

Basi di Teoria delle Code

da <http://escher07.altervista.org>

Generalità

Questo documento introduce alcuni concetti di Teoria delle Code in modo da poter affrontare meglio l'argomento della modellizzazione di reti a commutazione di pacchetto. E' un testo solo introduttivo : si rimanda al paragrafo "Bibliografia e Links" per gli approfondimenti.

Catene di Markov

Un Processo Stocastico è una variabile aleatoria funzione del tempo ovvero $\{X_t\}$, con $t \in T$. Tipicamente si fa riferimento a processi stocastici ad evoluzione tempo discreta ovvero dove $T=\{0,1,2,\dots\}$, ovvero dove la v.a cambia valore ad ogni scatto del "clock" temporale.

Le Catene di Markov sono processi stocastici tempo discreti che hanno le seguenti proprietà :

1. ad ogni istante la probabilità di transizione da un qualunque stato i allo stato j è indipendente da tutti gli stati assunti precedentemente.
2. Le probabilità di transizione da uno stato all'altro non dipendono dal tempo in cui si osserva la transizione stessa

... dove per stato del sistema si intende ogni possibile valore che può assumere la variabile aleatoria $\{X_t\}$.

Una Catena di Markov è detta a stati finiti nel caso in cui ammetta un numero finito Q di stati. La catena in questione può essere descritta dalle seguenti variabili :

$N(t)$: numero degli utenti del sistema, cioè la funzione tempo continua a gradino ottenuta dalla $\{X_t\}$ ovvero valore dello stato del sistema all'istante t . Equivalentemente al posto della variabile aleatoria $N(t)$ si può conoscere $\underline{\Pi}(t)$, ovvero vettore delle probabilità che all'istante t ci siano $i=0\dots Q-1$ utenti nel sistema.

$\underline{P}(t_1,t_2)$: matrice delle transizioni di stato dall'istante t_1 all'istante t_2 ovvero delle $p_{ij} = \text{Prob}\{X_{t_2} = j | X_{t_1} = i\}$ (dove $P\{A|B\}$ è la probabilità di A condizionata al precedente verificarsi di B definita come $P\{A|B\} = P\{AB\} / P\{B\}$ con A e B eventi non indipendenti).

In particolare : una Catena di Markov è nota se è noto $\underline{\Pi}(t)$ e questo lo possiamo trovare tramite $\underline{P}(t_1,t_2)$ essendo:

$$\underline{\Pi}(t_2) = \underline{\Pi}(t_1) * \underline{P}(t_1,t_2)$$

Questa equazione ci permette di trovare come si evolve il vettore riga (non colonna come di solito, così mi risparmio un po' di simboli di trasposta...) delle probabilità da un istante ad un altro. Se aggiungiamo l'ipotesi di Catena di tempo omogenea (la probabilità di

transizione non dipende dagli istanti scelti ma solo dalla differenza fra questi) si può prendere $t_2=t$ e $t_1=0$.

$$\underline{\Pi}(t) = \underline{\Pi}(0) * \underline{P}(t)$$

Il sistema è risolto se è noto l'andamento di $\underline{P}(t)$. Questo si può trovare tramite la seguente equazione differenziale:

$$\frac{d\underline{P}(t)}{dt} = \underline{P}(t) * \underline{Q}$$

Dove:

$$\underline{Q} = \lim_{Dt \rightarrow 0} \frac{\underline{P}(Dt) - I}{Dt}$$

Si arriva a questa espressione attraverso la definizione di derivata di $\underline{P}(t)$ e di $\underline{P}(t+Dt)$ sviluppato utilizzando la proprietà di Chapman-Kolmogorov per le catene di Markov secondo cui in generale $\underline{P}(t_1, t_3) = \underline{P}(t_1, t_2) * \underline{P}(t_2, t_3)$ e tenendo conto della stazionarietà $\underline{P}(t+Dt) = \underline{P}(t) * \underline{P}(Dt)$. Ne deriva che:

$$\frac{d\underline{\Pi}(t)}{dt} = \underline{\Pi}^T(t) * \underline{Q}$$

... dunque nota la matrice \underline{Q} è nota la nostra catena. Dalla sua definizione e dalle proprietà della matrice identità abbiamo che:

$$q_{ii} = \lim_{Dt \rightarrow 0} \frac{p_{ii}(Dt) - 1}{Dt}$$

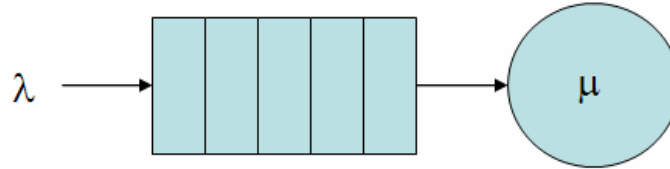
$$q_{ij} = \lim_{Dt \rightarrow 0} \frac{p_{ij}(Dt)}{Dt}$$

