

ファジィ時系列モデルを利用したキーワード検索数の動向予測手法の提案

長沼 博則[†] 藤田 ハミド[†] 樽松 理樹[†] 羽倉 淳[†]

岩手県立大学ソフトウェア情報学部[†]

1. はじめに

近年、時系列データからの動向予測を適用する分野が増え、株価などのチャート、テクニカル分析の他、在庫管理等で利用されている¹⁾。これに伴い予測精度向上の重要度は増している。また近年、企業は広告の場をインターネットに移してきている。平成27年の総務省の調査²⁾によると検索結果上位に表示されることによるサイト誘導効果が事業者間に広がった結果、PC向け検索連動型広告の市場規模は増加してきている。そのためSEO対策として、キーワードの選択は重要であるが、有効なものを予測することは困難である。

以上のような背景から、本研究では Shyi-Ming Chen の論文³⁾を参考に、キーワード検索数の時系列データから作成し、粒子群最適化(PSO)を適用したファジィ時系列モデルを用いる動向予測手法の提案を行う。

2. 提案手法

本手法では図1に示すように、初めに時系列データから、分割点・領域定義、ファジィ論理関係、重みの定義、予測値、評価値出力を行い、ファジィ時系列モデルを作成する。次に PSO (粒子群最適化)により、このモデルを最適化する。

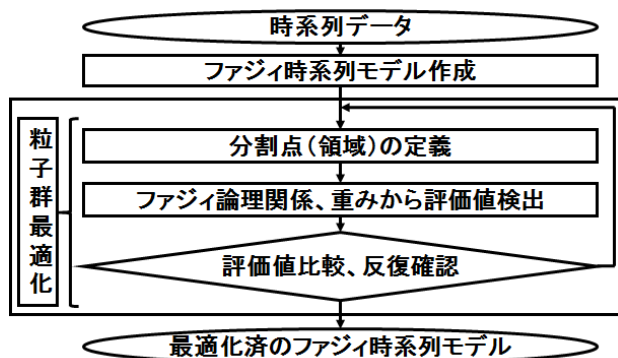


図1 処理の流れ

2.1 分割点、領域の定義

時系列データの最大値 D_{max} 、最小値 D_{min} をもとに D_1 と D_2 のマージンをとった範囲 $U = [D_{min} - D_1, D_{max} + D_2]$ を決定する。その範囲を等間隔に

$n + 1$ 個分割したものを $[x_{min}, x_{id,1}, x_{id,2}, \dots, x_{id,n}, x_{max}]$ とする。ここで $x_{id,j}$ は現在の粒子番号 id 、分割点番号 j の分割点を示す。 $D_{min} - D_1, D_{max} + D_2$ を x_{min}, x_{max} とする。

隣接する分割点の間にある領域 u_j を用いて、ファジィセットを定義する。例としてファジィセット A_i を式(1)で示す。0, 0.5, 1のメンバーシップ値はそのファジィセットがどこの領域を示しているかを表す。

$$\tilde{A}_i = 0/u_1 + \dots + 0.5/u_{i-1} + 1/u_i + 0.5/u_{i+1} + \dots + 0/u_{n+1} \quad (1)$$

2.2 ファジィ論理関係、重みの定義

時系列データにファジィセットを時系列ごとに振っていき、そして予測する値のファジィセット A_t と2つ前と1つ前のファジィセット A_{t-2}, A_{t-1} からファジィ論理関係 $A_{t-2}, A_{t-1} \rightarrow A_t$ を定義する。

このとき、1つ前と2つ前の大小関係から upGroup, equalGroup, downGroup の3種類のファジィ論理関係グループを定義し、グループごとに出力する重み $W_{id,i} = [w_{i1}, w_{i2}]$ を定義する。ここで w_{i1}, w_{i2} は、それぞれグループ i に属している $t-2, t-1$ 時の重みを示す。重みを定義する条件として $w_{i,j} \in [0,1], w_{i1} + w_{i2} = 1, 1 \leq i \leq 3, 1 \leq j \leq 2$ がある。

2.3 予測値、評価値出力

予測方法として、時系列 t の予測値 $F_{id,t}$ は定義された重みと時系列 $t-2, t-1$ の値 R_{t-2}, R_{t-1} を式(2)で求める。

$$F_{id,t} = R_{t-2} \times w_{i1} + R_{t-1} \times w_{i2} \quad (2)$$

また、この分割点や重みに対する評価値 RMSE を式(3)から求める。

$$RMSE = \sqrt{\sum_{t=3}^{t_{max}} \frac{(F_{id,t} - R_t)^2}{t_{max}} - 2} \quad (3)$$

2.4 PSO

PSO を利用し、ファジィ時系列モデルの分割点と重みを最適化していく。これにより精度の向上を試みる。PSO は粒子の情報を変えて作成した粒子を集め、粒子群を作成する。群の中で良粒子を選択し、前回作成した粒子群の良粒子と比べ評価し、最適化していく。

A Proposal of a Trend Prediction Method for Keywords Searching using Fuzzy Time Series Model

[†]NAGANUMA HIRONORI [†]FUJITA HAMIDO

[†]KUREMATSU MASAKI [†]HAKURA JUN

[†]Faculty of Software and Information Science, Iwate Prefectural University

2. 4. 1 分割点, 領域の粒子群作成

2. 1 で定義した分割点情報 $x_{id,j}$ をもつ $X_{id} = [x_{id,1}, x_{id,2}, \dots, x_{id,n}]$ を初期粒子とする. また群内良粒子 $P_{id} = [p_{id,1}, p_{id,2}, \dots, p_{id,n}]$, 反復内良粒子 $X_{gBest} = [X_{gBest}, X_{gBest}, \dots, X_{gBest}]$ を初期値は初期粒子で定義する.

