Modelli di Computazione Affettiva - Analisi Teorica

Alessandro Costella

Università degli Studi di Milano

1 Introduzione

Il presente documento si propone di analizzare modelli e algoritmi utilizzati nell'esperimento descritto in "Modeling Development of Multimodal Emotion Perception Guided by Tactile Dominance and Perceptual Improvement" pubblicato da Horii e Nagai all'interno di *IEEE* nel Novembre 2018. Il paper riporta un esperimento atto a simulare, tramite l'utilizzo di un'architettura percettrone appositamente progettata, il processo infantile di apprendimento della facoltà di discriminare emozioni per esplorarne alcuni aspetti secondo gli autori trascurati. Nell'analisi che segue verrà introdotto il modello emotivo utilizzato dai ricercatori nella produzione del database utilizzato nell'apprendimento. Dopodiché sarà data particolare attenzione al tipo di macchina utilizzata nell'esperimento, una Restricted Boltzmann Machine opportunamente modificata per simulare il differente apporto della dominanza tattile e dello sviluppo percettuale durante la crescita di un infante nel primo anno di vita, e nello specifico agli aggiustamenti introdotti dagli autori per rappresentarne accuratamente determinati aspetti.

2 Il dataset

Obiettivo dell'esperimento analizzato è quello di confermare due ipotesi formulate dai ricercatori: l'importanza, nello sviluppo della capacità di differenziazione delle emozioni, del contributo della dominanza tattile e dello sviluppo percettuale, in particolare in un contesto di interazione tra una figura "tutoriale" adulta che "insegni" tale capacità e un infante nel primo anno di vita che la debba sviluppare. Per verificare queste ipotesi si è fatto uso di apprendimento semi-supervisionato e di un dataset creato ad hoc, acquisendo i dati tramite un sistema di tre sensori - una telecamera, un microfono e un sensore tattile morbido - e il contributo di un operatore che simulasse comportamenti tipici dell'interazione con l'infante. Ognuno dei tre stimoli è stato selezionato da un insieme prestabilito (e.g. "alta voce", "bassa voce", "basso tono" e altri per lo stimolo uditivo). La terna poi è stata, nella frazione del database dedicata alla parte supervisionata dell'apprendimento, associata a un'opportuna etichetta. Risulta immediatamente evidente dalle etichette (fig. 1) che sia stato fatto uso del modello emotivo discreto.

TABLE I
DESCRIPTION OF DATASET SIMULATING INFANT–CAREGIVER INTERACTION

Emotional state	Visual stimuli	Auditory stimuli	Tactile stimuli (emotional valence)	Number of data
Joy	Smiling face	Pitch rise in voice	Stroke (positive)	150
Surprise	Surprised face	Loud voice	Touch (zero emotion)	150
Anger	Angry face	Loud voice	Pinch (negative)	125
Disgust	Worried face	Low tone voice	Pinch (negative)	125
Sadness	Tearful face	Quite voice	Weak pat (zero emotion)	125
Fear	Frighten face	High frequency voice	Pat (negative)	125
Neutral	Neutral face	Neutral voice	Touch (zero emotion)	125

Fig 1. Tabella delle etichette.

2.1 Il modello emotivo

La teoria delle emozioni discrete - che coincide col modello più utilizzato allo stato dell'arte - è attribuita, nella sua forma attuale, allo psicologo statunitense Paul Ekman. Riprendendo ciò che Darwin per primo aveva sostenuto riguardo le emozioni, ossia che esse fossero retaggio di comportamenti ancestrali la cui utilità comunicativa avrebbe avvantaggiato l'individuo e la specie nel processo di selezione naturale, Ekman non solo sostiene che esse appartengano alla sfera della comunicazione non verbale, ma fu il primo a verificare l'esistenza di "vocabolario emotivo universale" composto di emozioni fondamentali associate a espressioni facciali svincolate dal contesto culturale, in accordo con la teoria evolutiva. Egli tentò di verificare tale ipotesi in primis esponendo immagini di differenti espressioni facciali a persone provenienti da diverse culture civilizzate e in secundis, nel '67, adattando l'esperimento e riproponendo tali espressioni facciali ad abitanti indigeni della Papua Nuova Guinea per eliminare il bias generato dall'esposizione agli stessi elementi letterari e cinematografici che i partecipanti al primo esperimento avrebbero potuto portare con sé.

