Improving SMS Customer Experience through a Transformer-based Chatbot

Aaron Wang & Brian Hsu

Faculty Advisor – Professor Bertsimas | CVS Manager – Amy Wen

CVS Health

Massachusetts Institute of Technology

Motivation

CVS's Refill Reminders program texts patients to refill their prescriptions, but the current model only understands simple keywords such as Yes/No/Help. Messages outside of this predefined dictionary are invalid and are sent a default response.

> Would you like to refill your supply of TYL? Reply YES or NO.
>
> Not yet, I still have some left
>
> Sorry, I don't understand. Please reply YES or NO.

These invalid messages make up ~10% of all incoming texts, and results in unfulfilled refills and unhappy customers. Our objective is to build a chatbot with strong intent detection to vastly improve customer experience.

Challenges

- Though we have hundreds of thousands of texts, they are completely unlabeled
- The texts can have odd and unfavorable structures (e.g., very short texts, people names, medicine names)
- The texts can contain plenty of irrelevant samples (e.g., mis-sent messages, gibberish strings)

10,000,000 Ignored Messages, Annually

Data

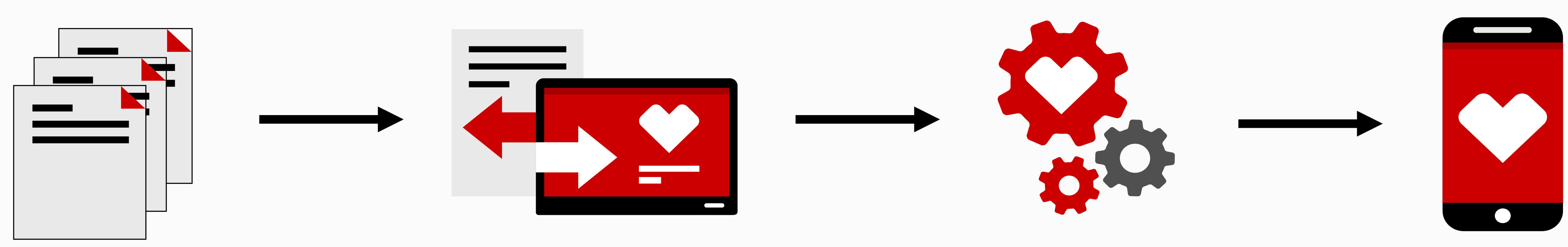
We manually labeled over 3,000 invalid texts using non-ambiguous guidelines, and ultimately determined that there were 22 granular intents. These 22 intents are then combined into 7 types of messages that CVS will respond with.

Why 3,000 messages? Refill Reminders is just one of many SMS programs at CVS. We wanted to label *just enough* data to develop a strong model, but also ensure that future models for other programs can be developed quickly with minimal manual work.

Action	Examples
Yes	yess okay, lisiniprol please
No	Nooope, no ibuprofen
Get Rx Name	tylenol 5mg, what is the name
Reschedule	maybe later, im out of town
Contact Help	u have wrong number, im not with cvs
No Action	okay thank u, driving cant text
Resend Prompt	whats for dinner today, 991k0193k1

Modeling

When designing the chatbot model, we explored several models in recent NLP literature. The best approach in both speed and performance was a custom pre-trained DistilBERT. Additionally, this model is flexible and can be painlessly applied to other CVS programs by only adapting the last step of fine-tuning..



1. We start with a pre-trained DistilBERT model (trained on general web data)
2. Then, we adaptively fine-tune and further pre-train the model on a large corpus of CVS text data (n = 500k)
3. Next, we fine-tune the model for classification on 22 labels
4. Finally, we design a response tree to answer patient queries

Accuracy 84% on 22 Intents **Speed** 100+ Texts/Second

It also beat out several other approaches we tested, including

Unsupervised Techniques: Zero-Shot Learning, SBERT Clustering

Semi-Supervised Techniques: Siamese Neural Network, GAN-BERT

Supervised Techniques: Bag-of-Words Random Forest, SVM