2021

Measuring Investor Sentiment by combining ML and photo from news

指导教师：林健武老师
汇报人：徐嘉营 徐有珩 李晨霄
时间：2021.10.21

目录
CONTENTS

PART ONE

Introduction

Background

Model

Structure

单击此处输入标题名称

This is a good space for a short subtitle

PhotoPe (photo pessimism)

daily market-level investor sentiment index
how visual content in news relates to financial markets.

Model

convolutional neural networks (CNNs),
to classify a large sample of news photos
based on sentiment.

Relationship

1. Based on behavioral model, investors' sentiment predicts market return reversal
2. correlation and effectiveness of information embedded in photo and text.

Additional insight

We have many PowerPoint templates that has been specifically designed to help anyone that is stepping into the world of PowerPoint for the very first time.

PART TWO

Data & Variable

Photo Classification

This is a good space for a short subtitle

CNNs

key model of the research area of mage classification
Achieve high accuracy with sufficient training data

Google inception model (V3 model)

a pretrained model with over 1000 different classes,

Transfer learning

Fine-tune the original model by replacing the final layer with small sample.

Verify

Training set with crowding source of information
Achieve 87% accuracy

Investors' sentiment index

Variables Construction

photope

$$PhotoPes_t = \frac{\sum_i (Neg_{it})}{n_t}$$

Neg ---whether photo i on day t is predicted to have negative sentiment.
Nt-- the number of photos on date t .

TextNeg--pessimism score for each article i on day t
Nt--the number of articles in date t

$$TextPes_t = \frac{\sum_i (TextNeg_{it})}{n_t},$$

textpe

Market Return

The daily returns on the CRSP value-weighted (VWRETD) index, the S&P 500 Index (SPX), the SPDR S&P 500 ETF (SPY), the Dow Jones Industrial Average Index (INDU), and the SPDR Dow Jones Industrial Average ETF (DIA).

Summary statistics

Panel A: Summary statistics of sentiment variables

Variable	*N*	*Mean*	*Median*	*P25*	*P75*	*Std dev*
PhotoPes	3048	0.228	0.222	0.180	0.270	0.077
TextPes	3048	0.686	0.681	0.646	0.722	0.056

Panel B: Summary statistics of market returns

R_t(%)	*N*	*Mean*	*P50*	*P25*	*P75*	*Std dev*
VWRETD	3048	0.045	0.081	-0.391	0.586	1.332
SPX	3048	0.042	0.070	-0.380	0.570	1.335
SPY	3048	0.049	0.070	-0.370	0.580	1.327
INDU	3048	0.039	0.060	-0.390	0.550	1.283
DIA	3048	0.047	0.070	-0.370	0.550	1.296

Panel C: Correlations between sentiment variables

	PhotoPes
TextPes	0.079***
	<0.01

PART THREE

Empirical Results

Behavioral Finance

1、Inverstors are irrational. Biases such as extrapolation and overconfidence exsist
2、limits to arbitrage prevent rational investors from correcting price instantly.
Prediction : market return reversal

To learn this circumstance, we must study the pessimism embedded in photos

The impact of PhotoPes on market returns

$$R_t = \beta_1 L5(PhotoPes_t) + \beta_2 L5(R_t) + \beta_3 L5\left(R_t^2\right) + \beta_4 X_t + \varepsilon_t$$

Rt denotes daily log returns on the VWRETD index, the S&P 500 Index (SPX), the SPDR S&P 500 ETF (SPY), the Dow Jones Industrial Average Index (INDU), and the SPDR Dow Jones Industrial Average ETF (DIA).