L'esperimento portò Ekman a definire un insieme di sei emozioni fondamentali: gioia, rabbia, paura, disgusto, sorpresa e tristezza. Queste, insieme alla settima voce "Neutral", coincidono con le etichette utilizzate da Horii e Nagai nell'esperimento.

3 La Rete Neurale

Si procede ora a una panoramica sulle reti di Boltzmann e in particolare su quelle ristrette, prima di proseguire con l'analisi degli accorgimenti apportati alla rete utilizzata nell'esperimento.

3.1 Percettroni Multistrato

Una Boltzmann Machine è una Rete Neurale di tipo Percettrone Multistrato, forse la classe più semplice di reti neurali ma di comprovato valore pratico. Sono costituite da un insieme di neuroni strutturati, come il nome suggerisce, in strati che svolgono differenti mansioni. Di questi almeno due esistono necessariamente: uno strato di input e uno strato di output. Il flusso di dati è unidirezionale, dall'input verso l'output - motivo per cui tali reti sono anche dette "feedforward" - pertanto ogni dato viene fornito alla rete necessariamente allo strato di input, elaborato e fatto avanzare verso il successivo fino a quando la sua ultima elaborazione viene mostrata in output, che è pertanto una funzione dell'input. Tale funzione è la combinazione successiva delle funzioni applicate dai neuroni (o nodi) di tutti gli strati della rete: ognuno di questi nodi è associato a un vettore di n pesi, con n lunghezza del vettore di input, un offset che ha funzione di bias e una funzione di attivazione f, la cui scelta è parte della progettazione di ogni rete, da applicare alla somma pesata degli input. Si può pertanto esprimere l'output v di ogni nodo in una Rete Neurale feedforward come segue:

$$v = f(\sum_{i=1}^{n} w_i \cdot x_i + b) \tag{1}$$

Nell'utilizzo di una rete di questo tipo per la risoluzione di un problema di classificazione i nodi di output riportano in genere differenti probabilità, e sono associati ognuno a una differente classe di appartenenza. Il nodo che presenta in output la probabilità più alta è quello alla cui classe associata viene assegnato l'input appena elaborato.

3.2 Restricted Boltzmann Machines

Le Reti di Boltzmann Ristrette, o RBMs, sono un tipo di Rete Stocastica a Energia costituita di due strati, di cui lo strato di input è detto *visibile* e lo strato successivo è detto *nascosto*, che ricevono un valore binario in input e restituiscono un valore Bernoulliano in output. Ogni nodo dello strato visibile è collegato a tutti i nodi dello strato nascosto, qualificando la rete come rete feedforward di tipo *fully connected*. La particolarità che distingue una Rete di Boltzmann Ristretta da una normale Rete di Boltzmann è l'impossibilità di collegamenti tra differenti nodi dello strato nascosto: nessun neurone può essere collegato a nodi dello stesso strato, pertanto ogni collegamento neurale è necessariamente un collegamento da un nodo v_i dello strato visibile a un noto h_j dello strato nascosto.

Nelle reti a energia, seguendo il modello di reti neurali ricorsive proposto da Hopfield, viene associato allo stato in cui il sistema si trova un valore di energia la cui formulazione dipende dal tipo di rete. Nelle Reti di Boltzmann Ristrette tale energia è definita come:

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j v_i w_{i,j} h_j \tag{2}$$

con

- v_i unità dello strato visibile.
- h_j unità dello strato nascosto.
- a_i pesi di bias dello strato visibile.
- b_j pesi di bias dello strato nascosto.
- $w_{i,j}$ pesi associati al collegamento dal nodo visibile v_i al nodo nascosto w_j.

