

アノテーション付き時系列データからの 知識発見システムの開発

学生員 杉村 博* 正員 松本 一教*

Development of Knowledge Discovery System from Annotated Time Series Data

Hiroshi Sugimura*, Student Member, Kazunori Matsumoto*, Member

(2011年6月20日受付, 2011年11月19日再受付)

This paper proposes a knowledge discovery system from annotated time series data, they are expressed as sequences of numerical values. They generally have a lot of important information in background, but it is not included in data. Analysis methods without background information have limitations. Several studies propose meta data approaches, which is often expressed as a short text, the techniques are insufficient for our purpose. We therefore develop a method that uses annotations which are compact expressions of background information. Subsequences obtained by using domain knowledge are organized into groups based on a distance measure. Among the groups some of them are identified as important features. In order to measure importance, we develop a method that uses global and local frequencies of subsequences. This idea is similar to the TF*IDF method, which is used in text mining. A subsequence that represents an important group is regarded as a feature pattern. In addition, we extract association rules over feature patterns and annotations. We introduce a new concept, called max allowance length, to focus only on influential annotations to a pattern. We demonstrate an effect of the proposed method by using financial data.

キーワード：アノテーション, 特徴抽出, クラス分類, 時系列データ, 知識情報処理

Keywords: Annotation, Feature extraction, Classification, Time series data, Knowledge engineering

1. まえがき

情報システムは社会生活において当然のように利用され、それに伴って膨大な量のデータが多彩な形式で蓄積されており、この多様なデータに対応するデータマイニング技術の重要性が高まっている。とくに、センサデータのように時間とともに記録された数値のシーケンス（時系列データ）だけを直接人間が扱うことは難しく、機械的な処理による知識抽出が望まれている。さらに、一般にデータの背景には重要な情報が内在しており、時系列データ以外の人間による解釈や背景情報が伴わなければ抽出できない情報がある。

大量のデジタルコンテンツを有効利用するため、テキス

トや音楽などに対してメタ情報（アノテーション）を付与し利用する研究が盛んに行われており⁽¹⁾、従来困難であった検索やコンテンツ簡約など多くのタスクを実現するための助けになることが知られている。時系列データにアノテーションを付与して、商品の売上げに対する宣伝効果や薬剤投与による病状の変化を知識抽出する場面が考えられるが、そのような研究はまだ提案されていない。

論文(2)は、時系列データとテキストの協調により時系列データを要約した文書を作成する方法について述べており、株価データとニュース記事を用いて、表示している時系列データの時間軸の縮尺をユーザが変更すると、ニュース記事から時系列データを説明するテキストを自動生成し、ニュース記事が要約される。ニュース記事を時系列データに付与したアノテーションとみなせるが、この方法で抽出されたテキストは、時系列データを説明するためのもので、新しい知識となるようなデータとは考えられない。

時系列データのみを使用して知識抽出を行う研究として、文献(3)は時系列データから頻出するパターンを発見する手法を提案しており、獲得した頻出パターンを特徴パターンと見

* 神奈川工科大学 工学研究科 情報工学専攻
〒243-0292 神奈川県厚木市下荻野 1030
Course of Information and Computer Sciences, Graduate School of Engineering, Kanagawa Institute of Technology
1030, Shimo-ogino, Atsugi-shi, Kanagawa 243-0292, Japan

Table 1. Table of time series data

| Index | Time | Value |
|-------|-------------|-------|
| 1 | 1 May, 2011 | 25298 |
| 2 | 2 May, 2011 | 24876 |
| 3 | 3 May, 2011 | 25352 |
| 4 | 4 May, 2011 | 25005 |
| 5 | 5 May, 2011 | 25329 |

なしている。しかし、頻出パターンは一般的に既知の情報である可能性が高く、必ずしも解析者への新しい知識となるわけではない。したがって、この手法で発見したパターンは重要でないか、または興味深いパターンではない場合が多い。

文献(4)は、時系列データから背景知識をもったユーザにより指定されたパターンを抽出する手法が提案されている。この手法は、2つの部分時系列データの手書き入力パターンとその組み合わせ方を指定することで、時系列データベースから条件を満たすパターンを抽出する。この方法でユーザは関心と一致する特徴パターンを発見できるが、この手法は前提条件として取り扱う時系列データについての知識をユーザがすでに獲得している必要があるため、知識が不十分なユーザの場合には有用なデータを獲得できない。

