ACTA UNIVERSITATIS APULENSIS No 10/2005

Proceedings of the *International Conference on Theory and Application of Mathematics and Informatics ICTAMI 2005* - Alba Iulia, Romania

PROBABILISTIC FINITE STATE AUTOMATA AND TIME SERIES ANALYSIS

Ciresica Jalobeanu

Abstract. Probabilistic finite state automata generating strings of letters is used to represent the symbolic time series. A symbolic time series as a model is developed, and a measure of the corresponding approximation error is proposed. The symbolic time series depends essentially on the segmentation of the initial series. An operator of segmentation using thresholds generates a structure of segments. Based on the past data, the probability of transitions and states of the probabilistic finite state automata are estimated. The forecasting probability in a symbolic time series is defined. The conditions for an optimal forecasting are analyzed.

Keywords: probabilistic finite state automata, forecasting, time series, symbolic time series.

Introduction

In some previous papers [2,3], we have developed a syntactic method as an alternative for other models used in time series analysis. The aim was to find solutions for forecasting some time series proposed at the 1992 international contest - Santa Fe Institute Studies in the Science of Complexity [8].

In our study we have considered a time series, graphically represented, as a digital picture, and the segments of the time series, as objects of the picture. We have taken as patterns the contour of objects preserving the monotony. A set of pattern primitives - a finite free set - have been chosen, and their labels, as a set of free symbols, have been considered as an alphabet Σ. In this way, a word from the language L ($L \subset \Sigma^*$), corresponds to an object, from the picture, while a set of words correspond to the whole series. This set of words has been used in order to construct a symbolic time series.

The main aim in studying time series is to use them in forecasting. In order to solve this problem, we have used the probabilistic finite state automata providing the forecast probability.

1. Probabilistic generation of strings

1.1. Probabilistic finite state automata

Definition 1.1. *$A = (Q, \Sigma, \delta.q_0, F)$ is a finite state automata if*
Q - is a finite set of states
Σ - is an alphabet
$\delta \subseteq Q \times \Sigma \times Q$ - is the set of transitions
q_0 - is the initial state
F - is the set of final states
The language $T(A) \subset \Sigma^*$, accepted by the automata is a regular language.

Definition 1.2. *A probabilistic finite state automata is*

$$A = (Q, \Sigma, \delta, I_A, F_A, P_A),$$

where:
$P_A : \delta \to [0, 1]$ are the transitions probabilities
$I_A : Q \to [0, 1]$ are the initial states probabilities
$F_A : Q \to [0, 1]$ are the final states probabilities
I_A, P_A, F_A are functions such that:

$$\sum_{q \in Q} I_A(q) = 1 \quad \text{and} \quad \forall\ q \in Q,\ F_A(q) + \sum_{a \in \Sigma, q' \in Q} P_A(q, a, q') = 1.$$

It is assumed that $P_A(q, a, q') = 0$ for all $(q, a, q') \notin \delta$.

Computing the probability of a string $x \in \Sigma^*$, that A generates let us consider

$$\theta(x) = (s_0, x_1, s_2, x_2, \ldots, s_{k-1}, x_k, s_k)$$

a path for $x = x_1 \ldots x_k$ in A , i.e. the sequence of transitions

$$(s_0, x_1, s_1), (s_1, x_2, s_2), \ldots, (s_{k-1}, x_k, s_k) \in \delta.$$

Definition 1.3. *The probability of generating a path*

$$\theta(x) = (s_0, x_1, s_1, x_2, \ldots, s_{k-1}, x_k, s_k)$$

is defined with:

$$P_A(\theta(x)) = I_A(s_0) \prod_{j=1}^{k} P_A(s_{j-1}, x_j, s_j) F_A(s_k).$$

Let Θ_A be the set of all valid path (having the probability greater than zero) generated by A.

DEFINITION 1.4. *The probability of generating x by A is*

$$P_A(x) = \sum_{\theta_i \in \Theta(x)} P(\theta_i).$$

The automata A defines a distribution D on Σ^* if

$$\sum_{x \in \Sigma^*} P_A(x) = 1.$$

In the sequel the indices A will be omitted.

1.2. Words and sub-words generated in probabilistic finite state automata

The probability of reaching state q and generating the prefix $x_1 \dots x_i$, using the sequence of states $S = (s_0, s_1, \dots, s_i)$, is

$$\forall\ q \in Q,\ \alpha_x(i, q) = \sum_{S \in \Theta(x_1 \dots x_i)} I(s_0) \prod_{j=1}^{i} P(s_{j-1}, x_j, s_j) \rho(q, s_j),$$

for $0 \leq i \leq |x|$, where $\rho(q, s_j) = 1$, if $q = s_j$ and $\rho(q, s_j) = 0$ if $q \neq s_j$.

