

個別銘柄の連動類似度を用いた株価予測

竹之内研究室
1015100 鈴木雄士

目的

- ・株式取引での利益を増やす
 - 株価の連動性を用いて予測
 - 機械学習を用いることで株価予測の精度を向上
- ・翌日の株価の上昇,下落を予測する

方法

1.特徴量作成

- ・類似度測定
- ・クラスタリング

2.学習モデル作成

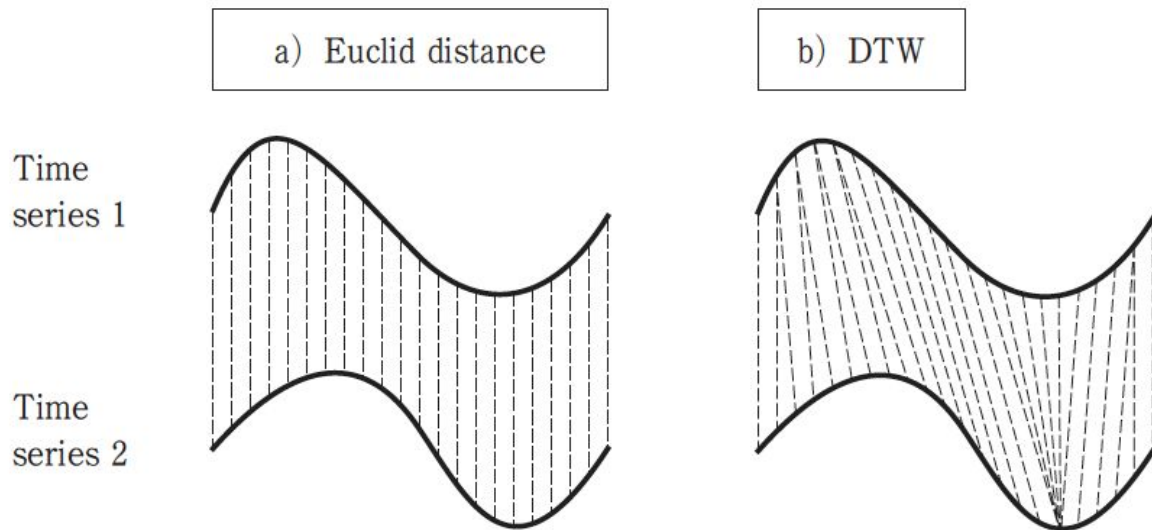
- ・クラスタ平均

3.データの予測・評価

類似度

- ・距離尺度に**DTW距離**を使用[1]

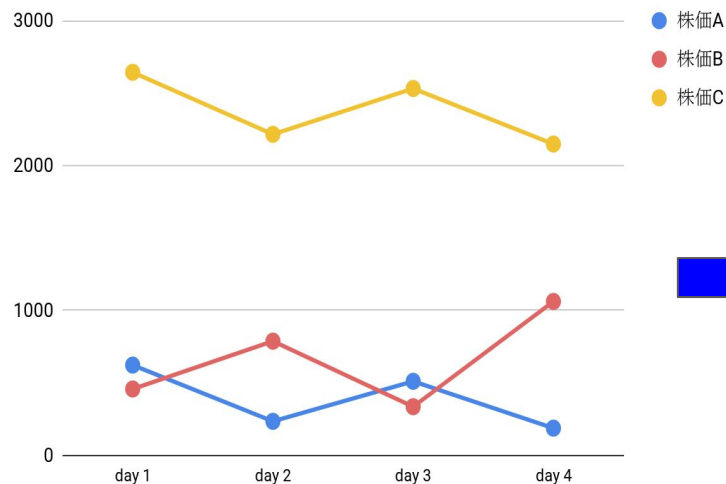
図1 ユークリッド距離 (a) と DTW (b) による時系列の対応付けの比較



