

Tận dụng Mô hình Ngôn ngữ Lớn cho Dự báo Chuỗi Thời gian Tài chính Có thể Giải thích Được

Nhóm: Seminar

Giới thiệu các thành viên trong nhóm

- Vũ Văn Tới - 23020427 - Trưởng nhóm
- Nguyễn Năng Thịnh - 23020439 - Thành viên
- Mai Phan Anh Tùng - 23020433 - Thành viên
- Phan Mạnh Quân - 23020419 - Thành viên

Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội
Viện Trí tuệ nhân tạo
Môn học: Seminar khoa học

Mục lục

TÓM TẮT	3
CÁC KHÁI NIỆM CCS	4
1 GIỚI THIỆU	5
2 CÁC NGHIÊN CỨU LIÊN QUAN	7
2.1 Phương pháp Thống kê/Kinh tế lượng Truyền thống	7
2.2 Kỹ thuật Học máy	7
2.3 Mô hình Ngôn ngữ Lớn (LLMs)	7
3 PHƯƠNG PHÁP LUẬN	9
3.1 Dữ liệu	9
3.1.1 Dữ liệu Giá Cổ phiếu	9
3.1.2 Dữ liệu Hồ sơ Công ty	9
3.1.3 Dữ liệu Tin tức Tài chính/Kinh tế	10
3.2 Suy luận Zero-shot/Few-shot Dựa trên Chỉ dẫn với LLMs	10
3.3 Huấn luyện Tinh chỉnh Dựa trên Chỉ dẫn với Open LLaMA	10
4 THÍ NGHIỆM	14
4.1 Cài đặt thí nghiệm	14
4.1.1 Khoảng thời gian dữ liệu	14
4.1.2 Mô hình cơ sở (Baseline Models)	14
4.1.3 Chỉ số đánh giá (Evaluation Metrics)	15
4.2 Đánh giá hiệu suất	15
4.2.1 Dự báo giá cổ phiếu	15
4.2.2 Chất lượng giải thích	16
5 KẾT LUẬN	17
TÀI LIỆU THAM KHẢO	18

TÓM TẮT

Bài báo này trình bày một nghiên cứu mới về việc tận dụng khả năng hiểu biết và lý luận vượt trội của các Mô hình Ngôn ngữ Lớn (LLMs) để dự báo chuỗi thời gian tài chính một cách có thể giải thích được. Việc ứng dụng các mô hình học máy vào chuỗi thời gian tài chính gặp nhiều thách thức, bao gồm khó khăn trong việc suy luận và diễn giải thông tin xuyên suốt chuỗi, trở ngại trong việc tích hợp tín hiệu đa phương thức từ tin tức lịch sử, đồ thị tri thức tài chính, v.v. . . , cũng như vấn đề trong việc giải thích và minh bạch hóa kết quả mô hình. Trong nghiên cứu này, chúng tôi tập trung vào các cổ phiếu thuộc chỉ số NASDAQ-100, sử dụng dữ liệu giá cổ phiếu lịch sử công khai, siêu dữ liệu công ty và các tin tức kinh tế/tài chính trong quá khứ. Chúng tôi tiến hành các thí nghiệm để minh họa tiềm năng của LLM trong việc cung cấp một giải pháp thống nhất cho các thách thức nói trên. Các thí nghiệm bao gồm việc thử nghiệm suy luận zero-shot/few-shot với GPT-4 và huấn luyện theo hướng dẫn (instruction-based fine-tuning) với mô hình LLM mã nguồn mở Open LLaMA. Chúng tôi cho thấy phương pháp của mình vượt trội hơn một số mô hình cơ bản, bao gồm mô hình ARMA-GARCH cổ điển và mô hình cây tăng cường gradient (gradient-boosting tree). Thông qua kết quả so sánh hiệu suất và một vài ví dụ, chúng tôi nhận thấy rằng LLM có thể đưa ra quyết định một cách có lý lẽ bằng cách suy luận từ cả tin tức văn bản và chuỗi thời gian giá, trích xuất thông tin chuyên sâu, tận dụng thông tin xuyên chuỗi và khai thác kiến thức sẵn có được tích hợp bên trong LLM. Ngoài ra, chúng tôi cũng cho thấy rằng một LLM mã nguồn mở như Open LLaMA, sau khi được tinh chỉnh theo hướng dẫn, có thể hiểu yêu cầu để tạo ra các dự báo có thể giải thích và đạt được hiệu suất hợp lý, mặc dù vẫn thấp hơn so với GPT-4.

CÁC KHÁI NIỆM CCS

- Phương pháp tính toán → Phân tích chuỗi thời gian; Xử lý ngôn ngữ tự nhiên;
- Trí tuệ nhân tạo có khả năng giải thích.

Chương 1

GIỚI THIỆU

Sự phát triển nhanh chóng của công nghệ Học máy (ML) và Trí tuệ nhân tạo (AI) trong vài năm gần đây đã mở ra nhiều cơ hội cũng như thách thức trong nhiều lĩnh vực khác nhau, bao gồm cả thị trường tài chính. Đặc biệt, nhiệm vụ dự báo chuỗi thời gian tài chính – một yếu tố then chốt trong việc ra quyết định chiến lược và xây dựng chính sách – đã chứng kiến nhiều đổi mới công nghệ, từ các kỹ thuật thống kê/kinh tế lượng cổ điển đến học máy và học sâu. Tuy nhiên, bất chấp những tiến bộ này, việc triển khai các mô hình ML/AI trong tài chính vẫn gặp phải một số thách thức cố hữu.

