

# 정보 흐름 네트워크를 활용한 국제 금융 시장 지수의 분석 및 국내 시장 지수 등락 예측

최인수 · 김우창<sup>†</sup>

한국과학기술원 산업 및 시스템공학과

## Analyzing Global Financial Market Indices and Predicting Fluctuations of the Korean Market Index Using Information Flow-Based Network Analysis

Insu Choi · Woo Chang Kim

Department of Industrial & Systems Engineering, Korea Advanced Institute of Science & Technology

The stock market has been extensively studied as one of the important research areas of economics and finance. In particular, research on analyzing and predicting stock markets based on stock price and rate of change data is the most active topic in the financial sector, and specifically, predicting stock prices and markets as a whole is one of the important factors for investors to establish optimal investment strategies. In this study, a causal network was constructed based on the information flow of major financial market indices using the concept of transfer entropy. In addition, the financial market was analyzed using the configured network, and the predictive power of KOSPI's fluctuation could be almost maintained using a cluster-based smaller number of whole data when predicting the global financial market index based on the fluctuation of information flows. To conduct this experiment, transfer entropy, which measures the amount of reduction in information uncertainty, was used as an indicator of measuring causal relationships. Specifically, an information flow network was constructed using efficient transfer entropy, an effective indicator in adjusting the finite size effect that may occur when measuring transfer entropy. In other words, we analyzed the causal relationships between global financial indices and predicted KOSPI fluctuation using effective transfer entropy. As a result, it was confirmed that the financial market index could be analyzed using a causal network using efficient transfer entropy, and the improved prediction results could be confirmed using fewer data columns in predicting fluctuations in the domestic financial market using the configured network.

**Keywords:** Financial Market, Information Flow, Transfer Entropy, Network Analysis

### 1. 서론

주식 시장은 경제학 및 금융의 중요한 연구 분야 중 하나로 광범위하게 연구되어 왔다. 특히 주가와 수익률 데이터를 기반으로 한 증시를 분석하고 예측하는 연구는 금융 분야 내에서 가장

활발히 다루어지고 있는 주제이다. 특히, 구체적으로는 주가 및 시장 전반에 대한 예측은 투자자들이 최적의 투자전략을 수립하기 위한 필수 과제 중 하나이다. 이러한 주식 시장의 예측은 시계열과 관련된 가장 중요하고 도전적인 문제 중 하나로 평가받는다(Chen *et al.*, 2017). Fama(1970)에 의해 주창되었으며

<sup>†</sup> 연락저자 : 김우창 교수, 34141 대전광역시 유성구 대학로 291 E2-2 #3107, 한국과학기술원 산업 및 시스템공학과, Tel : 042-350-3129, Fax : 042-350-3110, E-mail: wkim@kaist.ac.kr

2022년 2월 10일 접수; 2022년 3월 18일 수정본 접수; 2022년 5월 23일 게재 확정.

Fama (1991)에 의해 수정된 개념인 무작위 경로를 따르기 때문에 예측할 수 없다는 효율적 시장 가설(efficient market hypothesis)이 확립되었음에도 예측을 통한 수익 발생 모형에 대한 연구는 지속적으로 이루어지고 있다(Weng *et al.*, 2017). 실제로 Fama(1991), Malkiel(2003), Atsalakis and Valavanis(2009), Kumar *et al.*(2016)의 연구 결과에서 요약된 것과 같이 이러한 금융 시장이 효율적이지 못하다는 연구 역시 꾸준히 이루어지고 있으며 시간이 지남에 따라 수익을 창출할 수 있는 예측 모형은 시장이 효율적이지 못함을 의미하는 강력한 증거를 나타낼 뿐만 아니라 금융 포트폴리오 운영에 있어서 매우 중요한 요소이기에 상기한 것과 같이 그 중요성이 역설되고 있다(Henrique *et al.*, 2019).

이러한 금융 시장의 분석 및 예측에 있어 상관 분석은 두 변수의 유사성 정도를 측정하기 위한 간단하고 좋은 지표이다. 상관관계 기반 시계열 분석에 대한 여러 연구가 상관 관계를 바탕으로 금융 시장의 특성에 대해 밝혀낸 바 있다(Plerou *et al.*, 2002; Kim *et al.*, 2011; Kumar and Deo, 2012). 또한, 금융 시장이 복잡계 네트워크로 상정되며 이에 관한 연구가 이루어지면서 Pearson 상관계수 등과 같은 선형 상관관계 외에도 모든 금융 시장의 움직임이 동조화(synchronization)되는 것이 아니라 일정한 시간에 걸쳐서 이루어짐을 확인하게 되면서 금융 시장에 대해서 선형적인 상관관계 외에도 인과관계(causality)와 같이 시차를 둔 형태 등 다른 방법으로 정량화 및 시각화하려는 시도 역시 발생하기 시작했다. 이에 연구자들은 이러한 흐름에서 금융 시장의 비선형적 의존성과 인과관계 지표에 관해서도 관심을 갖고 연구하기 시작했다.

그 결과 시계열 사이의 인과관계를 정의하기 위해 Granger (1969)에 의해 Granger 인과관계가 도입되었으며 이는 시계열의 선형적 인과관계를 측정하는 지표로 널리 사용되기 시작하였다. 이러한 Granger 인과관계는 일반적으로 벡터자기회귀모형(VAR, vector autoregression model)을 전제로 하며 이러한 벡터자기회귀모형은 데이터의 정규성을 바탕으로 해야 한다는 조건이 있다. 하지만 여러 연구자는 가격 데이터나 수익률 데이터는 일반적으로 정규성을 만족하지 않으며 금융 데이터가 실제 경험적 분포가 정규 분포가 아닌 다른 분포, 구체적으로는 Pareto 분포(Pareto distribution), t-분포(t-distribution), Cauchy 분포(Cauchy distribution), Laplace 분포(Laplace distribution) 등에 가깝다는 연구 결과를 제시한 바 있다(Mandelbrot and Taylor, 1967; Bouchaud, 2001; Linden, 2001; Pierou *et al.*, 2001; Cowan and Sergeant, 2001; Young *et al.*, 2006; Young, 2008; Almonte *et al.*, 2011; Beitinger *et al.*, 2015; Lausberg *et al.*, 2020).

이에 기존의 연구들은 주로 중심극한정리에 기반하여 데이터의 정규성을 전제하고 Granger 인과관계를 측정하였다. 하지만 이러한 정규성이 전제되어야 하는 Granger 인과관계의 문제에 대해서 여러 대안이 제시되었으며 일부 연구들은 시스템 내에서 통계 물리학의 개념을 활용하기 시작했다. 경제학 및 금융 분야 역시 경제 또는 금융 시스템에서 금융 시장 간의

관계를 정량화하기 위해 복잡계 시스템 및 정보 이론과 같은 통계 물리학의 개념을 활용하여 측정하려는 여러 연구가 이루어졌다(Mantegna and Stanley, 1999; Noh, 2000; Bonanno *et al.*, 2003; Zunino *et al.*, 2008; Chi *et al.*, 2010; Bekiros *et al.*, 2017).

이러한 흐름에서 Schreiber(2000)가 전이 엔트로피(TE) 개념을 제시하였다. 전이 엔트로피는 Shannon(1948)이 제시한 정보 엔트로피로부터 발생하며, 특정 변수에서 변수로의 정보 불확실성의 감소를 측정하는 비모수적 척도이다. 전이 엔트로피는 시스템 내의 비대칭 상호작용을 포착하고 주행 및 반응 요소를 효율적으로 구분할 수 있는 능력 때문에 신경 과학 및 사회 과학 분야의 연구에 활용되어 왔다. 예를 들어, Vicente *et al.*(2011)은 신경과학 분야에서 전이 엔트로피가 비선형 상호작용에 대한 효율적 연결을 탐지하는 능력을 향상시킬 수 있음을 확인하였고, Faes *et al.*(2013)은 전이 엔트로피를 활용한 심혈관계 및 신경학적 시계열의 분석이 매우 효과적이며, 이를 통해 전이 엔트로피를 활용한 생리학적 메커니즘 조사를 위한 프레임워크를 제시하였다. Kim *et al.*(2016)은 전이 엔트로피를 이용하여 통계적 가정 없이 복잡계 사회적 네트워크 내에서 광역 수준의 정보 전이를 기반으로 한 인과관계 분석이 가능함을 확인하였다.

이러한 흐름에서 예측과 인과관계는 매우 밀접한 관련이 있으며, 상기한 것과 같은 인과관계 네트워크 기반의 연구가 이루어짐에 따라 본 연구에서는 금융 시장 지수의 인과관계 네트워크를 구성하고 여기에 금융 시장 지수의 예측을 접목시키고자 하였다. 이러한 방법을 통해 본 실험에서는 한국을 대표하는 금융 시장 지수인 KOSPI를 타깃으로 금융 시장 지수에 대한 예측력 향상의 여부와 더불어 특정 금융 지수와 밀접한 더 적은 수의 데이터만을 이용하여 더 많은 데이터를 이용한 것보다 더 나은 효과를 낼 수 있는지를 확인하고자 하였다. 이러한 금융 시장 네트워크를 예측에 활용한 예는 Kwon *et al.*(2005)이 개별 주식 단위에서 교차-상관계수를 이용한 네트워크를 구성한 뒤 유전 알고리즘 및 순환 신경망을 활용한 예측을 통해 매수 및 매도 전략을 구성한 연구와 Liu *et al.*(2019)이 금융 지수의 부실성을 금융 시장 네트워크와 그래디언트 부스팅 기계 학습 모형을 이용하여 예측한 연구 등이 있다.

본 연구에서는 이러한 흐름에서 전이 엔트로피 개념을 활용하여 주요 금융 시장 지수들의 정보 흐름을 바탕으로 한 인과관계 네트워크를 구성하고 구성된 네트워크의 연결 여부를 바탕으로 국제 금융 시장 지수를 예측할 때 국내 주식 시장의 대표적인 지수인 KOSPI의 등락에 대한 예측력이 개선될 수 있는지 확인하는 것을 목표로 하였다.

본 연구의 구성은 다음과 같다. 제2장에서는 데이터에 대해 설명하며 제3장에서는 연구 방법론에 대해 설명한다. 제4장에서는 제3장의 연구 방법론을 바탕으로 한 금융 시장 지수 네트워크를 구성한다. 제5장에서는 제4장에서 구성된 네트워크를 바탕으로 하여 국내 금융 시장의 예측 결과에 대해서 분석하였다. 제6장에서는 연구의 결론, 한계점 및 의의에 대해 설명한다.

