
現在進めている研究について

関西学院大学 理工学部 数理科学科 森本 孝之

統計数理研究所「数学協働プログラム」

政治・社会事象の数的分析に関するスタディグループ

2016 年度第 1 回会合

2016 年 10 月 23 日（日）@ 明治大学

現在進行中の研究は大きく分けて 2 つ

1. 経験類似度を利用した金融市場分析

Empirical Similarity

経験類似度の枠組みでは、意思決定者が予測モデルの成功確率の評価を行わずに、専ら類似性のみに依拠して将来を予測できる。

2. ニュース解析の金融市場分析への応用

News Analytics

ビッグデータとテキスト解析技術を用い、金融市場分析を行う。

1. 経験類似度を利用した金融市場分析

- 事例ベース意思決定理論に基礎を置いた経験類似度（ES, Empirical Similarity）という概念を適用することにより，異なるモデルから生じるボラティリティ予測値を結合する.
- 経験類似度の枠組みでは，意思決定者が予測モデルの成功確率の評価を行わずに，専ら類似性のみに依拠して将来を予測することができる.
- 具体的には，過去のモデル予測値と対応するボラティリティの実現値との距離を定量化することによって，予測の組合せの重みを決定する.
- そして，決定された重みを用い将来のボラティリティを予測する.

類推的推論

- 類推による推論は，過去の経験に基づき，未来の出来事を予測する基本的な方法の 1 つ (Gilboa et al. 2011).
- 帰納的推論の論理的妥当性に疑問を持ち，また類推的推論を議論したことで有名な Hume (1748) は，将来について過去から学ぶ基本思想.
- 不確実な状況において，意思決定者は将来の確率を評価できないが，将来について過去から学びその類似性に基づき思考することは可能.



事例ベース意思決定理論

- 不確実性下の意思決定におけるフォンノイマン・モルゲンシュテルンの期待効用理論では、意思決定者は起こりうる状態をすべて列挙した状態空間から計算される期待効用を最大化すべく行動（尾山，2012）。
- しかし、意思決定者が状態空間を完全に把握していると想定するにはあまりにも無理がある状況も多々あり、そのような状況で人々は過去の経験からの類推に基づいて現時点での行動を決めるであろう。
- これが Gilboa and Schmeidler（1995，2001）の提唱する事例ベース意思決定理論である（尾山，2012）。
- 類似性に基づく推論は、医学、法律、ビジネス、政治、あるいは人工知能における意思決定に幅広く応用（Gilboa and Schmeidler，2001）。

経験類似度

- y_t は時点 t での家具の骨董品の価格, x_t はそのスタイル, 製造年, 大きさ, などといった特性値.
- 過去の値 x_i と 現在の値 x_t を結合し, y_t を評価.
- ヒュームに従うならば, 過去の条件 x_i が現在の値 x_t と似ているか, あるいは似ていないか, を表す類似度の考えが必要.
- y_t の予測において, より似ていない条件の下で得られた観測値よりも, より似ている条件の下で得られた観測値に, より大きい重み.
- 「スタイル, 製造年, 大きさ, 発売時期」に関して, 過去の観測値が現在の観測値に似ているほど, この観測値により大きい重み.

類似度関数 s

- 類似度関数 s を仮定し, データベース $(x_i, y_i)_{i \leq n}$ と新しいデータ点 $x_t = (x_t^1, \dots, x_t^d)$ が与えられたならば, y_t の類似度予測子は

$$y_t^s = \frac{\sum_{i < t} s(x_i, x_t) y_i}{\sum_{i < t} s(x_i, x_t)} \quad \text{あるいは} \quad y_t^s = \frac{\sum_{i \neq t} s(x_i, x_t) y_i}{\sum_{i \neq t} s(x_i, x_t)}$$

- 類似度関数 s については, 任意の関数形で表すことが可能.
- 例えば Billot et al. (2008)

$$s_w(x, x') = \exp \left\{ - \sum_{j=1}^d w_j (x_j - x'_j)^2 \right\}$$

統計モデル

- 仮説検定により定性的結果を得るために、統計モデルに組み込む.

