

AI를 활용한 공모주 투자전략

AI/Big Data Strategy | 2022. 3. 15

2021년 국내 주식시장에서는 역대 최대 규모의 IPO가 이루어졌습니다. IPO 시장 과열에 대한 우려가 확대되고 있어 보다 효과적인 IPO 투자가 요구되는 시기입니다. 이에 AI 방법론을 활용하여 IPO 종목의 시가수익률과 상장 후 주가 추세에 대한 예측을 시도하였습니다.



김규진 (퀀트전략)
02)2229-6737, kyoujin.kim@nhqv.com

국내 IPO 시장 과열에 따른 공모주 주가 변동성 확대 우려

2021년 IPO 규모가 최대치를 달성하며 IPO 시장이 활황세를 보이고 있다. 공모주 투자는 안전하면서도 높은 수익을 기대할 수 있는 투자로 인식되고 있다.

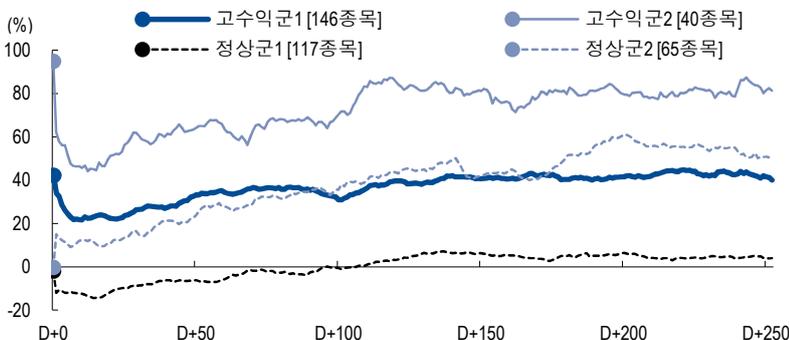
다만, 시장의 과열 양상으로 공모주의 주가 변동성도 확대되고 있다. 이에 보다 효과적인 공모주 투자 전략을 수립할 필요가 있다.

AI를 활용한 공모주 시가수익률 예측 모델과 이를 활용한 투자전략

AI 기술을 활용해 공모주의 공모가 대비 시가수익률이 20% 이상일 확률을 측정하는 모델을 구현하였다. 검증 결과 모델의 정확도는 약 70%를 기록하였다.

모델이 상승 전망한 공모주에 투자하여 상장 첫날 시초가에 매도 시, 모든 IPO에 참여하는 경우 대비 투자수익률이 11.5%p 개선되었다. 또한 모델을 활용해 공모주의 시가상승률의 원인을 과잉수요와 저평가발행 요인으로 분해할 경우, 공모주의 1년 장기주가 전망에도 유효하였다.

군집화된 공모주의 상장 후 주가 추세



주 1: K-Means Clustering(Appendix A 참조)을 통하여 구분한 8개 군 중에서 표본 수가 많은 4개 군(전체 중 78%)의 평균 수익률 그래프.

주 2: 군집화 대상은 2013년부터 2021년까지 상장한 공모주 중 SPAC을 제외한 종목의 상장시장 대비 상대수익률

자료: FnGuide, NH 투자증권 리서치본부

I. 국내 IPO 종목의 주가수익률 패턴

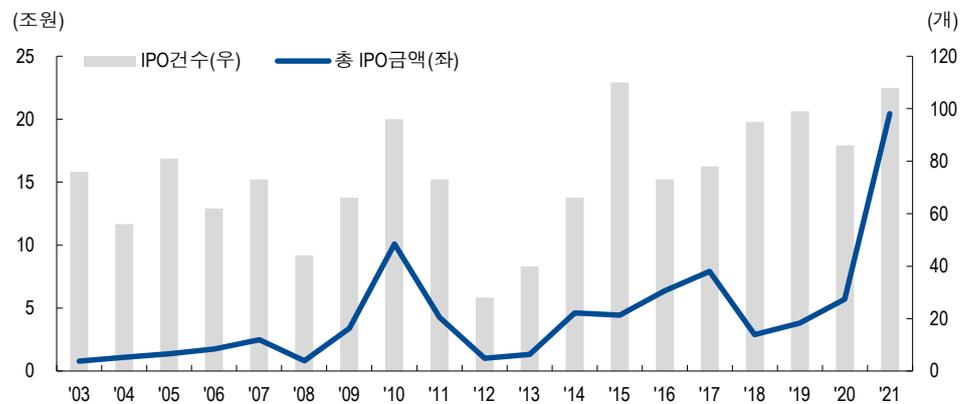
1. 국내 IPO 시장 추세와 성과 특성

IPO 시장은 신규상장, 재상장, 신주상장, 변경상장 중에서 신규상장에 해당한다. 신규상장에는 일반투자자를 대상으로 하는 공모상장, 소수 투자자를 대상으로 하는 사모상장, 주식을 신규로 발행하지 않는 직상장이 있으나, 코스피/코스닥에서의 상장 대부분은 공모상장을 통해 이루어진다.

국내 IPO 시장은 기업공개를 의무화하고 공개법인에 특혜를 주는 ‘기업공개촉진법’이 제정된 1972년 12월을 기점으로 제도화되었을 것으로 추정한다(금융감독원 사이버홍보관 ‘우리나라 금융발전의 역사’ 참고). 1998년 4월 1일에는 적정 가치 발견 기능 및 주식 배분 기능과 관련된 수요예측, 청약 등의 인수업무에 대한 규정이 제정되었다(이관영, 국내 신규공모절차 및 관련 제도의 변화, 2018, 224pg).

국내 IPO 시장의 규모는 2003년 7,800억원에서 2021년에는 20조4,460억원까지 가파르게 성장했다(그림1 참조).

그림1. 국내 IPO 시장 가파르게 성장



자료: FnGuide, NH 투자증권 리서치본부

공모주 주가 왜곡은 초기 성과 과잉, 장기 성과 부진으로 나타남

IPO 시장에서 수요예측과 청약 과정의 역할은 공모주의 적정 가치를 찾는 것에 있다. 그러나 시장의 불완전성으로 인하여 적정 가치 발견 기능에 문제가 발생하는 경우가 있으며, 상장 초기는 물론 장기적으로도 공모주 주가에 왜곡이 발생한다.

주가의 왜곡은 크게 두 가지 형태로 나타난다. 초기 성과 과잉과 1년 이상의 장기 성과 부진이다. 두 가지 현상이 겹치게 되는 경우 공모주의 주가 변동성이 증가하여 투자자의 손실로 이어질 수 있다. 초기 성과 과잉의 원인으로는 IPO 시장의 과잉 수요와 공모가를 낮게 설정하는 정보비대칭 현상, 시장의 비효율성이 지목되고 있으며, 장기 성과 부진의 원인은 과잉 수요의 해소와 과대 포장으로 설명된 바 있다(표1 참조).

표1. 공모주 주가 왜곡 설명 이론

분류	이론	설명
초기 성과 과잉	유행 효과-과잉수요 (Fad)	높은 IPO 초기 발행 수익률의 원인은 과잉수요(Aggarwal and Rivoli, 1990; Ritter, 1991; Levis, 1993; 임응기, 이성규, 1995; 윤평식, 김철중, 1997)
	정보비대칭 (Information asymmetry)	공모주 시장의 역선택 문제를 해결하기 위하여 주간사가 공모주를 저가로 발행(Rock, 1986)
	시장정상화 (Market Stabilization)	상장 초기 성과 과잉은 공모주 유통시장에서 공모주 초기 가격이 발행가격 이상으로 형성되는 시장 비효율성에서 기인(Ruud, 1993; Schultz and Zaman, 1994)
장기 성과 부진	유행 효과-수요해소 (Fad)	과잉수요로 발생한 높은 IPO 초기 수익률이 과잉수요가 해소되며 장기 저성과로 이어짐
	과대포장 (Window Dressing)	장기수익률(1년, 2년, 3년)의 하락은 기업 상장시 실제 기업내용을 과대포장하기 때문 (Teoh, Welch and Wong, 1998; 임병균, 1997; 김영규, 김영혜, 2000)

자료: 강원, "IPO가 우회상장보다 정보효율성이 더 높은가? : 코스닥시장을 중심으로. 재무관리연구", 2010, 121-156., NH투자증권 리서치본부

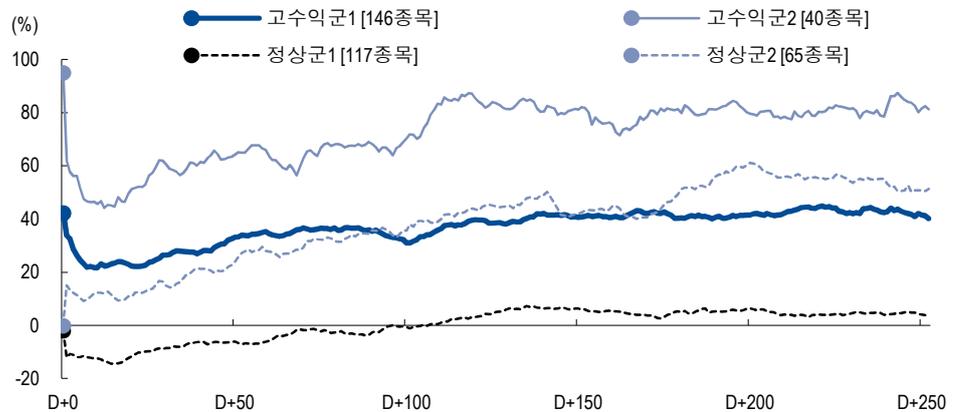
2. IPO 종목의 주가수익률 패턴

AI(인공지능) 기술 중 하나인 군집화 분석을 통해 한국 IPO 시장에서 발생하는 초기 성과 과잉과 장기 성과 부진 패턴을 확인할 수 있다. 군집화된 패턴은 크게 두 가지로 구분되었다. ①시초가가 공모가 대비 높게 형성된 고수익군과 ②시초가가 공모가와 유사하게 형성된 정상군이다(그림2 참조).

고수익군1과 2는 상장 직후 각각 평균 42%, 95%의 시가수익률을 기록하였다. 그러나 공모가 대비 1년 후 주가수익률이 각각 42%에서 22%까지, 95%에서 44%까지 저하되었다. 특히, 상장 직후 한달 간 공모주 과잉 수요 효과가 해소되며 큰 폭으로 주가가 하락하는 양상을 확인할 수 있다. 공모주 평균 대비 상대수익률 격차도 지속적으로 악화되는 것을 확인할 수 있다(다음페이지 그림3, 4 참조).

평균 시가수익률이 -2%, 0%를 기록하였던 정상군1과 2는 고수익군에 대비하여 상장 직후 낮은 변동성을 보이며 시가수익률 형성에 과잉수요가 미친 영향이 적은 것으로 추정된다. 장기적으로는 공모주 평균과 유사한 수익률을 보이거나 공모주 평균 대비 아웃퍼폼하였다(다음페이지 그림5, 6 참조).