Forward algorithm for calculating this probability:

$\alpha_x(0, q) = I(q)$

$\alpha_x(i, q) = \sum_{q' \in Q} \alpha_x(i-1, q') P(q', x_i, q),\ 1 \leq i \leq |x|$

Then

$$P(x) = \sum_{q \in Q} \alpha_x(|x|, q) F(q).$$

The probability of generating the suffix $x_{i+1} \dots x_{|x|}$, from the state q, on the sequence of states $S = (s_i, s_{i+1}, \dots, s_{|x|})$ is :

$$\beta_x(i, q) = \sum_{S \in \Theta(x_{i+1}, \dots, x_{|x|}} \rho(q, s_i) \prod_{j=i+1}^{|x|} P(s_{j-1}, x_j, s_j) F(s_{|x|}),$$

where $\rho(q, s_i) = 1$ if $q = s_i$ and $\rho(q, s_i) = 0$ if $q \neq s_i$.

The backward algorithm:

$\beta_x(|x|, q) = F(q)$

$\beta_x(i, q) = \sum_{q' \in Q} \beta_x(i+1, q') P(q', x_i, q),\ 1 \leq i \leq |x| - 1$

Then

$$P(x) = \sum_{q \in Q} I(q) \beta_x(0, q).$$

2. Symbolic time series

In order to construct the codification of a time series by words from a language, we will summarize some basic notions.

2.1. Alphabet, words, operation with words

Let Σ be a finite set of symbols used as an alphabet. We will denote by Σ^+ the set of words obtained on Σ using the concatenation operation (denoted by $\cdot$). Then $(\Sigma^+, \cdot)$ is the free semi-group generated by Σ. We will aid the null word λ and consider $\Sigma^* = \Sigma^+ \cup \{\lambda\}$.

The $(\Sigma^*, \cdot, \lambda)$ is the free semi-group with unity generated by Σ. A language L is a subset of Σ^*. Let $W \subset \Sigma^+$ be a set of words. We define the set of prefixes of the words from W as the set:

$$Preff(W) = \{u | u \in \Sigma^+,\ \exists\ v \in \Sigma^+,\ uv \in W\}$$

and the set of suffixes as:

$$Suff(W) = \{v | v \in \Sigma^+,\ \exists\ u \in \Sigma^+,\ uv \in W\}.$$

2.2. Segmentation of a time series

In order to construct a symbolic time series by labeling the segments of a time series we introduce a segmentation operator J, in respect with the threshold h_i, for $i = 1, 2, \ldots, q$ ($h_i \in R$).

A time series is a set of observations x_t ($x_t \in R$), each one being recorded at a time t. The sequence

$$X = \{x_1, x_2, \ldots, x_N\}$$

is a time series containing the measurements on a discreet time interval $\{t_1, t_2, \ldots, t_N\} \subset [0, T]$.

As we have already mentioned, we will consider the envelopes of the time series, and the picture containing objects - i.e. sequences of non-void values, and intervals with null values. We will keep the separating points of these regions in a set of segmentation points. Then, we will add to this set the points, which separate the monotone parts of the envelopes, and the points which determine sub-patterns from a free set (a free set has the property that no element of the set can be expressed as a combination of other elements).

Let us consider a segmentation operator $J_h : X \to \{0, h\}$, where $J_h(x_i) = h$ if $x_i \geq h$ and $J_h(x_i) = 0$ otherwise.

Considering a sequence of thresholds, $H = \{h_1, h_2, \ldots, h_q\}$, such that $h_1 > h_2 > \ldots > h_q > 0$, ($h_i \in R$, $i = 1, 2, \ldots, q$, for $q < N$), we can combine the sequence of J_h operators in a single one, $J : X \to \{h_1, h_2, \ldots, h_q, 0\}$, taking for each $x_i \in X$, $J(x_i) = \max\{J_h(x_i) | h \in H\}$.

$J(X)$ is a time series associated to X, and having a reduced set of values, only. The shape of $J(X)$, and the role of J for a proper segmentation of X, depends essentially by the threshold selection.

