

TRƯỜNG ĐẠI HỌC BÁCH KHOA

KHOA KHOA HỌC & KỸ THUẬT MÁY TÍNH



LUẬN VĂN TỐT NGHIỆP

Phát triển công cụ dự đoán xu hướng
giá ngắn hạn các đồng tiền mật mã
bằng kỹ thuật học máy

SVTH:

Vũ Quang Nam

GVHD:

Nguyễn An Khương

Nguyễn Lê Thành

Ngày 15 tháng 5 năm 2019

Mục lục

Danh mục hình vẽ	4
1 Giới thiệu	1
1.1 Giới thiệu đề tài nghiên cứu	1
1.2 Mục tiêu và phạm vi đề tài	1
1.2.1 Mục tiêu	1
1.2.2 Phạm vi đề tài	1
1.3 Tiến độ thực hiện	2
2 Tổng quan về lĩnh vực nghiên cứu	5
2.1 Những yếu tố tác động đến giá trị đồng tiền mã hóa	5
2.1.1 Cung và cầu của thị trường	5
2.1.2 Tin tức trên các phương tiện thông tin đại chúng	5
2.1.3 Quy định của chính phủ	5
2.1.4 Chính sách của các tổ chức	6
2.1.5 Các vấn đề kỹ thuật	6
2.2 Nhu cầu sử dụng tiền mã hoá của mỗi hệ sinh thái	6
3 Các khái niệm cơ bản có liên quan tới đề tài	7
3.1 Các khái niệm về tài chính	7
3.1.1 Tính thanh khoản (Liquidity)	7
3.1.2 Nhiễu (Noise)	8
3.2 Các khái niệm về xác suất	8
3.2.1 Hàm mật độ xác suất (Probability density function)	8
3.2.2 Hàm phân phối biên (Marginal distribution)	9
3.2.3 Nhiễu dữ liệu	10
3.2.4 Biến ẩn (Latent variable)	10
3.2.5 Phương pháp suy luận biến phân (Variational Inference)	10

Danh mục hình ảnh

3.1 Phân phối biên giá mở/đóng dữ liệu đã xử lý	9
---	---

Chương 1

Giới thiệu

Giới thiệu đề tài nghiên cứu

Hiện nay, tiền mã hóa đã trở nên phổ biến, đa dạng với nhiều sàn giao dịch khác nhau. Đồng tiền mã hóa có tỷ giá thay đổi theo thời gian, việc tìm xu hướng giá ngắn hạn tính theo giờ phút không bị ảnh hưởng nhiều bởi các yếu tố bên ngoài như các dự báo, các quy định của chính phủ. Từ dữ liệu cụ thể là tổng hợp của các giao dịch trên các sàn trực tuyến việc tìm ra một giải thuật có thể dự đoán xu hướng giá của các giao dịch tiếp theo với nguyên tắc đề cao khách quan so với kinh nghiệm bản thân là một vấn đề mới mẻ. Vậy nên tôi quyết định chọn đề tài **Dự đoán xu hướng giá ngắn hạn các đồng tiền mật mã bằng kĩ thuật học máy**.

Mục tiêu và phạm vi đề tài

Mục tiêu

Mục tiêu của luận văn này là xây dựng một công cụ dự đoán xu hướng giá ngắn hạn các đồng tiền mật mã bằng kĩ thuật học máy. Dữ liệu đầu vào là các thông tin về lịch sử giá các đồng tiền ảo trong các phiên giao dịch.

Phạm vi đề tài

- Tìm hiểu và nghiên cứu về lý thuyết học máy thống kê (statistical machine learning)