$$y_t = \frac{\sum_{i < t} s_w(x_i, x_t) y_i}{\sum_{i < t} s_w(x_i, x_t)} + \varepsilon_t \text{ あるいは } y_t = \frac{\sum_{i \neq t} s_w(x_i, x_t) y_i}{\sum_{i \neq t} s_w(x_i, x_t)} + \varepsilon_t$$

- 例えば、経済主体による価格形成過程を考えてみると、この経済主体は、過去において既に価格が決定されている他の商品との類似度に応じて、不動産や美術品のような商品の価格を決定するだろう.
- これは、経済主体が価格を決定する際の思考過程のモデルとみなせる.
- このような相互依存性は、空間統計学の応用分野として地理学、社会学、あるいは政治学データにおいても自然に解釈できる.

経験類似度によるモデル組合せ手法

- 経験類似度によるモデル組合せ手法は、予測モデルの組合せの重みを決定するために、異なるモデルの直近の予測力に関する情報を利用する。
- その他の確率的手法と比較して、経験類似度によるモデル組合せ手法を用いる利点として、以下の 3 点が考えられる。
 - (1) モデルの成功確率の知識を必要としない。
 - (2) 経済主体の選好に予測モデルの重みを関連づけられる。
 - (3) 意思決定者が予測値と実現値間の類似度をどのように評価するかをデータから明らかにできる。

ES 概念に基づいた予測の線形結合

- Golosnoy et al. (2014) は, Gilboa et al. (2006) による ES 概念に基づいた予測の線形結合を, 次のように定式化した.

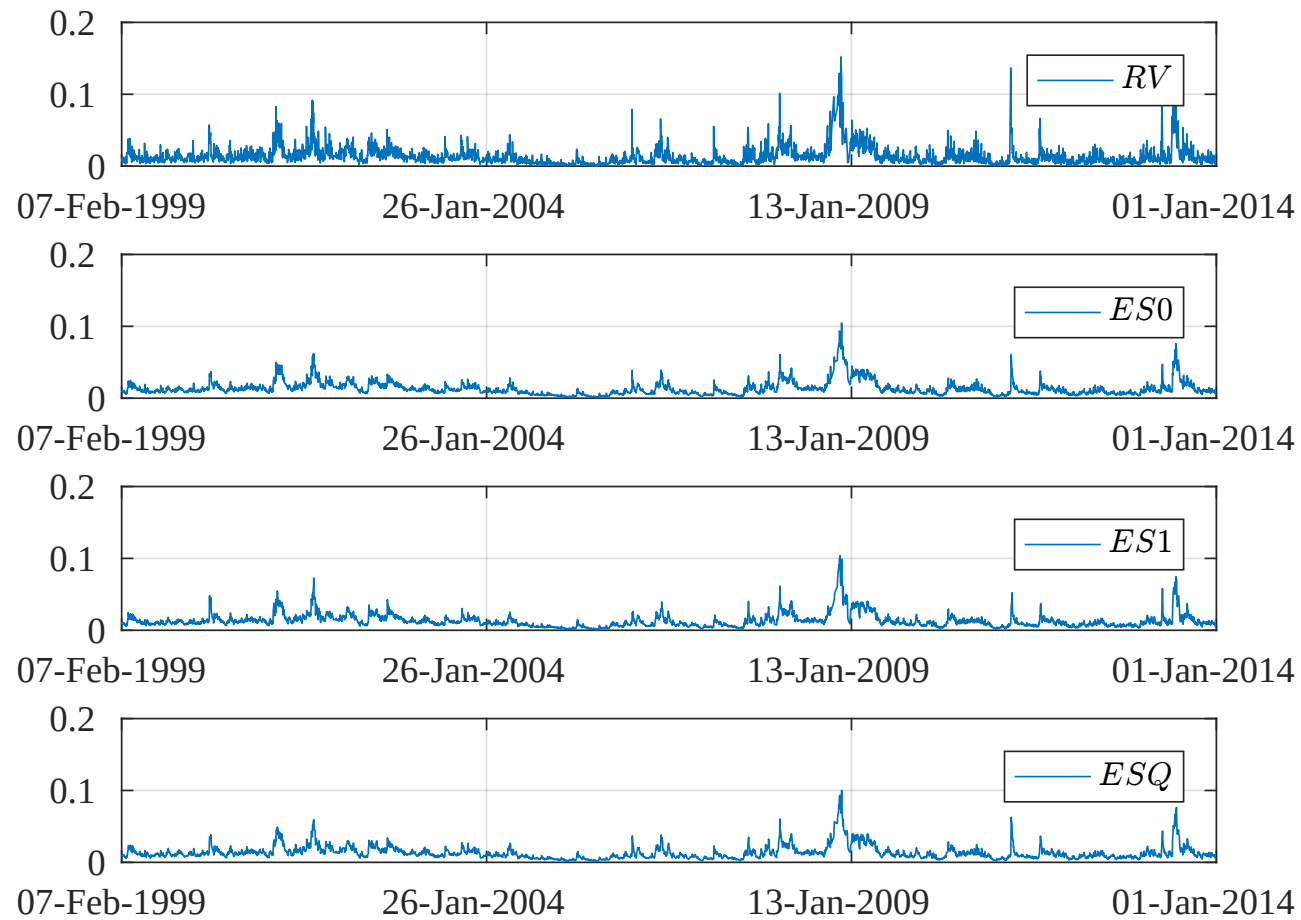
$$y_t = \sum_{j=1}^p \phi[y_{t-1}, x_{t-2}^j] x_{t-1}^j + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2)$$

