

Dự đoán Chuỗi Thời gian Tài chính sử dụng Mô hình Ngôn ngữ Lớn

Vũ Văn Tới

Nguyễn Năng Thịnh

Mai Phan Anh Tùng

Phan Mạnh Quân

Viện trí tuệ nhân tạo - Đại học Công nghệ, Đại học Quốc gia Hà Nội

Tóm tắt nội dung

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. Nghiên cứu này 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. Nghiên cứu 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. Qua đó cho thấy phương pháp được tiến hành nghiên cứu 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, nghiên cứu đã cho 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, nghiên cứu 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.

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 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 những biến thể, 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, nghiên cứu tập trung vào chuỗi giá cổ phiếu NASDAQ-100.

- Nghiên cứu 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 “chain of thought” giúp cải thiện hiệu suất trong hầu hết các thử nghiệm.
- Nghiên cứu 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). Qua đó cho thấy 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).

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), mô hình không gian trạng thái và bộ lọc Kalman, mô hình khuếch tán, mô hình Vector Hiệu chỉnh Sai số (VECM), và mô hình Cân bằng tổng thể ngẫu nhiên động (DSGE).

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 mô hình ngôn ngữ lớn như GPT-3 và GPT-4, LLaMA (bao gồm Alpaca, Vicuna) 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.

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 nghiên cứu 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).

Nghiên cứu trình bày thiết kế có cấu trúc của prompt cho LLMs và áp dụng mô hình GPT-4 hiện đại cho các tác vụ suy luận zero-shot và few-shot. Đối với huấn luyện tinh chỉnh, sử dụng mô hình mã nguồn mở Open LLaMA. Nghiên cứu cũng áp dụng kỹ thuật “Chuỗi Tư duy” (Chain of Thoughts - COT), 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

Nghiên cứu tiến hành tải dữ liệu giá cổ phiếu hàng ngày NASDAQ-100 từ Yahoo Finance bằng gói yfinance. Trong nghiên cứu này, đầu tiên là 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, tiến hành 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%, “Di” ($i = 5, 4, 3, 2, 1$) nghĩa là giảm từ $(i - 1)\%$ đến $i\%$, “Ui” 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

Nghiên cứu 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 1 là ví dụ về prompt dùng để yêu cầu GPT-4 tạo hồ sơ công ty.

Mô tả:

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 sáng tạo, công ty sản xuất nhiều sản phẩm đa dạng bao gồm iPhone, iPad, Mac, Apple Watch, cùng các dịch vụ như iCloud, App Store và nhiều hơn nữa.

Các yếu tố tích cực:

- *Kinh tế vĩ mô:* Tăng trưởng kinh tế, chi tiêu tiêu dùng tăng
- *Kinh doanh:* Doanh số mạnh, mở rộng 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, khách hàng trung thành
- *Pháp lý/Quy định:* Các chính sách thuận lợi, thành công trong đăng ký bằng sáng chế

Các yếu tố tiêu cực:

- *Kinh tế vĩ mô:* Lạm phát, suy thoái kinh tế, hạn chế thương mại
- *Kinh doanh:* Giảm doanh số, báo cáo lợi nhuận kém, thiếu đầu tư mới
- *Công nghệ:* Tốc độ đổi mới chậm, trì hoãn ra mắt sản phẩm
- *Thương hiệu:* Tổn hại đến uy tín thương hiệu, mất lòng trung thành của khách hàng
- *Pháp lý/Quy định:* Điều tra chống độc quyền, chính sách bất lợi, tranh chấp bằng sáng chế

Hình 1: Ví dụ về hồ sơ công ty của một cổ phiếu, bao gồm mô tả công ty và các yếu tố tích cực/tiêu cực chung ảnh hưởng đến giá cổ phiếu của công ty đó.

