RAIRO. RECHERCHE OPÉRATIONNELLE

BRUNO VISCOLANI

A queueing problem in pattern recognition

RAIRO. Recherche opérationnelle, tome 18, nº 1 (1984), p. 71-88
<http://www.numdam.org/item?id=RO_1984__18_1_71_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 18, n° 1, février 1984, p. 71 à 88).

A QUEUEING PROBLEM IN PATTERN RECOGNITION (*)

by Bruno VISCOLANI (¹)

Abstract. — *Patterns arrive at random instants to a sequential recognizer and possibly form a queue. Each pattern is recognized through one or more consecutive stages following a definite descipline; the times spent by a pattern in the different stages are assumed to be independent random variables. The resulting model of the recognizer is a Markovian queueing system with a service discipline which aims to minimize both the queue-size and the error probability. The traffic intensity of the patterns and the probabilities of entry in the different stages are the parameters which determine the equilibrium probabilities of the system. From these, then, the error probability and other characteristics of interest follow. Afterwards the hypothesis of service-time independency is removed and other special cases are also considered.*

Keywords: Queueing systems, pattern recognition, error probability, equilibrium behaviour.

Résumé. — *Un discriminateur séquentiel classifie des formes, qui arrivent à instants aléatoires et peuvent constituer une file d'attente. Une discipline de service determine la classification de chaque forme par un ou plusieurs stades consécutifs; les durées des permanences d'une forme dans des stades différents sont supposées être des variables aléatoires indépendantes. Le modèle ainsi choisi est un système Markovien avec file d'attente et avec une discipline de service pour rendre minimale soit la longueur de la file, soit la probabilité d'erreur. La densité du trafic des formes et les probabilités de passage par les stades successifs déterminent les probabilités d'observer le système dans ses différents états en condition d'équilibre. La probabilité d'erreur et d'autres caractéristiques intéressantes en suivent. Successivement on quitte l'hypothèse de l'indépendance des temps de service et on étudie aussi d'autres cas particuliers.*

INTRODUCTION

We consider *patterns* from a set $\mathscr{P}$ and the *recognition problem* defined by a partition $\{\mathscr{P}_1, \mathscr{P}_2, \ldots, \mathscr{P}_k\}$ of $\mathscr{P}$. A *classification* $d(x)$ of a pattern $x \in \mathscr{P}$ is a number from $\{1, 2, \ldots, k\}$ and $d(x)$ is said to be *correct* if $d(x) = i$ and $x \in \mathscr{P}_i$, otherwise it is said to be *wrong*. Let us indicate by $d^*(x)$ the correct classification of pattern x, for every $x \in \mathscr{P}$.

A *sequential pattern recognizer* Σ is a device which, to every pattern x, associates a sequence of classifications $d_1(x), d_2(x), \ldots, d_m(x), \ldots$ The recognizer Σ takes a time t_m to perform the classification $d_m(x)$, $m \geqq 1$, and we

(*) Received October 1982.

(¹) Seminario Matematico dell'Università, Via Belzoni, 7-35131 Padova, Italy.

R.A.I.R.O. Recherche opérationnelle/Operations Research, 0399-0559/1984/71/$ 5.00

assume that t_m is a random variable. In the main model considered here, the t_m are independent exponential random variables.

A case of dependency among the classification times is also discussed.

Σ, after that $d_m(x)$ is computed, starts computing the next classification $d_{m+1}(x)$ of the same pattern x, with probability α_m, or stops the actual classification sequence, with probability $1-\alpha_m$.

In the latter case the recognizer Σ is ready to begin the classification sequence for a possible new pattern. Moreover, the recognizer Σ may be interrupted at every instant during its operation, except when it is computing the first classification, d_1.

We may think in general that the recognizer Σ performs the classification $d_m(x)$ using information carried by the first *m features* of the pattern x. A detailed discussion of such recognizers is found in Fu [1].

The number of features of a pattern may vary and it may have a maximum, say M: in such case we have that $\alpha_M=0$ and that any classification sequence has length $\leqq M$.

The recognizer Σ has a characteristic *error probability sequence* $e_1, e_2, \ldots, e_m, \ldots$; where e_m is the probability that $d_m(x)$ is not the correct classification of the pattern x:

$$e_m = \Pr(d_m \neq d^*).$$

In other terms, e_m is the probability of misrecognition after m steps of the recognition procedure. We assume that $e_1<1$ and that $e_m \geqq e_{m+1}$, $m \geqq 1$.

Patterns arrive to Σ as a Poisson birth process. If Σ is busy classifying a pattern x, the arriving patterns join a queue.

The goals of the recognizer are:

(i) to keep the queue-size as little as possible;

(ii) to minimize the error probability.

In order to satisfy the above requirements, the following *recognition discipline* is adopted.

(A) If the queue is non-empty, then Σ computes only the classification d_1 for the pattern which is actually processed; then it starts classifying the next pattern in queue.