L5 transforms a variable into a row vector consisting of five lags of that variable; and Xt is a set of exogenous variables that includes an intercept, day-of-the-week indicators (except for Monday), and an indicator variable for whether time t is in a recession period

The impact of PhotoPes on market returns

	Panel A: PhotoPes									
	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat
$PhotoPes_{t-1}$	-0.042*	-1.837	-0.041*	-1.803	-0.040*	-1.787	-0.046**	-2.182	-0.047**	-2.183
$PhotoPes_{t-2}$	0.055**	2.004	0.051*	1.886	0.046*	1.726	0.043*	1.687	0.038	1.502
$PhotoPes_{t-3}$	-0.033	-1.324	-0.030	-1.213	-0.030	-1.294	-0.024	-1.053	-0.025	-1.142
$PhotoPes_{t-4}$	0.030	1.299	0.024	1.047	0.026	1.143	0.030	1.387	0.033	1.487
$PhotoPes_{t-5}$	0.057**	2.137	0.059**	2.228	0.056**	2.119	0.057**	2.193	0.054**	2.103
Sum t-1 to t-5	0.067		0.063		0.058		0.060		0.053	
Sum t-2 to t-5	0.109		0.104		0.098		0.106		0.100	
	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	2.272	0.132	2.081	0.149	1.700	0.192	1.979	0.160	1.644	0.200
$\chi^2(1)$[Sum t-2 to t-5 = 0]	6.615**	0.010	6.200**	0.013	5.466**	0.019	6.973***	0.008	6.257**	0.012
Adj. *R*-squared	0.033		0.038		0.029		0.042		0.040	
N	3044		3044		3044		3044		3044	

- $PhotoPest_{t-1}$ is negatively related to market returns in all specifications
- the average impact of a one standard deviation shift in PhotoPes on the next day's VWRETD is 4.2 bps
- the reversal is concentrated on lags two and five
- the initial decline in returns is followed by a complete reversal.

PhotoPes and sentiment embedded in text

Complement or Substitute?

$$R_t = \beta_1 L5(PhotoPes_t) + \beta_2 L5(TextPes_t) + \beta_3 (PhotoPes \times TextPes)_{t-1} + \beta_4 L5(R_t) + \beta_5 L5(R_t^2) + \beta_6 X_t + \varepsilon_t,$$

- To support the substitutive perspective, we expect to find that the coefficient for the interaction term $(\text{PhotoPes} \times \text{TextPes})_{t-1}$ is positive.
- In all five specifications, the coefficients for $(\text{PhotoPes} \times \text{TextPes})_{t-1}$ are positive and significant at the 10% level, supporting the substitutive hypothesis.

$$R_t = \beta_1 L5(TextPes_t) + \beta_2 L5(R_t) + \beta_3 L5(R_t^2) + \beta_4 X_t + \varepsilon_t,$$

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
$TextPes_{t-1}$	-0.071*	-1.663	-0.083*	-1.904	-0.087**	-1.977	-0.085**	-2.002	-0.086**	-1.998
$TextPes_{t-2}$	-0.056	-1.466	-0.061	-1.590	-0.065*	-1.661	-0.062*	-1.686	-0.061	-1.628
$TextPes_{t-3}$	-0.007	-0.150	0.001	0.018	-0.004	-0.078	0.005	0.108	-0.003	-0.069
$TextPes_{t-4}$	0.021	0.527	0.017	0.436	0.021	0.540	0.027	0.709	0.029	0.756
$TextPes_{t-5}$	0.107**	2.280	0.116**	2.448	0.124***	2.612	0.110**	2.408	0.116**	2.521
Sum t-1 to t-5	-0.006		-0.010		-0.011		-0.005		-0.005	
Sum t-2 to t-5	0.065		0.073		0.076		0.080		0.081	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	0.054	0.816	0.112	0.738	0.120	0.729	0.041	0.840	0.030	0.862
$\chi^2(1)$[Sum t-2 to t-5 = 0]	1.697	0.193	2.171	0.141	2.454	0.117	2.812*	0.094	2.931*	0.087
Adj. R-squared	0.028		0.035		0.028		0.038		0.039	
N	3044		3044		3044		3044		3044	

- The average impact of a one standard deviation shift in TextPes on the next day's INDU is 8.5 bps.
- Markets take more time to reflect information in news text compared to photos

Attention and Photo

- The relation between the pessimism embedded in news and market returns varies depending on the presence of salient news photos.