3.1.3 Dữ liệu Tin tức Tài chính/Kinh tế

Nghiên cứu sử dụng Google Custom Search API để thu thập 5 tin tức hàng đầu mỗi tuần liên quan đến từng cổ phiếu trong danh sách NASDAQ-100. Sau đó sử dụng GPT-4 để tạo bản tóm tắt và trích xuất từ khóa từ mỗi bài báo thu thập được. Một ví dụ về prompt và phản hồi của GPT-4 được minh họa ở Hình 2.

Bên cạnh đó, một phương pháp tương tự cũng được áp dụng để thu thập 5 tin tức hàng đầu mỗi tuần liên quan đến tình hình kinh tế vĩ mô và tài chính.

Để giảm kích thước đầu vào tiếp tục sử dụng GPT-4 để tạo ra bản tóm tắt tổng hợp (meta summary) và từ khóa cho mỗi tuần, dựa trên tất cả các bản tóm tắt tin chính và từ khóa của tuần đó. Trong các thí nghiệm của bài báo này, nghiên cứu chỉ sử dụng bản tóm tắt tổng hợp và từ khóa. Một ví dụ về bản tóm tắt tổng hợp và từ khóa được minh họa ở Hình 3, chúng có hình thức tương tự như ví dụ ở Hình 2 nhưng được rút gọn hơn nhiều.

Prompt [(mẫu), những phần bên trong dấu { } là chỗ cần điền]

Vui lòng tóm tắt dữ liệu tin tức dưới đây (mặc dù có thể nhiều) được trích xuất từ HTML của trang web, và trích xuất các từ khóa của tin tức. Văn bản tin tức có thể rất nhiều do được trích xuất từ HTML. Hãy đưa ra câu trả lời có định dạng như sau:

Summary: ...

Keywords: ...

Tin tức này được cho là liên quan đến cổ phiếu {symbol}. Bạn có thể ghi "N/A" nếu đoạn văn bản nhiều không chứa thông tin liên quan để trích xuất.

News: {news}

Phản hồi của GPT-4:

Tóm tắt: Apple Inc. đang trì hoãn các phúc lợi cho một số bộ phận văn phòng và mở rộng nỗ lực cắt giảm chi phí, cùng với xu hướng tinh giản hoạt động như các công ty đồng ngành. Công ty đề xuất hoãn việc chi trả tiền thưởng hoặc thăng chức trong tháng Tư, với kế hoạch thanh toán dời đến tháng Mười. Apple cũng đang hạn chế tuyển dụng và để trống thêm các vị trí sau khi nhân viên rời đi. Điều này đang gây ra lo ngại về khả năng cắt giảm nhân sự trong nội bộ nhân viên.

Từ khóa: Apple, tiền thưởng, tuyển dụng, cắt giảm chi phí, thăng chức, cắt giảm nhân sự, nhân viên.

Hình 2: Một ví dụ về phần tóm tắt tin tức và từ khóa được trích xuất từ một bài báo cho công ty sở hữu cổ phiếu (trong trường hợp này là AAPL), được tạo bởi GPT-4.

Tóm tắt:

Cổ phiếu của Apple Inc. đang thể hiện sức mạnh tương đối gia tăng, dù vẫn chưa đạt đến một mốc chuẩn quan trọng. Công ty đã được đánh giá là một trong những cổ phiếu hoạt động tốt nhất nhờ doanh số 1,3 tỷ chiếc iPhone và lợi nhuận ấn tượng 19,4 tỷ USD trong quý vừa qua. Khi chỉ số NASDAQ tăng mạnh với Apple là một cổ phiếu vốn hóa lớn, Apple tiếp tục mở rộng các biện pháp cắt giảm chi phí, hoãn các phúc lợi, hạn chế tuyển dụng và để trống nhiều vị trí, làm dấy lên lo ngại về việc cắt giảm nhân sự. Hiệu suất chung của thị trường chứng khoán cũng được thúc đẩy nhờ khoản bơm vốn trị giá 30 tỷ USD vào 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, bơm vốn 30 tỷ USD, iPhone, lợi nhuận.

