全体的傾向と局所的特徴に基づく時系列データの言葉による表現
—標準偏差による振動の表現—

Expression of Time Series in Natural Language with Global Trends and Local Features
—Expression of Oscillation with Standard Deviation

馬野 元秀　小泉 尚之　岡村 光洋
Motohide Umano　Naoyuki Koizumi　Okamura Mitsuhiro
大阪府立大学 大学院理学系研究科 情報数理科学専攻
Department of Mathematics and Information Sciences
Graduate School of Science, Osaka Prefecture University

1. はじめに

我々人間は、時系列データを理解するのに、まず全体を大雑把に理解してから局所的な特徴を見つけることにより、時系列データを理解していると思われる。昨年の報告会では、この考えに基づいて、全体的な傾向と局所的な特徴を言葉で表現する方法を提案した。

しかし、この方法ではデータの細かい振動的な変化をうまく表現することができない。そこで、この報告では、標準偏差により時系列データの細かい振動的な変化を表現する方法を検討する。

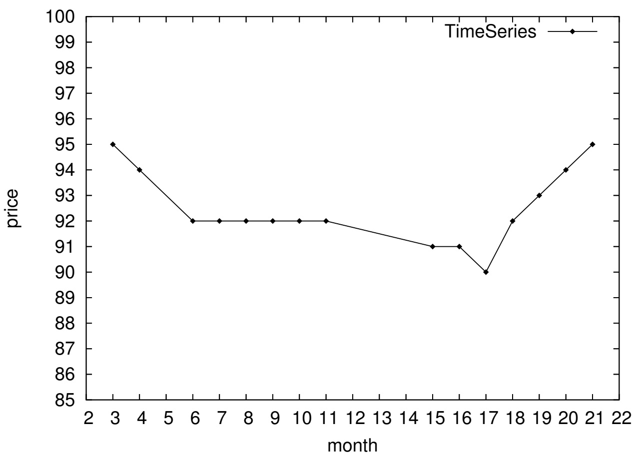

図 1: データセットから得られた時系列データ

2. 時系列データの全体的な傾向

時刻 t_a から t_b までの時系列データを考え、値の最小値と最大値を、それぞれ、x_a と x_b とする。例として、「動向情報の要約と可視化に関するワークショップ」[1][2] における研究用データセットの中の「レギュラーガソリンの全国平均店頭価格」の新聞記事データから手作業で取り出した時系列データ (図 1) を用いる。まず、全体的傾向を求める方法を説明しよう [3]。

(1) ファジィ集合による時系列データの期間分割

まず、時間を前期、中期、後期に分けて考える。このとき、前期、中期、後期は図 2 のようなファジィ集合と考える。各ファジィ集合は区分的に 2 次曲線によ

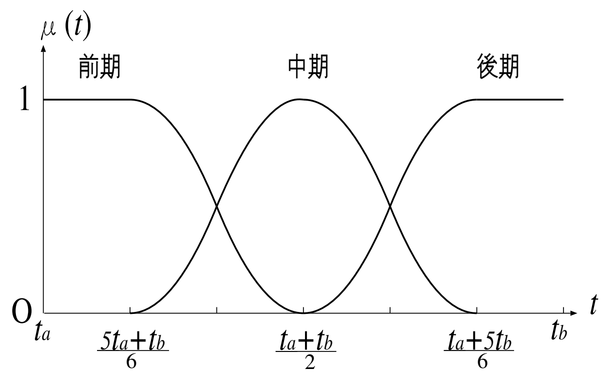

図 2: 期間のファジィ集合

表 1: 全体的傾向の表

前→中 ＼ 中→後	大きく減少	中ぐらい減少	少し減少	ほぼゼロ	少し増加	中ぐらい増加	大きく増加
大きく減少	急激に減少	急激に減少	穏やかに減少	穏やかに減少	やや下に凸	下に凸	下に凸
中ぐらい減少	急激に減少	急激に減少	穏やかに減少	穏やかに減少	やや下に凸	下に凸	下に凸
少し減少	穏やかに減少	穏やかに減少	穏やかに減少	だいたい一定	やや下に凸	やや下に凸	やや下に凸
ほぼゼロ	穏やかに減少	穏やかに減少	だいたい一定	だいたい一定	だいたい一定	緩やかに増加	穏やかに増加
少し増加	やや上に凸	やや上に凸	やや上に凸	だいたい一定	穏やかに増加	穏やかに増加	穏やかに増加
中ぐらい増加	上に凸	上に凸	やや上に凸	穏やかに増加	穏やかに増加	急激に増加	急激に増加
大きく増加	上に凸	上に凸	やや上に凸	穏やかに増加	穏やかに増加	急激に増加	急激に増加