$$R_t = (E_t)[\beta_1 L5(PhotoPes_t) + \beta_2 L5(TextPes_t) + \beta_3 (PhotoPes \times TextPes)_{t-1} + \beta_4 L5(R_t) + \beta_5 L5(R_t^2)]$$
$$+ (1 - E_t)[\gamma_1 L5(PhotoPes_t) + \gamma_2 L5(TextPes_t) + \gamma_3 (PhotoPes \times TextPes)_{t-1} + \gamma_4 L5(R_t) + \gamma_5 L5(R_t^2)] + \beta_6 X_t + \varepsilon_t$$

- Et is an indicator variable that takes a value of one if day t is in the top or bottom decile of PhotoPes.

	(1)				(2)				(3)			
	$VWRETD_t$				SPX_t				SPY_t			
	E_t=Salient photo period				E_t= Salient photo period				E_t= Salient photo period			
Variables	β	t-stat	γ	t-stat	β	t-stat	γ	t-stat	β	t-stat	γ	t-stat
$PhotoPes_{t-1}$	-0.070**	-2.479	-0.015	-0.332	-0.064**	-2.295	-0.016	-0.365	-0.063**	-2.260	-0.015	-0.342
$TextPes_{t-1}$	0.047	0.900	-0.070*	-1.883	0.031	0.606	-0.081**	-2.220	0.030	0.585	-0.080**	-2.137
$(PhotoPes \times TextPes)_{t-1}$	0.034	1.524	0.070	1.450	0.029	1.312	0.065	1.362	0.030	1.403	0.060	1.270
$PhotoPes_{t-2}$	0.100***	3.282	-0.034	-0.813	0.099***	3.316	-0.041	-0.978	0.094***	3.173	-0.044	-1.079
$PhotoPes_{t-3}$	-0.020	-0.659	-0.017	-0.411	-0.017	-0.565	-0.020	-0.498	-0.015	-0.526	-0.019	-0.489
$PhotoPes_{t-4}$	0.046*	1.710	0.009	0.216	0.042	1.571	-0.001	-0.033	0.042	1.632	0.003	0.066
$PhotoPes_{t-5}$	0.047	1.469	-0.009	-0.225	0.049	1.538	-0.004	-0.098	0.045	1.383	-0.006	-0.166
$TextPes_{t-2}$	0.044	0.684	-0.066*	-1.795	0.043	0.664	-0.071*	-1.912	0.051	0.761	-0.072*	-1.931
$TextPes_{t-3}$	-0.128**	-2.373	0.013	0.316	-0.113**	-2.174	0.022	0.538	-0.114**	-2.146	0.020	0.486
$TextPes_{t-4}$	-0.079	-1.412	0.002	0.057	-0.083	-1.488	-0.001	-0.018	-0.079	-1.488	-0.002	-0.050
$TextPes_{t-5}$	0.182***	3.133	0.059	1.520	0.186***	3.255	0.064	1.638	0.184***	3.135	0.065*	1.694
Sum t-1 to t-5 *PhotoPes*	0.103		-0.066		0.109		-0.082		0.103		-0.081	
Sum t-2 to t-5 *PhotoPes*	0.173		-0.051		0.173		-0.066		0.166		-0.066	
Sum t-1 to t-5 *TextPes*	0.066		-0.062		0.064		-0.067		0.072		-0.069	
Sum t-2 to t-5 *TextPes*	0.019		0.008		0.033		0.014		0.042		0.011	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 *PhotoPes*=0]	2.905*	0.088	0.602	0.438	3.224*	0.073	0.958	0.328	2.851*	0.091	0.941	0.332
$\chi^2(1)$[Sum t-2 to t-5 *PhotoPes*=0]	9.777***	0.002	0.480	0.489	9.696***	0.002	0.823	0.364	8.943***	0.003	0.833	0.362
$\chi^2(1)$[Sum t-1 to t-5 *TextPes*=0]	0.437	0.509	1.842	0.175	0.405	0.525	2.247	0.134	0.525	0.469	2.443	0.118
$\chi^2(1)$[Sum t-2 to t-5 *TextPes*=0]	0.041	0.840	0.028	0.867	0.115	0.734	0.079	0.778	0.187	0.666	0.052	0.820
Adj. R-squared	0.075				0.086				0.071			
N	3044				3044				3044			

- During days when photos are salient, PhotoPes dominates and TextPes is not significant. In contrast, during days when photos are not salient, TextPes dominates and PhotoPes is not statistically significant.

