[研究論文]

時系列データクエリ言語と遺伝的アルゴリズムを基にした知識発見システム

杉村博[1]・松本一教[2]

1 博士後期課程情報工学専攻
2 情報工学科

Knowledge Discover System based on a Temporal Query Language and the Genetic Algorithm

Hiroshi SUGIMURA[1], Kazunori MATSUMOTO[2]

Abstract

This paper proposes a knowledge discovery system that aims to predict future behavior of time-series data. The points of this system are twofold. First, we develop a flexible query language, which is named temporal query language, TQL for short. Similar to the regular expression on the text strings, TQL can express a combination of patterns using temporal operators. Second, we develop a powerful data mining algorithm that is based on decision tree learning. The algorithm starts with a given set of patterns, which is called clues, and then express time-series data in terms of the clues. In order to make a better prediction, we develop a mechanism that improves the quality of the clues. The essential idea of the mechanism is based on the genetic algorithm. This paper describes details on the system and results of the experiment.

Keywords: data mining, time-series data, dynamic time warping, decision tree, genetic algorithm

1. はじめに

本研究は時系列データの未来の動きを予測するための知識の抽出を目的としたデータマイニング技術について提案する. 時系列データとは時間の経過に沿って記録したデータで, 実社会においてこのような形式のデータは様々な場面で頻出し, 蓄積されている. これらを解析し未来の動きを予測することで, 近年問題となっているゲリラ豪雨や, 大地震の発生, 経済状態などの変動を予測できると考えられ, 重要なテーマとなっている.

このような時系列データの未来の動きを予測するためにデータマイニングの技術が開発されている. データマイニングとは, 統計学, パタン認識, 機械学習などによってデータベース内にある大量のデータに対して網羅的に処理を行うことで知識を取り出す技術である. 発見的な知識獲得が可能という特徴があり, 様々な技術が研究され, いくつもの研究が成功をおさめている[1, 2, 3].

本研究では決定木学習[4]を用いて知識の抽出を行う. 決定木学習は IF-THEN スタイルの知識を抽出できる利点があり, 成功事例が多数報告されている[5, 6]. しかし, 従来の決定木学習では時系列データを属性として扱えないため, このようなデータを含むデータ集合に対して適用する場合には前処理が必要である. もっとも単純な前処理方法の 1 つとして, 時系列データを計測値の平均値や中央値で置き換える方法が考えられるが, この方法では時系列データの動きを無視しており, 形が大きく異なる時系列データであっても同一のデータとみなされてしまう欠点がある. 時系列データには数値と数値の間に連続的な関係が存在していると見ることができるが, このように数値間のつながりを扱うことができない値への変換を行って決定木学習を適用するような方法では, 真に時系列データに存在する情報のすべてを扱えているとは言えない[7, 8].

山田ら[8, 9]は時系列決定木を提案し，実データによる成功事例を報告しているが, このアルゴリズムは医療などの 1 つの状態に対して多面的な方法で計測したような, 複数の時系列データを属性としてもつデータに対して有効な手法である. 本研究では 1 つの時系列データから知識を抽出することを目的としているため, 単純にこの手法を適用することはできない.

また，杉山ら[10]は長大な時系列データをスライドウィンドウによって部分時系列データとして, それらをクラスタリングして代表シーケンスとした後, 決定木学習を行うことによって時系列データの未来の動きの予測を行っている．しかし，このような部分時系列データに対して k-means などによるクラスタリングを適用した場合に，その中心はサイン曲線のようになるため, このような手法には意味がないという研究もある[11].

そのような背景の中で, 本研究では時系列データを解析するための補助的なパタンを与えることによって, 予測精度の高い決定木を作成する方法を開発した. この方法は入力したパタンと時系列データとの相違度を属性値とする方法を提案している[12, 13, 14]. 本研究ではユーザがシステムに与えた補助的なパタンをクルーと呼ぶ. しかし，この手法ではユーザが良い特徴的なパタンを与える必要があり, 知識のない人間にとってはそのようなパタンを入力することは難しい. そこで本研究では，この特徴を自動的に発見する手法について提案する.

