

# 다중 라벨을 활용한 코스피 200 주가지수선물 딥러닝 트레이딩 시스템 개발

정민찬, 손우진, 한예찬, 김재윤\*  
순천향대학교

jminchan05@gmail.com, kaldenza21@sch.ac.kr, ychan.sch@gmail.com, \*kimym38@sch.ac.kr

## Development of a Deep Learning Trading System for KOSPI 200 Index Futures Using Multi-Label Classification

Jung Minchan, Shon Woojin, Han Yechan, Kim Jaeyun\*  
Soonchunhyang Univ.

### 요 약

최근 다양한 인공지능 기술을 활용하여 매매를 자동화하는 트레이딩 시스템이 등장하고 있으며 발전하고 있다. 본 논문은 딥러닝 기반 예측 구간 생성 및 Multi Label 프레임워크를 활용한 선물 트레이딩 시스템을 제시한다. KOSPI 지수의 1분봉 데이터를 사용하여, LSTM 앙상블 모델로 예측 구간을 생성하고 다중 라벨을 조합한 트레이딩 전략을 제안한다. 트레이딩 시뮬레이션 결과 전통적인 전략보다 승률, Profit factor 와 같은 성능 지표가 개선된 결과를 확인하였다.

### I. 서론

최근 다양한 인공지능 기술을 활용하여 매매를 자동화하는 트레이딩 시스템에 관한 연구가 진행되고 있다 [1]. 이러한 인공지능 기반 트레이딩 시스템은 인간의 감정 개입 없이 거래를 수행하여 손실을 최소화하고, 더불어 자동 거래로 인해 투자자의 시간을 절약할 수 있다는 장점이 있다. 트레이딩 시스템은 과거의 주가 데이터에서 매매 규칙을 발견한 뒤, 발견한 규칙에 따라 백테스팅(backtesting)을 통해 검증을 수행하는 절차를 가진다. 대다수의 연구들은 매매 시점의 단위를 1 일로 하여 학습 및 백테스팅을 수행하였다 [2]. 이러한 연구들은 효과적인 결과를 도출하였지만 시시각각 변화하는 금융시장에서 하루에 나타나는 수많은 변동을 하나의 포인트로 일반화하는 한계가 존재한다. 특히, 데이트레이딩(day trading)을 위한 트레이딩 시스템은 1 일봉 데이터가 아닌 1 분봉 데이터 이용할 때 더 높은 성능의 모델을 구축할 수 있다는 점에서, 본 연구는 주식시장에서 1 일 단위 데이터가 아닌 1 분봉 데이터를 수집 및 사용한다.

또한, 백테스팅을 수행하기 위해 학습 데이터를 구축하는 과정에서 주가가 상승할지 하락할지 정답을 부여하는 Labeling 작업이 수행되는데 대부분의 회귀 모형을 사용하는 기존 연구에서는 회귀 모형의 예측 값을 기준으로 다양한 Labeling 기법을 시행하였다. [3] 그러나, 이러한 방법은 복잡한 금융시장에서 필연적으로 노이즈를 학습한 예측 모델의 예측 값에 따라 Labeling 이 결정된다는 한계가 존재한다. 따라서 본 연구에서는 회귀 모형의 예측 구간을 활용하여 거래 전략을 제안하고 거래 성과를 확인하고자 한다.

### II. 본론

본 연구의 프레임워크는 Fig. 1 과 같다. 제안된 트레이딩 방법은 LSTM 앙상블 모델을 이용한 예측

구간 생성 및 Multi Label 을 활용한 데이 트레이딩 시스템이며 5 분간 주가지수 선물의 흐름을 이용하여 가격을 예측한다. 모델 학습 기간은 2022 년 2 월부터 2023 년 10 월으로 1 년 8 개월 데이터를 사용하였으며 테스트 기간은 2023 년 11 월부터 2024 년 3 월까지 5 개월로 구성하였다.