I valori dei q_{ij} si possono trovare con delle ipotesi sul processo che determina le transizioni di stato. Qui di seguito vedremo un esempio concreto. Notiamo prima di proseguire che se quello che interessa è il valore stazionario di $\underline{\Pi}$ l'equazione differenziale precedente si riduce al sistema lineare :

$$\underline{\Pi} * \underline{Q} = \underline{0}$$

Processi di Nascita e Morte

Consideriamo ora una catena di Markov siffatta:



In sintesi è un sistema dove arriva un certo numero di richieste nell'unità di tempo con distribuzione nota. Queste vengono prima posizionate in una coda fino a riempirla completamente, scartando le eccedenti e poi vengono via via servite con il numero di richieste prelevate dalla coda nell'unità di tempo che ha parimenti distribuzione nota.

Un sistema di questo tipo si chiama "processo di nascita/morte" perché, essendo sia il riempimento della coda che il suo svuotamento di un elemento alla volta, identificato lo stato del sistema con numero di elementi (richieste) presenti nel sistema (inteso come richieste in arrivo + richieste presenti in coda dal passo precedente) sono possibili solo le seguenti transizioni:

$$\begin{aligned} k \rightarrow k+1 & \quad \text{"Nascita"} \\ k \rightarrow k-1 & \quad \text{"Morte"} \end{aligned}$$

dove si hanno $M = Q + 2$ stati possibili con Q capacità della coda, identificati in:

$k=0$	nessun elemento nel sistema
$k=1$	un elemento nel sistema (ovvero coda con un posto prima del prelievo)
..	
$k=Q$	Q elementi nel sistema (ovvero coda piena prima del prelievo)
$k=Q+1$	$Q+1$ elementi nel sistema (ovvero coda piena e un elemento all'ingresso)

Nota : per l'ultimo stato tenere conto che gli elementi entrano uno alla volta, quindi anche se la sorgente ne produce ad esempio $Q+n$ con coda vuota si arriva allo stato $Q+1$ che poi viene mantenuto per tutti i successivi ingressi di tutti gli n eccedenti.

Il numero di arrivi ed evasioni nell'unità di tempo sono variabili aleatorie. Si ipotizza che queste variabili abbiano distribuzione di Poisson ovvero risultati, detto X il numero di arrivi e U il numero di uscite nell'unità di tempo (ovvero nel nostro "clock" ad esempio un secondo):

$$\begin{aligned} P\{X = k\} &= \lambda^k e^{-\lambda} / k! \\ P\{U = k\} &= \mu^k e^{-\mu} / k! \end{aligned}$$

In un sistema di questo tipo, valendo per le variabili di Poisson che:

$$p_{ij}(Dt) = \lambda_{ij}Dt + o(Dt)$$

con λ_{ij} che è la frequenza media del passaggio dallo stato j allo stato i (con $i \neq j$) si ha che:

$$q_{ij} = \lambda_{ij} \quad i \neq j \quad i, j = 0 \dots M$$

Si ha inoltre che:

$$p_{ii}(Dt) = 1 - \sum_{i \neq j} p_{ij}(Dt) = 1 - \sum_{i \neq j} \lambda_{ij} Dt + o(Dt)$$

Per cui:

$$q_{ii} = - \sum_{i \neq j} \lambda_{ij} \quad i = 0 \dots M$$

La matrice Q è quindi perfettamente nota. Esplicitamente possiamo ottenerla con la seguente osservazione supplementare:

- La nascita (ovvero un incremento dello stato) si ottiene con un arrivo, di tasso λ
- La morte (ovvero una diminuzione dello stato) si ottiene con una partenza, di tasso μ
- I tassi λ_{ij} con i e j diversi più di 1 sono nulli perché sono permesse solo transizioni di uno stato alla volta