Một trong những thách thức nằm ở khả năng suy luận và diễn giải xuyên chuỗi – một khía cạnh quan trọng để hiểu các mẫu thời gian và đưa ra dự đoán chính xác. Các phương pháp hiện tại bao gồm phân tích tương quan chuỗi thời gian và phân cụm. Gần đây, học sâu đã được sử dụng để học các mối quan hệ ẩn phức tạp giữa các chuỗi thời gian. Tuy vậy, các phương pháp hiện có vẫn chưa nắm bắt hiệu quả được các phụ thuộc phức tạp đặc trưng của dữ liệu chuỗi thời gian. Ngoài ra, sự khác biệt trong thiết kế, triển khai và yêu cầu dữ liệu của các phương pháp này cũng tạo ra rào cản cho việc ứng dụng rộng rãi trong lĩnh vực tài chính.

Một trở ngại đáng kể khác là xử lý dữ liệu thời gian tài chính đa phương thức phức tạp – vốn vượt xa các chuỗi số đơn thuần. Dữ liệu có thể bao gồm nhiều nguồn khác nhau như tin tức lịch sử, đồ thị tri thức tài chính, hoạt động mạng xã hội và các chỉ báo thị trường khác. Gần đây đã có một số nỗ lực sử dụng suy luận thống kê, mạng RNN/CNN với nhúng văn bản, mạng nơ-ron đồ thị, v.v. để tích hợp thông tin phức tạp này.

Cuối cùng nhưng không kém phần quan trọng, vấn đề về khả năng diễn giải và minh bạch là một thách thức lớn đối với độ tin cậy của các mô hình học máy và học sâu. Phần lớn các mô hình học sâu hiện nay hoạt động như “hộp đen”, gần như không cung cấp thông tin gì về quá trình ra quyết định. Sự thiếu minh bạch này gây ra lo ngại về độ tin cậy của kết quả, đặc biệt trong lĩnh vực nhạy cảm như tài chính – nơi liên quan đến các khoản đầu tư và tài sản lớn. Một số nghiên cứu gần đây đã cố gắng hiểu dự đoán từ mô hình học sâu thông qua điểm chú ý (attention score), nhưng những thông tin đó vẫn chưa dễ hiểu đối với con người và đòi hỏi nhiều công sức diễn giải.

Sự phát triển gần đây của các Mô hình Ngôn ngữ Lớn (LLMs) mang đến cho chúng ta một công cụ mạnh mẽ để giải quyết đồng thời tất cả những thách thức nói trên một cách thống nhất và linh hoạt. Thứ nhất, LLM có thể học được các mối quan hệ phức tạp giữa các chuỗi. LLM hiện là những mô hình dựa trên Transformer mạnh mẽ nhất, và đã có rất nhiều nghiên cứu cho thấy các mô hình Transformer có khả năng học các mối liên hệ phức tạp trong các chuỗi văn

bản và giải các bài toán định lượng. Do đó, có cơ sở để kỳ vọng rằng LLM có thể hiểu được các phụ thuộc phức tạp giữa các chuỗi thời gian số kết hợp với dữ liệu văn bản theo thời gian. Thứ hai, LLM thể hiện khả năng suy luận vượt trội trên dữ liệu đa phương thức. Về bản chất, LLM được huấn luyện để học từ nhiều loại dữ liệu khác nhau, bao gồm cả dữ liệu số, văn bản và dữ liệu có cấu trúc. Nhờ được huấn luyện trên kho văn bản khổng lồ từ internet, LLM có thể xử lý dữ liệu đa dạng – rất phù hợp cho dự báo tài chính, nơi thông tin quan trọng thường đến từ các nguồn khác nhau như dữ liệu thị trường dạng số, bài báo tài chính, hay bài đăng mạng xã hội. Cuối cùng, LLM là những “người giải thích tự nhiên” – có thể tạo ra các lời giải thích dễ hiểu cho con người để cung cấp cái nhìn sâu sắc về quyết định mà mô hình đưa ra. Một ưu điểm chính của LLM là khả năng sinh văn bản tự nhiên mạch lạc, theo ngữ cảnh và toàn diện. Điều này cho phép chúng đưa ra các giải thích dễ hiểu cho con người. Thêm vào đó, thông qua kỹ thuật “chuỗi suy nghĩ” (Chain-of-Thought – COT), LLM còn có thể tạo ra các suy luận chi tiết theo từng bước để tiết lộ quá trình ra quyết định.

Các đóng góp chính của bài báo được tóm tắt như sau: Bài báo này khám phá một cách tiếp cận mới để nghiên cứu tiềm năng của LLM trong nhiệm vụ quan trọng là dự báo chuỗi thời gian tài chính có thể giải thích được. Cụ thể, chúng tôi tập trung vào chuỗi giá cổ phiếu NASDAQ-100. Theo hiểu biết của chúng tôi, hiện chưa có nghiên cứu công khai nào về chủ đề này. Chúng tôi thực hiện thử nghiệm với sự kết hợp giữa kỹ thuật suy luận zero-shot/few-shot trên mô hình AI tiên tiến nhất hiện nay là GPT-4, và kỹ thuật tinh chỉnh theo hướng dẫn sử dụng Open LLaMA. Kết quả thực nghiệm cũng cho thấy kỹ thuật “chuỗi suy nghĩ” giúp cải thiện hiệu suất trong hầu hết các thử nghiệm. Chúng tôi so sánh phương pháp tiếp cận dựa trên LLM với các phương pháp hiện có, bao gồm mô hình ARMA-GARCH và mô hình cây tăng cường (gradient boosting tree). Chúng tôi chứng minh rằng ngay cả suy luận zero-shot bằng GPT-4 cũng có thể vượt qua mô hình boosting-tree với khoảng 300 đặc trưng (features).