サポートベクタマシン (SVM) によって時系列データを分類する手法が文献(5)で提案されている。この方法は高い分類精度を獲得しているが、SVM で得られたモデルは、人間が理解することが困難であるという問題がある。人間が理解できないような知識を抽出する方法では、アノテーションとの有効な連携ができない。

本論文ではアノテーションを付与された時系列データから、新たな知識として特徴パターンとアノテーションとの相関ルールを抽出する方法を提案する。まず時系列データは領域知識によって切り出され、重要性を示す統計値に基づいて特徴パターンを発見し、特徴パターンとアノテーションとの相関ルールを抽出する。有効な相関ルールを抽出するために、本論文では相関ルールマイニングで用いられる最小信頼度と、新たな閾値である最大許容長を用意した。この2つの閾値に基づいて特徴パターンとアノテーションとの有効な相関ルールを抽出できることを示す。さらに、発見した知識を利用するための一手法として、獲得した知識に基づいて時系列データを分類する方法について述べる。

2. アノテーション時系列データ

数値のシーケンスとアノテーションを記録するための2つのテーブルがデータベースに保存される。アノテーションはある一点の観測時間を指定する方法と、期間を指定する方法の2種類が考えられるが、点を示すアノテーションは開始と終了時間が同じ範囲とみなすことで表現できる。Table 1とTable 2にデータ構造を示す。Table 1に観測時間と値が記録され、Table 2に時間に関連付けられたアノテーションが記録される。

Fig. 1にアノテーション時系列データの概要を示す。この図はある会社の株価変動を示しており、TV commercial

Table 2. Table of annotations

| Index | From | To | Annotation |
|-------|--------------|--------------|---------------|
| 1 | 15 May, 2011 | 15 May, 2011 | News Release |
| 2 | 20 May, 2011 | 11 Jun, 2011 | TV commercial |
| 3 | 21 Jun, 2011 | 21 Jun, 2011 | News Release |
| 4 | 5 July, 2011 | 5 July, 2011 | News Release |

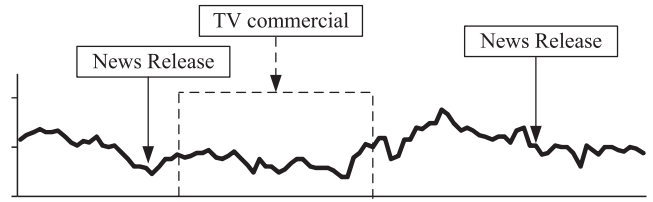


Fig. 1. Annotated time series data

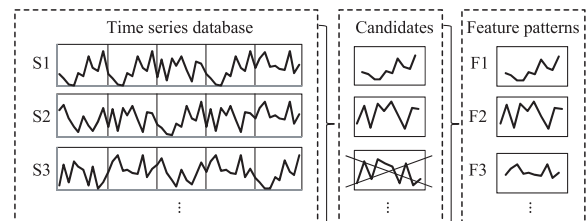


Fig. 2. Feature pattern extraction

が範囲を表しており、News Releaseは点のアノテーションを示している。

3. 特徴パターンと相関ルールの抽出

上述した通り、単純に頻出するパターンは一般的に既知の情報である可能性が高く、必ずしも解析者への新しい知識となるわけではない。このように単一の時系列データ内において頻出パターンが重要であることは明確であるが、一方、データベース全体で頻出するパターンは単一の時系列データを特定するための特徴という観点では重要ではない。Fig. 2に抽出すべき特徴パターンの概要を示す。データ S1 にとってパターン F1 は重要であるが、データ S2 にとってはパターン F2 が重要である。

〈3・1〉 特徴パターンの重要度 TF*IDF は、テキストマイニングの分野で重要な単語を発見するために使用される技術で、ある単語が文書に対してどれくらい重要であるかを評価するのに使用される統計的な測定値である⁽⁶⁾。単語の重要性は、文書内でのその単語の頻度に比例して増加するが、データベース全体の頻度で相殺される。同様に、ある特定の文書にのみ現れる単語は、データベース内の多くの文書に現れる単語よりも重要であると考えられる。TF*IDF は次の式(1)で定義される。

$$\text{TF*IDF}(w_a, t_b) = \text{tf}(w_a, t_b) \times \text{idf}(w_a) \cdots \cdots (1)$$

