do to install ?	you re not . i m not

Figure 6. Table showing the comparison of results obtained on the three different models

thank you!