定性情報と企業価値

三輪 宏太郎

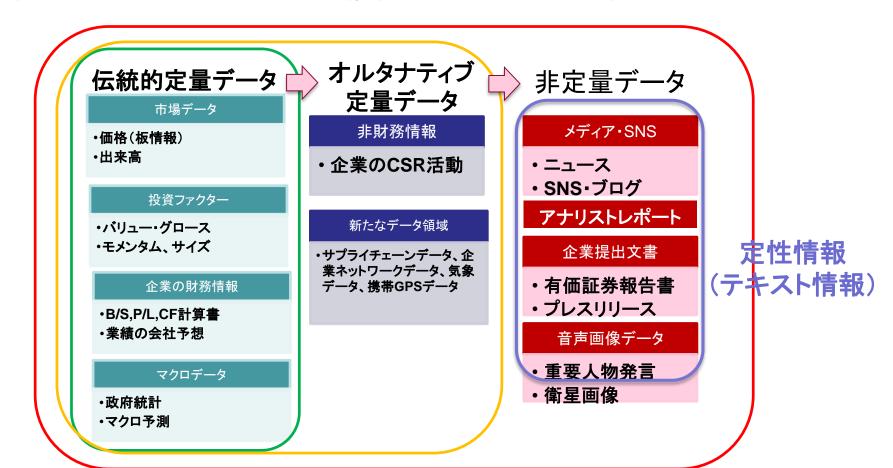
九州大学 経済学研究院

Dec 5,2020



情報と企業価値

✓ 株式価格がどのような情報を織り込むかは、資産価格理論の長年の 関心事だが、(分析される)情報の範囲は近年急拡大している



定性情報の分析プロセス

✓ 定性情報はそのままでは、定量情報である企業価値(株価)との関連性を分析できない。

(定性情報からの特徴量抽出)

- ✓ 定性情報のクレンジング
- ✓ 形態素解析(日本語の場合)
- ✓ 定性情報のベクトル化 (Bag of Words)
- ✓ 特徴量抽出(TF-IDF, Word embedding)

(目的に合わせた情報抽出)

- ✓ 極性(ポジネガ)
- ✓ 不確実性
- ✓ 可読性

Etc...



極性の計測

辞書的手法

✓ 既存の金融辞書使用(Henry, 2008; Loughran and McDonald, 2011)※独自作成することもある

極性=(ポジティブ単語数-ネガティブ単語数)/(ポジティブ単語数+ネガティブ単語数)

- ✓構文解析をして、係り受けの考慮、ネガティブの否定など に対処することもある
- ✓因果性の明確さなどからファイナンス研究ではよく使われる

極性の計測

機械学習系

- ✓ 学習あり
 - SVM
 - ✓ 回帰による分類。線形入力素子を利用して2クラス(ポジティブ・ネガティブ)のパターン識別機を構成
 - サイーブベイズ
 - ✓ 特徴間に強い独立性を仮定したうえで、ベイズの定理を使って分類
- ✓ 学習なし

LDA:文書中に出現している単語の種類と出現頻度に基づいて、トピックを教師なし学習で推定(分類)する。

深層学習モデル系

1. 有価証券報告書・決算発表コメント

以下の情報は将来の業績に対する予測力を持つ

有価証券報告書のMD&A (Feldman et al. 2010)

✓ MD&A(経営者による財政状態および経営成績の分析)の文書センチメント

Model	1	2	3	4	5	6	
Intercept	-0.001	-0.001	-0.001	-0.002	-0.002	-0.002	
Significance	0.001	0.001	0.001	0.036	0.038	0.035	
Accruals	0.0002	0.0003	0.0002	-0.0005	-0.0004	-0.0004	
Significance	0.707	0.613	0.699	0.411	0.470	0.423	
SUE				0.0093	0.0094	0.0094	
Significance				0.001	0.001	0.001	
(Pos-Neg)	0.0022			0.0014			
Significance	0.001			0.016			
Pos		0.0016			0.0009		
Significance		0.002			0072		
Neg			-0.0016			-0.0011	
Significance			0.002			0.017	Ш
Average R-							
Square	0.16%	0.13%	0.14%	1.5%	1.5%	1.5%	



2. ニュース (一般メディア)

✓ DJNSの悲観表現が、個別銘柄(SP500構成銘柄)の株価に対して影響を与えている。(Tetlock et al. 2008)