- また, 重み $\phi[\cdot, \cdot]$ は, 観測されたデータの過去の値に依存し, 現在の実現値の代理変数と j 番目のモデルの予測値間の距離は

$$\phi[y_t, x_{t-1}^j] = \frac{\theta[y_t, x_{t-1}^j]}{\sum_{k=1}^d \theta[y_t, x_{t-1}^k]}$$

のように算出される.

ボラティリティと経験類似度モデル



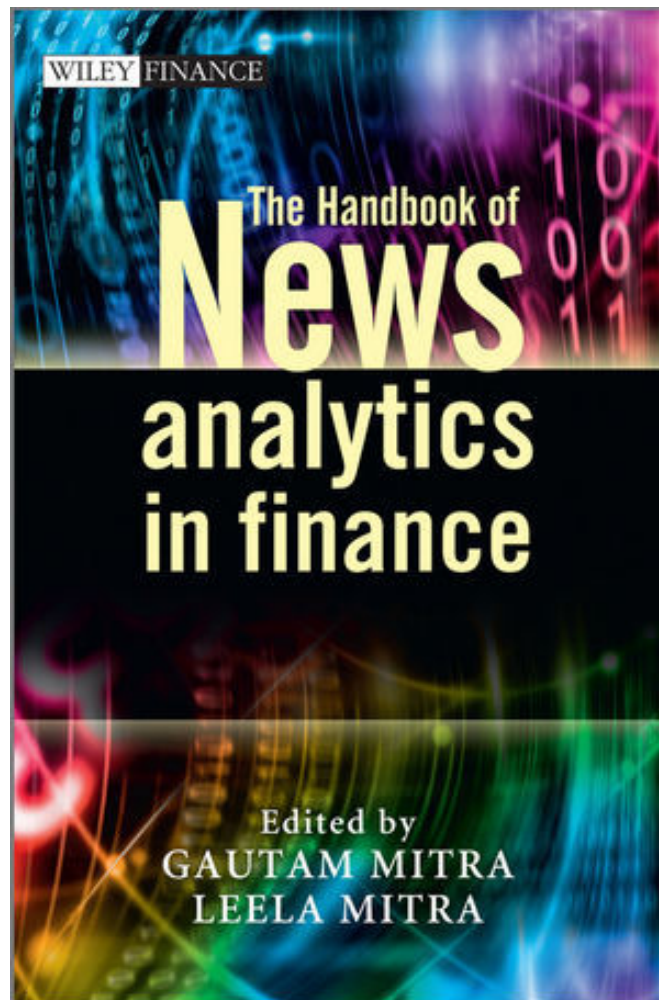
関連する文献 (Empirical Similarity)