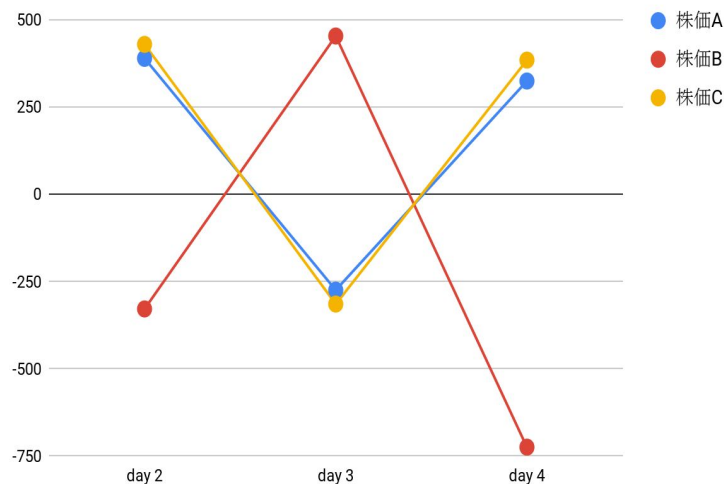
- 株価時系列の**形状**に基づく類似尺度を使用

→ 値をそのまま使うのではなく,**前日との差分**の時系列データに加工する

元データ



前日差分



クラスタリング

- ・x-means法[2]

クラスタの分割前と分割後でBIC値(ベイズ情報量規準)を比較し、
値が改善しなくなるまで分割

→ 最適なクラスタ数を決定

実験データ

- ・日経平均225銘柄(始値,終値,出来高)
- ・2018/4/1～2018/10/1までの125日分
- ・前半7割を学習用トレーニングデータに, 後半3割を評価用テストデータ

クラスタ平均値を用いた株価予測

1. クラスタリング

2. 各クラスタごとに

- ・x日分のクラスタ平均を使って翌日の株価の上下を予測
- ・学習アルゴリズムには決定木, ランダムフォレスト, サポートベクターマシン, 勾配ブースティング

クラスタ平均値を用いた株価予測

1. クラスタリング

2. 各クラスタごとに

- ・x日分のクラスタ平均を使って翌日の株価の上下を予測
- ・学習アルゴリズムには決定木, ランダムフォレスト, サポートベクターマシン, 勾配ブースティング
- ・時間アンサンブル予測
 - 短期, 中期, 長期の予測を組み合わせる

評価方法

$$\text{期待値} = \frac{\max \{ \text{株価上昇回数}, \text{株価下落回数} \}}{\text{株価上昇回数} + \text{株価下落回数}}$$

$$\text{予測精度} = \frac{\text{期待値を上回ったクラスタ数の合計}}{\text{クラスタ数の合計}}$$

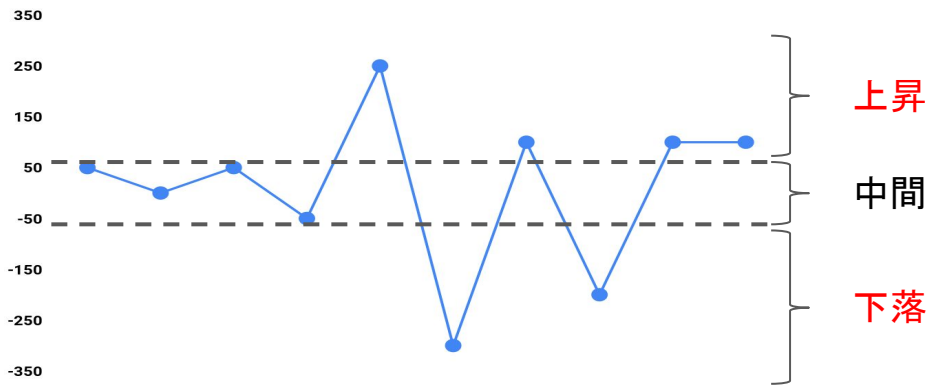
学習アルゴリズム	学習日数			
	1 日	5 日	25 日	1 日 5 日 25 日
決定木	36%	31%	26%	26%
ランダムフォレスト	26%	26%	31%	26%
サポートベクターマシン	31%	26%	26%	31%
勾配ブースティング	31%	15%	31%	15%

表1.終値のクラスタ平均を用いた予測精度

2段階予測

・ほんの少しの上昇下落のような**ノイズ**を取り除いて学習させることで予測精度を向上

→ 翌日の株価の変動が**大きい**(1%以上の上昇または-1%以上の下落)か**小さい**(1%未満の上昇または-1%未満の下落)かというデータを予測モデルに組み込む