Hình 3: Một ví dụ về bản tóm tắt tổng hợp và từ khóa của một tuần, được cô đọng từ tất cả các bản tóm tắt và từ khóa của công ty trong tuần đó.

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

Trong suy luận zero-shot và few-shot, các mô hình ngôn ngữ lớn (LLM) thể hiện khả năng tạo ra phản hồi mà không cần bất kỳ ví dụ bổ sung nào (zero-shot) hoặc dựa trên một số rất ít ví dụ ngoài tập huấn luyện gốc (few-shot). Trong thí nghiệm suy luận zero-shot/few-shot của nghiên cứu, nghiên cứu sử dụng một prompt dựa trên hướng dẫn. Cấu trúc của prompt, được minh họa trong Hình 4, bao gồm các hướng dẫn, hồ sơ công ty,

một chuỗi tóm tắt tin tức/từ khóa lịch sử theo thời gian đan xen với chuỗi dữ liệu giá cổ phiếu được phân loại, cùng với các ví dụ học few-shot đa chuỗi (cross-sequence few-shot learning).

Để tránh lặp lại không cần thiết trong văn bản prompt, nghiên cứu có chủ ý cung cấp các ví dụ học few-shot từ các cổ phiếu tương tự với cổ phiếu đang quan tâm. Thiết kế này cũng giúp chứng minh rằng LLM có thể xem xét thông tin đa chuỗi từ nhiều cổ phiếu khác nhau. Để xác định các cổ phiếu tương tự, nghiên cứu hỏi GPT-4 câu như: "Liệt kê 3 cổ phiếu NASDAQ tương tự nhất với AAPL". Một câu trả lời điển hình, ví dụ "MSFT, GOOGL, AMZN", thể hiện khả năng hiểu biết của LLM về các mối quan hệ giữa các thực thể và khái niệm tài chính khác nhau. Khi sử dụng LLM, nghiên cứu ngầm tận dụng kiến thức sâu rộng của nó về các thực thể và khái niệm tài chính.

Cấu trúc prompt và các hướng dẫn đã được điều chỉnh dựa trên thực nghiệm. Ví dụ, nghiên cứu chia phần hướng dẫn thành hai phần, đặt ở đầu và cuối prompt, giúp mô hình nhận biết tốt hơn nhiệm vụ của nó: dự đoán tóm tắt và từ khóa cho tuần tiếp theo, thay vì tóm tắt dữ liệu lịch sử. Tóm tắt và từ khóa được dự đoán này đóng vai trò như lời giải thích cho dự đoán lợi nhuận cổ phiếu tương ứng.

Nghiên cứu cũng thử nghiệm phương pháp Chain-of-Thoughts, tức là ý tưởng "suy nghĩ từng bước một", bằng cách thêm hướng dẫn "Bạn có thể suy luận từng bước trước khi hoàn thiện kết quả không?" vào cuối prompt. Đáng ngạc nhiên là điều này đã cải thiện hiệu suất đáng kể (xem Mục 4.2). Kết quả của quá trình suy nghĩ từng bước phản hồi cho Hình 4 được minh họa trong Hình 5, nơi thể hiện rõ GPT-4 nhận ra một điểm quan trọng trước đây bị bỏ sót về "báo cáo lợi nhuận" khi các bước suy luận được tạo ra rõ ràng.

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

Chỉ dẫn: Dự báo lợi suất cổ phiếu (biến động giá) cho tuần tới của mã cổ phiếu, dựa trên hồ sơ công ty, tóm tắt tin tức hàng tuần gần đây, từ khóa, và lợi suất cổ phiếu trong quá khứ, và tùy chọn các ví dụ từ các cổ phiếu tương tự.

