

금융 시계열을 위한 기하학적 국면 인식: 곡률 기반 형태학적 라벨링 접근

Geometric Regime Recognition for Financial Time Series: A Curvature-based Morphological Labeling Approach

유민우¹
Yu Minwoo

¹건국대학교 컴퓨터공학부
E-mail: snowypainter@konkuk.ac.kr

요약

본 연구는 곡률·기울기·가속도, 밴드 내 상대위치, 거래량·이벤트 신호를 통합한 형태학적 라벨러로 주가 시계열의 5국면(Accumulation - EarlyUp - Peak - Distribution - LateDown)을 해석 가능하게 정의하고, LSTM 기반 전이 확률과 멀티태스크(리바운드 확률·예상 수익률·최대낙폭) 정책 점수를 결합해 통온리 EOD 전략을 구성한다. 롤링 워크포워드로 학습·검증을 분리하고, 거래비용·슬리피지·노출 제약을 반영한 테스트에서 $PF \approx 3.91$, $Calmar \approx 21.48$, $Beta \approx 0.17$ 을 확인했다.

키워드 : 형태학적 국면 인식, 곡률 특징, 약지도 기반 확률 학습, 멀티태스크 정책, 주가 시계열

1. 서론

단일 지표는 국면 문맥을 충분히 설명하지 못한다. 곡선의 형태와 이벤트 맥락이 필요하다.

확률적 레짐 모델(regime switching model)은 해석이 어렵다 [1]. 딥러닝 기반 분류는 블랙박스 특성을 갖는다 [2].

목표는 기하학적 라벨링을 통해 해석을 확보하고, 확률 학습과 정책 점수로 실용 성능을 얻는 것이다.

2. 방법론

2.1 형태학적 라벨링

제안 라벨러는 곡선 형태(기울기·이산 곡률), 밴드 내 상대 위치(pos), 변동성·모멘텀, 거래량 질감 및 이벤트 신호(재유입, 반등 후보, 소진, 붕괴 위험 등)를 우선순위 규칙으로 결합해 Accumulation - EarlyUp - Peak - Distribution - LateDown의 5국면을 부여한다.

동률 시에는 Accumulation 국면으로 설정하도록 설계하였다. Peak·Distribution·LateDown과 같은 고확신 레짐에서는 이미 정책 점수 및 포지션 제약이 강하게 작동하므로, 예측 기반이 모호한 시점에 위험노출을 확대하지 않도록 Accumulation을 기본 상태로 정의하였다.

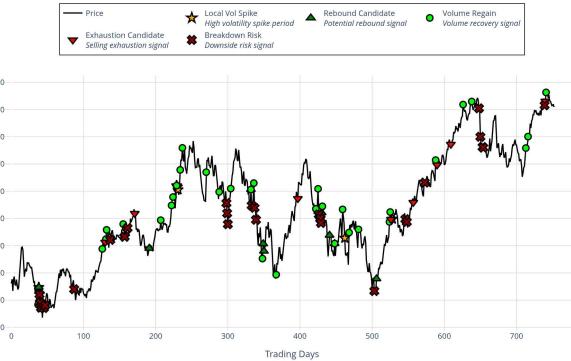


그림 1. AAPL 차트에 형태학적으로 라벨링된 이벤트 시각화

Fig. 1. Example visualization of morphological events on AAPL daily price series.

Fig. 1은 AAPL 일봉에 대해 자동 산출된 이벤트 심볼(▽ Exhaustion, × Breakdown Risk, ☆ Volatility Spike, △ Rebound, ○ Volume Regain)을 중첩한 예시로, 저점 인근에서 Rebound/Volume Regain이 밀집하고 고점 후반에 Breakdown 계열이 분포함을 보여준다. 이 도식은 국면 라벨의 해석 근거를 시각적으로 확인하게 해 주며, 이후 확률 학습 및 정책 점수 산정의 입력 근거가 된다.

2.2 시퀀스 학습

입력 피쳐는 기울기·곡률·가속도 등을 포함한 18개 항목이며, 각 시점에서 20일 구간(Window=20)을 하나의 입력 시퀀스로 사용하였다. 목표 라벨은 2.1절에서 정의한 5가지 형태학적 국면(Accumulation, EarlyUp, Peak, Distribution, LateDown)이다.

감사의 글 : 본 연구는 2025년 건국대학교 컴퓨터 공학부 학부생 개인 연구로 수행되었습니다. 모델 수식의 정제 및 PyTorch 기반 시퀀스 - 정책 파이프로 이어지는 풀 아키텍처 구현 과정에서 OpenAI의 GPT-5 모델을 연구 보조 도구로 활용하여 수식 검증에 도움을 받았습니다.

시계열의 장기 의존성을 학습하기 위해 LSTM 구조를 사용하였다 [3]. 은닉 차원은 64로 설정하였으며, Dropout 비율은 0.1으로 과적합을 완화하였다. 출력층에는 5차원 Softmax 함수를 적용하여 각 시퀀스가 각 국면에 속할 확률을 산출한다.