- Title: "Analogical Legal Reasoning: Theory and Evidence"
- 題目:「類推による法的論法：理論と証拠」
- Author: Joshua C. Teitelbaum (Georgetown University)
- Published in: *American Law and Economics Review*, Volume 17, Issue 1, Spring 2015, Pages 160–191

概要 (Teitelbaum, 2015)

- 類推による法的論法の形式的モデルを提案し実データに適用する.
- 新しい判例の結果は過去の判例の加重平均として表される.
- この加重平均は過去の影響を受け, 事実の類似性 (空間的距離) と先例の職権 (司法序列における地位) に依存する.
- 米国の海難事故の判例の時系列に対して行なった実証分析により, このモデルの有用性が示される.
- さらに本研究により, 上級の裁判所において決定された判例は, 下級の裁判所において決定された判例よりも影響が大きいことが立証される.

2. ニュース解析の金融市場分析への応用



News analytics

From Wikipedia, the free encyclopedia
News analytics are usually derived through automated text analysis and applied to digital texts using elements from natural language processing and machine learning such as latent semantic analysis, support vector machines, “bag of words” among other techniques.

https://en.wikipedia.org/wiki/News_analytics

ロイターニュース (2014年 02月 6日 22:50 JST)

勝ち残るための経営体制へ＝三菱自次期会長 | 国内 | Reuters - Mozilla Firefox

Firefox tvガイド

REUTERS ロイター

最新ニュース 米株続伸し3指数が揃って1%強上昇、アップル自社株買いも支援

株価、ニュース、ビデオ

ログイン

加藤夏希

“NISA”で投資するなら

松井証券

現在地: ホーム > ニュース > 国内 > 記事

日経平均: 14462.41 307.29 ▲ | 米ドル: 102.33 | DOW: 15794.08 ▲

PR 取引所 F X*くりつく 3 6 5 "始めるなら安心・信頼の太和証券/手数料0円

勝ち残るための経営体制へ＝三菱自次期会長

2014年 02月 6日 22:50 JST

記事印刷 | ブックマーク | 1ページに表示

[-] 文字サイズ [+]

おすすめ ツイート チェック

国内

特集 アベノミクスの波紋

原発・エネルギー政策

スポーツ

エンタテインメント

世界のこぼれ話

ロイター特集

コラム

ブログ 討論x討論

ビデオ

写真

不動産

キャリア

その他

RSS

Twitter

Facebook

Google+

アプリ & 携帯サイト

ニュースメール

会社概要

金融

メディア

Customer Zone

金融製品に関するお問合せ

広告掲載について

編集長のおすすめ

米雇用の伸び、1月は予想大きく下回る11.3万人

1月米雇用統計、予想下回る：識者はこうみる

Windows XP サポート終了間近! Windows XP → 7 or 8

乗りカエル・買いカエル

パソコン取りで、最大5,000円OFF実施中!

iyama 15P5000-i7-FE

74,980円~ パソコン工房

東京都知事選こうみる

細川氏勝利なら安倍おろし

田中秀征氏(元衆院議員)は自民党が都知事選で負けた場合、党内で安倍首相降ろしが始まると予想。記事の全文 | 特集ページ

再稼働に危機感、東電改革に物申す＝細川氏

反原発選挙リク、細川都知事で円安副作用も

細川氏の主張、時代に逆行するリク＝武者氏

「東京都知事選」関連記事一覧

国内ニュースランキング

西武HDとサーベラ

来年度早期のIPOで

【東京 6日 ロイター】三菱自動車(7211.T: 株価、ニュース、レポート)は6日、新経営体制の発表に伴い記者会見した。

次期会長に就任する益子修社長は、公募増資の完了により優先株処理、業績好調で普通株式への復配にもめどがついたことで「無事、再生に区切りをつけ、競争のスタートラインに立てた」と述べ、「競争に勝ち残り、生き残っていくためには強力な経営体制が必要」と新体制移行の狙いを語った。

今回新たに最高経営責任者(CEO)と最高執行責任者(COO)の職を設け、益子氏がCEOを、社長に内定した相川哲郎常務がCOOを兼務する。益子氏は「多くの投資家から経営の継続性を求められた」とCEO・COO体制をとった理由を説明。益子氏が「戦略や経営方針の決定、長期計画策定の責任」を、相川氏は「戦略を効率的かつ確実に実行する役割」をそれぞれ担う。

