

パターン形状のファジィ抽出による時系列データの予測

宮城 雅夫* 宮城 隼夫**

A Prediction of Time-series Data by Fuzzy Sampling of Pattern Form

Norio MIYAGI and Hayao MIYAGI

Abstract : The field of forecasting is considered as an application of time-series analysis even if the data is linear or nonlinear. To obtain the forecasted values from observed data exerts a big influence on the decision-making support system of the control of machine etc. The nonlinear data appear as the random enumerated data. However we sometimes find that the pattern of past appearance repeats itself when we try to observe these data locally. From this point of view, we propose a way of forecasting nonlinear data from the pattern of past appearance using fuzzy theory. The advantages of the method are that we can forecast the next data by small numbers of previous data, and react to some differences, considering the ambiguous nature of the given data.

Keywords : Fuzzy forecasting, Series of the nonlinear data

1. 緒言

時系列解析の目的の一つは予測であり、観測されたデータから予測値を得ることは人間の意思決定や機械の制御などに多大な影響を与える。どのような予測を行う際にも、何の手掛かりもなく予測することはできない。手掛かりのほとんど全ては過去の経験の中に求められるが、最も確実に扱い易いのは時系列に整理された過去のデータである。

ファジィ推論は、if~then 形式でルールを記述するルールベース型であり、条件部が完全に一致しなくても結論を導き出せるという特長を持ち、2値論理では不可能なことが可能になると言われている⁽¹⁾。このファジィ推論を予測問題に用いると、ある程度の予測は可能になる。しかしながら、人間の意思などが複雑に絡み合った時系列データに関して予測を行う場合、ルールを厳密に記述しなければならないため、ルール数が膨大になってしまう。さらにデータ毎にデータ特有の特徴が異なるために、ルールをデータ毎に作成し直

さなければならないという問題点がある。

また、実測値の時系列データを基に、その背後にある原則性を調べる解析方法の一つであるARモデルをファジィ化した例もある⁽²⁾⁽³⁾⁽⁴⁾が、自己相関係数の値が1以上あるいは-1以下になるため、 α -cutによる調整を必要としている。このため、結果には曖昧さが増す傾向がある。

一方、複雑な時系列データでも局所的に観察してみると、過去に現れたデータパターンが再び現れている場合がある⁽⁵⁾⁽⁶⁾。本研究では、この点に着目し、時系列データの予測値を過去の出現パターンから求めるファジィ抽出法を提案する。提案するファジィ抽出法とは、過去に現れたデータをデータパターンに分割し、過去に描いた軌道の中で、現在描いている軌道に最も似たデータパターンをmax・minファジィ合成演算を用いて探すというものである。本手法において、扱うデータは具体的な実測値ではなく、データの変化量を適用する。それは、一般的に差分をとると不規則変動のみが階差系列に含まれるという特長を活かすためであり、また、データの値そのものの並びを一つのパターンとするのではなく、値はどうであれ変化の形状を一つのパターンとみなすためである。本手法によれば、どのような時系列データに対しても一つの予測手順を確立することができる。

*機械工学科

College of Technology, Dai-Ichi University

**琉球大学工学部 (〒903-0129 沖縄県中頭郡西原町字千原1)

Faculty of Engineering, Ryukyu University

2. 諸定義

2.1 ファジィ集合

全体集合 X 上のファジィ集合 A は、次の式(2.1)にて定義されるメンバーシップ関数 μ_A によって特性づけられた集合である⁽¹⁾⁽⁶⁾⁽⁹⁾。

$$\mu_A : X \rightarrow [0, 1] \tag{2.1}$$

メンバーシップ関数 $\mu_A(x)$ は A における要素 x の帰属度を表す。このとき $\mu_A(x)$ の値が 1 に近ければ、 x の A に属する度合いが大きく、反対に 0 に近ければ、 x の A に属する度合いが小さいことを示している。

2.2 ファジィ集合演算

2つのファジィ集合 A, B の共通集合 $A \cap B$ は式(2.2)のように定義される。

$$A \cap B \Leftrightarrow \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \tag{2.2}$$

また、2つのファジィ集合 A, B の和集合 $A \cup B$ は式(2.3)のように定義される。

$$A \cup B \quad \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \tag{2.3}$$