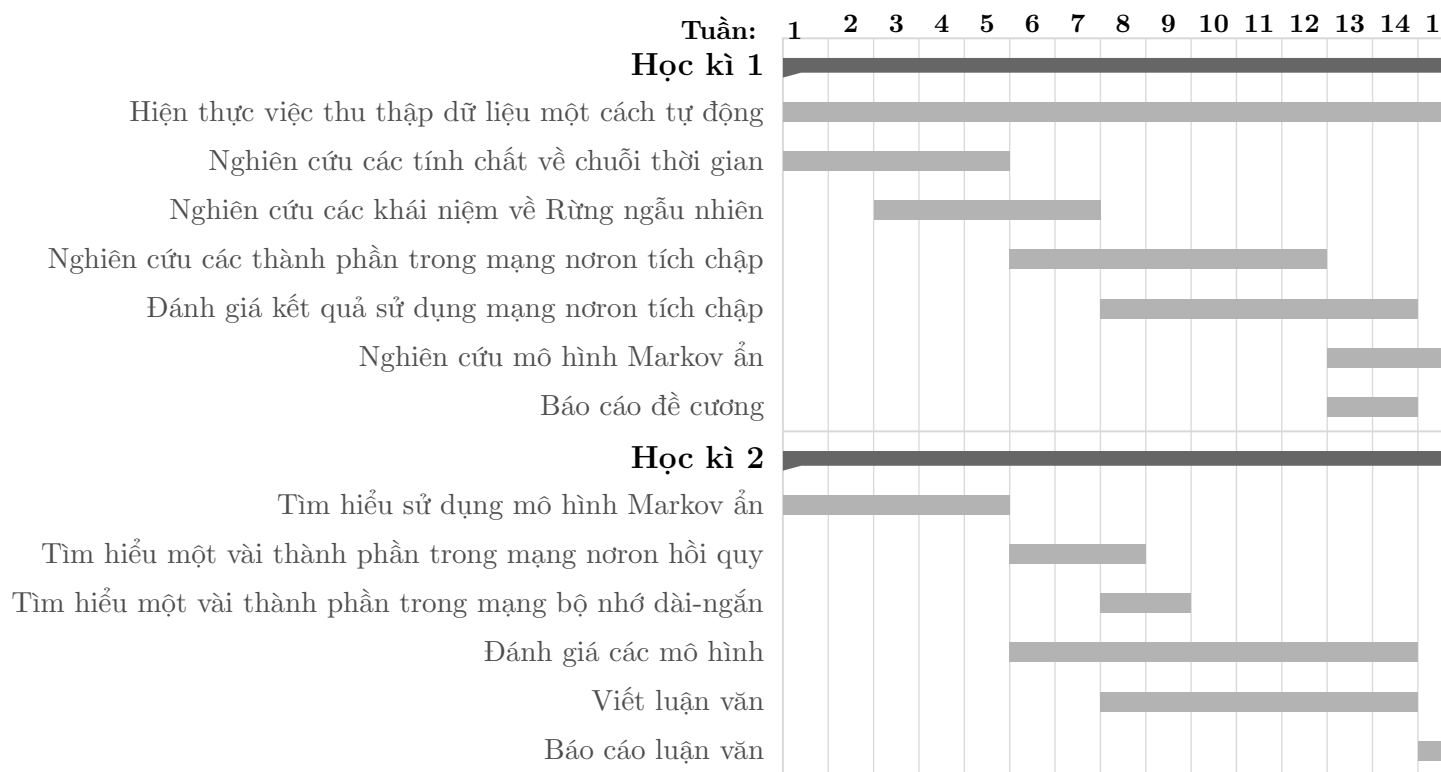
- Xây dựng mô hình dự đoán về xu hướng tăng giảm, dự đoán giá của các đồng trong thời gian ngắn hạn.

Các đối tượng nghiên cứu trong đề tài:

- Tìm hiểu một vài loại đồng tiền mã hóa, và các sàn giao dịch.
- Một vài tài liệu liên quan tới lý thuyết thống kê hiện đại.
- Tìm hiểu một vài mô hình trong học máy: hồi quy logistic, rừng ngẫu nhiên, mạng nơron.
- Sử dụng ngôn ngữ Python, R và một số thư viện để hiện thực mô hình.
- Xây dựng công cụ dự đoán giá một cách tự động.

Tiến độ thực hiện

Trong phần này, tác giả xin trình bày lịch trình công việc đã thực hiện đề tài trong học kỳ I và lịch trình dự kiến hiện thực đề tài trong quá trình làm luận văn chính thức ở học kỳ II dưới dạng biểu đồ Gantt sau đây.



Chương 2

Tổng quan về lĩnh vực nghiên cứu

Những yếu tố tác động đến giá trị đồng tiền mã hóa

Cung và cầu của thị trường

Trong nguyên tắc chính của kinh tế nếu người ta mua một đồng tiền, giá trị của đồng tiền sẽ tăng lên và nếu người ta bán đồng tiền, giá sẽ giảm.

Tin tức trên các phương tiện thông tin đại chúng

Các sự kiện chính trị và kinh tế trên toàn thế giới ảnh hưởng đến cách mà con người phản ứng với các dự đoán giá, tin tức cảnh báo về rủi ro tác động chính lên cung-cầu.

Quy định của chính phủ

Có 4 cấp độ quản lý tiền ảo hiện nay đang được các nước thực thi, cụ thể:

- Cấm trên diện rộng.
- Cấm trong lĩnh vực tài chính ngân hàng (trong đó có Trung Quốc, Nga).
- Cảnh báo rủi ro đối với người sử dụng, đầu tư.
- Chấp nhận như một phương tiện thanh toán (các nước chấp nhận đồng bitcoin gồm có Mỹ, Canada, Úc, Liên minh châu Âu, Phần Lan [2]).

Chính sách của các tổ chức

Facebook, Google và Twitter đã ngăn chặn khách hàng và người dùng sử dụng dịch vụ cryptocurrency.