## 2. 데이터 및 연구 방법론

### 2.1 데이터 설명 및 수집

본 연구에서는 2016년 1월 1일부터 2021년 11월 30일까지의 기간을 연구를 위한 데이터 수집의 기간으로 상정하였다. 이 중 금융 시장 지수 간의 수익률의 흐름에 대한 네트워크의 도식화를 위한 실험 기간 및 기계 학습의 학습을 위한 실험 기간은 2016년 1월 1일부터 2020년 12월 31일이었다. KOSPI의 학습된 모형에 대한 테스트 기간은 2021년 1월 1일부터 2021년 11월 30일로 설정하였다. 실험에 사용된 국제 금융 시장 지수는 <Table 1>과 같이 2020년 기준 국제 금융 시장의 시총을 기준으로 연구 기간 내 데이터가 존재하는 해당 금융 시장을 관리하는 상위 20개 국가와 이를 대표하는 금융 지수 20개를 선정하였으며 여기에 기타 주요 지수(EuroStoxx, 금, 원유, 비트코인)를 선택하여 총 24개의 데이터 열을 생성하였다. 또한 시각화의 편의성을 위하여 각 국가 및 금융 시장 지수는 가장 우측에 있는 열과 같이 약어로 표기하였다. 이 때 국가의 경우 ISO 3166-1 Alpha-3 코드를 사용하였다. 본 연구에서는 연구에서 활용되는 국가별로 최대 시차 차이가 24시간 이상이 나는 경우에는, 일별 데이터의 일자가 달라짐에 따라 본 연구에서 정의하고자 하는 인과관계의 시차에 대한 전제가 필요하나 본 연구에서 사용된 금융 시장 간의 최대 시차 차이는 24시간보

다 작으므로 인과관계의 계산에는 유의미한 차이가 없을 것으로 판단하고 실험을 진행하였다.

### 2.2 기초통계량

2.1절에서 언급된 데이터의 로그 수익률에 대한 기초 통계량 값은 위와 같다. <Table 2>에서 \*, \*\*, \*\*\*은 각각  $\alpha=0.1$ ,  $\alpha=0.05$ ,  $\alpha=0.01$  수준에서 해당 검정통계량이 통계적으로 유의미함을 의미한다. 그 결과, 정규성은 본 논문에서 전이 엔트로피를 활용하게 된 중요한 이유 중 하나이므로 결과의 강건성을 위하여 두 가지 검정, Jarque-Bera 검정과 Shapiro-Wilk 검정을 시행하였으며, 그 결과 24개의 모든 데이터 열에 대해서 두 검정 모두에서 정규성을 만족하는 데이터 열은 없음을 Jarque-Bera 검정의 검정 통계량 JB와 Shapiro-Wilk 검정의 검정 통계량  $W$ 을 통해 확인할 수 있다. 또한 정상성(stationarity)을 측정하기 위하여 Augmented Dickey-Fuller 검정을 실시하였으며, Augmented Dickey-Fuller 검정의 검정통계량 DF를 통해 24개의 모든 데이터 열이 정상성을 만족하지 않음을 확인하였다. 이러한 통계 검정 결과는 본 연구에서 후술할 내용과 같이 Granger 인과관계와 비교해 통계적 성질에 대한 전제로부터 자유로운 전이 엔트로피 개념을 이용하여 금융 시장 네트워크를 구축하는 데에 타당한 통계적 근거로 활용될 수 있다.

Table 1. Selected global market indices(Worldbank, 2021)

Country	Financial index	Abbreviation
United States	Standard and Poor's 500	USA(S&P500)
China	Shanghai Stock Exchange Composite Index	CHN(SSE)
Japan	Nikkei 225	JPN(NIKKEI225)
Hong Kong	Hang Seng Index	HKG(HSI)
United Kingdom	Financial Times Stock Exchange 100	GBR(FTSE100)
Canada	Standard and Poor's Toronto Stock Exchange Index	CAN(TSX)
India	Bombay Stock Exchange Sensex Index	IND(SENSEX)
Saudi Arabia	Tadawul All Share	SAU(TAS)
Germany	Deutscher Aktienindex	DEU(DAX)
South Korea	Korea Composite Stock Price Index	KOR(KOSPI)
Switzerland	Swiss Market Index	CHE(SMI)
Australia	Standard and Poor's Australian Securities Exchange 200	AUS(ASX200)
Iran	Tedpix	IRN(TEDPIX)
South Africa	South Africa Top 40	ZAF(SA40)
Brazil	Bovespa	BRA(BVSP)
Spain	Iberian Index 35	ESP(IBEX35)
Russian Federation	Moscow Exchange Index	RUS(MOEX)
Singapore	Financial Times Stock Exchange Singapore	SGP(FTWISGPL)
Thailand	Stock Exchange of Thailand Index	THA(SET)
Indonesia	Indonesia Stock Exchange Index	IDN(IDX)

Table 2. Descriptive Statistics

Financial Index	Average	Standard Deviation	Minimum	Maximum	Q1	Median	Q3	Skewness	Kurtosis	W	p-value (W)	JB	p-value (JB)	DF	p-value (DF)
US(S&P 500)	0.0006	0.0116	-0.1277	0.0897	-0.0029	0.0008	0.0053	-1.1422	22.7413	0.7978	0.0000***	32139.2102	0.0000***	-11.8845	0.0000***
China(SSE Composite Index)	0.0001	0.0113	-0.0804	0.0555	-0.0049	0.0006	0.0055	-0.9475	7.0195	0.9148	0.0000***	3140.8410	0.0000***	-17.5039	0.0000***
Japan(NIKKEI 225)	0.0003	0.0128	-0.0825	0.0773	-0.0054	0.0006	0.0065	-0.2022	5.7324	0.9290	0.0000***	1967.6890	0.0000***	-15.5226	0.0000***
Hong Kong(Hang Seng)	0.0001	0.0117	-0.0572	0.0492	-0.0061	0.0008	0.0069	-0.4237	1.9633	0.9752	0.0000***	275.1089	0.0000***	-39.0400	0.0000***
UK(FTSE 100)	0.0001	0.0106	-0.1151	0.0867	-0.0045	0.0006	0.0052	-1.0709	16.2227	0.8810	0.0000***	16561.3404	0.0000***	-12.5651	0.0000***
Canada(S&P TSX)	0.0003	0.0103	-0.1318	0.1129	-0.0030	0.0008	0.0043	-2.0842	50.6303	0.6883	0.0000***	158379.3691	0.0000***	-12.3829	0.0000***
India(BSE Sensex)	0.0005	0.0114	-0.1410	0.0859	-0.0043	0.0008	0.0061	-1.6197	25.4506	0.8330	0.0000***	39649.2574	0.0000***	-11.0492	0.0000***
Saudi Arabia (Tadawul All Share)	0.0003	0.0107	-0.0868	0.0683	-0.0042	0.0007	0.0057	-1.1379	10.8445	0.8844	0.0000***	7460.4110	0.0000***	-13.0756	0.0000***
Germany(DAX)	0.0002	0.0138	-0.1202	0.0923	-0.0058	0.0007	0.0071	-1.2129	13.2195	0.8754	0.0000***	10901.4922	0.0000***	-14.2278	0.0000***
Republic of Korea (KOSPI)	0.0003	0.0106	-0.0877	0.0825	-0.0040	0.0007	0.0058	-0.2985	9.9300	0.9023	0.0000***	5883.4861	0.0000***	-23.6737	0.0000***
Switzerland(SMI)	0.0002	0.0095	-0.1013	0.0678	-0.0043	0.0006	0.0050	-1.0183	13.1323	0.9081	0.0000***	10863.4355	0.0000***	-39.0720	0.0000***
S&P ASX 200)	0.0002	0.0102	-0.1020	0.0677	-0.0039	0.0008	0.0051	-1.3616	16.1390	0.8522	0.0000***	16636.5258	0.0000***	-9.7237	0.0000***
Iran(TEDPIX)	0.0022	0.0133	-0.0634	0.0438	-0.0028	0.0009	0.0069	-0.0420	2.2277	0.9351	0.0000***	291.4920	0.0000***	-12.3667	0.0000***
South Africa(South Africa Top 40)	0.0002	0.0122	-0.1045	0.0906	-0.0056	0.0007	0.0066	-0.5285	10.1308	0.9167	0.0000***	6337.1450	0.0000***	-13.7896	0.0000***
Brazil(Bovespa)	0.0006	0.0171	-0.1599	0.1302	-0.0074	0.0011	0.0096	-1.2992	16.3979	0.8720	0.0000***	16670.5274	0.0000***	-13.4445	0.0000***
Spain(IBEX 35)	-0.0001	0.0131	-0.1515	0.0823	-0.0059	0.0004	0.0061	-1.7886	22.2852	0.8605	0.0000***	31917.0870	0.0000***	-13.4321	0.0000***
Russia(MOEX)	0.0003	0.0160	-0.0888	0.0883	-0.0080	0.0000	0.0095	-0.1656	2.3573	0.9762	0.0000***	348.7780	0.0000***	-37.8681	0.0000***
Singapore(FTSE Singapore)	0.0001	0.0099	-0.0826	0.1015	-0.0042	0.0000	0.0046	0.0134	17.5827	0.8618	0.0000***	19661.4793	0.0000***	-15.4846	0.0000***
Thailand(SET Index)	0.0002	0.0101	-0.1143	0.0765	-0.0039	0.0005	0.0047	-1.9308	26.7956	0.8065	0.0000***	43661.3450	0.0000***	-9.5879	0.0000***
Indonesia(Jakarta Composite Index)	0.0003	0.0102	-0.0681	0.0970	-0.0046	0.0007	0.0054	-0.0934	10.0150	0.9147	0.0000***	5952.4901	0.0000***	-12.2923	0.0000***
EU(Eurostoxx 50)	0.0002	0.0120	-0.1324	0.0883	-0.0047	0.0005	0.0056	-1.3793	17.3493	0.8663	0.0000***	19359.1366	0.0000***	-12.8298	0.0000***
Gold/USD	0.0003	0.0085	-0.0589	0.0469	-0.0039	0.0006	0.0048	-0.3593	4.3211	0.9553	0.0000***	1220.4851	0.0000***	-38.0032	0.0000***
Bitcoin/USD	0.0023	0.0406	-0.4973	0.2276	-0.0132	0.0022	0.0195	-0.9125	13.3728	0.9051	0.0000***	16313.0964	0.0000***	-32.4521	0.0000***
WTI/USD	0.0008	0.0308	-0.2822	0.3196	-0.0109	0.0021	0.0127	0.2014	28.0467	0.7704	0.0000***	50766.9886	0.0000***	-9.6270	0.0000***

### 3. 연구 방법론

#### 3.1 전이 엔트로피

변형이 없는 기본적인 형태의 그래인저 인과관계를 산출하기 위해서는 정규성, 정상성, 선형성과 같은 데이터의 통계적 성질에 대한 가정이 전제된다. 하지만 2장의 결과에서 확인할 수 있듯 본 연구에서 사용된 데이터 세트는 전부 정규성을 전혀 만족하지 못하였으며, 이는 금융 데이터의 변화량 데이터가 일반적으로 정규성을 만족하지 못한다는 기존의 연구 결과를 따른다고 볼 수 있다. 이러한 결과에 따라 본 연구에서는 위에 언급된 가정 없이 사용할 수 있는 방법론을 활용하고자 하였고, 그 결과 정보 이론을 기반으로 하여 인과관계를 측정하는 Schreiber(2000)가 제시한 전이 엔트로피(transfer entropy) 개념을 활용하였다. 전이 엔트로피는 Shannon 엔트로피를 기반으로 두 변수 사이의 정보 흐름을 확인하기 위한 비모수 측정 방법론이다. 전이 엔트로피는 인과관계를 정보의 흐름에 따른 해소된 불확실성을 측정한다는 아이디어를 바탕으로 시스템 내 인과관계를 정량화하고 선행 변수

(source)로부터 종속 변수(target)에 대한 인과관계를 크기를 측정한다. 이러한 전이 엔트로피는 정보 또는 물리학 분야를 넘어서 신경 과학, 전기 공학, 화학 공학 분야에서도 인과관계를 측정하는 지표로써 사용되고 있다. 엔트로피 이론은 데이터의 성질과 무관하게 모형을 활용할 수 있으며, 비선형적 인과관계 역시 측정할 수 있다는 장점이 있으며 이러한 장점은 엔트로피를 기반으로 한 인과관계 지표인 전이 엔트로피 역시 유효하다.