ファジィ理論では、あるファジィ量 a と b の \min あるいは \max を表す演算として、式(2.4)で表す略記法が用いられる。

$$\min(a, b) = a \wedge b, \quad \max(a, b) = a \vee b \tag{2.4}$$

2.3 ファジィ数

任意の $x_1 \in X, x_2 \in X$ と $0 \leq \lambda \leq 1$ であるような任意の λ に対して、式(2.5)が成立するとき、ファジィ集合 A は凸であるという。

$$\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_2)) \tag{2.5}$$

また、実数直線 R^1 上で定義された正規かつ凸ファジィ集合で、特にメンバーシップ関数が区分的に連続なものをファジィ数という。

「だいたい m ぐらい」というファジィ数 M のメンバーシップ関数 $\mu_M(x)$ の例としては、三角型メンバーシップ関数

$$\mu_M(x) = \max(0, 1 - |x - m|/a) \quad a > b \tag{2.6}$$

あるいは釣鐘型のメンバーシップ関数

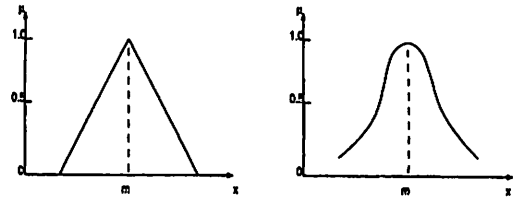
$$\mu_M(x) = e^{-bx - m} \quad b \geq 1 \tag{2.7}$$

などがよく用いられる。式(2.6), (2.7)のメンバーシップ関数の例を図1に示す。

2.4 ファジィ関係

直積 $X \times Y = \{(x, y) | x \in X, y \in Y\}$ におけるファジィ関係 R は(2.8)で示すメンバーシップ関数 μ_R によって特性づけられたファジィ集合であり、メンバーシップ関数 $\mu_R(x, y)$ は x と y の関係の度合いを表す。

$$\mu_R : X \times Y \rightarrow [0, 1] \tag{2.8}$$



(1)三角型のメンバーシップ関数 (2)釣鐘型のメンバーシップ関数

図1. 三角型と釣鐘型のメンバーシップ関数の例

$X \times Y$ におけるファジィ関係 R と、 $X \times Y$ におけるファジィ関係 S に対して、 R と S の $\max \cdot \min$ 合成は $X \times Z$ におけるファジィ関係となり、これを ROS と書き表すと、式(2.9)のようになる。

$$ROS \Leftrightarrow \mu_{ROS}(x, y) = \max_y \min(\mu_R(x, y), \mu_S(y, z)) \\ = \bigvee_y \mu_S(x, y) \wedge \mu_T(y, z) \tag{2.9}$$

2.5 相関係数

母集団から抽出された標本は

$$x^1, x^2, \dots, x^N$$

という数値からなる。ただし、 N は標本の大きさを表す。

大きさ N の標本を特徴づける数値として、標本平均値は式(2.10)で定義される。

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x^i \tag{2.10}$$

また、標本平均値の回りにばらつく尺度として、標本分散値が式(2.11)で定義される。

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x^i - \bar{x})^2 \tag{2.11}$$

また、式(2.11)の平方根 s を標本標準偏差値という。

組なったデータ

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

において、 $|x_i|$ と $|y_i|$ のばらつき方の関係を見る量として、それぞれの平均からのずれを掛け合せたものの平均、すなわち

$$\tilde{s}_{xy} = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y} \tag{2.12}$$

を共分散という。ただし、 x と y はそれぞれ $|x_i|$ と $|y_i|$ の平均である。また、相関係数 r_{xy} は式(2.13)によって定義される。

$$\frac{\tilde{s}_{xy}}{\tilde{s}_x \tilde{s}_y} \tag{2.13}$$

また、式(2.13)において \tilde{s}_x と \tilde{s}_y はそれぞれ $|x_i|$ と $|y_i|$ の標準偏差である⁽⁶⁾⁽⁷⁾。

3. ファジィ抽出法

3.1 ファジィ推論

通常、2値論理における推論の方法は、「AならばB」という含意を用いて行われる。「Aならば、Bが真」であるとき、「Aが真」ならば「Bが真」であることを推論するもので、つぎのように記述される。

$$((A \rightarrow B) \text{ and } A) \rightarrow B \quad (3.1)$$

一方、ファジィ理論を用いた予測は、ファジィ推論を用いて行う場合が多く、ファジィ制御、ファジィエキスパートシステム、ファジィ診断などの分野で重要な役割を果たしてきている。 x, y を対象名、 A, A' ; B, B' をそれぞれ全体集合 U ; V におけるファジィ集合を表すものとすれば、一般のファジィ推論は if~then ルールを用いて次のように記述される⁽¹⁾。

