

# 방산주 주가를 활용한 안보 위기 모니터링 index 구축

최민, 류성원<sup>a</sup> and 조성준<sup>b</sup>

[blank line]

<sup>a</sup> 서울대학교 공과대학 산업공학과

08826, 서울시 관악구 관악로 1

Tel: +82-2-880-7025, E-mail: minchoi@dm.snu.ac.kr

Tel: +82-2-880-7025, E-mail: lyusungwon@dm.snu.ac.kr

<sup>b</sup> 서울대학교 공과대학 산업공학과 및 산업시스템혁신연구소

08826, 서울시 관악구 관악로 1

Tel: +82-2-880-6275, E-mail: zoon@snu.ac.kr

## Abstract

국내 안보/국방 위기 상황 시 주식 시장은 그 영향이 가장 빠르게 반영되는 분야 중 하나다. 특히 KOSPI 종목들 중에서도 방산주로 분류되는 종목들의 주가는 위기 상황이 즉각적으로 반영되는데, 본 논문에서는 방산주 주가 움직임을 통해 안보 위기의 파급력을 정량적으로 분석하고자 한다. 정교한 분석을 위해 *encoder-decoder architecture*에 *disentanglement method*를 적용해 주가 변화 중 안보 위기의 영향력만을 추출해 냈다. 이를 지수화 시켜 안보 위기 모니터링 index로 정의한다. 추후 새로운 위기 상황이 발생했을 때 형성된 방산주의 주가 움직임에서 안보 위기 모니터링 index를 뽑아낼 수 있다. 이를 과거 이벤트 때와 비교해 어느 시점과 유사했는지 확인할 수 있고, 추후 사회 파급력을 예측할 수 있겠다.

## Keywords:

국가 위기관리; 안보 위기; 국방 위기; 주식시장; 방산주; Unsystematic & Systematic Risk; disentanglement; adversarial training; encoder-decoder architecture

## 1. Introduction

국가 위기관리 및 안보 위기에 대한 선행 연구는 국가 위기 유형을 정의하고 안보 위기관리 시스템의 발전 방향을 논의하는 방식이 주를 이룬다. 기존 연구 [1]에서는 우리나라의 전통적인 안보 위기 관리에 대한 문제점을 탐색해보고 그 발전 방안으로 1) 전면전을 대비한 대통령의 긴급명령 발동을 위한 법적인 요건 개정, 2) 동원 제도 전/평시 일원화 등을 제시했다(2005년, 김용석). 또한 [2]에 따르면 한국의 국가 위기 관리 체계는 개념적으로 전통 안보, 재난,

국가 핵심 기반으로 나누어져 관장된다. 현 시스템의 한계점으로 국가 위기 상황에 대한 논의가 정부 관료제 틀 안에서 이루어진다는 점과 최근 빠르게 변화하고 있는 위기 발생 양상에 대한 대응이 부족하다는 점을 꼽았다(2014, 부형욱/최수은) 대부분의 선행 연구가 국가 위기 상황에 대한 정책적 접근과 해결 방안 제시로 구성됐다.

본 연구는 전통 안보 위기 상황을 대상으로 특정 이벤트의 파급력을 정량적으로 측정하는 것을 목표로 한다. 이를 위해 주식 시장 데이터를 활용했다. 2010년 연평도 포격 도발, 2013년과 2017년 북한의 전쟁 위협 등 국내 안보 위기 상황 시 주식 시장은 그 영향이 가장 빠르게 반영되는 분야 중 하나다. 특히 방산주로 분류되는 특정 종목군들은 주가에 국방/안보 위기가 빠르게 반영되는데, 본 논문에서는 방산주 주가 움직임을 통해 국가 안보 위기 상황의 파급력을 예측하고자 한다.

