

차트 패턴 형성 전조의 비지도 유형화와 확률적 매핑을 통한 중장기 주가 방향 예측에 관한 연구

김진주¹, 정민찬², 김재윤*

순천향대학교

¹wlswn@sch.ac.kr, ²alscks3279@sch.ac.kr, *kimym38@sch.ac.kr

Forecasting Medium- to Long-Horizon Stock Direction via Unsupervised Typing and Probabilistic Mapping of Chart-Pattern Precursors

Kim Jinju¹, Jung Minchan², Kim Jaeyun*

Soonchunhyang Univ.

요약

본 논문은 차트 패턴 완성 직전 전조를 연속형 시퀀스와 이산형 이벤트로 이중 표현하고 비지도 군집으로 유형화하여 전조 패턴과 차트 패턴의 경험적 확률 지도를 구축하여 거래일 예측에 투영하는 방법을 제안한다. 2000-01-01부터 2025-07-15까지의 OHLCV를 70:30으로 분할하여 누수 없이 평가했다. 장기 지평(H)에서 정확도와 F1, Brier가 개선되었고 $L \geq 20$ 에서 spinning top 등 공통 전조 문법이 일관되게 관찰되었다. 본 연구는 차트 패턴을 사후 식별이 아닌 사전 예측 가능한 사건으로 다루며 실무 적용을 위한 해석 가능성을 함께 제시한다.

I. 서론

금융 시장 연구에서 캔들 차트 패턴 인식은 시장 상태를 진단과 매매 의사결정에 널리 활용되어 왔다.[1] 최근에는 패턴 탐지를 자동화하고 정량화하려는 시도가 증가하면서 캔들 차트와 지표 시퀀스를 입력으로 가격 수준이나 방향을 예측하는 연구도 활발히 진행되고 있다. 그러나 기존 연구들의 다수는 패턴이 형성된 뒤 그 성과를 평가하거나 개별 캔들 차트와 지표의 조합으로 가격 자체를 곧바로 예측하는 것에 초점을 둔다.[2]

차트 패턴은 패턴이 완성된 이후에 눈에 띄지만 실제로는 그 이전 단계에서 서서히 만들어진다. 변동성이 서서히 수축하고 가격의 리듬이 바뀌며 추세가 교체될 조짐이 차곡차곡 쌓인다.[3][4] 본 연구는 바로 그 완성 직전의 가격과 거래량, 흐름 등과 같은 이하 전조를 정량화하고 유형화하여 이후 어떤 패턴이 등장할지와 향후 H(30, 60, 90, 120, 150, 180, 200) 거래일에 대한 수익 방향을 확률적으로 예측하는 체계를 제안한다.

본 논문의 핵심 기여는 패턴 이전 전조 시퀀스의 정식화-대규모 정량화, 전조 조합의 빈도-성과 연결로 신호 식별, 패턴 발생 예측과 중장기 수익률 방향 예측까지의 단일 파이프라인 결합이며 이를 통해 차트 패턴을 사후 식별이 아닌 사전 예측이 가능한 사건으로 다루는 관점을 제시한다.

II. 본론

1. 방법론

전조는 연속형과 이산형의 이중 표현으로 구성되며 시계열 시퀀스와 벡터인 연속형 표현에서는 종목·시대별 절대 규모 차이는 제거되고 형상과 리듬을 보존하기 위해 전조 구간의 OHLCV는 특성별 Min-Max 정규화로 0에서 1 사이의 스케일에 맞춘다. 이벤트열과 콤보인 이산형 표현에서는 같은 전조 구간에서 의미 있는 전조 이벤트를 시간순으로 기록하며 연속 쌓인 콤보를 집계하여 전조의 문법을 추출한다.

