

Dart Quant Strategy Series

End-to-End

Pairs Trading Strategy

Choi Munseok, 12th

Choi Wonjun, 13th

May

2024

목차

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

21

1. 소개
2. 사전연구 검토
3. 방법론과 데이터
   1. 주성분분석
   2. 군집화 알고리즘
   3. 성능 평가
   4. 칼만필터(Kalman Filter)
   5. 정상성 검정
   6. 다중 페어트레이딩
   7. 데이터 설명
4. 실증연구
   1. 분석기간
   2. 군집화 기반 앙상블 페어선택 알고리즘
   3. 칼만필터를 활용한 스프레드 동적 추정 알고리즘
5. 결론
6. 부록

**DART 2024 First Semester Pairs Trading Project**

**Choi Munseok**

**cmschs0301@knu.ac.kr**

**Choi Wonjun**

**song1002@knu.ac.kr**

**[Github]**

**https://github.com/DART-KNU/Pairs-Trading-from-Beginning-to-End**

**Cover image source : Space Center Houston**

1. **Introduction**

오늘날

1. **Literature Review**

**III. Methodology and Data**

**3.1 Principal Component An**

**IV. 실증연구**

**4.1 분석기간**

Train/Validation/Test 데이터셋으로 6:2:2 비율로 나눠 연구를 진행하였다.

**4.2 군집화 기반 앙상블 페어선택 알고리즘**

일반적으로 페어트레이딩에서 페어선택 방법은 모든 유가증권 간의 조합을 고려하여 모든 가능한 후보 페어를 선택하는 것이다. 사용가능한 유가증권이 n개라고 했을 때, 가능한 조합의 수는 이다. 이때, 두가지 문제가 발생하게 된다. 첫째, 모든 가능한 조합에 대한 평균회귀 테스트의 시간복잡도는 이다. 따라서 고려되는 유가증권이 늘어날수록, 계산량이 급격하게 증가하게 된다. 둘째, 다중 검정 문제가 발생한다. 하나의 테스트에서 귀무가설을 기각할 확률, 즉 제 1종 오류를 저지를 확률은 이다. 하나의 테스트에서 귀무가설을 기각하지 않을 확률은 이며, 모든 m개의 테스트에서 귀무가설을 기각하지 않을 확률은 이다. 적어도 하나의 테스트에서 귀무가설을 기각할 확률은 가 된다. 즉, 각각의 가설에 대해 유의수준(1종오류)인 검정을 동시에 수행할 경우 전체 오류율이 매우 커지게 된다. 해당부분은 Family-wise type 1 error rate(FWER)로 잘 알려져 있다. 다중 비교 문제의 영향을 완화하는 방법은 두가지이다. 첫째, 본페로니 교정(Bonferroni correction)과 같은 다중 보정테스트를 적용하는 방법이다. 본페로니 교정은 유의수준을 총 검정의 수로 나누어서 개별 검정을 보정하는 방법이다. 하지만, Harlacher[8]는 본페로니 보정이 페어선택에 있어 너무 보수적이 되어 실제로 페어의 발견을 방해한다는 것을 밝혔다. 둘째, 통계적 테스트의 수를 줄이는 것이다. 본 장에서는 Simão Moraes Sarmento[9]가 제안한 군집화 기반 앙상블 페어선택 알고리즘을 사용하여 통계적 테스트의 수를 줄인다. 또한, 해당 알고리즘에 대한 저자의 개인적인 생각을 바탕으로 수정한 새로운 방법론과 성능을 비교한다. 군집화 기반 앙상블 페어선택 알고리즘은 (1) 비슷한 위험요소에 노출된 유가증권끼리 묶여 함께 움직일 가능성이 높으며 (2) 필요한 통계적 테스트의 수가 급감하는 장점이 있다.

**4.2.1 알고리즘1**

Simão Moraes Sarmento[9]가 제안한 알고리즘은 비지도 학습 알고리즘을 사용하여 의미 있는 자산 군집을 추론하여 쌍을 선택하는 것이다. 알고리즘은 다음과 같다.

1. 차원축소 2) 군집화 알고리즘 적용 3) 페어 선택

**4.2.2 알고리즘2**

Simão Moraes Sarmento[9] 제안한 군집화 기반 앙상블 페어선택 알고리즘은

**4.2.3 실증분석**

앞서 소개한 1번 알고리즘과 2번 알고리즘을 상관계수를 제곱한 값 즉 결정계수를 사용하여 평가한다. 상관계수 자체는 두 변수 간의 선형 관계의 강도와 방향을 나타낸다. 시계열 데이터의 군집화 관점에서 결정계수를 사용하여 군집 내 종목들이 얼마나 유사한지를 평가한다. Train 데이터 셋에 알고리즘을 적용하여 군집화를 마친 후 Validation 데이터셋에서 알고리즘의 성능을 평가한다. 알고리즘의 성능을 5회 측정하여 평균한 값은 <표5>와 같다.

