

Supplementi al progetto GuessIT

Lorenzo Monacelli, Mattia Miotto

30 settembre 2016

1 Calcolo dell'entropia

“Guess It” è un *game with a purpose*, un gioco in grado di stimare l'entropia del sistema in base alle scelte del giocatore. In questa sezione mostreremo il dettaglio di come avviene il calcolo dell'entropia. Il giocatore, quando gioca e punta il proprio budget, sta dando una stima personale delle probabilità di cosa si possa trovare sotto la casella coperta[1].

Sia b la frazione di budget che il giocatore punta sul contenuto vincente all' n -esimo passo:

$$b(x_n | x_{n-1} x_{n-2} \cdots x_0)$$

Questa probabilità è stimata dal giocatore dopo che sono state scoperte le $n-1$ caselle precedenti. Il budget dopo una scommessa sarà pari al triplo del budget puntato sulla casella vincente:

$$B_{n+1} = 3 \cdot B_n \cdot b(x_n | x_{n-1} \cdots x_0)$$

Quindi dopo N scommesse, il budget sarà:

$$B_N = 3^N B_0 \underbrace{b(x_N | x_{N-1} \cdots x_0) \cdots b(x_0)}_{b(x_N, x_{N-1}, \cdots, x_0)}$$

$$B_N = 3^N B_0 b(x_N, \cdots, x_0)$$

Il budget finale non dipende dal particolare ordine con il quale vengono piazzate le scommesse, per questo motivo la scelta dell'ordine è lasciata al giocatore. Alla fine del gioco il programma stima la quantità G come:

$$G = \frac{1}{N} \ln \frac{B_N}{B_0} = \ln 3 + \frac{1}{N} \ln b(x_N, \cdots, x_0)$$

Mediando su tanti tentativi si ottiene

$$\begin{aligned} \langle G \rangle &= \ln 3 + \frac{1}{N} \langle \ln b(x_N, \cdots, x_0) \rangle = \\ &= \ln 3 + \frac{1}{N} \left[\langle \ln p(x_N, \cdots, x_0) \rangle - \left\langle \ln \frac{p(x_N, \cdots, x_0)}{b(x_N, \cdots, x_0)} \right\rangle \right] \end{aligned}$$

Se p è la vera distribuzione di probabilità delle caselle vincenti allora si avrà:

$$\langle G \rangle = \ln 3 - S - D(p|b)$$

Dove S è l'entropia del sistema, mentre $D(p|b)$ è la divergenza di Kullback-Leibler tra le distribuzioni p e b . Poiché la divergenza è sempre un valore positivo vale la disuguaglianza:

$$\langle G \rangle < \ln 3 - S$$

Il guadagno medio massimo ottenibile dal giocatore permette allora di calcolare S :

$$S = \min_b [\ln 3 - \langle G(b) \rangle]$$

Abbiamo ricondotto il problema del calcolo dell'entropia ad un problema di minimizzazione sulla strategia del giocatore. La risoluzione computazionale di questo problema di minimo ha tempi proibitivi, poiché bisognerebbe sondare tutte le possibili strategie e calcolare quale di queste è in grado di minimizzare H . La mente umana invece è in grado di avvicinarsi molto alla soluzione utilizzando l'intuizione propria del cervello.

2 Utilità del calcolo dell'entropia

In meccanica statistica l'entropia è la funzione di stato che descrive interamente il sistema fisico in esame nell'ensemble microcanonico (in cui il sistema è isolato). In un sistema microcanonico occorre conoscere l'Hamiltoniana H , una funzione dello stato del sistema e di eventuali parametri esterni (come il volume), che descrive le interazioni tra individui. Dall'Hamiltoniana si può ricavare la densità degli stati nello spazio delle fasi:

$$\Omega(E, V, N) = \int d^3p d^3q \delta [E - H(\vec{p}_1, \dots, \vec{q}_1, \dots V)]$$

Dalla densità degli stati si calcola l'entropia:

$$S(E, V, N) = k_b \ln \left(\frac{\Omega(E, V, N) \Delta}{\Gamma_0} \right)$$

Nota S è possibile ricavare tutta la termodinamica del sistema [2, 3]:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E} \right)_{V, N} \quad - \frac{\mu}{T} = \left(\frac{\partial S}{\partial N} \right)_{E, V} \quad \frac{P}{T} = \left(\frac{\partial S}{\partial V} \right)_{N, E}$$

Nello studio dei sistemi come il modello preda predatore ci manca la conoscenza delle interazioni tra individui, ossia non conosciamo l'Hamiltoniana che descrive correttamente il sistema. Tuttavia, anche se conoscessimo H , il calcolo della densità degli stati per ricavare l'entropia sarebbe complesso e costoso dal punto di vista computazionale.

Lo scopo del programma è quello di calcolare S direttamente dalle configurazioni, aggirando il calcolo della densità degli stati. Purtroppo per ottenere la corretta definizione dell'energia del sistema occorre conoscere l'Hamiltoniana, infatti:

$$E(S, V, N) = \langle H \rangle$$

Per una corretta descrizione del sistema è quindi necessario un modello efficace dell'Hamiltoniana del sistema. Questo è possibile attraverso i recenti

algoritmi di massima entropia [4, 5, 6], che sono in grado trovare la migliore Hamiltoniana in un sottospazio opportunamente scelto dello spazio di tutte le Hamiltoniane. Attraverso questi algoritmi è possibile legare l'Hamiltoniana ai parametri del sistema (capacità di predazione, vita media delle prede, capacità riproduttiva, mobilità di prede e predatori nell'ambiente, fitness ecologica, etc.). Questi parametri prendono il posto del volume, mentre il numero di particelle corrisponde allo spazio disponibile nel sistema (il numero di caselle). È così possibile definire un sistema di equazioni di stato in grado di descrivere l'ecosistema come se fosse un gas di particelle, permettendo di predire come reagisce l'ecosistema a eventuali stimoli esterni, come l'introduzioni di nuove prede o predatori.

Riferimenti bibliografici

- [1] C. E. Shannon, "Prediction and entropy of printed english," *Bell system technical journal*, 1951.
- [2] R. Kubo, "Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems," *Journal of the Physical Society of Japan*, vol. 12, no. 6, pp. 570–586, 1957.
- [3] D. Ter Haar, "Foundations of statistical mechanics," *Reviews of Modern Physics*, vol. 27, no. 3, p. 289, 1955.
- [4] W. Bialek and R. Ranganathan, "Rediscovering the power of pairwise interactions," *arXiv preprint arXiv:0712.4397*, 2007.
- [5] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, "Identification of direct residue contacts in protein–protein interaction by message passing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 1, pp. 67–72, 2009.
- [6] T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan, "Maximum entropy models for antibody diversity," *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, pp. 5405–5410, mar 2010. [Online]. Available: <http://dx.doi.org/10.1073/pnas.1001705107>