본 연구는 전체적으로 다음과 같이 진행된다. 슬라이딩 윈도우와 피봇

규칙으로 플래그, 변동성 수축, 더블탑, 더블바텀, 헤드 앤 숄더 등 고전 패턴을 자동 탐지하여 각 사례의 시작과 종료 시점을 확보한다. 탐지된 각각의 사례마다 t start 직전 L을 잘라서 연속형과 이산형 표현을 생성하고 전조 직후 예측 지평(H)별 미래 수익률도 계산하여 방향 라벨을 부여한다. 학습 구간의 전조 1D 시퀀스 L을 표준화한 뒤에 K 평균으로 군집화하며 K는 일반적으로 H 범위에서 실루엣 점수가 최대가 되도록 자동 선택한다. 전조 유형인 각 군집에 대해 이후 관측된 차트 패턴의 분포, 지평별 상승 하락 확률, 빈번한 전조 이벤트 콤보 Top N, 대표 시퀀스 등을 요약한다. 평가 구간의 매 거래일마다 각 종목의 최근 L의 전조 시퀀스를 만들고 학습 시 저장한 표준화 파라미터로 변환한 뒤에 학습 단계의 센트로이드 중 가장 가까운 것을 찾는다. 해당 (L, 전조 유형) 키로 확률 매핑을 조회하여 그날의 패턴과 방향 확률을 얻으며 L이 여러 개인 경우에는 다중 L 앙상블을 적용한다. 동일한 종목과 날짜에 대해 거래일 기준 지평을 적용하여 실제 수익 방향을 계산하며 구간 내 실제 발생한 패턴이 있으면 패턴으로 없으면 방향으로 적용하여 정답 패턴으로 병합한다. 예측 확률과 실제 방향을 비교하여 전체, 종목별에 대해 정확도, 정밀도, 재현율, F1, Brier, ROC AUC를 지평별로 산출하고 전조 콤보의 장기 성과, 전조와 패턴 분류 성능도 함께 정리하여 어떤 전조 콤보가 실증적으로 유의미한지 함께 도출한다.

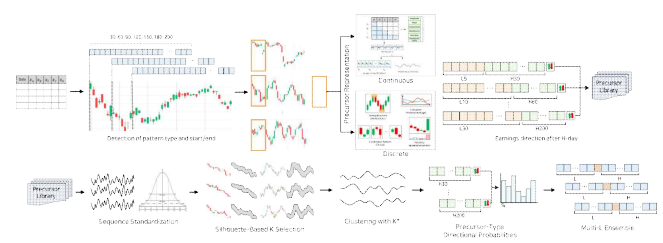


그림1. 전조 유형화 기반 중장기 방향 예측 전체 프레임워크

2. 실험 결과

연구 대상은 다수 종목·지수의 일간 OHLCV이다. 전체 기간은 2000-01-01부터 2025-07-15이고 학습과 테스트 기간은 70:30으로 분할 했다. 전조 창 의 길이는 L[5, 10, 15, 20, 25, 30]이며 예측 지평은 H(30, 60, 90, 120, 150, 180, 200) 거래일로 설정했다. 표준화 파라미터 추정, 군집화, k 선택 및 모든 학습 단계는 학습 구간에서만 수행하여 테스트 구간에 대한 정보 누출을 원천 차단했다. 전조와 이중 표현에 대해서는 어떤 패턴의 시작일을 t start라 할 때 전조는 직전 L 구간 [t start - L, t start - 1]이며 미래 정보 누출을 막고 실제 예측에 필요한 최소 정보를 남기기 위해서 이 구간만을 사용했다. 테스트 시에는 학습된 표준화, 센트로이드, 확률 매 평만을 사용하여 진행하였다.

전조 단일 이벤트의 분포를 보면 창 길이 L이 길어질수록 특정 전조가 뚜렷이 부각 된다. L5에서는 PRE_TREND_UP(L-day Positive Slope)이 639건으로 우세하고 L10에서는 CDLENGULFING(Engulfing Pattern), L15에서는 CDLHIKKAKE(Hikkake Pattern)이 각각 최빈 전조로 나타났다. 반면 L≥20부터는 CDLSPINNINGTOP(Spinning Top)이 압도적으로 늘어나 L20, 25, 30에서 최다를 기록했다. 이는 창이 길수록 봉폭 축소와 결정 지연 같은 미세한 리듬 변화가 더 자주 포착된다는 것을 시사한다.

표1. 룩백(L)별 최빈 전조(Top1 precursor)와 발생 건수

L	precursor_pattern	occurrences
5	PRE_TREND_UP	639
10	CDLENGULFING	747
15	CDLHIKKAKE	1438
20	CDLSPINNINGTOP	2526
25	CDLSPINNINGTOP	3817
30	CDLSPINNINGTOP	5147

전조 패턴과 차트 패턴의 짝지음을 살펴보면 여러 패턴에서 CDLSPINNINGTOP(Spinning Top)이 전조로 가장 자주 동반된다. DNTL(Uptrend Line)에서는 1,002건, FLAGU(Bullish Flag) 2,620건, UPTL(Downtrend Line) 1,306건(으로 관찰됐다. FLAGD(Bearish Flag)에서는 CDLLONGLINE(Long Line Candle)이 최빈 전조로 2,005건에 달한다. HNSD(Head And Shoulders)와 HNSU(Inverse Head And Shoulders) 역시 전조의 평균 발생 횟수가 약 1.68~1.69회 수준에 모여서 단일 전조 이벤트가 한 전조 구간에서 소수 회 반복되는 경향을 보여준다. 패턴의 종류가 달라도 봉의 균형 붕괴 전의 머뭇거림(spining top)과 긴 실체(long line) 같은 형상적 징후가 공통된 문법으로 기능한다는 것을 확인할 수 있다.