		$Return_{+1,+1}$			$FFCAR_{+1,+1}$	
Stories Included	DJNS	WSJ	All	DJNS	WSJ	All
neg	-0.0277 (-3.67)	-0.0105 (-1.24)	-0.0221 (-3.72)	-0.0320 (-4.83)	-0.0102 (-1.37)	-0.0253 (-4.88)
$FFCAR_{0,0}$	0.0285 (5.28)	0.0229 (2.92)	0.0246 (5.43)	0.0259 (5.00)	0.0224 (2.94)	0.0226 (5.19)
$FFCAR_{-1,-1}$	-0.0272 (-3.63)	-0.0154 (-2.17)	-0.0222 (-4.21)	-0.0254 (-3.86)	-0.0106 (-1.68)	-0.0190 (-4.13)

3. ソーシャルメディア

- StockTwitのネガティブコメントに、株価が反応(ニュース記事のセ ンチメントをコントロール後も)(Chen et al. 2014)
- 決算直前のTwitterコメントの極性が、決算発表時の株価反応を 予測する (Bartov et al. 2018)

	(1)	(2)	(3)
$NegSA_{i,t}$	-0.379 (-2.24)	-0.332 (-2.03)	-0.320 (-1.98)
NegSA-Comment _{i,t}		-0.194 (-3.44)	-0.196 (-3.55)
$I(SA-Comment_{i,t})$		0.001 (0.25)	0.001 (0.17)
$NegDJNS_{i,t}$			-0.254 (-1.44)
$I(DJNS_{i,t})$			0.009 (1.33)
77 1	0.003	0.003	0.003

Chen et al. (2014) Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media

4. アナリストレポート(Huang et al. 2014)

定量情報(銘柄推奨度、業績予想)の影響を控除しても、テキストの極性は株価に強い影響を与えている。

	BUY		HOLD		SELL	
OPN	CAR	# of Obs.	CAR	# of Obs.	CAR	# of Obs.
5 (Top Quintile)	0.92%	42,847	0.89%	16,268	1.01%	1,631
4	0.45%	38,300	0.35%	20,458	0.26%	2,610
3	0.22%	34,320	0.01%	23,264	0.07%	3,806
2	-0.08%	29,153	-0.30%	25,896	-0.54%	4,940
1 (Bottom Quintile)	-0.85%	22,383	-1.24%	31,614	-1.49%	6,902
All OPN Ouintiles	0.26%		-0.21%		-0.52%	
Top - Bottom	1.77%***		2.13%***		2.50%***	

テキスト研究の方向性

テキストデータは、非定量データの中では、比較的多くの研究がなされている

- 取得可能なテキストデータは膨大であり、拡大し続けているため、未検証のテキストデータは膨大にある
- 情報科学, 言語学を中心に分析手法の精緻化が為されており、既に十分分析されているテキストデータでも、 新たな知見が得られる可能性がある

⇒実証ファイナンス分野の中では、ポテンシャルの高い研究 分野

テキスト研究の方向性

- i. インフォーマルなテキスト
- ii. 新たなテキスト分析手法
- iii. 意見の多様性
- iv. 日本語テキスト

1インフォーマルなテキスト

- ✓有価証券報告書や金融当局発行の文書などのフォーマルが文 章の分析が中心
- ✓SNSなどのインフォーマルな文書・文字起こしなどを対象とした 分析は遅れている
- ✓インフォーマルな方が、推敲を重ねたフォーマルな文書より本音 が出やすいという側面もある



1インフォーマルなテキスト

Executives' Honest Opinions

By Kotaro Miwa

投資家ミーティングのQ&Aセッションの経営陣の発言(A)のトーンに注目

- トーン(過去のミーティングとの変化)が、株価に対してインパクトを与える
- ポジティブ(ネガティブ)なトーンほど、アナリストがその後業績 予想を上方(下方)修正する
- フォーマルなコメントである、MD&A(ミーティング冒頭の経営陣からの説明)のトーンは情報価値がない

- ✓手法に関する頑健性を重視する為、伝統的手法(e.g. 辞書的手法)などが好まれてきた
- ✓学習ありの機械学習+深層学習系は、適用例が少ない
- ✓これら手法の活用により、新たなテキスト情報と資産価値の間に関係が見えてくることもある

辞書的手法 機械学習系

深層学習系

学習あり学習なし



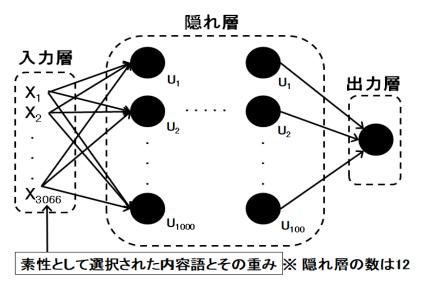
アナリストレポートのトーンの情報価値

(平松賢士 三輪宏太郎 酒井浩之 坂地泰紀)

- ✓ 深層学習手法を駆使し、日本株を対象としたレポートのトーンの情報価値を検証。
- ✓ 株価はレポートのトーンに強く反応しているのみならず、遅延反応までも発生。
- ⇒レポートのトーンは有益な情報を保持しているのみならず、株価はこの情報を即座には織り込んでいない事が示唆された。