Các vấn đề kỹ thuật

Vì đồng tiền mã hóa có thể bị hack thành công vào tài khoản hoặc tấn công máy chủ, có thể làm giảm tỷ giá hối đoái, dẫn đến giá giảm.

Nhu cầu sử dụng tiền mã hoá của mỗi hệ sinh thái

- Số thành viên tham gia vào hệ sinh thái (Số người đến khu vui chơi mua vé tham gia các trò chơi trong đó bằng tiền A).
- Số lượng dịch vụ trong hệ sinh thái (Khu vui chơi có càng nhiều trò chơi thì nhu cầu sử dụng tiền A càng tăng); Và các nền tảng như Ethereum luôn mở cho các đối tác tạo các dịch vụ gia tăng trên đó giống như khu vui chơi cho phép đối tác bên ngoài vào tổ chức trò chơi ở trong.
- Số người đầu cơ: Những người nhận thấy nhu cầu tiền mã hoá của một hệ sinh thái tăng dần sẽ mua để nắm giữ chờ tăng giá thì bán ra. (Giống như phe vé bóng đá ngày trước mua vé chờ sát trận nhu cầu tăng vọt thì bán ra. Khu vui chơi thì ít có nhóm này vì lượng vé không bị giới hạn).
- Số người bán bên ngoài chấp nhận tiền mã hoá: Một số người bán nhận thấy tính thanh khoản của tiền mã hoá và giá trị tăng dần của nó nên đã chấp nhận khách hàng thanh toán các hàng hoá dịch vụ của mình bằng loại tiền này (Nhà hàng bên cạnh khu vui chơi có thể chấp nhận khách hàng thanh toán bằng tiền A).

Chương 3

Các khái niệm cơ bản có liên quan tới đề tài

Các khái niệm về tài chính

Tính thanh khoản (Liquidity)

Khái niệm về tính thanh khoản dùng để chỉ mức độ mà một tài sản có thể được mua hoặc bán trên thị trường mà không làm ảnh hưởng nhiều đến giá thị trường. Khái niệm tính thanh khoản được chia thành 2 loại: tính thanh khoản thị trường (liquid market) và tính thanh khoản về tài sản (liquid asset). Thị trường có tính thanh khoản cao ám chỉ rằng trong thị trường thường xuyên có các nhà đầu tư sẵn sàng giao dịch. Một tài sản có tính thanh khoản cao đồng nghĩa với việc tài sản đó có thể chuyển đổi sang tiền mặt một cách dễ dàng. Đối với thị trường tiền mã hóa, để so sánh tính thanh khoản giữa các sàn trong cùng một thời điểm hoặc tính thanh khoản của một sàn tại những thời điểm khác nhau có 3 yếu tố quan trọng:

- Lượng đồng giao dịch trong ngày.
- Số lượng lệnh mua/bán dựa trên danh sách lệnh (order book) được công khai dựa theo các sàn như Coinbase Pro [3], Binance, Bittrex, ...
- Lượng chênh lệch giữa giá yêu cầu của bên bán và giá bỏ thầu của bên mua (bid/ask spread).

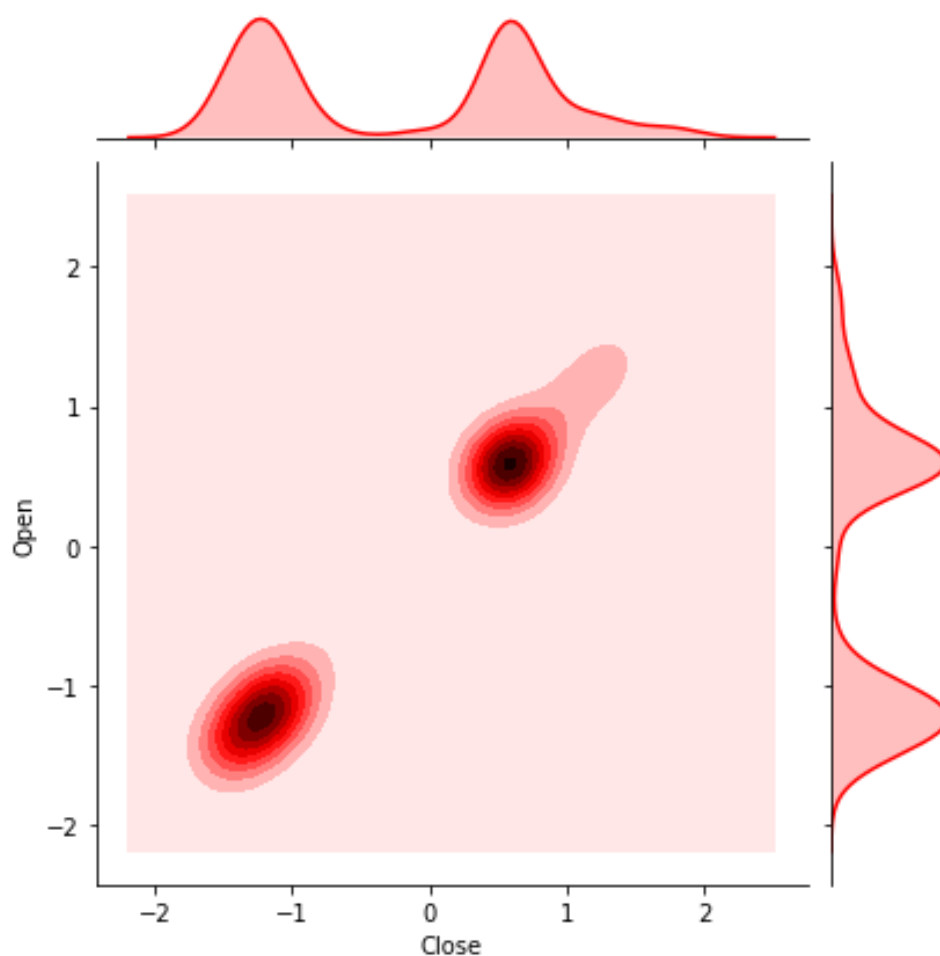
Nhiều (Noise)