정교한 분석을 위해 Disentanglement method를 적용해 주가 움직임에 반영된 시장 전체의 위험(Systematic risk)과 종목/업종 고유의 위험(Unsystematic risk)을 분리하고자 했다. 여기서 시장 위험을 제거한 Residual risk를 안보 위기 이벤트의 심각성(파급력)으로 정의한다. Residual risk에는 종목과 업종 고유의 위험이 녹아있기 때문에 개별 종목 이슈 영향력을 보정하기 위해 대상 종목을 6종목으로 뒀다.

모델은 크게 1st step의 *encoder\_1*, *classifier*과 2nd step의 *encoder\_2*, *decoder*, *adv\_classifier*로 구성된다. 첫 번째 단계에서는 방산주 주가에서 시장과 관련된 성분을 추출해냈고 두 번째 단계에서는 adversarial training 방식을 적용해 업종 고유의 위험을 추출했다. 학습 과정에서 사용한 loss는 reconstruction loss와 adv discriminator loss의 조합으로 구성된다.

## 1.1 Related Works

### Stock Systematic & Unsystematic Risk

포트폴리오 이론(Portfolio Theory) [3] 에 따르면 주식 투자 시 투자자가 감수하는 위험은 Systematic & Unsystematic Risk로 쪼개질 수 있다. Market Risk라고도 불리는 Systematic Risk는 시장 전체의 위험을 대변하며 종목 분산을 통해서 줄일 수 없는 위험을 말한다. 반면 Unsystematic Risk는 개별 회사나 업종 고유의 불확실성 때문에 생기는 위험으로 종목 혹은 업종 분산 투자를 통해 낮출 수 있는 위험이다. 본 논문에서는 주가에 반영되는 방산 업종만의 고유한 위험(Unsystematic Risk)을 뽑아내 이를 국방 위기 과급력으로 정의했다. 안보 위기 이벤트 발생 시 위기의 심각성과 과급력이 높을수록 방산주 주가에 반영되는 업종 고유의 risk도 높아질 것이기 때문이다.

### Disentanglement

Disentanglement는 [4]에서와 같이 neural networks를 활용해 이미지, 음성, 영상 등 데이터에서 content와 style을 구분하는 task를 의미한다. 초기 연구는 manifold learning method를 활용해 appearance에서 형태를 분류해 내는 방식으로 이뤄졌다. 최근에는 variational auto encoders를 통해 digit의 contents와 style을 구분하는 [5]와 같은 연구가 등장하였다. 본 연구에서 사용한 method는 [6]에서 제안한 two-step disentanglement이다. 주어진 dataset(data, label)에 대해 data에서 label과 correlated되어 있는 부분과 아닌 부분을 구분하는 알고리즘이다. 이를 주가 dataset에 적용해 encoder, decoder와 분류기를 학습시켜 주가에 반영된 market, residual risk를 구분하는 것이 목표다.

### Encoder-Decoder Architecture

Encoder-Decoder는 다수의 NLP [7] (Natural Language Process) 및 이미지 [8] 와 관련된 선행연구에서 활용되는 구조이다. 연구 [7] 에서 encoder-decoder의 일반적인 구조를 정의했다. Encoder는 가변적인 size의 주어진 data(source language)를 fixed size vector Z로 압축하는 것이다. Decoder는 input data의 정보를 압축한 Z로부터 다시 input size와 동일한 output을 생성해내는 단계이다. 이 구조는 figure1에 도식화해서 표기했다. Encoder와 decoder에서는 압축, 재생성을 위해 input 데이터의 특성에 따라 MLP(MultiLayer Perceptron), RNN 등을 적용하게 된다. 본 논문에서는 시계열 특성을 가지고 있는 주가 데이터를 input으로 사용했다. 따라서, encoder와 decoder에 MLP와 RNN 알고리즘을 적용해 실험했으며 상대적으로 accuracy가 높았던 MLP를 최종적으로 선택했다.