このとき、 $\text{tf}(w_a, t_b)$ は文書 t_b 内の単語 w_a の出現頻度であり、 $\text{idf}(w_a)$ は α をデータベース内の文書の数、 β を単語 w_a が一回以上現れる文書の数とすると、 $\log \frac{\alpha}{\beta}$ で定義される。

本論文ではこの手法を時系列データに適用する。データ

ベース内の1つの時系列データを1つの文書とみなし、全ての時系列データから抽出した各部分時系列データを単語とみなす。部分時系列データの出現頻度を計算することで、重要な部分時系列データを発見できる。

一般に、部分時系列データは網羅的に獲得するためにスライドウィンドウで切り出すが、この手法で得られる部分時系列データの総数は膨大になり、それら各データの出現頻度もゼロに近いものとなる。そこで、切り出した全ての部分時系列データから幾つかの代表的な部分時系列データを発見し、その類似度または距離を計算することでグループ分けを行い、グループごとに頻度を計算する。グループはその中心となる1つの部分時系列データで代表する。この代表的な部分時系列データの発見とグループ分けのためにクラスタリングを行う。

クラスタリングには k-means⁽⁷⁾を用いる。この手法は、分割の良さの評価関数を定め、その評価関数を最適にする分割を探索する。評価基準は部分集合 G_i に含まれる各データ v と G_i のセントロイド (重心) g_i の間の距離である。重心を $g_i (i = 1, \dots, k)$ 、 $D(P, Q)$ をデータ P とデータ Q の間の距離を測定する距離関数としたとき、評価基準は次の式 (2) で定義される。

$$\sum_{i=1}^k \sum_{v \in G_i} D(v, g_i) \dots \dots \dots (2)$$

単純なスライドウィンドウにより機械的に切り出した部分時系列データに対し、ユークリッド距離を用いて k-means を行うと意味がないセントロイドが出力されるという指摘がなされている⁽⁸⁾。機械的な部分時系列の切り出しという枠組みの中で、この問題を解決する方法が提案されているが⁽⁹⁾、本論文ではスライドウィンドウのサイズとスライド幅を機械的に固定するのではなく、領域的な知識に基づいて決定する方法を用いる。文献 (10) では対象領域の知識を使ってデータの変化に注目したスライドウィンドウのサイズとスライド幅の決定を行い、不等間隔となるスライドウィンドウを用いることで意味のあるクラスタリングを行っており、本論文でも数値の変化に基づいた不等間隔の切り出しを行う。実際の切り出し方についてはデータに依存するため、詳細は実験で説明する。

k-means において距離関数 $D(P, Q)$ は一般的に属性ベクトル間のユークリッド距離で求められるが、ユークリッド距離はシーケンスの各要素を独立して比較するため、長さの異なるシーケンスのペアの距離を計算できない。そこで距離測定関数 $D(P, Q)$ は Dynamic Time Warping (DTW) を用いる。DTW⁽¹¹⁾ は2つの時系列データの距離を最小化するように時間軸を伸縮させる変換処理で、DTW 距離は動的計画法に基づいて計算される。Fig. 3 にユークリッド距離と DTW 距離の違いを示す。

長さ i のシーケンス $P = p_1, \dots, p_i$ と長さ j のシーケンス $Q = q_1, \dots, q_j$ との DTW 距離 $D(P, Q)$ は、下記の式 (3) で定義される。