(B) If the queue is empty and a new pattern x arrives:

(B.1) if Σ is idle, then it starts computing the classification sequence of the pattern x just arrived;

(B.2) if Σ is computing d_1 for the last pattern arrived before x, then it continues the operation following rule A (queue is no more empty);

(B.3) if Σ is computing d_m, $m > 1$, for the last pattern arrived before x, then it interrupts that operation and starts computing the classification sequence of the new pattern (thus Σ breaks the actual classification sequence at the element d_{m-1}).

The recognizer Σ, with its rules for fast recognition of patterns in queue, appear to be a Markovian queueing system. It is indeed a modified $M/M/1$ queueing system. Our aim is to study the equilibrium behaviour of the system. Besides typical characteristics of queueing systems, we are naturally led to study the error probability and the size of the classification sequences.

Moreover we have to deal with *waste time*, i. e. that part of working time that Σ spends computing any classification d_m, $m > 1$, but without finishing that operation because of a new pattern arrival.

In present work no attention is given to the alternative recognition discipline which would suggest to conclude any classification, once it is started.

The basic notions of Queueing theory are found, for example, in the books by Conolly [2] or by Kleinrock [3].

2. A QUEUEING SYSTEM

We assume that all classification times of any pattern are independent and exponentially distributed random variables, whose mean is $1/\mu$. The rate of the Poisson (pattern-) arrival process is λ. The set of the possible states of the system Σ is represented by:

$$S = \{(0, m): m \geqq 0\} \cup \{(n, 1): n \geqq 1\}.$$

where:

(i) (0, 0) means that Σ is idle;

(ii) $(0, m)$ means that there is one customer (pattern) which is having the m-th service (classification d_m) and the queue is empty;

(iii) $(n, 1)$ means that there are $n+1$ customers in the system, one of which is having the first service (classification d_1).

The state-transition-rate diagram is given in figure 1. Let us denote by p_{nm} the equilibrium probability for Σ to be in state $(n, m) \in S$ and put $\rho = \lambda/\mu$. By inspection of the state-transition-rate diagram we establish the following equilibrium equations:

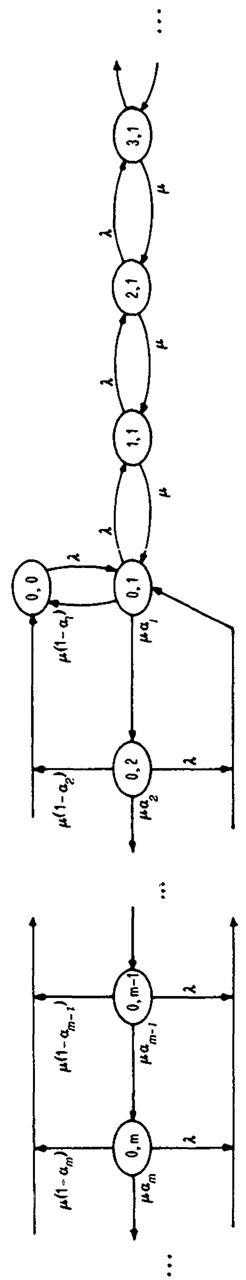

Figure 1. — State-transition-rate diagram of Σ.

$$\rho p_{00}=\sum_{m=1}^{\infty}(1-\alpha_m)\,p_{0m};\tag{1}$$

$$(1+\rho)\,p_{01}=\rho p_{00}+\rho\sum_{m=2}^{\infty}p_{0m}+p_{11};\tag{2}$$

$$(1+\rho)\,p_{0m}=\alpha_{m-1}p_{0m-1},\qquad m\geqq 2;\tag{3}$$

$$(1+\rho)\,p_{n1}=\rho p_{n-1,1}+p_{n+1,1},\qquad n\geqq 1.\tag{4}$$

Moreover it must hold that:

$$\sum_{m=0}^{\infty}p_{0m}+\sum_{n=1}^{\infty}p_{n1}=1.\tag{5}$$

The following definitions will be useful:

$$\chi_0=1;\tag{6}$$

$$\chi_m=\frac{\alpha_m}{1+\rho}\chi_{m-1},\qquad m\geqq 1;\tag{7}$$

$$\chi^*=\sum_{m=0}^{\infty}\chi_m.\tag{8}$$

We observe that $(1+\rho)^m\chi_m=\prod_{i=1}^{m}\alpha_i$ is the probability that the classification d_{m+1} is required, to recognize any pattern, provided that there is a sufficiently large delay between the start of the classification d_1, with empty queue, and the next pattern arrival.