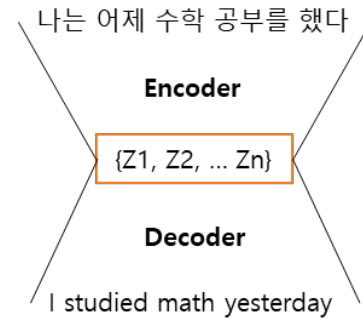


Figure 1 – Encoder Decoder Architecture

## 2. Research Methods

### The Data

연구에 활용한 방산주는 국내 업종 분류 상 우주 항공과 국방으로 분류되는 6개 종목(빅텍, 스펜코, 휴니드, 퍼스텍, S&T 중공업, 한화에어로스페이스-구 한화테크윈)이다. 우주항공과 국방 종목 중 사업 보고서의 사업 부문을 확인해 방산업을 영위하고 있는 업체만으로 분석 대상을 추렸다.

확보한 시계열 데이터는 2004년 1월 1일 이후 각 종목의 일간 시가/종가 자료이며, 가격을 로그 수익률로 변환해 모델에 투입했다. 본 연구에서는 증권사 API 제공 limit으로 일 단위 데이터를 분석했으나 주제의 특성과 딥러닝 알고리즘의 특성 상 좀 더 긴 시계열의 분, 틱 데이터를 투입한다면 보다 정교한 분석이 가능할 것이다.

### Model Structure

모델은 크게 1) 주가에서 시장 영향력(S)만을 뽑아내 분류하는 first step, 2) 나머지 업종 영향력(Z)을 압축하고 추출된 S, Z를 decoder에 태워 주가 흐름을 reconstruction하는 second step으로 나뉜다.

우선 첫 번째 단계(Figure 2)에서 encoder\_1은 방산주 주가 수익률을 보고 동기간 시장(KOSPI)이 어떤 국면인지 구분하는 task를 수행하게 된다. 시장 국면은 KOSPI 수익률을 기준으로 10개의 class로 나눠 정의했다. 학습 과정을 통해 encoder\_1은 시장 국면을 잘 맞출 수 있도록 시장 영향력을 S로 압축하게 된다.

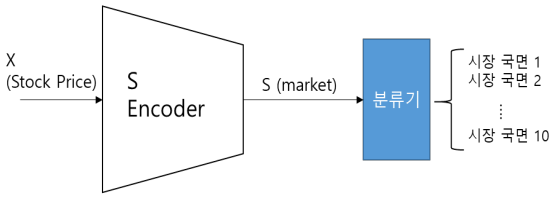


Figure 2 - First step: encoder & classifier

두 번째 단계(Figure 3)에서는 input 데이터를 encoder\_2에 태워 Z에 업종 고유의 영향력을 담는다. 여기서 업종 영향력을 잘 추출하기 위해 adversarial training method를 사용하는데, 1단계에서 학습시킨 classifier를 활용한다. Z를 분류기에 넣어서 시장 국면을 맞추는 경우 패널티(discriminator loss)를 부여해 encoder\_2에서 Z에 시장 예측에 도움이 될 수 있는 성분을 담지 않게 된다.

최종적으로 결과로 나온 두 S, Z 성분을 decoder에 투입해 x'을 reconstruction한다. decoder의 parameter는 input x와 output x'의 차이를 줄이는 방향으로 학습된다. 즉 전체 모델의 loss function은 아래 equation (1) 처럼 reconstruction loss 와 discriminator loss의 합이다.

$$\text{Min}\{L_{reconstruction} - \lambda * L_{adv\_classifier}\}, \lambda > 0 \quad (1)$$