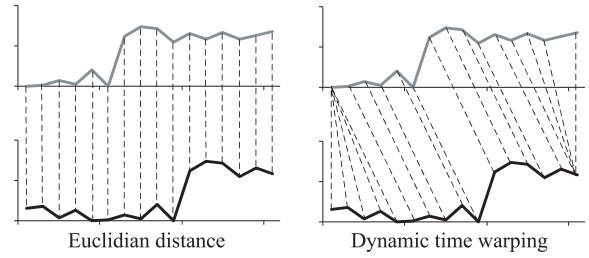


Fig. 3. Distance

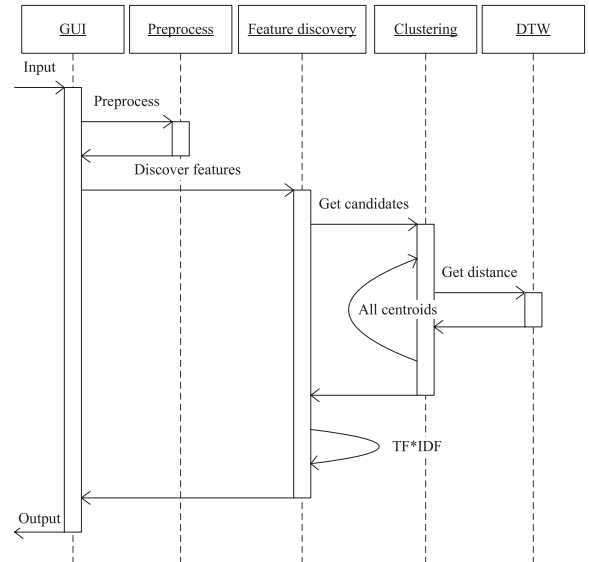


Fig. 4. Feature discovery

$$D(P, Q) = f(p_i, q_j) = \min \begin{cases} f(p_i, q_{j-1}) + s \\ f(p_{i-1}, q_j) + t \\ f(p_{i-1}, q_{j-1}) + u \\ \dots \dots \dots \end{cases} (3)$$

この時、 s と t は時間軸に沿ってシーケンスを伸縮するコストを示す、そして、 u は値の違いによる距離コストを示す。

〈3・2〉 特徴パターン抽出 Fig. 4 に、上述した手法を組み合わせることで時系列データから特徴パターンを抽出するアルゴリズムを示す。入力されたデータはデータクリーニングや部分時系列データに切り出すなどの前処理を行う。次にデータから特徴発見を行う。特徴発見では部分時系列データをクラスタリングして、特徴候補を見つける。このクラスタリングの距離計算のために DTW を用いる。獲得した特徴候補から TF*IDF で特徴パターンを発見する。

〈3・3〉 相関ルール抽出 発見した特徴パターンとアノテーションとの相関ルールを抽出するために、まず、発見した特徴パターンと各アノテーションの出現する時間を基にして長さを計算する。この長さは特徴パターンとアノテーションとの間の期間である。範囲を示すアノテーションの場合には最短の長さを使用する。Fig. 5 に特徴パターン P_1 とアノテーション A_1, A_2 との長さを示す。例に示す通り、特徴パターンの後にアノテーションが出現する場合とアノテーショ

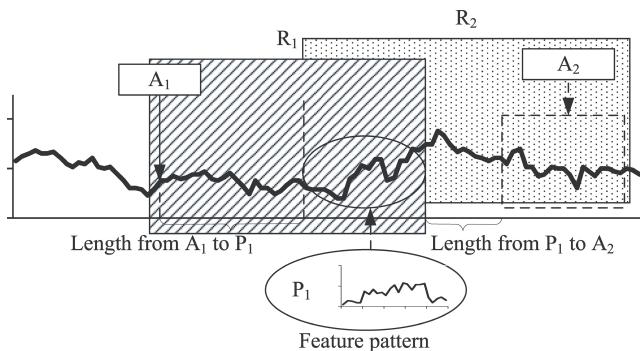


Fig. 5. Length of intervals

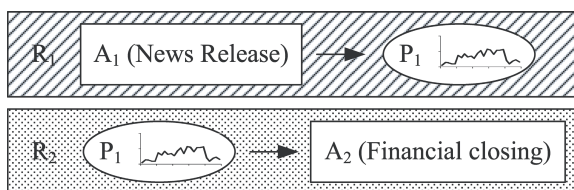


Fig. 6. Example of association rules

ンの後に特徴パターンが出現する場合の2通りを調べる必要がある。

特徴パターンとアノテーションが離れている場合にはその2つのデータには関係がないと判断できる。そこで、最大許容長 D_{\max} を定め、 D_{\max} 以上離れている特徴パターンとアノテーションの間には関係がないものとみなす。これにより特徴パターンとアノテーションとの組み合わせの数を削減できる。このようにして作成したルールはFig. 6のようになる。 R_1 はアノテーション A_1 が出現した後、特徴パターン P_1 が出現する相関ルールを示し、 R_2 は特徴パターン P_1 が出現した場合にアノテーション A_2 が出現する相関ルールを示す。