이러한 전이 엔트로피를 계산하기 위해서는 우선 정보 엔트로피의 정의가 필요하다. 본 연구에서는 Shannon(1948)이 제시한 정보 엔트로피 개념을 활용하였으며, Shannon의 정보 엔트로피  $H(X)$ 는 확률공간  $\Omega$ 에서 확률질량함수  $p(x)$ 를 갖는 확률변수  $X$ 에 대해서 식 (1)과 같이 정의된다. 이러한 엔트로피를 바탕으로 전이 엔트로피는 두 시계열 변수  $X_t$ 와  $Y_t$ 에 대해서 식 (2)와 같이 정의된다. 이 때 식 (2)의  $k$ 와  $l$ 은 각 시계열  $X_t$ 와  $Y_t$ 에서 전이 엔트로피의 계산에 활용되는 시차(lag)를 의미하며,  $x_t^{(k)} = \{x_t, x_{t-1}, \dots, x_{t-k+1}\}$ ,  $y_t^{(l)} = \{y_t, y_{t-1}, \dots, y_{t-l+1}\}$  인 시계열을 의미한다.

$$H(X) = - \sum_{x \in \Omega} p(x) \log_2 p(x) \quad (1)$$

$$\begin{aligned} TE_{Y \rightarrow X}^{(k,l)}(t) &= H(X_{t+1} | X_t, \dots, X_{t-k+1}) \quad (2) \\ &- H(X_{t+1} | X_t, \dots, X_{t-k+1}, Y_t, \dots, Y_{t-l+1}) \\ &= \sum_i p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 p(x_{t+1} | x_t^{(k)}, y_t^{(l)}) \\ &- \sum_i p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 p(x_{t+1} | x_t^{(k)}) \\ &= \sum_i p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 \frac{p(x_{t+1} | x_t^{(k)}, y_t^{(l)})}{p(x_{t+1} | x_t^{(k)})} \end{aligned}$$

본 연구에서는 두 시계열의 시차  $k$ 와  $l$ 에 대해서  $k=l=1$ 인 경우를 상정하였다. 이는 가장 짧은 시차 형태로 이러한 시차 형태를 사용하는 것은 약형 효율적 시장가설(efficient market hypothesis), 즉 무작위 행보 이론(random walk behavior)을 가정하기 때문이다(Bekiros *et al.*, 2017). 약형 효율적 시장가설은 가격 움직임, 물량, 수익 데이터가 주가에 영향을 미치지 않고 향후 방향을 예측하는 데 사용될 수 없다는 주장이며 이러한 형태의 가정을 할 경우 가장 짧은 시차를 통해 TE를 계산하게 되는 것이 가장 적합한 시차로 해석될 수 있다. 이렇게 될 경우 전이 엔트로피는 다음과 같이 표현될 수 있다.

$$\begin{aligned} TE_{Y \rightarrow X}^{(1,1)}(t) &= \sum_i p(x_{t+1}, x_t, y_t) \log_2 \frac{p(x_{t+1} | x_t, y_t)}{p(x_{t+1} | x_t)} \quad (3) \\ &= \sum_i p(x_{t+1}, x_t, y_t) \log_2 \frac{p(x_{t+1}, x_t, y_t) p(x_t)}{p(x_{t+1}, x_t) p(x_t, y_t)} \end{aligned}$$

이러한  $k=l=1$  시차 하에서 본 연구에서는 엔트로피를 기반으로 한 인과관계 측정 지표인 전이 엔트로피(transfer entropy, TE)의 발전된 형태로 효율적 전이 엔트로피(effective transfer entropy, ETE)를 활용하고자 하였다. 이러한 효율적 전이 엔트로피는 유한 시계열에서 발생할 수 있는 유한크기효과(finite size effect)를 제거한 형태이다. 이는 Marschinski and Kantz(2002)에 의해 처음 제시되었으며 여러 가지 형태가 있으나 본 연구에서는 Boba *et al.*(2015)에서 제시한 두 시계열을 모두 섞어 계산된 식 (4)과 같은  $Z$  값과 같은 형태의 효율적 전이 엔트로피 값을 사용하기로 하였으며, 총 1,000회의 시뮬레이션을 실시하였다. 식 (4)의  $\overline{TE}_{shuffle}$ 은 두 시계열을 모두 섞어 1,000회 반복된 전이 엔트로피의 평균값  $\sigma_{shuffle}$ 은 두 시계열을 모두 섞어 1,000회 반복된 계산된 전이 엔트로피의 평균값에 대한 표준 편차를 의미한다.

$$ETE = Z = \frac{TE - \overline{TE}_{shuffle}}{\sigma_{shuffle}} \quad (4)$$

전이 엔트로피는 금융 분야에서 주로 2000년대 후반부터 금융자산과 금융 시장의 인과관계를 확인하는 데 사용됐다.

2000년대에 Marschinski and Kantz(2002)는 독일 DAX Xetra Stock Index(DAX)와 미국 Dow Jones Industrial Average Index 사이의 인과관계를 보고하였으며, Kwon and Yang(2008)은 정보 이전의 방향을 보여준 결과, 시장 지수가 미국 증시의 개인 주식에 영향을 미친다는 것을 발견했다. 2010년대 이후에는 Dimpfl and Peter(2013)는 신용위험의 가격 책정에 대한 회사 채 시장에 대한 신용부도스왑 시장의 인과관계와 VIX와 iTraxx Europe의 시장위험과 신용위험 사이의 동적 관계를 위기 이전, 위기 이후 기간의 관점에서 분석했다. Sensoy *et al.*(2014)은 효율적 전이 엔트로피를 사용하여 몇몇 신흥국가의 환율과 주가 간 정보 흐름의 강도와 방향을 조사했다. 전이 엔트로피를 기반으로, Bekiros *et al.*(2017)는 미국 주식 및 상품 시장에서 네트워크 역할을 전이 엔트로피를 기반으로 산출하였다. 그리고 Lim *et al.*(2017)은 전이 엔트로피를 기반으로 신용부도스왑 시장과 미국의 주식시장 간의 정보 흐름을 구조를 분석하였다. 최근에 Jang *et al.*(2019)은 TE를 사용하여 비트코인, 금, S&P 500 지수, 미국 달러 간의 인과관계를 연구하였으며 Yue *et al.*(2020, 2020)은 중국 금융 상품 간의 수익률의 인과관계에 대해 TE를 이용하여 분석하고, 또한 미국과 중국의 금융 시장에서의 정보 흐름에 대해 비교 분석했다. 이러한 선행 연구를 바탕으로 본 연구에서는 전이 엔트로피가 본 연구의 목적인 금융 시장 지수 네트워크의 구축과 이를 바탕으로 한 KOSPI의 예측에 있어 적합하다고 판단하였고, 이를 인과관계의 지표로 활용하였다.

이때, 전이 엔트로피는 두 시계열 데이터의 확률분포를 이산화한 후 산출이 가능한데 본 연구에서는 가장 보편적으로 사용되는 시계열 데이터 이산화 방법론인 동일한 간격으로 정의된 히스토그램을 활용하였으며, 이 때 히스토그램 구간의 수는 주어진 데이터에 대해 히스토그램 기반으로 결합 엔트로피를 계산할 때 수치해석적으로 최적의 히스토그램 구간 개수 제시한 Hacine-Gharbi and Ravier(2018)의 연구를 참조하였다.

### 3.2 네트워크 분석

#### 3.2.1 네트워크 밀도(Network Density)

네트워크의 밀도는 네트워크 노드들 사이에서 연결된 정도를 의미한다.  $D_i$ 를 금융 시장 지수  $i$ 의 금융 시장 지수 네트워크의 밀도로 정의할 때, 금융 시장 지수 네트워크 내의 주식의 수가  $n_i$ 개라 할 때, 발생 가능한  $n_i(n_i - 1)$ 개의 인과관계 중 통계적으로 유의미한 인과관계의 개수  $k_i$ 의 비율로 정의한다. 이를 요약하면 식 (5)과 같다.

$$D_i = \frac{k_i}{n_i(n_i - 1)} \quad (5)$$

#### 3.2.2 연결 중심성(Degree Centrality)

연결 중심성은 전이 엔트로피에 의해 인과관계를 갖는 주

식의 개수를 집계하는 것으로 내향 연결 정도 중심성과 외향 연결 정도 중심성으로 구분된다.

$$N_j^I: (S \rightarrow j) = \frac{1}{N-1} \sum_{i \neq j} (i \rightarrow j) \quad (6)$$

$$N_j^O: (j \rightarrow S) = \frac{1}{N-1} \sum_{i \neq j} (j \rightarrow i) \quad (7)$$

식 (6)의  $N_j^I$ 는 내향 연결 중심성(in-degree centrality)으로 구성된 금융 시장 지수 네트워크  $S$  내에서 효율적 전이 엔트로피에 의해 영향력을 받는 다른 금융 시장 지수의 개수를 의미하며 식 (7)의  $N_j^O$ 는 외향 연결 중심성(out-degree centrality)으로 금융 시장 지수 네트워크  $S$  내에서 효율적 전이 엔트로피에 의해 영향력을 행사하고 있는 다른 금융 시장 지수의 개수를 의미한다. 이 때  $i \neq j$ 이다.

### 3.2.3 페이지랭크(PageRank)

중심성은 네트워크 분석에 있어서 가장 많이 사용되는 지표로서, 노드가 가지는 영향력을 측정하기 위해 중심에 위치하는 정도를 측정된 지표이다. 이 중 Brin et al.(1999)이 고안한 PageRank는 자사 검색엔진에서 웹페이지의 순위를 측정하기 위해 고안한 중심성 측정 알고리즘으로, 네트워크상의 경로들을 고려하여 노드의 중심성을 결정하는 고유벡터 기반 위계 중심성 지표 중 가장 유명한 방법이다. 특히, 유사한 측정 방법인 고유벡터 중심성(eigenvector centrality)과 Katz 중심성(Katz centrality) 등의 지표의 단점을 보완한 방법으로 알려져 있다. PageRank는 유향 그래프(directed graph)의 연결 분석 방법으로, 각 노드의 영향 그리고 현재 및 미래의 정보 흐름의 세 요소를 모두 측정할 수 있다.

PageRank를 계산하는 방법은 다음과 같다. 네트워크에  $N$ 개의 노드가 있을 때, 이 네트워크는  $\mathbf{A} \in \mathbb{R}^{N \times N}$  라는 인접행렬(adjacency matrix)로 나타낼 수 있다. 이 인접행렬을 통해 각 노드의 PageRank를 원소로 하는 열벡터  $\mathbf{r} \in \mathbb{R}^N$ 를 아래와 같이 수학적으로 계산할 수 있다.

$$\mathbf{r} = (1 - \alpha)(\mathbf{I} - \alpha \mathbf{A}^T \mathbf{D}^{-1})^{-1} \mathbf{1} \quad (8)$$

식 (8)에서  $\mathbf{I} \in \mathbb{R}^{N \times N}$ 는 단위행렬이며,  $\mathbf{1} \in \mathbb{R}^N$ 는 모든 원소가 1인 열벡터이고,  $\mathbf{D} \in \mathbb{R}^{N \times N}$ 는  $\max(K_{out}^i, 1)$ 를 원소로 하는 대각행렬이다. 이때  $K^{i_{out}}$ 는 노드  $i$ 에서 뺀어나가는 유향 선분의 개수를 의미한다.  $\alpha$ 는 0에서 1 사이의 값을 가지는 감쇠치(damping factor)로, 본 연구에서는 PageRank를 고안한 연구에서 사용된 통상적인 값인 0.85로 정했다(Brin et al., 1999). 위의 식을 이용하면, 거듭제곱 반복법(power method)을 통해 수치적으로 PageRank의 값을 도출해낼 수 있게 된다. PageRank의 값이 클수록 그 노드의 중심성이 강하다는 뜻이기 때문에 본 연구에서 구성할 금융 시장 네트워크에서 해당 금융 지수가 갖는 영향력이 강하다는 의미로 해석할 수 있다.