Let us denote by

$$U = \{\mu_1, \mu_2, \ldots, \mu_s\} \quad \text{and} \quad \mu_1 < \mu_2 < \ldots < \mu_s$$

the $J(X)$ steps, i.e. the first indices where the values are changing. We will take as points of segmentation the values $\mu_1, \mu_2, \ldots, \mu_s$. The points $\{(\mu_i, \mu_{i+1}) | i = 1, 2, \ldots, s-1\}$ define the segments where the pattern primitives are defined. We will label the pattern primitives with symbols from the alphabet Σ.

Let us consider the segmentation points and their corresponding values in $X : (\mu_i, x_k)$, $i = 1, \ldots, s$ and $k \in \{1, \ldots, N\}$.

DEFINITION 2.1. *The point (μ_i, x_k) is a cut point if $|x_k - x_{k+1}| > K$, or $|x_k - x_{k-1}| > K$, where K is a chosen value (for instance $K > x_k/2$).*

2.3. The construction of the symbolic time series

In order to construct a symbolic time series, corresponding to the segmented time series, we will attach to every segment a letter, as a label. If a segment has a level zero, then it will be labeled by λ. Let us denote by W the set of words occurring in the codification of the initial time series.

Then, the symbolic time series will be

$$S_m = \sum_{i=1}^{m} (w_i^{p_i} + \lambda^{q_i}) \tag{2.1}$$

where:

m - is the number of different words corresponding to the "objects"

w_i - is a word

p_i - natural number of successive occurrence of w_i

q_i - natural number of null intervals between two objects.

As usual, we denoted by $|w|$ the length of the word w.

Here, we have used the following operations with letters and words:

a) The concatenation $(\cdot)$. Two letters can be concatenated if they are the labels of two adjacent segments without a cut-point in between.

b) The sum of words $(+)$, for a union of ordered words.

Example 1.

The objects from a picture have the description:

$$S = g + \lambda + fg + efg + cdefg + abcdefg$$

when $a, b, c, d, e, f, g \in \Sigma$.

In [2] we have studied the error in approximation with symbolic time series. We found the possibility to decide if a segmentation is acceptable in respect with a tolerable error. In the sequel we will consider only acceptable segmentations.

3. Predictibility in symbolic time series

3.1. Reconstruction by interpolation

In the paper [3] we have studied the symbolic time series that can be reconstructed by interpolation. Let us present, first of all, how we can formulate an interpolation problem for symbolic time series.

Let us consider a segmented time series, with $\mu_1, \ldots, \mu_m$ some cut-points, and $a_1, a_2, \ldots, a_r$, the letters placed at the right of these points, $a_i \in \Sigma$, $|\Sigma| = r$, r symbols used to label the segments of the series.

If we know the pairs (μ_k, a_k), $k = 1, \ldots, m$, $a_k \in \Sigma$, the interpolation problem consist in finding a symbolic time series having in the points $\mu_1, \ldots, \mu_m$, the patterns labeled with $a_1, a_2, \ldots, a_m$.

In this case the pairs (μ_k, a_k), $k = 1, \ldots, m$, will be called interpolation points.

THEOREM 3.1. *A symbolic time series interpolates the points* $(\mu_1, a_1), \ldots,$ (μ_m, a_m), *where* $\mu_1, \ldots, \mu_m$ *are cut-points, if the symbolic time series*

$$\sum_{i=1}^{m}(w_i + \lambda^{q_i}),\ w_i \in W,\ q_i \in N$$

accomplished the following conditions:

(1) For every $j = 1, \ldots, m$ *there exist* $w_j \in W$, *such that* $\Phi(w_j) = a_j$, *where* $\Phi(w_j)$ *is the first letter of the word* w_j.

(2) $|w_j| + |\lambda^{q_j}| \leq \mu_{j+1} - \mu_j$, *for* $j = 1, \ldots, m-1$.

If there is a symbolic time series interpolating the segmented time series, we will say that the original time series can be reconstructed by interpolation.

THEOREM 3.2. *(on reconstruction by interpolation) The segmented time series* Y *can be reconstructed by interpolation if there exists a set of words* W *such that every term of the symbolic time series*

$$\sum_{i=1}^{m} w_i + \lambda^{q_i}$$

has the property that

1)$\Phi(w_i) = a_i$, *for* $i = 1, \ldots, m$, *and for all* i, j, $\Phi(w_i) \neq \Phi(w_j)$, *if* $i \neq j$,

2) $|w_j| + |\lambda^{q_j}| = \mu_{j+1} - \mu_j$, *for* $j = 1, \ldots, m-1$.