さらに, 本論文ではデータベースから関係する時系列データを検索するための手法についても提案する. 機械学習は与えられたデータすべてを用いて解析を行うが, 与えられるデータの選択は人間によって行われる. このときに選択されたデータに依存して, 得られる知識の有用性が異なる．このため，データ選択は重要な位置づけであるといえる.

データ選択の方法として, ユーザによって手書き入力されたシーケンスに類似する時系列データを検索し, 収集する方法[15]がある．しかし，1 つのシーケンスとの相違度によって検索を行うだけでは, 曖昧さがシーケンス全体での相違度でしか表現できず, 検索の際に指定する相違度の閾値の設定が困難である．また，検索したいシーケンス中の特に重要視する部分を特別に扱うことができない. そこで本研究では複数のシーケンスの組み合わせによって検索をおこなうためのクエリ記述方法として, 時系列データクエリ言語を提案する.

2. 時系列クエリ言語

1 つのシーケンスとその相違度による検索では柔軟な検索ができないため, 複数のシーケンスとそれぞれの相違度によって検索する．Fig. 1 に検索するシーケンスを複数のシーケンスとして分け, 曖昧さを変えた場合の検索結果の違いを示す.

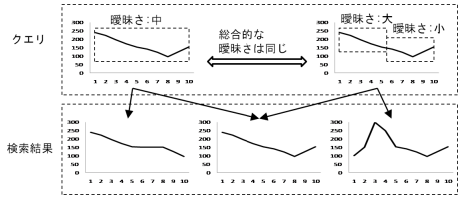


Fig. 1 検索方法の違いによる検索結果の違い

複数のシーケンスの組み合わせによるクエリを表現するために，本研究では時系列データクエリ言語(TQL, Temporal Query Language)を提案する．この言語はシーケンスの出現順序を XPath のような"/"を用いてつなげる経路表現として記述し, 出現距離はワイルドカードを用いて正規表現のように記述する．Table 1 にクエリの構文を, Table 2 にワイルドカード表現の概要を示す.

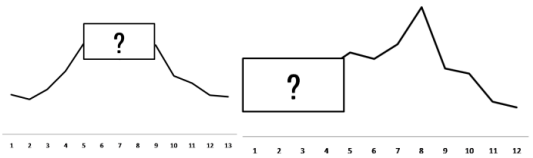
Table 1 時系列データクエリの構文

Query	=	Path | Path Op Query
Path	=	Pattern | Pattern “/” Path
Pattern	=	String | Wildcard
Op	=	“and” | “or”
Wildcard	=	Table 2 を参照

Table 2 ワイルドカード表現

.	何か 1 つのデータにマッチ
*	0 以上の連続にマッチ
+	1 以上の連続にマッチ
?	0 か 1 個にマッチ
{n, m}	n 個以上，m 個以下の連続にマッチ

TQL を用いることでシーケンスの出現順序と出現距離を定義して柔軟な検索が可能となる. たとえば Fig. 2 の(a)に示すように上昇した値が下降するに至るまでの中間部分にはどのような関係や頻出シーケンスがあるか発見するために, または Fig. 2 の(b)に示すようにピークになる前にはどのような共通的な形状があるのかを調査するために用いることができる.



(a) (b)

Fig. 2 柔軟な検索

Fig. 3 に 2 つのシーケンスを定義して TQL を記述し, 検索を行った結果の例を示す.

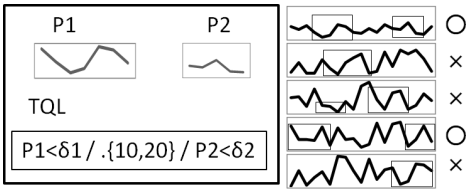


Fig. 3 TQL による検索結果

このクエリは「P1 に類似したデータが出現した後，P2 と類似したデータが出現するデータを検索する, このとき

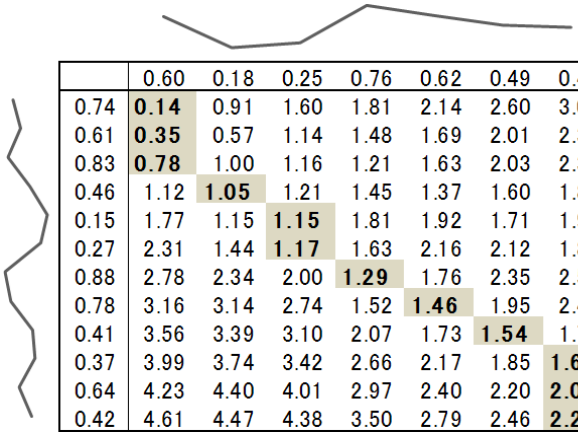
P1 と P2 の間には 10 から 20 個の何かしらのデータが出現している」ということを意味している．クエリに記述した δ1，δ2 は P1，P2 に対する曖昧さを表しており，各シーケンスと検索データとのマッチを行う際に，類似していると認識するための閾値である．相違度の計算方法は次節で説明する．