Chương 2

CÁC NGHIÊN CỨU LIÊN QUAN

Lĩnh vực dự báo chuỗi thời gian tài chính đã là đối tượng nghiên cứu sâu rộng, với nhiều phương pháp khác nhau đã được đề xuất qua các năm. Trong khi các phương pháp thống kê truyền thống và kỹ thuật học máy đã đóng góp đáng kể vào lĩnh vực này, sự ra đời của các mô hình ngôn ngữ lớn (LLMs) mang đến tiềm năng mới và quan trọng.

2.1 Phương pháp Thống kê/Kinh tế lượng Truyền thống

Các phương pháp thống kê/kinh tế lượng truyền thống từ lâu đã là nền tảng trong dự báo chuỗi thời gian tài chính. Các kỹ thuật như mô hình ARMA-GARCH được sử dụng rộng rãi vì khả năng mô hình hóa sự phụ thuộc và cụm biến động trong chuỗi thời gian tài chính. Các mô hình này đã được mở rộng và điều chỉnh theo nhiều cách để nắm bắt tốt hơn sự phức tạp của thị trường tài chính. Các phương pháp thống kê/kinh tế lượng phổ biến khác bao gồm: mô hình Vector Tự hồi quy (VAM) [73], mô hình không gian trạng thái và bộ lọc Kalman [12], mô hình khuếch tán [6], mô hình Vector Hiệu chỉnh Sai số (VECM) [27], và mô hình Cân bằng tổng thể ngẫu nhiên động (DSGE) [54].

2.2 Kỹ thuật Học máy

Với sự phát triển của học máy, nhiều mô hình đã được áp dụng cho dự báo tài chính. Các cây quyết định, máy vector hỗ trợ (SVM), v.v. đã được nghiên cứu tích cực trong dự báo chuỗi thời gian tài chính. Gần đây hơn, các kỹ thuật học sâu như Mạng nơ-ron hồi tiếp (RNN), Mạng nơ-ron tích chập (CNN), và mô hình Transformer đã được áp dụng cho nhiệm vụ này, cho thấy khả năng nắm bắt các mối quan hệ phi tuyến phức tạp trong dữ liệu.

2.3 Mô hình Ngôn ngữ Lớn (LLMs)

Sự phát triển gần đây của các mô hình ngôn ngữ lớn (LLMs) đã mở ra nhiều khả năng mới cho dự báo chuỗi thời gian tài chính. Các LLMs như GPT-3 [5] và GPT-4 [43], LLaMA [58] (bao gồm Alpaca [57], Vicuna [10]) thể hiện khả năng xuất sắc trong suy luận và hiểu các mối phụ thuộc phức tạp trong dữ liệu không đồng nhất, cũng như khả năng tạo ra lời giải thích dễ hiểu

cho con người. Tuy nhiên, việc ứng dụng LLM vào dự báo chuỗi thời gian tài chính có kèm theo giải thích vẫn là một lĩnh vực còn mới mẻ, và bài báo này nhằm đóng góp vào lĩnh vực đang nổi lên này.

Chương 3

PHƯƠNG PHÁP LUẬN

Trong nghiên cứu này, chúng tôi tập trung vào chuỗi thời gian giá cổ phiếu NASDAQ-100, được bổ sung bởi dữ liệu metadata về công ty và các tin tức tài chính có liên quan đến cả cổ phiếu cụ thể và bối cảnh tài chính/kinh tế rộng hơn. Mục tiêu chính của chúng tôi là dự báo lợi suất cổ phiếu hàng tuần/hàng tháng (được định nghĩa là phần trăm thay đổi giá cổ phiếu từ đầu đến cuối tuần/tháng) kèm theo giải thích. Trọng tâm này phù hợp với thể mạnh của các Mô hình Ngôn ngữ Lớn (LLMs). Chúng tôi trình bày thiết kế có cấu trúc của prompt cho LLMs và áp dụng mô hình GPT-4 hiện đại [44] cho các tác vụ suy luận zero-shot và few-shot. Đối với huấn luyện tinh chỉnh, chúng tôi sử dụng mô hình mã nguồn mở Open LLaMA [18]. Chúng tôi cũng áp dụng kỹ thuật “Chuỗi Tư duy” (Chain of Thoughts - COT) [38, 64], vốn đã được chứng minh là cải thiện hiệu quả của LLMs trong các nghiên cứu khác.

3.1 Dữ liệu

3.1.1 Dữ liệu Giá Cổ phiếu

Chúng tôi tải dữ liệu giá cổ phiếu hàng ngày NASDAQ-100 từ Yahoo Finance bằng gói yfinance. Trong nghiên cứu này, chúng tôi đầu tiên chuẩn hóa chuỗi thời gian giá số thành chuỗi thời gian thay đổi phần trăm, sau đó phân loại thay đổi phần trăm này thành các nhóm. Ví dụ, trong dự báo hàng tuần, chúng tôi phân loại thay đổi giá giữa tuần này và tuần trước thành 12 nhóm: “D5+”, “D5”, ..., “U5+”, trong đó “D5+” nghĩa là giá giảm hơn 5%, “D_{*i*}” ($i = 5, 4, 3, 2, 1$) nghĩa là giảm từ $(i - 1)\%$ đến $i\%$, “U_{*i*}” nghĩa là tăng trong khoảng tương tự. Đối với dự báo theo tháng, số lượng nhóm có thể lên đến 10, với các nhãn tương ứng như “D10+”, “U10+”.