In view of the above definitions, equations (3) and (1) give respectively:

$$p_{0m}=\chi_{m-1}p_{01},\qquad m\geqq 1;\tag{9}$$

$$\rho p_{00}=(1+\rho-\rho\chi^*)p_{01}.\tag{10}$$

From equation (4) we obtain that:

$$p_{n1}=\rho^n p_{01},\qquad n\geqq 0.\tag{11}$$

Substitution of (9), (10) and (11) in (5) gives:

$$p_{01}=\rho(1-\rho),\tag{12}$$

so that we obtain:

$$p_{n1}=(1-\rho)\,\rho^{n+1},\qquad n\geqq 0;\tag{13}$$

$$p_{00}=(1-\rho)(1+\rho-\rho\chi^*); \tag{14}$$

$$p_{0m}=\rho(1-\rho)\,\chi_{m-1}, \qquad m\geqq 2. \tag{15}$$

It is easily verified that equation (2) is satisfied. The condition for our system to be ergodic is, as for $M/M/1$, that:

$$\rho<1. \tag{16}$$

We note that p_{00}, given by (14), is always non-negative; in fact, from $0\leqq\alpha_m\leqq 1$ it follows that:

$$\chi^*=1+\sum_{m=1}^{\infty}(1+\rho)^{-m}\prod_{i=1}^{m}\alpha_i\leqq\sum_{m=0}^{\infty}(1+\rho)^{-m}=1+\rho^{-1}, \tag{17}$$

where equality holds only if $\alpha_m=1$, for every m.

The system Σ behaves indeed just like the $M/M/1$ queueing system as far as the queue is non-empty. Σ differs from $M/M/1$ only in that it takes advantage of part of the $M/M/1$'s idle time, working to reduce the error probability.

3. PATTERNS TRAFFIC

Let us consider first the number N of patterns in the system and define:

$$v_n=\Pr\{N=n\}, \qquad n\geqq 0. \tag{18}$$

We have the probability generating function:

$$v(z)=\sum_{n=0}^{\infty}v_n z^n=(1-\rho)\left[1+\rho-\rho\chi^*+\rho\chi^* z+\rho^2\frac{z^2}{1-\rho z}\right]. \tag{19}$$

The average number of patterns in the system is:

$$\bar{N}=\frac{\rho}{1-\rho}\left[(1-\rho)^2\chi^*+(2-\rho)\rho\right], \tag{20}$$

which ranges from $\rho/(1-\rho)$ to $1-\rho+\rho/(1-\rho)$ as $\chi^*\in[1, 1+\rho^{-1}]$. The lower bound $\rho/(1-\rho)$ corresponds to $\chi^*=1$ and is of course the value of $\bar{N}$ in $M/M/1$.

The variance of N may well be grater or less than the value $\rho(1-\rho)^{-2}$, which holds for $M/M/1$, i.e. when $\chi^*=1$. From Little's formula we derive that the average time $\overline{T}$ that a pattern spends in the recognizer (in queue and during its recognition) is:

$$\overline{T}=\frac{1}{\lambda}\,\overline{N}=\frac{1/\mu}{1-\rho}\,[(1-\rho)^2\,\chi^*+(2-\rho)\,\rho]. \tag{21}$$

We see that, if $\alpha_1 \neq 0$, then $\chi^* > 1$, and $\overline{T}$ is greater than the average time spent by a customer in $M/M/1$. Moreover the ratio between the two average times tends to 1 as $\rho \to 1$ and to χ^* as $\rho \to 0$.

As far as the queue-size N_q is concerned, Σ behaves just like $M/M/1$. In fact, let:

$$q_n=\Pr\{N_q=n\}, \qquad n \geqq 0, \tag{22}$$

then we have that:

$$q_0=\sum_{m=0}^{\infty} p_{0m}=1-\rho^2, \tag{23}$$

$$q_n=(1-\rho)\,\rho^{n+1}, \qquad n \geqq 1. \tag{24}$$

These probabilities are independent of χ^* and the average queue-size is:

$$\overline{N}_q=\frac{\rho^2}{1-\rho}. \tag{25}$$

Little's formula tells us that the average time spent in queue is:

$$\overline{T}_q=\frac{1}{\mu}\cdot\frac{\rho}{1-\rho}. \tag{26}$$

It is easy to see that the time T_q, that a pattern has to wait in queue, is null if Σ is in a state $(0, m)$, $m \neq 1$, at the time of the pattern-arrival; and T_q is an Erlang E_{n+1} random variable, with mean $(n+1)/\mu$, if Σ is in $(n, 1)$, $n \geqq 0$, when the pattern joins the queue.

4. TIME EXPLOITATION

The system Σ is:

(i) *busy* when it is in a state $(n, 1)$, $n \geqq 0$;

(ii) *idle* when it is in the state $(0, 0)$;

(iii) *not busy nor idle* when it is in a state $(0, m)$, $m \geqq 2$.

In case (iii), in fact, Σ is working, but a pattern x which arrives finds Σ ready to start at once its own classification sequence.

The fraction of time when Σ is busy is ρ, as in $M/M/1$; the one when Σ is idle is $(1-\rho)(1+\rho-\rho\chi^*)$. The remaining fraction $\rho(1-\rho)(\chi^*-1)$ of time is spent partly for *effective work*, which gives better classifications, partly for *ineffective work*, which is interrupted before that the pursued classification is reached.