### 3.2.4 군집 분석

네트워크의 군집을 생성하기 위하여 다음과 같은 Dugué and Perez(2015)의 유향 네트워크에 대한 Louvain 알고리즘에 관한 연구를 활용하였다. 해당 연구에서는 유향 네트워크의 군집화를 위하여 다음과 같은 네트워크의 모듈성(modularity)을 최대화하는 방식으로 활용하였다. 이 때 모듈성(modularity)은 네트워크 내에서 상대적으로 밀접한 관계를 가지고 있는 하위 집단을 찾기 위해 주로 사용되는 척도로, 모듈성이 크다는 것은 집단 내 정점(금융 시장 지수) 사이의 관계가 집단 간 금융들 사이의 관계에 비해 강한 것을 의미한다. 다시 말해, 하위 그룹이 명확하게 분류될수록 모듈성의 값이 큼을 의미한다.

$$Q = \frac{1}{w} \sum_{i,j} (A_{ij} - \gamma \frac{d_i^+ d_j^-}{w}) \delta_{c_i, c_j} \quad (9)$$

$\mathbf{A}$ 는 인접행렬,  $c_i$ 는 노드  $i$ 의 군집,  $d_i^+$ 와  $d_i^-$ 는 외향 및 내향 가중치를 의미하며  $w = \mathbf{1}^T \mathbf{A} \mathbf{1}$ 이며  $\delta$ 는 Kronecker의 델타이다. 마지막으로,  $\gamma$ 는 해상도 모수(resolution parameter)이다.

## 4. 실험 결과

### 4.1 네트워크 구성 결과 및 해석

본 연구에서는 ETE로 상정된  $Z$ 의 값을 통해 산출된 p-값에

Table 3. Network-level Network Measures of Global Financial Market Index Network

Attribute	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Total Connected Nodes	24	24	24
Total Connections	314	257	140
Network Density	56.88%	46.56%	25.36%
Maximum Degree Centrality	40	37	25
Minimum Degree Centrality	16	10	2
Average Degree Centrality	26.17	21.42	11.67



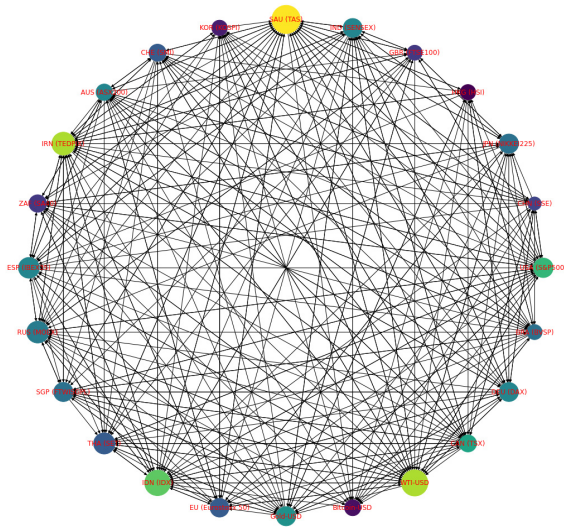


Figure 1. Global Financial Market Index Network( $\alpha = 0.1$ )

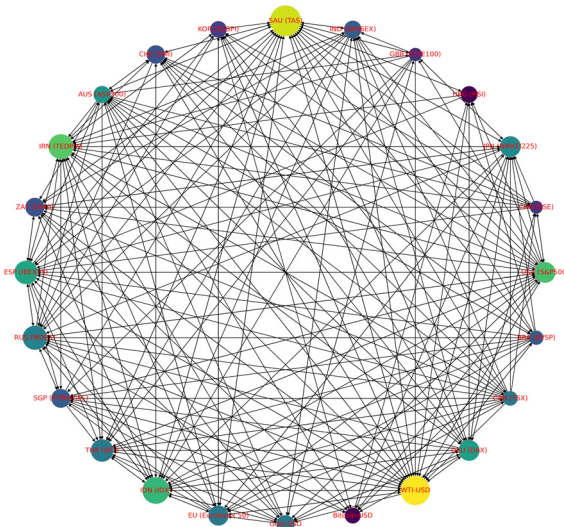


Figure 2. Global Financial Market Index Network( $\alpha = 0.05$ )

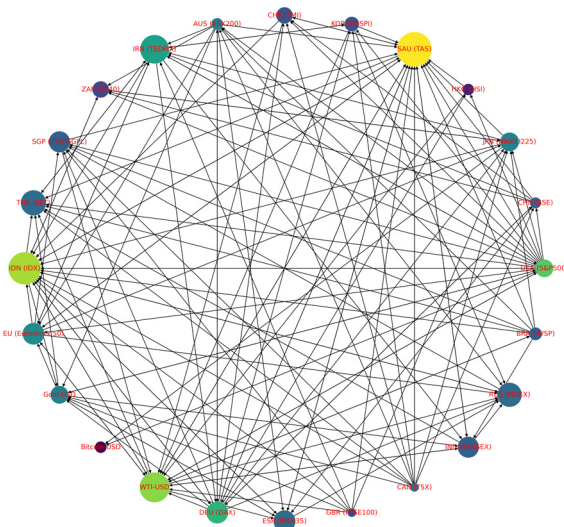


Figure 3. Global Financial Market Index Network( $\alpha = 0.01$ )

대해 유의 수준  $\alpha = 0.1, 0.05, 0.01$ 에 대해서 위와 같은 세 가지 네트워크를 구성하였다. 네트워크 전체를 기준으로 한 주요 지표는 다음과 같다. 이 때, 24개의 노드에 대해서 가능한 총 연결의 개수는  $24 \times 23$ 개, 즉 552개이다. 또한, 후술할 예측 실험을 위해 데이터 유출(data leakage)을 방지하고자 위의 그래프는 2016년 1월 1일부터 2020년 12월 31일까지의 5개년 치의 데이터에 대한 국제 금융 시장 지수 네트워크이다.

총 가능한 연결 수의 50%를 기준으로 할 때,  $\alpha = 0.1$ 일 때의 네트워크는 뻘뻘한(dense) 네트워크,  $\alpha = 0.05$  및  $\alpha = 0.01$ 일 때의 네트워크는 성긴(sparse) 네트워크로 분류할 수 있다. 그리고 모든 유의 수준에 대한 네트워크에서 누락된 노드, 즉 국제 금융 시장 지수 중 인과관계적으로 고립된 시장 지수가 없음을 확인할 수 있다.

이 중, 본 연구에서는 연구 결과의 강건성을 위하여  $\alpha = 0.01$ 을 기준으로 한 네트워크를 활용하여 후술할 실험들을 진행하였다.  $\alpha = 0.01$  수준에서 구성된 국제 금융 시장 네트워크의 주요 네트워크 지표에 대해서 노드 수준의 분석 결과는 다음과 같다.

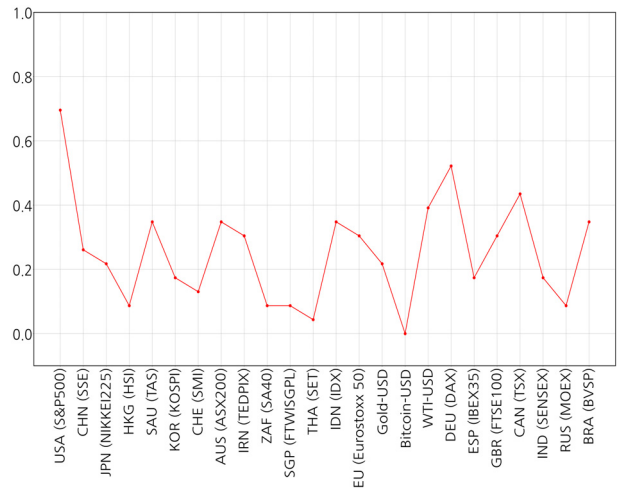


Figure 4. In-Degree of Global Financial Market Index Network

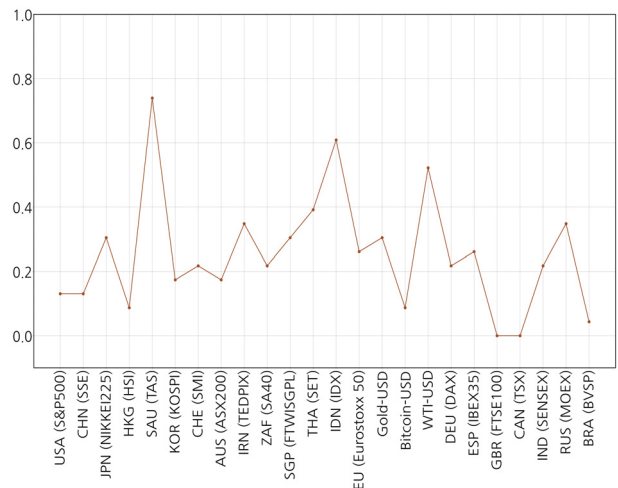


Figure 5. Out-degree of Global Financial Market Index Network

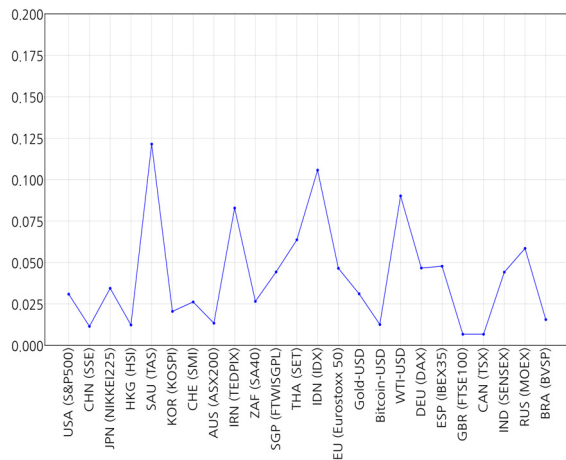


Figure 6. PageRank of Global Financial Market Index Network

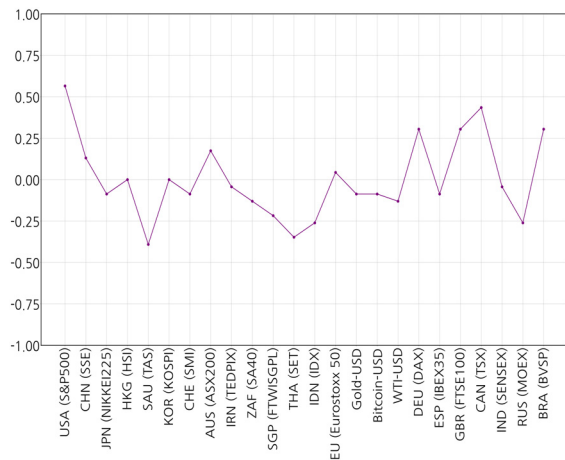


Figure 7. Net Degree(Out-Degree - In-Degree) of Global Financial Market Index Network