In example 1, $W = \{g, fg, efg, cdefg, abcdefg\}$ with the cut points $(\mu_1, \ldots, \mu_6)$. If we take as interpolation points (μ_1, g), (μ_2, f), (μ_3, e), (μ_4, c), (μ_5, a) then the series can be reconstructed by interpolation.

DEFINITION 3.1. *The symbolic time series* S_m *is partially predictible if knowing the first letter in a word, there is a unique word, or sub-word, in* W, *beginning with the respective letter.*

It is straightforward the following result:

THEOREM 3.3. *If a symbolic time series can be reconstructed by interpolation then it is partially predictable.*

3.2. Probabilistic finite state automata generating the interpolation points

Let us consider a symbolic time series partially predictable. What is impredictible is the next word, after a cut point. If we would have some information

on the probability distribution of the interpolation points, we could predict with a probability the next interpolation point.

We will consider the sequence of interpolating points $z_i = (\mu_i, a_i)$, $i = 1, \ldots, k$, where by μ_i we denote the position of the interpolating point. We will associate with this string a probabilistic finite state automata.

Considering that from historical data we can estimate the following probabilities: initial states probabilities, transitions probabilities and final states probabilities I_A, P_A, F_A, we will suppose that

$$\forall\ q \in Q_A,\ F_A(q) + \sum_{a\in\Sigma, q'\in Q_A} P_A(q, a, q') = 1 \quad \text{and} \quad \sum_{q\in Q_A} I_A(q) = 1.$$

In that case we have a probabilistic finite state automata generating the interpolation points of a symbolic time series.

Let us consider $z = z_1 \ldots z_i$ as a word and $z_{i+1} \ldots z_n$ its suffix i.e. a possible continuation of the word. We have the possibility to calculate the probability of generating the suffix, i.e. the probability of a possible continuation of the word z. As we already have seen in 1.2., using the backward algorithm we find

$$\beta_z(i, q) = \sum_{S\in\Theta(z_{i+1},\ldots,z_n)} \rho(q, s_i) \prod_{j=i+1}^{n} P(s_{j-1}, z_j, s_j) F(s_n)$$

the probability of generating the suffix $z_{i+1} \ldots z_n$ beginning with the state q.

This is the probability of the continuation of the word z. When it is associate with a symbolic time series, we can find the probability of the forecasting using the next $n - i$ interpolation points $z_{i+1} \ldots z_n$.

The probabilistic finite state automata is a suitable device for defining dynamically the probabilities, as the automata would learn the new probabilities.

4. Conclusions

In searching for conditions for predictability of a time series, we have seen that a symbolic time series can be partially extrapolated if it can be reconstructed by interpolation. If a time series can be reconstructed by interpolation, then knowing the probabilities for each interpolation point, the series can be extrapolated with the corresponding probability given by the most probable interpolation point.

References

[1] Brockwell, P., Davis, R., *Time Series Theory and Methods*, Springer Verlag, Berlin, 1991.
[2] Jalobeanu, C., *Pattern Segmentation for Time Series*, Acta Technica Napocensis, Series Appl. Math. and Mech., 42, 1, 1999, 11-22.
[3] Jalobeanu, C., *Classification of the Time Series Using Syntactic Analysis*, Conference on Analysis, Functional Equations, Approximation and Convexity, in Honour of Prof. E. Popoviciu, Cluj-Napoca, 1999, 104-118.
[4] Jalobeanu, C., *Time Series Syntactic Analysis and Extrapolation*, International Symposium on Forecasting, Washington D.C., Session 1.10, 1999.
[5] Ge, X. Smith, P., *Deformable Markov Model Templates for Time-Series Pattern Matching*, Technical report UCI-ICS 00-10, Dep. Of Inform. And Computer Science, Univ. California, Irvine, 2000.
[6] Singh, S., *Pattern Recognition Modelling in Time Series Forecasting*, Cybernetics and System - An International Journal, 31, 1, 2000.
[7] Vidal, E., Tholland, F., de la Higuera, C., Casacuberta, F., Carraasco, R.C., *Probabilistic Finite Automata*, Part I, IEEE Transaction PAMI (in print).
[8] Weigend, A.S., Gershenfeld, N.A. (eds), *Time Series Prediction, Forecasting the Future and Understanding the Past*, Addison Wesley Publ. Comp., Reading, New-York, 1994.

Ciresica Jalobeanu
Department of Mathematics
Technical University of Cluj-Napoca