We observe that the idle time become null when χ^* reaches its maximum $1+\rho^{-1}$, and the fraction (iii) of time is $1-\rho$.

The event that a pattern arrives before that the actual classification is performed, or $\{T_\lambda < T_\mu\}$ with $T_\lambda \sim \lambda e^{-\lambda t}$, $T_\mu \sim \mu e^{-\mu t}$, has probability:

$$\Pr\{T_\lambda < T_\mu\} = \int_0^\infty \mu e^{-\mu t} \int_0^t \lambda e^{-\lambda s}\, ds\, dt = \rho(1+\rho)^{-1}. \tag{27}$$

We can now determine the fraction of time spent by Σ doing ineffective work:

Pr (Σ is doing ineffective work)

$$= \sum_{m=2}^{\infty} p_{0m} \cdot \Pr(\text{a pattern arrives before } m\text{-th classification is performed})$$

$$= \rho^2 \frac{1-\rho}{1+\rho} \sum_{m=2}^{\infty} \chi_{m-1} = \rho^2 \frac{1-\rho}{1+\rho}(\chi^*-1). \tag{28}$$

This probability is an increasing function of χ^* and ranges from 0 to $\rho(1-\rho)(1+\rho)^{-1}$.

Let Σ conclude the classification d_m at epoch t_0 and start computing the classification d_{m+1}: let us assume that at epoch t_0+T, $T>0$, a pattern arrives, but the classification d_{m+1} is not concluded. T is the time spent in ineffective work (waste time) and it has the following conditional probability distribution function:

$$\begin{aligned} F(t \mid T>0) &= \Pr\{T \leqq t \mid T>0\} = \Pr\{T_\lambda \leqq t \mid T_\lambda < T_\mu\} \\ &= \Pr\{T_\lambda \leqq t \text{ and } T_\lambda < T_\mu\}.(\Pr\{T_\lambda < T_\mu\})^{-1} \\ &= \rho^{-1}(1+\rho)\,[\Pr\{T_\lambda < t \mid T_\mu > t\}.\Pr\{T_\mu > t\} \\ &\quad + \Pr\{T_\lambda < T_\mu \mid T_\mu < t\}\,\Pr\{T_\mu < t\}] \\ &= \rho^{-1}(1+\rho)[\Pr\{T_\lambda < t\}\Pr\{T_\mu > t\} + \Pr\{T_\lambda < T_\mu \text{ and } T_\mu < t\}] \\ &= \rho^{-1}(1+\rho)\left[\int_0^t \lambda e^{-\lambda z}\,dz . \int_t^\infty \mu e^{-\mu s} ds + \int_0^t \mu e^{-\mu s} \int_0^s \lambda e^{-\lambda z} dz\, ds\right] \\ &= 1 - e^{-(\lambda+\mu)t}. \end{aligned} \tag{29}$$

Thus we have obtained that the waste time T is exponentially distributed provided that $T>0$, with mean:

$$E[T \mid T>0]=\frac{1}{\mu} \cdot \frac{1}{1+\rho}. \tag{30}$$

Let us denote by A the event that a pattern is recognized with waste time. Then Σ wastes time in the recognition of the fraction $\Pr(A)$ of the patterns which arrives to it. Therefore $\lambda \Pr(A)$ is the rate of patterns recognized with waste time and finally $\lambda \Pr(A) E[T \mid T>0]$ is the fraction of total time which is wasted, i. e. the probability that Σ is doing ineffective work. Thus, by (28) we have that:

$$\lambda \Pr(A) E[T \mid T>0]=\rho^2(1-\rho)(1+\rho)^{-1}(\chi^*-1),$$

and, by (30), that:

$$\Pr(A)=\rho(1-\rho)(\chi^*-1). \tag{31}$$

5. ERROR PROBABILITY

Let L be the event that a pattern has no queue behind itself when it starts being processed for d_1; let $\bar{L}$ be its contrary. If L does not occur for a given pattern x, then x will get only the classification d_1, with error probability e_1.

For a fixed pattern x, the occurrence of L depends on the state of Σ when the pattern x arrives.

Let $\Pr(L \mid (n, m))$ be the probability of L, given that Σ was in (n, m) when x arrived.

First we have that:

$$\Pr(L \mid (0, m))=1, \qquad m \geqq 0, \qquad m \neq 1, \tag{32}$$

because the pattern x, just arrived, starts at once being classified and in a Poisson birth process we cannot have two births at the same epoch, which would make a queue to rise.