3.1.2 Dữ liệu Hồ sơ Công ty

Chúng tôi sử dụng GPT-4 để tạo mô tả công ty và các yếu tố tích cực/tiêu cực tổng quát có thể ảnh hưởng đến giá cổ phiếu công ty đó. Hình 3.1 là ví dụ về prompt dùng để yêu cầu GPT-4 tạo hồ sơ công ty.

3.1.3 Dữ liệu Tin tức Tài chính/Kinh tế

Chúng tôi sử dụng Google Custom Search API để lấy 5 tin tức hàng đầu hàng tuần của từng cổ phiếu NASDAQ-100. Sau đó, dùng GPT-4 để tóm tắt và trích xuất từ khóa từ từng bài báo. Phương pháp tương tự được áp dụng để thu thập tin tức hàng đầu về kinh tế vĩ mô và tình hình tài chính mỗi tuần. Để giảm kích thước đầu vào, chúng tôi tiếp tục tạo “tóm tắt meta” và từ khóa hàng tuần bằng GPT-4 dựa trên tất cả các tóm tắt tin tức hàng đầu của tuần. Chúng tôi chỉ sử dụng tóm tắt meta và từ khóa trong các thí nghiệm của bài báo này.

3.2 Suy luận Zero-shot/Few-shot Dựa trên Chỉ dẫn với LLMs

Trong các thử nghiệm zero-shot và few-shot, LLMs thể hiện khả năng tạo phản hồi hoặc không cần ví dụ bổ sung (zero-shot), hoặc dựa trên số lượng ví dụ tối thiểu (few-shot). Chúng tôi sử dụng prompt có cấu trúc chỉ dẫn, bao gồm phần chỉ dẫn, hồ sơ công ty, chuỗi tóm tắt tin tức theo thời gian kết hợp với chuỗi thời gian giá cổ phiếu đã được phân loại, và các ví dụ học few-shot giữa các chuỗi. Để tránh lặp lại không cần thiết trong prompt, chúng tôi cung cấp các ví dụ few-shot từ các cổ phiếu tương tự cổ phiếu đang xét. Thiết kế này cũng giúp chứng minh LLM có thể xem xét thông tin giữa các chuỗi cổ phiếu khác nhau. Ví dụ, để xác định các cổ phiếu tương tự, chúng tôi yêu cầu GPT-4 trả lời: “Liệt kê 3 cổ phiếu NASDAQ giống nhất với AAPL”, với phản hồi điển hình là “MSFT, GOOGL, AMZN”. Điều này cho thấy LLM hiểu rõ mối quan hệ giữa các thực thể tài chính. Cấu trúc prompt đã được tinh chỉnh thực nghiệm. Chẳng hạn, chúng tôi chia phần chỉ dẫn thành hai đoạn, đặt ở đầu và cuối prompt, giúp mô hình nhận diện rõ hơn nhiệm vụ là dự đoán tóm tắt và từ khóa của tuần tới, thay vì chỉ tóm tắt dữ liệu lịch sử. Tóm tắt và từ khóa được dự đoán chính là lời giải thích cho dự đoán lợi suất cổ phiếu tương ứng. Chúng tôi cũng thử nghiệm kỹ thuật Chuỗi Tư duy bằng cách thêm chỉ dẫn: “Bạn có thể suy luận từng bước trước khi đưa ra kết quả cuối cùng không?” vào cuối prompt. Bất ngờ thay, điều này giúp cải thiện hiệu suất đáng kể (xem Mục 4.2). Kết quả minh họa cho quy trình suy luận được thể hiện ở Hình 5, trong đó GPT-4 phát hiện một điểm quan trọng từng bị bỏ qua: “báo cáo lợi nhuận”.

3.3 Huấn luyện Tinh chỉnh Dựa trên Chỉ dẫn với Open LLaMA

Chúng tôi thực hiện huấn luyện tinh chỉnh dựa trên chỉ dẫn với mô hình Open LLaMA 13B để đánh giá khả năng của một mô hình mã nguồn mở so với GPT-4, đặc biệt sau khi tinh chỉnh. Trong chế độ zero-shot, Open LLaMA 13B có xu hướng sao chép nội dung prompt thay vì thực thi hiệu quả các chỉ dẫn, do đó không thể xử lý prompt như Hình 4 nếu không được tinh chỉnh. Việc tinh chỉnh dựa trên chỉ dẫn gần đây được chứng minh là hiệu quả trong việc định hướng quá trình huấn luyện của mô hình với các chỉ dẫn cụ thể. Chúng tôi tạo một tập dữ liệu gồm 30 nghìn ví dụ dự báo hàng tuần và 7 nghìn ví dụ dự báo hàng tháng, lấy từ dữ liệu lịch sử 5 năm (từ tháng 6 năm 2017 đến tháng 6 năm 2022). Khác với GPT-4 hỗ trợ prompt lên đến 8K tokens, chúng tôi phải nén prompt xuống còn 1K tokens cho việc tinh chỉnh Open LLaMA do giới hạn phần cứng. Đối với mỗi ví dụ huấn luyện, chúng tôi sử dụng GPT-4 để rút gọn toàn bộ tóm tắt/từ khóa tin tức lịch sử thành một cặp tóm tắt/từ khóa ngắn gọn hơn. Đồng thời, các phần “Hồ sơ Công ty” và “Ví dụ Dự báo” trong prompt cũng được cô đọng lại thành các đoạn văn súc tích. Mặc dù lý tưởng nhất là Open LLaMA có thể tự thực hiện toàn bộ quy trình thử

