PERGAMON

Chaos, Solitons and Fractals 11 (2000) 2325–2330

CHAOS SOLITONS & FRACTALS

www.elsevier.nl/locate/chaos

Quantum model of emerging grammars

Michail Zak

Ultracomputing Group, Jet Propulsion Laboratory, California Institute of Technology, Mail Stop 126-347, Pasadena, CA 91109, USA

Accepted 28 July 1999

Abstract

A special class of quantum recurrent nets (QRNs) simulating Markov chains with absorbing states is introduced. The absorbing states are exploited for pattern recognition: each class of patterns is attracted to a unique absorbing state. Due to quantum interference of patterns, each combination of patterns acquires its own meaning: it is attracted to a certain combination of absorbing states which is different from those of individual attractions. This fundamentally new effect can be interpreted as formation of a grammar, i.e., a set of rules assigning certain meaning to different combinations of patterns. It appears that there exists a class of unitary operators in which each member gives rise to a different artificial language with associated grammar.

One of the oldest and most challenging problems is to understand the process of language formation. In this note, based upon recent successes in quantum information theory [1], and in particular, upon a concept of quantum recurrent nets (QRNs) [2], a new phenomenological formalism for pattern recognition and grammar formation is proposed.

A quantum recurrent network consists of a conventional quantum network augmented with a classical measurement and quantum reset operation. The design of a one-dimensional quantum recurrent network is shown in Fig. 1.

An initial state, $|\psi\rangle$, is fed into the network, transformed under the action of a unitary operator, U, subjected to a measurement, indicated by the measurement operator $M\{\ \}$, and the result of the measurement is used to control the new state fed back into the network at the next iteration. One is free to record, duplicate or even monitor the sequence of measurement outcomes, as they are all merely bits and hence constitute classical information. Moreover, one is free to choose the function used during the reset phase, including the possibility of adding no offset state whatsoever. Such flexibility makes the QRN architecture remarkably versatile. To simulate a Markov process, it is sufficient to return just the last output state to the next input at each iteration.

For a proof-of-concept, we will start with the following unitary N-dimensional operator $\overset{0}{U}$

$$\overset{0}{U} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} & 0 & \cdots & 0 \\ u_{21} & u_{22} & \cdots & u_{2n} & 0 & \cdots & 0 \\ u_{n1} & u_{n2} & \cdots & u_{nn} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}, \quad n < N, \tag{1}$$

which maps the ith eigenvector into a jth eigenvector

E-mail address: zak@solstice.jpl.nasa.gov (M. Zak).

0960-0779/00/$ - see front matter
PII: S 0 9 6 0 - 0 7 7 9 (9 9) 0 0 1 5 3 - 8

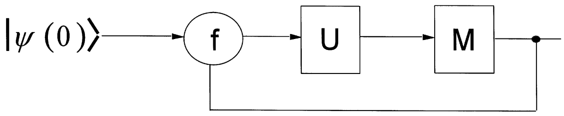

Fig. 1. A one-dimensional quantum recurrent network.

$$\left\{00\ldots 0\underset{\substack{\uparrow\\ i}}{1}0\ldots 0\right\} \to \left\{0\ldots 00\underset{\substack{\uparrow\\ j}}{1}0\ldots 0\right\} \tag{2}$$

with the probability

$$p_i^j = |u_{ji}|^2. \tag{3}$$

Eq. (3) is modified to the following [2]

$$p_i^j = \frac{\left|\sum_{k=1}^{n} u_{jk}a_k + u_{ji}\right|^2}{\left|\sum_{k\neq i} a_k^2 + (a_i+1)^2\right|^2} \tag{4}$$

if each result of the measurement is combined with an arbitrary offset vector

$$|\psi'\rangle = \{a_1 \ldots a_n\}. \tag{5}$$

It should be emphasized that the sum of the output vector in (2) and the offset vector (5) is first calculated, normalized, and then the corresponding quantum re-entering state is prepared.

For the purpose of pattern recognition, the offset vector will be chosen as follows:

$$|\psi_0'\rangle = \begin{cases} \{a_1, a_2, \ldots, a_N\} & \text{if } i \leqslant n, \\ 0 & \text{if } i > n, \end{cases} \tag{6}$$

where i is defined by Eq. (2).