テキストデータの抽出

勝ち残るための経営体制へ三菱自次期会長

三菱自動車7211は6日、新経営体制の発表に伴い記者会見した。

次期会長に就任する益子修社長は、公募増資の完了により優先株処理、業績好調で普通株式への復配にもめどがついたことで「無事、再生に区切りをつけ、競争のスタートラインに立てた」と述べ、「競争に勝ち残り、生き残っていくためには強力な経営体制が必要」と新体制移行の狙いを語った。

今回新たに最高経営責任者CEOと最高執行責任者COOの職を設け、益子氏がCEOを、社長に内定した相川哲郎常務がCOOを兼務する。益子氏は「多くの投資家から経営の継続性を求められた」とCEO・COO体制をとった理由を説明。益子氏が「戦略や経営方針の決定、長期計画策定の責任」を、相川氏は「戦略を効率的かつ確実に実行する役割」をそれぞれ担う。

データ記述言語 JSON (JavaScript Object Notation)

```
{ "_id" : "52c10f8f", "create_date" : 1293807600000,  
  "author" : "Reuters", "title" : "勝ち残るための経営体制へ三菱  
  自次期会長", "body" : "三菱自動車7211は6日、新経営体制の発表に  
  伴い記者会見した。次期会長に就任する益子修社長は、公募増資の完了に  
  より優先株処理、業績好調で普通株式への復配にもめどがついたことで  
  「無事、再生に区切りをつけ、競争のスタートラインに立てた」と述べ、  
  「競争に勝ち残り、生き残っていくためには強力な経営体制が必要」と新  
  体制移行の狙いを語った。今回新たに最高経営責任者C E Oと最高執行責  
  任者C O Oの職を設け、益子氏がC E Oを、社長に内定した相川哲郎常務  
  がC O Oを兼務する。益子氏は「多くの投資家から経営の継続性を求めら  
  れた」とC E O・C O O体制をとった理由を説明。益子氏が「戦略や経営  
  方針の決定、長期計画策定の責任」を、相川氏は「戦略を効率的かつ確実に  
  実行する役割」をそれぞれ担う。 "
```