nghiệm từ đầu đến cuối, bao gồm cả nhiệm vụ nén lời nhắc (prompt compression) để phục vụ cho việc fine-tuning, nhưng hiện tại chúng tôi vẫn phải sử dụng GPT-4. Lý do là vì khả năng tóm tắt theo kiểu zero-shot của mô hình Open LLaMA 13B trên thực tế kém hơn đáng kể so với GPT-4. Các bản tóm tắt và từ khóa được trích xuất bởi mô hình Open LLaMA 13B thường không đạt yêu cầu sử dụng. Sau khi được fine-tune, mô hình Open LLaMA 13B thể hiện khả năng hiểu lệnh tốt hơn đáng kể, dẫn đến việc tạo ra dự báo cùng lời giải thích có tính mạch lạc cao. Điều này được minh họa trong Hình 6. Theo kết quả trong Mục 4.2, đối với bài toán phân loại nhị phân, hiệu suất của mô hình Open LLaMA khá cạnh tranh so với GPT-4. Tuy nhiên, chúng tôi nhận thấy mô hình Open LLaMA có xu hướng đưa ra những dự đoán cực đoan hơn, như U5+ hoặc D5+, dẫn đến sai số bình phương trung bình (MSE) cao hơn.

Ví dụ về tóm tắt (OpenLLaMA fine-tuned): Tóm tắt: Cổ phiếu của Apple Inc. đang thể hiện sức mạnh tương đối gia tăng, mặc dù vẫn chưa đạt đến một mốc chuẩn quan trọng. Công ty được đánh giá là một trong những cổ phiếu hoạt động tốt nhất nhờ vào việc đã bán được 1,3 tỷ chiếc iPhone và đạt lợi nhuận đáng kể 19,4 tỷ USD trong quý vừa qua. Khi chỉ số NASDAQ tăng mạnh với sự đóng góp của Apple với tư cách là công ty vốn hóa lớn, Apple cũng đang mở rộng các biện pháp cắt giảm chi phí, hoãn các phúc lợi, hạn chế tuyển dụng và để trống các vị trí, làm dấy lên lo ngại về việc cắt giảm nhân sự. Hiệu suất chung của thị trường chứng khoán được thúc đẩy nhờ vào gói hỗ trợ tiền gửi trị giá 30 tỷ USD dành cho các công ty đang gặp khó khăn. Từ khóa: Apple Inc., cổ phiếu, sức mạnh tương đối, cắt giảm chi phí, cắt giảm nhân sự, NASDAQ, gói hỗ trợ tiền gửi 30 tỷ USD, iPhone, lợi nhuận.

Lời nhắc (Prompt):

Tạo một mô tả ngắn gọn cho công ty AAPL (Apple Inc.). Đồng thời liệt kê các yếu tố tích cực và tiêu cực có thể ảnh hưởng đến giá cổ phiếu của công ty; trình bày ngắn gọn và sử dụng từ khóa. Cân nhắc đến các yếu tố đa dạng như tình hình kinh tế vĩ mô (ví dụ: lạm phát, tăng trưởng CPI), yếu tố kinh doanh (ví dụ: doanh số, đầu tư, sản phẩm), yếu tố công nghệ (ví dụ: đổi mới sáng tạo), và các yếu tố khác. Sử dụng định dạng sau: Description: ..., Positive Factors: ..., Negative Factors: ...

Phản hồi từ GPT-4:

Mô tả (Description): AAPL (Apple Inc.) là một công ty công nghệ nổi tiếng toàn cầu có trụ sở tại Cupertino, California. Được biết đến với thiết kế tinh tế và công nghệ tiên tiến, công ty sản xuất nhiều loại sản phẩm như iPhone, iPad, Mac, Apple Watch, và cung cấp các dịch vụ như iCloud, App Store và nhiều dịch vụ khác.

Các yếu tố tích cực (Positive Factors):

- *Kinh tế vĩ mô: Tăng trưởng kinh tế, chi tiêu tiêu dùng gia tăng*
- *Kinh doanh: Doanh số mạnh, mở rộng sang các thị trường mới, báo cáo lợi nhuận tích cực, các thương vụ mua lại chiến lược*
- *Công nghệ: Đổi mới sáng tạo, ra mắt sản phẩm mới, tiến bộ trong trí tuệ nhân tạo và thực tế tăng cường*
- *Thương hiệu: Thương hiệu toàn cầu mạnh, cơ sở khách hàng trung thành*
- *Pháp lý/Quy định: Thay đổi chính sách tích cực, phê duyệt bằng sáng chế thành công*

Các yếu tố tiêu cực (Negative Factors):

- *Kinh tế vĩ mô: Lạm phát, suy thoái kinh tế, hạn chế thương mại*
- *Kinh doanh: Doanh số giảm, báo cáo lợi nhuận kém, thiếu đầu tư mới*
- *Công nghệ: Đổi mới chậm, trì hoãn trong việc ra mắt sản phẩm*
- *Thương hiệu: Tổn hại danh tiếng thương hiệu, mất lòng trung thành từ khách hàng*
- *Pháp lý/Quy định: Điều tra chống độc quyền, thay đổi chính sách bất lợi, tranh chấp bằng sáng chế*
- *Môi trường: Không đạt mục tiêu bền vững, tác động môi trường tiêu cực*

Hình 3.1: Một ví dụ về hồ sơ công ty của cổ phiếu bao gồm phần mô tả công ty và các yếu tố tích cực/tiêu cực chung ảnh hưởng đến giá cổ phiếu của công ty.