Now the probability of the mapping (2) performed by the unitary operator $\overset{0}{U}$ and the offset vector (6) can be obtained by combining Eqs. (3) and (4), and the transition matrix for the corresponding Markov chain is

$$P_1 = \begin{pmatrix} P_1^1 & \cdots & P_1^n & P_1^{n+1} & \cdots & P_1^N \\ P_2^1 & \cdots & P_2^n & P_2^{n+1} & \cdots & P_2^N \\ P_n^1 & \cdots & P_n^n & P_n^{n+1} & \cdots & P_n^N \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad 0 \leqslant P_i^j < 1, \qquad \sum_{j=1}^{N} P_i^j = 1. \tag{7}$$

This chain has n transient states T_p $(p = 1, 2, \ldots, n)$ and $N_1 - n$ absorbing states A_γ $(\gamma = n+1, n+2, \ldots, N)$, and therefore, regardless of an initial state, the stochastic process eventually will be trapped in one of the absorbing states A_k. However, the probability that it will be a prescribed state A_γ depends upon the initial state. Indeed, as follows from theory of Markov chains, the probability f_p^k of absorption into A_k from T_p satisfies the system of equations

$$f_p^k = \sum_{j=0}^{n} p_p^j f_j^k \quad \text{for } p = 0, 1 \ldots, n; \;\; k = n+1, n+2, \ldots, N. \tag{8}$$

Consequently, by appropriate choice of $\overset{0}{U}$ and $|\psi_0'\rangle$ in Eqs. (1) and (6), one can divide all the initial states into $N_1 - n$ groups such that each state of the group is absorbed (with a sufficiently high probability) into the same prescribed state. Such a performance can be interpreted as pattern classification if each eigenvector introduced to the QRN is associated with the corresponding patterns.

We will not go into mathematical details here in order to focus attention upon formation of an artificial language instead. For that purpose, suppose that each run of the quantum device is repeated ℓ times while $\ell \leqslant n$ independent measurements are collected and fed back into QRN. Then, instead of mapping (2), one arrives at the following:

$$\frac{1}{\sqrt{\ell}}\left\{0\ldots 0\underset{\substack{\uparrow\\ i_1}}{1}0\ldots 0\underset{\substack{\uparrow\\ i_\ell}}{1}0\ldots 0\right\} \to \frac{1}{\sqrt{\ell}}\left\{00\ldots 0\underset{\substack{\uparrow\\ j_1}}{1}0\ldots 00\underset{\substack{\uparrow\\ j_m}}{1}0\ldots 0\right\}, \quad m \leqslant \ell. \tag{9}$$

This corresponds to evolution of k different patterns introduced to QRN simultaneously.

One can generalize Eq. (4) to the following

$$P_{i_1\ldots i_l}^{j_1\ldots j_m} = \frac{\prod_{\alpha=1}^{m}\left|\sum_{k=1}^{n} U_{j_\alpha k} a_k + \sum_{\beta=1}^{\ell} U_{j_\alpha i_\beta}\right|^2}{\left|\sum_{k\neq i} a_k^2 + \sum_{\beta=1}^{\ell}\left(a_{i_\beta}+1\right)^2\right|^{2m}}, \quad m \leqslant \ell, \tag{10}$$

by considering how each of the recurrent states combined with the offset vector (5) evolves under the action of $\overset{0}{U}$.

Eq. (10) defines the probability of transition from the set of inputs $i_1 \ldots i_\ell$ to the set of outputs $j_1 \ldots i_m$.

If $m = \ell$, and the offset vector is expressed by Eq. (6), the transition probability matrix P_ℓ can be presented in the form similar to P_1 in Eq. (7):

$$P_\ell = \begin{pmatrix} p_{11\ldots 1}^{11\ldots 1}\cdots & P_{11\ldots 1}^{nn\ldots n} & \cdots & P_{11\ldots 1}^{N^\ell N^\ell \ldots N^\ell} \\ p_{nn\ldots n}^{11\ldots 1}\cdots & P_{nn\ldots n}^{nn\ldots n} & \cdots & P_{nn\ldots n}^{N^\ell N^\ell \ldots N^\ell} \\ 0\cdots & 010 & \cdots & 0 \\ 0\cdots & \cdots & \cdots & 01 \end{pmatrix}. \tag{11}$$