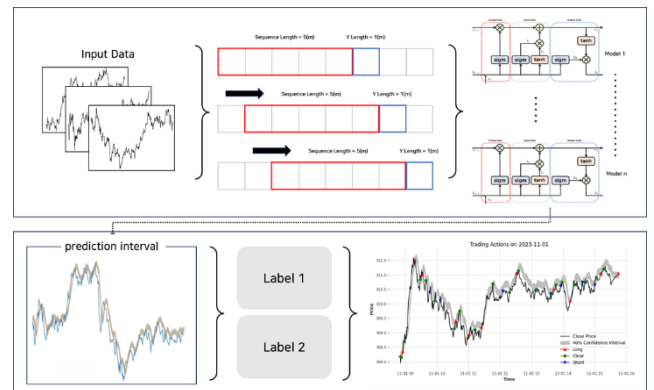


Fig. 1 Prediction Multi Label 트레이딩 프레임워크

### 2.1 Data

본 연구는 KOSPI 200 선물지수를 바탕으로 제안 모델을 적용하였다. 2022 년 2 월 9 일 ~ 2024 년 3 월 29 일 (27 개월) 기간의 데이터를 1 분봉 단위로 open API 를 활용하여 총 211,269 개의 데이터를 수집하였다. 훈련 기간은 2022 년 2 월 9 일 ~ 2023 년 10 월 31 일(22 개월)의 데이터로 테스트 기간은 2023 년 11 월 1 일 ~ 2024 년 3 월 29 일(5 개월)의 데이터로 설정하였다.

### 2.2 LSTM (Long Short-Term Memory)

본 연구에서는 LSTM 앙상블 모델을 사용하여 예측 구간을 생성하는 데 중점을 두었다. LSTM 은 장기 의존성을 학습할 수 있는 RNN 계열의 모델 중 하나로 이러한 장점으로 인해 많은 양의 timestep 에서도 빠르고

효율적으로 학습할 수 있다 [4]. 이러한 LSTM 모델 15 개를 앙상블 하여, 예측 구간을 추정하였다.

### 2.3 Multi Label

트레이딩을 시스템을 개발하기 위해선 예측 결과를 활용한 라벨링 시스템이 필요하다. 본 논문에서는 2 개의 라벨링을 생성하여 조합하는 Multi Label(ML) 시스템에 대해서 제안한다. Label 1 은 실제값이 예측 예측 구간안에 포함되어있을 때 신호를 생성하고 Label 2 는 국소 최소값(current min)을 실시간으로 판단하여 현재의 변화율을 계산하고 임의의 임계값과 비교하여 상승, 하락 추세 신호를 생성한다. Multi Label 은 Label 1 과 Label 2 를 조합하여 Long, Short(포지션 진입), Exit(포지션 청산) 신호를 생성하였다.

### 2.4 트레이딩 시뮬레이션

Multi Label 을 사용하여 트레이딩 시뮬레이션 방법은 다음과 같다. 8시 45분에 장이 열리고 난 후 5분동안은 장의 상황을 파악한 후 진입 신호가 발생하면 발생 시간 증가로 진입하고 이후 청산 신호가 나오기 전까지 보유하며 청산 신호 이후 진입 신호 전까지는 거래를 하지 않는다. 트레이딩 성능을 비교하기 위해 투자 성과 지표(총 4 개)를 측정하였으며 거래 승률(=이익이 발생한 횟수/총 거래 횟수), 거래 횟수, Max Drawdown(MDD, 최대 낙폭), Profit factor(총 수익/총 손실)을 계산하였다.

