

自然言語処理による金融関連情報の高精度解析技術

Natural Language Processing for Financial Data

Fei Cheng, Yusuke Miyao

金融活動におけるデータとは？

- ニュースなど**多様な**テキスト
- **非構造化**データ
- **膨大な**情報量
- 解析が**難しい**



自然言語処理とは？

自然言語テキストを**コンピュータ**に自動解析させる

- テキスト分類 (Text Classification)
- 情報抽出 (Information Extraction)
- **ディープラーニング技術** (Deep Learning)

自然言語処理による金融関連データの解析技術

経済記事からの不祥事報道検知 (テキスト分類)

日経新聞 (2018/4/26)

ゆうちょ銀・佐護副社長退任へ ゴールドマン出身

金融機関

2018/4/26 11:42 | 日本経済新聞 電子版

ゆうちょ銀行の佐護勝紀副社長 (50) が退任することが26日、明らかになった。佐護氏はゴールドマン・サックス証券出身。金融庁参与などを経て2015年にゆうちょ銀に転じ、16年に副社長に就いた。超低金利が続くなか、国債に偏重していたゆうちょ銀の運用改革を主導してきた。3年間で改革の道筋を整えたことを区切りで退く。

→ 不祥事でない

日経新聞 (2018/4/26)

米司法省、中国ファーウェイを捜査か 米紙報道

2018年4月26日 10:07

【ワシントン=風山太成】複数の米メディアは25日、米司法省が中国の通信機器大手、華為技術 (ファーウェイ) を捜査していると報じた。米国製品をイランに違法に輸出した疑いがあるという。米政府は同様の違法行為を巡って中興通訊 (ZTE) に制裁を科すと決めたばかり。中国の通信企業に圧力を強めており、米中貿易摩擦の新たな火種となっている。

→ 不祥事

時間情報解析、タイムライン自動生成 (情報抽出)

2018.1.17
グーグル、新データセンターと海底ケーブルでクラウドインフラ拡大へ

Googleは米国時間1月16日、2018年内にデータセンターを新設し、2019年には新しい海底ケーブルを敷設して、既存のクラウドネットワークを拡大することをブログで発表した。

同社は、2018年に5つの地域にデータセンターを追加する。第1四半期にモントリオールとオランダ、続いてロサンゼルスに新設した後、フィンランドと香港にも年内に新設する。これによって同社は、5大陸にわたる18の地域にデータセンター

2018.1.18
グーグルで今も続く不適切動画の問題YouTube、より厳格な審査基準を導入

米グーグルは、1月16日、同社傘下の動画共有サービス「YouTube」で、不適切動画排除の取り組みを強化するため、より厳格な基準を設けると発表した。

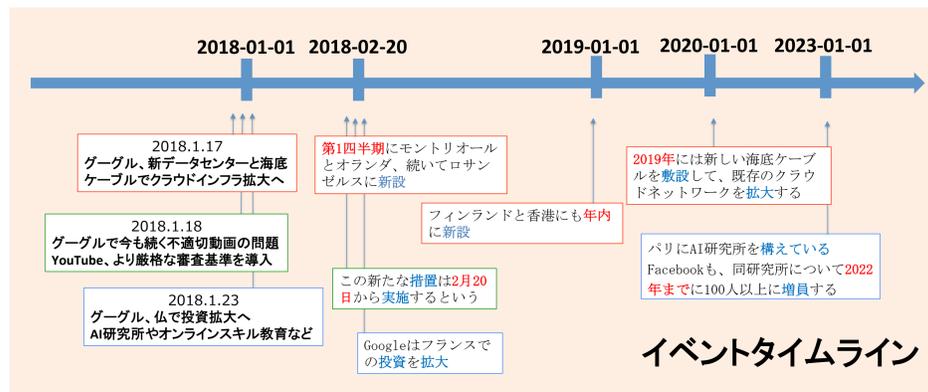
厳しい基準はもう1つある。チャンネルに登録している利用者の数が、1000人に満たない場合も、広告を表示しないというものだ。

この新たな措置は2月20日から実施するという。

2018.1.23
グーグル、仏で投資拡大へAI研究所やオンラインスキル教育など

Googleの最高経営責任者 (CEO) であるSundar Pichai氏は、フランスのEmmanuel Macron大統領と現地時間1月22日に会談した後、同国での投資を拡大する取り組みを発表した。

Bloombergによると、既にパリにAI研究所を構えているFacebookも、同研究所について2022年までに100人以上に増員する



これまでの研究成果

経済記事からの不祥事報道検知

Jason Bennett, 野原 崇史, Fei Cheng, 石田 隆, 宮尾 祐介
言語処理学会第24回年次大会2018

経済記事を自動分類するため、特に以下の3点を重視した実務的な End-to-End システムの設計・構築を行なった

- 解釈性の担保: 分類の根拠が理解できる
- システムの頑健性: 特定期間の特徴 (企業名など) に依存しない
- 再学習の必要性: 新たな種類の不祥事に対応

停止	リコール	カルテル	遅れる
賠償	回収	粉飾	漏れ
解任	不具合	漏れる	損失
障害	無配	事故	トラブル

「不祥事」記事に関連すると認識された特徴量 (解釈性)

イベントの時間情報の自動認識

Fei Cheng, 宮尾 祐介
言語処理学会第24回年次大会2018, NAACL2018

Googleは米国時間1月16日、2018年内にデータセンターを新設し、2019年には新しい海底ケーブルを敷設して、既存のクラウドネットワークを拡大することをブログで発表した。

同社は、2018年に5つの地域にデータセンターを追加する。第1四半期にモントリオールとオランダ、続いてロサンゼルスに新設した後、フィンランドと香港にも年内に新設する。これによって同社は、5大陸にわたる18の地域にデータセンターを新設する。

- タイムライン生成には、イベントの発生時間を認識する必要がある
- 時間情報認識の新手法を提案

	完全一致	部分一致
既存手法	0.442	0.553
提案手法	0.586	0.811