Così ad esempio per una coda ad una posizione ovvero per un sistema a $M=3$ stati (0,1,2) avremo

$$q_{01} = (\text{passaggio dallo stato 0 allo stato 1}) = \text{arrivo} = \lambda$$

$$q_{02} = (\text{passaggio dallo stato 0 allo stato 2}) = \text{non possibile} = 0$$

$$q_{10} = (\text{passaggio dallo stato 1 allo stato 0}) = \text{partenza} = \mu$$

$$q_{12} = (\text{passaggio dallo stato 1 allo stato 2}) = \text{arrivo} = \lambda$$

$$q_{20} = (\text{passaggio dallo stato 2 allo stato 0}) = \text{non possibile} = 0$$

$$q_{21} = (\text{passaggio dallo stato 2 allo stato 1}) = \text{partenza} = \mu$$

Riguardo ai termini diagonali avremo che:

$$q_{00} = (i=0) = -\sum_{j \neq 0} \lambda_{ij} = -[\lambda_{01} + \lambda_{02}] = -[\lambda + 0] = -\lambda$$

$$q_{11} = (i=1) = -\sum_{j \neq 1} \lambda_{ij} = -[\lambda_{10} + \lambda_{12}] = -[\mu + \lambda] = -(\lambda + \mu)$$

$$q_{22} = (i=2) = -\sum_{j \neq 2} \lambda_{ij} = -[\lambda_{20} + \lambda_{21}] = -[0 + \mu] = -\mu$$

Si ottiene una matrice come questa

$$\underline{Q} = \begin{pmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda + \mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix}$$

Questa matrice è singolare (la colonna centrale è la somma delle altre due, cambiata di segno) quindi per poter risolvere il sistema lineare $\underline{\Pi}^T * \underline{Q} = \underline{0}$ ed avere una sola soluzione si aggiunge la seguente condizione (normalizzazione delle probabilità):

$$\pi_0 + \pi_1 + \pi_2 = 1$$

Ed il gioco è fatto. In particolare si ottiene un sistema come questo (dei tre prodotti riga-colonna non considero quello centrale, visto che darebbe una equazione linearmente dipendente con le altre e più incasinata):

$$\begin{aligned} -\lambda * \pi_0 + \mu * \pi_1 &= 0 \\ \lambda * \pi_1 - \mu * \pi_2 &= 0 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

Ovvero:

$$\begin{aligned} \lambda * \pi_0 &= \mu * \pi_1 \\ \lambda * \pi_1 &= \mu * \pi_2 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

Da cui:

$$\begin{aligned} \pi_1 &= (\lambda/\mu) * \pi_0 \\ \pi_2 &= (\lambda/\mu) * \pi_1 = (\lambda/\mu)^2 * \pi_0 \\ \pi_0 &= 1 - (\lambda/\mu) * \pi_0 - (\lambda/\mu)^2 * \pi_0 \end{aligned}$$

Ovvero posto $\rho_0 = \lambda/\mu$ si ha:

$$\begin{aligned} \pi_0 &= 1 / [1 + \rho_0 + \rho_0^2] = 1 - \rho_0 / [1 - \rho_0^3] \\ \pi_1 &= \rho_0 * \pi_0 \\ \pi_2 &= \rho_0^2 * \pi_0 \end{aligned}$$

In questo modo, noti λ e μ ovvero il fattore di carico conosciamo il vettore delle probabilità stazionarie del sistema. Mostriamo ora qui di seguito come ricavare da questo le altre grandezze di interesse del sistema.

Riguardo al tempo di attesa T , ovvero al tempo passato nel sistema, esiste un opportuno Teorema (di Little) che mette in relazione il suo valor medio atteso $E[T]$ con il valor medio del numero di utenti $E[N]$ ed il tasso di arrivo λ , essendo:

$$E[N] = \lambda * E[T]$$

Da qui, essendo gli stati identificati col numero di utenti nel sistema :

$$E[N] = \sum_{i=0}^M i * \pi_i$$

abbiamo nel nostro caso $E[N] = 0 * \pi_0 + 1 * \pi_1 + 2 * \pi_2 = \rho_0 * \pi_0 + 2 * \rho_0^2 * \pi_0$ cioè:

$$E[N] = \pi_0 [\rho_0 + 2\rho_0^2]$$

Essendo quindi il nostro ritardo medio :

$$E[T] = \pi_0 \frac{\rho_0 + 2\rho_0^2}{\lambda}$$

Un sistema nascita-morte quale quello descritto è indicato convenzionalmente come di tipo M/M/1/1 (Tipo Arrivi/Tipo Servizi/Numero Serventi/Capacità della Coda) con la cosiddetta notazione di Kendall dove M sta per Markov ed indica processo di tipo di Poisson.