前提1 If x is A then y is B

前提2 x is A'

結論 y is B'

A, A', B, B' はファジィ集合であり、 A と A' 、 B と B' は必ずしも一致するわけではない。すなわち、ファジィ推論は条件部が完全に一致しなくても結論を導き出せるという特徴を持つ。2値論理では不可能なことがファジィ推論では可能であるという有効性が伺える。

また、ファジィ推論の一般化として、ファジィ制御やファジィ診断などにおいてよく用いられている多重ファジィ推論形式は次のように表現される⁽¹⁾。

前提1 If x is A_1 and y is B_1 then Z is C_1 else

前提2 If x is A_2 and y is B_2 then Z is C_2 else

⋮

前提 n If x is A_n and y is B_n then Z is C_n else

前提 $n+1$ x is A' and y is B'

結論 Z is C'

しかしながら、このような if~then ルールを用いて時系列データの予測を行う場合、これらの if~then ルールを厳密に記述しなければならないため、膨大な数の if~then ルールが必要になる。また存在する様々な時系列データに対し、データ特有の特徴は異なるために、それぞれの時系列データ毎にルールを構成し直さなければならない。このような欠点から、汎用性に欠けると考えられる。

そこで、本手法では、過去の出現パターンをファジィ理論を用いてデータパターン化し、max・min 演算により適合度を求めて、予測を行うファジィ抽出法を提案する。過去の出現パターンから予測を行うことは、

人が過去の経験から予測を行うことに相当すると考えられることができる。

3.2 前準備

前準備として以下の変数を定義する。ただし T は時刻を表す。

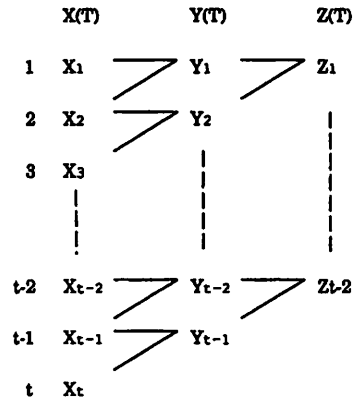


図2. データ X_t, Y_t, Z_t の概念図

- ・ $X(T)$: (実測値) \hat{X}_t : 時刻 t の予測値
- ・ $Y(T)$: (X_t の差分) \hat{Y}_t : 時刻 t の予測値
- ・ $Z(T)$: (Y_t の差分) \hat{Z}_t : 時刻 t の予測値

データ X_t, Y_t, Z_t のデータの概念を図に表すと、図2のようになる。

すなわち、 Y_{t-1} は式(3.2)で、 Z_{t-2} は式(3.3)でそれぞれ定義される。

$$Y_{t-1} = X_t - X_{t-1} \quad (3.2)$$

$$Z_{t-2} = Y_{t-1} - Y_{t-2} \quad (3.3)$$

よって、予測値 \hat{X}_{t+1} は次の式(3.4)、式(3.5)から求めることができる。

$$\hat{Y}_t = \hat{X}_{t+1} - X_t \quad (3.4)$$

$$\hat{Z}_{t-1} = \hat{Y}_t - \hat{Y}_{t-1} \quad (3.5)$$

さらに、式(3.4)、式(3.5)を変形すると、式(3.6)になる。

$$\hat{X}_{t+1} = X_t + Y_{t-1} + \hat{Z}_{t-1} \quad (3.6)$$

予測値を求めるために、過去の時系列データから類似したパターンを捜し出すという方法は、図3に示すように、過去の時刻 $n' (< n)$ でデータを調べ、観測データ x_{n-1} に等しい($x_{n'-1} = x_{n-1}$)あるいは近い($x_{n'-1} \cong x_{n-1}$)データを探す。類似したパターンが見つければ、 $x_{n'-1}$ の次の x_n を用いて $\hat{x}_n = x_n$ と予測値が決定できることになる。

