

SageCrypto: A Chatbot that Knows Cryptocurrency

Cheng Shu

University of Missouri, Kansas City

December 14, 2021

Outline

- 1 Overview
- 2 Tech Stack
 - Overview
 - Back-end
 - Front-end
- 3 Demo
 - Limitations and Future Plans

Objectives

What is SageCrypto

A Telegram chatbot based on a finetuned DialoGPT model.

Highlights

- Good understanding of the domain knowledge of cryptocurrency
- Can conduct multi-turn dialogue

The Model

Back-end: Fine-tuned DialoGPT Model

- Trained with Reddit data
- Hosted on HuggingFace

Front-end: Telegram bot

- Flask API
- Hosted on Heroku

Microsoft DialoGPT

- Large-scale Pretrained Response Generation Model
- GPT-2 based
- Pretrained small (117M) medium (345M) and large (762M) model
- Trained on 147M multi-turn dialogue from Reddit discussion thread
- Requires finetuning for domain-specific topics

A Example Dialogue with DialoGPT-large

```
>> User:What do you know about bitcoin?  
CryptoBot: You don't need to know anything about bitcoin to know that it's not a good investment.  
>> User:Why?  
CryptoBot: Because it's a speculative investment. The market is unpredictable.  
>> User:What about ethereum?  
CryptoBot: That's the one I've been looking at.  
>> User:Is it better than bitcoin?  
CryptoBot: I haven't tried it yet.  
>> User:What do you suggest me buy then?  
CryptoBot: Why do you want to buy?
```

My Data

Subreddit Communities

- Cryptocurrency 4.1m members
- ethereum 1.2m
- CryptoMarkets 676k
- CryptoCurrencyNews 121k

Processing

- Pulled from Reddit API using PRAW
- Top 20 of most popular thread from the targeted subreddits
- Cleaned and transformed

Results

Finetuned models

- Up to six previous comments and the pinned title used as context
- Finetuning with large model is infeasible with Google Colab
- Two models trained: small and medium

Hugging Face Host

- Hugging Face model repository to host ML models
- Tag as conversational
- Inference API for easy access to any hosted model

Telegram

de facto official IM

- Encrypted messenger service
- Popular among cryptocurrency communities for communication and marketing
- Easy API for 3rd party bot developers

BotFather

- Basic info
- Basic functionality

Flask API

telegram

- code endpoint
- Credential management

huggingface hub

- connects hugging face models
- response in json

Demo

Limitations

Training is limited by Colab

- Limited GPU resources
- Data needs more cleaning

fine-tuned model is unstable

- Stupid looping
- unpredictable behavior locally vs remotely
- Too crypto oriented
- slow to respond
- About 70% of the time has logic

Future

Integrate pricing info

Telegram channel manager

Other types of bot

- Discord
- Whatsapp