Lời nhắc (Prompt) [mẫu, các phần nằm trong dấu [] là chỗ cần điền thông tin]:

Vui lòng tóm tắt đoạn văn bản tin có thể chứa nhiều nhiều bên dưới, được trích xuất từ mã HTML của một trang web, và trích xuất các từ khóa của bản tin. Đoạn tin này có thể rất nhiều do được trích từ HTML. Hãy đưa ra câu trả lời theo định dạng sau: Summary: ..., Keywords: ... Bản tin này được cho là liên quan đến cổ phiếu {symbol}. Bạn có thể ghi "N/A" nếu văn bản chứa quá nhiều nhiều và không có thông tin phù hợp để trích xuất. News: {news}

Phản hồi từ GPT-4:

Summary: Apple Inc đang trì hoãn một số phúc lợi cho các bộ phận thuộc khối văn phòng và mở rộng nỗ lực cắt giảm chi phí, cùng bước đi với các công ty đồng ngành trong việc tinh giản hoạt động. Công ty dự kiến sẽ không chi trả thưởng hoặc thăng chức trong tháng Tư, và sẽ chỉ thực hiện các khoản chi trả này vào tháng Mười. Apple cũng đang hạn chế tuyển dụng và để trống thêm các vị trí khi có nhân viên rời đi. Điều này đang gây ra lo ngại về việc cắt giảm nhân sự trong nội bộ nhân viên.

Keywords: Apple, thưởng, tuyển dụng, cắt giảm chi phí, thăng chức, sa thải, nhân viên

Hình 3.2: Một ví dụ về bản tóm tắt tin tức và từ khóa được trích xuất từ một bài báo liên quan đến công ty (trong trường hợp này là AAPL), được tạo ra bởi GPT-4. Bản tin gốc nằm tại: đây. Mẫu lời nhắc trong ví dụ này là một khuôn mẫu. Các đoạn văn trong dấu [] chỉ là chú thích và không thực sự xuất hiện trong lời nhắc khi gửi đến mô hình LLM. Chúng tôi sẽ điền mã cổ phiếu và văn bản bản tin cụ thể vào các phần được đánh dấu bởi {}.

Chương 4

THÍ NGHIỆM

4.1 Cài đặt thí nghiệm

Chi tiết dữ liệu sử dụng trong thí nghiệm được mô tả tại Mục 3.1. Trong bài báo này, chúng tôi tập trung vào dự báo lợi suất cổ phiếu NASDAQ-100.

4.1.1 Khoảng thời gian dữ liệu

- Khoảng thời gian huấn luyện/fine-tune: Dữ liệu 5 năm, từ 12/06/2017 đến 05/06/2022. Dữ liệu này được dùng để huấn luyện các mô hình cơ sở và fine-tune mô hình Open LLaMA 13B.
- Khoảng thời gian đánh giá: Gồm 52 tuần, từ 06/06/2022 đến 04/06/2023. Dữ liệu trong khoảng này được dùng để đánh giá các mô hình cơ sở, thử nghiệm zero/few-shot với GPT-4 và đánh giá mô hình Open LLaMA 13B sau khi fine-tune.

4.1.2 Mô hình cơ sở (Baseline Models)

Để đánh giá hiệu suất phương pháp của chúng tôi, chúng tôi sử dụng các mô hình cơ sở gồm:

- Heuristic baseline: Dự đoán bằng bin xuất hiện nhiều nhất trong lịch sử trước tuần dự đoán.
- Mô hình ARMA-GARCH ($AR = MA = 1$).
- Mô hình cây tăng cường độ dốc (gradient boosting tree) dùng thư viện LightGBM.

Các mô hình cơ sở được huấn luyện trên dữ liệu huấn luyện/fine-tune và đánh giá trên dữ liệu đánh giá. Với mô hình cây tăng cường, các đặc trưng (features) gồm khoảng 300 đặc trưng, bao gồm:

- Dữ liệu chuỗi thời gian giá cổ phiếu hàng ngày: giá mở cửa, đóng cửa, cao nhất, thấp nhất và khối lượng giao dịch.

- Thống kê cửa sổ trượt: trung bình, trung vị, min, max và độ lệch chuẩn với kích thước cửa sổ là 2, 5, 10, 30, 60, 90 ngày.
- Thông tin ngành của cổ phiếu và lợi nhuận lịch sử thu thập từ Alpha Vantage (<https://www.alphavantage.co/documentation/>).

4.1.3 Chỉ số đánh giá (Evaluation Metrics)

Chúng tôi thực hiện dự báo lợi suất cổ phiếu hàng tuần và hàng tháng (4 tuần = 1 tháng, tổng cộng 13 “tháng” trong 52 tuần). Ba chỉ số đánh giá được sử dụng:

- Độ chính xác nhị phân (Binary precision): Đánh giá khả năng dự đoán đúng chiều hướng tăng/giảm của giá cổ phiếu.
- Độ chính xác theo bin (Bin precision): Đánh giá khả năng dự đoán chính xác bin cụ thể như “D5+”, “D4”, ..., “U5+”.
- Sai số bình phương trung bình (MSE): Dựa trên thứ tự bin liên tiếp (-6 cho “D5+”, ..., 0 cho “U1”, ..., 5 cho “U5+”), đánh giá độ chênh lệch trung bình giữa dự đoán và thực tế, giúp xác định xu hướng dự đoán cực đoan của mô hình.