次に、この相関ルールの信頼度を求める。一般的な相関ルールマイニングではトランザクションデータベースを扱い、 $X \rightarrow Y$ という相関ルールはデータベースのトランザクション内に X が出現したときに同時に Y が出現する確率が高いことを示す。このとき、相関ルールの信頼度 Confidence は $\frac{X \text{ と } Y \text{ を含むトランザクション数}}{X \text{ を含むトランザクション数}}$ で計算される。この方法を用いて、条件部 X が出現する回数を $\# [X]$ 、条件部 X から最大許容長 D_{\max} までの距離に帰結部 Y が出現する回数を $\# [X \cup Y]$ とすると、ルールの信頼度は $\text{Confidence} = \frac{\# [X \cup Y]}{\# [X]}$ で計算できる。ユーザは最小信頼度 C_{\min} を定義し、 C_{\min} 以上のルールのみを抽出することで、信頼できるルールを獲得できる。

4. 分類器の作成

本論文では、発見した知識を用いて時系列データの分類を行う方法についても示す。発見された特徴パターンとアノテーションからなる相関ルールを時系列データ分類の際の着目点として用いてクラス分類を行うために、分類ルールを抽出する。

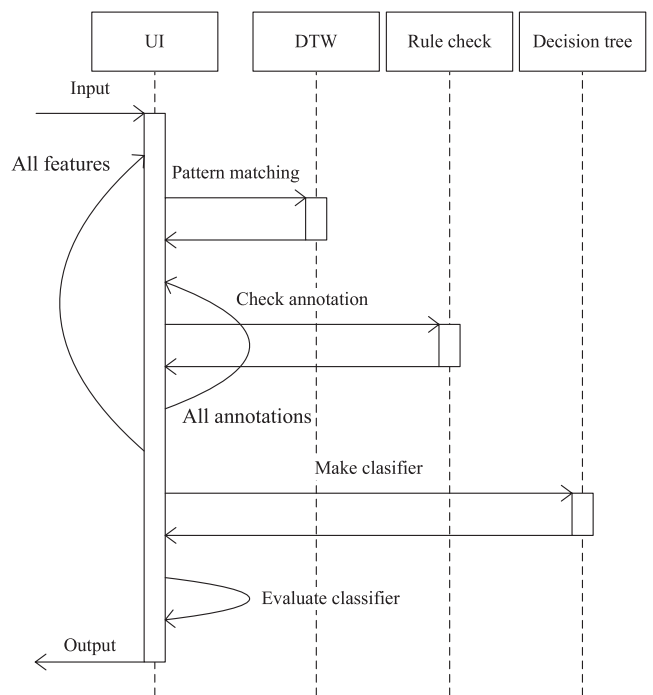


Fig. 7. Classifier extraction

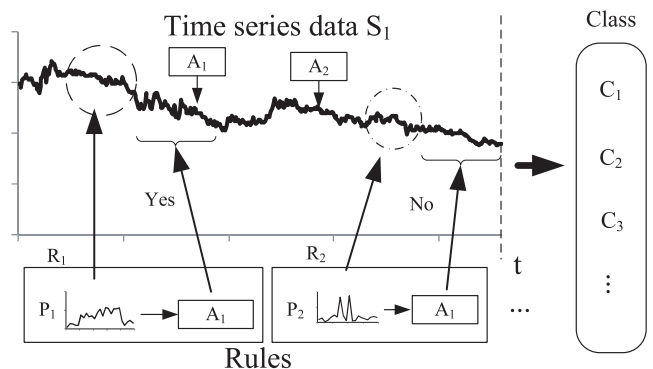


Fig. 8. Making of training data

分類ルールの抽出はFig. 7に示すアルゴリズムで実行される。まず、教師データとなるクラスに関連付けられた時系列データと、特徴パターン抽出で獲得した全ての特徴パターンと相関ルールが与えられる。そして、各特徴パターンと時系列データとのDTW距離を計算する。ルールに用いられている特徴パターンが存在するとき、そのパターンから最大許容長 D_{\max} までのアノテーションを確認し、そのルールが成立するか否かを調べる。ルールが成立している場合に Yes、成立していない場合に No とする。Fig. 8に、このルールの成立を確認する動作の概要を示す。 R_1, R_2 は相関ルールを示し、 P_1 は相関ルール R_1 に使われている特徴パターン、 P_2 は相関ルール R_2 に使われているパターンを示す。Fig. 9にトレーニングデータのインスタンスの例を示す。ルールの成立を確認する動作は、発見された特徴パターンすべてに対して行う。特徴パターンと相関ルールに基づいたトレーニングデータを用いて機械学習を行うことで分類ルールを抽出する。Fig. 10にこのトレーニングデータから作成した決