1. クラスタリング

2. 各クラスごとに

- ・x日分のクラスタ平均

- ・ランダムフォレスト

- ・ $g(x)$: 翌日の株価が大きく変動するのか小さく変動するのかを予測

→ $f1(x)$: $g(x)$ によって翌日の株価が大きく変動すると判断された株価を対象として, 次の時点での上下を予測

→ $f2(x)$: $g(x)$ によって翌日の株価が小さく変動すると判断された株価を対象として, 次の時点での上下を予測

評価方法

$$\text{期待値} = \frac{\max \{ \text{株価上昇回数}, \text{株価下落回数} \}}{\text{株価上昇回数} + \text{株価下落回数}}$$

$$\text{予測精度} = \frac{\text{期待値を上回ったクラスタ数の合計}}{\text{クラスタ数の合計}}$$

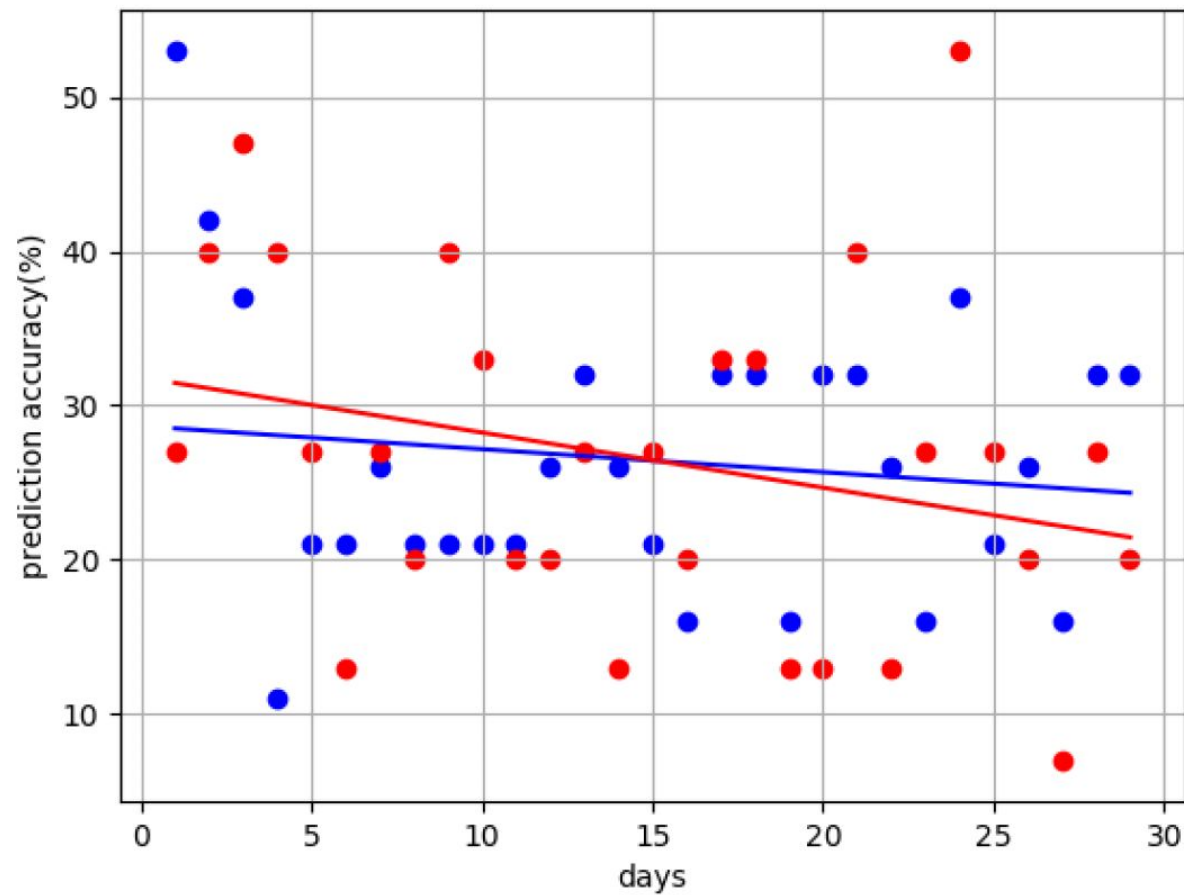


図1.学習日数と予測精度

個別銘柄株価 + クラスタ平均株価予測

- ・予測対象の銘柄の株価予測をする際に、銘柄の株価時系列データに予測対象の銘柄の属する**クラスタのデータ**を加えることで予測精度を向上

1. クラスタリング

2. 各クラスタごとに

- ・x日分の対象銘柄の株価とクラスタ平均を使って翌日の株価の上下を予測

- ・ランダムフォレスト

評価方法

$$\text{評価量1} = y_t - \hat{y}_t$$

$$\text{評価量2} = y_t - \tilde{y}_t$$

y_t 対象銘柄の株価にクラスタ平均のデータを追加した予測精度

\hat{y}_t 対象銘柄の株価のみの予測精度

\tilde{y}_t チャンスレベルの予測精度(訓練データにおける株価の上昇回数と下落回数で多い方を予測としてテスト期間を予測)