Which information is more effectively transmitted by photos?

$$R_t = (F_t)[\beta_1 L5(PhotoPes_t) + \beta_2 L5(TextPes_t) + \beta_3 (PhotoPes \times TextPe)s_{t-1} + \beta_4 L5(R_t) + \beta_5 L5(R_t^2)] + (1 - F_t)[\gamma_1 L5(PhotoPes_t) + \gamma_2 L5(TextPes_t) + \gamma_3 (PhotoPes \times TextPes)_{t-1} + \gamma_4 L5(R_t) + \gamma_5 L5(R_t^2)] + \beta_6 X_t + \varepsilon_t,$$

- F_t is an indicator variable that takes a value of one if day t has an above-median fear score

- The coefficient for the pessimism embedded in photos is roughly 2.8 times larger during periods of elevated fear compared to periods of little fear, while the coefficient for TextPes is similar during both periods

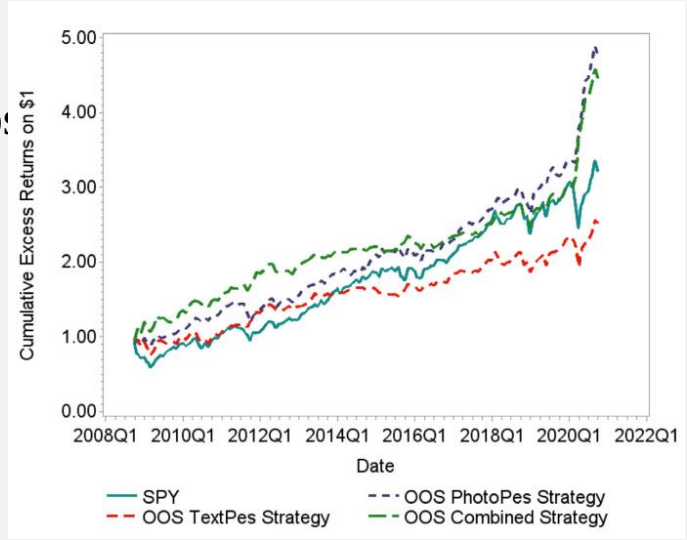
traumatic

Application

Invest either in SPY or risk- free asset

- based on the pessimism embedded in news photos
- based on pessimism embedded in text
- involves pessimism from both text and photos

- The reversal pattern for PhotoPes starts on day t + 2， however, the reversal pattern for TextPes does not turn positive until day t + 4.



	Strategy One	Strategy Two	Strategy Three
Times	1992	1891	1221
Sharpe Ratio	0.88	0.53	0.6
Excess Return	5.8	3.7	4.7

Validation of PhotoPes

Limits to arbitrage

- The average impact of a one standard deviation shift in PhotoPes on the next day's return on the highest and lowest volatility-sorted portfolio is 7.1 and 3.5 bps
- PhotoPes has a stronger impact on small companies compared to large companies.

The impact of PhotoPes on trading volume

$$V_t = \beta L5(V_t) + \gamma X_t + \varepsilon_t$$

→ Remove the time trend

$$\tilde{V}_t = \beta_1 L5(PhotoPes_t) + \beta_2 L5(|PhotoPes_t|) + \beta_3 L5(R_t) + \beta_4 L5(R_t^2) + \varepsilon_t$$

- One standard deviation increase in PhotoPes is associated with a moving future trading volume of 0.080 standard deviations

PART FOUR

Conclusion

THANKS