Second, we observe that $\Pr(L \mid (n, 1))$, $n \geqq 0$, is the probability that no pattern arrives until $n+1$ consecutive classifications are performed. Independency of the times of the $n+1$ classifications and the memoryless property of the exponential distribution of interarrival times lead to:

$$\Pr(L \mid (n, 1))=[\Pr(L \mid (0, 1))]^{n+1}, \qquad n \geqq 0. \tag{33}$$

From:

$$\Pr(L \mid (0, 1)) = \int_0^\infty \mu e^{-\mu t} \int_t^\infty \lambda e^{-\lambda s}\, ds\, dt = (1+\rho)^{-1}, \tag{34}$$

we derive that:

$$\Pr(L \mid (n, 1)) = (1+\rho)^{-n-1}, \qquad n \geqq 0. \tag{35}$$

Finally, the unconditional probability that a pattern starts being classified with empty queue is:

$$\Pr(L) = 1 - \rho^2. \tag{36}$$

For a fixed pattern x and for $m \geqq 1$, let us define the following events:
D_m^+: the classification sequence contains d_m;
D_m: the classification sequence stops with d_m.
Let us further define, for $m \geqq 1$:
A_m: a pattern arrives during d_m computation.
In the case that L occurred, we obtain that:

$$\Pr(D_1^+ \mid L) = 1;$$

$$\Pr(D_2^+ \mid L) = \Pr(\overline{A}_1) . \Pr(d_2 \text{ is required} \mid D_1^+) . \Pr(\overline{A}_2) = (1+\rho)^{-2} \alpha_1;$$

$$\Pr(D_{m+1}^+ \mid L) = \Pr(D_m^+ \mid L) . \Pr(d_{m+1} \text{ is required} \mid D_m^+) . \Pr(\overline{A}_{m+1})$$
$$= (1+\rho)^{-1} \alpha_m \Pr(D_m^+ \mid L), \qquad m \geqq 2.$$

Thus, from definition (6) and (7) it follows that:

$$\Pr(D_m^+ \mid L) = \begin{Bmatrix} 1, & m=1; \\ (1+\rho)^{-1} . \chi_{m-1}, & m \geqq 2. \end{Bmatrix} \tag{37}$$

Now we are able to determine:

$$\Pr(D_1 \mid L) = \Pr\{A_1 \text{ or } \overline{A}_1 \text{ and } [(d_2 \text{ is not required} \mid D_1^+) \text{ or } (d_2 \text{ is required} \mid D_1^+) \text{ and } A_2]\}$$
$$= (1+\rho)^{-1} (\rho + \chi_0 - \chi_1);$$

$$\Pr(D_m \mid L) = \Pr(D_m^+ \mid L) . \Pr\{(d_{m+1} \text{ is not required} \mid D_m^+) \text{ or } (d_{m+1} \text{ is required} \mid D_m^+) \text{ and } A_{m+1}\}$$
$$= [1 - (1+\rho)^{-1} \alpha_m] . \Pr(D_m^+ \mid L), \qquad m \geqq 2.$$

Explicitly we have the result:

$$\Pr(D_m \mid L) = (1+\rho)^{-1}(\chi_{m-1} - \chi_m + \rho\delta_{m1}), \qquad m \geqq 1, \tag{38}$$

where $\delta_{m1} = 1$ if $m = 1$ and $\delta_{m1} = 0$ if $m \neq 1$.

On the other hand, if L did not occur, we have that:

$$\Pr(D_m \mid \overline{L}) = \delta_{m1}, \qquad m \geqq 1. \tag{39}$$

Finally, from (36), (38) and (39) we obtain the unconditional probabilities:

$$\Pr(D_m) = \rho\delta_{m1} + (1-\rho)(\chi_{m-1} - \chi_m), \qquad m \geqq 1. \tag{40}$$

The error probability then is given by:

$$e = \sum_{m=1}^{\infty} e_m \Pr(D_m) = \rho\, e_1 + (1-\rho) \sum_{m=1}^{\infty} e_m (\chi_{m-1} - \chi_m). \tag{41}$$

Another characteristic of interest is the expected size of the classification sequence, which, by (40), is:

$$\overline{m} = \rho + (1-\rho) \sum_{m=1}^{\infty} m(\chi_{m-1} - \chi_m) = \rho + (1-\rho)\chi^*. \tag{42}$$

REMARK. — We can now determine directly the probability, $\Pr(A)$, that a pattern is recognized with waste time. We obtained $\Pr(A)$ as result (31) in the preceding section by means of indirect arguments.

We have the following identity:

$$\Pr(A) = \Pr(A \mid \overline{L}) \Pr(\overline{L}) + \Pr(A \mid L) \Pr(L) = (1-\rho^2) \Pr(A \mid L), \tag{43}$$

because $\overline{L}$ implies that no classification d_m, $m \geqq 2$, is started: $\Pr(A \mid \overline{L}) = 0$.