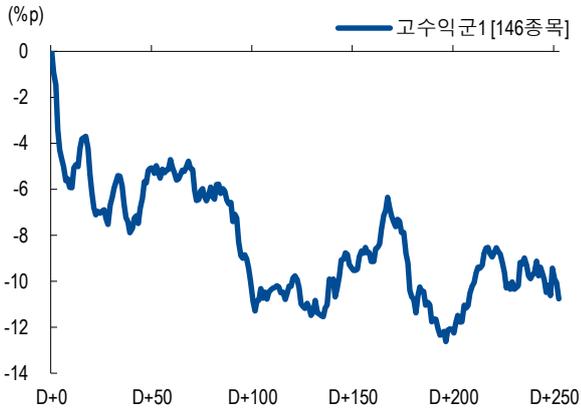
그림2. 군집화된 공모주의 상장 후 주가 추세



주 1: K-Means Clustering(Appendix A 참조)을 통하여 구분한 8개 군 중에서 표본 수가 많은 4개 군(전체 중 78%)의 평균 수익률 그래프.

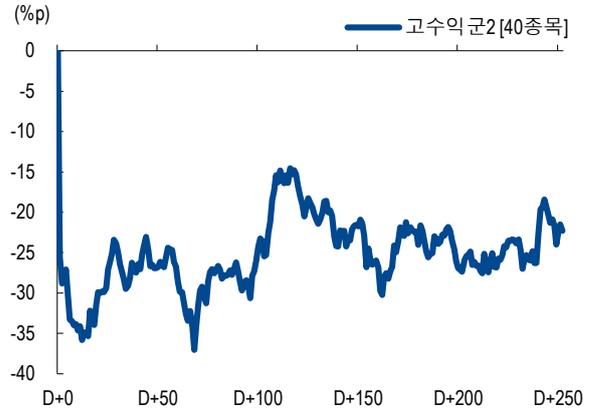
주 2: 군집화 대상은 2013년부터 2021년까지 상장한 공모주 중 SPAC을 제외한 종목의 상장시장 대비 상대수익률
자료: FnGuide, NH투자증권 리서치본부

그림3. 상장 이후 공모주 평균 대비 고수익군1의 상대수익률



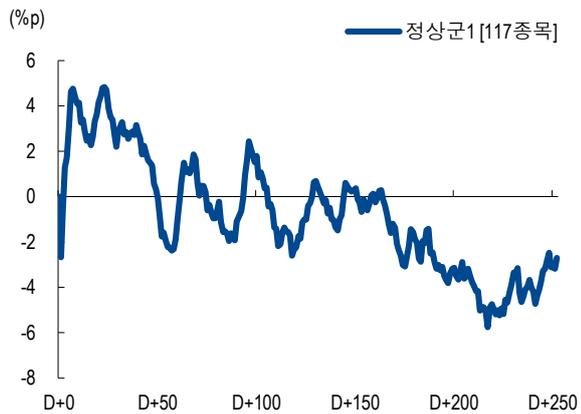
자료: FnGuide, NH투자증권 리서치본부

그림4. 상장 이후 공모주 평균 대비 고수익군2의 상대수익률



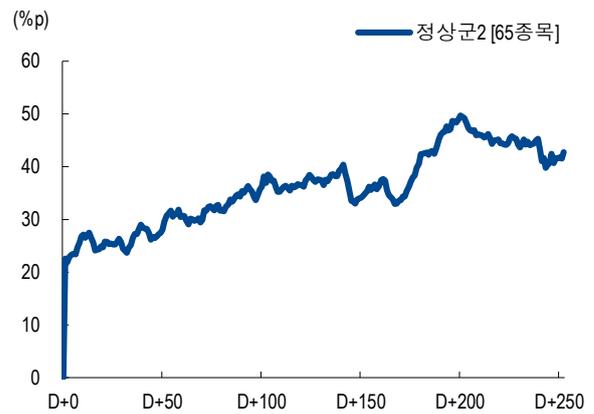
자료: FnGuide, NH투자증권 리서치본부

그림5. 상장 이후 공모주 평균 대비 정상군1의 상대수익률



자료: FnGuide, NH투자증권 리서치본부

그림6. 상장 이후 공모주 평균 대비 정상군2의 상대수익률



자료: FnGuide, NH투자증권 리서치본부

IPO 시장 과열 국면마다 장단기 성과 격차 심화

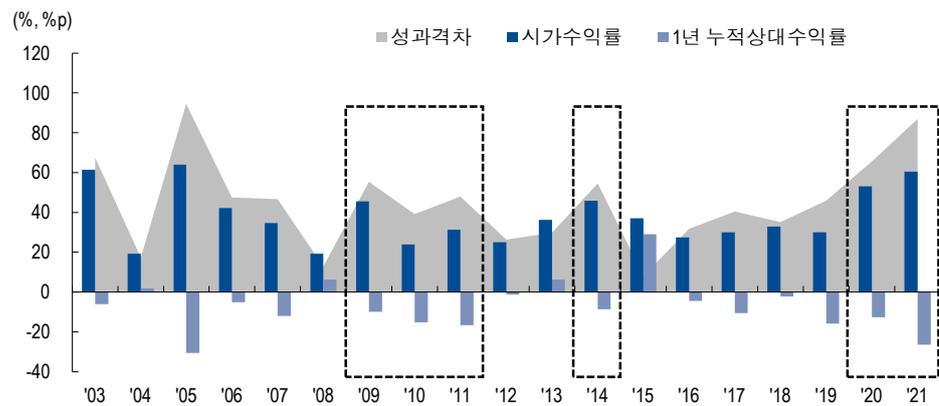
장단기 성과 격차는 IPO 시장의 과열 국면마다 심화되는 추세를 보여왔으며, 이 때마다 금융당국은 제도적인 준비를 통하여 왜곡을 제거하기 위해 노력해왔다(그림7 참조).

그 예시로 2009~2011년 IPO 시장 과열 국면을 살펴보면, 공모주는 높은 시가수익률을 보인 반면, 저조한 1년 누적상대수익률을 보여 큰 폭의 성과 격차가 발생하였다. 정부는 적정 공모가 산정을 위한 제도 정비와 공모기업 정보 공시를 시행하였다.

2014~2015년에는 코스닥 상장 장려를 위하여 금융당국이 상장 규정을 완화하면서 IPO 시장 과열 현상이 나타났으며, 역시 큰 폭의 성과 격차가 발생하였다. 다만, 시장개입으로 인한 과열 국면은 이듬해 즉각적으로 해소되었다.

2020년 과열 국면에서도 시가수익률과 1년 누적상대수익률 간 격차가 확대되었으며, 금융당국이 수요예측 단계에서의 적정 가치 발견 기능을 정상화하기 위해 2022년 2월 규정 개선안을 발표한 바 있다.

그림7. 과열 국면마다 시가수익률과 1년 누적상대수익률 간 격차 심화



주: 1년 누적상대수익률은 종목별 (시초가 대비 1년 후 주가 성과)에서 (상장 시장의 성과)를 제외한 값
 자료: FnGuide, NH 투자증권 리서치본부

3. 주가 패턴에 영향을 미치는 요인

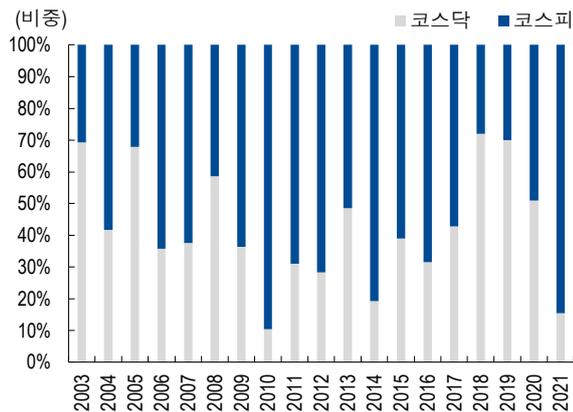
한국 IPO 시장은 상장시장별(코스피/코스닥)로 상장하는 종목에 요구되는 재무안정성이 다르다. 대기업에 속하는 계열사의 경우 상장 시 상대적으로 투자자의 높은 관심을 받는 경우가 많다. 한편 코스닥에는 특례상장 제도가 존재한다. 이와 같은 범주변수가 시가수익률에 미치는 영향이 클 것으로 판단된다.

1) 상장시장

장단기 성과 격차는 코스닥 > 코스피 순

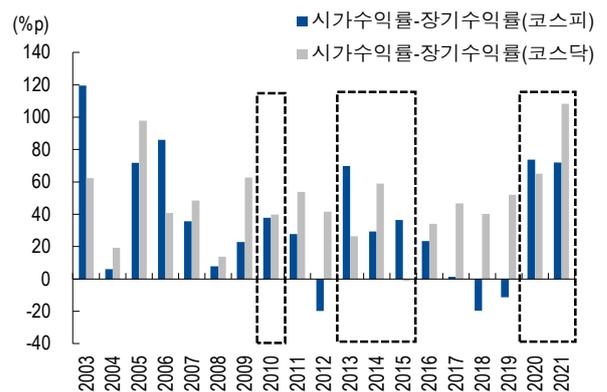
시가수익률과 장기수익률 간 격차는 코스닥 상장 종목이 코스피 상장 종목 대비 모든 기간에서 전반적으로 높은 수준을 보이는 것으로 확인된다. 다만, 코스피 상장 종목은 2010년, 2014년, 2021년 등 IPO 시장 과열 국면에서 공모금액 비중이 높아지는 추세를 보이며, 이 시기 일시적으로 장단기 성과 격차가 확대되었다(그림9 참조).

그림8. 상장시장별 공모금액 비중



자료: FnGuide, NH투자증권 리서치본부

그림9. 상장시장별 장단기 성과 격차



자료: FnGuide, NH투자증권 리서치본부

표2. 상장시장별 평균 공모주수와 수익률

분류	코스피 상장 종목	코스닥 상장 종목
연도별 평균 상장종목수(개)	10	52
연도별 평균 공모비중(%)	59.6	40.4
평균 시가수익률(%)	29.0	39.9
평균 1년 누적수익률(%)	-6.8	-7.1
평균 3년 누적수익률(%)	-25.0	-10.3
평균 장단기 성과 격차(%p)	35.3	47.9

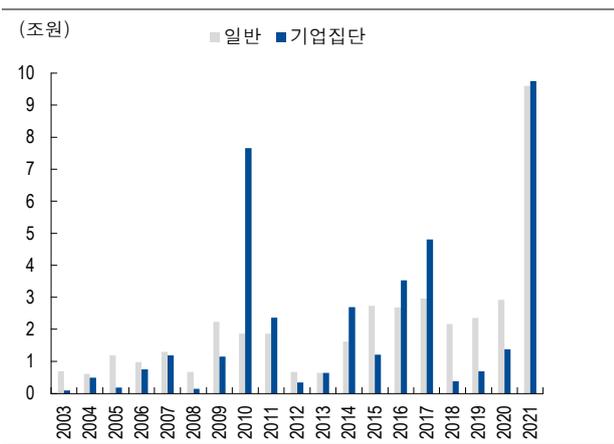
자료: NH투자증권 리서치본부

2) 기업집단 포함 여부

장단기 성과격차는
일반 > 기업집단 순

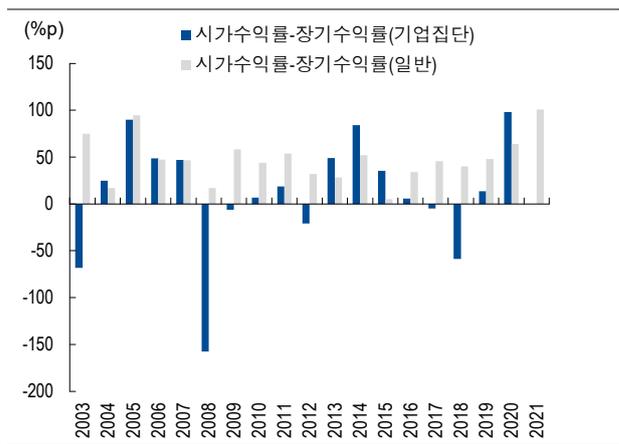
기업집단 포함 여부는 상장시장 구분보다 명확한 격차를 보였다. 기업집단 공모주는 시가수익률이 일반 공모주 대비 높았으나, 장기 저성과 효과가 없었다. 반면, 일반 공모주의 경우 장기저성과 효과가 존재하여 기업집단 공모주 대비 장단기 성과 격차가 더 큰 것으로 조사되었다(그림10,11, 표3 참조).