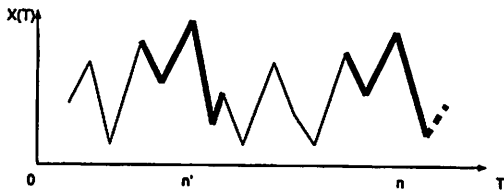


図3. 類似パターンの概念図

提案するファジィ抽出法は、どのような時系列データにおいてもデータを局所的に見ると、過去の出現パターンが再び現れるということを基本にしている。過去の出現パターンを基にメンバーシップ関数を作成し、過去の出現パターンと現在の出現パターンとがどれくらいマッチしているかという適合度を max・min 合成演算により求める。これは予測値が最も適合度が高いパターンの軌道のように描くはずであるという仮定に基づいている。

3.3 ファジィ抽出法の予測手順及び適用例

提案するファジィ抽出法は次の1~7の手順で行われる。

1. 実測値から差分データを作成
2. 差分データからさらに差分データを作成
3. データパターン作成
4. メンバーシップ関数作成
5. 適合度の計算
6. 適合度が最も大きいデータパターンの選択
7. 予測値と信頼度の決定

提案するファジィ抽出法は具体的な数値例を用いて説明する。用いるデータは表1で示す時系列データであり、この時系列データから、第16番目の値 \hat{X}_{16} を予測する。表中の t は時刻を表し、 X_t はその時刻の値を表す。また、表1のデータをグラフに表すと、図4になる。

表1. 与えられた時系列データ

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X_t	1	7	6	8	3	4	2	6	5	7	2	7	6	9	3	\hat{X}_{16}

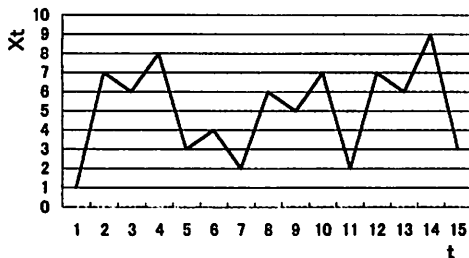


図4. 与えられた時系列データのグラフ

●差分データ作成

表1で与えられた時系列データから式(3.2)より表2の差分データを作成する。

表2. X_t から作成された差分データ Y_t

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y_t	-6	-1	2	-5	1	-2	4	-1	2	-5	5	-1	3	-6	\hat{Y}_{15}

●差分データ Y_t からさらに差分データ Z_t を作成する

式(3.5)より表3.3の差分データ Z_t を作成する。

表3. Y_t から作成された差分データ Z_t

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Z_t	5	3	-7	6	-3	6	-5	3	-7	10	-6	4	-9	\hat{Z}_{14}

●データパターン作成

表2の時系列データから図5で示すようなデータパターンを作成する。この数値例では、 $n = 4$ とする。 n の値は4なので、 $D_1, D_2, D_3, \dots, D_{11}$ とデータパターンは全部で11個になる。 $D_1 \sim D_{10}$ が過去の出現パターンであり、 D_{11} が現在の出現パターンである。

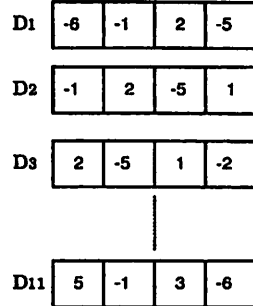


図5. 作成されたデータパターン

●メンバーシップ関数作成

図5のように作成したデータパターン $D_1 \sim D_{10}$ から式(3.7)に基づき三角形メンバーシップ関数を作成する。その様子を図6に示す。

$$\mu(y_i) = \max(0, 1 - |x - y_i| / a) \quad a > 0 \quad (3.7)$$

●適合度の計算

作成したデータパターン $D_1 \sim D_{10}$ のメンバーシップ関数と D_{11} のメンバーシップ関数との適合度をそれぞれ、式(2.9)の max・min 合成演算より適合度 P_i を求める式(3.8)を用いて求める。データパターン $D_1 \sim D_{10}$ は過去の出現パターンであり、 D_{11} は現在の出現パターンである。