分割点を移動させる変数 $V_{id} = [v_{id,1}, v_{id,2}, \dots, v_{id,n}]$ を定義する. $v_{id,j}$ は範囲が決めており, 初期はその中をランダムに出力する. 2回目以降は, 群内の良粒子情報, 反復内の良粒子情報を考慮しつつ式(4)で求める.

$$v_{id,j} = k \times [\omega \times v_{id,j} + c_1 \times rand \times (p_{id,j} - x_{id,j}) + c_2 \times Rand \times (x_{gBest,j} - x_{id,j})] \quad (4)$$

最大値を超えた場合は最大値を, 最小値を下回った場合は最小値を出力する. 範囲は式(5)で示す.

$$\frac{-0.2 \times (x_{max} - x_{min})}{2} \leq v_{id,j} \leq \frac{0.2 \times (x_{max} - x_{min})}{2} \quad (5)$$

$Rand$, $rand$ は0から1の範囲で出力する乱数であり, 異なる値である. 係数 c_1 , c_2 , k , ω の最大値, 最小値は任意とし, 式(6),(7)で示す. 式(7)において $iter$ は反復回数である.

$$k = 2 / \left| 2 - \varphi - \sqrt{\varphi^2 - 4\varphi} \right| \quad \varphi = c_1 + c_2 \quad (6)$$

$$\omega = (\omega_{max} - \omega_{min}) \times \frac{(iter_{max} - iter)}{iter_{max}} \quad (7)$$

分割点の更新を式(8)で示す.

$$x_{id,j} = x_{id,j} + v_{id,j} \quad (8)$$

$x_{id,j}$ の範囲は x_{min} から x_{max} であり x_{max} を超えた場合は式(9), x_{min} を下回った場合は式(10)で求める.

$$x_{id,j} = \begin{cases} x_{max} - (1/2 \times rand \times (x_{max} - x_{min})) & (9) \\ x_{min} + (1/2 \times rand \times (x_{max} - x_{min})) & (10) \end{cases}$$

新たな分割点でファジィ論理関係, ファジィ論理関係グループを定義し, これを新たな粒子の情報とする.

2. 4. 2 重みの粒子群作成

2. 2 で定義した重み $w_{i,j}$ をもつ $Z_{id}^* = [z_{id^*,1}, z_{id^*,2}]$ を初期粒子とする. この重みを変更して, 分割点と同じく様々な粒子, 粒子群を作成していく.

2. 4. 3 ベスト粒子の更新

新たな分割点, 重みで予測値と評価値を出力し, 評価値が前回粒子よりも良ければ群内良粒子を更新する. また群作成が終了し, 群内良粒子が反復内良粒子より評価値が良い場合, 反復内良粒子を更新する.

分割点情報, 重み情報を初期値に戻し, 群作成を繰り返していき, 反復終了時に反復内良粒子を最適化されたモデルとして出力する.

2. 5 予測値算出

予測対象の時系列データに対して作成したモデルを適用した結果から式(2), 時系列データの最終の値を用いて, 予測値を出力する.

3. 評価実験

本研究では, MATLAB ver2016a⁴⁾内の作成したプログラムを用いて実験を行う. 評価方法は, Google trend⁵⁾の2011年12月25日から2015年12月13日までのキーワード検索数推移データ, 約200件をトレーニングデータ, 2015年12月20日から2016年12月18日までのキーワード検索数推移データ, 約50件をテストデータとする. 選択したキーワードとしては, 身近なキーワードとして岩手に関連したキーワードや季節, 時期によっての変化があるもの, 検索数上位キーワードで時系列データが穏やかなもの, 各年度の流行語など無作為に選択した. PSOを適用したファジィ時系列モデルと最適化していないファジィ時系列モデルを比べて評価する. 分割点を変更した粒子数は10個, 反復は100回, 重みを変更した粒子数は10個, 反復は10回で行った. 初期値として等分割点数は8, 分割領域数は9, マージンをとるための D_1 , D_2 はそれぞれ1, 重みはそれぞれ0.5, 係数 $c_1 = c_2 = 2.05$, $k = 0.7298$, ω は最大値0.9, 最小値0.4としている.

結果として, 前日と当日の変化率が大きい時系列データは精度が悪く, 最適化してもテストデータでは初期値のファジィ時系列モデルを超えられなかった. 値の最大値, 最小値で作成していたモデルを値が取りうる可能性がある最大値, 最小値で作成する必要がある.

4. 終わりに

本研究はキーワード検索数の時系列データからファジィ時系列モデルを作成し, PSOによってモデル最適化を行い, 最適化されたモデルによる動向予測手法を提案した. 今後の課題として, 重みの算出方法, 最適化手法などを変更し, 様々な時系列データで対応できるようにすることが挙げられる.

参考文献

- 1) Furong Ye, Liming Zhang, Defu Zhang, Hamido Fujita, "A novel forecasting method based on multi-order fuzzy time series and technical analysis", Information Sciences(2016)Vol 367-368, 1 November 2016, Pages 41-57
- 2) 広告手段としてのインターネット利用, 企業ホームページ開設, 情報通信白書 (総務省 平成27年) <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc121220.html>, (閲覧日2016年12月20日)
- 3) Shyi-Ming Chen, Bui Dang Ha Phuong, "Fuzzy Time Series Forecasting Based on Optimal Partitions of Intervals and Optimal Weighting Vectors", Knowledge-Based Systems(2016), Available online 29 Nov 2016, Doi:10.1016/j.knsys.2016.11.019
- 4) MathWorks, MATLAB <https://jp.mathworks.com/>(閲覧日2016年1月12日)
- 5) Google Trend <https://www.google.co.jp/trends/> (閲覧日2016年12月20日)