This means that the corresponding ℓ-variate stochastic process has n^ℓ transient states T_p $(p = 1, 2, \ldots, n^\ell)$ and $N^\ell - n^\ell$ absorbing states A_γ, and therefore, $\binom{n}{\ell}$ combinations of ℓ different patterns (in the form of normalized sums of ℓ different eigenstates) are mapped onto $N_c \leqslant \binom{n}{\ell}$ different classes. Hence, the total number of pattern combinations which can be classified by the QRN is

$$S = \sum_{\ell=1}^{n}\binom{n}{\ell} = 2^n. \tag{12}$$

Now the performance of the QRN can be given the following interpretation. As soon as the unitary matrix U and the offset vector $|\psi'\rangle$ are chosen (see Eqs. (1) and (6)), all the transition matrices P_k $(k = 1, 2, \ldots, \ell)$ are uniquely defined (see Eqs. (4), (7), (10) and (11)). It should be noticed that these matrices do not have to be implemented: they exist in an abstract mathematical space being induced by $\overset{0}{U}$ and $|\psi_0'\rangle$. If only one measurement is fed back $(\ell = 1)$, then the transition matrix (7) manipulates by basic patterns-eigenstates which can be identified with "letters" of an alphabet: by mapping each eigenvector into a corresponding class, it assigns a certain meaning to the letter. If ℓ independent measurements are fed back $(1 < \ell \leqslant n)$, then the transition matrix (11) assigns certain meanings to combinations of letters, i.e., to ℓ-letter "words". In order to understand the rules of these assignments, i.e., the "grammar", let us turn to Eq. (10). As follows from there, in general

$$P_{i_1\ldots i_\ell}^{j_1\ldots j_\ell} \neq P_{i_1}^{j} \otimes \cdots \otimes P_{i_\ell}^{j_\ell}, \tag{13}$$

i.e., an ℓ-variate stochastic process is not simply the product of ℓ underlying one-dimensional stochastic processes, and the difference

$$\Delta_{i_1\ldots i_\ell}^{j_1\ldots j_\ell} = \left|P_{i_1\ldots i_\ell}^{j_1\ldots j_\ell} - P_{i_1}^{j_1} \otimes P_{i_2}^{j_2} \otimes \cdots \otimes P_{i_\ell}^{j_\ell}\right| \tag{14}$$

expresses the amount of "novelty", or new information created by interaction between different patterns via quantum interference. Formally Eq. (14) resembles quantum entanglement which is also responsible for creation of a new information; however, actually this entanglement is not quantum: it is a correlation between several classical stochastic processes generated by quantum interference.

It should be recalled that classical neural nets where patterns are stored at dynamical attractors, do not have a grammar: any combination of patterns is meaningless unless their storage is specially arranged, and that would require actual implementation of an exponential number of new attractors (see Eq. (12)).

Thus, each unitary operator $\overset{0}{U}$ having the structure (1) and supplied with an offset vector $|\psi_0'\rangle$ of the type (6) generates a new grammar. Since the structure (1) is preserved under matrix products, new operators

$$\overset{0^{1}}{U} = \overset{0}{U}_1 \overset{0}{U}_2, \quad \overset{0^{11}}{U} = \overset{0}{U}_1 \otimes \overset{0}{U}_2 \tag{15}$$

as well as their linear combinations represent new grammar. In particular, if the time period of each run of the QRN is increased in q times, then the effective unitary operator will be different from the original one

$$\overset{0^{1}}{U} = \overset{0^{q}}{U} \tag{16}$$

and thereby a set of new languages can be generated by the same quantum "hardware". The second equation in (15) opens up a possibility to build a high-dimensional operator $\dot{U}$ from low-dimensional components of the same structure.

It is worth mentioning that not every language of the possible set of languages is useful. Indeed, the performance of the QRN, and in particular, the assignments of pattern combinations to specific absorbing states is probabilistic. It is reasonable to require that for each selected patterns combination, the corresponding absorbing probability distribution over all possible states has a well-pronounced preference for a certain state; otherwise a word would lose its stable meaning. (It should be noticed that small overlapping of absorbing states is acceptable: it makes the language more colorful by incorporating double-meaning to some words.) As mentioned earlier, stability of the meaning of the basic patterns, i.e., letters, can be achieved by an appropriate choice of $\overset{0}{U}$ and $|\psi_0'\rangle$ based upon solutions of Eq. (8). However, as soon as $\overset{0}{U}$ and $|\psi_0'\rangle$ are fixed, there is no further control over stability of words' meaning since all the transition matrices P_ℓ are already predetermined (see Eqs. (10) and (11)). In this situation, one can characterize the effectiveness of the language by the ratio ζ of the number W of useful words to the total number of words S

$$\zeta = \frac{W}{S}, \quad S \sim \mathrm{O}(2^n). \tag{17}$$

Hence, to maximize ζ one has to identify such a solution to Eq. (8) which simultaneously stabilizes the meanings of all the letters as well as most of the words. Obviously, in general, this problem is hard.