Chất lượng lời giải thích dự báo (tóm tắt/từ khóa tuần/tháng tới) được đánh giá bằng chỉ số ROGUE-1 và ROGUE-2, so sánh với tóm tắt/từ khóa thực tế do GPT-4 tạo ra dựa trên tin tức thực tế của tuần/tháng tiếp theo.

4.2 Đánh giá hiệu suất

Kết quả thí nghiệm được tóm tắt ở Bảng 4.1 và Bảng 4.2. Tóm lại, kết quả cho thấy các mô hình LLM hiệu quả trong dự báo chuỗi thời gian tài chính, đặc biệt là GPT-4 với few-shot kết hợp Chain-of-Thoughts (COT) cho kết quả tốt nhất cả về độ chính xác và chất lượng giải thích. Ngoài ra, kỹ thuật COT liên tục cải thiện hiệu suất, và việc fine-tune các mô hình mã nguồn mở như Open LLaMA dựa trên hướng dẫn có thể đạt hiệu suất khá so với GPT-4.

4.2.1 Dự báo giá cổ phiếu

Từ kết quả trong Bảng 4.1, chúng tôi nhận thấy rằng cả GPT-4 và mô hình Open LLaMA 13B đều vượt trội hơn so với mô hình ARMA-GARCH và mô hình cây tăng cường độ dốc (gradient-boosting tree) về cả độ chính xác nhị phân (binary precision) và độ chính xác theo bin (bin precision). Đặc biệt, GPT-4 thể hiện hiệu suất vượt trội trong cả hai thiết lập zero-shot và few-shot, trong đó thiết lập few-shot kết hợp với kỹ thuật Chain-of-Thought (COT) đạt hiệu suất tốt nhất. Xét về sai số bình phương trung bình (MSE), “GPT-4 few-shot with COT” cũng đạt mức sai số thấp nhất, cho thấy nó không chỉ dự đoán đúng hướng biến động giá mà còn ước lượng tốt hơn về mức độ thay đổi. Mô hình Open LLaMA 13B, sau khi được tinh chỉnh (fine-tuning), cho thấy hiệu suất cạnh tranh so với GPT-4 về độ chính xác nhị phân. Tuy nhiên, độ chính xác theo bin lại thấp hơn rõ rệt, cho thấy mô hình này chưa có khả năng suy luận chi tiết đủ tốt để chọn đúng bin. Ngoài ra, mô hình này cũng có xu hướng đưa ra các dự đoán cực đoan hơn, dẫn đến MSE cao hơn tương đối.

Bảng 4.1: So sánh hiệu suất giữa các mô hình cơ sở và các mô hình ngôn ngữ lớn (LLMs) trong dự báo giá cổ phiếu theo tuần/tháng.

Mô hình	Hàng tuần			Hàng tháng		
	Độ chính xác nhị phân	Độ chính xác theo nhóm	MSE	Độ chính xác nhị phân	Độ chính xác theo nhóm	MSE
Most-Frequent Historical Bin	50.7%	16.4%	43.5	51.4%	17.2%	155.1
ARMA-GARCH	52.4%	11.1%	22.1	50.5%	6.2%	90.1
Gradient Boosting Tree Model	60.8%	26.4%	24.3	56.4%	17.7%	85.6
GPT-4 Zero-Shot	64.5%	31.2%	20.5	64.8%	26.0%	60.1
GPT-4 Few-Shot	65.8%	32.7%	20.6	65.3%	26.5%	58.2
GPT-4 Few-Shot w/COT	66.5%	35.2%	18.7	69.5%	28.6%	50.4
Open LLaMA (13B) Fine-Tuned	62.2%	26.5%	23.3	60.1%	22.6%	63.3
Open LLaMA (13B) Fine-Tuned w/COT	64.7%	30.7%	21.0	62.2%	24.4%	63.5

4.2.2 Chất lượng giải thích

Bảng 4.2 cho thấy chất lượng của các phần giải thích được tạo ra bởi các mô hình ngôn ngữ lớn (GPT-4 và Open LLaMA đã tinh chỉnh), được đánh giá bằng điểm số ROUGE-1 và ROUGE-2 đối với cả phần tóm tắt (S) và từ khóa (K) của tin tức. Một lần nữa, kết quả cho thấy “GPT-4 few-shot with COT” đạt điểm ROUGE cao nhất, chứng tỏ nó tạo ra các giải thích phù hợp và chính xác nhất cho các dự đoán. Open LLaMA, sau khi được tinh chỉnh với COT, cũng cho thấy chất lượng giải thích hợp lý, tương đương với kết quả của GPT-4 không sử dụng COT.

Bảng 4.2: Đánh giá chất lượng giải thích bằng cách sử dụng điểm số ROUGE, với bản tóm tắt/chiết xuất từ khóa của GPT-4 cho các tin tức hàng đầu thực sự mỗi tuần từ tìm kiếm Google được sử dụng làm dữ liệu chuẩn.

