Detecting and Explaining Causes From Text for a Time Series Event

by Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, and Eduard Hovy

読む人：宮澤 彬

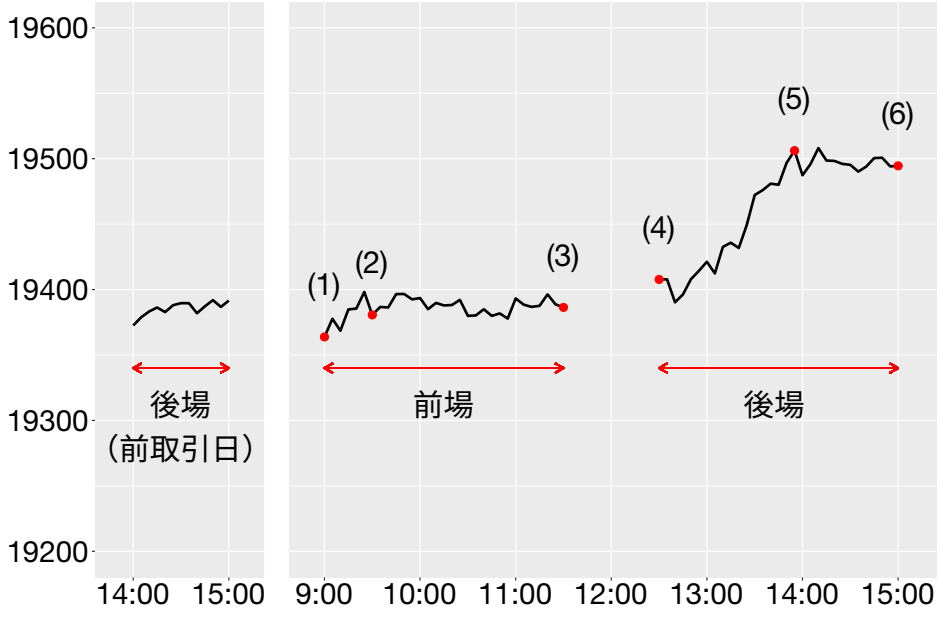
September 17, 2017

読もうと思った理由

メタファーの論文があまりなかった😐.

産総研 PLU グループで時系列データの分析をやっており，それに関連してそうだったから.

cf. “Learning to Generate Market Comments from Stock Prices” by Murakami et al. (2017).



概況テキスト
(1) 日経平均、続落で始まる
(2) 日経平均、上げに転じる
(3) 日経平均、続落 前引けは 5 円安の 19,386 円
(4) 日経平均、午後は上昇で始まる
(5) 日経平均、上げ幅 100 円超える
(6) 日経平均、反発 大引けは 102 円高の 19,494 円

概要 I

株価の変動の**要因**を説明するテキストを生成する研究.

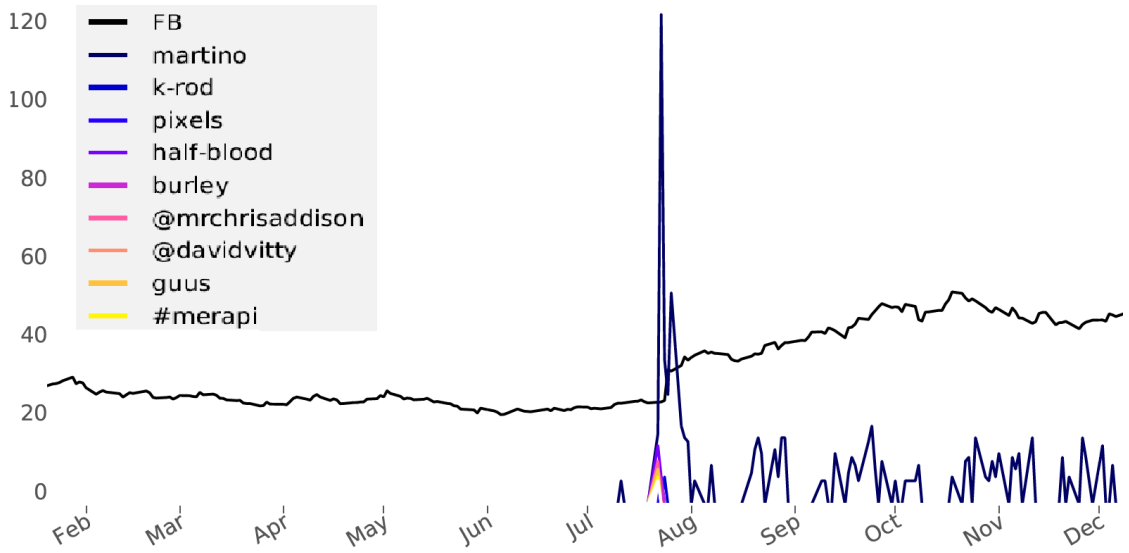


Figure: Facebook の株価と原因となる特徴量.