3． パタンマッチング

システムはパタンマッチングによってシーケンスと時系列データとの相違度を計算する．もっとも簡単な計算手法の 1 つにユークリッド距離がある．この手法は計測値数が異なる場合に適用できず，時間軸のずれを許容できないため，人間の直感に反する結果を生じてしまう場合がある[9]．そこで本研究では Dynamic Time Warping (DTW) [16] によって相違度の計算を行う．

DTW は，パタンの要素間に定義された相違度に基づいて，パタンの伸縮まで考慮に入れたマッチング方式である．マッチングの際に必要なプロパティは時間軸のずれに対応したコストと，値の一致度に対応したコストである．ここで計算した距離が TQL によって入力された閾値以下のときに，パタンにマッチしたと認められる個所となる．

時系列データ $A = a_1, a_2, ..., a_i$ と $B = b_1, b_2, ..., b_j$ 間の DTW に基づく距離 $G(A,B)$を求めることを考える．このとき A, B の対応付けをワーピングパスと呼び，距離は ij 平面上の格子として考える．Fig. 4 にワーピングパスの対応を行った例を示す．時間軸のずれのコストを q, r，値の不一致に対するコストを s とした場合，距離 G は式(1)から求められる．



	0.60	0.18	0.25	0.76	0.62	0.49	0.46
0.74	**0.14**	0.91	1.60	1.81	2.14	2.60	3.09
0.61	**0.35**	0.57	1.14	1.48	1.69	2.01	2.37
0.83	**0.78**	1.00	1.16	1.21	1.63	2.03	2.38
0.46	1.12	**1.05**	1.21	1.45	1.37	1.60	1.80
0.15	1.77	1.15	**1.15**	1.81	1.92	1.71	1.90
0.27	2.31	1.44	**1.17**	1.63	2.16	2.12	1.89
0.88	2.78	2.34	2.00	**1.29**	1.76	2.35	2.52
0.78	3.16	3.14	2.74	1.52	**1.46**	1.95	2.48
0.41	3.56	3.39	3.10	2.07	1.73	**1.54**	1.78
0.37	3.99	3.74	3.42	2.66	2.17	1.85	**1.63**
0.64	4.23	4.40	4.01	2.97	2.40	2.20	**2.01**
0.42	4.61	4.47	4.38	3.50	2.79	2.46	**2.23**

Fig. 4 ワーピングパス

$$G(A,B) = g(i,j) = \min \begin{cases} g(i,j-1) + q \\ g(i-1,j) + r \\ g(i-1,j-1) + s \end{cases} \quad (1)$$

4． 知識抽出

本研究で作成するシステムの概要を Fig. 5 に示す．

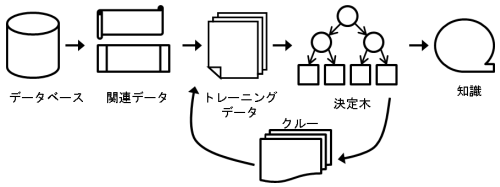


Fig. 5 システムの概要

システムはデータベースから TQL によって関係データを収集する．関係データは標準化され，クルーとの相違度を基にしたトレーニングデータに変換する．このトレーニングデータへの変換については後述する．トレーニングデータから決定木学習によって知識を抽出し，評価を行う．トレーニングデータへの変換から知識抽出，評価までの操作をクルーの自動改良を行いながら終了条件を満たすまで繰り返す．