パターン形状のファジィ抽出による時系列データの予測

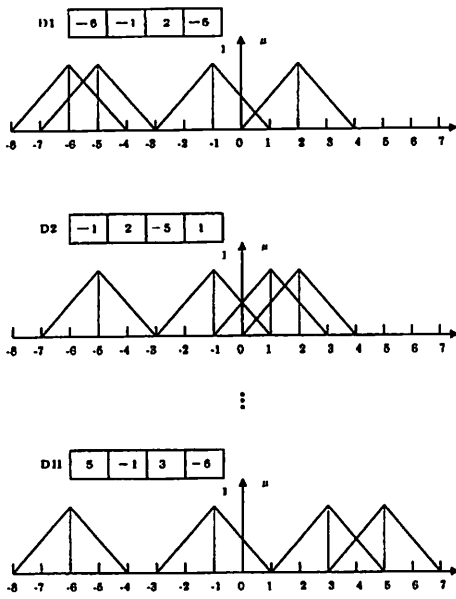


図6. データパターンとメンバーシップ関数

1つの例として、D11とD7との適合度を求める様子を図7に示す。

$$P_i = \frac{1}{n} | p_i + \sum_{j=1}^{n-1} p_{i+j} | \quad (3.8)$$

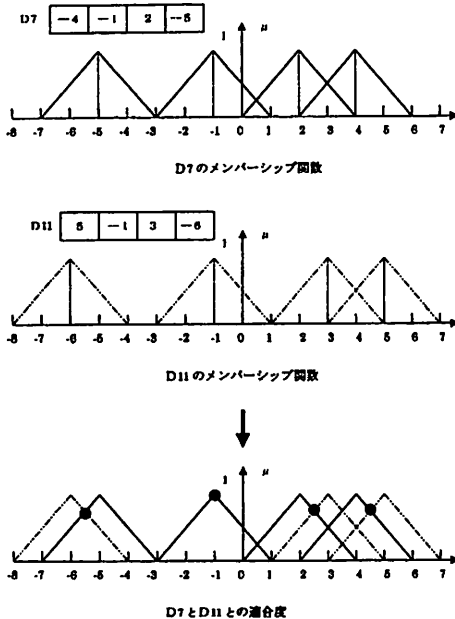


図7. max・min 演算により得られた適合度

図7において式(3.8)より得られた適合度はデータパターンの要素毎にそれぞれ黒の丸印部分になる。それぞれの演算結果は次のようになる。

$$p_7=0.75 \quad p_8=1 \quad p_9=0.75$$

$$p_{10}=0.75 \quad P_7=0.8125$$

適合度 $P_7=0.8125$ は、 D_7 が D_{11} に0.8125の割合で似ていることを表す。

• 適合度が最大になるデータパターンを抽出

適合度が最大になるデータパターンを抽出すると、 D_7 になる。求める Y_{15} は $Y_{i+n}=Y_{7+4}=Y_{11}$ より $Y_{15}=5$ 、適合度は0.8175となる。この数値例の場合、式(3.8)より得られる適合度が等しいデータパターンは存在しないので、 Z_t の時系列データを用いた計算は行わない。補足すると、2番目に適合度が大きいデータパターンは D_1 であり、 Y_{15} の値は1、適合度は0.625である。