対象銘柄の株価のみとの比較

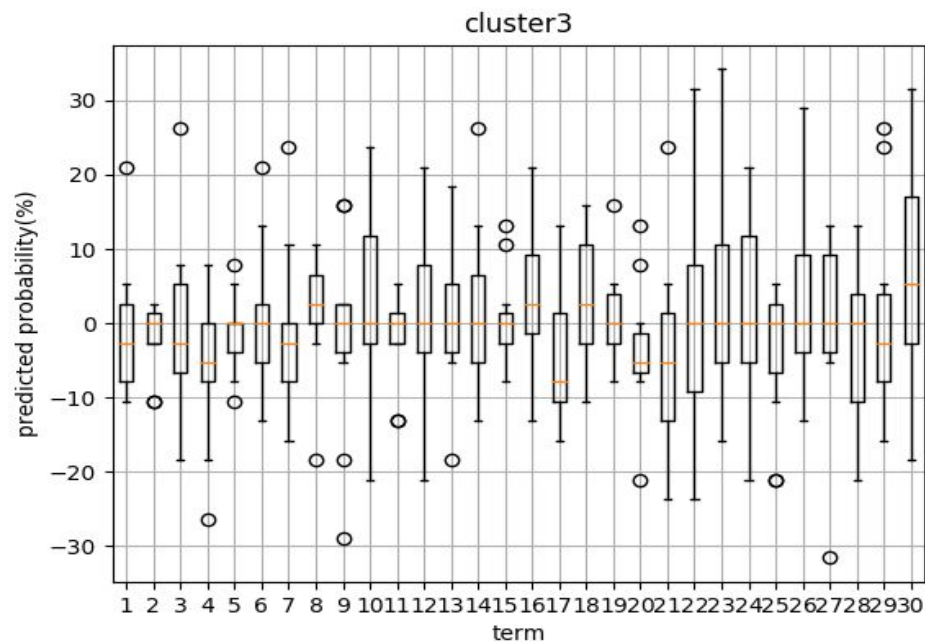


図2. クラスタ3の学習日数と評価量 1

チャンスレベルとの比較

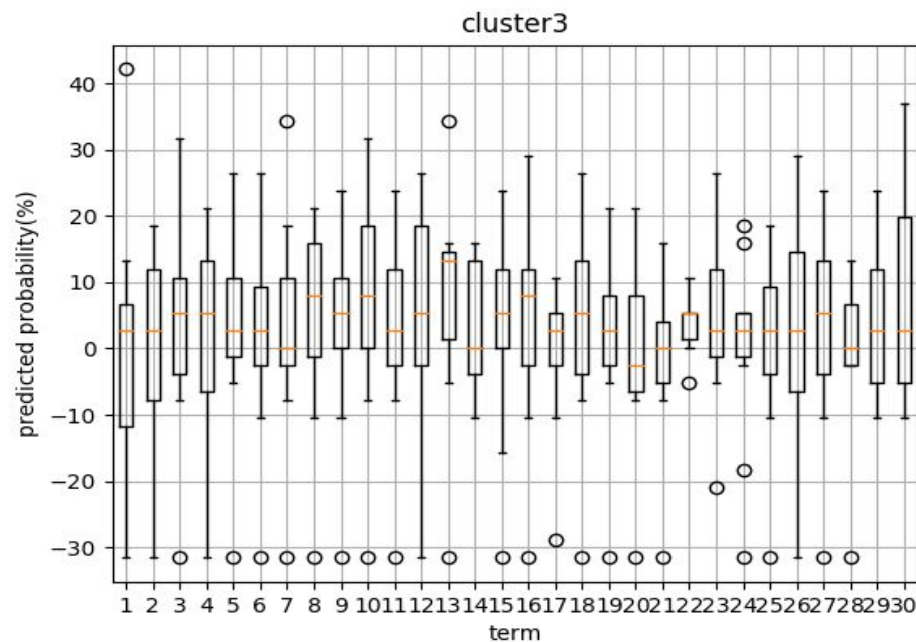


図3. クラスタ3の学習日数と評価量 2

結論

- ・クラスタ平均を用いることで予測精度が上がる傾向は見られなかった
- ・株価が自身の業績だけではなく様々な外部要因の影響を受けて形成されていることや過去の情報だけでは説明できない動きをするためだと考える
- ・他の統計量など株価からのデータの抽出方法を変えたり、為替や経済指標などの株価以外のデータを用いる

参考文献

- [1].中川慧, 今村光良, 吉田健一: 株価変動パターンの類似性を用いた株価予測, 第31回人工知能学会全国大会, 2D1-1, 2017.
- [2].小沢育実, 関和広: 関連銘柄同定のための時系列データ類似度尺度の提案, 第78回情報処理学会全国大会, 6H-08, 2016.
- [3].志津綾香, 松田真一: クラスタ分析におけるクラスタ数自動決定法の比較, アカデミア情報理工学編11, 17-34, 2011.
- [4].赤池弘次: AICとMDLとBIC, オペレーションズリサーチ, 1996年7月号, 375-378, 1996.