(モデル)

- ✓ 銘柄推奨度が変化したアナリストレポートから、特徴語を抽出
- ✓ 銘柄推奨度が上がったアナリストレポートの場合は1、銘柄推奨度が下がったアナリストレポートの場合は0とした学習データを使い学習させる



(既存手法との比較)

- ✓ 作成された文書の特性の株価インパクト(及び事後リターン)を観測
- ✓ 先行研究のナイーブベイズによる手法を結果を比較

(深層学習)	全期間	(ナイーブベイズ)	全期間
TONE	0.0179 ***	TONE	0.0031 ***
	(24.00)		(8.41)
銘柄推奨度	0.0025 ***	銘柄推奨度	0.0023 ***
	(18.59)		(16.69)
銘柄推奨度変化	0.0152 ***	銘柄推奨度変化	0.0189 ***
	(22.92)		(27.79)
EPS 予想変化率	0.0078 ***	EPS予想変化率	0.0087 ***
	(8.77)		(9.40)
目標株価変化率	0.0270 ***	目標株価変化率	0.0452 ***
	(12.66)		(21.82)

- ✓ほとんどの研究がコンセンサス(平均的な極性)を分析
- ✓Informal系のテキストは多種多様な意見を内包してい るが、外れた意見が分析対象に入ることはほぼない



九州大学

三輪 宏太郎

✓ コンセンサス

17

ソーシャルメディアなどのインフォーマルなコメントは、

- ┌─ 大多数を占める、情報価値のないコメント
- ┥ − 少数の情報価値を持つコメント

コンセンサス(平均トーン)は、大多数の無価値なツイートの影響を受ける

SNSの価値は、少数の情報価値を持つコメントにあるとしたら、これらを抽出して情報価値を分析する必要がある

- ロ尖った意見(Extreme Opinions)に注目
- SNSの尖った意見は、会社や製品の弱点や強みを特定する上で有用な情報をもたらす (Almatarneh and Gamallo, 2018)

Extreme Opinions on Social Media

By Kotaro Miwa

個別銘柄に対するツイートの中で、最も強気なツイートと弱気なツイート各々のクロスセクショナルな分布に注目

- 1. 極端なツイートのトーンは株価への予測力がある
- 2. 極端なツイートのトーンは、アナリストの業績予想・目標 株価修正への予測力がある(ファンダメンタル情報)
- 3. 平均的なトーン(コンセンサスは)、一時的なミスプライシングを発生させるが、ファンダメンタル情報などは内包しない(=投資家の非合理的なセンチメント)

各銘柄、日次頻度で計算されるBloomberg's firm-level Twitter sentiment を使用

特に2分間隔で計算されるトーンの1日の最大値と最小値 に注目.

$$Twitter_{i,t}^{Extreme} = \frac{Twitter_{i,t}^{Highest} + Twitter_{i,t}^{Lowest}}{2}$$

Fama-Macbeth regression :
$$Ret_{i,t} = \alpha + \sum_{k=1}^{5} \beta_{1,k} Twitter_{i,t-k}^{Extreme} + \sum_{k=1}^{5} \beta_{2,k} Twitter_{i,t-k}^{Mean} + (Controls) + \varepsilon_{i,t}$$

- ✓ $Twitter_{i,t-1}^{Extreme}$ (Extreme_Opinion) と翌日リターンの関係は有意に正の関係(その後修正されない)
- ✓ $Twitter_{i,t-1}^{Mean}$ (Consensus) と翌日リターンの関係は有意に正の関係だが、その後大きく修正 (一時的)

	Raw		Risk Adjusted		
$Twitter_{i,t-1}^{Extreme}$	0.0019 ***	(4.31)	0.0018 ***	(3.93)	
$Twitter^{Extreme}_{i,t-2}$	-0.0003	(0.64)	-0.0002	(0.43)	
$Twitter_{i,t-3}^{Extreme}$	-0.0004	(0.95)	-0.0008	(1.72)	
$Twitter^{Extreme}_{i,t-4}$	-0.0004	(1.25)	-0.0006	(1.32)	
$Twitter_{i,t-5}^{Extreme}$	0.0002	(0.45)	-0.0004	(0.98)	
$Twitter_{i,t-1}^{Mean}$	0.0010 ***	(4.87)	0.0011 ***	(4.67)	
$Twitter^{Mean}_{i,t-2}$	-0.0008 ***	(3.61)	-0.0006 **	(2.28)	
$Twitter^{Mean}_{i,t-3}$	0.0000	(0.13)	0.0002	(0.60)	
$Twitter_{i.t-4}^{Mean}$	-0.0003	(1.21)	-0.0003	(1.13)	
$Twitter^{Mean}_{i.t-5}$	0.0001	(0.42)	0.0002	(0.79)	