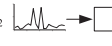
| | | | | | | |
|-------|--------------|---|---------------|----------|--|-------|
| | R_1 | P_1  | \rightarrow | A_1 | | |
| | R_2 | P_2  | \rightarrow | A_1 | | |
| | DTW distance | Existing | DTW distance | Existing | | Class |
| S_1 | 50.0 | Yes | 230.9 | No | | C_1 |

Fig. 9. Example of an training instance

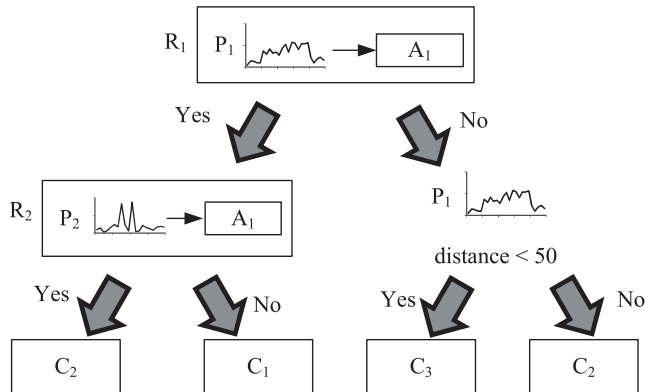


Fig. 10. Example of decision tree

定木の例を示す。

5. 実験

実験データには東京株式市場の25社の2006年1月5日から2006年12月29日までの1年分の株価の終値を用いる。アノテーションは各社のホームページの記事を用いて、決算、発売、お詫びの3種類を人手で付与したところ全合計752個となり、各社平均で決算6.4個、発売21.4個、お詫び0.3個となった。スライドウィンドウは領域知識に基づき不等間隔で行い、サイズ20(1か月)、スライド幅5(1週間)を基準とし、ローソク足に従って陽線・陰線が連続している区間は区切らないように最終的なサイズとスライド幅を決定する。結果的にスライドウィンドウのサイズは最小20, 最大27, 平均20.9となり、スライド幅は最小4, 最大6, 平均4.8データのスライド幅となった。特徴パターンとアノテーションとの相関ルールでは、最大許容長 D_{\max} を20(1か月)、最小信頼度 C_{\min} は0.90とした。決定木はC4.5アルゴリズム⁽¹²⁾を使用し、DTWは時間軸のコストを50.0, 縦軸のコストを $|p - q|$ とする。Fig. 11はクラスタ数 k と抽出された特徴パタンの数を示し、Table 3は獲得した相関ルールの個数を示す。

本実験ではFig. 12に概要を示すように、過去1か月(20営業日)の株価データから、未来1週間(5営業日)の動向を予測するための知識獲得を目的とする。未来の動向については、1年分の実験データ中での1週間の株価変動を集めてk-meansによりクラスタリングすることにより、Fig. 12に示す C_1, \dots, C_5 までの5つのパタンのいずれかとして扱うことにした。クラスタリング手法の性質により、5つの変動パターン(クラスタ)に属する事例数はほぼ均等に近くなる。実際、最大事例数パターンでは556事例、最小では398事例であり、ほぼ20%ずつの等しい分布となっている。

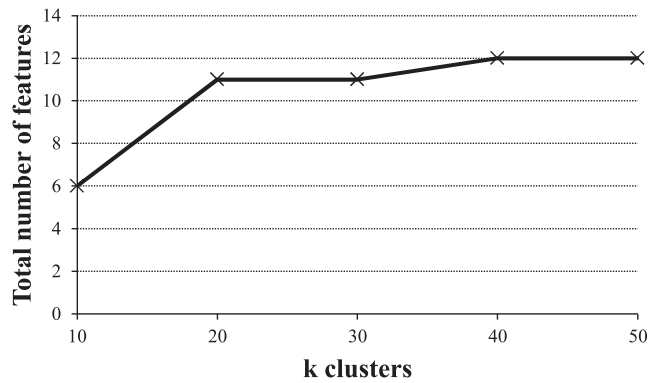


Fig. 11. Total number of clusters

Table 3. Total number of rules

| clusters k | rules |
|-------------|-------|
| 10 clusters | 13 |
| 20 clusters | 18 |
| 30 clusters | 25 |
| 40 clusters | 27 |
| 50 clusters | 23 |