Xu hướng được biểu diễn bằng các nhãn "D5+", "D5", "D4", "D3", "D2", "D1", "U1", "U2", "U3", "U4", "U5", "U5+", trong đó:

- "D5+" nghĩa là giá giảm hơn 5%,
- "D5" là giá giảm từ 4% đến 5%,
- "D4" là giá giảm từ 3% đến 4%,
- "U5+" nghĩa là giá tăng hơn 5%,
- "U5" là giá tăng từ 4% đến 5%,
- "U4" là giá tăng từ 3% đến 4%, v.v.

Hồ sơ công ty: {company_profile} [mô tả công ty, các yếu tố tích cực/tiêu cực chung như trong Hình 1.]

Tin tức gần đây: Tin tức được sắp xếp từ tuần cũ nhất đến tuần gần nhất.

8 tuần trước:

{meta_news_summary_company} [tóm tắt tin tức và từ khóa cho công ty trong tuần]

{meta_news_summary_macro} [tóm tắt tin tức và từ khóa cho kinh tế vĩ mô và tình hình tài chính trong tuần]

...

Tuần trước:

...

Ví dụ dự báo: {few_shot_learning_examples_from_similar_stocks} [Sử dụng ví dụ từ 3 cổ phiếu tương tự nhất làm ví dụ học ít-shot...]

Bây giờ hãy dự đoán Tóm tắt, Từ khóa, và Lợi suất cổ phiếu cho tuần tới.

Phần Tóm tắt/Từ khóa được dự đoán phải giải thích được dự báo lợi suất cổ phiếu. Bạn cần dự đoán điều có thể xảy ra trong tuần tới, không chỉ tóm tắt quá khứ. Lợi suất tuần tới không bắt buộc phải giống tuần trước.

Sử dụng định dạng:

Summary: ...

Keywords: ...

Stock Return: ...

Phản hồi của GPT-4 [Dự báo tuần từ 01/05/2023 đến 07/05/2023]:

Tóm tắt: Cổ phiếu Apple dự kiến sẽ có những biến động trong tuần tới trong bối cảnh các tin tức hỗn hợp...

Từ khóa: AAPL, kính thực tế hỗn hợp, mở rộng sang Ấn Độ, tài khoản tiết kiệm lãi cao...

Lợi suất cổ phiếu: U1

Dữ liệu thực tế [Tóm tắt từ tin tức thật từ 01/05/2023 đến 07/05/2023 và lợi suất thực tế]:

Tóm tắt: Cổ phiếu Apple đã ghi nhận mức tăng đáng kể sau khi vượt qua kỳ vọng của Phố Wall...

Từ khóa: Apple, cổ phiếu, Phố Wall, quý tài chính thứ hai, iPhone, doanh thu dịch vụ...

Lợi suất cổ phiếu: U3

Hình 4: Cấu trúc lời nhắc (prompt) được sử dụng cho các thí nghiệm trong bài báo này với các mô hình ngôn ngữ lớn (LLMs).

Lời nhắc: {Giống lời nhắc trong Hình 4.}

Bạn có thể suy luận từng bước trước khi đưa ra dự đoán cuối cùng không?

Phản hồi của GPT-4 [Dự báo cho tuần từ 01/05/2023 đến 07/05/2023]:

Dựa trên thông tin được cung cấp, chúng ta có thể thấy rằng Apple (AAPL) đã có nhiều diễn biến tích cực và một vài mối quan ngại tiềm tàng trong thời gian gần đây.

Các diễn biến tích cực:

- Doanh số và lợi nhuận gần đây khá mạnh, và công ty có vẻ đang mở rộng sang các thị trường mới như Ấn Độ.
- Kính thực tế hỗn hợp sắp ra mắt và dịch vụ "Apple Pay Later" cho thấy Apple vẫn tiếp tục đổi mới và mở rộng sang các lĩnh vực sản phẩm mới.
- Các biện pháp cắt giảm chi phí và kế hoạch sử dụng 100% cobalt tái chế cho

thấy Apple đang nỗ lực cải thiện hiệu quả và tính bền vững, điều này có thể thu hút một số nhà đầu tư.