그림10. 기업집단별 공모금액 추이



자료: FnGuide, NH투자증권 리서치본부

그림11. 기업집단별 장단기 성과 격차



자료: FnGuide, NH투자증권 리서치본부

표3. 기업집단별 평균 공모주수와 수익률

분류	기업집단 공모주	일반 공모주
연도별 평균 상장종목수(개)	5	57
연도별 평균 공모비중(%)	43.1	56.9
평균 시가수익률(%)	36.1	38.1
평균 1년 누적수익률(%)	22.8	-8.7
평균 3년 누적수익률(%)	13.0	-13.1
평균 장단기 성과 격차(%p)	11.4	47.5

자료: NH투자증권 리서치본부

3) 코스닥 소속부 구분

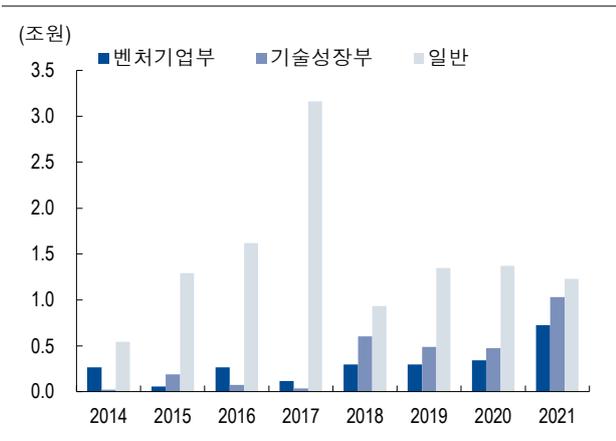
장단기 성과 격차는 벤처기업부 > 일반 > 기술성장부 순

코스닥 시장 상장사는 우량기업부, 중견기업부, 벤처기업부, 기술성장기업부 총 4부제로 구분된다. 기술성장특례상장 시에는 기술성장기업부로, 일정 재무기준 충족 시 벤처기업부 또는 우량기업부로, 그외에는 중견기업부로 등록된다.

기술성장특례상장 제도는 기술 또는 성장성을 인정받은 기업의 상장조건을 완화시켜주는 제도이다. 2015년 이후 적용대상이 모든 섹터로 확대되면서 기술성장특례를 활용한 상장 케이스는 꾸준히 증가하고 있으며, 규모 기준 2021년 코스닥 상장의 35%를 차지하였다(그림12 참조). 한편, 2017년에 최대치를 기록했던 우량기업부 공모금액은 감소하는 추세를 보이고 있다. 즉, 최근 코스닥 시장에서는 상장기업의 퀄리티보다는 성장성을 기준으로 상장이 이루어지고 있는 것으로 추정된다.

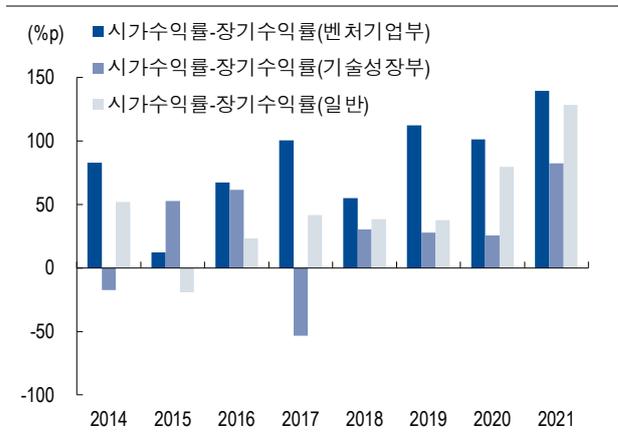
벤처기업부, 기술성장기업부 편입 기업은 그 외 기업 대비 확연한 장단기 성과 격차를 보인다(그림13, 표4 참조). 벤처기업부 공모주는 일반 공모주 대비 높은 시가수익률을 보이지만, 큰 폭의 장기 저성과를 기록하며 큰 장단기성과 격차를 보였다. 반면, 기술성장특례를 통해 상장한 공모주는 상대적으로 낮은 시가수익률, 양의 장기수익률을 보이며 작은 장단기 성과 격차를 기록하였다.

그림12. 소속부별 공모금액 추이



자료: FnGuide, NH투자증권 리서치본부

그림13. 소속부별 장단기 성과 격차



자료: FnGuide, NH투자증권 리서치본부

표4. 소속부별 평균 상장종목수와 수익률

분류	벤처기업부	기술성장부	일반
연도별 평균상장종목수(개)	12	13	29
연도별 평균공모비중(%)	17.2	18.2	64.6
평균 시가수익률(%)	55.1	31.8	39.8
평균 1년 누적수익률(%)	-26.3	6.7	-3.0
평균 3년 누적수익률(%)	-16.6	8.5	-5.3
평균 장단기 성과 격차(%p)	83.8	26.2	47.8

자료: NH투자증권 리서치본부

4) 업종 구분

IT, 산업재, 건강관리
업종의 장단기 성과
격차가 큰 편

공모주가 속한 업종에 따라서도 장단기 성과 격차가 발생하였다. IT, 산업재, 건강관리 업종은 타 업종 대비 장단기 성과 격차가 확대될 확률이 높은 것으로 관측되었다(표5 참조). 특히, IT와 건강관리 업종은 전반적으로 높은 수준의 장단기 성과 격차가 지속되었다(다음페이지 그림14,16,18 참조).

표5. 업종별 장단기 성과 격차와 분포 특성

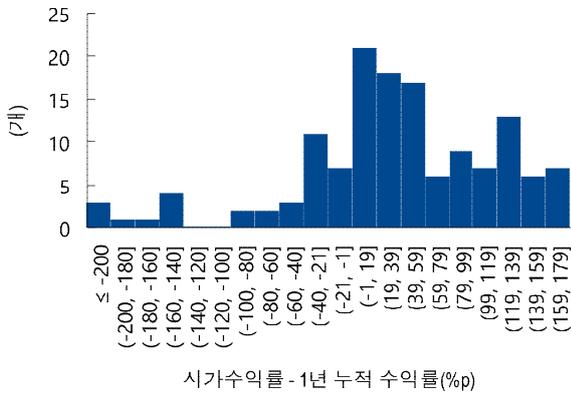
업종구분	왜도	첨도	평균 장단기 성과 격차(%p)
산업재	-5.69	50.39	27.7
건강관리	-2.58	10.72	32.7
IT	-1.09	4.72	50.1
커뮤니케이션서비스	-0.51	0.73	44.4
필수소비재	-0.30	2.47	84.2
소재	-0.10	1.42	56.5
경기관련소비재	0.02	0.75	48.3
금융	0.52	2.90	44.1

주1: 왜도가 (-) 방향으로 높을수록 장단기 성과 격차(시가수익률 - 1년 누적 수익률)가 높은 종목이 많음

주2: 일반적으로 첨도가 높을수록 종목 분포가 평균값에 집중. 그러나 왜도의 음의 절대값이 높고 첨도가 높다면, 종목들의 장단기 성과 격차가 높은 쪽으로 넓고 길게 분포되어 있음을 의미