4.1 トレーニングデータ

決定木学習で時系列データを扱うために，時系列データの特徴を表すための補助的なパタンであるクルーとの相違度によるトレーニングデータに変換する．属性はクルー，属性値は時系列データとクルーとの相違度とする．この相違度の計算には前述した DTW を用いる．さらに未来の動きによってクラスを付与する．この方法によって決定木学習で時系列属性を扱うことができるようになる．作成したトレーニングデータの例を Fig. 6 に示す．

Time-series \ Clues	C1	C2	C3	Class
S1	7.49	8.24	7.77	Up
S2	7.81	7.77	8.91	Down
S3	7.29	8.40	8.85	Down

Fig. 6 トレーニングデータ

4.2 特徴の発見と改良

決定木の予測精度は属性として与えるクルーに依存する．属性として与えるクルーは人間が与えることで高い予測精度の決定木を作成できることが先行研究によって分かっているが[12]，この方法では知識を持たないユーザにはよいクルーを与えることができない．そこで本研究では属性として与えるクルーを自動的に発見し，改良も行うことによって高い予測精度の決定木を作成する．このクルーの発見と改良のメカニズムには遺伝的アルゴリズムを用いる．遺伝的アルゴリズムでは遺伝子を評価し，遺伝子を操作することによって最適解を求めていく．

まず，乱数を用いて遺伝子となるクルーを複数生成して初期遺伝子集団とする．次に各遺伝子の適合度を計算する．選択では適合度の高い遺伝子を残し，適合度の低い遺伝子を削除する，交差では適合度の高い遺伝子 2 つを掛け合わせて新しい子の遺伝子を生成する，突然変異では遺伝子の一部をランダムに変化させる．これら 3 種のオペレータに

よって遺伝子集団を改良する. システムは終了条件を満たすまでこの操作を実行する．Fig. 7 に遺伝的アルゴリズムの概容を示す.

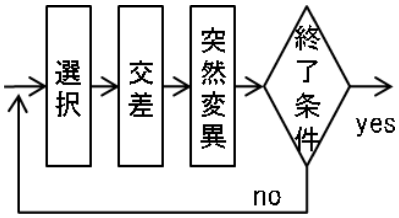
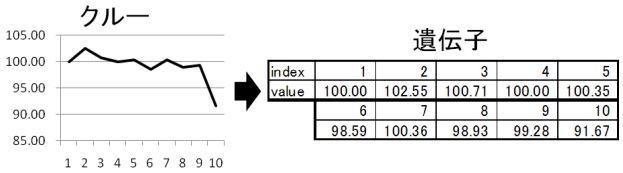


Fig. 7 遺伝的アルゴリズムの概要

4.3 クルーの遺伝子表現

本研究では，1 つの遺伝子が 1 つのクルーに対応している. 遺伝子は数値のシーケンスによって, 時系列データで注目すべきグラフ形状を表す．それぞれの数値は 1 つ前のデータからの変化率を示す. 前データのない最初の値は固定値の 100%とする．このようにして，システムは入力されたすべての時系列データに対してクルーを統一的に扱うことができる. Fig. 8 に 1 つのクルーに対応する遺伝子表現の例を示す.



index	1	2	3	4	5
value	100.00	102.55	100.71	100.00	100.35
	6	7	8	9	10
	98.59	100.36	98.93	99.28	91.67

Fig. 8 クルーの遺伝子表現

4.1 遺伝子の適合度

遺伝子の適合度には, 対応するクルーとの相違度によってトレーニングデータを分類した際の獲得情報量 Gain(X) を用いる. 獲得情報量は式(2)で計算できる．ここで，T は分割前の集合であり，ある分割によって n 個の部分集合 $T_i(1 < i < n)$ 個に分割されるものとする

$$\mathrm{Gain}(X) = \mathrm{info}(T) - \mathrm{info}_x(T) \tag{2}$$

$$\mathrm{info}_x(T) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times \mathrm{info}(T_i) \tag{3}$$

info(S) は情報エントロピーと呼ばれる量で，式(4)で計算できる．S は集合を示し，$\mathrm{freq}(C_j, S)$ はクラス C_j に属する S 内の要素の数である．k はクラスの数である.

$$\mathrm{info}(S) = -\sum_{j=1}^{k} \frac{\mathrm{freq}(C_j, S)}{|S|} \times \log_2 \frac{\mathrm{freq}(C_j, S)}{|S|} \tag{4}$$

4.2 選択

本研究の選択ではルーレット選択を用いる. ルーレット選択とは, 適応度に比例した割合で選択確率を変えて遺伝子を選択する方法で，John Henry Holland が最初に提案したときに使われた最も基本的な選択方式である[17]. ルーレット選択を用いることにより, 適応度の高い遺伝子ほど選ばれる確率が大きくなるが, 適応度の低い遺伝子でも次世代の遺伝子として選択される可能性が残される. これにより, 適応度が高い遺伝子だけが残ることによって局所解にとらわれやすくなるのを防ぐ. 本研究では選択を次の手順で行う.