We note that, for $m \geqq 3$, it holds:

$$\begin{aligned} \Pr(A \mid D_{m-1}^+, L) = {} & \Pr(d_m \text{ is required} \mid D_{m-1}^+) \\ & \times [\Pr(A_m) + \Pr(\overline{A}_m) \Pr(A \mid D_m^+, L)] \\ & = \alpha_{m-1} [\rho(1+\rho)^{-1} + (1+\rho)^{-1} \Pr(A \mid D_m^+, L)]. \end{aligned} \tag{44}$$

Let us define:

$$\sigma_m = \frac{\alpha_m}{1+\rho} [\rho + \Pr(A \mid D_m^+, L)], \qquad m \geqq 1. \tag{45}$$

Then equation (44) is rewritten as:

$$\sigma_m = \frac{\alpha_m}{1+\rho}(\rho+\sigma_{m+1}), \qquad m \geqq 1, \tag{44'}$$

and:

$$\Pr(A \mid D_m^+, L) = \sigma_m, \qquad m \geqq 2. \tag{44''}$$

When $m=1$ we have that:

$$\Pr(A \mid D_1^+, L) = \Pr(\bar{A}_1).\sigma_1 = (1+\rho)^{-1}.\sigma_1. \tag{46}$$

Moreover it holds that:

$$\Pr(A \mid L) = \Pr(A \mid D_1^+, L).\Pr(D_1^+ \mid L) = \Pr(A \mid D_1^+, L). \tag{47}$$

Now (44') gives:

$$\sigma_1 = \chi_1 \left\{ \rho + \frac{\alpha_2}{1+\rho} \left\{ \rho + \frac{\alpha_3}{1+\rho} \{ \rho + \dots \} \right\} \right\} = \rho \sum_{m=1}^{\infty} \chi_m = \rho(\chi^* - 1). \tag{48}$$

Finally we obtain that:

$$\Pr(A) = (1-\rho^2)(1+\rho)^{-1} \rho(\chi^* - 1) = \rho(1-\rho)(\chi^* - 1), \tag{49}$$

which is the same result that we obtained in (31), preceding Section.

6. DEPENDENT CLASSIFICATION TIMES

It seems quite natural to assume that classification times of different patterns are independent random variables. On the other hand, dependency of some kind may exist among the times required for the classifications $d_1(x)$, $d_2(x)$, ..., $d_m(x)$, ... for a fixed pattern x.

Let t_m, $m \geqq 1$, be the time required to perform the classification d_m. As before we assume that t_1 is exponentially distributed with mean $1/\mu$.

Let the sequence:

$$\tau_1 = 1, \qquad \tau_2, \dots, \qquad \tau_m, \dots, \tag{50}$$

be given. We assume that t_m is proportional to t_1, with ratio τ_m:

$$t_m = \tau_m t_1, \qquad m \geqq 2. \tag{51}$$

Let us define the numbers:

$$\sigma_m = \sum_{i=1}^{m} \tau_i, \qquad m \geqq 1. \tag{52}$$

As before, T_λ and T_μ indicate exponential random variables with means $1/\lambda$ and $1/\mu$ respectively. The events L, D_m and A_m, $m \geqq 1$, are the ones defined in Section 5.

Under the actual dependency hypotheses we have that:

$$\Pr(D_m \mid \overline{L}) = \delta_{m1}, \qquad m \geqq 1; \tag{53}$$

$$\Pr(D_1 \mid L) = \Pr(A_1) + \Pr(\overline{A}_1)\,[(1-\alpha_1) + \alpha_1 \Pr(A_2 \mid \overline{A}_1)]; \tag{54}$$

and, putting $A_0 = \emptyset$:

$$\Pr(D_m \mid L) = \prod_{j=1}^{m} \Pr(\overline{A}_j \mid \overline{A}_{j-1}) \times \prod_{j=1}^{m-1} \alpha_j [(1-\alpha_m) + \alpha_m \Pr(A_{m+1} \mid \overline{A}_m)], \qquad m \geqq 2. \tag{55}$$

We note that:

$$\prod_{j=1}^{m} \Pr(\overline{A}_j \mid \overline{A}_{j-1}) = \Pr\left(\prod_{j=1}^{m} \overline{A}_j\right)$$
$$= \Pr \text{ (no pattern arrives before that classification } d_m \text{ is computed)}$$
$$= \Pr\left\{ T_\lambda > \sum_{i=1}^{m} t_i \right\} = \Pr\{ T_\lambda > \sigma_m T_\mu \}$$
$$= (1+\rho\sigma_m)^{-1}, \qquad m \geqq 1, \tag{56}$$

which gives:

$$\Pr(D_m \mid L) = \frac{\rho}{1+\rho}\delta_{m1} + (1+\rho)^{m-1} \chi_{m-1} \frac{(1+\rho\sigma_{m+1}) - \alpha_m(1+\rho\sigma_m)}{(1+\rho\sigma_m)(1+\rho\sigma_{m+1})}, \qquad m \geqq 1. \tag{57}$$

Thus, from (53) and (57) we obtain:

$$\Pr(D_m) = \rho\delta_{m1} + (1-\rho^2)(1+\rho)^{m-1} \chi_{m-1} \frac{(1+\rho\sigma_{m+1}) - \alpha_m(1+\rho\sigma_m)}{(1+\rho\sigma_m)(1+\rho\sigma_{m+1})}, \qquad m \geqq 1. \tag{58}$$

Using (58) we are now able to compute the error probability and the expected size of the classification sequence, as we did for (41) and (42) of Section 5.