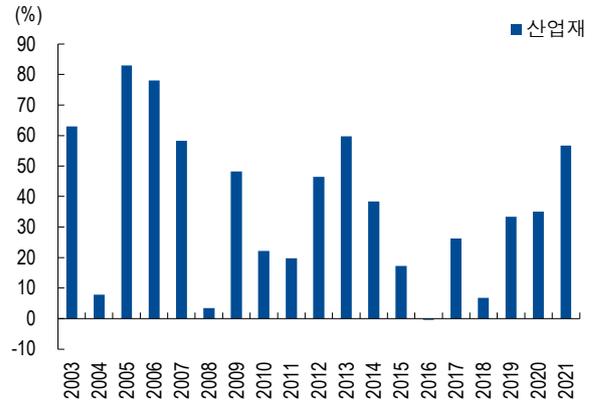
자료: NH투자증권 리서치본부

그림14. 산업재 업종 공모주의 장단기 성과 격차 히스토그램



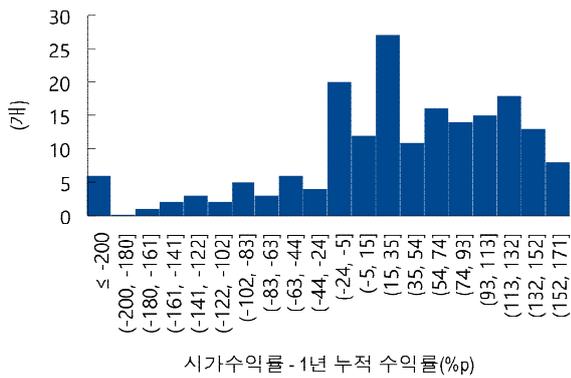
자료: FnGuide, NH투자증권 리서치본부

그림15. 산업재 업종 의 연도별 평균 장단기 성과 격차



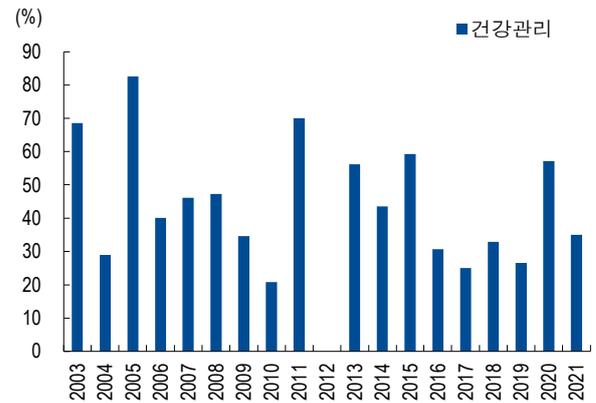
자료: FnGuide, NH투자증권 리서치본부

그림16. 건강관리 업종 공모주의 장단기 성과 격차 히스토그램



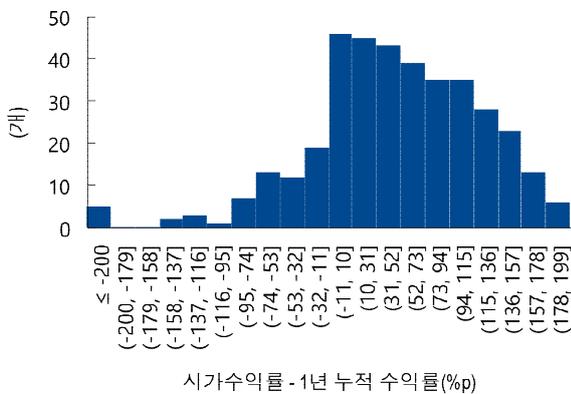
자료: FnGuide, NH투자증권 리서치본부

그림17. 건강관리 업종 의 연도별 평균 장단기 성과 격차



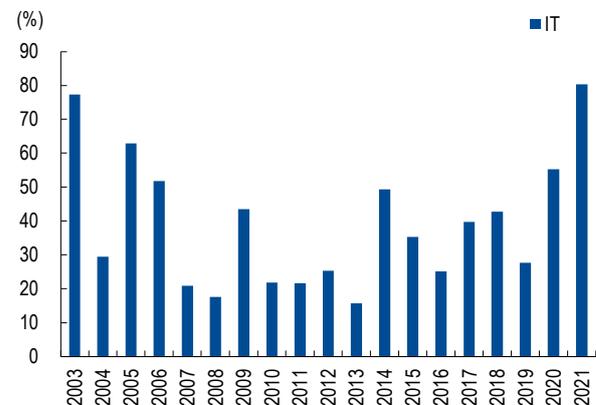
자료: FnGuide, NH투자증권 리서치본부

그림18. IT 업종 공모주의 장단기 성과 격차 히스토그램



자료: FnGuide, NH투자증권 리서치본부

그림19. IT 업종 의 연도별 평균 장단기 성과 격차



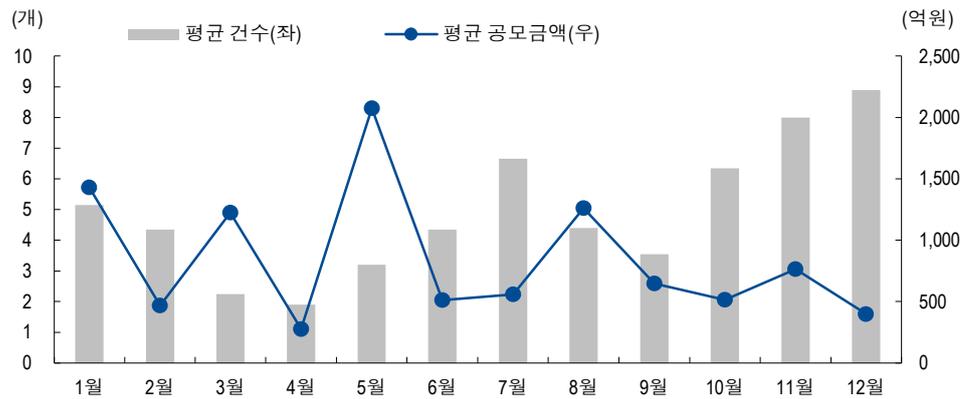
자료: FnGuide, NH투자증권 리서치본부

5) 상장월 구분

7, 9, 12월 장단기 성과 격차가 컸음

시기별로는 6~8월, 10~12월에 IPO 상장 건수가 가장 많았다(그림20 참조). 1, 5, 8월에는 상대적으로 대형 공모주가 상장되었으며 장단기 성과 격차도 함께 커졌다. 분포의 특성에서는 7, 9, 12월에 다른 월 대비 큰 폭의 장단기 성과 격차를 보일 가능성이 더 높았다(표6 참조).

그림20. 월별 상장 공모주수 및 공모 금액 평균



자료: FnGuide, NH 투자증권 리서치본부

표6. 월별 장단기 성과 격차와 분포 특성

월	총 건수	왜도	첨도	장단기 성과 격차(%p)
1월	104	-1.54	4.25	-24.7
2월	86	-0.44	0.03	-58.7
3월	49	-0.10	-0.67	-76.4
4월	43	-0.46	0.15	-79.3
5월	75	-0.31	-0.36	-91.5
6월	107	-1.13	3.29	-64.2
7월	142	-2.90	15.67	-67.9
8월	99	0.06	-0.60	-85.5
9월	80	-2.57	9.86	-42.8
10월	133	-0.26	0.56	-66.6
11월	167	-0.60	0.91	-35.8
12월	223	-6.06	57.02	-47.9

주1: 왜도가 (-) 방향으로 높을수록 장단기 성과 격차(시초가 수익률 - 1년 누적 수익률)가 높은 종목이 많음
 주2: 일반적으로 첨도가 높을수록 종목 분포가 평균값에 집중. 그러나 왜도의 음의 절대값이 높고 첨도가 높다면, 종목들의 장단기 성과 격차가 높은 쪽으로 넓고 길게 분포되어 있음을 의미
 자료: NH투자증권 리서치본부

4. 제도적 한계점

현재 금융투자협회는 ‘증권 인수업무 등에 관한 규정’을 통해 수요예측 및 공모가 결정 과정에서 적절한 가치 평가가 이루어지도록 규정하고 있으나, 실질적으로는 주관회사와 수요예측 참여기관의 자율성에 큰 비중을 두고 있다(표7 참조). 이에 따라 향후에도 공모가 저평가와 과잉 수요로 인한 공모주 초기 성과 과잉 및 장기 성과 부진 현상의 발생 가능성을 배제하기 어렵다.

따라서 공모주 수요예측 및 청약에 참여함에 있어 해당 종목의 주가 왜곡 가능성을 선제적으로 판단하여 투자전략을 계획하고 투자성과를 개선하는 것이 중요할 것으로 판단한다.

표7. 신규 공모주의 공모가 결정 규정

조항	상세설명
주식의 공모가격 결정	<p>주식의 공모가격 결정 방법</p> <ol style="list-style-type: none"> 1. 인수회사와 발행회사가 협의하여 단일가격으로 결정 2. 기관투자자를 대상으로 수요예측을 실시. 그 결과를 감안하여 인수회사와 발행회사 협의 하에 공모가격 결정 3. 기관투자자로부터 경매의 방식으로 입찰가격과 수량을 제출받은 후 최저공모가격 이상의 입찰에 대해 해당 입찰자가 제출한 가격 결정 4. 기관투자자로부터 경매의 방식으로 입찰가격과 수량을 제출받은 후 산정한 단일가격으로 결정 <p>수요예측 방법</p> <ol style="list-style-type: none"> 1. 발행회사의 우리사주 조합원 및 인수회사는 배제. 수요 예측 참여자는 고유재산과 위탁재산을 구분하여 참여 2. 수요예측 참여자별 가격 및 수량의 정보는 누설 금지 3. 수요예측 참여 대상자는 기관투자자로 한정
주식의 배정	<p>공모주식의 청약자 유형별 배정 비율</p> <ol style="list-style-type: none"> 1. 우리사주 조합원: 20% 2. 일반청약자: 25% (우리사주 조합원 청약비율이 20% 미만인 경우, 일반청약자에 최대 5% 추가 배분) (균등방식 배정을 선택한 경우, 일반 청약 배정 수량 중 50% 이상을 모든 일반청약자에게 동등하게 배정) 3. 고위험고수익투자신탁: 5% (코스닥시장 상장을 위한 경우, 벤처기업투자신탁 30% 이상 배분) 4. 기관투자자: 잔여비율 <p>배정 결정 사유</p> <ol style="list-style-type: none"> 1. 인수회사 및 인수회사의 이해관계인 (위탁재산으로 청약하는 집합투자회사, 투자일임회사, 신탁회사 제외) 2. 발행회사의 이해관계인 3. 발행회사 또는 인수회사와 공모관련 이해관계가 있는 자 4. 주금납입능력을 초과하는 물량 또는 현저히 높거나 낮은 가격을 제시하는 등 수요예측 참여에 진실성이 낮다고 판단되는 자 5. 자신이 대표주관업무를 수행한 발행회사의 기업공개를 위하여 금융위원회에 제출된 증권신고서의 "주주에 관한 사항"에 주주로 기재된 주요주주에 해당하는 기관투자자 및 창업투자회사등

자료: 금융투자협회 법규정보시스템, '증권 인수업무 등에 관한 규정', NH 투자증권 리서치본부

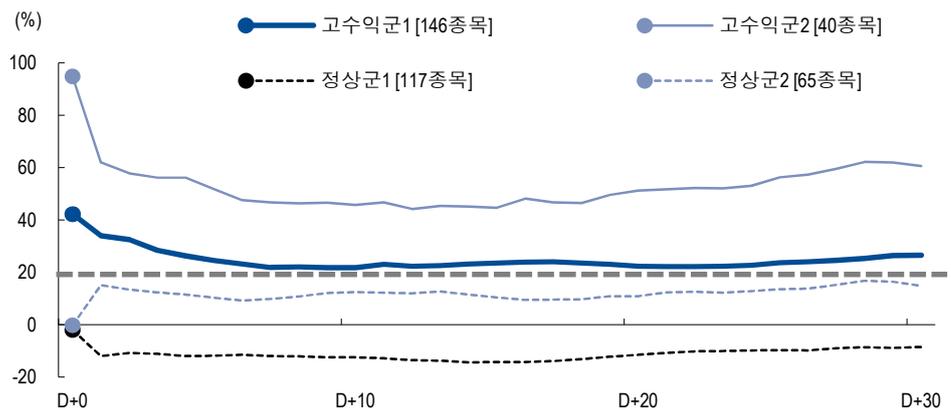
II. 시를 활용한 IPO 시가수익률 예측 모델

1. 모델 목표

초기 성과 과잉은 공모가 저평가와 과잉 수요가 중복될 경우 더욱 큰 폭으로 발생하며, 장기 성과 부진으로 이어질 가능성이 높다. 따라서 모델은 공모주 시가 상승/안정을 분류하는 기능을 수행하여 초기 성과 과잉 여부를 판단하고, 이를 공모가 저평가와 과잉 수요로 분해하여 장기 성과 부진 여부 판단에 도움을 주는 것을 목표로 한다.

모델의 학습 데이터로 선정한 2013~2019년 상장 종목의 절반은 시가수익률이 20% 미만을 기록하였다. 앞서 살펴본 군집분석에서도 직관적으로 20%를 기준으로 정상군과 고수익군이 구분되는 것을 확인할 수 있다(그림21 참조). 이에 따라 시초가 수익률 20% 미만을 안정 공모주, 20% 이상을 상승 공모주로 분류한다.

그림21. 시가수익률 20%를 기준으로 고수익군과 정상군 구분



자료: FnGuide, NH 투자증권 리서치본부

2. 변수 선정

시가수익률에 영향을 미치는 요인은 재무변수, 상장 당시 시장상황 등 종목별 개별 변수와 상장시장, 업종구분, 상장일 등 범주변수로 구분된다.

개별변수로는 IPO 지표, 재무 지표, 최근 시초가 수익률, 최근 예탁금 규모, 경제지표, 감성지표, 투자주체별 거래대금 비중 등 총 67개 변수를 선정하였다(표8 참조). 기관 투자자의 경우에는 실질적으로 수요예측 단계에서 공모주 투자 규모가 결정되므로 수요예측일을 기준으로 얻을 수 있는 정보만을 개별변수에 포함시켰다.