り定義されている [4]。

(2) 重み付き平均値の算出

前期、中期、後期における代表値として、期間を表わすファジィ集合のよる重み付き平均値を求める。これは、

$$m_{\text{期間}} = \frac{\sum x(t)\mu_{\text{期間}}(t)}{\sum \mu_{\text{期間}}(t)} \tag{1}$$

となり、図 1 の場合は、$m_{\text{前期}}$ が 92.91、$m_{\text{中期}}$ が 91.75、$m_{\text{後期}}$ が 92.46 となる。

(3) 変化量のファジィ集合

まず、前期から中期への、中期から後期への変化量を計算すると、図 1 の時系列データの場合は、$d_{\text{前期}\to\text{中期}} = -1.16$、$d_{\text{中期}\to\text{後期}} = 0.71$ となる。

次に、図 3 のような変化量に対応するファジィ集合を用い、変化量 $d_{\text{前期}\to\text{中期}}$ と $d_{\text{中期}\to\text{後期}}$ と一致するファジィ集合とその一致度 (メンバシップ値) を求める。一致度が正になるファジィ集合 (または一定の値以上になるもの) の名前が変化量を表現する言葉となる。図 1 の時系列データの場合は、$d_{\text{前期}\to\text{中期}}$ に対して、「中ぐらい減少」が 0.08、「少し減少」が 0.92 言える。同様に、$d_{\text{中期}\to\text{後期}}$ に対しては、「ほぼゼロ」が 0.044、「少し増加」が 0.956 言える。

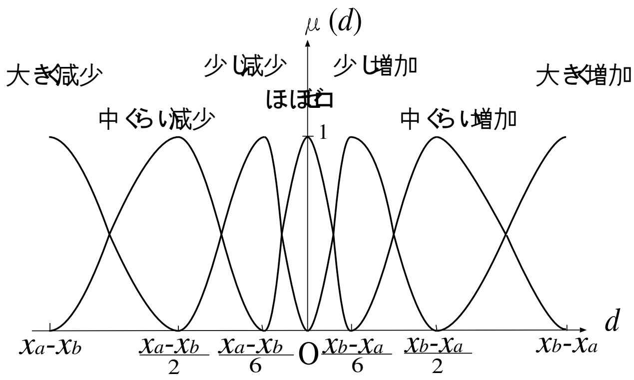

図 3: 変化量のファジイ集合

(4) 全体の評価と確信度

最後に全体的な傾向の表現を得るために、表 1 を用いる。表 1 は、縦に前期から中期への変化量のファジィ集合を、横に中期から後期への変化量のファジィ集合をとり、表の対応する要素に全体的傾向を表わすファジィ集合を入れてある。変化量のファジィ集合のすべての組合せについて全体的傾向を求め、さらに各変化量の一致度の積をその全体的傾向の一致度とする。そして、同じ表現の全体的傾向が複数の組合わせから得られた場合は、各一致度の和をその表現の一致度とする。

図 1 の時系列データの場合には、前期から中期への変化 {0.08/中ぐらい減少, 0.92/ 少し減少 } と中期から後期への変化 {0.044/ほぼゼロ，0.956/少し増加 } の組合せを考える。表 1 により、「だいたい一定」の一致度が 0.04、「やや下に凸」の一致度が 0.88+0.0765 = 0.9565、「穏やかに減少」の一致度が 0.0035 となる。

この方法は、前期から中期への変化量と中期から後期への変化量を用いて、表 1 のようなファジィルールを用いて、×-+ 法でファジィ推論をしていることに

なる (結論部が数値ではないので、重心をとることはできない)。

3. 時系列データの局所的特徴

時系列データの局所的な特徴は、時系列データと全体的傾向を比べて、違いが顕著な部分であると考えられる。

(1) 全体的傾向を表現する時系列データ

時系列データと全体的傾向を比べて違いを見つけるために、まず全体的傾向を表現する時系列データをつくる。時系列データは時間と値の組の集合できまるので、全体的傾向は期間と代表値の組で決まると考えられる。すなわち、いま考えている例では、$\{$(前期, $m_{前期}$), (中期, $m_{中期}$), (後期, $m_{後期}$)$\}$ で決まることになる。これを 1 変数のファジィルールと考えると、図 2 のようにすべての t で、ファジィ集合に対するメンバシップ値の和が 1 になる場合には、

$$\begin{aligned} x(t) = \mu_{前期}(t) \cdot m_{前期} \\ +\mu_{中期}(t) \cdot m_{中期} \\ +\mu_{後期}(t) \cdot m_{後期} \end{aligned}$$