### 2.5 트레이딩 시뮬레이션 결과

트레이딩 시뮬레이션 결과는 Table 1 과 같다. 본 논문에서는 제안한 예측 구간을 활용한 Multi Label 트레이딩과 벤치마크 전략인 Buy & Hold 와 비교해보면 승률 Profit factor, MDD 등의 모든 성과지표들에서 개선된 결과를 확인할 수 있었다. 이와 같은 결과는 제안 모델이 선물 시장의 추세와 적절한 시점을 찾아 포지션 진입, 청산을 통해 거래를 진행한 것으로 판단된다. Fig 2 는 Multi Label 이 적용된 테스트 기간의 포지션 진입 청산 시점을 보여준다.

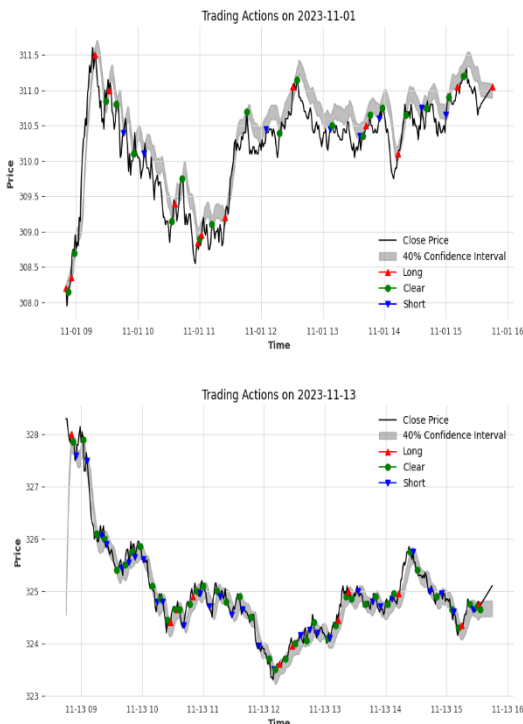


Fig. 2 Multi Label 트레이딩 시뮬레이션

Table 1. 트레이딩 시뮬레이션 결과

트레이딩	거래 횟수	거래 승률	Profit Factor	Max Drawdown
ML ( $\pm 5\%$ )	249	0.63	1.15	-97.45%
ML ( $\pm 10\%$ )	431	0.65	1.35	-35.15%
ML ( $\pm 20\%$ )	756	0.71	1.57	-14.73%
ML ( $\pm 30\%$ )	2161	0.68	2.21	-16.52%
ML ( $\pm 40\%$ )	2352	0.70	3.56	-21.27%
Day_Long	98	0.45	0.86	-294.4%
Day_Short	98	0.54	1.15	-575.44%

### III. 결론

본 연구에서는 LSTM 딥러닝을 활용한 예측 구간 생성 및 다중 라벨을 적용한 선물 트레이딩 시스템을 제안하였다. KOSPI 200 주가지수 선물의 1 분봉 데이터를 기반으로 LSTM 앙상블 모델을 이용하여 예측 구간을 생성하고, 이를 조합한 다중 라벨을 통해 거래를 수행하는 전략을 제안하였다. 트레이딩 시뮬레이션 결과, 전통적인 Buy & Hold 방식과 비교하여 모든 성과지표에서 더 우수한 결과를 보였다.

### ACKNOWLEDGMENT

본 연구는 2024 년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구 결과로 수행되었음. (2021-0-01399) 또한, 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A2C1092808).

### 참 고 문 헌

- [1] Huang, B., Huan, Y., Xu, L. D., Zheng, L., & Zou, Z. (2019). Automated trading systems statistical and machine learning methods and hardware implementation: a survey. *Enterprise Information Systems*, 13(1), 132-144. <https://doi.org/10.1080/17517575.2018.1493145>
- [2] 이우식. (2017). 딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측. *한국데이터정보과학회지*, 28(2), 287-295.
- [3] Kuo, R.J., Chen, C.H., Hwang, Y.C. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets and Systems*, 118(1), 21-45. [https://doi.org/10.1016/S0165-0114\(98\)00399-6](https://doi.org/10.1016/S0165-0114(98)00399-6)
- [4] S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.