• 予測値と信頼度の決定

式(3.3)より

$$\hat{Y}_1 = \hat{X}_{1+1} - X_1$$

$$\hat{Y}_{15} = \hat{X}_{16} - X_{15}$$

$$\hat{Y}_{16} = 5 + 3 = 8$$

以上のことから、表1の数値例の場合、得られた予測値 X_{16} と信頼度は表4のようになる。

表4. 予測値と信頼度

予測値 \hat{X}_{16}	信頼度
8	0.8125

表1のグラフと得られた予測値 \hat{X}_{16} を含めたグラフを図8に示す。

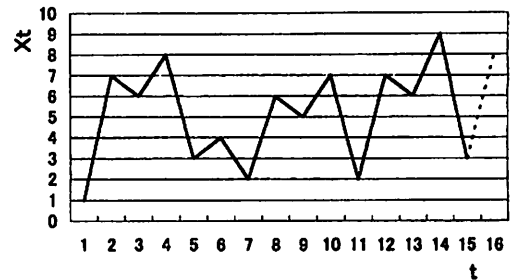


図8. 予測値を含んだグラフ

4. データおよびシミュレーション

シミュレーションプログラム用のデータには、ある商品の1日の値段を用いた。データは表6に示すように t を時刻とし、その時刻のデータを X_t とする。ここ

で、データ数は200個である。データの特徴として標本平均と標本分散を表5にまとめる。

表5. 標本平均値と標本分散値

	データ
標本平均値	1103.31
標本分散値	1502.36
標本標準偏差値	38.76

表6. データ

t(日)	1	2	3	4	5	6	7	8	9	10	11	12
X _t (円)	1043	1043	1040	1028	1023	1030	1027	1035	1026	1021	1000	1016
t(日)	13	14	15	16	17	18	19	20	21	22	23	24
X _t (円)	1015	1002	1017	1022	1027	1047	1038	1050	1040	1050	1050	1042
t(日)	25	26	27	28	29	30	31	32	33	34	35	36
X _t (円)	1054	1070	1083	1088	1062	1062	1064	1064	1072	1085	1083	1077
t(日)	37	38	39	40	41	42	43	44	45	46	47	48
X _t (円)	1073	1065	1063	1068	1087	1079	1079	1079	1071	1080	1080	1070
t(日)	49	50	51	52	53	54	55	56	57	58	59	60
X _t (円)	1076	1076	1082	1092	1075	1068	1077	1075	1073	1080	1074	1086
t(日)	61	62	63	64	65	66	67	68	69	70	71	72
X _t (円)	1087	1076	1081	1083	1081	1099	1097	1099	1105	1105	1119	1116
t(日)	73	74	75	76	77	78	79	80	81	82	83	84
X _t (円)	1118	1165	1158	1133	1121	1117	1120	1104	1075	1063	1072	1065
t(日)	85	86	87	88	89	90	91	92	93	94	95	96
X _t (円)	1077	1074	1074	1102	1097	1104	1112	1116	1120	1120	1125	1128
t(日)	97	98	99	100	101	102	103	104	105	106	107	108
X _t (円)	1128	1127	1127	1138	1127	1118	1121	1112	1112	1113	1115	1120
t(日)	109	110	111	112	113	114	115	116	117	118	119	120
X _t (円)	1119	1116	1116	1108	1117	1124	1123	1124	1118	1121	1125	1112
t(日)	121	122	123	124	125	126	127	128	129	130	131	132
X _t (円)	1110	1109	1113	1117	1118	1116	1118	1115	1110	1115	1123	1112
t(日)	133	134	135	136	137	138	139	140	141	142	143	144
X _t (円)	1112	1129	1123	1116	1115	1112	1109	1103	1105	1108	1107	1100
t(日)	145	146	147	148	149	150	151	152	153	154	155	156
X _t (円)	1094	1095	1106	1107	1104	1106	1150	1125	1125	1154	1143	1137
t(日)	157	158	159	160	161	162	163	164	165	166	167	168
X _t (円)	1151	1134	1139	1146	1158	1165	1174	1177	1173	1173	1172	1172
t(日)	169	170	171	172	173	174	175	176	177	178	179	180
X _t (円)	1166	1149	1150	1159	1144	1146	1154	1137	1138	1131	1138	1140
t(日)	181	182	183	184	185	186	187	188	189	190	191	192
X _t (円)	1143	1139	1143	1153	1146	1145	1147	1140	1140	1140	1122	1124
t(日)	193	194	195	196	197	198	199	200				
X _t (円)	1129	1132	1136	1132	1135	1129	1127	1130				

表6のデータに対して、一つのデータパターンの要素数を表すnの値を4つのケース(n=4, 5, 10, 20)に分けてシミュレーションを行った。総データ数は200個であり、t>20に対し、m個のデータからm+1個目の値を予測する。図9~12に、実測値と予測値のグラフを示す。ただし、折れ線グラフ中の実線は実測値を表し、破線は予測値を表す。

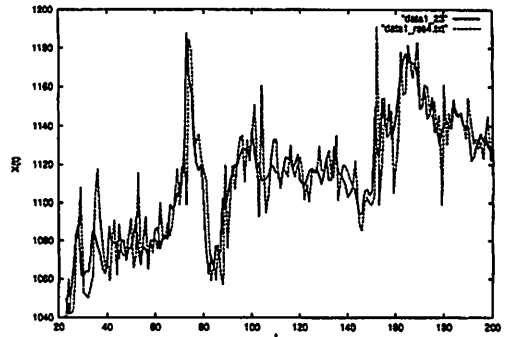


図9. 実測値と予測値 (n=4)