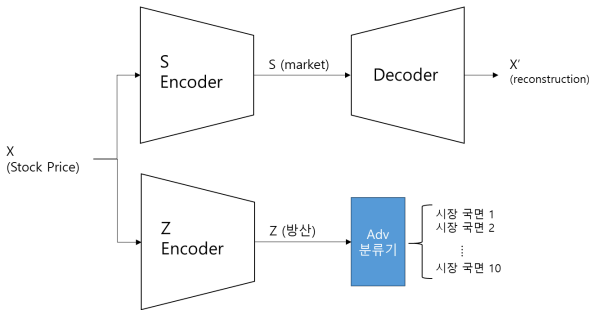


Figure 3 - Second step: encoder-decoder & adv classifier

### Training Details

Encoder와 Decoder는 MLP(MultiLayer Perceptron)를 적용했으며, layer 수와 hidden size를 바꿔가며 loss를 확인했다. 분류기의 classification loss는 cross-entropy, encoder-decoder의 loss는 MSE를 사용했다. Data의 양이 많지 않아 epoch은 최대 100회로 제한했으며, step 1 epoch은 train-test error를 확인해 test error가 증가하는 30회에서 early stopping하였다. 학습 단계에서는 Adam Optimizer를 사용했고, learning rate는 0.0001로 두었다.

### 3. Results

모델에 대한 정량적 평가는 train/test loss 및 accuracy로 확인했다. 정성적으로는 output으로 나온 시장 영향력(S)와 방산주 고유의 위험(Z)을 clustering해 실제 과거 국방 위기 상황 시기의 data들의 특징이 있는지 확인했다.

16차원에 mapping한 방산주 정보(Z)를 1차원으로 변환해 안보 위기 모니터링 index를 정의했다. 이 모니터링 지수를 시계열로 나타내 위에서와 마찬가지로 국방 위기 상황의 지수 변화를 확인했다.

#### Train & Test Loss

첫 번째 단계에서의 분류기는 MLP의 layer 수, hidden 수, output 수 등 parameter를 바꿔가며 시장 예측 정확도를 높이려고 했다. (4-layer, 128 hidden, 16 out - 아래 figure4의 skyblue line) 조합에서 test accuracy가 0.206으로 가장 좋았다.

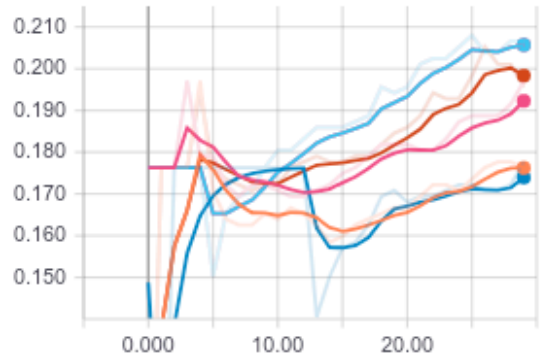


Figure 4 - First step : test accuracy

두 번째 단계에서 total loss는 epoch을 늘릴수록 점진적으로 낮아지는 모습을 보였다. Step 1과 마찬가지로 parameter 세팅을 (4-layer, 128 hidden, 16 out - 아래 figure5의 skyblue line)로 두었을 때 loss가 가장 빠르게 낮아지는 모습을 보였다.

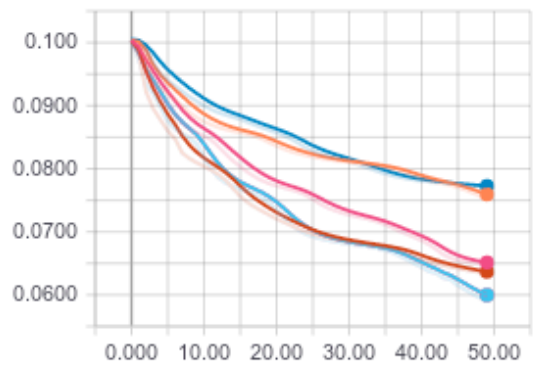


Figure 5 - Second step : test loss

### Clustering Images

정성적 평가를 위해 모델에서 추출한 각 시기의 S와 Z를 클러스터링했다. 우선, step 1의 분류기가 잘 학습되었는지 보기 위해 Z를 3차원 공간에 mapping 시켜 시장 급락/급등기의 정보가 S(시장 영향력)에 잘 압축되었는지 확인했다. Figure 6에서 알 수 있듯이 급락기(class 0), 급등기(class 9)의 데이터들이 3차원 구의 반대 파트에 투영되었다