- Việc thắng kiện trong vụ tranh chấp với cơ quan chống độc quyền của Anh giúp công ty tránh được các khoản phạt tiềm năng.
- Cổ phiếu AAPL gần đây có xu hướng tăng, với nhiều tuần liên tiếp tăng (U5, U4, U3, U2), và chỉ có một ngày giảm nhẹ (D1).

Các mối quan ngại tiềm tàng:

- Apple đang cắt giảm chi phí, trì hoãn thưởng và giảm tốc độ tuyển dụng ở một số bộ phận, có thể ảnh hưởng đến tinh thần làm việc và năng suất của nhân viên.
- Ý kiến tiêu cực từ nhà phân tích Toni Sacconaghi về việc Phố Wall có thể đánh giá quá cao doanh thu quý 3 và quý 4 có thể ảnh hưởng đến giá cổ phiếu.

Bên cạnh những yếu tố trên, xu hướng chung của thị trường chứng khoán dường như đang tăng, được minh chứng bằng sự tăng điểm gần đây của Dow Jones, S&P 500 và Nasdaq. Trong vài tuần gần đây, hiệu suất cổ phiếu của các công ty công nghệ lớn như Microsoft và Alphabet cũng rất mạnh. Hơn nữa, các diễn biến tích cực trong môi trường kinh tế vĩ mô, bao gồm sự gia tăng chi tiêu tiêu dùng, cũng mang lại tín hiệu khích lệ.

Giờ đây, chúng ta có thể đưa ra dự đoán về Tóm tắt, Từ khóa, và Lợi suất cổ phiếu cho tuần tới. Tuy nhiên, cần lưu ý rằng đây chỉ là dự đoán dựa trên thông tin hiện có và kết quả thực tế có thể khác.

Tóm tắt: Các bản xem trước kính thực tế hỗn hợp của Apple đã nhận được phản hồi tích cực từ các chuyên gia trong ngành, giúp tăng niềm tin của nhà đầu tư. Tuy nhiên, những lo ngại về các biện pháp cắt giảm chi phí và nguy cơ tuyển dụng chậm lại có thể tạo ra sự không chắc chắn. Phố Wall kỳ vọng báo cáo lợi nhuận mạnh, nâng cao tinh thần thị trường. Bên cạnh đó, việc mở rộng sang Ấn Độ của Apple vẫn tiếp tục cho thấy tiềm năng.

Từ khóa: Apple, kính thực tế hỗn hợp, niềm tin nhà đầu tư, cắt giảm chi phí, tuyển dụng, Phố Wall, báo cáo lợi nhuận, mở rộng, Ấn Độ.

Lợi suất cổ phiếu: U2. Dựa trên các xu hướng tích cực và những diễn biến đầy hứa hẹn, cổ phiếu có khả năng tăng giá. Tuy nhiên, những lo ngại về cắt giảm chi phí có thể làm giảm đà tăng, dẫn đến dự đoán mức tăng từ 1% đến 2% trong tuần tới, tức là “U2.”

Hình 5: GPT-4 đưa ra các bước suy luận chi tiết nếu chúng ta chỉ cần thêm câu lệnh "Bạn có thể suy luận từng bước trước khi đưa ra dự đoán cuối cùng không?" vào cuối lời nhắc trong Hình 4.

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

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.

4 THÍ NGHIỆM

4.1 Cài đặt thí nghiệm

4.1.1 Cửa sổ thời gian dữ liệu

Chi tiết dữ liệu sử dụng trong thí nghiệm được mô tả tại Mục 3.1. Bài báo này tập trung vào dự báo lợi nhuận cổ phiếu NASDAQ-100.

- **Cửa sổ 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.
- **Cửa sổ 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 nghiên cứu 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:

1. **Dữ liệu chuỗi thời gian giá cổ phiếu hàng ngày:** giá mở cửa, đóng cửa, cao nhất, thấp nhất và khối lượng giao dịch.
2. **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.
3. **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)

Nghiên cứu 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 1 và Bảng 2**:

- **Bảng 1:** So sánh hiệu suất của các phương pháp dựa trên LLM và mô hình cơ sở trong dự báo lợi suất cổ phiếu.
- **Bảng 2:** Đánh giá chất lượng lời giải thích được tạo bởi các LLM.

