Development and evaluation of a natural language conversational bot for identifying appropriate clinician referral from patient narratives

IUPUI
INDIANA UNIVERSITY
School of Informatics
and Computing

Sriharsha Tummala, Saptarshi Purkayastha, Ph.D., Josette Jones, Ph.D.

Department of BioHealth Informatics, School of Informatics and Computing, Indiana University Purdue University Indianapolis

Introduction

- Recent years have seen a significant increase in automated conversational agent chatbots. Conversational agents like chatbots for health may provide timely and cost-effective support in clinical care.
- Some studies show that chatbots could have an impact on patient engagement. Additionally, health systems are attempting to connect with patients over social networks, mainly where specialists are limited
- By 2025, the Association of American Medical Colleges estimates that the United States will have a shortfall of 61,700-94,700 physicians and critical shortage in many specialties, delaying available appointments by months in many cases.
- Thus, we need innovative solutions that can manage the time of limited specialists appropriately.
- Recent research has demonstrated that deep-learning methods are superior for natural language classification tasks compared to other machine learning methods.
- The primary objective of this study was to develop a telegram chatbot which reads patient narratives and acts as a conversational agent by redirecting the case to the appropriate specialist.
- Besides simply working on improving conversational capabilities of chatbots, we developed a novel method for referring the cases to specialists based on their responses to previous cases on a social network group.
- As far as we know, no other chatbot has the level of accuracy or referral system like our developed chatbot.

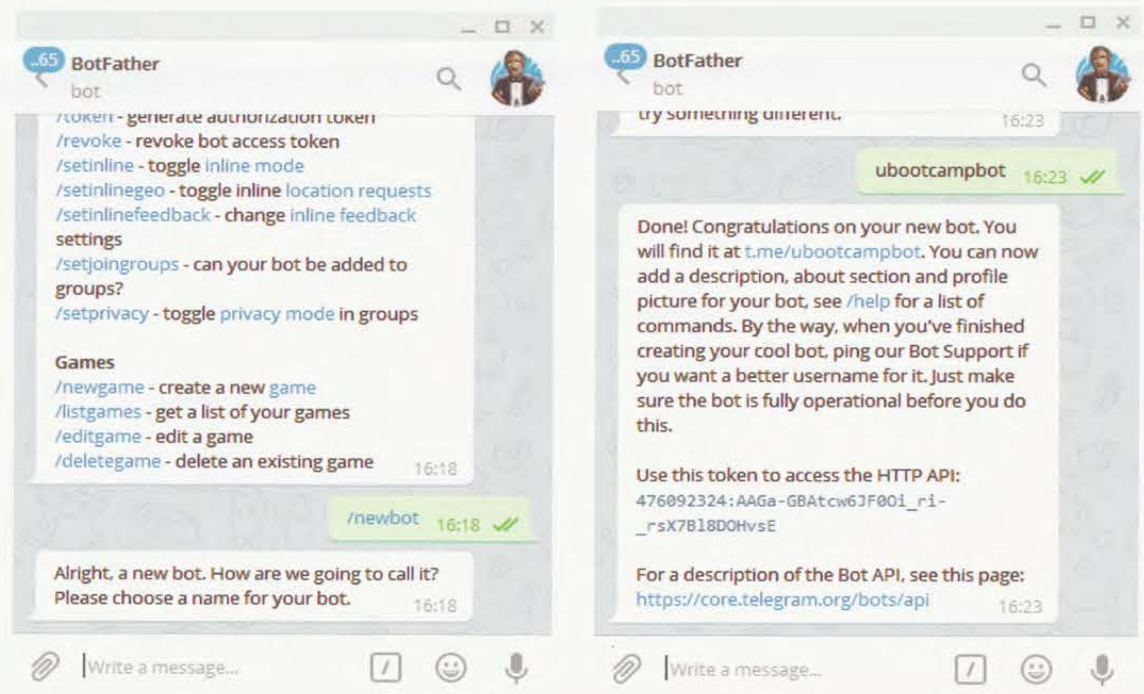
Methodology

- Data is collected from Facebook consists of 1890 clinicians. The data is 2 years old and it is deidentified
- 568 actors identified
- Inclusion Criteria: Must have made at least 1 post or comment
- 1,143 (top level) posts and 8,606 comments made to these posts
- *Commauth* and *Postcontent* are extracted as primary columns from the data set and used for model development. Data was 3x imputed to get appropriate size for training a deep neural network.