위의 4개 노드 수준의 네트워크 지표 결과에서는 다음과 같은 점을 주목할 수 있다. 우선 전체 지표에서 미국의 S&P 500 지수가 외향 연결 중심성에 대해서 높은 순위에 위치하고 있는 것을 통하여 미국의 S&P 500 지수는 실제 금융 시장에서 알려진 것과 유사하게 본 연구의 국제 금융 시장 지수 네트워크에서 정보 전달자로서 핵심적인 하고 있음을 알 수 있다. 그리고 <Figure 7>에서 외향 연결 중심성으로부터 내향 연결 중심성의 값을 뺀 값을 확인했을 때, 미국의 S&P 500의 지수의 값이 양의 값으로 가장 큰 값을 가짐에 따라 타 시장으로의 영향력이 가장 크다는 것을 확인할 수 있다. 이러한 결론은 기존의 연구 결과와 유사하다(Graham et al., 2013; Oatley et al., 2013; Gokmenoglu and Fazlollahi, 2015; Alareeni and Hamdan, 2020). 반면 사우디아라비아는 가장 높은 내향 연결 중심성을 가지며 또한 순 연결 중심성(net degree)에서 가장 낮은 값을 가지는 것을 확인할 수 있다. 이는 다른 지수의 움직임에 영향을 주는 것에 비해 받는 영향이 매우 큰 것을 의미한다. 이와 더불어 동남아시아와 중동 국가 금융 시장 지수의 경우 주로 내향 연결 중심성이 높으며 이러한 내향 중심적 영향력에 따라 PageRank에서도 상위권을 차지하고 있다. 이는 동남아시아와 중동 국가의 금융 시장 지수는 다른 국가의 금융 시장 지수의 움직임에 의해 영향을 많이 받음을 의미하며 동남아시아와 중동 국가의 금융 시장의 규모가 다른 상위권 금융 시장에 비해 규모가 작아 주요 금융 시장의 움직임에 영향을 받는다는 기존의 통념과 흐름을 같이 한다고 볼 수 있다. 구체적으로 PageRank의 값을 기준으로 1위(사우디아라비아, 0.121), 2위(인도네시아, 0.106), 4위(이란, 0.082), 5위(태국, 0.064)의 순이다(3위는 WTI/USD, 0.090). 이외에도 영국의 FTSE 100 지수와 캐나다의 S&P/TSX 지수는 외향 연결 중심성에서 0의 값을 보임에 따라 외부로부터 다른 금융 시장 지수의 움직임으로부터 정보를 전달만 받는 수신자의 역할을 하고 있음을 확인할 수 있다. 단, 이는 선형적인 분석이 아니기 때문에 통념적으로 알려진 선형 금융 시장 내 영향력과 다른 결과가 도출될 수 있다.

#### 4.2 네트워크 군집의 구성

3.2.4절에서 언급한 것과 같이 Dugué(2015)의 유향 네트워크에 대한 Louvain 알고리즘을 활용하여 다음과 같이 3개로 나뉘어진 네트워크 군집 결과를 확인할 수 있다. 구체적인 군집 분석 결과는 <Table 4>와 같다.

#### 4.3 네트워크 군집을 이용한 KOSPI의 등락 예측 실험

##### 4.3.1 실험 설명

본 연구에서는 기계 학습 모형을 활용하여 4.2절에서 생성된 군집을 이용하여 KOSPI 등락 예측을 실시하였다. 실험 실시를 위해 다음과 같은 실험들을 구상하였다. 이 때 KOSPI 데이터는 왜곡할 수 있어 벤치마크 실험을 제외하고 예측에 직접적으로 포함시키지 않았다.

Benchmark: KOSPI만을 이용한 예측

Experiment 1: KOSPI를 제외한 실험 내 전체 국제 시장 지수 데이터열을 이용한 예측

Experiment 2: 군집 1에 속한 국제 시장 지수 데이터열을 이용한 예측

Experiment 3: 군집 2에 속한 국제 시장 지수 데이터열을 이용한 예측

Experiment 4: 군집 3에 속한 국제 시장 지수 데이터열을 이용한 예측

Experiment 5: 군집 2와 3에 속한 국제 시장 지수 데이터열을 이용한 예측

모형의 성능을 측정하는 실험 기간은 2.1절에서 상기한 것과 같이 2021년 1월 1일부터 2021년 11월 30일까지이며, 2016년 1월 1일부터 2020년 12월 31일까지의 데이터를 학습 데이터로 활용하였다.



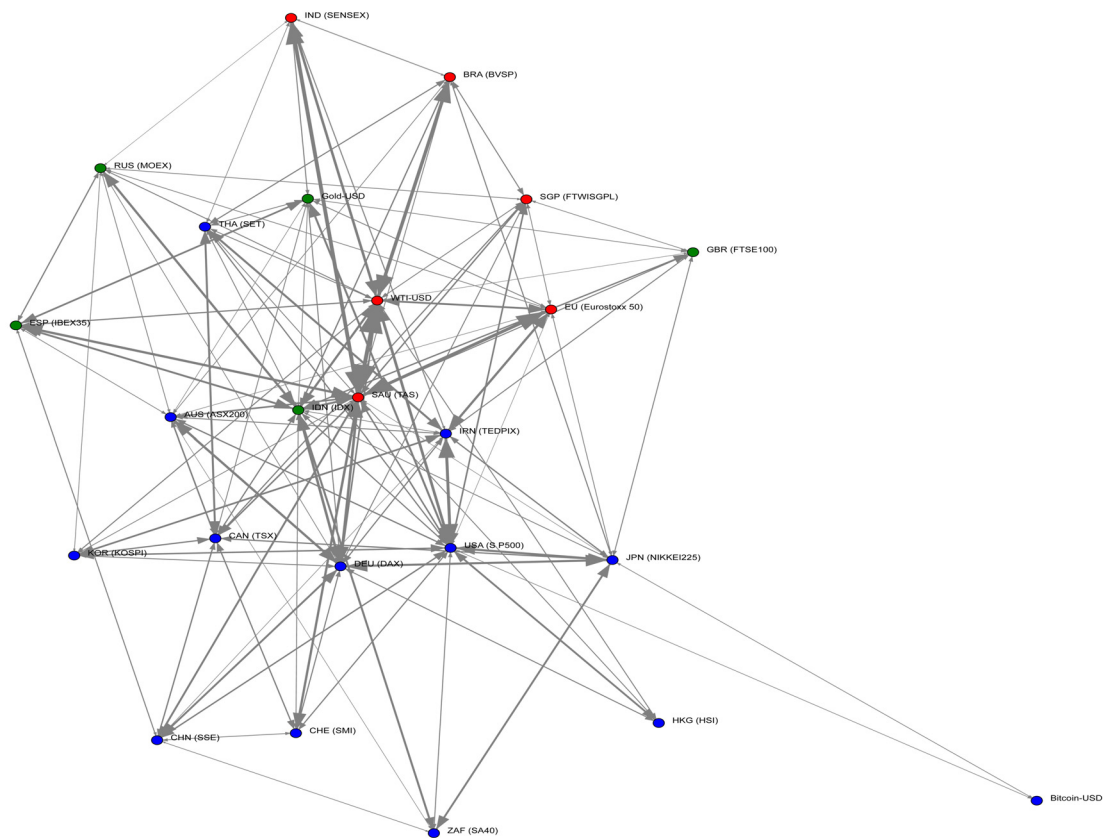


Figure 8. Global Financial Market Index Network Including Clustering Results

Table 4. Clustering results

Country	Financial Market Index	Abbreviation	# of Cluster
United States	Standard and Poor's 500	USA(S&P500)	1
China	Shanghai Stock Exchange Composite Index	CHN(SSE)	1
Japan	Nikkei 225	JPN(NIKKEI225)	1
Hong Kong	Hang Seng Index	HKG(HSI)	1
United Kingdom	Financial Times Stock Exchange 100	GBR(FTSE100)	2
Canada	Standard and Poor's Toronto Stock Exchange Index	CAN(TSX)	1
India	Bombay Stock Exchange Sensex Index	IND(SENSEX)	3
Saudi Arabia	Tadawul All Share	SAU(TAS)	3
Germany	Deutscher Aktienindex	DEU(DAX)	1
South Korea	Korea Composite Stock Price Index	KOR(KOSPI)	1
Switzerland	Swiss Market Index	CHE(SMI)	1
Australia	Standard and Poor's Australian Securities Exchange 200	AUS(ASX200)	1
Iran	Tedpix	IRN(TEDPIX)	1
South Africa	South Africa Top 40	ZAF(SA40)	1
Brazil	Bovespa	BRA(BVSP)	3
Spain	Iberian Index 35	ESP(IBEX35)	2
Russian Federation	Moscow Exchange Index	RUS(MOEX)	2
Singapore	Financial Times Stock Exchange Singapore	SGP(FTWISGPL)	3
Thailand	Stock Exchange of Thailand Index	THA(SET)	1
Indonesia	Indonesia Stock Exchange Index	IDN(IDX)	2
EU	EuroStoxx 50	EuroStoxx 50	3
-	Gold/USD	Gold/USD	2
-	Bitcoin/USD	Bitcoin/USD	1
-	WTI/USD	WTI/USD	3

학습을 위해 포함되는 데이터의 형태는 국제 시장 지수의 일일 수익률의 형태로 포함이 되었으며, 등락 분류를 위하여 타깃은 KOSPI 지수의 일일 수익률이 양수거나 0이면 1, 음수면 0으로 변환하여 이진 분류 문제(binary classification problem)로 정의하였다. 그리고 실제 테스트 단계에서는 2021년 1월 1일과 2021년 11월 30일 사이의 영업일 사이의 임의의 시점  $t$ 에 대해서  $t-1$  시점의 금융 시장 지수 데이터만을 이용하여 예측을 실시하였다.

이러한 KOSPI의 등락을 예측하기 위한 기계학습 모형에는 그래디언트 부스팅 알고리즘 기반 모형인 XGBoost(Chen and Guestrin, 2016), Light gradient boosting machine(Ke *et al.*, 2017, 이하 LightGBM), 그리고 CatBoost 알고리즘(Prokhorenkova *et al.*, 2018)을 사용하였다. 이외에도 로지스틱 회귀 모형(logistic regression), K-평균(K-means) 분류기, 다중 퍼셉트론(multi-layer perceptron) 분류기, 가우시안 과정(Gaussian process) 분류기, 서포트 벡터 머신(support vector machine), 트리(decision tree), 그리고 랜덤 포레스트(random forest)와 같이 총 10개의 분류 모형들을 사용하였으나 예측 성능이 좋지 않아 상기한 세 모형을 제외한 7개의 모형에서 예측 성능이 일괄 예측이 되거나(모두 상승할 것으로 예측하는 경우 또는 모두 하락할 것으로 예측하는 경우), 예측 성능이 50%보다 낮은 경우 등이 발생하여 해당 모형들을 제외하고 유의미하다고 판단된 XGBoost, LightGBM, 그리고 CatBoost 세 가지 모형을 통해서 결과를 서술하였다. XGBoost, LightGBM, 그리고 CatBoost 모형은 본 연구에서 활용되는 정형 데이터에 대해서 매우 좋은 분류 성능을 보이는 트리 기반의 모형들로 알려져 있으며 이러한 점에 착안하여 본 연구에서는 해당 모형들을 사용하였다. 또한, 초모수(hyperparameter) 최적화를 위하여 Python의 Optuna(Akiba *et al.*, 2019) 패키지를 활용하였다. 그리고 국제 시장 지수의 예측 영향도를 확인하고 시각화하기 위하여 SHAP(Shapley additive explanations, Lundberg *et al.*, 2017; 2017) 방법 중 실험에서 활용하는 기계 학습 모형이 모두 트리를 기반으로 하므로 TreeSHAP를 활용하였다. 기계 학습 모형과 SHAP에 대한 설명은 다음과 같다.