概要 II

より詳しくいうと株価の変動に対する**因果の説明** (causal explanation) を生成する.

party $\overset{cut}{\longmapsto}$ budget_cuts $\overset{lower}{\longmapsto}$ budget_bill
$\overset{decreas}{\longmapsto}$ republicans $\overset{caus}{\longmapsto}$ obama $\overset{leadto}{\longmapsto}$ facebook_polls
$\overset{caus}{\longmapsto}$ facebook's stock $\downarrow$

Figure: 因果の説明の例.

ここでいう「因果」は必ずしも「原因と結果」ではない.

- Granger 因果性
- *agent* vs. *patient* のような意味合いでの「因果」

テキスト特徴量を利用した因果性検出

与えられた時系列 y に対して，**因果性** (causality) が最大となるようなテキストから抽出した特徴量を探す．

$$\arg\max_{F \subseteq X} C\left(y, \Phi(F, y)\right) \tag{1}$$

ここで記号の意味は以下の通り．

- C: 因果性
- X: テキストから抽出した特徴量全体の集合
- $F := \{f_i\}_i \subseteq X$
- Φ: $\{f_i\}_i$ の線型結合を作る関数（対象の時系列 y に依存）

Granger 因果性 I

Granger 因果性 (Granger, 1988) という時系列解析の割と古典的な手法を用いる[1].

背景にある考えは「ある時系列 x が対象の時系列 y の原因となっているならば, x の情報を使ったほうが y_t をより正確に予測できる」というもの.

[1] Hamilton (1994)（定番の教科書）にも載っている.

Granger 因果性 II

時系列 y が以下の外生変数付きの自己回帰モデルに従うとする.

$$y_t = \alpha y_{t-\ell} + \beta(f_{1,t-\ell} + \cdots + f_{k,t-\ell}) + \varepsilon_t \tag{2}$$

もし $\beta_k = 0$ ならば f_k は予測に役立っていない.

$\Delta(\beta_{y,f,\ell})$ をラグ ℓ で特徴量 f をいれたときの β の分散の増減とする. このとき

$$C(y, F) = \sum_{k,\ell} \Delta(\beta_{y,f,\ell})$$

を F について最大化する.

テキスト特徴量

3 つの特徴量を利用する.

- F_{words}
  日毎の N グラムの出現頻度の列. 定常的なものは除外.
  (例) $x^{\text{Michael\_Jordan}} = (12, 51, \ldots)$
- F_{topic}
  LDA により得られた各トピックの頻出語の出現頻度の列.
  (例) $x^{\text{healthcare}}, \text{healthcare} = \{\text{insurance}, \text{obamacare}, \ldots\}$
- F_{senti}
  ポジティブな語とネガティブな語の割合の列.

因果関係のグラフ（CGraph）の構築 I

原因の時系列 x と結果の時系列 y が与えられたとき，それらをつなぐ因果の経路を求めたい．

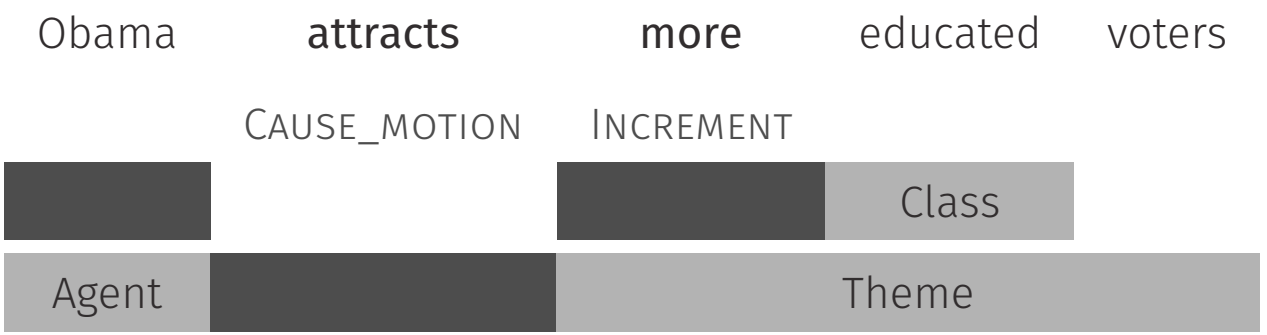
$$(x \mapsto e_1), (e_1 \mapsto e_2), \ldots, (e_t \mapsto y) \tag{3}$$

原因と結果をノード，関係をエッジとみなした *cause-effect graph* (CGraph) を作る．

（例）obama $\overset{\textit{Cause\_motion}}{\longmapsto}$ more educated voters

因果関係のグラフ（CGraph）の構築 II

1. SEMAFOR で New York Times などの文を解析し FrameNet のラベルを付与する.



2. Cause_change や Cause_motion などの関係を抽出し，それぞれの始点と終点で原因・関係・結果の組を作る.
（例）(obama, Cause_motion, more educated voters)
3. 上の手順で集めた組を文字列マッチングで繋げることでグラフを構築する．このとき Freebase のノードとリンクさせ，表層は異なるが意味としては近いものがまとまるようにする.