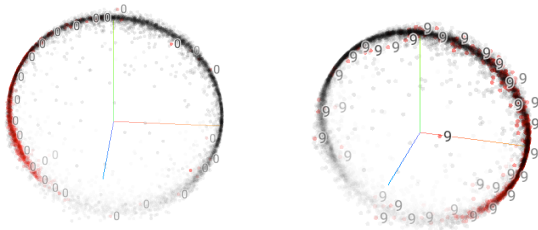


Figure 6 - 시장 영향력(S) 클러스터

방산업 고유의 특성이 담긴 Z의 mapping 결과도 확인했다(Figure 7). 2차원 평면상에서 Z는 포물선 형태를 그리며 포물선 꼭지점에 대부분의 데이터가 몰려 있는 형태를 가진다.

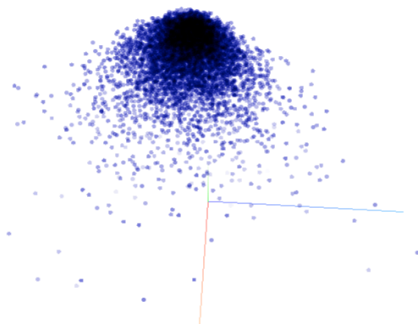


Figure 7 - 업종 영향력(Z) 3D Projection(PCA)

과거 안보 위기 상황의 Z가 평상시 데이터와 구분되는 특징이 있다면 방산주의 업종 영향력을 압축한 Z를 안보 위기를 모니터링 하는 바로미터로 사용할 수 있을 것이다.

Figure 8은 과거 주요 북한의 도발 이벤트의 Z를 labeling한 것이다. 연평도 포격(2010), 정전 무효 선언 및 개성공단 폐쇄(2013) 시기의 Z는 데이터가 sparse한 포물선의 아랫부분에 분포하는 경향을 보인다.

또한 Figure 9은 남한(미국)과 북한의 대립이 격화되었던 2017년 9월과 올해 초 이후 화해/평화 노선을 타기 시작한 2018년 1월의 Z분포를 보여준다. 안보 위기 상황에서 Z는 데이터가 dense하게 모여있는 영역에서 sparse한 영역으로 튀어나가는 행태를 보였다.



Figure 8 - 북한 주요 도발 이벤트 시기의 Z 분포

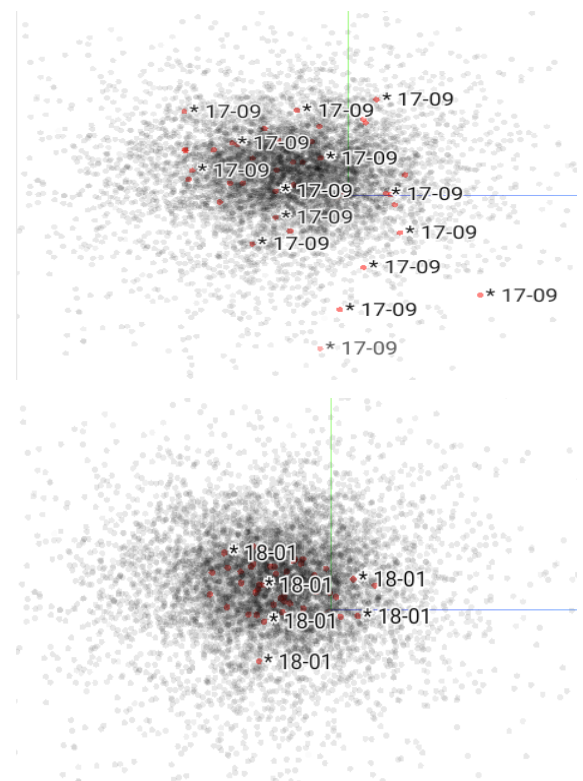


Figure 9 - 갈등 심화기(17/09), 완화기(18/01) Z 비교

Z 분포의 경향성을 확인한 뒤 Z를 안보 위기 index로 변환해 과거 추이를 살펴봤다. 아래 Figure 10은 2017년 1월~2018년 6월까지의 안보 위기 index를 나타낸다.