### 4.3.2 XGBoost

Chen and Guestrin(2016)이 개발한 XGBoost는 Friedman(2001)이 제안한 부스팅 그래디언트 기법을 사용한 알고리즘이다. 한 분류기를 순차적으로 개선해나감으로써 보강력한 분류기를 생성하는 트리 그래디언트 부스팅 기법을 적용한 앙상블 알고리즘의 한 종류이며, XGBoost의 최장점은 모든 시나리오에서 확장 가능(scalable)하다는 점이며 데이터 예측 및 분류 프로젝트에서 가장 활발하게 활용되는 알고리즘 중 하나이다.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (10)$$

$$\begin{aligned} Z &= \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (11) \\ &= \sum_i l(\hat{y}_i, y_i) + \sum_i (\gamma T + \frac{1}{2} \lambda \|w\|^2) \end{aligned}$$

식 (10)은 트리의 앙상블 모형을 나타내는 식이며,  $F$ 는 모든 가능한 CART(classification and regression trees)의 집합 공간이며,  $K$ 는 트리의 개수이다.  $f_k$ 는 독립된 각각의 트리와 파생되는 각 잎(leaf)의 가중치로 대응할 수 있다. 이 때 각각의 잎의 점수를 합산 및 비교하여 최종 예측을 하게 된다. 식 (11)은 XGBoost 모형의 정규화된 목적함수이다.  $l(\hat{y}_i, y_i)$ 는 예측값과 목표값의 차이를 측정하는 미분 가능한 볼록(convex) 손실 함수이며,  $\Omega(f_k)$ 는 정규화 항이며 모델의 복잡도를 조정하여 최종 학습된 가중치를 매끄럽게(smooth) 만듦으로써 과적합 문제의 발생을 막는 CART 함수들이다.  $\gamma T$ 는 CART의 잎의 개수,  $\frac{1}{2} \lambda \|w\|^2$ 는 CART의 잎에 배정된 점수를 의미한다.

### 4.3.3 LightGBM

LightGBM은 Ke *et al.*(2017)이 개발한 그래디언트 부스팅 기반의 기계 학습 모형으로 그래디언트 기반 원-사이드 샘플링(gradient-based one side sampling, GOSS)과 변수 개수를 줄이기 위해 상호배타적인 변수들을 묶는 배타적 특성 결합(exclusive feature bundling, EFB)의 두 가지 접근법을 사용한 모형이다. 트리가 수직적으로 확장되는 특징을 가진다. 이러한 특징 덕분에 LightGBM은 타 기계 학습 모형들보다 일반적으로 효율적이고 정확한 성능을 제공하는 것으로 알려져 있다. Sun *et al.*(2019)에서는 LightGBM의 장점은 빠른 훈련 속도, 낮은 메모리 소비량, 우수한 모델 정확도로 귀결되며 해당 논문에서는  $T$ 개의 트리에 대해서 추정된 LightGBM의 함수는 다음과 같이 정의된다고 언급한다.

$$Y_t = \sum_{h=1}^T f_h(x) \quad (12)$$

이 때  $f_h(x)$ 는 트리이며 뉴턴 방법(Newton's method)을 이용하여 목적 함수를 추정하게 된다.

### 4.3.4 CatBoost

CatBoost는 Prokhorenkova *et al.*(2018)이 제안한 새로운 경사 강하 기반 알고리즘이다. 이 지도 기계 학습 알고리즘은 의사결정 트리의 그래디언트 부스팅 기법을 사용하여 범주형 데이터를 분류하는 데에 특화되어 있다. 트리는 학습 데이터 세트에 대해서 유사한 인스턴스로 분할함으로써 생성된다. Prokhorenkova *et al.*(2018)은 CatBoost는 범주형 기능을 처리하기 위해 순서형 부스팅을 위시로 한 알고리즘을 채용하였으며, CatBoost가 성능 면에서 타 부스팅 그래디언트 기법보다 나은 측면이 있음을 언급하고 있다. 다음과 같이 정의된 식 (13)의 트리  $h$ 에 대해서 새로운 트리에 대한 CatBoost의 그래디언트 부스팅 과정은 다음과 같이 쓸 수 있다.

$$h(\mathbf{x}) = \sum_{j=1}^J b_j \mathbf{1}_{\{x \in R_j\}} \quad (13)$$

$$h' = \arg \min \frac{1}{N} \sum (-f'(\mathbf{X}_k - y_k) - h(\mathbf{X}_k))^2 \quad (14)$$

$t = \frac{n^+ - y_k + ap}{n - 1 + a}$  인 분기 기준을 의미하며,  $a$ 는 분할 속성으로 임의의 양수,  $p$ 는 데이터 세트의 목표 변수의 평균값을 의미한다.  $\mathbf{X}_k$ 는  $N$ 개의 투입 변수에 대한 랜덤 벡터(random vector)이며  $Y_k$ 는 목표 변수, 그리고 함수  $f$ 는 뉴턴 방법에 의한 최소 제곱 근사 함수이다. 또한 CatBoost는 효율성을 향상시키고 실행 속도를 향상시키며 과적합 문제를 해결하기 위해 망각 결정 방법(oblivious tree method)을 사용한다. 이 방법은 트리를 분할할 때 동일한 분할 기준이 전체 트리 단계에서 적용되는 것으로, 이는 균형적인 트리를 생성하며 과적합을 막을 수 있다는 장점이 있다.

4.3.5 TreeSHAP

기계 학습은 시계열 데이터를 예측하는 데 큰 잠재력을 가지고 있다. 그러나 연구자들은 보통 그들의 예측을 설명하지 않는 데, 이것은 기계 학습의 채택에 대한 장벽이다. 이 문제를 극복하기 위해 Lundberg and Lee(2017)는 XGBoost, LightGBM, CatBoost 및 Scikit-learn 패키지의 트리 기반 모델을 포함한 다양한 기술에 대한 예측을 해석하기 위한 SHAP 접근방식을 제안했다. SHAP는 사용자가 복잡한 모델의 예측을 해석하는 데 도움이 되는데, 이러한 개념의 근간이 되는 새플리 가치(Shapley value)는 미국의 경제학자 로이드 새플리의 해에 의해 처음 제안되었으며 게임 이론의 개념에 기초하고 있다(Shapley, 1953). 예측에 대한 각 기능의 영향을 계산하여 특정 입력  $X$ 의 예측을 설명할 수 있다는 것으로 설명 가능한 인공지능(explainable artificial intelligence, XAI) 기법의 한 가지 방법으로 각광받고 있다. 게임 이론을 기반으로 각 데이터 열별 조건부 기댓값을 결합하여 추정된 Shapley 값의 평균 다음과 같이 계산된다.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \quad (15)$$

$\phi_i$ 는 데이터  $i$ 에 대한 새플리 값이며,  $N$ 은 총 투입 변수의 집합이다.  $S$ 는 총 투입 변수에서  $i$ 번째 변수를 제외한 모든 집합이며  $v(S)$ 는  $i$ 번째 데이터를 제외한 나머지 부분 집합이 결과에 공헌한 기여도이며  $f_x(S \cup \{i\})$ 는  $i$ 번째 데이터를 포함한 전체 기여도이다.

Lundberg *et al.*(2018)는 XGBoost를 포함한 그래디언트 부스팅 기반 모형에서 SHAP 계산에 특화된 TreeSHAP 기법을 제안했다. 해당 기법은 의사결정 트리, 랜덤 포레스트 및 경사 부스팅 트리와 같은 트리 기반 기계 학습 모델에서 TreeSHAP은 기존 SHAP보다 신속한 계산을 위한 대안으로 도입되었다.

본 연구에서 선정한 세 가지 기계 학습 모형이 모두 트리 기반 모형임에 따라 TreeSHAP 방법을 SHAP 값을 계산하였으며 다음과 같은 SHAP 값의 절댓값의 평균값(mean absolute SHAP values)을 이용하여 데이터열 별 특성 중요도를 계산하였다.

$$I_j = \frac{1}{N} \sum_{i=1}^N |\phi_j^{(i)}| \quad (16)$$

4.3.6 실험 결과

Optuna를 이용한 초모수 최적화 결과, 각 모형별로 다음과 같은 초모수가 사용되었다. 이 때 학습 성과를 평가하기 위한 지표로는 F1 점수(F1 score)를 사용하였다.

실험 결과를 주요 성과 지표를 통해 요약하면 다음과 같다. 실험은 모형별로 무작위 난수 발생 시드를 변경하여 Optuna로

Table 5. Experiment Results(average accuracy of 100 time simulations)

ML Algorithm	Benchmark	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
XGBoost	0.4961	0.5659	<b>0.5736</b>	0.4729	0.4651	0.5426
LightGBM	0.4806	<b>0.5814</b>	0.5659	0.4496	0.5194	0.5271
CatBoost	0.4651	0.5504	<b>0.5659</b>	0.4806	0.5194	0.5271
HardVotingClassifier (with equal-weights)	0.4806	<b>0.5736</b>	0.5659	0.4651	0.5039	0.5504

Table 6. Experiment Results(average F1 score of 100 time simulations)

ML Algorithm	Benchmark	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
XGBoost	0.5752	<b>0.6164</b>	0.6043	0.5584	0.5175	0.6093
LightGBM	0.5442	<b>0.6250</b>	0.6056	0.5298	0.5694	0.5793
CatBoost	0.6102	<b>0.6234</b>	0.6216	0.5839	0.6026	0.6013
SoftVotingClassifier (with equal-weights)	0.5732	<b>0.6309</b>	0.6164	0.5605	0.5616	0.6133

20회 학습을 진행한 후 해당 초모수들을 사용하여 각 100회를 진행하였으며 각 성과 평가 지표의 평균값 및 표준편차를 표기하였다. 또한 선정된 세 모형을 동일한 비율로 사용한 간접 투표 분류기(soft voting classifier) 결과도 실험을 평균적인 성능을 확인하고자 포함하였다.

그 결과, 전체 데이터열을 사용한 Experiment 1과 원래 KOSPI가 포함된 군집 1 내의 데이터열을 이용한 Experiment 2가 가장 좋은 결과를 보이는 것을 확인하였다. 단, F1 점수를 기준으로는 전체 데이터열을 사용한 Experiment 1이 가장 좋은 성능을 보이는 것을 확인할 수 있다. 또한 KOSPI를 단독으로 사용한 벤치마크의 경우 일종의 동전 던지기(coin tossing) 문제에 가깝거나 이보다 더 낮은 예측율을 보임에 따라 기존의 연구와 같이 약형 효율적 시장가설(weak form of the efficient market hypothesis)과 거의 유사한 결과를 얻어낼 수 있음을 확인하였다. 즉, 무작위 행보(random walk)에 가까운 움직임을 보였다고 해석할 수 있다. 군집 2와 군집 3을 이용한

Experiment 3, 4의 결과가 가장 좋지 않은 것을 확인했으며 군집 1을 제외한 두 군집을 합쳐 실험한 Experiment 5의 경우에도 Experiment 1 및 Experiment 2보다 좋은 정확도를 얻지 못했다. 정확도와 F1 점수를 기준으로 상대적으로 좋은 결과가 도출된 Experiment 1과 Experiment 2에서의 각 모형별 TreeSHAP 기반의 예측 중요도는 다음과 같다.