で計算できる (メンバシップ値の和が 1 にならないときは、メンバシップ値の和で割ればよい)。

図 1 の時系列データより得られる全体的傾向の時系列データを図示すると、図 4 の点線のようになる。図には比較のために実線でもとの時系列データも図示してある。

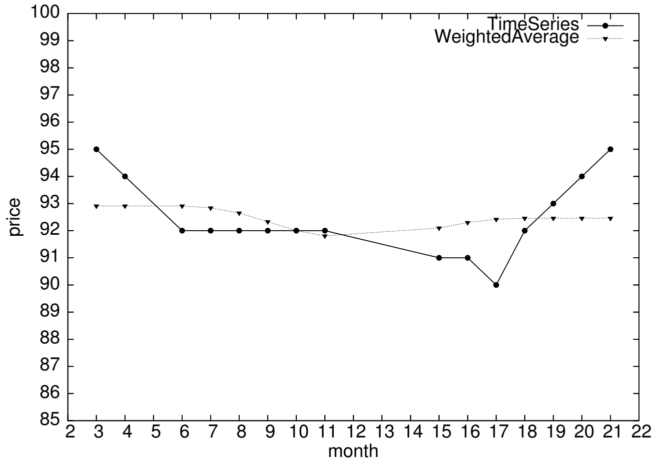

図 4: 全体的傾向のグラフ

(2) 局所的特徴

時系列データの局所的な特徴は、時系列データと全体的傾向を比べて、差の絶対値が大きい部分を見つけ、差をファジィ集合で表現する。図 1 の時系列データの場合には、差の絶対値が最も大きいのは時刻 $t = 21$ で 2.544 であるので、後期において、0.552 で「中ぐらい増加」、0.448 で「少し増加」となり、2 番目に大きいのが時刻 $t = 17$ で -2.416 であるので、後期において、0.596 で「少し減少」、0.404 で「中ぐらい減少」となる。このときには図 3 のファジィ集合を使用する。

以上を総合すると、図 1 の時系列データに対する言葉による表現は、一致度の大きなものを選ぶと、「全体的にはやや下に凸であるが、後期において中ぐらい増加している」と「全体的にはやや下に凸であるが、後期において少し減少している」となる。この例の場合は、局所的特徴が矛盾しているように思われるが、後期に値が大きく変化しており、いずれも正しいことが分かる。

4. 細かい振動的な変化の表現の検討

株や為替などの時系列データは、細かく振動しながら全体的にある傾向を持ったり局所的な特徴を持ったりする。通常はこのような振動の解析はフーリエ変換を用いて周波数の分布に変換して行なうが、今回は平均値からのずれである分散と標準偏差について考える。

ファジィ集合のメンバシップ値で重み付けられた各期間の分散 $V_{期間}$ と標準偏差 $\sigma_{期間}$ は式 (1) の重み付き平均 $m_{期間}$ を用いて、

$$V_{期間} = \sigma^2_{期間} = \frac{\sum(x(t) - m_{期間})^2 \mu_{期間}(t)}{\sum \mu_{期間}(t)} \qquad (2)$$

と定義される [5]。

分散や標準偏差の値からどのようにして言葉を作るかはよく分かっていないので、実際に時系列のグラフをいくつか描いて、標準偏差を計算してみよう (ファジィ集合の重みは使わないとする)。図 5〜図 7 のグラフはすべてデータの数が同じで、平均も同じである。図 5 のグラフでは、標準偏差は (a) を 2 とすると、(b) が 3 で、(c) が 1 で、(d) が 2 となる。(a) と (d)