形態素解析

オープンソース 形態素解析エンジン Mecab

% mecab

すもももももももものうち

すもも 名詞, 一般, *, *, *, *, すもも, スモモ, スモモ

も 助詞, 係助詞, *, *, *, *, も, モ, モ

もも 名詞, 一般, *, *, *, *, もも, モモ, モモ

も 助詞, 係助詞, *, *, *, *, も, モ, モ

もも 名詞, 一般, *, *, *, *, もも, モモ, モモ

の 助詞, 連体化, *, *, *, *, の, ノ, ノ

うち 名詞, 非自立, 副詞可能, *, *, *, うち, ウチ, ウチ

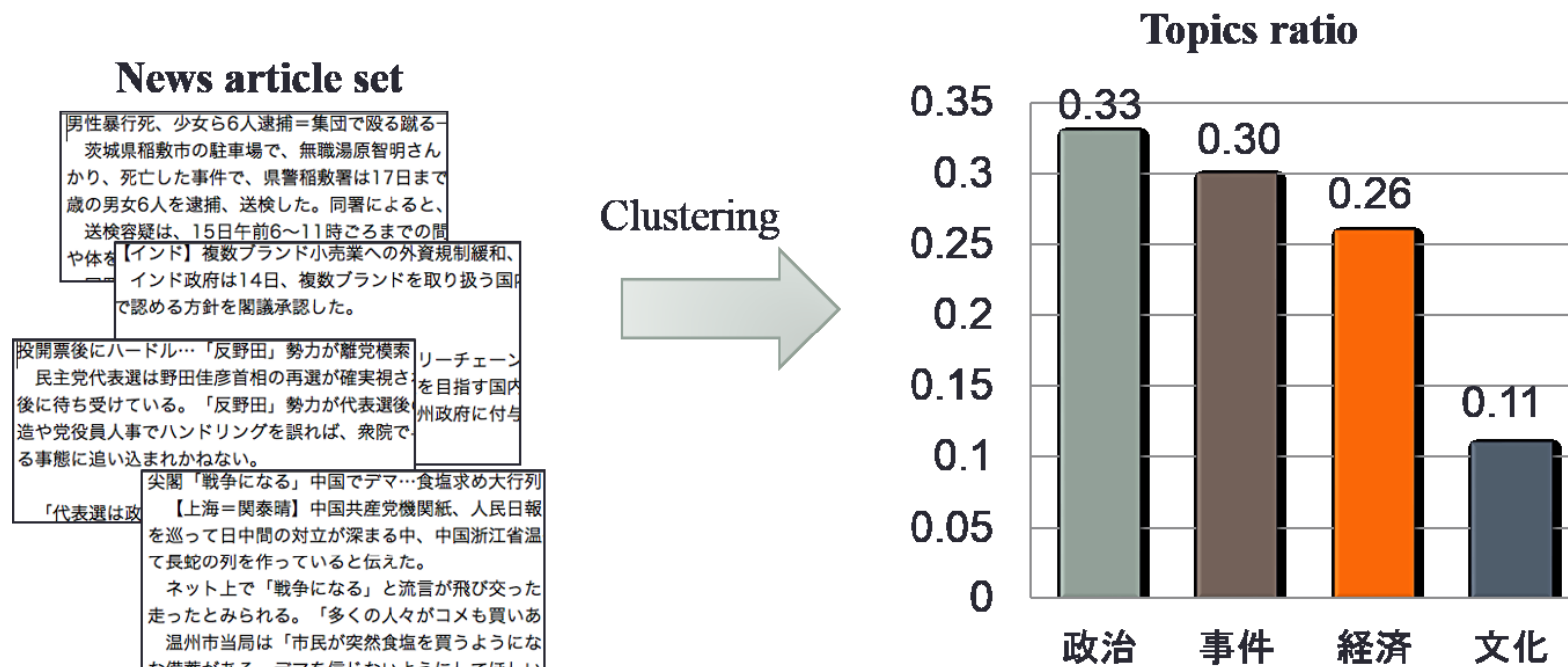
EOS

形態素解析の結果

経営	会長	普通	体制	益子	体制	相川
体制	就任	株式	移行	社長	理由	戦略
三菱自	益子	復配	狙い	内定	説明	効率
次期	修	無事	今回	相川	益子	确实
会長	社長	再生	新た	哲郎	戦略	実行
三菱自動車	公募	区切り	最高	常務	経営	役割
経営	増資	競争	経営	兼務	方針	益子
体制	完了	競争	責任	益子	決定	三菱自
発表	優先株	強力	最高	多く	長期	今後
記者	処理	経営	執行	投資	計画	成長
会見	業績	体制	責任	経営	策定	重要
行次期	好調	必要	職	継続	責任	