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.

	hàng tuần			hàng tháng		
	Binary Precision	Bin Precision	MSE	Binary Precision	Bin Precision	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

Bảng 1: So sánh hiệu suất giữa các mô hình cơ bản và các mô hình ngôn ngữ lớn (LLM) trong dự báo giá cổ phiếu theo tuần/tháng.

	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	0.2500	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

Bảng 2: Đánh giá chất lượng giải thích bằng cách sử dụng điểm ROUGE, với bản tóm tắt/trích xuất từ khóa của GPT-4 từ các tin tức hàng đầu thực sự mỗi tuần (tìm từ Google Search) làm dữ liệu chuẩn.

4.2.1 Dự báo giá cổ phiếu

Kết quả trong Bảng 1 cho 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.

4.2.2 Chất lượng giải thích

Bảng 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.

5 KẾT LUẬN

Bài báo này 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. Bài báo 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 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 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,
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

- [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] Thomas Conlon, Heather J Ruskin, and Martin Crane. 2009. Cross-correlation dynamics in financial time series. *Physica A: Statistical Mechanics and its Applications* 388, 5 (2009), 705–714.
- [12] Piet De Jong and Ben Zehnwirth. 1983. Claims reserving, state-space models and the Kalman filter. *Journal of the Institute of Actuaries* 110, 1 (1983), 157–181.
- [13] Alexiei Dingli and Karl Sant Fournier. 2017. Financial time series forecasting-a deep learning approach. *International Journal of Machine Learning and Computing* 7, 5 (2017), 118–122.
- [14] Feike C Drost and Theo E Nijman. 1993. Temporal aggregation of GARCH processes. *Econometrica: Journal of the Econometric Society* (1993), 909–927.
- [15] Vijay Prakash Dwivedi and Xavier Bresson. 2020. A generalization of transformer networks to graphs. arXiv preprint arXiv:2012.09699 (2020).
- [16] Jianqing Fan. 2005. A selective overview of nonparametric methods in financial econometrics. *Statist. Sci.* (2005), 317–337.

- [17] Christian Francq and Jean-Michel Zakoian. 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 4 (2004), 605–637.
- [18] Xinyang Geng and Hao Liu. 2023. OpenLLaMA: An Open Reproduction of LLaMA. <https://github.com/openlm-research/openllama> MGhahramani and AThavaneswaran. 2006. *Financial applications of ARM* Am –543.
- [19] Parameswaran Gopikrishnan, Vasiliki Plerou, Yan Liu, LA Nunes Amaral, Xavier Gabaix, and H Eugene Stanley. 2000. Scaling and correlation in financial time series. *Physica A: Statistical Mechanics and its Applications* 287, 3-4 (2000), 362–373.
- [20] Wolfgang Härdle, Helmut Lütkepohl, and Rong Chen. 1997. A review of nonparametric time series analysis. *International statistical review* 65, 1 (1997), 49–72.
- [21] Jan S Henneke, Svetlozar T Rachev, Frank J Fabozzi, and Metodi Nikolov. 2011. MCMC-based estimation of Markov Switching ARMA–GARCH models. *Applied Economics* 43, 3 (2011), 259–271.
- [22] Altaf Hossain and Mohammed Nasser. 2011. Comparison of the finite mixture of ARMA-GARCH, back propagation neural networks and support-vector machines in forecasting financial returns. *Journal of Applied Statistics* 38, 3 (2011), 533–551.
- [23] Tsung-Yu Hsieh, Suhang Wang, Yiwei Sun, and Vasant Honavar. 2021. Explainable multivariate time series classification: a deep neural network which learns to attend to important variables as well as time intervals. In *Proceedings of the 14th ACM international conference on web search and data mining*. 607–615.
- [24] Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, and Honggang Zhang. 2019. Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine* 57, 6 (2019), 114–119.
- [25] Shima Imani, Liang Du, and Harsh Shrivastava. 2023. Mathprompter: Mathematical reasoning using large language models. arXiv preprint arXiv:2303.05398 (2023).
- [26] Søren Johansen. 1995. *Likelihood-based inference in cointegrated vector autoregressive models*. OUP Oxford.
- [27] Norberto Ritzmann Júnior and Julio Cesar Nievola. 2018. A generalized financial time series forecasting model based on automatic feature engineering using genetic algorithms and support vector machine. In *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.
- [28] Nont Kanungsukkasem and Teerapong Leelanupab. 2019. Financial latent Dirichlet allocation (FinLDA): Feature extraction in text and data mining for financial time series prediction. *IEEE Access* 7 (2019), 71645–71664.
- [29] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017).