因果関係のグラフ（CGRAPH）を使った因果推論

CGRAPH は巨大なので効率的な探索の方法を提案している.

1: $\mathcal{S} \leftarrow y, d = 0$
2: **while** $(\mathcal{S} = \emptyset)$ or $(d > D_{\max})$ **do**
3: $\quad \left\{e^1_{-d}, \ldots, e^k_{-d}\right\} \leftarrow \mathrm{BreadthFirstSeach}_{\mathrm{back}}(\mathcal{S})$
4: $\quad$ **for** j in $\{1, \ldots, k\}$ **do**
5: $\quad\quad$ **if** $\sum_\ell \mathbf{C}\left(y, e^j_{-d}, \ell\right) < \varepsilon$ **then**
6: $\quad\quad\quad \mathcal{S} \leftarrow e^j_{-d}$

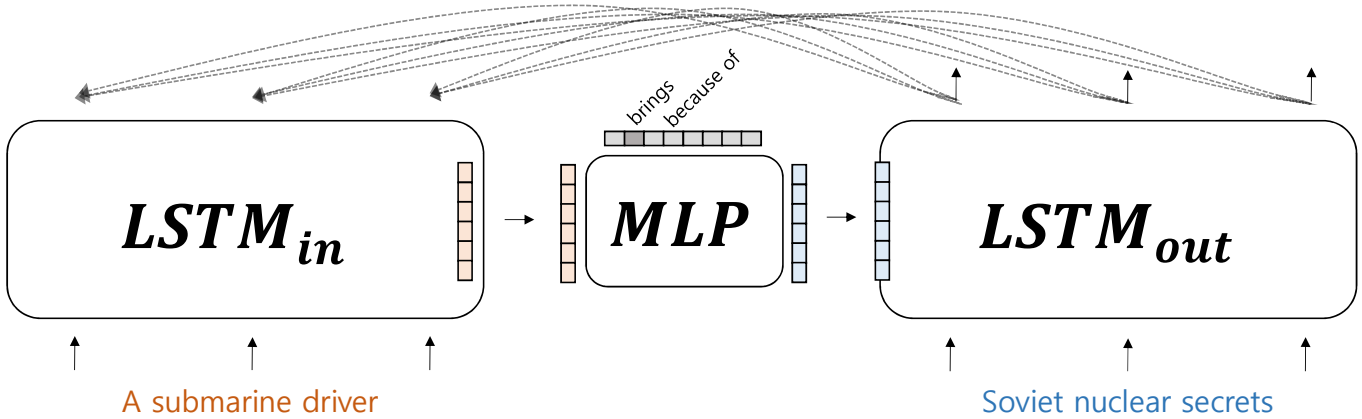
Granger 因果性が閾値以上になるまで結果の方から遡っている.

ニューラルネットワークを使った因果推論

前述の手法ではグラフに記述したい事象やそれらの間の関係が存在しないことが多い.

これを解決するため encoder-decoder を使う.

- Encoder への入力：原因
- Decoder への入力 /Decoder からの出力：結果



実験

テキストデータ（原因）

- 2008 年から 2013 年までのツイート (1 mil. tweets/day)
- 2010 年から 2013 年までのブログとニュース (Google's news API)

数値データ（結果）

- 2001 年から執筆当時までの株価データ（NASDAQ と NYSE の 6,200 銘柄）
- 2012 年の大統領選挙の投票データ (USA Today, Huffington Post, Reuters, etc.)