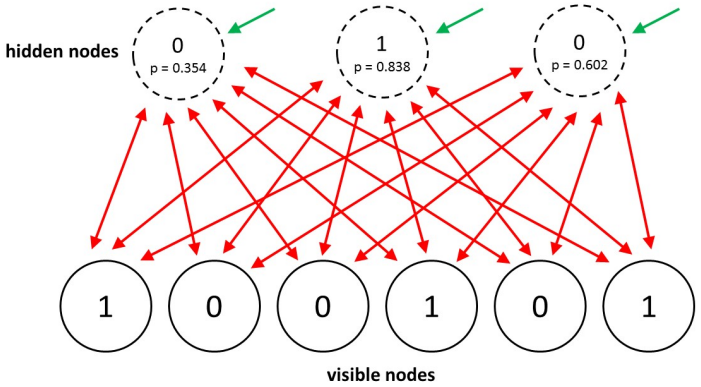

Fig 2. Esempio di RBM.

La RBM può essere dunque rappresentata come una rete di neuroni binari stocastici osservabili connessi a unità stocastiche nascoste non osservabili. La probabilità del sistema di trovarsi in un dato stato è funzione della sua energia:

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)} \tag{3}$$

con Z una funzione di partizione ottenuta sommando tutti i possibili $e^{-E(v,h)}$ per normalizzare le probabilità in modo da mantenerne unitaria la somma. Si noti inoltre che l'assenza di connessioni intra-strato, causa d'esistenza delle Reti Ristrette, garantisce l'indipendenza tra i nodi dello stesso strato. Questo permette alcune affermazioni sulle probabilità degli stati dei nodi. Infatti, con m nodi nello strato visibile e n nodi nello strato nascosto, si ha che le probabilità condizionate di un nodo visibile dato uno stato nascosto e un nodo nascosto dato uno stato visibile sono rispettivamente:

$$P(v|h) = \prod_{i=1}^{m} P(v_i|h) \tag{4}$$

$$P(h|v) = \prod_{j=1}^{n} P(h_j|v) \tag{5}$$

che ci permette infine di esprimere le probabilità di attivazione dei singoli nodi della rete:

$$p(h_j = 1|v) = \sigma(b_j + \sum_{i=1}^{m} w_{ij} v_i) \tag{6}$$

per i nodi dello strato nascosto e

$$p(v_i = 1|h) = \sigma(a_i + \sum_{j=1}^{n} w_{ij} h_j) \tag{7}$$

per i nodi dello strato visibile. σ rappresenta qui la funzione di attivazione, che nel caso dell'esperimento in analisi coincide con una sigmoide logistica. Tale funzione di attivazione è considerata lo standard nelle Reti di Boltzmann.

Si ricorda che i valori di uscita dei nodi sono Bernoulliani, pertanto $h_j = 1$ e $v_i = 1$ coincidono con l'unico stato di attivazione del nodo. L'aggiornamento della rete è in genere operato tramite metodi di discesa del gradiente tradizionali.

4 Struttura dell'esperimento

Passiamo ora a osservare la macrostruttura proposta da Horii e Nagai nel loro esperimento, oltre alle modifiche apportate dagli scienziati al classico modello RBM per ottenere la simulazione del processo di apprendimento di un neonato. Si tratta nello specifico di tre accorgimenti importanti, il primo atto a rappresentare la continuità dei segnali e i successivi a rappresentare due precise caratteristiche dell'apprendimento infantile e a verificare ognuna delle due ipotesi avanzate dai ricercatori.

4.1 Modello cognitivo

Il modello di Horii e Nagai suddivide l'apprendimento da parte dell'infante in due sottoprocessi: un processo sensoriale di acquisizione dei segnali e uno di riconoscimento emotivo. Inoltre viene imposta l'ipotesi di isolamento sensibile, si assume cioè che gli unici segnali ricevuti dal neonato siano quelli funzionali all'apprendimento di nostro interesse.