- [30] Kyoung-jae Kim. 2003. Financial time series forecasting using support vector machines. *Neurocomputing* 55, 1-2 (2003), 307–319.
- [31] Boris Kovalerchuk and Evgenii Vityaev. 2005. Data mining for financial applications. *Data Mining and Knowledge Discovery Handbook* (2005), 1203–1224.
- [32] Bjoern Krollner, Bruce J Vanstone, Gavin R Finnie, et al. 2010. Financial time series forecasting with machine learning techniques: a survey. In *ESANN*.
- [33] Pedro Lara-Benítez, Manuel Carranza-García, and José C Riquelme. 2021. An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems* 31, 03 (2021), 2130001.
- [34] Tim Leung and Theodore Zhao. 2021. Financial time series analysis and forecasting with Hilbert–Huang transform feature generation and machine learning. *Applied Stochastic Models in Business and Industry* 37, 6 (2021), 993–1016.
- [35] Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, et al. 2022. Solving quantitative reasoning problems with language models. arXiv preprint arXiv:2206.14858 (2022).
- [36] Wei Li and Jian Liao. 2017. A comparative study on trend forecasting approach for stock price time series. In *2017 11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID)*. IEEE, 74–78.
- [37] Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. 2023. Let’s Verify Step by Step. arXiv preprint arXiv:2305.20050 (2023).
- [38] Junmei Ma and Xinli Yu. 2013. Research on the Pricing of the Basket Credit Default Swap. *Mathematical Computation* 2, 4 (2013).
- [39] Romit Maulik, Arvind Mohan, Bethany Lusch, Sandeep Madireddy, Prasanna Balaprakash, and Daniel Livescu. 2020. Time-series learning of latent-space dynamics for reduced-order model closure. *Physica D: Nonlinear Phenomena* 405 (2020), 132368.
- [40] Alexey Natekin and Alois Knoll. 2013. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics* 7 (2013), 21.
- [41] Nam Nguyen and Brian Quanz. 2021. Temporal latent auto-encoder: A method for probabilistic multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 9117–9125.
- [42] OpenAI. 2023. GPT-4 Technical Report. ArXiv abs/2303.08774 (2023).
- [43] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]
- [44] Ping-Feng Pai and Chih-Sheng Lin. 2005. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* 33, 6 (2005), 497–505.