	label	text
0	boudhayan.dm	Dr MpSingh, Dr Swagatawhat would you suggest f...
1	durga.prasan	Dr MpSingh, Dr Swagatawhat would you suggest f...
2	durga.prasan	Ideally the rectal polyp should have been snar...
3	durga.prasan	My suggestion is that all cancers should be de...
4	swagata.brahmachari	Ya surely .As correctly pointed out by Durga P...

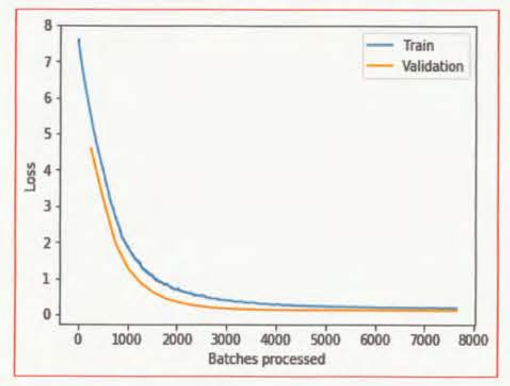
- Deep learning library *fastai* is used for the model development and training.
- Tokenizer was used to tokenize the important vocabulary. Pretrained weights from Wikipedia trained model was downloaded as part of transfer learning.
- Data was divided into training and target using TextDataBunch
- The language model was trained for 30 epochs, which resulted in an accuracy of 97% with a converging training (0.190) and validation loss (0.119).
- The language model creation process resulted in a neural network with an embedding size of 400, 3 layers, 1150 hidden activations per layer.

Software developed



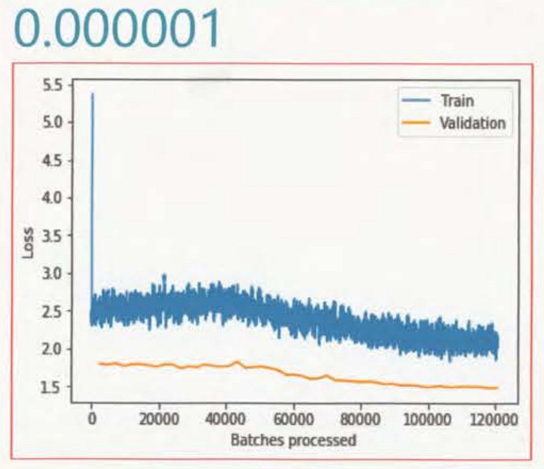
Results

- Losses in language model at 30 epochs

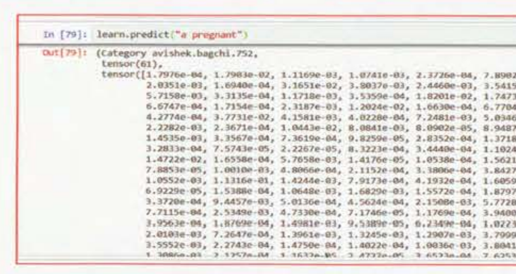


18	0.247654	0.125226	0.973907
19	0.232066	0.123492	0.973918
20	0.219138	0.122613	0.974127
21	0.220072	0.121861	0.974219
22	0.214477	0.120913	0.974325
23	0.203683	0.120074	0.974313
24	0.195483	0.119843	0.974327
25	0.197264	0.120266	0.974234
26	0.196699	0.118628	0.974576
27	0.190719	0.118809	0.974291
28	0.193372	0.118906	0.974326
29	0.189111	0.118985	0.974323
30	0.190446	0.119288	0.974294