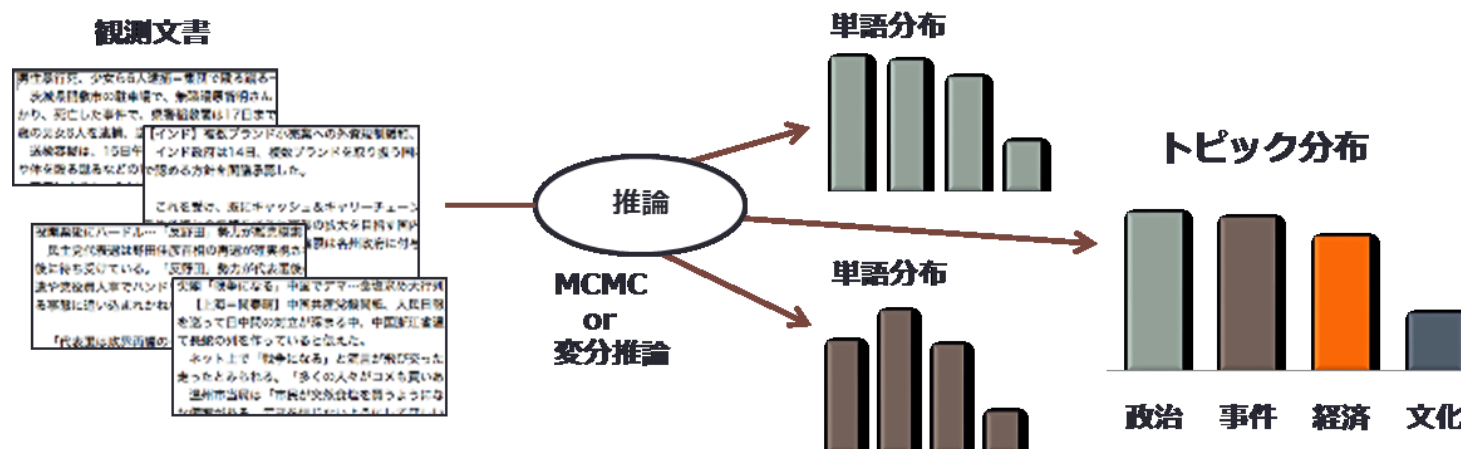
トピック（話題）

- 単語ではなくトピック（話題）を指標として使用.
- ニュース記事中に含まれるトピック（話題）の比率を分析.



トピックモデル

- 文書をトピック（話題）でクラスタリングする機械学習の手法。
- 文書がトピックおよび単語を表す確率分布に従って生成されたと仮定。
- 観測した文書からトピック分布，単語分布を推論。



潜在的ディリクレ配分法

- θ_d を文書 $d \in D$ のトピック分布とする.
- ϕ_z をトピック z の単語分布とする.
- 多項分布

$$\text{Multinomial}(n|\theta) = \binom{N}{n_1 \cdots n_K} \prod_{k=1}^K \theta_k^{n_k}, \quad \sum_{k=1}^K n_k = N$$

- ディリクレ分布 ($\alpha = (\alpha_1, \cdots, \alpha_K)$, $\alpha_k > 0$)

$$\text{Dirichlet}(\theta|\alpha) = \frac{\Gamma(A)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_K)} \prod_{k=1}^K \theta_k^{\alpha_k - 1}, \quad A = \sum_{k=1}^K \alpha_k$$

潜在的ディリクレ配分法の流れ

1. For $k \in \{1 \dots Z\}$:

- 各トピックの単語分布を生成 $\phi_z \sim \text{Dirichlet}(\phi|\beta)$

2. For each document $d \in D$:

- 各文書のトピック分布を生成 $\theta_d \sim \text{Dirichlet}(\theta|\alpha)$

- For each word $w_i \in d$:

(a) 多項分布からトピックを生成 $z_i \sim \text{Multinomial}(\theta_d)$

(b) 多項分布から単語を生成 $w_i \sim \text{Multinomial}(\phi_{z_i})$

ここで α, β は各ディリクレ分布の超パラメータである.

トピック毎の単語分布

トピック 1		トピック 2		トピック 3	
日経	0.109	円	0.208	先物	0.091
平均	0.107	現在	0.069	上昇	0.079
続伸	0.043	週末	0.053	下落	0.077
東証	0.038	一時	0.040	米	0.072
中心	0.037	上値	0.038	原油	0.067
投信	0.035	可能	0.036	安値	0.058
国内	0.032	時間	0.034	反落	0.031
東京	0.032	反発	0.033	前日	0.025
寄り付き	0.031	後半	0.030	日本語	0.022
大手	0.027	連休	0.024	石油	0.021

トピックスコア

- ある特定のトピックについての時系列データを得るため, ここでは次のようなトピックスコアを考える.
- 時点 t における i 番目のトピックに対するスコアは

$$s_t^{(i)} = \sum_{j=1}^{D_t} \theta_{t,j}^{(i)}$$

と定義される.

- ここで D_t は時点 t における文書の数, $\theta_{t,j}^{(i)}$ は時点 t における文書 j のトピック分布の i 番目の成分を, それぞれ表す.