21

- ✓多くの研究が英文テキストの分析で、日本語テキスト の分析は少ない
- ✓情報学系、言語学系は一定量あり、日本語特有の分析手法が整備されている
- ✓単なるEvidence from Japanだと、ジャーナルのランクは妥協が必要だが、日本語テキストの意義を十分示せれば、ある程度のランクのジャーナルも狙える(テキスト系論文に限ったことではないが。。。)

Language Barriers in Analysts' Reports

By Kotaro Miwa

- ✓ 日本株の日本語レポートと米株の英語レポートのテキスト極性 (トーン)に対する株価の反応の違いを比較
 - 1. 米国株のレポートのトーンに対しては株価は即座に織り込む
 - 2. 日本株のレポートのトーンに対しては、株価の遅延反応がみられる
 - 3. 遅延反応は、英訳が伴っている日本株レポートに対しては見られない(即座に織り込む)
 - ⇒言葉の壁が、テキスト情報に対す遅延反応を引き起こす

✓株価の反応の違い

- 米国株のレポートのトーンに対しては株価は即座に織り込む
- 日本株のレポートのトーンに対しては、株価の遅延反応がみられる

CAR[0,1]

0.0495 ***

0.0261 ***

0.0017 ***

0.8674 **

0.2464 ***

0.0024 ***

-0.0306 ***

-0.0013 **

Yes

0.0010

8.97%

64999

0.0038 ***

Dependent variables

(5.82)

(6.17)

(5.59)

(2.45)

(11.03)

(7.77)

(3.58)

(1.96)

(3.20)

(0.31)

			_
iables			•
CA	R[2,5	0]	
-0.0075		(0.39)	•
-0.0013		(0.12)	
0.0011		(0.85)	
0.7868	**	(2.02)	
0.0139		(0.74)	
0.0020	***	(2.82)	
0.0029		(0.13)	
-0.0109	***	(3.69)	
0.0121	**	(2.28)	_
			•
Yes			
0.0530	***	(4.06)	
1.16%			
64999			

(日本株式レポート)

			Dependent	Variables		
	C	AR[0,	1]	CAR[2,50]		
TONE _N	0.0945	***	(10.22)	0.0481	***	(2.71)
TONE _P	0.0155	***	(4.47)	-0.0068		(0.84)
REC	0.0038	***	(9.42)	-0.0015		(0.97)
EPS_REV	0.0456	***	(3.02)	-0.0348		(1.57)
TP_REV	0.0617	***	(16.61)	-0.0038		(0.45)
SUE	0.0018	***	(4.21)	0.0034	***	(3.42)
PCAR	-0.0228	***	(3.55)	-0.0224		(1.21)
MV	-0.0016	***	(3.31)	-0.0106	***	(4.20)
BM	0.0033	***	(3.58)	0.0041		(1.22)
Controls for						_
Industry and Year	Yes			Yes		
Effects						
Intercept	0.0154	***	(3.30)	0.0467		(1.83)
Adjusted R2	3.40%			0.62%		
N	36943			36943		

24

TONE_N

TONE_P

EPS_REV

TP REV

REC

SUE

MV

BM

Controls for Industry and

Year Effects Intercept

Adjusted R2

PCAR

宏太郎

• 遅延反応は、英訳が伴っている(日本語レポート発行日24時間以内に英文レポートが発行された)日本株レポートに対しては見られない (即座に織り込む)

		Dependent Variables					
		C	AR[0,1	.]	CA	R[2,50]
英訳な	TONE_J _N	0.0981	***	(9.49)	0.0524	***	(2.71)
英訳付	さ TONE_E _N	0.0807	***	(4.30)	0.0313		(0.83)
	$TONE_{P}$	0.0155	***	(4.47)	-0.0068		(0.84)
	REC	0.0038	***	(9.42)	-0.0015		(0.97)
	EPS_REV	0.0456	***	(3.02)	-0.0348		(1.57)
	TP_REV	0.0617	***	(16.60)	-0.0039		(0.45)
	SUE	0.0018	***	(4.21)	0.0034	***	(3.43)
	PCAR	-0.0228	***	(3.55)	-0.0224		(1.20)
	MV	-0.0016	***	(3.32)	-0.0106	***	(4.20)
	BM	0.0033	***	(3.58)	0.0041		(1.22)
	Controls for						
	Industry and Year				Yes		
	Effects						
	Intercept	0.0154	***	(3.33)	0.0468		(1.83)
	Adjusted R2	3.40%			0.62%		
	N	36943			36943		



三輪 宏太郎

25