표8. 개별변수 목록

변수	항목	기준
IPO 지표 (횡단면)	상장시장 대비 시가총액 비중(%)	시가총액 / 상장시장 시가총액
	신주 모집 비율(%)	신주 모집 주식수 / 공모주식수
	구주 매출 비율(%)	구주 매출 주식수 / 공모주식수
	유통비율(%)	유통가능주식수 / 상장주식수
	우리사주 비율(%)	우리사주 배정물량 / 공모주식수
	기관투자자 비율(%)	기관투자자 배정물량 / 공모주식수
	일반투자자 비율(%)	일반투자자 배정물량 / 공모주식수
	상장주선사 수(개)	상장주선사 수
재무 지표 (연간 시계열, 3년)	상대 자본총계 비율(%)	시가총액 대비 지배자본총계 / 상장시장 시가총액 대비 지배자본총계
	상대 매출액 비율(%)	시가총액 대비 매출액 / 상장시장 시가총액 대비 매출액
	상대 영업이익 비율(%)	시가총액 대비 영업이익 / 상장시장 시가총액 대비 영업이익
	상대 순이익 비율(%)	시가총액 대비 지배순이익 / 상장시장 시가총액 대비 지배순이익
	상대 현금흐름 비율(%)	시가총액 대비 영업현금흐름 / 상장시장 시가총액 대비 영업현금흐름
	상대 순이익률(%)	순이익률 / 상장시장 순이익률
	상대 ROE(%)	지배 ROE / 상장시장 지배 ROE
	상대 ROA(%)	ROA / 상장시장 ROA
	상대 부채비율(%)	부채비율 / 상장시장 부채비율
	상대 유통비율(%)	유통비율 / 상장시장 유통비율
	상대 판관비율(%)	판관비율 / 상장시장 판관비율
IPO 시장 수익률 지표 (시계열, 최근 10개)	시초가 수익률(%)	최근 10개 종목 시초가 수익률
IPO 시장 자금 지표 (일간 시계열, 20일)	시총 대비 예탁금 비율(%)	전체시장 예탁금 / 전체시장 시가총액

자료: NH투자증권 리서치본부

표8. 개별변수 목록(계속)

변수	항목	기준
경제 지표 (월간 시계열, 12개월)	미국장단기금리차(%p)	미국 국채 10년물 금리 - 미국 국채 3개월물 금리
	한국장단기금리차(%p)	한국 국채 10년물 금리 - 한국 국채 3년물 금리
	한국실질금리(%)	한국 콜 금리 - 물가상승률
	재고순환지표(P)	재고순환지표
	경제심리지수(P)	경제심리지수
	기계류내수출하지수(%)	기계류내수출하지수(선박제외) %y-y
	건설수주액(%)	건설수주액(실질) %y-y
	수출입물가비율(%)	수출입물가비율
	광공업생산지수(%)	광공업생산지수 %y-y
	서비스업생산지수(%)	서비스업생산지수(도소매업제외) %y-y
	건설기성액(%)	건설기성액(실질) %y-y
	소매판매액지수(%)	소매판매액지수 %y-y
	내수출하지수(%)	내수출하지수 %y-y
	수입액(%)	수입액(실질) %y-y
	비농림어업취업자수(%)	비농림어업취업자수 %y-y
	생산자제품재고지수(%)	생산자제품재고지수 %y-y
	소비재수입액(%)	소비재수입액(실질) %y-y
	취업자수(%)	취업자수 %y-y
	통화 대비 주식시장시총(%)	전체 주식시장 시가총액 / M2 통화량
	주식시장 옵션 프리미엄(P)	KOSPI 200 콜옵션 프리미엄
	주식시장 변동성(%)	KOSPI 실현변동성
	주식시장 ROE(%)	자기자본이익률(12m fwd)
	주식시장 Payout(%)	배당성향(12m fwd)
	주식시장 EP(%)	당기순이익(12m fwd)/시가총액
	주식시장 5yEP(%)	5년평균 당기순이익(12m fwd)/시가총액
	주식시장 배당률(%)	배당수익률(12m fwd)
주식시장 BP(%)	순자산(12m fwd)/시가총액	
감성 지표 (주간 시계열, 13주)	부정 감성(P)	Negative Fear Sadness Anger Disgust
	긍정 감성(P)	Positive Joy Anticipation Trust Surprise
수급 지표 (일간 시계열, 20일)	기관 콜옵션 거래비중(%)	기관 일간 콜옵션 순매수대금 / 주체별 콜옵션 매수대금 절대값 합
	외인 콜옵션 거래비중(%)	외인 일간 콜옵션 순매수대금 / 주체별 콜옵션 매수대금 절대값 합
	개인 콜옵션 거래비중(%)	개인 일간 콜옵션 순매수대금 / 주체별 콜옵션 매수대금 절대값 합
	기관 풋옵션 거래비중(%)	기관 일간 풋옵션 순매수대금 / 주체별 풋옵션 매수대금 절대값 합
	외인 풋옵션 거래비중(%)	외인 일간 풋옵션 순매수대금 / 주체별 풋옵션 매수대금 절대값 합
	개인 풋옵션 거래비중(%)	개인 일간 주식 순매수대금 / 주체별 주식 매수대금 절대값 합
	기관 주식 거래비중(%)	기관 일간 주식 순매수대금 / 주체별 주식 매수대금 절대값 합
	외인 주식 거래비중(%)	외인 일간 주식 순매수대금 / 주체별 주식 매수대금 절대값 합
개인 주식 거래비중(%)	개인 일간 주식 순매수대금 / 주체별 주식 매수대금 절대값 합	

주: 감성지표는 코스피 지수와 상관관계가 높음. 세부사항은 김규진, 'AI와 대체데이터를 활용한 KOSPI 예측 모델', 2021 참조

자료: NH투자증권 리서치본부

범주변수로는 앞서 성과 특성 요인으로 파악된 상장시장 구분, 기업집단 포함 여부, 코스닥 소속부 구분, 산업구분, 상장월 구분 총 5가지 변수를 선정하였다(표9 참조).

표9. 범주변수 목록

변수	항목
상장시장 구분	코스피/코스닥 (2)
기업집단 포함 여부	포함/비포함 (2)
코스닥 소속부 구분	일반/벤처기업부/기술성장부 (3)
업종 구분	산업재/건강관리/IT/커뮤니케이션서비스/필수소비재/소재/경기관련소비재/금융/유틸리티/에너지 (10)
상장월 구분	1월 - 12월 (12)

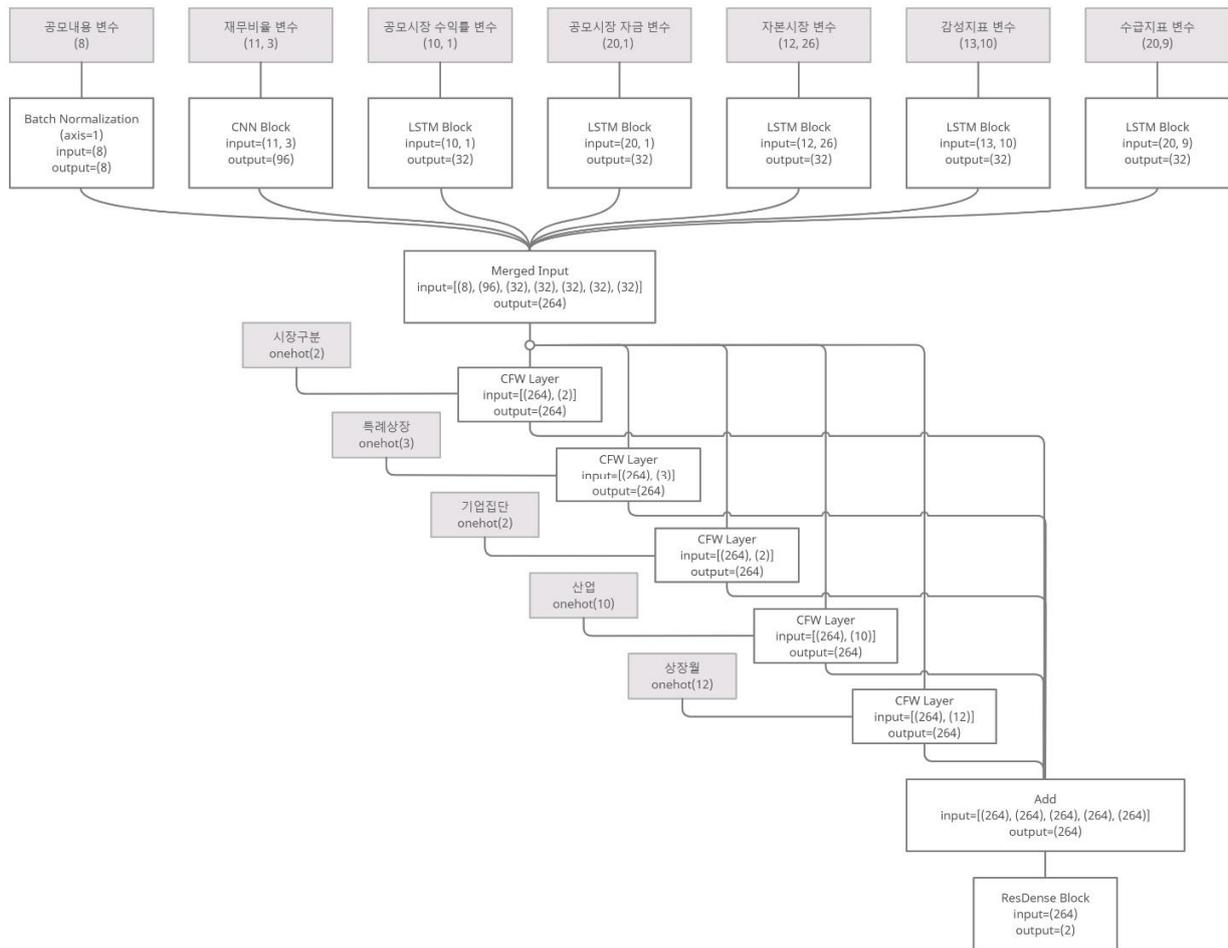
자료: NH 투자증권 리서치본부

3. 모델 설정

앞서 언급한 바와 같이 본 AI 모델은 분류기 모델을 사용하여 공모주 시가수익률을 상승/안정 2종으로 분류 예측하는 것을 목표로 한다.

정답으로 사용되는 예측의 결과가 명확하게 주어지므로 학습 방법으로는 지도학습을 사용하여 최적화를 진행하였다.

그림22. AI를 활용한 IPO 시가수익률 예측 모델 순서도



자료: NH투자증권 리서치본부

모델 제작에 중점적으로 고려한 요소는 아래와 같다.

1. 횡단면 변수와 서로 다른 주기와 길이의 시계열 변수의 통합

- 횡단면 변수는 정규화 후 통합
- 최근 3년 재무제표 변수는 CNN Block으로 전처리 후 통합(다음페이지 그림 23, Appendix B, E 참조)
- 각 주기별 시계열은 변수 성질별로 각각 LSTM Block으로 전처리 후 통합(다음페이지 그림24, Appendix C, E 참조)

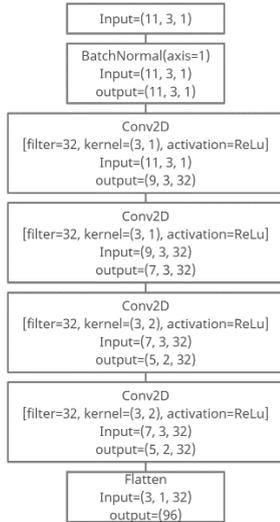
2. 범주변수가 개별변수 비중에 영향을 줄 때 변수의 통합

- 범주변수별로 개별변수의 비중을 조정해주도록 CFW(Categorical Feature Weighting) Custom Layer를 사용
- CFW Layer는 범주변수별로 개별변수에 부여하는 가중치를 설명 가능하도록 설계(Appendix D, E 참조)

총 557개 표본(2012년 중반~2022년 초반) 중, 최근 표본인 30%(167개, 2019년 후반~2022년 초반)를 테스트 표본으로 선정하였다. 나머지 390개 표본 중 임의의 30%(117개)는 검증변수로 선택하였고, 나머지 70%(273개)를 학습 표본으로 선정하였다.