In order to demonstrate the existence of effective emerging grammars, consider the following example.

Suppose that in Eqs. (1) and (6)

$$U = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 & 0 \\ -\sin\varphi & \cos\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad a = (0, 0, a_3 a_4), \tag{18}$$

where φ, a_3 and a_4 are real.

Then, applying Eq. (4) one finds the elements of the transition matrix P_1 (see Eq. (7)):

$$\begin{aligned} &P_1^1 = P_2^2 = \gamma_1 \cos^2\varphi, \quad P_1^2 = P_2^1 = \gamma_1 \sin^2\varphi, \\ &P_1^3 = P_2^3 = \gamma_1 a_3^2, \quad P_1^4 = P_2^4 = \gamma_1 a_4^2 \\ &P_3^1 = P_3^2 = P_3^4 = P_4^1 = P_4^2 = P_4^3 = 0, \quad P_3^3 = P_4^4 = 1, \quad \gamma_1 = \frac{1}{a_3^2 + a_4^2 + 1}. \end{aligned} \tag{19}$$

As follows from Eq. (19), there are two transient states (T_1 and T_2), and two absorbing states (A_3 and A_4).

Introducing four input patterns

$$|\psi_1\rangle = \{1000\}, \quad |\psi_2\rangle = \{0100\}, \quad |\psi_3\rangle = \{0010\}, \quad |\psi_4\rangle = \{0001\} \tag{20}$$

as well as their images in the probabilistic space

$$\pi_1 = \{1000\},\ \pi_2 = \{0100\}, \ldots, \pi_1 = |\psi_i\rangle, \tag{21}$$

first one can write down trivial mapping

$$|\psi_3\rangle \to \pi_3 \to A_3,\ f_3^3 = 1; \quad \text{and} \quad |\psi_4\rangle \to \pi_4 \to A_4,\ f_4^4 = 1. \tag{22}$$

Other transitions

$$|\psi_1\rangle \to \pi_1 \to A_3,\ |\psi_1\rangle \to \pi_4 \to A_4,\ |\psi_2\rangle \to \pi_2 \to A_3 \quad \text{and} \quad \pi_2 \to A_4 \tag{23}$$

are more complex, and they can be found from Eq. (8):

$$f_1^2 = P_1^1 f_1^3 + P_1^2 f_2^3 + P_1^3, \quad f_2^3 = P_2^1 f_1^3 + P_2^2 f_2^3 + P_2^3, \tag{24}$$

whence

$$f_1^3 = \frac{a_3^2}{a_3^2 + a_4^2}, \quad f_2^3 = \frac{a_4^2}{a_3^2 + a_4^2}. \tag{25}$$

Similarly one finds

$$f_1^4 = f_1^3, \quad f_2^4 = f_2^3. \tag{26}$$

Thus, if

$$a_3 \cong a_4, \tag{27}$$

the patterns $|\psi_1\rangle$ and $|\psi_2\rangle$ do not have any meaning; with the same probability they can be absorbed by the states A_3 or A_4. However, if

$$a_3 \gg a_4, \quad \text{or} \quad a_3 \ll a_4, \tag{28}$$

the same patterns are absorbed by only one state A_3 or A_4 and that assigns certain meaning to each of them.

For mapping combinations of patterns (20), one has to repeat twice each measurement before feeding it back. Now the input pattern's combinations will be the following:

$$|\psi_{12}\rangle = |\psi_{21}\rangle = \frac{1}{\sqrt{2}}\{1100\},\ |\psi_{13}\rangle = |\psi_{31}\rangle = \frac{1}{\sqrt{2}}\{1010\}, \ldots \tag{29}$$

but their image in the probabilistic space will be different from (20)

$$\pi_{12} = \pi_1 \otimes \pi_2,\ \ \pi_{13} = \pi_1 \otimes \pi_3, \ldots \tag{30}$$

Instead of listing all the 64 elements of the matrix P_2 (see Eqs. (10) and (11)), we will concentrate upon those which will be used in our analysis. First of all

$$P_{ii}^{\alpha\beta} = \begin{cases} 0 & \text{if } \alpha \neq i,\ \beta \neq i,\ i = 3, 4, \\ 1 & \text{otherwise,} \end{cases} \qquad P_{ij}^{\alpha\beta} = \begin{cases} 0 & \text{if } \alpha \neq i,\ \beta \neq j, \\ 1 & \text{otherwise.} \end{cases} \tag{31}$$