표 2. 차트 패턴별 최빈 전조와 발생 통계

chart	precursor	occurrences	avg count
DNTL	CDLSPINNINGTOP	1002	1.776
FLAGD	CDLLONGLINE	2005	1.777
FLAGU	CDLSPINNINGTOP	2620	1.719
HNSD	CDLSPINNINGTOP	2001	1.676
HNSU	CDLSPINNINGTOP	2157	1.688
UPTL	CDLSPINNINGTOP	1306	1.784

예측 성능은 지평이 길어질수록 점진적으로 개선되었다. 방향 예측 기준으로 정확도, 정밀도, F1은 H=30일에 0.512, 0.602, 0.581에서 출발하여 H=200일에는 0.728, 0.728, 0.843으로 상승했다. Brier 점수는 0.252에서

0.207로 감소하여 확률 예측의 평균제곱오차가 줄었다. ROC_AUC는 전 지평에서 0.50대 초반에 머무르는 것을 보여주며 값이 크지는 않지만 분류 지표가 개선되는 양상을 보였다.

표 3. 거래일 지평(H)별 방향 예측 성능

H	accuracy	precision	f1	brier	roc_auc
30	0.512	0.602	0.581	0.252	0.496
60	0.548	0.645	0.652	0.246	0.504
90	0.578	0.671	0.693	0.242	0.510
120	0.688	0.689	0.815	0.226	0.504
150	0.708	0.709	0.830	0.217	0.505
180	0.722	0.723	0.839	0.211	0.506
200	0.728	0.728	0.843	0.207	0.514

III. 결론

본 연구는 패턴 완성 이후가 아닌 완성 직전의 전조 패턴을 분석 단위로 삼아서 전조를 연속형 시퀀스와 이산형 이벤트로 이중 표현하고 비지도 군집으로 유형을 추출한 뒤에 각 전형에 귀결되는 경험적 결과인 패턴과 방향 분포를 확률 지도로 고정하여 날마다 투영하는 방법을 제시했다.

룩백(L)이 커질수록 전조의 형상적 문법이 뚜렷해지며 이는 봉폭 축소 후 결정과 지연 후 추세 전환의 전야라는 공통 서사를 시사한다. 서로 다른 차트 패턴이라도 공통 전조가 반복적으로 동반되어 패턴 간에 공유되는 상위 문법이 존재함을 보여준다. ROC-AUC는 높지 않지만 캘리브레이션과 임계값 조정을 통해 운영 지표가 실무적으로 의미 있는 수준으로 개선될 수 있음을 확인했다.

향후 연구에서는 지평과 레짐별 확률 보정과 임계값 적응화로 운영 지표를 추가 개선, 전조 문법의 길이 확장 및 순서 민감도 모델링,동적 룩백 선택과 다중 L 가중 양상을 학습을 추가하여 진행할 것이다. 또한 거래비용과 슬리피지, 체결 제약과 비용을 포함한 실거래 백테스트로 등으로 확장 시켜서 실전 성과 검증을 통하여 확장 시킬 예정이다.

ACKNOWLEDGMENT

“본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구 결과로 수행되었음”(2021-0-01399)

참 고 문 헌

- [1] Chen, Jun-Hao, and Yun-Cheng Tsai. "Dynamic deep convolutional candlestick learner." arXiv preprint arXiv:2201.08669 (2022).
- [2] Zhu, Min, Yu Guo, and Yuping Song. "A parallel hybrid neural networks model for forecasting returns with candlestick technical trading strategy." Expert Systems with Applications 255 (2024): 124486.
- [3] Shahbandari, Lida, Elahe Moradi, and Mohammad Manthouri. "Stock price prediction using multi-faceted information based on deep recurrent neural networks." arXiv preprint arXiv:2411.19766 (2024).
- [4] Steinbacher, Matej. "Predicting stock price movement as an image classification problem." arXiv preprint arXiv:2303.01111 (2023).