- [5].佐藤賀一: テクニカル分析に基づくペイトレードの有効性と日本の株式市場の効率性, 行動経済学第10巻, 22-49, 2017.
- [6].荒木大, 小島昌一: 数値データによる決定木の帰納学習, 人工知能学会誌, 7巻6号, 992-1000, 1992.
- [7].海野一則, 山田隆志, 寺野隆雄: 機械学習を用いたポートフォリオの最適化, 第27回人工知能学会全国大会, 1D5-6, 2013.

x-means法

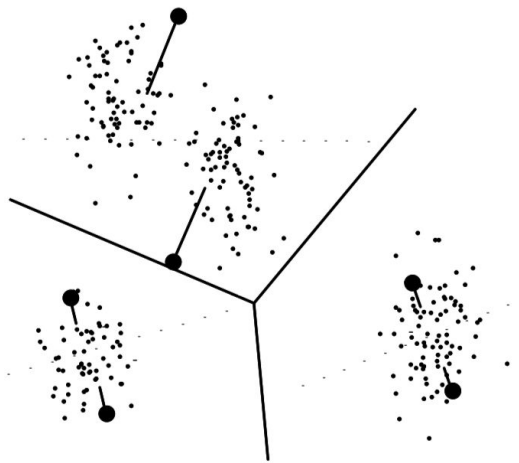


Figure 3: The first step of parallel local 2-means. The line coming out of each centroid shows where it moves to.

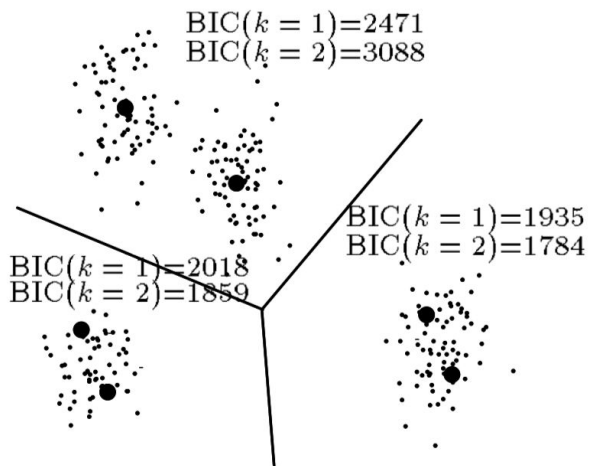


Figure 4: The result after all parallel 2-means have terminated.

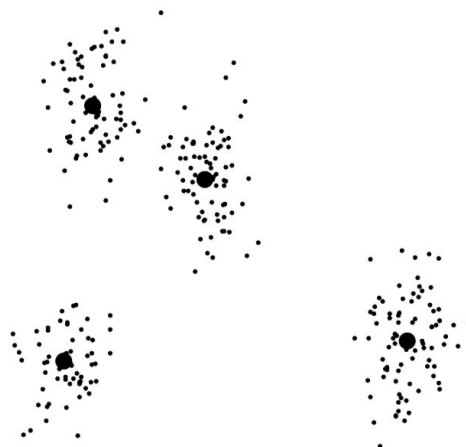


Figure 5: The surviving centroids after all the local model scoring tests.