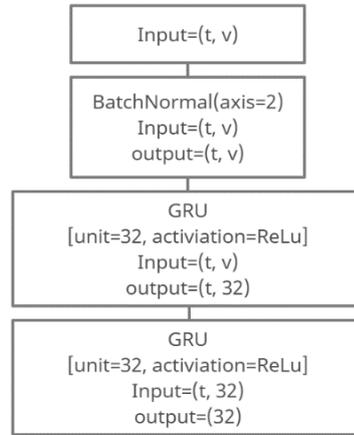
손실값은 Categorical Cross Entropy식으로 계산하였다. 시작학습속도는 $3E-4$ 로 설정하였으며, 학습 스케줄은 Cosine Annealing with Warm Restart를 사용하되 Restart 시 Decaying 계수와 주기 증가계수는 0으로 설정하였다.

그림23. CNN Block Plot



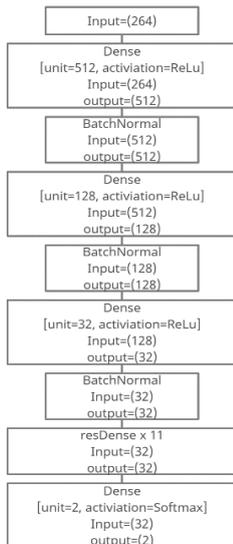
자료: NH투자증권 리서치본부

그림24. LSTM Block Plot



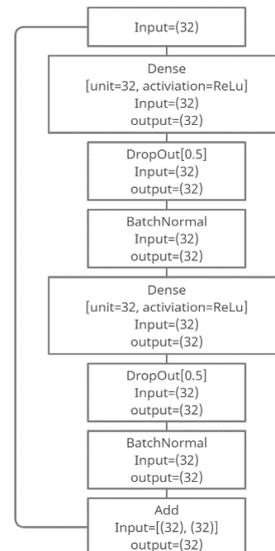
자료: NH투자증권 리서치본부

그림25. ResDense Block Plot



자료: NH투자증권 리서치본부

그림26. resDense Plot



자료: NH투자증권 리서치본부

4. 예측 모델 성과

2019년 10월 22일부터 2022년 1월 20일까지 상장한 총 167개의 종목을 기준으로 Out of Sample 검정을 한 결과 예측률은 70.1%를 기록하였다. 안정에 대한 예측정확도는 65%인 반면, 상승에 대한 예측정확도는 72%를 기록하여 상승에 대한 예측정확도가 더 높은 것으로 확인되었다(표10,11 참조).

성과 측면에서는 모든 공모주에 투자하였을 때 대비 상승 전망 공모주에만 투자하였을 때 평균 시초가 수익률이 11.5%p 높았다(표12 참조).

표10. 총기간 모델 예측률

매트릭스		예측		
		안정	상승	
실제	안정	36	19	재현도 72.3
	상승	31	81	
		정밀도 81.0		정확도 70.1

F1-Score : 76.4

주: 정밀도(예측 상승 중 일치할 확률), 재현도(실제 상승 중 일치할 확률), 정확도(전체 사례 중 일치할 확률), F1-Score(재현도와 정밀도의 조화평균)

자료: NH투자증권 리서치본부

표11. 2020년 이후 대형 IPO 종목

코드	기업명	예측	실제	시가수익률(%)	1년수익률(%)
A259960	크라프트	안정	안정	-10.0	-39.0
A293490	카카오게임즈	상승	상승	100.0	53.8
A302440	SK 바이오사이언스	상승	상승	100.0	12.7
A361610	SK 아이이테크놀로지	상승	상승	100.0	-42.4
A373220	LG 에너지솔루션	안정	상승	99.0	-26.3
A323410	카카오뱅크	안정	상승	38.0	-12.0
A377300	카카오페이	안정	상승	100.0	-22.8
A352820	하이브	안정	상승	100.0	14.7

자료: NH투자증권 리서치본부

표12. 예측 시나리오별 투자 수익률

Case	시가 수익률(%)	3개월 수익률(%)	6개월 수익률(%)	1년 수익률(%)	시가-1년 수익률(%p)
모든 공모주에 투자한 경우	51.8	-2.9	-2.4	-9.1	60.9
Case1. 예측(상승), 실제(상승)	77.7	-10.9	-9.8	-22.7	100.3
Case2. 예측(상승), 실제(안정)	1.8	-0.2	-12.2	1.2	0.7
예측(상승)	63.3	-8.9	-10.3	-18.0	81.3
Case3. 예측(안정), 실제(상승)	74.1	-19.5	-25.8	-37.4	111.5
Case4. 예측(안정), 실제(안정)	0.7	28.7	41.0	35.2	-34.5
예측(안정)	34.6	6.1	9.3	1.1	33.5

자료: NH투자증권 리서치본부

장단기 성과 격차 요인은 앞서 정리한 바와 같이 저평가발행과 과잉수요 요인으로 분해가능하다. 모델이 저평가발행 여부를 정확히 예측할 수 있다고 가정한다면, 모델을 통하여 공모주의 상장 이후 추가흐름 역시 추정할 수 있다(표13 참조).

추정한 추가흐름이 실제 시가수익률, 1년 누적 수익률 평균과 일치하며 장기 추가흐름 전망의 유효성을 보여준다.

표13. 예측 시나리오별 향후 추가흐름 추정

Case	저평가발행	과잉수요	향후 추가흐름 추정	실제 추가흐름
예측(상승), 실제(상승)	O	알 수 없음	정상가치 회복 후, 과잉수요 해소에 따른 하락 가능	시가수익률(+77.7%) 1년수익률(-23.7%)
예측(상승), 실제(안정)	O	X	정상가치 회복 후, 장기 우상향	시가수익률(+1.8%) 1년수익률(1.2%)
예측(안정), 실제(상승)	X	O	정상평가 되었으나 과잉수요로 인한 고평가 발생, 장기 하락	시가수익률(+74.1%) 1년수익률(-37.4%)
예측(안정), 실제(안정)	X	X	정상평가 되었으며, 시장 무관심으로 정상평가 유지, 장기 우상향	시가수익률(+0.7%) 1년수익률(+37.9%)

자료: NH투자증권 리서치본부

5. 범주변수별 공모가 저평가에 미치는 영향 분석

앞서 설계한 Categorical Feature Weighting Layer의 기능은 범주변수별로 개별요인의 비중값을 가산/감산하는 기능을 한다. 따라서 각 범주변수별 가산/감산 정도를 통해 개별변수별 시가수익률 기여도를 파악할 수 있다.

상장시장 범주변수의 경우, 코스피는 코스닥 대비 상대적으로 재무 지표가 시가수익률에 영향을 미쳤다. 반대로 경제 지표와 같은 시황변수의 영향은 적었다. 코스닥은 감성지표의 영향이 높았으며 예탁금 규모나 투자자별 수급요인은 영향이 적었다(표14 참조).

표14. 상장시장 범주변수 가산/감산변수

상장시장	가산변수	감산변수
코스피	재무지표	경제지표
코스닥	감성지표	예탁금 규모, 수급지표

자료: NH투자증권 리서치본부

코스닥 소속부 범주변수의 경우, 벤처기업부 소속기업은 재무안정성이 이미 반영되어 있으므로 상대적으로 예탁금 규모, 최근 시가수익률 등 시황변수가 중요변수로 반영되었다. 기술성장기업의 경우, 재무 지표가 중요변수로 반영된 반면, 시황변수의 영향은 적었다(표15 참조).

표15. 코스닥 소속부 범주변수의 가산/감산변수

코스닥 소속부	가산변수	감산변수
일반	감성지표	예탁금 규모
벤처기업부	예탁금 규모, 최근 시가수익률	수급지표
기술성장기업부	재무지표	예탁금 규모, 감성지표

자료: NH투자증권 리서치본부

기업집단 범주변수의 경우, 기업집단 소속 공모주는 IPO 지표(유통비율, 투자자별 배정비율 등)가 중요변수로 작용하였다. 반면, 일반 공모주는 개별 종목의 재무지표가 시가수익률 결정에 영향을 미쳤다(표16 참조).

표16. 기업집단 범주변수별 가산/감산변수

기업집단	가산변수	감산변수
일반	재무지표	수급지표
기업집단	IPO 지표	감성지표

자료: NH투자증권 리서치본부

그 외 업종별, 상장월 범주변수별 가산/감산변수 목록은 아래와 같다(표17,18 참조, 가산/감산 정도는 Appendix F 참조)

표17. 업종 범주변수별 가산/감산변수

업종	가산변수	감산변수
IT	재무지표	경제지표, 감성지표
건강관리	-	IPO 지표, 경제지표
경기관련소비재	재무지표	예탁금, 수급지표
금융	-	-
산업재	-	IPO 지표
소재	경제지표	IPO 지표, 재무지표
커뮤니케이션서비스	-	경제지표, 재무지표, 감성지표
필수소비재	IPO 지표, 예탁금, 감성지표	-

자료: NH투자증권 리서치본부

표18. 상장월 범주변수별 가산/감산변수

상장월	가산변수	감산변수
1월	경제지표	-
2월	수급지표	재무지표, 예탁금
3월	IPO 지표, 감성지표	예탁금
4월	-	IPO 지표, 최근 시가수익률
5월	경제지표	감성지표
6월	경제지표, 수급지표	IPO 지표
7월	-	-
8월	경제지표	-
9월	예탁금, 감성지표	-
10월	경제지표	IPO 지표

자료: NH투자증권 리서치본부

6. 향후 상장 공모주 전망

종목코드	종목명	시장구분	수요예측일	상장예정일	상승확률	가산변수	감산변수
A388050	지투파워	코스닥	2022-03-17	2022-04-01	98.9	재무지표, 감성지표	예탁금 규모, 수급지표
A310210	보로노이	코스닥	2022-03-14	2022-03-31	99.7	재무지표, 감성지표	예탁금 규모, 경제지표
A396300	세아메카닉스	코스닥	2022-03-10	2022-03-25	100.0	감성지표, 경제지표	예탁금 규모, 수급지표
A366030	공구우먼	코스닥	2022-03-07	2022-03-23	100.0	감성지표, 재무지표	예탁금 규모, 수급지표
A388720	유일로보틱스	코스닥	2022-02-24	2022-03-18	89.6	최근 시가수익률, 감성지표	예탁금 규모, IPO 지표

주: 상기 확률은 특정 지표에 의해 추출되었으며, 당사 전망과는 무관함

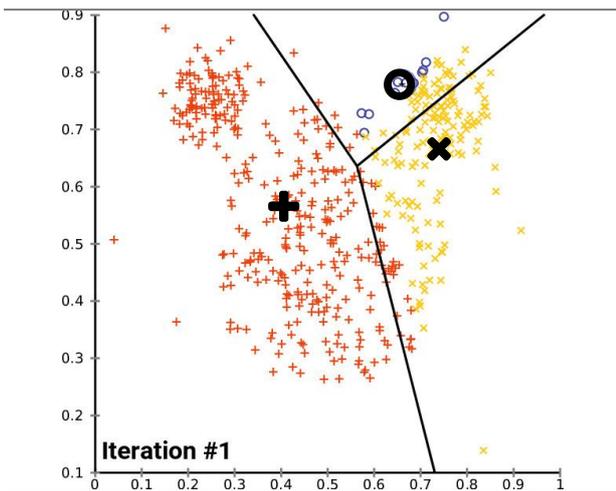
III. Appendix

Appendix A. K-Means Clustering

Clustering은 주로 비지도학습을 사용한 군집화 방법론을 의미한다. 이 때 군집화란 표본들을 특정 규칙에 의하여 몇 개의 집단으로 묶는 것을 의미한다. K-Means Clustering은 평균(Means)을 기준으로 K개의 집단으로 표본을 군집화시키며, 다음과 같은 과정을 따른다.