|  |  |  |
| --- | --- | --- |
|  | | R-Squared |
| 1번 알고리즘 | Kmeans | 0.1396 |
| DBSCAN | 0.214 |
| Hierarchical | 0.1005 |
| 2번 알고리즘 | Kmeans | 0.1788 |
| DBSCAN | NaN |
| Hierarchical | 0.1881 |

<표 5> 군집화 기반 앙상블 페어선택 알고리즘 성능비교

1번 알고리즘에서 DBSCAN을 사용했을 때 성능이 가장 좋게 나온 것을 확인할 수 있다. 하지만 이때, 군집의 수가 1개이고 군집내 종목의 수가 2이므로 본 연구의 목적에 부합하지 않는다. 따라서 두번째로 성능이 좋으며 군집의 개수와 군집내 종목의 개수가 어느정도 보장되는, 2번 알고리즘의 Hierarchical을 사용하여 분석을 진행한다. 군집내 구성종목은 <표6>과 같다.

|  |  |
| --- | --- |
| 2번 알고리즘 – Hierarchical Clustering | |
| 군집1 | 삼성전자, SK하이닉스, 삼성SDI, LG전자, LG, 한화에어로스페이스, 삼성전기 |
| 군집2 | 동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산 |
| 군집3 | 삼화콘덴서, 삼화전기, 삼화전자 |
| 군집4 | NI스틸, 동양철관, DSR제강, 문배철강 |
| 군집5 | 신한지주, 한국전력, 기업은행, SK텔레콤, KT 강원랜드, 한국가스공사 |
| 군집6 | 현대차, 기아, 현대모비스 |

<표 6> 군집내 구성종목

Validation Set으로 군집내 페어가 선정되는지 또한 확인하였다. 군집3과 군집5를 제외한 모든 군집에서 many-to-many 페어가 선정이 되었으므로, 군집 5를 제외한 나머지 군집으로 거래를 진행한다.

**4.3 칼만필터를 활용한 스프레드 동적 추정 알고리즘**

**4.3.1 one-to-one 페어 선정 및 반감기 계산**

**4.3.2 연관 규칙 분석과 이분 그래프 분할 알고리즘**

**4.3.3 연관 규칙 분석과 이분 그래프 분할 알고리즘**

**4.3.4 칼만필터를 활용한 스프레드 동적 추적 알고리즘**

**4.3.5 실증분석**

**가. 거래 수수료 1%, 보유기간: 반감기**

|  |  |
| --- | --- |
| 군집 | 수익률 |
| 1번 | -8.4474% |
| 2번 | 33.4439% |
| 4번 | -20.0266% |
| 6번 | -12.8396% |

Many-to-Many 페어별 동일 가중 포트폴리오 구성시 포트폴리오 총 수익률 : -7.8697

1번 군집

텍스트, 라인, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

2번 군집

텍스트, 라인, 그래프, 스크린샷이(가) 표시된 사진

자동 생성된 설명

4번 군집

텍스트, 라인, 번호, 그래프이(가) 표시된 사진

자동 생성된 설명

5번 군집

텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

**VI. 부록**

X. Reference

[1] Simão Moraes Sarmento, and Nuno Horta. “Proposed Pairs Selection Framework” A Machine Learning based on Pairs Trading Investment Strategy (2020): 21-35 [[LINK](https://link.springer.com/chapter/10.1007/978-3-030-47251-1_3)]

[2] Hubel, David H., and Torsten N. Wiesel. ＂Receptive fields, binocular interaction and functional Architecture in the cat＇s visual cortex.＂ The Journal of physiology 160.1 (1962): 106. [[LINK](https://link.springer.com/chapter/10.1007/978-3-030-47251-1_3)]

[8] Harlacher M (2016) Cointegration based algorithmic pairs trading. PhD thesis, University of st.Gallen

[9] Moraes Sarmento, S., Horta, N. (2021). Proposed Pairs Selection Framework. In: A Machine Learning based Pairs Trading Investment Strategy. SpringerBriefs in Applied Sciences and Technology(). Springer, Cham. https://doi.org/10.1007/978-3-030-47251-1\_3

본 자료는 경북대학교 금융 데이터분석학회 DART의 제작물로서 모든 저작권은 작성한 학회의 조사분석담당자 본인에게 있습니다. 본 자료는 학회의 동의 없이 어떠한 경우도 변형, 복제, 배포, 전송, 대여할 수 없습니다. 본 자료에 수록된 내용은 학회 및 조사분석담당자가 신뢰할 만한 분석 및 자료로부터 얻은 것이나, 본 학회는 그 정확성과 완전성을 보장할 수 없습니다.