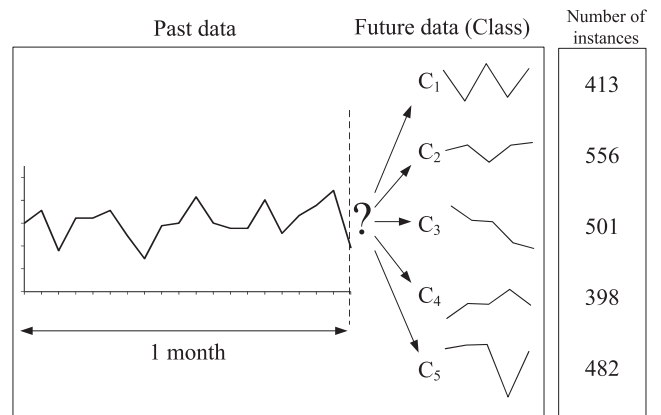


Fig. 12. Past data and future data

したがってランダムな予測の精度も約20%となる。なお、ここで使った5つの変動パターンは、文献(13)(14)から売買に有効利用できるものとなっている。

既存手法との比較のために、部分時系列データの各時間軸のポイントを属性としてみなして機械学習を行う。固定したスライドウィンドウを用いてサイズ20で切り出し、20個の属性を持つトレーニングデータを作成する。このデータを用いて、決定木、ニューラルネットワーク、SVMの3種類の既存手法との比較を行う。ニューラルネットワークでは12層の隠れレイヤを用いて500回の学習を行う。さらに、本研究の手法における比較として、TF*IDFを適用しない場合の分類精度についても比較を行う。精度はすべて10分割交差検定で求める。Table 4に分類精度の結果を示す。

Fig. 11からクラスタ数 k に従って特徴パタンの総数が増加することがわかり、Table 3からクラスタ数 k に従って相関ルールの個数が増加することがわかる。クラスタ数増加とともにデータ解析の詳細度が上がるため、クラスタ数

Table 4. Result of accuracy

| Method | Accuracy |
|------------------------|----------|
| Decision tree | 30.89 % |
| Neural Network | 30.04 % |
| Support Vector Machine | 27.91 % |
| 10 clusters | 26.66 % |
| 20 clusters | 29.53 % |
| 30 clusters | 29.14 % |
| 40 clusters | 31.32 % |
| 50 clusters | 31.45 % |
| 10 clusters + TF*IDF | 34.60 % |
| 20 clusters + TF*IDF | 46.13 % |
| 30 clusters + TF*IDF | 69.96 % |
| 40 clusters + TF*IDF | 65.23 % |
| 50 clusters + TF*IDF | 69.22 % |

k の数に従って、作られる分類器の精度も上昇することが Table 4 からわかる。さらに、TF*IDF 適用前と比べて大きく精度の差が出たことで、特徴的な形状の発見が分類精度向上に大きく貢献できたと思われる。

TF*IDF によって獲得できる知識は、例えば多くの企業に共通して「発売」→「値上がりボタン」というルールが高い信頼度で抽出された。しかし、成長企業は製品発売とは独立に値上がり傾向が強く、単純な値上がりは抽出されずに、一部に値下がりを含むボタンが抽出された。このように成長企業の値下がりの特徴を見いだせることから、本手法は意外性を含む知識発見を成し得ていると考えられる。

6. おわりに

本論文では時系列データにアノテーションを付与したデータから知識を発見する手法について提案した。まず、領域的な知識に基づき部分時系列データを抽出し、特徴パターンとしての重要性を計算する。このために、テキストマイニングでよく用いられる TF*IDF を時系列データに適用する手法を開発した。次に、獲得した特徴パターンとアノテーションとの相関ルールを発見する。最大許容長と最小信頼度という 2 つの指標を基にして、有用な相関ルールを抽出できることを示した。さらに、獲得した特徴パターンと相関ルールを属性として時系列データを分類する手法についても提案した。作成した分類器は、従来の機械学習よりも高い分類精度の知識となる。