학습은 시계열 순서를 유지한 워크포워드(rolling walk-forward) 방식으로 수행하였고, 전체 구간을 일정 비율로 슬라이딩하며 각 윈도마다 학습·검증을 반복하고, 폴드별 성능 지표를 집계하였다.

최적화에는 AdamW를 사용하였고, 각 폴드 종료 시 최신 스냅샷을 로드해 추론 안정성을 확보하였다. 손실 함수는 라벨-스무딩 0.08이 적용된 크로스 엔트로피 로스를 사용하며 클래스 불균형은 배치단위 오버샘플링을 통해 완화한다.

2.3 멀티테스크 정책 헤드

정책 헤드는 LightGBM 기반 부스팅 트리 회귀기로 구성하였다 [4]. 입력은 레짐 확률과 핵심 피쳐이며, 향후 수익(5D), 최대 낙폭, 리바운드 여부를 공동 추정한다.

$$S = P_r \cdot \frac{E[ret_{5D}]}{100} - \lambda \cdot \max(0, \frac{-E[dd_{5D}]}{100}) \quad (1)$$

식 (1)의 첫 항 $P_r \cdot (E[ret_{5D}]/100)$ 은 LSTM이 추정한 ‘단기 반등(regime recovery)’ 확률 P_r 과 향후 5일 누적수익 기대값 $E[ret_{5D}]$ 의 곱으로, 보상 기대를 확률적으로 가중한 항이다. 둘의 단위를 맞추기 위해 100으로 나누어 정규화한다.

두 번째 항 $\lambda \cdot \max(0, -E[dd_{5D}]/100)$ 은 기대 최대낙폭 $E[dd_{5D}]$ 가 음수일 때에만(하방 위험이 존재할 때에만) 패널티를 부과하여 위험 회피 강도 λ 로 가중한다.

결과 점수 S 는 종목 내 60일 끝에 Z-정규화 후 분위수 임계값으로 LONG/FLAT/SHORT를 결정한다. 현재 구현에서는 Peak 확률이 높을 때 SHORT 신호가 강화되도록 설계되었다.

2.4 백테스트 설정

과도한 모델 탐색으로 인한 데이터 스누핑을 방지하기 위해 워크포워드(rolling walk-forward) 검증 기반의 백테스트 현실 검정을 적용하였다 [5].

본 연구의 신호는 매 거래일 종가(EOD)에 산출하며, 다음 거래일 시가(T+1 open)에 체결한다.

거래비용은 매수·매도 각각 10 bps로, 체결 슬리피지는 5 bps의 고정 값을 반영하였다.

포트폴리오는 롱-온리로 동일가중을 기본으로 하며, 종목별 포지션 상한과 포트폴리오 총 노출 상한을 적용해 과도한 집중을 방지한다.

평가 지표는 Profit Factor, Calmar, Sharpe, 베타 등으로 구성하였다.

3. 실험 결과

3.1 성능 요약

표 1. 백테스트 성능 지표
Table 1. Backtest Metrics and Values

Metric	Value
Profit Factor (PF)	3.91
Calmar Ratio	21.48
Beta	0.17
Win Rate (%)	50.4

집계 결과, Profit Factor는 약 3.91, Calmar는 약 21.48, 베타는 약 0.17을 기록했으며, 승률은 50.4%, 승률 자체는 평범하지만, 정책 점수와 위험 제약의 결합으로 손익비 개선과 낮은 베타를 동시에 달성했다. 이는 단일 지표 의존 없이 보상 - 위험 균형을 직접 최적화한 효과로 해석된다.

PF·Calmar의 상승과 베타의 하락이 동시에 관측되어, 하방 통제력과 시장 민감도 완화가 병행되었음을 시사한다.

다만 해당 집계는 티커당 평균 거래 횟수가 약 5.7회로 매우 적은 샘플에 기반하며, 일부 집계(예: 연평균 변동성)에서 수치적 오류가 관측되어 해당 지표는 해석에 유의해야 한다.

4. 결론

기존의 레짐 전환 모델 [1] 및 딥러닝 기반 분류 [2]에 비해, 본 연구는 형태학적 해석 가능성과 정책 기반 결합을 제안하였다. EOD 운용 시나리오에서 실거래 제약을 내재화한 성능을 보고하였다. 결과적으로 단순 승률이 아닌 손익비와 하방 통제를 개선하면서 낮은 베타를 달성하였다.

참 고 문 헌

- [1] J. D. Hamilton, “A new approach to the economic analysis of nonstationary time series and the business cycle,” *Econometrica*, vol. 57, no. 2, pp. 357 - 384, 1989.
- [2] H. Chen, Y. Li, and J. Zhang, “Deep regime classification for financial time series,” *IEEE Access*, vol. 11, pp. 45216 - 45228, 2023.
- [3] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735 - 1780, 1997.
- [4] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “LightGBM: A highly efficient gradient boosting decision tree,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
- [5] H. White, “A reality check for data snooping,” *Econometrica*, vol. 68, no. 5, pp. 1097 - 1126, 2000.