図9のようにデータに対してn=4の場合、tが100あたりまでは大きな誤差が見られる。100から150あたりまでは大きな誤差は見られず、予測精度が良いことがわかる。また、160あたりから再び誤差が見られる。

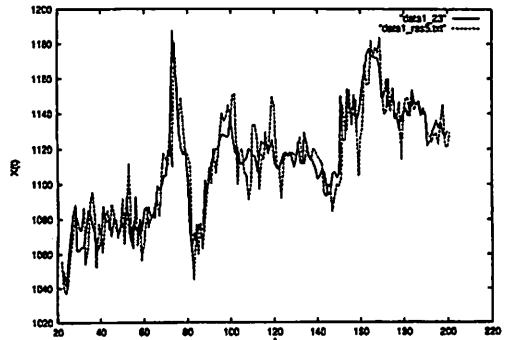


図10. 実測値と予測値 (n=5)

図10のようにn=5の場合、tが70あたりで大きな誤差が見られる。また120から150あたりまでは誤差は小さい。全体的に誤差は小さい。

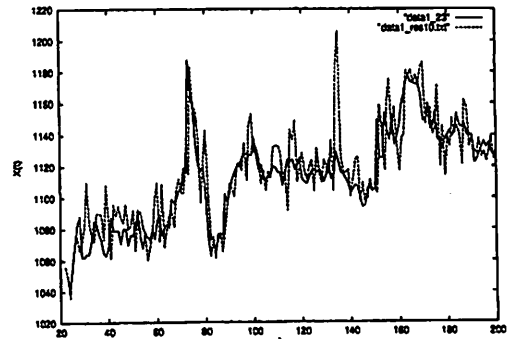


図11. 実測値と予測値 (n=10)

図11のようにn=10の場合、予測精度が良い箇所もあるが、全体的に大きな誤差が所々に見られる。

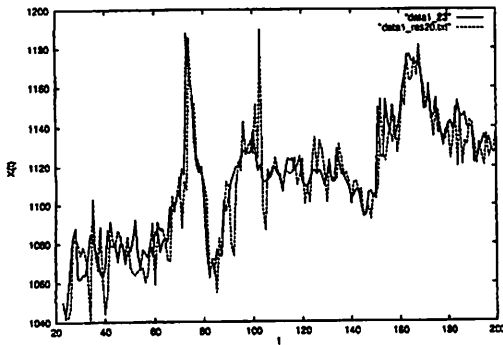


図12. 実測値と予測値 (n=20)

図12より、 $n=20$ の場合、所々に大きな誤差が見られるものの、 t が100以降では誤差が比較的小さく、予測精度の良さが伺える。

5. 結言

提案した本手法は、人間の主観的な思考や判断のあいまい性を定量的に取扱うことができるというファジィ理論の特長を活かした。このことにより、人間の意思が構成要素として入っているデータに対しても予測値を得ることができた。手法の有効性を確かめるために、提案したファジィ抽出法を実在の時系列データに適用し、1つのデータパターンの要素である n の値を変えながら、予測を行った。このシミュレーションの結果、予測誤差が大きい箇所がいくつか見られたが、ほとんど誤差がなく予測精度の良さが伺える箇所も見られた。これらの原因として、1つのデータパターンの要素数である n の値と三角形ファジィメンバーシップ関数の式(2.6)の a の値にあると考えられる。

参考文献

- [1] 坂和正敏：ファジィ理論の基礎と応用，森北出版（1989）
- [2] 豊浦由浩・藪内賢之・和多田淳三：ファジィARモデルによる日本経済の分析，第15回ファジィシステムシンポジウム，pp 405-406（1999）
- [3] 小沢和浩・渡辺匠・菅家正康：ファジィ自己回帰モデルとその応用，第12回ファジィシステムシンポジウム，pp 373-376（1996）
- [4] 和多田淳三・田中英夫・横山宏：ファジィ時系列モデルとその予測問題への応用，日本経営工学会誌，Vol. 34, No.3 pp180-186.（1983）
- [5] 松葉育雄：非線形時系列解析，朝倉書店（2000）
- [6] 廣松毅・浪花貞夫：経済時系列分析，朝倉書店（1995）
- [7] 佐藤幸男：信号処理入門，オーム社（1992）
- [8] 田中英男：ファジィモデリングとその応用，朝倉書店（1996）
- [9] 水本雅晴：ファジィ理論とその応用，サイエンス社（1989）