Mô hình	Hàng tuần				Hàng tháng			
	ROUGE-1 (S)	ROUGE-2 (S)	ROUGE-1 (K)	ROUGE-2 (K)	ROUGE-1 (S)	ROUGE-2 (S)	ROUGE-1 (K)	ROUGE-2 (K)
GPT-4 Zero-Shot	0.2212	0.0675	0.1295	0.0447	0.2528	0.0665	0.1335	0.0657
GPT-4 Few-Shot	0.2242	0.0526	0.1304	0.0454	0.2450	0.0634	0.1348	0.0644
GPT-4 Few-Shot w/ COT	0.2414	0.0543	0.2083	0.0869	0.2645	0.0758	0.2450	0.1025
Open LLaMA (13B) Fine-Tuned	0.2053	0.0395	0.0927	0.0324	0.2242	0.0474	0.1167	0.0520
Open LLaMA (13B) Fine-Tuned w/COT	0.2371	0.0434	0.1123	0.0425	0.2436	0.0536	0.1356	0.0834

Chương 5

KẾT LUẬN

Trong bài báo này, chúng tôi đã nghiên cứu ứng dụng của các Mô hình Ngôn ngữ Lớn (LLMs) trong lĩnh vực dự báo chuỗi thời gian tài chính, với trọng tâm đặc biệt là dự báo lợi suất cổ phiếu NASDAQ-100. Chúng tôi đã chứng minh cách các LLM, cụ thể là GPT-4 và Open LLaMA, có thể được sử dụng để tạo ra dự báo và cung cấp các giải thích dễ hiểu cho các dự đoán của chúng. Phương pháp của chúng tôi sử dụng các prompt có cấu trúc, bao gồm dữ liệu hồ sơ công ty, dữ liệu giá cổ phiếu lịch sử và dữ liệu tin tức tài chính, để hướng dẫn các LLM trong nhiệm vụ dự báo. Kết quả thực nghiệm cho thấy LLM có thể vượt trội hơn các mô hình thống kê truyền thống và các kỹ thuật máy học như mô hình ARMA-GARCH và mô hình cây tăng cường độ dốc. Đáng chú ý, việc tích hợp quy trình suy luận từng bước, dựa trên phương pháp Chain of Thought (COT), đã cải thiện đáng kể hiệu suất của các mô hình dựa trên LLM. Hơn nữa, các thí nghiệm tinh chỉnh (fine-tuning) với Open LLaMA đã chứng minh khả năng tinh chỉnh hiệu quả các LLM công khai cho nhiệm vụ này, từ đó giải quyết các thách thức vốn có như suy luận qua chuỗi thời gian, tích hợp tín hiệu đa phương thức và tính giải thích kết quả trong lĩnh vực tài chính.

Tóm lại, nghiên cứu sơ bộ của chúng tôi về ứng dụng LLM trong dự báo tài chính đã mang lại những kết quả đầy hứa hẹn ban đầu. Mặc dù vẫn còn ở giai đoạn đầu, những kết quả tích cực này tạo động lực mạnh mẽ cho các nghiên cứu tiếp theo trong lĩnh vực này. Khi chúng tôi tiếp tục mở rộng ứng dụng của LLM trong lĩnh vực tài chính, chúng tôi hình dung một tương lai nơi dự báo tài chính không chỉ chính xác hơn mà còn dễ hiểu và minh bạch hơn. Sự phát triển này có thể đóng góp đáng kể vào sự chuyển đổi của quá trình ra quyết định tài chính trên toàn ngành. Các nghiên cứu trong tương lai sẽ đi sâu hơn vào các phương pháp này, bao gồm nhưng không giới hạn ở: 1) mở rộng nghiên cứu để bao gồm nhiều chỉ số cổ phiếu hơn như SP500 và Russell 2000, 2) tích hợp nghiên cứu với nhiều loại dữ liệu hơn, chẳng hạn như chuỗi thời gian kinh tế vĩ mô, khối lượng giao dịch cổ phiếu và dữ liệu mạng xã hội, và 3) khám phá việc tinh chỉnh các mô hình công khai lớn hơn, ví dụ mô hình 30 tỷ tham số, nhằm nâng cao khả năng suy luận.

TÀI LIỆU THAM KHẢO

- [1] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, and Teh Ying Wah. 2015. Timeseries clustering—a decade review. *Information systems* 53 (2015), 16–38.
- [2] Torben Gustav Andersen, Richard A Davis, Jens-Peter Kreiß, and Thomas V Mikosch. 2009. *Handbook of financial time series*. Springer Science & Business Media.
- [3] M Suresh Babu, N Geethanjali, and B Satyanarayana. 2012. Clustering approach to stock market prediction. *International Journal of Advanced Networking and Applications* 3, 4 (2012), 1281.
- [4] Arash Bahrammirzaee. 2010. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications* 19, 8 (2010), 1165–1195.
- [5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [6] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. J. Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *ArXiv abs/2005.14165* (2020).
- [7] Cathy WS Chen, Feng-Chi Liu, and Mike KP So. 2011. A review of threshold time series models in finance. *Statistics and its Interface* 4, 2 (2011), 167–181.
- [8] Zheng Chen, Xinli Yu, Yuan Ling, Bo Song, Wei Quan, Xiaohua Hu, and Erjia Yan. 2018. Correlated anomaly detection from large streaming data. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 982–992.
- [9] Dawei Cheng, Fangzhou Yang, Sheng Xiang, and Jin Liu. 2022. Financial time series forecasting with multi-modality graph neural network. *Pattern Recognition* 121 (2022), 108218.
- [10] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. 2023. Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality. <https://vicuna.lmsys.org>

- [11] Søren Johansen. 1995. Likelihood-based inference in cointegrated vector autoregressive models. OUP Oxford.
- [12] OpenAI. 2023. GPT-4 Technical Report. ArXiv abs/2303.08774 (2023).
- [13] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023).
- [14] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. arXiv preprint arXiv:2201.11903 (2022).
- [15] Eric Zivot and Jiahui Wang. 2006. Vector autoregressive models for multivariate time series. Modeling financial time series with S-PLUS® (2006).