It is more interesting to find the conditional distributions of the time T spent by Σ in ineffective work, given that $T>0$ and that the event D_m occurred, $m\geqq 1$.

Let Σ conclude the classification d_{m-1} and start computing the classification d_m at epoch t_0; let us assume that at epoch t_0+T, $T>0$, a pattern arrives, but the classification d_m is not concluded (and will not be concluded in fact). The conditional probability distribution of the waste time T is:

$$\begin{aligned}F_{m-1}(t \mid T>0) &= \Pr\{T<t \mid D_{m-1}, T>0\} = \Pr\{T_\lambda<t \mid T_\lambda<\tau_m T_\mu\}\\ &= (\Pr\{T_\lambda<\tau_m T_\mu\})^{-1}.\Pr\{T_\lambda<t \text{ and } T_\lambda<\tau_m T_\mu\}\\ &= (\rho\tau_m)^{-1}(1+\rho\tau_m)[\Pr\{T_\lambda<t\}\Pr\{\tau_m T_\mu>t\}\\ &\qquad + \Pr\{T_\lambda<\tau_m T_\mu \text{ and } \tau_m T_\mu<t\}]\end{aligned}$$

$$=(\rho\tau_m)^{-1}(1+\rho\tau_m)\left[(1-e^{-\lambda t})e^{-\mu t/\tau_m}+\int_0^{t/\tau_m}\mu e^{-\mu s}\int_0^{\tau_m s}\lambda e^{-\lambda z}\,dz\,ds\right]$$

$$=1-e^{-((\mu+\lambda\tau_m)/\tau_m)t}. \quad (59)$$

The simplest case occurs when $\tau_m=1$, for every $m\geqq 1$, i.e. when all classification times, for a given pattern x, are equal to the one, t_1, spent in the first classification $d_1(x)$. Then it holds that $\sigma_m=m$, $m\geqq 1$, and:

$$\Pr(D_m)=\rho\delta_{m1}+(1-\rho^2)(1+\rho)^{m-1}\chi_{m-1}\frac{[1+(m+1)\rho]-\alpha_m[1+m\rho]}{[1+m\rho][1+(m+1)\rho]}, \qquad m\geqq 1.$$

As for the distribution of the waste time T, under the condition $T>0$, we have that, in this case, it does not depend on the point at which the classification sequence is interrupted. We have indeed that:

$$F_m(t \mid T>0)=\Pr\{T<t \mid D_m, T>0\}=1-e^{-(\lambda+\mu)t};$$

as we obtained in the case of independent classification times.

7. PARTICULAR CASES

Let us go back to our main model with independent exponentially distributed service (classification) times.

Let us assume that:

$$\alpha_m = \alpha, \qquad m \geqq 1, \tag{60}$$

where α is a fixed number and $0 < \alpha < 1$.

Definitions (6), (7) and (8) give:

$$\chi_m = \left(\frac{\alpha}{1+\rho}\right)^m, \qquad m \geqq 1; \tag{61}$$

$$\chi^* = \frac{1+\rho}{1+\rho-\alpha} = 1 + \frac{\alpha}{1+\rho-\alpha}. \tag{62}$$

The equilibrium probabilities of the states of Σ with empty queue, i. e. the ones depending on the parameter α, are as follows:

$$p_{00} = \frac{(1-\rho)(1+\rho)(1-\alpha)}{1+\rho-\alpha}; \tag{63}$$

$$p_{0m} = \rho(1-\rho)\left(\frac{\alpha}{1+\rho}\right)^{m-1}, \qquad m \geqq 2. \tag{64}$$

Of course, for the remaining states, (13) still holds, giving:

$$p_{n1} = (1-\rho)\,\rho^{n+1}, \qquad n \geqq 0.$$

The average number of patterns in the system and the average time that a pattern spends in the system are respectively:

$$\overline{N} = \rho\left[\frac{1}{1+\rho} + \frac{1-\rho}{1+\rho-\alpha}\alpha\right], \tag{65}$$

$$\overline{T} = \frac{1}{\mu}\left[\frac{1}{1-\rho} + \frac{1-\rho}{1+\rho-\alpha}\alpha\right]; \tag{66}$$

which are increasing functions of α.

As for the waste time, we have that:

$$\Pr(\Sigma \text{ is doing ineffective work}) = \rho^2 \frac{(1-\rho)\alpha}{(1+\rho)(1+\rho-\alpha)} \tag{67}$$

The probability that a pattern in recognized wasting some time is:

$$\Pr(A) = \frac{\rho(1-\rho)\alpha}{1+\rho-\alpha}. \tag{68}$$

The size m of the classification sequence is described by:

$$\Pr(D_m)=\rho\delta_{m1}+(1-\rho)\left(1-\frac{\alpha}{1+\rho}\right)\left(\frac{\alpha}{1+\rho}\right)^{m-1}, \qquad m\geqq 1. \tag{69}$$

Then the expected size of the classification sequence is:

$$\bar{m}=\frac{1+\rho-\alpha\rho}{1+\rho-\alpha}. \tag{70}$$

Let us further assume that the error probability sequence of Σ is given by:

$$e_m=e_0\,\varepsilon^m, \qquad m\geqq 1, \tag{71}$$

where e_0 and ε are fixed numbers in $(0, 1)$.