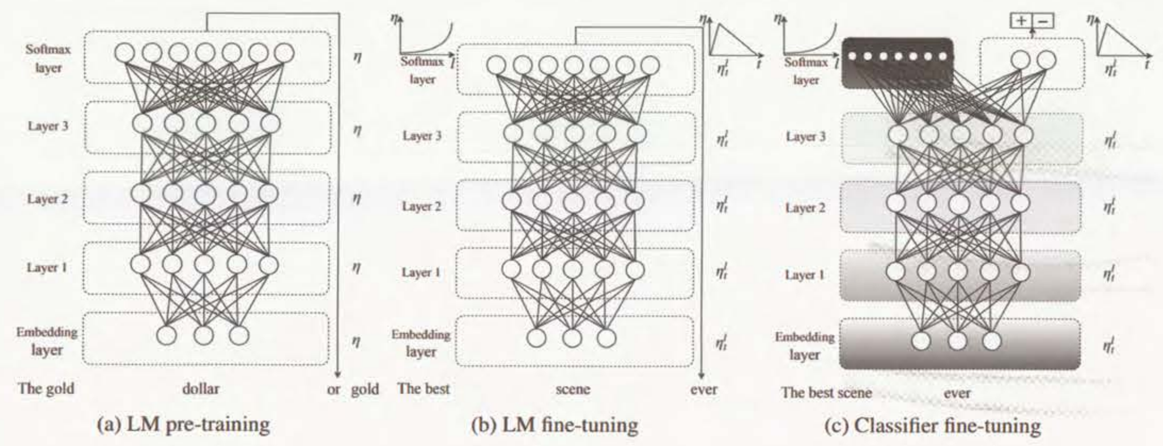
- The pre-trained vocabulary from Wikipedia is used for transfer learning for the NLP using the ULMFiT (Universal Language Model Fine-tuning for Text Classification) [1]
- We then trained a classified using the AWD-LSTM algorithm with a batch size of 16 and with 3 different learning rates from 0.01 to 0.000001



37	2.303326	1.538335	0.582263
38	2.151838	1.517407	0.589308
39	2.178674	1.516321	0.590758
40	2.214718	1.515671	0.587857
41	2.147552	1.494473	0.592623
42	2.185507	1.496596	0.589101
43	2.061003	1.510748	0.590344
44	2.202823	1.492539	0.592623
45	2.226420	1.497136	0.593867
46	2.070127	1.501744	0.591173
47	2.201834	1.499393	0.589930
48	2.257642	1.497376	0.593245
49	2.074065	1.485408	0.592830
50	2.156907	1.489073	0.593867



The ULMFiT transfer learning for NLP classification [1]



- The classifications have a 59% accuracy, but they are close to 100% accurate based on gold-standard human verification.
- Since there are more specialists of the same specialization in the social network, the classifier approximates the specialists, which was then verified manually to be accurate.

Limitations

- Data collected was very low in size to be suitable for Neural networks
- The data consists many of non-English words, images in the posts and typos which poses difficulty in training and developing the model

Future work and conclusion

- To deploy this on patient portals or a Facebook/telegram/WhatsApp group and evaluate the clinician response, and maybe later patient outcomes.
- We believe further research of this approach can help reduce burden, and ease the load of resource limited health systems.

Part of this work was supported through the Nvidia GPU grant to Dr. Purkayastha (2018), and the NSF XSEDE grant for computing resources.

References:

[1] Howard, Jeremy, and Sebastian Ruder. "Universal language model fine-tuning for text classification." In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 328-339. 2018.

[2] H. Wang, Q. Zhang, M. Ip and J. T. Fai Lau, "Social Media–based Conversational Agents for Health Management and Interventions," in Computer, vol. 51, no. 8, pp. 26-33, 2018. doi:10.1109/MC.2018.3191249