

# 大規模テキストを利用した経済指標分析手法に関する研究



目的: 大規模経済時系列データの分析

手法: テキストマイニングによる時系列データ予測やシミュレーション

## テキスト利用による人工市場のリアル化

## テキスト分析による経済時系列データ予測

**データマイニングと人工市場シミュレーションとの統合**

- 人工市場ソフトウェアの開発
- 複数の取引プログラムが参加する市場シミュレーション
- テキストマイニング結果より現実の経済環境を反映

**成果① 大規模介入の効果比較**

- 中央銀行の介入行動が外国為替市場の安定性に与える影響評価
- 政情・債券市場・金融政策に関するニュースがあったタイミングで介入すると70%以上為替安定化

**成果② 空売り規制の影響検証**

- 空売り規制が株式市場安定化にどのような効果をもたらすかを検証
- 短期的には安定化するが、長期的にはバブルを引き起こす可能性もあった

人工市場を用いた市場インパクトの評価システムの試作版

実際の介入行動(左)と提案された介入行動(右)のシミュレーションパス

空売り規制なし(左)と規制あり(右)のシミュレーションパス

**日本銀行 金融経済月報のテキストマイニング手法開発**

- テキストマイニングによる初の長期市場分析
- 日本国債、日経平均

**成果① 変動に強い市場動向推定**

- 既存の数値分析手法に比べ精度を平均48.4%、最大74.0%改善
- 変動が大きい時期は特に推定精度が高い(71.4~85.7%の精度で方向性を予測)

**成果② 運用テストによる評価法の開発**

- 月報発表から月末まで国債市場で運用
- テスト期間: 2008年1月~2009年5月
- 既存手法に比べて55~152%のリターン増加

**成果③ 英語テキスト分析への拡張**

- イングランド銀行のレポート、英国SWAP金利
- 正のリターン (71.0~193.3ベースポイント)

テキストマイニングによる日経平均株価の外挿予測結果。月次の価格変動の推定値と実際の値。

## 学習機能の改良(複利型強化学習)

### 複利型強化学習

- リターン型MDPを対象にした強化学習
- 複利リターンに基づく行動価値を最大化する行動規則を学習
- 対数割引複利リターンを最大化

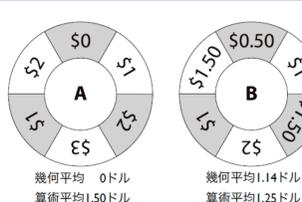
$$\rho_t = \prod_{k=0}^{\infty} (1 + R_{t+k+1}f)^{\gamma^k}$$

$$\log \rho_t = \sum_{k=0}^{\infty} \gamma^k \log(1 + R_{t+k+1}f)$$

$$Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a) = \sum_{s'} P_{ss'}^a (\log(1 + R_{ss'}^a f) + \gamma \max_{a'} Q^*(s', a'))$$

リターン  $R_t$   
割引率  $0 \leq \gamma \leq 1$   
投資比率  $0 \leq f < 1$

状態遷移確率  $P_{ss'}^a$   
期待リターン  $R_{ss'}^a$



- Aに全額賭け続けると\$0が出たときに破産する
- 複利リターンを最大化するにはBの方がいい

## 長期市場動向分析手法(CPR法)

① 共起関係に基づく主要単語の抽出

② 主成分分析による単語のグループ化

③ 重回帰分析による金利データの動向分析

$$r'_{i,t} = a_{0,t} + \sum_{j=1}^{100} a_{j,t} x_{j,t}$$

業種別の株価指数(超過リターン)

被説明変数  $Y_t$

説明変数  $X_{i,t}$

## 3つの観点からの手法改善

- テキストデータの**詳細な言語的分析**
- 大量テキストの解析**
- オンライン学習による効率的解析**

### (1)テキストデータの詳細な言語的分析

#### (a)外部知識

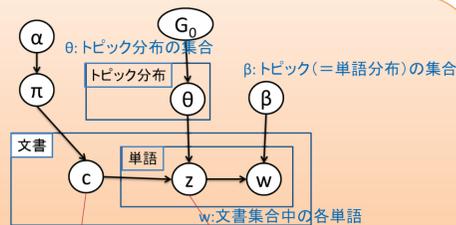
- 経済用語辞書(日経シソーラス)
- 業種別銘柄

#### (b)言語解析ツール

- 係り受け解析
- 固有名詞抽出

#### (c)テキストマイニングツール

- トピックモデル



文書トピック番号 単語トピック番号  
トピックモデルの例:  
Dirichlet Enhanced Latent Semantic Analysis (DELISA)

### (2)大量テキストの解析

- その日のすべての新聞記事
- Twitter上のテキスト
- Google N-gram

### (3)オンライン学習

- 教師あり学習(オンラインパーセプトロン等)
- 教師なし学習(オンライントピックモデル等)