Questa struttura richiama fortemente il modello Kantiano di acquisizione della conoscenza: con queste due fasi si identificano i primi due dei tre passi distinti nel processo conoscitivo descritti nella *Critica della Ragion Pura*, la *sensibilità*, che ha facoltà di discernere i fenomeni che abbiano valore conoscitivo tramite i nostri sensi, e l'*intelletto*, che utilizza ciò che la sensibilità fornisce per formulare dei giudizi. Tale struttura, nella Critica, è influenzata dal lavoro di Cartesio, che per primo, ne *Le passioni dell'anima* (dove per passione si intende tanto la sensazione quanto l'emozione, l'"essere paziente", anima in ricezione passiva), aveva dato una prima ipotetica classificazione delle emozioni fondamentali, riferendosi ad esse come "passioni riferite solo all'anima" e "passioni in senso stretto".

4.2 Architettura

Per simulare tramite software il modello cognitivo proposto i ricercatori hanno progettato un'architettura che simula ognuno dei tre canali sensoriali interessati (visivo, uditivo e tattile) con un'apposita RBM. La scelta dell'RBM è motivata da un'ampia letteratura a sostegno della bontà dello strumento nel simulare l'apprendimento umano, anch'esso basato sulla minizzazione dell'energia libera Questo primo livello è quello che riceve gli effettivi dati sensoriali e coincide con il primo dei due passi sopra descritti. I tre canali sono intesi come perfettamente distinti tra loro e vengono unificati dopo essere stati processati dalle rispettive Reti Neurali. Il secondo strato, quello emotivo, aggrega i segnali integrandoli con una successiva RBM nella sezione dell'architettura che prende il nome di "strato emotivo".

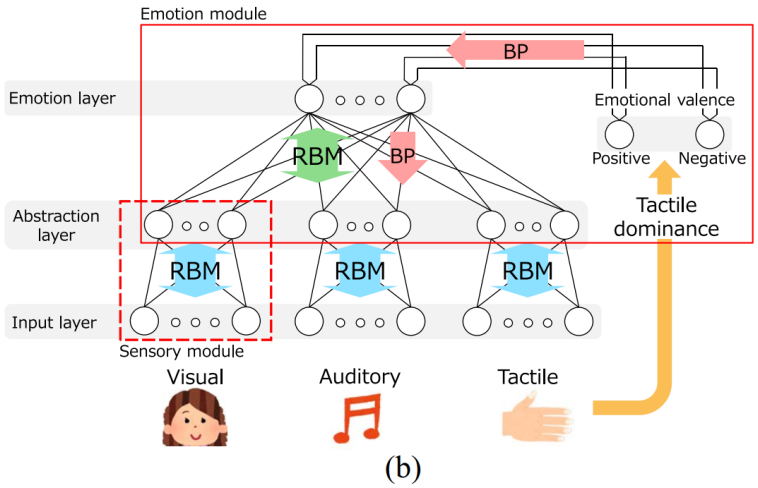

Fig 3. L'architettura proposta da Horii e Nagai.

4.3 Prima specifica: segnali continui

Il primo accorgimento è quello adottato per rappresentare la continuità dei segnali sensoriali in input. A questo scopo i segnali di input, normalmente binari, vengono sostituiti da distribuzioni Gaussiane $\mathcal{N}(\cdot|\mu, \sigma^2)$ a media μ e varianza σ^2. Questo altera la probabilità di attivazione tanto delle unità visibili quanto di quelle nascoste:

$$p(v_i = v|\mathbf{h}) = \mathcal{N}(v|a_i + \sum_{j=1}^{n} w_{ij}h_j, \sigma_i^2) \tag{8}$$

$$p(1|\mathbf{v}) = g(b_j + \sum_{i=1}^{m} \frac{1}{\sigma_i^2} v_i w_{ij}) \tag{9}$$

Anche l'aggiornamento deve essere adattato, in particolare nei gradienti dei pesi di interconnessione e dei bias a_i è necessario pesare il calcolo dei valori attesi con le varianze $\frac{1}{\sigma_i^2}$. Si rende dunque necessario anche aggiornare σ_i, poiché modula le probabilità $p(\mathbf{v})$ e $p(\mathbf{v}|\mathbf{h}, \theta)$ che concorrono al calcolo della perdita, di tipo cross-entropy. La variazione standard è stata aggiornata dagli autori tramite una funzione surrogata $z_i = log\sigma_i^2$ sempre tramite discesa del gradiente, con l'aspettativa che essa tenda a convergere a 0.