ボラティリティとトピックスコア

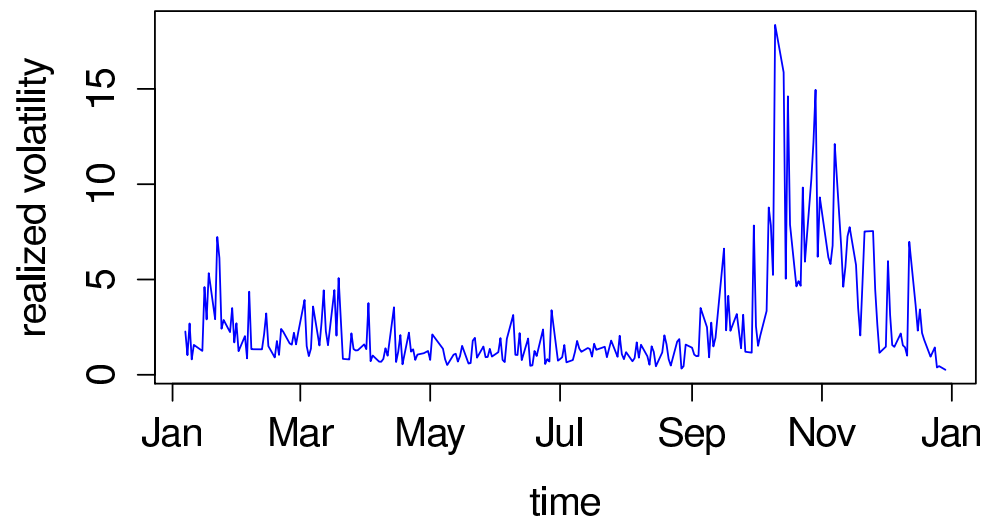


Figure1 ボラティリティ

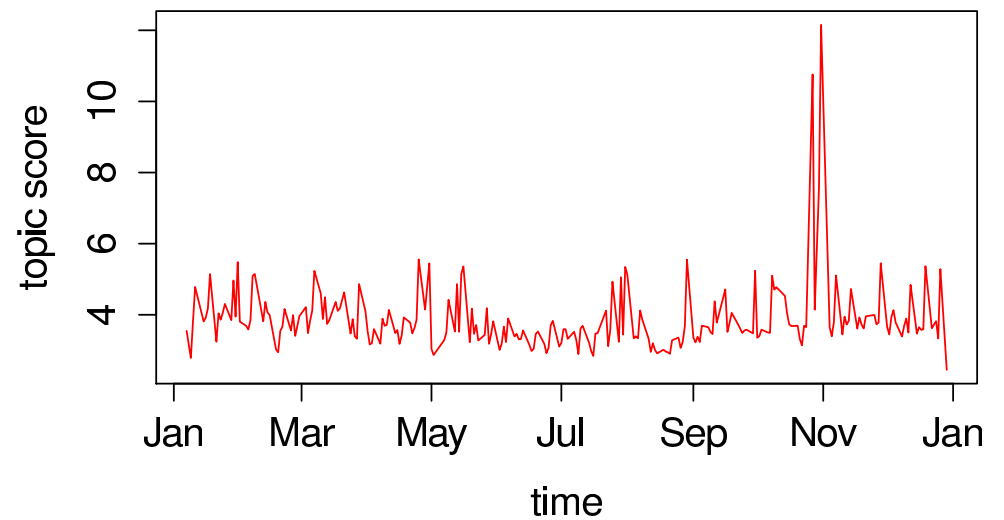


Figure2 トピックスコア

関連する文献 (News Analytics)

- Title: "Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding"
- 題目:「トピックモデルと文化の社会学的見地との親和性を活用する : 米国政府の芸術助成金の新聞報道への応用」
- Authors: Paul DiMaggio, Manish Nag and David Blei
- Published in: *Poetics*, Volume 41, Issue 6, December 2013, Pages 570-606

概要 (DiMaggio et al., 2013)

- 文化社会学に対するトピックモデルの有用性を示す.
- 1986 年から 1997 年までの米国政府の芸術助成金に関する 5 つの新聞の報道記事を評価するために LDA を用いる.
- 8000 近くの新聞記事の中から 12 のトピックが識別され検証される.
- 記事のニュアンスを把握するために, どのようにトピックモデルが用いられるかを示す.
- また, どのようにトピックモデルにより記事の出典ごとの差が明らかになるかを示す.