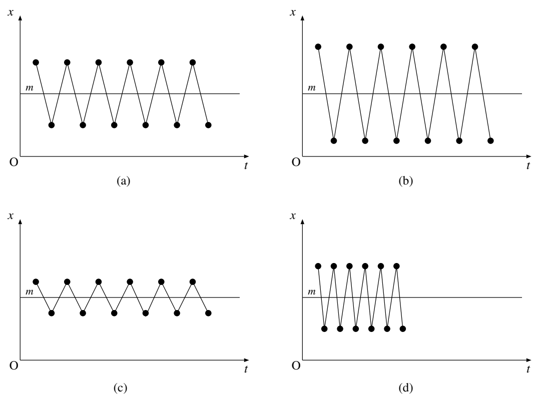

図 5: 平均が同じ時系列データ

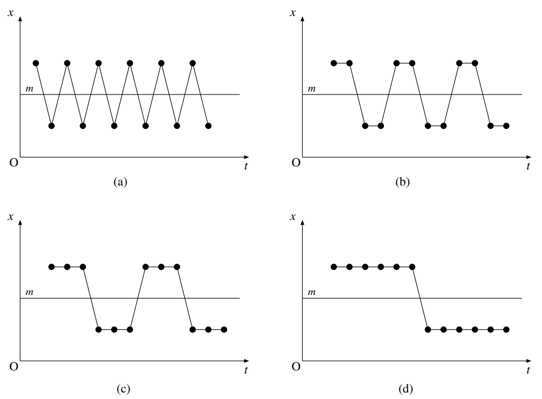

図 6: 標準偏差が同じ時系列データ

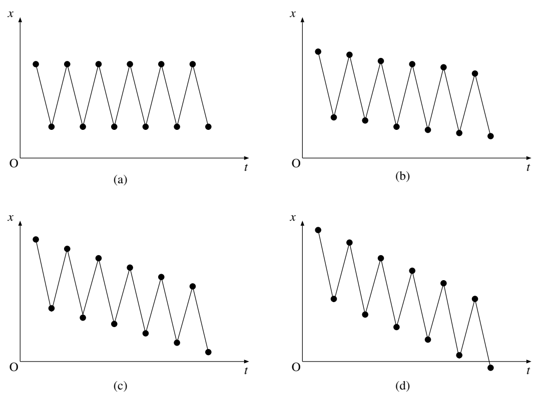

図 7: 分散が同じ時系列データ

が同じなので、標準偏差だけでは決められないので、フーリエ解析を使わないまでも時間あたりの極値または 0 となる回数や時間当りの変化量などを考慮する必要があると思われる。図 6 では、すべての標準偏差は同じであるが、印象はずいぶんと異なる。これも時間あたりの極値または 0 となる回数や時間当りの変化量などを使う必要があると思われる。図 7 では、標準偏差は (a) 、(b)、(c)、(d) の順に大きくなる。しかし、基準線として全体的傾向の線を思い浮かべてしまい、細かい振動については同じような言葉でよさそうにも思える。したがって、全体的傾向を表わす時系列データが持っている標準偏差分を考慮する必要があると思われる。

さらに、隣り合う期間の標準偏差をどのように考慮するかは重要な問題である。また、実際に言葉に表現するときには、今までの全体傾向と局所的特徴とどのように関連付けるかなど検討すべき項目は多い。

5. おわりに

本論文では、時系列データの全体的な傾向と局所的な特徴を言葉で表現する方法について、細かく振動する変化を標準偏差を利用して表現する方法の基本的な検討を行なった。今後、さらに検討を加えて、CommonLisp のシステムに組み込んでいきたい。

さらに、時系列データの時間や値があいまいである場合や、多くの時系列データからの規則抽出などについても、考えていきたい。

参考文献

[1] 加藤 恒昭、松下 光範、平尾 努 (2004): 「動向情報の要約と可視化に関するワークショップの提案」、情報処理学会 自然言語処理研究会、2004-NL-164 (15)、pp.89–94

[2] http://must.c.u-tokyo.ac.jp/must_index.html

[3] 馬野 元秀、小泉 尚之、篠原 貴之、瀬田 和久 (2006): 「全体的傾向と局所的特徴に基づく時系列データの言葉による表現」、第 22 回ファジィ システム シンポジウム、pp.343-346

[4] L.A. Zadeh (1975): "Fuzzy Logic and Approximate Reasoning," *Synthese*, Vol.30, pp.407–428

[5] L.A. Zadeh (1968): "Probability Measures of Fuzzy Events," *Journal of Mathematical Analysis and Applications*, Vol.23, pp.421–427

[問い合わせ先]
〒599–8531
堺市中区学園町 1 – 1
大阪府立大学 大学院理学系研究科
情報数理科学専攻
馬野 元秀
TEL: 072–254–9675
FAX: 072–254–9930
E-mail: umano@mi.s.osakafu-u.ac.jp