4.4 Seconda speifica: dominanza tattile

La dominanza tattile, di estrema importanza per un neonato, è causata dalla presenza delle cosiddette fibre nervose δA e delle fibre nervose C i cui assoni sono rispettivamente ricoperti e non ricoperti da mielina. Lle fibre nervose C, ad esempio, sono detti "afferenti tattili C" e contribuioscono, rispondendo a contatti leggeri sulla pelle, a esperire emozioni di valenza positiva, una funzione importante nell'interazione sociale.

Per simulare la presenza del fenomeno di dominanza tattile, il cui contributo è una delle due ipotesi fondamentali dello studio, i ricercatori hanno inserito, affianco alla RBM nello strato emotivo, uno strato di valenza emotiva composto di due neuroni che rappresentano i valori *positivo* e *negativo* del segnale tattile. L'associazione ai diversi segnali del valore in questione è osservabile nella tabella riportata in Fig.1, secondo modelli provenienti da studi precedenti. Si noti che il segnale "neutro" è ottenuto mantenendo inattivi entrambi i neuroni di valenza. Di fatto, la RBM multimodale

(quella che combina i tre stimoli nello strato emotivo) apprende in maniera supervisionata ad associare i suoi input a un'etichetta di valenza. L'architettura permette di scegliere se fare uso o meno di tale feature per poter simulare l'assenza di dominanza tattile.

4.5 Terza Specifica: sviluppo percettuale

Lo sviluppo delle facoltà percettuali nel primo anno di infanzia è simulato in conseguenza all'aggiornamento della varianza delle distribuzioni Gaussiane discusse al paragrafo IV.3: con la convergenza della stessa a valori via via minori le probabilità di attivazione dei neuroni dello strato nascosto delle macchine "sensore" (le tre RBM che agiscono da canali sensoriali) divengono meno incerte, e coprono sempre meno segnali di input, rendendosi via via più precise, coincidendo con accuratezza crescente con l'emozione rappresentata dagli stimoli di input. Anche questa funzione, nell'architettura, può essere sospesa in modo da verificare il contributo dello sviluppo percettuale. In questo caso-studio gli autori hanno impostato la varianza di tutti i neuroni dello strato di input a 0.01 mantenendola poi costante.

5 Osservazioni

La reiterazione dell'esperimento in quattro modalità - ottenute dalle diverse combinazioni di assenza e presenza delle feature descritte in IV.4 e IV.5 - sembra dimostrare che i fenomeni di dominanza tattile e di sviluppo percettuale siano fortemente influenti sul processo di apprendimento di distinzione delle emozioni solo quando compresenti, caso in cui tale influenza è molto significativa. Le soluzioni architetturali adottate dagli autori appaiono concettualmente semplici - soprattutto la simulazione dello sviluppo percettuale, "naturale conseguenza" della trasformazione degli ingressi in segnali continui - eppure si dimostrano di grande efficacia. Sono inoltre chiaramente distinte e isolabili, in modo da permetterne l'inclusione o esclusione dall'architettura durante l'addestramento per osservare situazioni patologiche - come in casi di assenza di nervi tattili osservata in alcuni neonati. Il modello teorico adottato è il più comune allo stato dell'arte e gode della solidità data da una lunga storia di contributi autorevoli. Gli strumenti utilizzati, in particolare le Restricted Boltzmann Machines, sono tra i più moderni e collaudati per simulare i processi di apprendimento umano.

References

1. T. Horii e Y. Nagai, "Modeling Development of Multimodal Emotion Perception Guided by Tactile Dominance and Perceptual Improvement"
2. Fiore et al., "Network anomaly detection with the restricted Boltzmann machine"
3. I. Kant, "Critica della Ragion Pura"
4. R. Decartes, "Le passioni dell'anima"