팜 선제타격 발언 등 트럼프와 김정은의 연이은 설전으로 noise가 많았던 2017년에는 안보 index가 출렁이며 값이 튀는 것을 확인할 수 있었다. 반면, 2018년 들어 화해 무드가 조성된 이후에는 안보 index의 평균적인 level이 내려가 안정되는 모습을 보인다. 본 논문에서 추출해낸 Z가 모니터링 지수로의 역할을 충분히 수행할 수 있다고 판단한다.

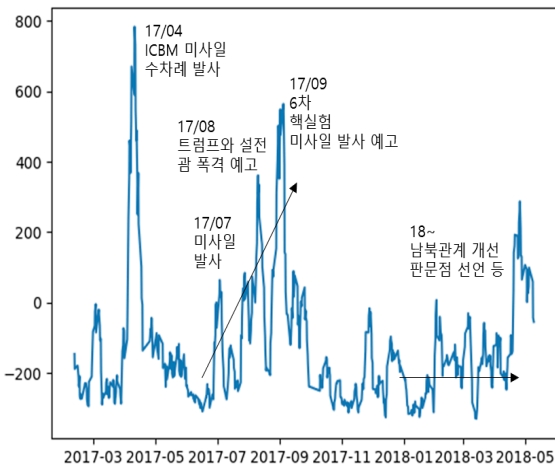


Figure 10 – 2017.01~2018.06 안보 모니터링 지수

#### 4. Conclusion

국가의 국방/안보 위기 상황은 정치적, 외교적 이유로 사전에 예측이 불가능한 영역이라고 할 수 있다. 따라서 국가는 이러한 위기 상황이 발생했을 때 그 파급력을 정확히 예측하고 대응하는 것이 중요하다. 이를 위해 본 논문에서는 방산주 주가 데이터에 딥러닝 알고리즘을 적용해 추출한 안보 위기 모니터링 index를 활용할 것을 제안한다. 제시한 모형을 정성적, 정량적으로 분석한 결과도 우수했다. 위기 상황에서 안보 위기 모니터링 index 및 Z의 변화를 clustering 이미지를 통해 확인할 수 있었다. 본 논문의 의의는 기존의 정성적인 위기 관리 연구에서 벗어나 안보 이벤트의 영향력에 대한 정량적인 평가를 시도한 것에 있다고 하겠다. 모델에 대한 정량적, 정성적 평가를 통해 그 실효성을 입증하였으나 몇가지 한계점이 존재해 이를 추후 연구 과제로 남겨두기로 한다. 우선, 톱, 분, 시간 단위의 데이터를 확보하지 못해 모델 학습에 사용한 dataset의 크기가 작았다. 데이터 수를 늘린다면 보다 의미 있는 결과를 추출할 수 있을 것이다. 또한 가격 데이터뿐만 아니라 거래량, 기술적 지표들을 투입 변수로 추가한다면 모델의 성능이 개선될 여지가 있다고 생각한다.

#### 5. References

[1] 김용석. (2005). 우리나라의 전통적 안보위기관리 발전방안. *한국위기관리논집*, 1(2), 1-14.  
 [2] 부형욱, & 최수은. (2014). 국가위기 양상의 변화와 대응방향. *국방정책연구*, 103, 123-152.  
 [3] Fama, E. F., & French, K. R. (1993). Common risk

factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56.

[4] Lloyd A. Fletcher and Rangachar Kasturi. A robust algorithm for text string separation from mixed text/graphics images. *IEEE transactions on pattern analysis and machine intelligence*, 10(6):910-918, 1988.  
 [5] Huang, F. J., Boureau, Y. L., & LeCun, Y. (2007, June). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*(pp. 1-8). IEEE.  
 [6] Hadad, N., Wolf, L., & Shahar, M. (2017). Two-Step Disentanglement for Financial Data. *arXiv preprint arXiv:1709.00199*.  
 [7] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.  
 [8] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*.