We obtain for the error probability:

$$e=e_0\,\varepsilon\left[\rho+(1-\rho)\frac{1+\rho-\alpha}{1+\rho-\alpha\varepsilon}\right]. \tag{72}$$

The results of this section assume that infinite classification sequences are allowed, or that patterns have possibly an infinite number of features.

The case of classification sequences bounded by a maximum size M can be easily handled substituting infinite sums by finite ones.

8. FINITE QUEUE

Let Σ have finite room for patterns to wait and let $K\geqq 0$ be the maximum queue-size allowed. This upper bound for the queue-size may be established in order to have a relatively high probability to find a single pattern in the recognizer.

Let $K\geqq 1$; the equilibrium equations are in this case the (1), (2) and (3) of Section 2, whereas the equations (4) and (5) become now:

$$(1+\rho)\,p_{n1}=\rho\,p_{n-1,\,1}+p_{n+1,\,1}, \qquad 1\leqq n\leqq K-1; \tag{4'}$$

$$p_{K1}=\rho\,p_{K-1,1}; \tag{4''}$$

$$\sum_{m=0}^{\infty}p_{0m}+\sum_{n=1}^{K}p_{n1}=1. \tag{5'}$$

The solution is:

$$p_{n1}=\frac{1-\rho}{1-\rho^{K+2}}\,\rho^{n+1}, \qquad 0\leqq n\leqq K; \tag{73}$$

$$p_{00}=\frac{1-\rho}{1-\rho^{K+2}}(1+\rho-\rho\chi^*); \tag{74}$$

$$p_{0m}=\frac{1-\rho}{1-\rho^{K+2}}\rho\chi_{m-1}, \qquad m\geqq 2. \tag{75}$$

We can note that, from the viewpoint of queueing systems, p_{K1} is the fraction of customers that are lost, then λp_{K1} is the rate at which patterns are rejected by Σ with no classification by any stage of the recognizer. On the other hand, from the viewpoint of recognition, patterns which arrive and find Σ in state $(K, 1)$ are not lost. In fact, a constant classification d_0 may be assigned to every pattern which finds no room to wait and be recognized by Σ. Since Σ gives the classification d_0 and does not spend time for that, then (73), (74) and (75) still hold. If we indicate by d^* the exact classifier, then:

$$e_0=\Pr\{d_0\neq d^*\}, \tag{76}$$

is the error probability of the classification d_0.

Let us put:

$$\pi_i=\Pr\{d^*=i\}=\Pr(\mathscr{P}_i), \qquad i\in\{1, 2, \ldots, k\}, \tag{77}$$

then it holds that:

$$e_0=1-\Pr\{d^*=d_0\}=1-\pi_{d_0}, \tag{78}$$

and we will choose d_0 so as to minimize e_0.

The actual expression for the error probability is:

$$e=p_{K1}e_0+[1-p_{K1}-\Pr(L)]e_1+\Pr(L)\sum_{m=1}^{\infty}e_m(\chi_{m-1}-\chi_m), \tag{79}$$

where L indicates the event that a pattern starts being classified (for d_1) with empty queue, and its probability is given by:

$$\Pr(L)=\sum_{n=0}^{K-1}\frac{1}{(1+\rho)^{n+1}}p_{n1}=\frac{\rho(1-\rho)}{1-\rho^{K+2}}\left[1-\left(\frac{\rho}{1+\rho}\right)^K\right]. \tag{80}$$

We can consider "true" classifications only the d_m, $m\geqq 1$, because only these ones cost some time of the system Σ and we may be interested in the average cost of recognition, possibly a function of the expected size of the "true" classification sequence:

$$\overline{m}=1-p_{K1}-\Pr(L)+\Pr(L)\chi^*=1+(\chi^*-1)\Pr(L)-p_{K1}. \tag{81}$$

Of course, the expected size $\overline{m}$ of the "true" classification sequence may well be less than 1.

Finally, in the case that $K=0$, that is to say that no queue is allowed, we have that:

$$p_{0m}=\frac{\rho}{1+\rho}\chi_{m-1}; \tag{82}$$

$$p_{00}=\frac{1+\rho-\rho\chi^*}{1+\rho}. \tag{83}$$

The discussion of this case is trivial.

REFERENCES

1. K. S. FU, *Sequential Methods in Pattern Recognition and Machine Learning*, Academic Press, New York, 1968.
2. B. W. CONOLLY, *Lecture Notes on Queueing Systems*, Ellis Horwood Ltd., Chichester, 1975.
3. L. KLEINROCK, *Queueing Systems*, Vol. I: *Theory*, Wiley-Interscience, New York, 1975.