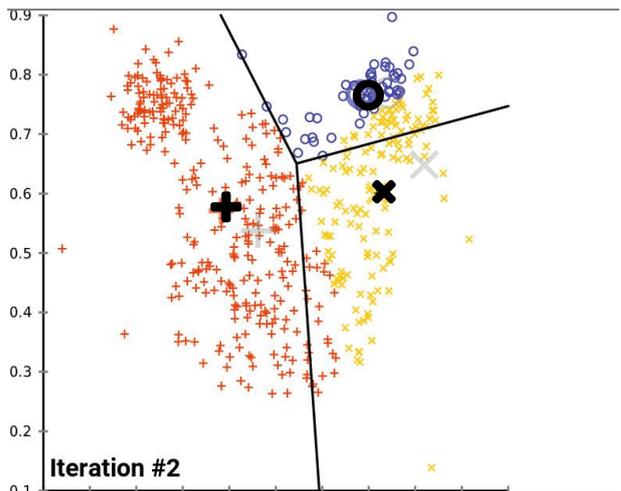
1. 평면상에 임의로 K개의 점을 찍는다
2. K개의 점들에 직선거리가 가장 가까운 점들을 배정(그림A1 참조)
3. 배정된 점들의 평균값으로 K개의 점들을 각각 이동(그림A2 참조)
4. 점들의 이동이 멈출 때까지 2, 3번 과정을 반복(그림A3 참조)

그림A1. 임의의 K개의 점에 가까운 점들을 배정



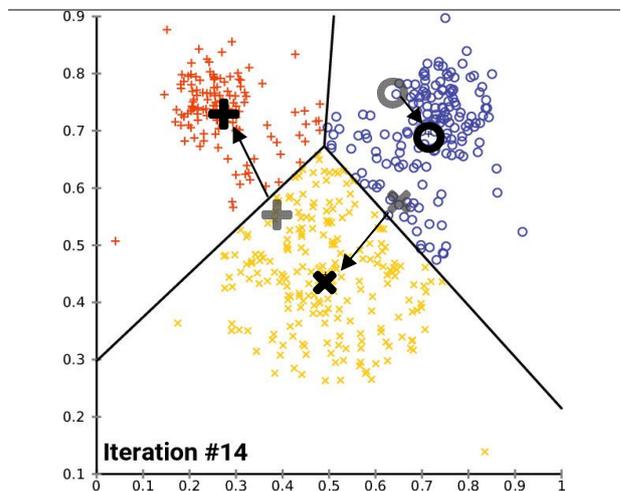
자료: Wikimedia Commons. "K-means convergence.gif", NH투자증권 리서치본부

그림A2. 배정된 점들의 평균으로 각 K개의 점들 이동



자료: Wikimedia Commons. "K-means convergence.gif", NH투자증권 리서치본부

그림A3. 과정을 반복하여 점들의 위치(K-Means) 수렴



자료: Wikimedia Commons. "K-means convergence.gif", NH투자증권 리서치본부

Appendix B. CNN (Convolution Neural Network)

AI/머신러닝 수행에 있어서 변수의 개수는 모델의 정확도 상승에 도움이 된다. 그러나 변수의 개수가 많을수록 학습을 시킬 때에 많은 시간과 자원이 소요된다. 어떤 경우에는 많은 변수 개수가 모델 학습을 방해하는 요인으로 작용해 성과를 저해하기도 한다. 이러한 문제는 간단한 이미지 파일을 분석할 때에도 나타난다.

이미지를 분석할 때 변수는 이미지의 가장 작은 단위인 픽셀에 해당한다. 아래 이미지는 총 271*308 픽셀로 이루어져 있으며, 해당 이미지를 분석하기 위해서는 총 8만개의 변수를 학습시켜야 한다. 색을 포함하여 분석하는 경우 변수는 $271*308*3=25$ 만개로 증가한다.

그림B1. 아래와 같은 간단한 이미지도 8만개의 픽셀로 구성되어있다.



자료: NH투자증권 리서치본부

변수의 개수 문제를 차치하더라도 그림의 분석을 위하여 2차원 데이터를 1차원 데이터로 변경하는 과정에서 그림 분석에 있어 중요한 위치 데이터가 손실되는 현상이 발생한다. 또한 그림의 단순 회전만으로도 그림을 분석하는 모델의 성능이 크게 저하될 수 있다.

따라서 이미지가 가지고 있는 변수를 압축하면서도, 위치데이터를 보존하고, 이미지 전반의 특성 또한 추출해낼 수 있는 분석 방법론이 요구된다.

그림B2. 2차원 데이터를 1차원으로 정렬시 위치 데이터 손실



자료: NH투자증권 리서치본부

그림B3. 그림의 단순 회전만으로도 모델의 예측 성능 저하

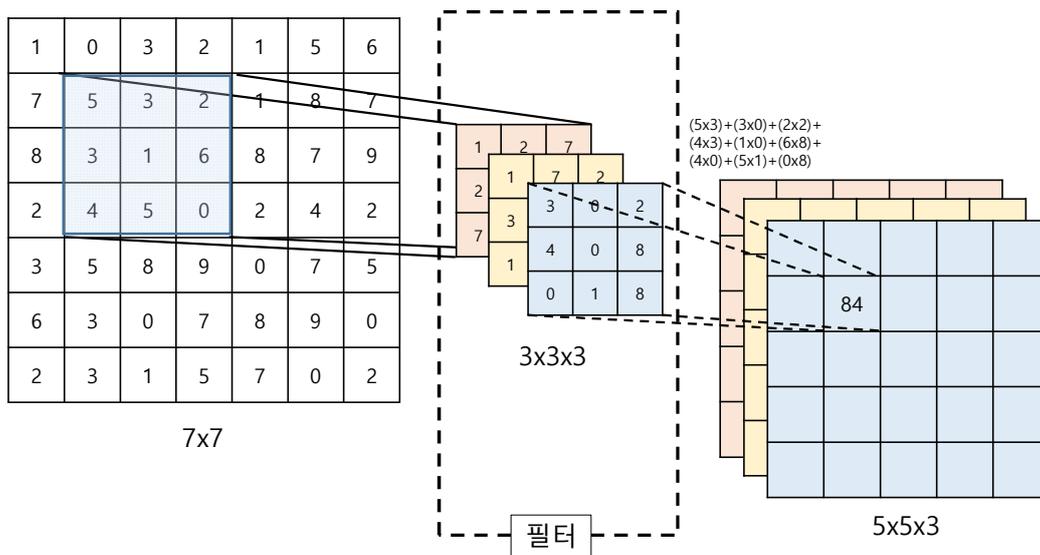


자료: NH투자증권 리서치본부

CNN(Convolution Neural Network)은 여러 개의 필터를 사용한 합성곱(Convolution)을 시행하여 그림의 영역별 특성을 압축하여 저장한다(그림B4 참조). 여러 개의 돋보기로 그림의 일부분을 조금씩 보는 것과 같은 이러한 방식을 통하여 CNN은 위치정보를 보존하면서도 그림의 종합적인 특성을 파악할 수 있다.

다수의 머신러닝 대회에서 CNN을 사용한 모델이 수상하는 등 이미지 처리 분야에서 CNN은 독보적인 모델 성능을 자랑하고 있다.

그림B4. 그림의 일정 영역의 특성 정보를 여러 개의 필터를 통하여 압축 저장

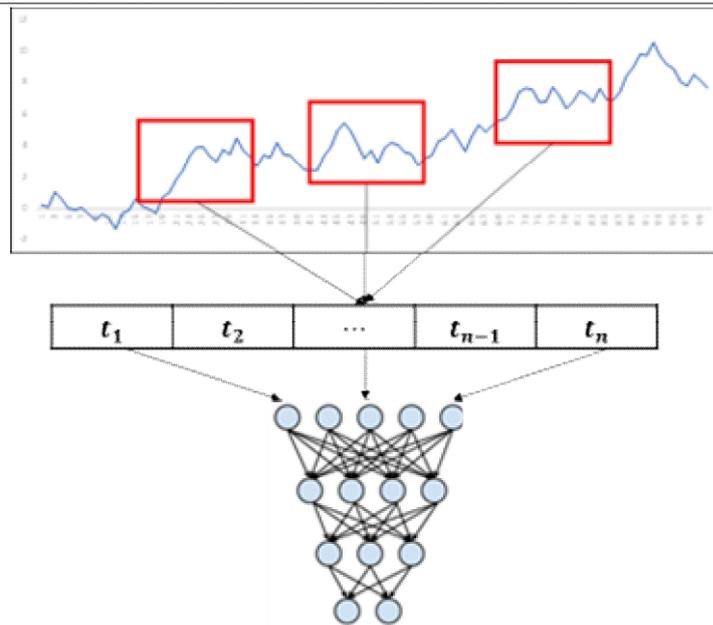


자료: NH투자증권 리서치본부

Appendix C. RNN (Recurrent Neural Network)

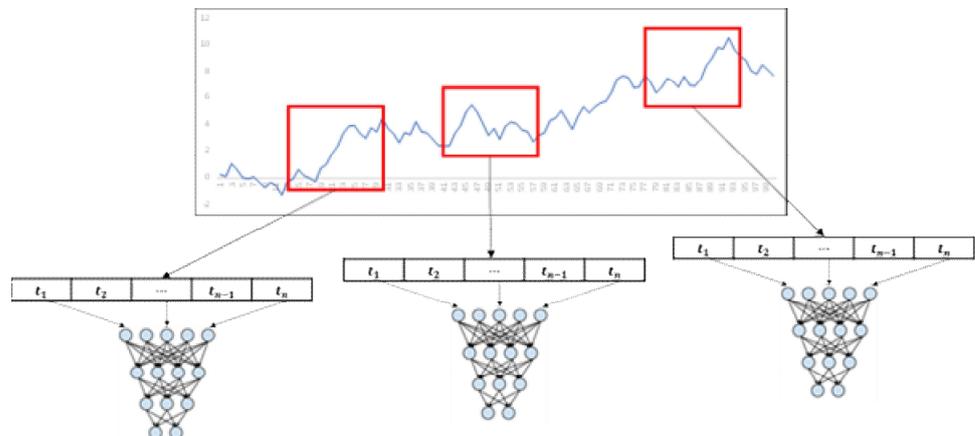
기존 신경망층의 각 노드(비중값)는 모든 시계열값을 동시에 입력받아 처리한다. 따라서 시계열의 순서를 인식하지 못하며, 시계열 특성별 이전 상태값의 반영이 일률적이다(그림C1 참조). 그러나 이를 보완하기 위하여 시계열의 특성별로 서로 다른 모델을 제작하는 것은 지나치게 소모적이다(그림C2 참조).