I risultati ottenuti possono essere facilmente generalizzati al caso di coda a Q posizioni ovvero al sistema M/M/1/Q. In tal caso avremo :

$$\pi_0 = \frac{1 - \rho_0}{1 - \rho_0^{Q+1}} =: 1 - \rho$$

Avendo definito il fattore di utilizzo ρ come $1 - \pi_0$ denominazione che ha un senso perché π_0 è la probabilità che il sistema sia vuoto quindi che il server sia inattivo. Di conseguenza potremo scrivere per $i=0...Q+1$

$$\pi_i = (1 - \rho)\rho_0^i$$

Ovvero:

$$E[T] = \frac{1 - \rho}{\lambda} \sum_{i=0}^{Q+1} i\rho_0^i$$

Il tutto si semplifica decisamente per coda ad infinite posizioni ovvero per il sistema M/M/1/ ∞ . In tal caso, come si vede dalla definizione di π_0 il fattore di utilizzo dinamico coincide con quello statico ovvero risulta:

$$\rho = \rho_0 = \lambda / \mu$$

In più la sommatoria a Q+2 termini nella precedente espressione diventa – in ipotesi di $\rho_0 < 1$ che si dimostra essere sufficiente per l'esistenza della soluzione stazionaria – la seguente serie convergente:

$$\sum_{i=0}^{\infty} i\rho_0^i = \frac{\rho_0}{[1 - \rho_0]^2}$$

Per cui in conclusione:

$$\pi_0 = 1 - \rho$$

$$\pi_i = (1 - \rho)\rho^i$$

$$E[T] = \frac{1 - \rho}{\lambda} \frac{\rho}{[1 - \rho]^2} = \frac{1}{\mu - \lambda}$$

Notiamo che la precedente è l'espressione analitica del valor medio del tempo di attesa T, che ovviamente è una variabile aleatoria. Si può dimostrare che tale variabile ha densità di probabilità esponenziale con valor medio $\mu - \lambda$.

Un cenno all'analisi dei Processi di Nascita e Morte nel caso non stazionario

Nel caso non stazionario il sistema risulta ugualmente descrivibile, solo che si tratta di risolvere un sistema differenziale (magari con l'aiusilio della trasformata di Laplace che trasforma le derivate in prodotti per "s") invece che uno lineare.

Si può dunque determinare $\underline{\Pi}(t)$ e quindi vedere come si evolvono le probabilità degli stati col tempo e sapere di conseguenza il valore medio pesato del numero di utenti nel sistema all'istante fra l'istante 0 e l'istante t, che dovrebbe – in ipotesi di processo ergodico, ovvero t.c. la media temporale è uguale alla media di insieme – essere uguale alla media pesata di tutti gli stati con le relative probabilità all'istante t:

$$\bar{N}(t) = \frac{1}{t} \int_0^t N(\tau) d\tau = \sum_{k=0}^{Q+1} k \pi_k(t)$$

Siccome il Teorema di Little vale anche fra i valori medi temporali delle grandezze coinvolte (vedi dimostrazione prima dell'ultimo passaggio al limite) si può dire che:

$$\bar{T}(t) = \frac{\bar{N}(t)}{\lambda}$$

... è il valore medio del ritardo all'istante T in funzione di t. Facendo un po' di calcoli si trova che per le catene M/M/1/1 che abbiamo studiato il cambiamento rispetto al caso stazionario è che il sistema da risolvere diviene questo:

$$-\lambda * \pi_0 + \mu * \pi_1 = s \pi_0 - \pi_0(0)$$

$$\lambda * \pi_1 - \mu * \pi_2 = s \pi_1 - \pi_1(0)$$

$$\pi_0 + \pi_1 + \pi_2 = 1$$

Risolvendo tale sistema nel dominio di Laplace ed antitrasformando si ottiene il vettore $\underline{\Pi}(t)$ – in funzione del suo valore iniziale $\underline{\Pi}(0)$ – e da questo i valori medi di N(t) e T(t) di cui in precedenza.

Bibliografia e Links

<http://cs.unibg.it/martigno/documenti/reti/2SistemiACoda.pdf>

<http://www.reti.dist.unige.it/reti/Documenti/CODE.pdf>

<http://www.dsi.unive.it/~balsamo/vp/VP.Mod4.pdf>

<http://home.dei.polimi.it/capone/Infrastrutture/materiale/Teoria%20del%20traffico.pdf>