(1) 前節で計算した各遺伝子の適合度を基にして，ルーレット選択のスケーリングを行う.

(2) ルーレット選択によって選択した遺伝子を次世代に残し，ルーレットから削除する.

(3) 手順 1，2 を繰り返し，あらかじめ定めておいた数の遺伝子を次世代に残す.

4.3 交差

交差とは適応度の高い 2 つの親の遺伝子を合成して，新しい子の遺伝子を作る操作である. 交差方法には一点交差，二点交差，一様交差などの手法がある．本研究では二点交差を用いて次世代の遺伝子を作成する. 二点交差は交差点をランダムで 2 か所選び，2 つの交差点に挟まれている部分を入れ替える手法である．二点交差によって，遺伝子総数があらかじめ定められた数になるまで新しい遺伝子を作成する．2 つの遺伝子が二点交差を行う様子を Fig. 9 に示す.

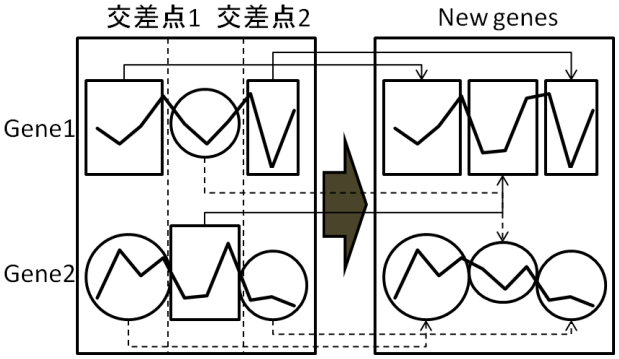


Fig. 9 二点交差

4.4 突然変異

突然変異は, ランダムに選択された遺伝子の一部をある確率でランダムに変化させる操作である. 突然変異確率が大きすぎるとランダムサーチに近くなるが, ある程度の変異は局所解を避けるために必要である. 交差だけでは親となる遺伝子の性質しか伝わらないために、場合によってはあまり適応度が高くないにもかかわらず同じような遺伝子ばかりが生き残る．このような場合に，それ以降の適合度の向上が見られなくなるため, 突然変異によるランダムな変化が重要となることがある．本研究では，遺伝子をランダムに選択し, その遺伝子の一部の数値に対してあらかじめ定められた範囲の乱数を加算, もしくは減算すること

で行う．Fig. 10 にこの操作の概要を示す．

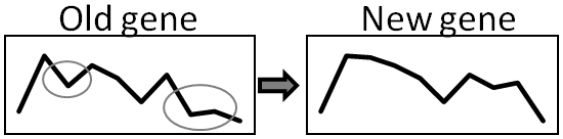


Fig. 10 突然変異

5． 実装と実験

実際にシステムを作成し，マイニングを行う．ユーザはシステムにクエリを入力する．システムはクエリに従い時系列データを検索し，収集する．収集したデータを前データとの比にすることで標準化する．標準化したデータに分類を付与してトレーニングデータを作成する．このトレーニングデータに対して遺伝的アルゴリズムと決定木学習によって特徴の発見と改良を行いながらパタンを抽出する．最後に，抽出したパタンの評価を行い，評価結果とともに出力する．

未来の動きは up, down，そして stay の 3 種類によって分類する．この 3 種類は未来の観測期間で期待できる変化量を x%として，Table 3 に示すように分類を行う．

Table 3 分類方法

Up	観測期間で x%以上の上昇が期待できる場合
Down	観測期間で x%以上の下降が期待できる場合
Stay	Up と Down の両方に含まれない場合

5.1 検索手法の違い

従来の手法によって類似データを収集して未来の動きの予測を行った場合と，提案したクエリである TQL を用いて収集したデータから未来の動きの予測を行った場合とを比較した結果を Table 4 に示す．