결론적으로, 정보 흐름 기반의 네트워크의 군집 내의 데이터열을 사용하여 KOSPI의 등락을 예측할 때 더욱 적은 데이터열로 모든 데이터열을 사용했을 때와 거의 준하는 예측 결과를 얻을 수 있음을 확인했다. 이는 KOSPI의 등락에 대해서 더욱 적은 데이터열로 동등하거나 또는 그 이상의 예측 정확도를 얻을 수 있음을 의미한다. 또한 KOSPI가 속한 군집 외의 국가들로 KOSPI를 예측하는 어떠한 실험 결과(Experiment 3, Experiment 4, Experiment 5)보다 KOSPI가 속하는 군집으로 예측하는 것이 정확도와 F1 점수를 기준으로 모든 기계학습 모형에서 좋은 예측 결과가 나오는 것을 확인할 수 있었다.

Table 7. Average Feature Importance of Experiment 1

XGBoost			LightGBM			CatBoost		
Index	Average Mean( SHAP ) Value	Rank	Index	Average Mean( SHAP ) Value	Rank	Index	Average Mean( SHAP ) Value	Rank
USA(S&P500)	0.4253	1	Gold/USD	0.3911	1	Gold/USD	0.1393	1
Gold/USD	0.4237	2	USA(S&P500)	0.3464	2	DEU(DAX)	0.1299	2
SAU(TAS)	0.3942	3	DEU(DAX)	0.3048	3	USA(S&P500)	0.1109	3
Bitcoin/USD	0.3915	4	RUS(MOEX)	0.2734	4	IND(SENSEX)	0.0966	4
DEU(DAX)	0.3722	5	SAU(TAS)	0.2624	5	BRA(BVSP)	0.0795	5
RUS(MOEX)	0.3246	6	IDN(IDX)	0.2600	6	CAN(TSX)	0.0721	6
IND(SENSEX)	0.3196	7	Bitcoin/USD	0.2558	7	AUS(ASX200)	0.0713	7
JPN(NIKKEI225)	0.2992	8	IND(SENSEX)	0.2424	8	JPN(NIKKEI225)	0.0690	8
CHN(SSE)	0.2982	9	IRN(TEDPIX)	0.1910	9	Bitcoin/USD	0.0617	9
IDN(IDX)	0.2649	10	JPN(NIKKEI225)	0.1902	10	IRN(TEDPIX)	0.0569	10
SGP(FTWISGPL)	0.2560	11	AUS(ASX200)	0.1848	11	HKG(HSI)	0.0541	11
CAN(TSX)	0.2393	12	CHN(SSE)	0.1798	12	SAU(TAS)	0.0524	12
AUS(ASX200)	0.2310	13	SGP(FTWISGPL)	0.1627	13	IDN(IDX)	0.0454	13
CHE(SMI)	0.2251	14	THA(SET)	0.1607	14	CHE(SMI)	0.0435	14
HKG(HSI)	0.2227	15	HKG(HSI)	0.1474	15	WTI/USD	0.0367	15
THA(SET)	0.2186	16	WTI/USD	0.1465	16	RUS(MOEX)	0.0323	16
IRN(TEDPIX)	0.2143	17	CHE(SMI)	0.1448	17	CHN(SSE)	0.0309	17
WTI/USD	0.1934	18	CAN(TSX)	0.1399	18	SGP(FTWISGPL)	0.0253	18
GBR(FTSE100)	0.1667	19	GBR(FTSE100)	0.1377	19	EuroStoxx 50	0.0211	19
BRA(BVSP)	0.1612	20	BRA(BVSP)	0.1254	20	GBR(FTSE100)	0.0200	20
ZAF(SA40)	0.1339	21	EuroStoxx 50	0.0941	21	ESP(IBEX35)	0.0173	21
ESP(IBEX35)	0.1207	22	ZAF(SA40)	0.0864	22	THA(SET)	0.0171	22
EuroStoxx 50	0.1185	23	ESP(IBEX35)	0.0838	23	ZAF(SA40)	0.0156	23

Table 8. Average Feature Importance of Experiment 2

XGBoost			LightGBM			CatBoost		
Index	Average Mean( SHAP ) Value	Rank	Index	Average Mean( SHAP ) Value	Rank	Index	Average Mean( SHAP ) Value	Rank
Bitcoin/USD	0.5628	1	Bitcoin/USD	0.4115	1	DEU(DAX)	0.1839	1
CAN(TSX)	0.5066	2	DEU(DAX)	0.3825	2	CAN(TSX)	0.1746	2
DEU(DAX)	0.4354	3	USA(S&P500)	0.3343	3	USA(S&P500)	0.1637	3
CHN(SSE)	0.4081	4	CAN(TSX)	0.3277	4	Bitcoin/USD	0.1517	4
USA(S&P500)	0.3950	5	CHN(SSE)	0.3004	5	JPN(NIKKEI225)	0.1325	5
HKG(HSI)	0.3828	6	HKG(HSI)	0.2842	6	CHN(SSE)	0.1265	6
THA(SET)	0.3644	7	IRN(TEDPIX)	0.2754	7	AUS(ASX200)	0.1254	7
AUS(ASX200)	0.3563	8	AUS(ASX200)	0.2689	8	IRN(TEDPIX)	0.1244	8
JPN(NIKKEI225)	0.3543	9	THA(SET)	0.2559	9	HKG(HSI)	0.1134	9
IRN(TEDPIX)	0.3374	10	JPN(NIKKEI225)	0.2449	10	CHE(SMI)	0.0897	10
CHE(SMI)	0.2602	11	ZAF(SA40)	0.1546	11	THA(SET)	0.0868	11
ZAF(SA40)	0.2419	12	CHE(SMI)	0.1535	12	ZAF(SA40)	0.0613	12

## 5. 결론

주식 시장은 경제학 및 금융의 중요한 연구 분야 중 하나로 광범위하게 연구됐다. 특히 주가와 수익률 데이터를 기반으로 증시를 분석하고 예측하는 연구는 금융 분야 내에서 가장 활발히 다루어지고 있는 주제이며 구체적으로 주가 및 시장 전반에 대한 예측은 투자자들이 최적의 투자전략을 수립하기 위한 중요한 요소 중 하나이다.

본 연구에서는 전이 엔트로피 개념을 활용하여 주요 금융 시장 지수의 정보 흐름을 바탕으로 한 인과관계 네트워크를 구성하였다. 그리고 구성된 네트워크를 활용하여 금융 시장을 분석하였으며, 또한 정보 흐름 네트워크를 금융 시장 지수를 예측에 KOSPI의 등락에 활용할 수 있는지 확인하고자 하였다.

이를 위하여 정보 불확실성의 감소량을 측정하는 전이 엔트로피를 인과관계의 측정 지표로 상정하였다. 그리고 전이 엔트로피 측정 시 발생할 수 있는 유한크기 효과(finite size effect)를 조정하는 데 있어서 효과적인 지표인 효율적 전이 엔트로피를 활용하여 정보 흐름 네트워크를 구성하였으며 이를 이용하여 금융 시장 지수 간의 인과관계를 분석하고 KOSPI의 등락 예측에 활용하였다. 그 결과, 금융 시장 지수를 효율적 전이 엔트로피를 이용한 인과관계 네트워크를 활용하여 금융 시장의 복잡계 네트워크 분석이 가능함을 확인하였고, 구성된 네트워크의 군집을 활용하여 국내 금융 시장 등락 예측에 있어 약 50%에 해당하는 적은 데이터 열을 활용하여 전체 데이터를 활용하는 것에 준하는 예측 결과를 확인할 수 있었다.

본 연구의 한계점은 다음과 같다. 우선 성능의 문제로 인하여 그래디언트 부스팅 알고리즘 기반의 기계 학습 방법론을 활용하여 예측 및 성과 측정을 진행하였으나 최신의 다양한 기계 학

습 방법론을 추가적으로 활용하여 이것이 그래디언트 부스팅 알고리즘과 같은 특정 형태의 기계 학습 기반 예측 방법론에 국한되지 않는다는 것을 보임으로써 결과의 강건성을 높일 필요가 있다. 또한, 본 연구는 국제 금융 시장 간의 인과관계를 중심으로 네트워크를 구성하였으나 금융 시장 변수 간의 관계에는 경제적, 사회적 관점에서 더욱더 많은 등락 결정 요인이 포함될 수 있으나 이러한 변수를 포함하여 연구하지 못했다는 제한점이 있다. 또한 기본적인 예측 모형의 성과를 생각했을 때 50% 미만의 60%를 갖 상회하는 F1 점수는 예측의 성과적 측면에서 불만족스러운 것으로 해석될 수 있다. 하지만 금융 시장이 점점 동적 복잡계 네트워크 시장이 됨에 따라 4년의 데이터로 구성된 네트워크가 중간에 변형이 일어남에 따라 이러한 네트워크의 연결성의 변형이 더 짧은 시간대에서 있음을 예측할 수 있다. 특히 2019년 COVID-19 시기 이후에 금융 시장의 네트워크가 변화했을 가능성이 여러 연구에서 제기되고 있다(Zhang *et al.*, 2020; Sansa *et al.*, 2020; So *et al.*, 2021). 따라서, 이러한 부분에 대해서 기존의 연구와 같이 이동창(moving window) 기법을 사용하면 변화적 동태를 고려할 경우 더 나은 결과를 얻을 수 있는 가능성이 있다(Turchencko *et al.*, 2011; Shynkevich *et al.*, 2017; Henrique *et al.*, 2018). 결론적으로 상기한 점을 보완한다면 더욱 결과의 강건성을 제고하고, 개선된 예측 결과가 도출될 수 있을 것으로 사료된다.

