6H-08

関連銘柄同定のための時系列データ類似度尺度の提案

小沢　育実[†]　関　和広[‡]
甲南大学 知能情報学部

1　はじめに

ある銘柄の株価が大きく変化したとき，そのグループ企業や取引企業等，他の銘柄の株価にも変化が生じることがある．本論文では，そのような関連銘柄を発見するための新しい類似度尺度を提案する．そして，1081社の過去の株価データを用いて実際に類似度計算を行い，結果を可視化・分析することで提案類似度の有用性について検証する．

2　既存の類似度/距離尺度

株価等の時系列データの類似性あるいは距離を測るため，これまでも多くの尺度が利用・提案されている．例えば，よく使われる単純な距離尺度として，ユークリッド距離が挙げられる．ユークリッド距離は，2つの時系列における同一時点の値を対応付けて，その差の2乗和の平方根を距離とする．ただし，定義上，2つの時系列の長さが異なる場合に距離を定義できない．

時系列の長さが異なっていても適用可能な距離尺度としては，Dynamic Time Warping（DTW）が広く使われている．DTWでは，一方の時系列のある時点と，もう一方の時系列の複数の時点を対応付けて距離を算出することができる．ただし，ユークリッド距離もDTWも，データの個々の値を基に時系列データ間の距離を定義しており，距離が近い時系列データ同士でも，その形状は必ずしも類似しているわけではない．例えば，時系列データのある区間で値が上昇し，もう一方の時系列データでは下降しているような場合でも，その区間の値そのものに差がなければ，類似していると判断されてしまう．よって，DTWなどの値に基づく距離尺度は，形状に着目して時系列データの類似性を測定したい場合は，適切な距離/類似度の尺度とはならない．

そこでKeoghら[2]は，時系列の形状に着目したDerivative Dynamic Time Warping（DDTW）を提案している．DDTWでは，形状の類似性を考慮するため，式(1)で定義される傾きを用いて時系列データを変換したのちに，DTWにより距離の算出を行う．

$$x'_i = \frac{(x_i - x_{i-1}) + ((x_{i+1} - x_{i-1})/2)}{2} \qquad (1)$$

3　類似度尺度の提案

本研究では，時間的なずれを持って変動する株価の類似性から関連銘柄の発見を試みる．そのため，前節で述べた距離尺度では，時系列データ間の適切な関係性を測定することができない．そこで，次のような特性を持った類似度尺度を新たに定義する．

- 一方の時系列を基準にしたとき，もう一方の時系列との対応付けは，必ず時間的な遅れを持つように制限する（図1）．
- 「基準データが上昇（下降）した際に，何日後かにもう一方でも上昇（下降）する」もしくは「基準データが上昇（下降）した際に，もう一方では下降（上昇）する」ような関連性を持つデータ間に高い類似度を与える．

基準となる時系列を $X = x_1, x_2, \ldots, x_i, \ldots, x_m$，比較対象とする時系列を $Y = y_1, y_2, \ldots, y_j, \ldots, y_n$ とおき，各点を対応付ける際は，$j > i$ という制限を設ける．すなわち，時系列 X の時点と時系列 Y の時点が同時間で対応しないようにし，必ず時系列 X の時点と何日後かの時系列 Y の時点を対応させる（図1）．

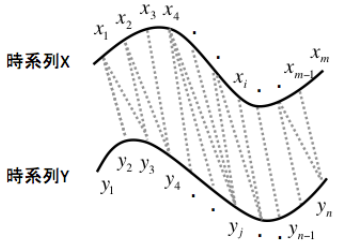

図 1: 時系列の対応付け．

また，株価の変動（上昇・下降）を考慮するためデータの形状に注目し，DDTWと同様に傾き（式(1)参照）

On similarity measure for identifying associated financial time series data
[†]Ikumi Ozawa
[‡]Kazuhiro Seki
Faculty of Intelligence and Informatics, Konan University

を用いる．そして，各時点での類似度を算出し，その中で最も類似度が高くなる対応付けを探索することで類似度 S を定義する．

$$S(X_m, Y_n) = \max \begin{cases} S(X_m, Y_{n-1}) + s(x'_m, y'_n) \\ S(X_{m-1}, Y_n) + s(x'_m, y'_n) \\ S(X_{m-1}, Y_{n-1}) + s(x'_m, y'_n) \end{cases} \quad (2)$$