Table 4 検索手法の違いによる精度

	類似検索	TQL
検索数	212	103
予測精度	40.09%	48.54%

5.2 従来手法との比較

従来手法として，単純に部分時系列データの各時刻の値を属性とした決定木と，杉山らが提案している属性の特徴をクラスタリングによって求める手法によって実験を行い，本研究の提案手法とを比較する．

杉山らが提案している手法では文献[10]にならい，データの収集とパラメータ設定を行った．すなわち，株価データはカブロボから取得した 2006 年 1 月 4 日の前場から 2006 年 12 月 29 日の後場までの 25 社のデータを用いる．属性には 17 種類のテクニカル指標を使い，各テクニカル指標の算出に必要なパラメータ値は一般的に用いられていることの多い値とし，代表パタン数は 4 とした．

本論文の手法では初期遺伝子として 1 つのクルーに対して 50.0 から 150.0 の幅で 10 個の乱数を発生させて作成した．この遺伝子が 5 個，10 個，20 個からなる 3 種類のセットを用意して実験を行う．遺伝的アルゴリズムによる改良の様子を調べるために，改良後の結果は「GA:」というプレフィックスをつけてあらわす．

Table 5 に 3 種類の手法によって作成した決定木を比較した結果を示す．数値はすべて 25 社の平均値である．また，予測精度の安定性を調べるために数値の分散を計算している．

Table 5 各手法による実験結果

手法	決定木サイズ	予測精度(%)	分散
単純手法	22.52	69.15	97.69
杉山手法	14.92	70.32	80.84
5 genes	20.68	60.47	66.29
GA: 5 genes	28.96	67.05	20.06
10 genes	34.44	67.09	10.49
GA: 10 genes	37.45	70.82	9.97
20 genes	33.48	67.00	14.27
GA: 20 genes	36.35	72.29	6.76

5.3 気象データによる実験

データはアメリカ合衆国 NOAA の National Climatic Data Center の Global Historical Climate Network(GHCN) [18]の世界の気象データ集から日本の観測点 10 か所のデータを抜き出した．1959 年 6 月 1 日から 2009 年 5 月 31 日まで 50 年間の平均気圧を用いる．杉山らの手法は株価データの指標を属性として用いるため気象データによる実験では取り扱わない．このため気象データでの実験では単純手法と本研究による手法とを比較する．Table 6 に比較結果を示す．

Table 6 平均気圧による実験結果

	予測精度(%)	分散	処理時間(m)
単純手法	99.79	0.01	0.15
5 genes	99.77	0.00	2.15
GA: 5 genes	99.78	0.00	7.50
10 genes	99.79	0.00	2.35
GA: 10 genes	99.81	0.00	14.50
20 genes	99.79	0.00	2.83
GA: 20 genes	99.83	0.00	28.30

6． 考察

精度について従来手法と比較を行う．Table 5 から，単純手法による予測精度は 69.15%に対し，杉山手法では 70.32%，本研究で提案する手法は GA: 5 genes で 67.05%, GA: 10 genes で 70.82%, GA: 20 genes で 72.29%であった．この結果から遺伝子 5 個程度では単純手法よりも予測精度が悪くなるが遺伝子数の増加に伴って精度は上昇し，遺伝子数 20 個で杉山手法よりも高い予測精度になることが

分かる．また，分散を見てみると単純手法では 97.69 に対し, 杉山手法は 80.84, 本研究で提案する手法は GA: 5 genes で 20.06，GA: 10 genes で 9.97，GA: 20 genes で 6.76 と安定性が高いことが分かる．また，GA 有りと無しの場合を比較すると 20 genes では 14.27 であったのに対して GA: 20 genes では 6.76 と精度の安定性が上昇していることが分かる．

処理時間については Table 6 から，単純な手法よりも多くの時間を必要とすることが分かる. これは遺伝的アルゴリズムによって何度も決定木を作成する必要があるためである. 今回の終了条件は 100 回の遺伝子改良を条件に行ったが, 一定以上の予測精度を条件にすれば処理時間を緩和できると考えられる．