- [45] Andrew J Patton. 2012. A review of copula models for economic time series. *Journal of Multivariate Analysis* 110 (2012), 4–18.
- [46] Baolin Peng, Chunyuan Li, Pengcheng He, Michel Galley, and Jianfeng Gao. 2023. Instruction tuning with gpt-4. arXiv preprint arXiv:2304.03277 (2023).
- [47] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A Nunes Amaral, and H Eugene Stanley. 1999. Universal and nonuniversal properties of cross correlations in financial time series. *Physical review letters* 83, 7 (1999), 1471.
- [48] Yuan Qi and Jing Xiao. 2018. Fintech: AI powers financial services to improve people’s lives. *Commun. ACM* 61, 11 (2018), 65–69.
- [49] Sangeeta Rani and Geeta Sikka. 2012. Recent techniques of clustering of time series data: a survey. *International Journal of Computer Applications* 52, 15 (2012).
- [50] Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. 2020. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems* 33 (2020), 12559–12571.
- [51] Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing* 90 (2020), 106181.
- [52] Li Shen, Yan Sun, Zhiyuan Yu, Liang Ding, Xinmei Tian, and Dacheng Tao. 2023. On Efficient Training of Large-Scale Deep Learning Models: A Literature Review. arXiv preprint arXiv:2304.03589 (2023).
- [53] Frank Smets and Raf Wouters. 2003. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European economic association* 1, 5 (2003), 1123–1175.
- [54] Wen Song and Shigeru Fujimura. 2021. Capturing combination patterns of long-and short-term dependencies in multivariate time series forecasting. *Neurocomputing* 464 (2021), 72–82.
- [55] Him Tang, Kai-Chun Chiu, and Lei Xu. 2003. Finite mixture of ARMA-GARCH model for stock price prediction. In *Proceedings of the Third International Workshop on Computational Intelligence in Economics and Finance (CIEF’2003)*, North Carolina, USA. 1112–1119.
- [56] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford Alpaca: An Instruction-following LLaMA model. https://github.com/tatsu-lab/stanford_alpaca. *HugoTouvron, ThibautLavril, GautierIzacard, XavierMartinet, Marie – AnneLachaux, TimothéeLacroix, BaptisteRozière, NamanGoyal, EricHambro, FaisalAzhar, et al. Openandefficientfoundationlanguage models. arXivpreprintarXiv* : 2302.13971(2023).
- [57] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume

- Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. ArXiv abs/2302.13971 (2023).
- [58] Theodore B Trafalis and Huseyin Ince. 2000. Support vector machine for regression and applications to financial forecasting. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, Vol. 6. IEEE, 348–353.
 - [59] Chih F Tsai and Sammy P Wang. 2009. Stock price forecasting by hybrid machine learning techniques. In *Proceedings of the international multiconference of engineers and computer scientists*, Vol. 1. 60.
 - [60] Manuel R Vargas, Beatriz SLP De Lima, and Alexandre G Evsukoff. 2017. Deep learning for stock market prediction from financial news articles. In *2017 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*. IEEE, 60–65.
 - [61] Jar-Long Wang and Shu-Hui Chan. 2006. Stock market trading rule discovery using two-layer bias decision tree. *Expert Systems with Applications* 30, 4 (2006), 605–611.
 - [62] 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).
 - [63] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. arXiv preprint arXiv:2303.17564 (2023).
 - [64] Haiqin Yang, Laiwan Chan, and Irwin King. 2002. Support vector machine regression for volatile stock market prediction. In *Intelligent Data Engineering and Automated Learning—IDEAL 2002: Third International Conference Manchester, UK, August 12–14, 2002 Proceedings 3*. Springer, 391–396.
 - [65] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. 2021. Do transformers really perform badly for graph representation? *Advances in Neural Information Processing Systems* 34 (2021), 28877–28888.
 - [66] Paul D Yoo, Maria H Kim, and Tony Jan. 2005. Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC’06)*, Vol. 2. IEEE, 835–841.
 - [67] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. 2019. Graph transformer networks. *Advances in neural information processing systems* 32 (2019).
 - [68] Jiawei Zhang, Haopeng Zhang, Congying Xia, and Li Sun. 2020. Graph-bert: Only attention is needed for learning graph representations. arXiv preprint arXiv:2001.05140 (2020).

- [69] Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. 2023. Multimodal chain-of-thought reasoning in language models. arXiv preprint arXiv:2302.00923 (2023).
- [70] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. arXiv preprint arXiv:2303.18223 (2023).
- [71] Eric Zivot and Jiahui Wang. 2006. Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®* (2006), 385–429.