## 참고문헌

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019, July), Optuna: A next-generation hyperparameter optimization framework,

- In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623-2631.
- Alareeni, B. A. and Hamdan, A. (2020), ESG impact on performance of US S&P 500-listed firms, *Corporate Governance: The International Journal of Business in Society*, **20**(7), 1409-1428.
- Altomonte, C., Barba Navaretti, G., Di Mauro, F., and Ottaviano, G. (2011), Assessing competitiveness: How firm-level data can help (No. 2011/16), Bruegel Policy Contribution.
- Atsalakis, G. S. and Valavanis, K. P. (2009), Surveying stock market forecasting techniques-Part II: Soft computing methods, *Expert Systems with Applications*, **36**(3), 5932-5941.
- Baitinger, E., Kutsarov, I., Maier, T., Storr, M., and Wan, T. (2015), A holistic approach to diversification management: The diversification delta strategy applied to non-normal return distributions, *Credit and Capital Markets*, **48**(1), 89-119.
- Bekiros, S., Nguyen, D. K., Junior, L. S., and Uddin, G. S. (2017), Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets, *European Journal of Operational Research*, **256**(3), 945-961.
- Boba, P., Bollmann, D., Schoepe, D., Wester, N., Wiesel, J., and Hamacher, K. (2015), Efficient computation and statistical assessment of transfer entropy, *Frontiers in Physics*, **3**, 10.
- Bonanno, G., Caldarelli, G., Lillo, F., and Mantegna, R. N. (2003), Topology of correlation-based minimal spanning trees in real and model markets, *Physical Review E*, **68**(4), 046130.
- Bouchaud, J. P. (2001), Power laws in economics and finance: Some ideas from physics, *Quantitative Finance*, **1**(1), 105.
- Chen, T. and Guestrin, C. (2016, August), Xgboost: A scalable tree boosting system, In *Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining*, 785-794.
- Chen, H., Xiao, K., Sun, J., and Wu, S. (2017), A double-layer neural network framework for high-frequency forecasting, *ACM Transactions on Management Information Systems (TMIS)*, **7**(4), 1-17.
- Chi, K. T., Liu, J., and Lau, F. C. (2010), A network perspective of the stock market, *Journal of Empirical Finance*, **17**(4), 659-667.
- Cowan, A. R. and Sergeant, A. M. (2001), Interacting biases, non-normal return distributions and the performance of tests for long-horizon event studies, *Journal of Banking and Finance*, **25**(4), 741-765.
- Dimpfl, T. and Peter, F. J. (2013), Using Transfer Entropy to Measure Information Flows between Financial Markets, *Studies in Nonlinear Dynamics and Econometrics*, **17**(1), 85-102.
- Dugué, N. and Perez, A. (2015), Directed Louvain: Maximizing modularity in directed networks (Doctoral dissertation), Université d'Orléans.
- Faes, L., Nollo, G., and Porta, A. (2013), Compensated transfer entropy as a tool for reliably estimating information transfer in physiological time series, *Entropy*, **15**(1), 198-219.
- Fama, E. F. (1970), Efficient capital markets: A review of theory and empirical work, *The Journal of Finance*, **25**(2), 383-417.
- Fama, E. F. (1991), Efficient capital markets: II, *The Journal of Finance*, **46**(5), 1575-1617.
- Friedman, J. H. (2001), Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, **29**(5), 1189-1232.
- Gokmenoglu, K. K. and Fazlollahi, N. (2015), The interactions among gold, oil, and stock market: Evidence from S&P500, *Procedia Economics and Finance*, **25**, 478-488.
- Graham, M., Kiviahio, J., and Nikkinen, J. (2013), Short-term and long-term dependencies of the S&P 500 index and commodity prices, *Quantitative Finance*, **13**(4), 583-592.
- Granger, C. W. (1969), Investigating causal relations by econometric models and cross-spectral methods, *Econometrica: Journal of the Econometric Society*, **37**(3), 424-438.
- Hacine-Gharbi, A. and Ravier, P. (2018), A binning formula of bi-histogram for joint entropy estimation using mean square error minimization, *Pattern Recognition Letters*, **101**(1), 21-28.
- Henrique, B. M., Sobreiro, V. A., and Kimura, H. (2018), Stock price prediction using support vector regression on daily and up to the minute prices, *The Journal of Finance and Data Science*, **4**(3), 183-201.
- Henrique, B. M., Sobreiro, V. A., and Kimura, H. (2019), Literature review: Machine learning techniques applied to financial market prediction, *Expert Systems with Applications*, **124**, 226-251.
- Jabeur, S. B., Mefteh-Wali, S., and Viviani, J. L. (2021), Forecasting gold price with the XGBoost algorithm and SHAP interaction values, *Annals of Operations Research*, 1-21.
- Jang, S., Yi, E., Kim, W. C., and Ahn, K. (2019), Information flow between bitcoin and other investment assets, *Entropy*, **21**(11), 1116.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T. Y. (2017), Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, 30.
- Kim, K. J. (2003), Financial time series forecasting using support vector machines, *Neurocomputing*, **55**(1-2), 307-319.
- Kim, M., Newth, D., and Christen, P. (2016), Macro-level information transfer in social media: Reflections of crowd phenomena, *Neurocomputing*, **172**, 84-99.
- Kumar, D., Meghwani, S. S., and Thakur, M. (2016), Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets, *Journal of Computational Science*, **17**, 1-13.
- Kumar, S. and Deo, N. (2012), Correlation and network analysis of global financial indices, *Physical Review E*, **86**(2), 026101.
- Kwon, O. and Yang, J. S. (2008), Information Flow between Composite Stock Index and Individual Stocks, *Physica A: Statistical Mechanics and Its Applications*, **387**(12), 2851-2856.
- Kwon, Y. K., Choi, S. S., and Moon, B. R. (2005, June), Stock prediction based on financial correlation, In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, 2061-2066.
- Lausberg, C., Lee, S., Müller, M., Oertel, C., and Schultheiß, T. (2020), Risk measures for direct real estate investments with non-normal or unknown return distributions, *Zeitschrift für Immobilienökonomie*, **6**(1), 3-27.
- Lim, K., Kim, S., and Kim, S. Y. (2017), Information Transfer Across Intra/Inter-Structure of CDS and Stock Markets, *Physica A: Statistical Mechanics and Its Applications*, **486**, 118-126.
- Linden, M. (2001), A model for stock return distribution, *International Journal of Finance and Economics*, **6**(2), 159-169.
- Liu, J., Wu, C., and Li, Y. (2019), Improving financial distress prediction using financial network-based information and GA-based gradient boosting method, *Computational Economics*, **53**(2), 851-872.
- Lundberg, S. M. and Lee, S. I. (2017), A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, 30.
- Lundberg, S. M. and Lee, S. I. (2017), Consistent feature attribution for tree ensembles, arXiv preprint arXiv:1706.06060.
- Lundberg, S. M., Erion, G. G., and Lee, S. I. (2018), Consistent individualized feature attribution for tree ensembles, arXiv preprint arXiv:1802.03888.
- Malkiel, B. G. (2003), The efficient market hypothesis and its critics, *Journal of Economic Perspectives*, **17**(1), 59-82.
- Mandelbrot, B. and Taylor, H. M. (1967), On the distribution of stock price differences, *Operations Research*, **15**(6), 1057-1062.



- Mantegna, R. N. and Stanley, H. E. (1999), *Introduction to econophysics: Correlations and complexity in finance*, Cambridge university press.
- Marschinski, R. and Kantz, H. (2002), Analyzing the information flow between financial time series, *European Physical Journal B - Condensed Matter and Complex Systems*, **30**(2), 275-281.
- Noh, J. D. (2000), Model for correlations in stock markets, *Physical Review E*, **61**(5), 5981.
- Oatley, T., Winecoff, W. K., Pennock, A., and Danzman, S. B. (2013), The political economy of global finance: A network model, *Perspectives on Politics*, **11**(1), 133-153.
- Plerou, V., Gopikrishnan, P., Gabaix, X., Amaral, L. A. N., and Stanley, H. E. (2001), Price fluctuations, market activity and trading volume, *Quantitative Finance*, **1**(2), 262.
- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., and Stanley, H. E. (2002), Random matrix approach to cross correlations in financial data, *Physical Review E*, **65**(6), 066126.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018), CatBoost: Unbiased boosting with categorical features, *Advances in Neural Information Processing Systems*, 31.
- Sandoval, L. (2014), Structure of a Global Network of Financial Companies Based on Transfer Entropy, *Entropy*, **16**(8), 4443-4482.
- Sansa, N. A. (2020), The Impact of the COVID-19 on the Financial Markets: Evidence from China and USA, *Electronic Research Journal of Social Sciences and Humanities*, **2**.
- Sensoy, A., Sobaci, C., Sensoy, S., and Alali, F. (2014), Effective transfer entropy approach to information flow between exchange rates and stock markets, *Chaos, Solitons and Fractals*, **68**, 180-185.
- Schreiber, T. (2000), Measuring information transfer, *Physical Review Letters*, **85**(2), 461.
- Shannon, C. E. (1948), A mathematical theory of communication, *Bell System Technical Journal*, **27**(3), 379-423.
- Shapley, L. S. (1953), A value for n-person games, *Contributions to the Theory of Games*, **2**, 307-317.
- Shynkevich, Y., McGinnity, T. M., Coleman, S. A., Belatreche, A., and Li, Y. (2017), Forecasting price movements using technical indicators: Investigating the impact of varying input window length, *Neurocomputing*, **264**, 71-88.
- So, M. K., Chu, A. M., and Chan, T. W. (2021), Impacts of the COVID-19 pandemic on financial market connectedness, *Finance Research Letters*, **38**, 101864.
- Sun, X., Liu, M., and Sima, Z. (2020), A novel cryptocurrency price trend forecasting model based on LightGBM, *Finance Research Letters*, **32**, 101084.
- Turchenko, V., Beraldi, P., De Simone, F., and Grandinetti, L. (2011, September), Short-term stock price prediction using MLP in moving simulation mode, In *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, IEEE, **2**, 666-671.
- Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011), Transfer entropy: A model-free measure of effective connectivity for the neurosciences, *Journal of Computational Neuroscience*, **30**(1), 45-67.
- Weng, B., Ahmed, M. A., and Megahed, F. M. (2017), Stock market one-day ahead movement prediction using disparate data sources, *Expert Systems with Applications*, **79**, 153-163.
- World Bank (2021), Market capitalization of listed domestic companies (current US\$) | Data. data.worldbank.org. Retrieved 2021-09-20.
- Young, M. S. (2008), Revisiting non-normal real estate return distributions by property type in the US, *The Journal of Real Estate Finance and Economics*, **36**(2), 233-248.
- Young, M. S., Lee, S. L., and Devaney, S. P. (2006), Non-normal real estate return distributions by property type in the UK, *Journal of Property Research*, **23**(2), 109-133.
- Yue, P., Cai, Q., Yan, W., and Zhou, W. X. (2020), Information Flow Networks of Chinese Stock Market Sectors, *IEEE Access*, **8**, 13066-13077.
- Yue, P., Fan, Y., Batten, J. A., and Zhou, W. X. (2020), Information Transfer between Stock Market Sectors: A Comparison between the USA and China, *Entropy*, **22**(2), 194.
- Zhang, D., Hu, M., and Ji, Q. (2020), Financial markets under the global pandemic of COVID-19, *Finance Research Letters*, **36**, 101528.
- Zunino, L., Tabak, B. M., Figliola, A., Pérez, D. G., Garavaglia, M., and Rosso, O. A. (2008), A multifractal approach for stock market inefficiency, *Physica A: Statistical Mechanics and its Applications*, **387**(26), 6558-6566.

## 저자소개

**최인수** : 경희대학교 산업경영공학과에서 학사학위를 취득, 한국과학기술원 산업 및 시스템공학과에서 석사학위를 취득하였다. 현재는 한국과학기술원 산업 및 시스템공학과 금융공학 연구실 박사과정에 재학 중이며 관심 연구분야는 금융공학, 금융 데이터사이언스, 경제물리학, 네트워크 분석, 기계 학습, 핀테크 및 금융 서비스이다.

**김우창** : 서울대학교 산업공학과에서 학사와 석사학위를 취득하고 프린스턴대학교에서 경영과학 및 금융공학 박사학위를 취득하였다. 현재 한국과학기술원 산업 및 시스템공학과 교수로 재직 중이며 추계적 최적화, 자산부채관리, 금융 데이터사이언스가 연구 분야이다.