Khái niệm nhiễu có quan hệ đối lập với khái niệm thông tin (information) với dữ liệu giá cung cấp đầy đủ thông tin, việc dự đoán dễ dàng và ngược lại với dữ liệu có nhiễu cao do bị ảnh hưởng bởi các yếu tố khác như phần đề cập tại phần 2.1. Nhiễu khiến những quan sát của các nhà đầu tư không được hoàn hảo, điều này dẫn thị trường có khả năng lưu động[noise-in-finance].

Các khái niệm về xác suất

Hàm mật độ xác suất (Probability density function)

Với các khối nền có các thành phần như giá mở, giá đóng, số lượng đồng giao dịch,..., ta có thể coi như các biến ngẫu nhiên liên tục tương ứng. Khái niệm hàm mật độ xác suất trong văn cảnh trên được hiểu như một hàm gồm các tham số thể hiện được mật độ phân bố của các biến ngẫu nhiên.



HÌNH 3.1 – Phân phối biên giá mở/đóng dữ liệu đã xử lý

Hình 3.1 thể hiện mật độ của phân phối đồng thời giữa giá đóng và giá mở của các khối nên được biểu diễn dưới dạng $p_{data}(Open, Close)$.

Hàm phân phối biên (Marginal distribution)

Với dữ liệu liên tục như trên, hàm phân phối biên đối với giá mở được biểu diễn dưới dạng:

$$p_{data}(Open) = \int_y p_{data}(Open, Close = y) dy = \int_y p_{data}(Open | Close = y) p_{data}(Close = y) dy$$
 Một cách trực quan, hàm phân phối trên được biểu diễn bởi đường biên bên trái Hình 3.1

Nhiều dữ liệu

Nhiều trong dữ liệu như phần đề cập trong phần 3.1.2 có thể được giảm thiểu bằng cách tìm hàm phân phối của nhiễu bằng thống kê, nếu phân phối của nhiễu có dạng phân phối chuẩn với trung bình là 0, nhiễu này được gọi là nhiễu trắng Gauss (white Gaussian noise). Việc giảm thiểu nhiễu trong dữ liệu làm mô hình trở nên dễ tìm được mẫu đặc trưng (pattern) hơn.

Biến ẩn (Latent variable)

Biến ẩn được hiểu theo cách trừu tượng là biến không thể quan sát trực tiếp[bishop] mà được suy luận từ biến quan sát được trong dữ liệu. Cụ thể hơn, với dữ liệu là giá của 100 ngày đầu một mô hình có khả năng tìm được quan hệ giữa giá ngày thứ 101 phụ thuộc nhiều vào giá ngày thứ 99 hơn so với ngày thứ 1, mô hình này được gọi là mô hình biến ẩn (latent variable model).

Phương pháp suy luận biến phân (Variational Inference)

Tài liệu tham khảo

- [1] Michael I. Jordan Andrew Y. Ng. “On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes”. In: *NIPS* (2001). URL: <https://papers.nips.cc/paper/2020-on-discriminative-vs-generative-classifiers-a-comparison-of-logistic-regression-and-naive-bayes.pdf>.
- [2] Prableen Bajpai. *Countries Where Bitcoin Is Legal Illegal*. <https://www.investopedia.com/articles/forex/041515/countries-where-bitcoin-legal-illegal.asp>. cited May 2019.
- [3] Trade Volume. *Cryptometer Live Order Book*. https://www.cryptometer.io/data/coinbase_pro/btc/usd. cited April 2019.