This means that there are four absorbing states: A_{33}, A_{34}, A_{43} and A_{44}; the rest 12 states (T_{12}, T_{13}, etc.) are transient. Here we will be interested only in the evolution of the pattern's combination $|\psi_{12}\rangle$ (see Eq. (29)) since it is the only one which entangles the patterns $|\psi_1\rangle$ and $|\psi_2\rangle$ (see Eq. (20)). (Other combinations: $|\psi_{13}\rangle$, $|\psi_{23}\rangle$, etc. are not entangled, and therefore, their evolution can be predicted from the previous analysis as direct products $|\psi_1\rangle \otimes |\psi_3\rangle, |\psi_2\rangle \otimes |\psi_3\rangle$, i.e., it does not have any novelty element.)

Thus, one obtains

$$
\begin{aligned}
&P_{12}^{11} = \gamma_2(\cos\varphi + \sin\varphi)^4, \quad P_{12}^{22} = \gamma_2(\cos\varphi - \sin\varphi)^4, \quad P_{12}^{12} = \gamma_2(\cos^2\varphi - \sin^2\varphi)^2 = P_{12}^{21},\\
&P_{12}^{13} = \gamma_2 a_3^2(\cos\varphi + \sin\varphi)^2, \quad P_{12}^{14} = \gamma_2 a_4^2(\cos\varphi + \sin\varphi)^2,\\
&P_{12}^{23} = \gamma_2 a_3^2(\cos\varphi - \sin\varphi)^2, \quad P_{12}^{24} = \gamma_2 a_4^2(\cos\varphi - \sin\varphi)^2, \quad P_{12}^{34} = P_{12}^{43} = \gamma_2 a_3^2 a_4^2,\\
&P_{12}^{33} = \gamma_2 a_3^4, \quad P_{12}^{44} = \gamma_2 a_4^4,
\end{aligned}
\tag{32}
$$

where

$$
\gamma_2 = \frac{1}{(a_3^2 + a_4^2 + 2)^2}. \tag{33}
$$

As follows from the last four equations in (32), there are direct transitions from the pattern combination $|\psi_{12}\rangle$ to the absorbing states. However, in addition to that, there exist many indirect transitions to the same states, for instance, $T_{12} \to T_{13} \to T_{33}$, $T_{12} \to T_{14} \to T_{44}$, etc. and these transitions include the entanglement effect which has maxima at $\varphi = \pm 1/\sqrt{2}$. As a result, the pattern combination $|\psi_{12}\rangle$ acquires a new meaning since it cannot be reduced to the direct product of the patterns $|\psi_1\rangle$ and $|\psi_2\rangle$.

The performance of this simple QRN becomes more sophisticated if the elements of U and a in Eq. (18) are complex numbers.

Utilizing the properties (15), one can represent a unitary operator $\overset{0}{U}$ in Eq. (1) in the form

$$
\overset{0}{U} = \left(U_1^{(1)} \otimes \cdots \otimes U_m^{(1)}\right) \cdot \left(U_1^2 \otimes \cdots \otimes U_m^{(2)}\right) \cdots \left(U_1^{(m)} \otimes \cdots \otimes U_m^{(m)}\right) \tag{34}
$$

gaining exponential dimensionality of U with linear resources.

Thus, it has been demonstrated that QRN is capable of creating emerging grammars by assigning different meanings to different combinations of letters. The paradigm is based upon quantum interference of patterns which entangles the corresponding Markov processes, and thereby, creates a new meaning depending upon how different patterns interact. The capacity of the language, i.e., the total number of words in it is exponential in n where n is dimensionality of the basic unitary operator. However, if this operator is presented as a direct product, then the number of words can be made double-exponential in the dimensionality of the quantum hardware.

The problems of hardware implementations of QRN have not been discussed in this note. However, since QRN operates by interleaving quantum evolution with measurement and reset operations, they are far less sensitive to decoherence than other designs of quantum computers.

Acknowledgements

The research described in this paper was performed by the Center for Space Microelectronics Technology, Jet Propulsion Laboratory and California Institute of Technology and was sponsored by the National Aeronautics and Space Administration and Office of Space Science and Technology.

References

[1] Bennett C. Quantum information and computations. Physics Today 1995;48:24–30.
[2] Zak M, Williams C. Quantum neural nets. Int J Theoret Phys 1998;37(2).