X_i は X の部分時系列を $x_1, x_2, \ldots, x_i$ とする（Y_j についても同様）．また，x' は式 (1) で定義される傾きである．各時点での類似度は，前述の類似度の特性を満たすよう，式 (3) のように定義する．

$$s(x'_i, y'_j) = x'_i \times y'_j \times \frac{1}{t^2} \quad (3)$$

なお，t は時系列 X と Y の時間差（$j - i$）を示しており，$\frac{1}{t^2}$ の因子により，時間差が少ない点の対応付けが優先される．この定義により，時系列 X に大きな変動があり，その後時系列 Y にも大きな変動があった際に，（Y が X に影響を受けたと考え）類似度の絶対値が大きくなる．なお，類似度の値が正で大きい場合は X と Y が同じ方向に変動しており，負で大きい場合は逆の方向に変動していることを示す．

4　評価実験

前節で提案した類似度尺度の有効性を検証するため，既存の距離尺度 DDTW と比較しつつ評価実験を行った．実験データとしては，1990 年 1 月から 2012 年 12 月の間，東証 1 部，2 部に上場していた 1081 社の株価データを用いた．また，株価の大きさに違いがありすぎる際，適切な類似度が算出できないため，各時系列を -1 から $+1$ に正規化した． 図 2 ではある 1 企業を基準データとして例に挙げ，それぞれの方法で算出した類似度の大きい上位 10 企業を表示している．

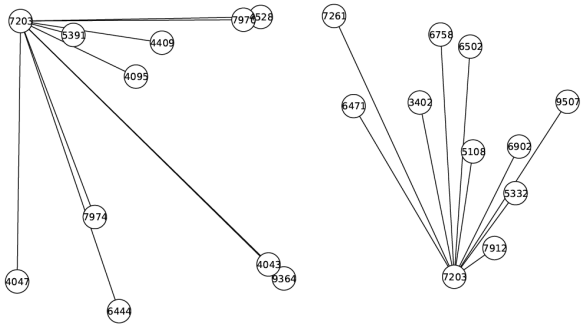

図 2: DDTW（左）と提案類似度尺度（右）の結果の比較．

一般的に，同じ業種は類似度が高いと考えられる．しかし，DDTW では同じ業種で時系列が似ていると考えられる銘柄同士の類似度が高いと判断されない場合があった．長い期間（何か月分）ずれていても似ていると判断される銘柄があったからだと考えられる（図 3 に例を示す）．2 つの銘柄間を実線で引いている部分が似た変化をしている部分であり，DDTW ではこのような長期間ずれている対応付けが多く見られた．これに

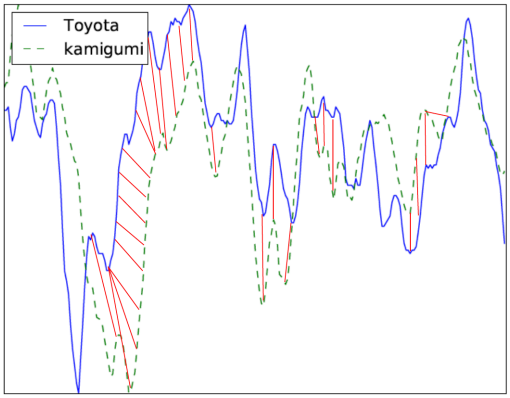

図 3: トヨタ（7203）と上組（9364）の株価．

対して，提案類似度尺度では，同じ業種で類似度が高いと判断される銘柄がいくつか見られた．同じ業種の類似性を示せたことから，異なる業種でも類似度が高いと判断された銘柄は，基準データの銘柄に影響を受けて株価の変動が生じている可能性がある．

5　まとめ

本研究では，ある銘柄の株価が変動したときに，その影響を受けて遅れて株価が変動する銘柄を発見することを目的とし，新しい類似度尺度を提案した．実験の結果，同じような業種であれば高い類似度が得られることが確認できた．一方，異なる業種でも類似度が高いと判断される銘柄も見つかり，今後，これらの銘柄の実際の関連性を検討していく必要がある．

謝辞

本研究の一部は，私立大学等経常費補助金特別補助「大学間連携等による共同研究」によるものである．

参考文献

[1] Eamonn Keogh. “Exact Indexing of Dynamic Time Warping,” in *Proc. of the 28th VLDB Conference*, pp. 406–417, 2002.

[2] Eamonn Keogh, Michael J. Pazzani. “Derivative Dynamic Time Warping,” in *Proc. of the 2001 SIAM International Conference on Data Mining*, pp. 1–11.