実験 1: テキストの特徴量と時系列データの関係性の分析

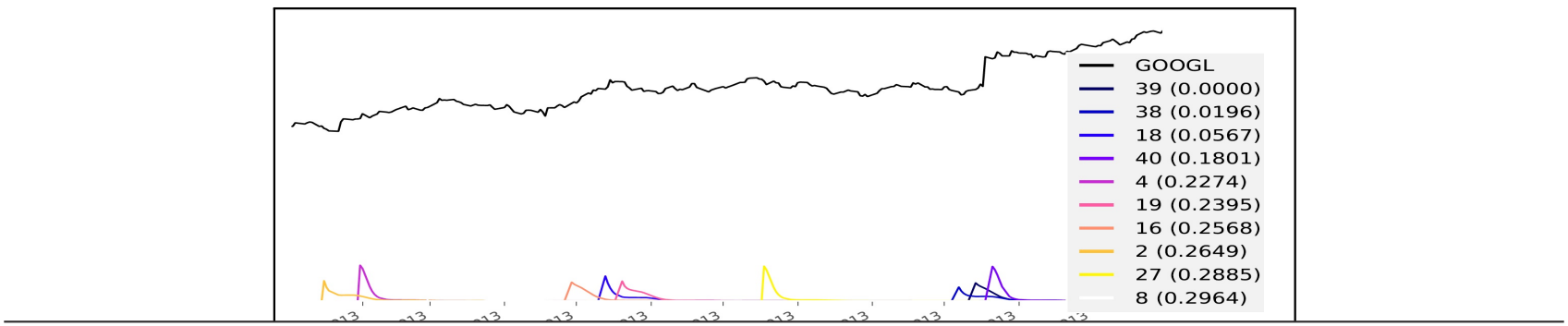


Figure: Google の株価と原因となる特徴量の比較.

実験 2: テキスト特徴量による数値予測

株価データ (Stock) と大統領選挙の票数予測 (Poll) のデータを使い予測を行う.

Data	Step	Time Series		Time Series + Text				
		SpikeM	LSTM	C_{rand}	C_{words}	C_{topics}	C_{senti}	C_{comp}
Stock	1	102.13	6.80	3.63	2.97	3.01	3.34	**1.96**
Stock	3	99.8	7.51	4.47	4.22	4.65	4.87	**3.78**
Stock	5	97.99	7.79	5.32	**5.25**	5.44	5.95	5.28
Poll	1	10.13	1.46	1.52	1.27	1.59	2.09	**1.11**
Poll	3	10.63	1.89	1.84	1.56	1.88	1.94	**1.49**
Poll	5	11.13	2.04	2.15	1.84	1.88	1.96	**1.82**

Table: 特徴量ごとの誤差.

ここに結論を書く.

実験 3: ニューラルネットによる推論 I

CGRAPH 上の原因・結果について，原因をニューラルネットで結果を予測するタスク．

Cause $\overset{cause}{\longmapsto}$ Efffect in CGRAPH	Beam Predictions
the dollar's $\overset{cause}{\longmapsto}$ against the yen	[1] $\overset{caus}{\longmapsto}$ against the yen [2] $\overset{caus}{\longmapsto}$ against the dollar [3] $\overset{caus}{\longmapsto}$ against other currencies
without any execise $\overset{leadto}{\longmapsto}$ against the yen	[1] $\overset{leadto}{\longmapsto}$ against the yen [2] $\overset{caus}{\longmapsto}$ against the dollar [3] $\overset{caus}{\longmapsto}$ against other currencies

実験 3: ニューラルネットによる推論 II

	BLEU @ 1	BLEU @ 3A	BLEU @ 5A
S2S	10.15	8.80	8.69
S2S + WE	10.15	8.80	8.69
S2S + WE + REL	10.15	8.80	8.69

Table: S2S: sequence to sequence, **WE**: word embedding, and **REL**: relation specific attention.

実験 4: 因果関係の説明

原因・結果のイベントを与えられたときに CGRAPH の経路を生成する.
ニューラルネットによって生成された因果関係の列の例

excess $\overset{match}{\longmapsto}$ excess_materialism $\overset{caus}{\longmapsto}$

people_make_films $\overset{make}{\longmapsto}$ money $\overset{changed}{\longmapsto}$ twitter $\overset{turned}{\longmapsto}$ facebook facebook's stock $\downarrow$

10

- `TSLA`: Tesla
- `MSFT`: Microsoft
- `GOOGL`: Alphabet
- `YHOO`: Altaba
- `ORCL`:

説明生成の評価

生成された説明について accuracy を人手で評価した.
→ ニューラルのほうがよい結果.

Reasoner	SYMB	NEUR
Accuracy (%)	42.5	57.5

定量的な評価は難しい.

まとめ

- 多様なデータで多様な手法を試している．SNS や経済のニュースなどの汚いデータをたくさん使ってるのがすごい．

参考文献 I

Granger, C. W. (1988). “Some recent development in a concept of causality”. In: *Journal of econometrics* 39.1-2, pp. 199–211.

Hamilton, J. D. (1994). *Time series analysis*. Vol. 2. Princeton university press Princeton.

Kang, D. et al. (2017). “Detecting and Explaining Causes From Text For a Time Series Event”. In: *Proceedings of the 2017 Conference on EMNLP*. Association for Computational Linguistics, pp. 2748–2757.

Murakami, S. et al. (2017). “Learning to Generate Market Comments from Stock Prices”. In: *Proceedings of the 55th Annual Meeting of the ACL*. Association for Computational Linguistics, pp. 1374–1384.