그림C1. 그림의 여러 부분을 일률적인 비중값으로 분석



자료: NH투자증권 리서치본부

그림C2. 시계열 특성별로 서로 다른 모델을 제작하는 것은 지나치게 소모적

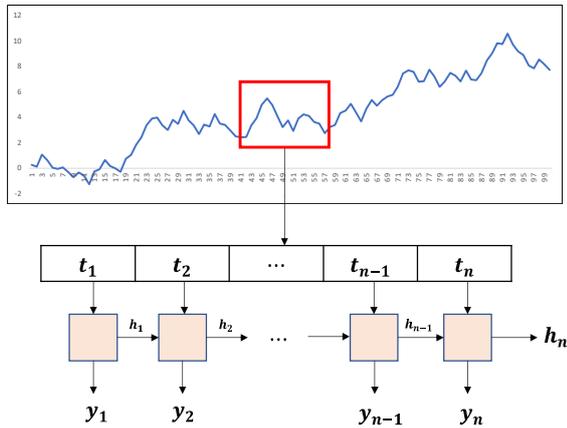


자료: NH투자증권 리서치본부

이를 해결하기 위하여 RNN(Recurrent Neural Network)은 이전의 값을 다음 블록에 전달하는 은닉상태값(h_n), 이전 은닉상태값과 현재시계열 값을 입력받아 계산된 출력값(y_n)을 출력한다. 출력된 이전 은닉상태값을 재귀적(Recurrent)으로 다음 상태값 계산에 반영하므로 시계열의 순서를 반영할 수 있다. 또한 현재상태값에 따라 이전 시계열의 반영이 달라지므로 시계열의 특성 역시 반영된다(그림C3 참조).

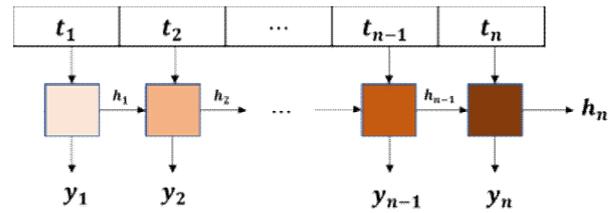
그러나 단순 RNN 구조 역시 입력하는 시계열의 길이가 길면 길수록 이전 시계열값의 반영이 어려워진다는 문제가 존재한다. RNN 블록 처리가 중첩적으로 이루어지므로 점차 희미해지는 현상이 발생하는 것이다(그림C4 참조). 따라서 현재에는 이전 상태값을 보존하면서 전달할 수 있는 통로를 추가한 LSTM(Long Short Term Memory) 구조의 RNN이 주로 사용되고 있다(그림C5 참조).

그림C3. 이전 은닉상태값을 재귀적으로 다음 상태값 계산에 반영하는 RNN 구조



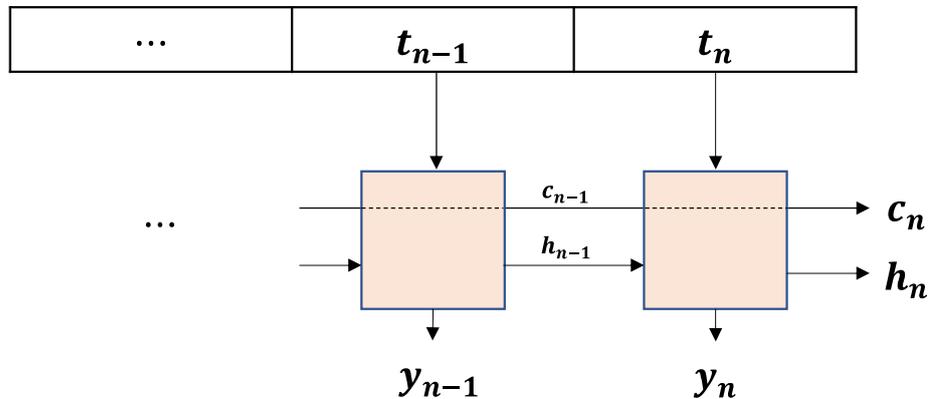
자료: NH투자증권 리서치본부

그림C4. 시계열의 길이가 길수록 이전 시계열값의 반영이 어려워지는 단점이 발생



자료: NH투자증권 리서치본부

그림C5. 이전 은닉상태값을 보존하면서 전달할 수 있는 통로 구조를 포함한 LSTM(Long Short Term Memory) 구조



자료: NH투자증권 리서치본부

Appendix D. CFW (Categorical Feature Weighting)

CFW Layer는 모델 성능의 개선보다는 모델의 설명 가능성에 초점을 두고 제작하였다. 학습 가능한 변수 W를 통하여 범주변수별 개별변수 가중치를 학습시킨다(그림D1,D2 참조).

CFW Layer의 추가는 모델 학습에 있어서 모델최종단 Fully Connected Layer 비중값의 Global Maximum 수렴에 어려움을 주는 경향이 있었다. 따라서 Residual Fully Connected Layer(본문에서는 ResDense로 표현)를 도입하여 수렴 성능 개선을 시도하였다. CFW Layer를 통해 조정된 비중값을 합하는 방식이 수렴 성능이 가장 좋았으며, 설명력 측면에서도 가장 유리하였다.

그림D1. CFW(Categorical Feature Weighting) Layer 표현식

$$Row - Wise \sum_1^N \{Diag(X_N) \cdot W_{N \times C} * OneHot(i_c)\} = Y_{N \times 1}$$

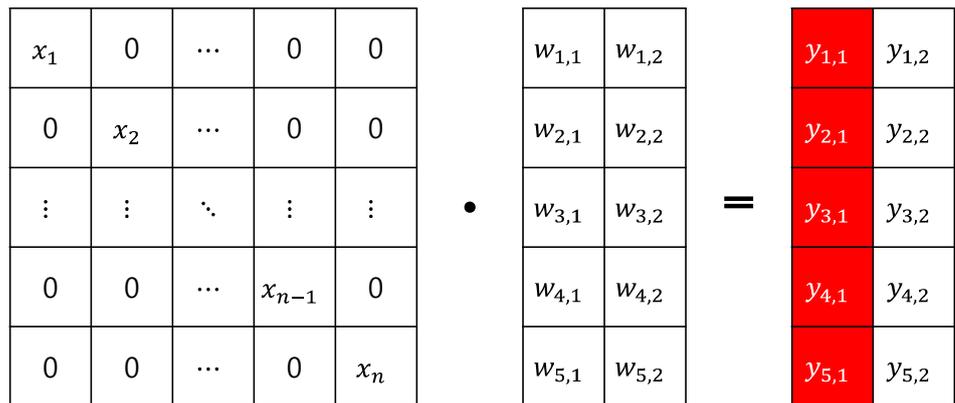
x: 입력변수

w: 훈련가능 변수

i: 총 C개의 범주변수

자료: NH투자증권 리서치본부

그림D2. CFW(Categorical Feature Weighting) Layer 다이어그램



자료: NH투자증권 리서치본부

Appendix E. 모델 구조별 결과 비교

모델의 F1 Score는 Random Forest 모델이 가장 좋았으나, CNN+LSTM+ResDense+CFW(Categorical Variable) 모델도 이와 유사한 수준의 성과를 보여주었다. 다만, CNN+LSTM+ResDense+CFW(Categorical Variable) 모델은 CFW Layer를 통하여 모델 설명 가능성 측면에서 더 우수하다.

표 E1. 모델별 정확도

모델 구성	재현도	정밀도	정확도	F1 Score
Vanilla Fully Connected Layer	69.84	39.29	47.90	50.29
+CNN	69.01	43.75	49.10	53.55
+CNN+LSTM	73.53	66.96	61.68	70.09
+CNN+LSTM+ResDense	73.95	78.57	67.07	76.19
+CNN+LSTM+ResDense +Categorical Variable	73.68	62.50	59.88	67.63
+CNN+LSTM+ResDense +CFW(Categorical Variable)	81.00	72.32	70.06	76.42
Random Forest	75.83	81.25	70.06	78.45

자료: NH투자증권 리서치본부

Appendix F. CFW 범주변수별 개별변수 가산/감산값

표 F1. 범주변수별 개별변수 가산/감산값

변수	구분	IPO 지표	재무지표	예탁금	이전시가수익률	경제지표	감성지표	수급지표
상장시장	코스닥	-0.014	0.001	-0.018	-0.004	-0.001	0.011	-0.012
	코스피	-0.022	0.013	0.000	-0.001	-0.009	0.009	-0.001
코스닥 소속부	일반	-0.005	-0.004	-0.011	0.001	-0.015	0.004	-0.006
	벤처	-0.003	-0.005	0.011	0.008	0.003	0.003	-0.010
	기술성장	-0.006	0.007	-0.015	0.002	0.003	-0.007	0.002
기업집단 포함여부	일반	0.015	0.007	-0.008	0.004	-0.004	-0.005	-0.007
	기업집단	0.023	-0.004	-0.007	-0.005	-0.002	-0.015	-0.003
업종	IT	-0.002	0.006	-0.002	-0.004	-0.013	-0.010	0.001
	건강관리	-0.020	0.003	-0.005	0.003	-0.009	0.000	0.002
	경기관련소비재	-0.009	0.011	-0.017	0.005	-0.001	0.009	-0.015
	금융	-0.008	0.004	-0.007	-0.007	-0.004	-0.008	0.005
	산업재	-0.025	0.000	0.004	-0.008	-0.002	-0.009	-0.008
	소재	-0.034	-0.010	0.005	0.000	0.011	0.006	0.001
	에너지	-0.004	-0.002	-0.008	-0.004	0.004	0.017	0.005
	유틸리티	0.008	0.003	0.004	-0.003	0.009	-0.012	0.009
	커뮤니케이션서비스	0.006	-0.007	0.001	-0.004	-0.011	-0.008	0.001
필수소비재	0.013	-0.006	0.016	-0.005	0.001	0.010	-0.005	
상장월	1월	-0.002	0.004	0.001	-0.007	0.025	0.005	-0.005
	2월	-0.002	-0.010	-0.012	0.006	-0.004	-0.002	0.009
	3월	0.029	0.003	-0.013	0.000	0.001	0.015	0.006
	4월	-0.018	0.007	0.001	-0.017	-0.002	0.004	0.000
	5월	0.009	-0.005	0.001	0.007	0.015	-0.010	-0.006
	6월	-0.022	-0.006	-0.008	0.004	0.012	0.003	0.014
	7월	0.008	0.003	-0.005	0.003	-0.005	0.005	0.006
	8월	0.012	0.002	-0.007	-0.012	0.010	-0.002	0.005
	9월	-0.010	0.003	0.019	-0.008	0.004	0.018	-0.005
	10월	-0.052	-0.004	-0.004	-0.001	0.013	0.009	0.007
	11월	0.002	0.002	-0.006	0.014	0.002	-0.009	-0.004
	12월	-0.018	0.003	0.008	-0.006	0.002	-0.012	-0.012
평균		-0.005	0.001	-0.003	-0.001	0.001	0.001	-0.001

자료: NH투자증권 리서치본부

Compliance Notice

- 당사는 동 자료를 기관투자자 또는 제3자에게 사전 제공한 사실이 없습니다.
 - 동 자료에 게시된 내용들은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭없이 작성되었음을 확인합니다.
-

고지 사항

본 조사분석자료에 수록된 내용은 당사 리서치본부의 금융투자분석사가 신뢰할 만한 자료 및 정보를 바탕으로 최선을 다해 분석한 결과이나 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 투자자의 투자판단을 위해 작성된 것이며 어떠한 경우에도 주식 등 금융투자상품 투자의 결과에 대한 법적 책임소재를 판단하기 위한 증빙자료로 사용될 수 없습니다. 본 조사분석자료는 당사의 저작물로서 모든 지적 재산권은 당사에 귀속되며 당사의 동의 없이 복제, 배포, 전송, 변형, 대여할 수 없습니다.