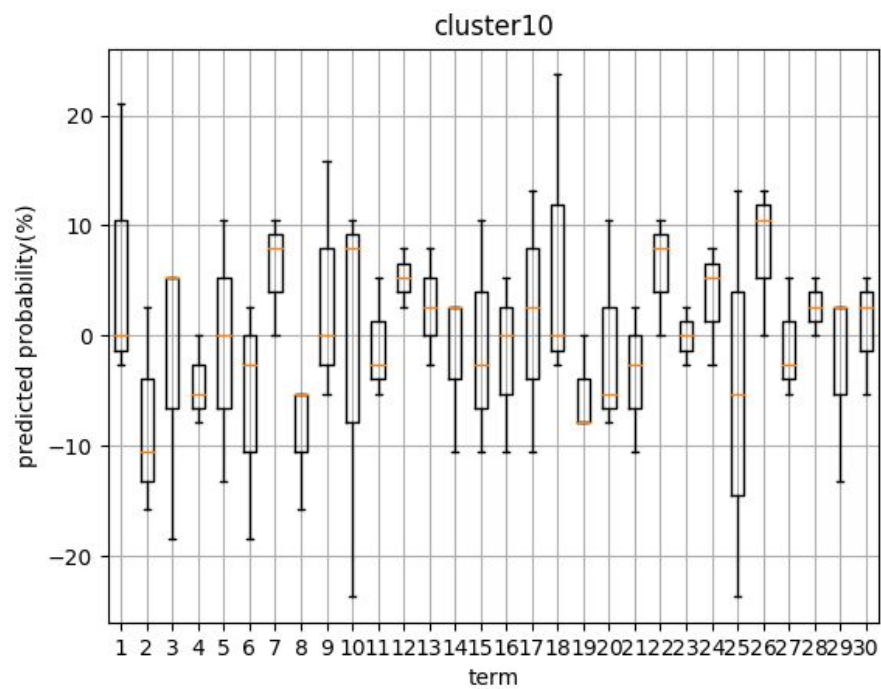


図2. クラスタ10の学習日数と評価量 1

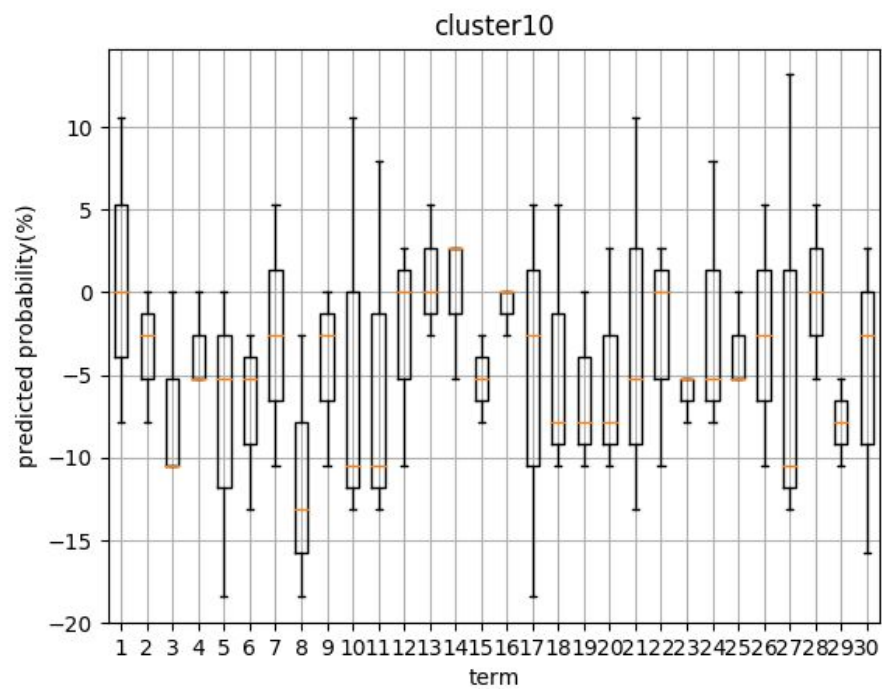


図3. クラスタ10の学習日数と評価量 2