今後の課題として、抽出した実際のデータや D_{\max} の決定方法などの設定値について専門家に意見を求めることでさらなる検討を行う予定である。

文 献

- (1) K. Nagao: "Digital content annotation and transcoding", Artech House Publishers (2003)
- (2) I. Kobayashi, C. Watanabe, and N. Okumura: "Intelligent Information Presentation Based on Collaboration between 2D Chart and Text: With an Example of Nikkei Stock Average Text and Its 2D Charts Presentation", *Transactions of Information Processing Society of Japan*, Vol.48, No.3, pp.1058–1070 (2007) (in Japanese)

小林一郎・渡邊千明・奥村奈穂子:「グラフとテキストの協調による知的な情報提示手法: 日経平均株価テキストとグラフの提示を例にして」, 情報処理学会論文誌 インタラクショナル技術の原理と応用特集, Vol.48, No.3, pp.1058–1070 (2007)

- (3) R. Agrawal and R. Srikant: "Mining sequential patterns", In *Data Engineering*, 1995. Proceedings of the Eleventh International Conference on., pp.3–14 (2002)
- (4) D. Rafiei and A. O. Mendelzon: "Querying Time Series Data Based on Similarity", *IEEE Transactions on Knowledge and Data Engineering*, Vol.12, No.5, pp.675–693 (2000)
- (5) A. Kampouraki, G. Manis, and C. Nikou: "Heartbeat Time Series Classification with Support Vector Machines", *Information Technology in Biomedicine, IEEE Transactions on*, Vol.3, No.4, pp.512–518 (2009)
- (6) K. S. Jones: "A Statistical Interpretation of Term Specificity and its Application in Retrieval", *Journal of Documentation*, Vol.28, pp.11–21 (1972)
- (7) T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu: "An Efficient k-means Clustering Algorithm: Analysis and Implementation", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol.24, No.7, pp.881–892 (2002)
- (8) E. Keogh and J. Lin, and W. Truppel: "Clustering of Time Series Subsequences is Meaningless: Implications for Previous and Future Research", *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)*, pp.115–122 (2003)
- (9) R. Fujimaki, S. Hirose, and T. Nakata: "Frequency Analysis of Subsequence Time Series Clustering and Evaluation of Phase Alignment Preprocessing", *Transactions of the Japanese Society for Artificial Intelligence*, Vol. 25, No.4, pp.540–548 (2010) (in Japanese)
- (10) 藤巻遼平・広瀬俊亮・中田貴之:「部分時系列クラスタリングの周波数解析と位相整列処理の評価」, 人工知能学会論文誌, Vol.25, No.4, pp.540–548 (2010)
- (11) 大崎美徳・阿部秀尚・北口真也・横井英人・山口高平:「不等間隔な標準化と量子化を用いた時系列パターン抽出手法の開発と評価」, 人工知能学会知識ベースシステム研究会, Vol.69, pp.39–43 (2005)
- (12) D. J. Berndt and J. Clifford: "Using Dynamic Time Warping to Find Patterns in Time Series", In *Proceedings of KDD-94: AAAI Workshop on Knowledge Discovery in Databases*, pp.359–370 (1994)
- (13) J. R. Quinlan: "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers (1993)
- (14) 天海源一郎:「ズバリわかる!株価チャートの読み方・使い方」, 成美堂出版 (2005)
- (15) 林 康史:「株価が読めるチャート分析入門—売買シグナルはどうつかむか!?初心者からセミプロまで使えるテクニカル分析の決定版」, かんき出版 (2000)

杉 村 博 (学生員) 1981 年 3 月 2 日生。2003 年神奈川工科大学工学部情報工学科卒業後、2 年間 IT 関連の企業に就職。その後、2007 年神奈川工科大学大学院情報工学専攻博士前期課程修了、現在、神奈川工科大学大学院情報工学専攻博士後期課程在籍。主にデータマイニングのシステム設計やアルゴリズムに関する研究に従事。人工知能学会、情報処理学会、IEEE 各会員。



松 本 一 教 (正員) 1961 年生。1984 年九大理学部数学科卒業、1986 年九大総合理工学研究科修士課程修了、1986 年から 2002 年まで (株) 東芝に勤務。理博。現在、神奈川工科大学情報学部情報工学科教授。主にデータマイニングのシステム設計やアルゴリズムに関する研究に従事。人工知能学会、情報処理学会、IEEE 各会員。