全体の考察として, 本研究で提案した手法は従来手法と比べて安定して高い予測精度の決定木を得られることが分かった．しかし，処理速度の点でリアルタイム性を要求するような状況で用いることは難しいと思われる. このような性質から, 例として株価予測を取り上げた場合にデイトレードと呼ばれる短期的に何度も取引を行うような目的には適用しにくく，中期・長期的な取引を目的とした場合において有効に利用できる技術になると考えられる．

7. おわりに

本研究ではクルーを用いて時系列データの未来の動きを予測するための知識抽出方法について提案した. クルーを用いることによって時系列データの特徴であるデータの関係を扱うことができる．

さらに, このクルーを遺伝的アルゴリズムによって改良することによって予測精度とその安定性を高める手法についても提案した．

また, データベースからの効率的なデータ収集方法として時系列データクエリ言語である TQL を提案した．この TQL を用いることによって 1 つのシーケンスとの類似時系列データを収集して未来の動きを予測するよりも, いくつかのシーケンスの組み合わせによって柔軟に関係データを検索できる. このため適切にデータを獲得することができ，知識獲得に大きく貢献できると期待される．

参考文献

[1] Guozhu Dong and Jian Pei: Sequence Data Mining, Springer, (2007).

[2] Ian H. Witten and Eibe Frank: Data Mining: Practical Machine Learning Tools And Techniques, Morgan Kaufmann Pub, (2005).

[3] Mark Last, Abraham Kandel and Horst Bunke: Data Mining In Time Series Databases, World Scientific, (2004).

[4] J. Ross Quinlan: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993).

[5] Kazuto Kubota, Akihiko Nakase, Hiroshi Sakai and Shigeru Oyanagi: Parallelization of Decision Tree Algorithm and its Performance Evaluation, IPSJ SIG Notes, Vol. 99, No. 66, pp. 161-166, (1999).

[6] Rakesh Agrawal, Tomasz Imielinski and Arun Swami: Database Mining: A Performance Perspective, IEEE Trans. on Knowledge and Data Engineering, Vol. 5, No. 6, pp. 914-925, (1993).

[7] 大桃諭, 陳漢雄, 古瀬一隆, 大保信夫：タイムワーピングに基づく時系列データの類似検索- 次元縮小による効率化,DBSJ Letters, Vol. 4, No. 1, pp. 17-20, (2005).

[8] 山田悠，鈴木英之進，横井英人，高林克日己: 時系列決定木による分類学習，第 17 回人工知能学会全国大会論文集, (2003).

[9] 山田悠，鈴木英之進，横井英人，高林克日己: 動的時間伸縮法に基づく時系列データからの決定木学習，IPSJ SIG Notes.ICS, Vol. 2003, No. 30, pp. 141-146, (2003).

[10] 杉山喜昭，平林悟，阿部秀尚，山口高平：時系列パターン抽出に基づく個人投資家意思決定支援システムの実現，第 22 回人工知能学会全国大会論文集, (2008).

[11] Eamonn Keogh, Jessica Lin and Wagner Truppel: Clustering of Time Series Subsequences is Meaningless: Implications for Previous and Future Research, in Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03), pp. 115-122, (2003).

[12] 杉村博，松本一教: ユーザ入力にもとづいた時系列データマイニングシステム，第 23 回 人工知能学会 全国大会論文集, (2009).

[13] Hiroshi SUGIMURA, Kazunori MATSUMOTO: Hints Driven Knowledge Discovery and Management, Proceedings of the iadis international conference information systems 2010, (2010).

[14] Hiroshi SUGIMURA, Kazunori MATSUMOTO: Datamining Tool with Exploratory Search and Feature Discovery, Proceedings of the iadis international conference intelligent systems and agents 2010, (2010).

[15] Martin Wattenberg: Sketching a graph to query a time-series database, CHI '01 extended abstracts on Human factors in computing systems, pp.381-382, (2001).

[16] Donald J. Berndt and James 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, Seattle, Washington, (1994).

[17] John H. Holland: Adaptation in Natural and Artificial Systems, University of Michigan Press, (1975).

[18] National Environmental Satellite and Data and Information Service: NNDC Climate Data Online, http://www.nesdis.noaa.gov/, (2009).