

LA VISIONE DELLA MACCHINA

Presentazione del lavoro

- Il sistema di visione è human-like, per realizzarlo mi sono basato soprattutto sugli studi di neurologia e di psicofisica. Negli anni Sessanta fu scoperto da Hubel e Wiesel che la corteccia visiva del gatto era costituita principalmente da celle che rispondevano agli angoli. Quanto vale per il gatto vale anche per gli animali superiori e per l'uomo. Asportare la corteccia visiva ad un animale superiore significa accecarlo, mentre un animale inferiore continua a poter agire nel suo ambiente: ciò perché, scendendo nella scala evolutiva, la visione mesencefalica ha peso via via maggiore. Gli animali superiori e l'uomo riconoscono comunque gli oggetti per mezzo della corteccia visiva e quindi codificano l'immagine principalmente in base agli angoli.
- Ho sviluppato questa idea (di considerare gli angoli come features) e la macchina riconosce, nel senso che attribuisce lo stesso nome a figure che una persona umana riconosce somiglianti, per esempio due R di foggia tipografica diversa, oppure individua le parti somiglianti e le parti non somiglianti fra figure. Quindi realizza il pattern recognition non solo topologico ma anche sintattico. Conseguentemente può formulare delle frasi del tipo: “la lettera P è come la lettera R senza una gambetta obliqua, a destra in basso”. In tal modo le due lettere, o in generale le forme, non potranno essere confuse.
- La macchina impara e memorizza la forma in un modo compresso, usualmente poche centinaia di bytes, perché essa ne coglie le caratteristiche essenziali ed elimina i particolari. Ad esempio, le lettere E dell'alfabeto, in varie foggie tipografiche, purché appaiano somiglianti ad un osservatore, avranno la stessa codifica (pattern). Invece la stessa lettera alfabetica in corsivo o quella in gotico avranno patterns diversi perché un osservatore umano non le percepisce come somiglianti. Tuttavia la macchina, oltre che generalizzare, può anche riconoscere la foggia tipografica di una lettera: dipende dall'insegnamento che le viene impartito.
- Circa l'estrazione del contorno, ho usato un metodo personale che permette di giungere velocemente agli angoli, quindi non è uno dei soliti gradienti (utilizzanti maschere di sobel, roberts,..) o un laplaciano. L'algoritmo funziona ma l'estrazione dei contorni “perfetti” nel mio lavoro non ha interesse in quanto essi si formano, si completano, nella fase di apprendimento. La macchina unisce i tratti di contorno derivanti dalla figura che osserva con quelli che ha in memoria e, nel caso, prolunga questi ultimi perché una figura può essere riconosciuta anche partendo da parti di contorno, così come viene riconosciuto un oggetto in parte coperto o viene riconosciuto un carattere in un manoscritto anche quando esso è unito al carattere successivo. L'estrazione dei contorni, su cui sono corsi fiumi di inchiostro, non mi pare un problema essenziale.
- Per i miei studi, uso un normale PC con una scheda di acquisizione immagini, collegata a una telecamera. Eseguo il trattamento delle immagini con una serie di programmi in turbo-pascal, che girano su un normalissimo personal computer “Pentium”. Tanto mi basta. Conosco le reti neurali, gli algoritmi e le logiche più strane ma, forse proprio perché non ho mai avuto grandi difficoltà in matematica, ritengo che essa sia spesso una mera tautologia

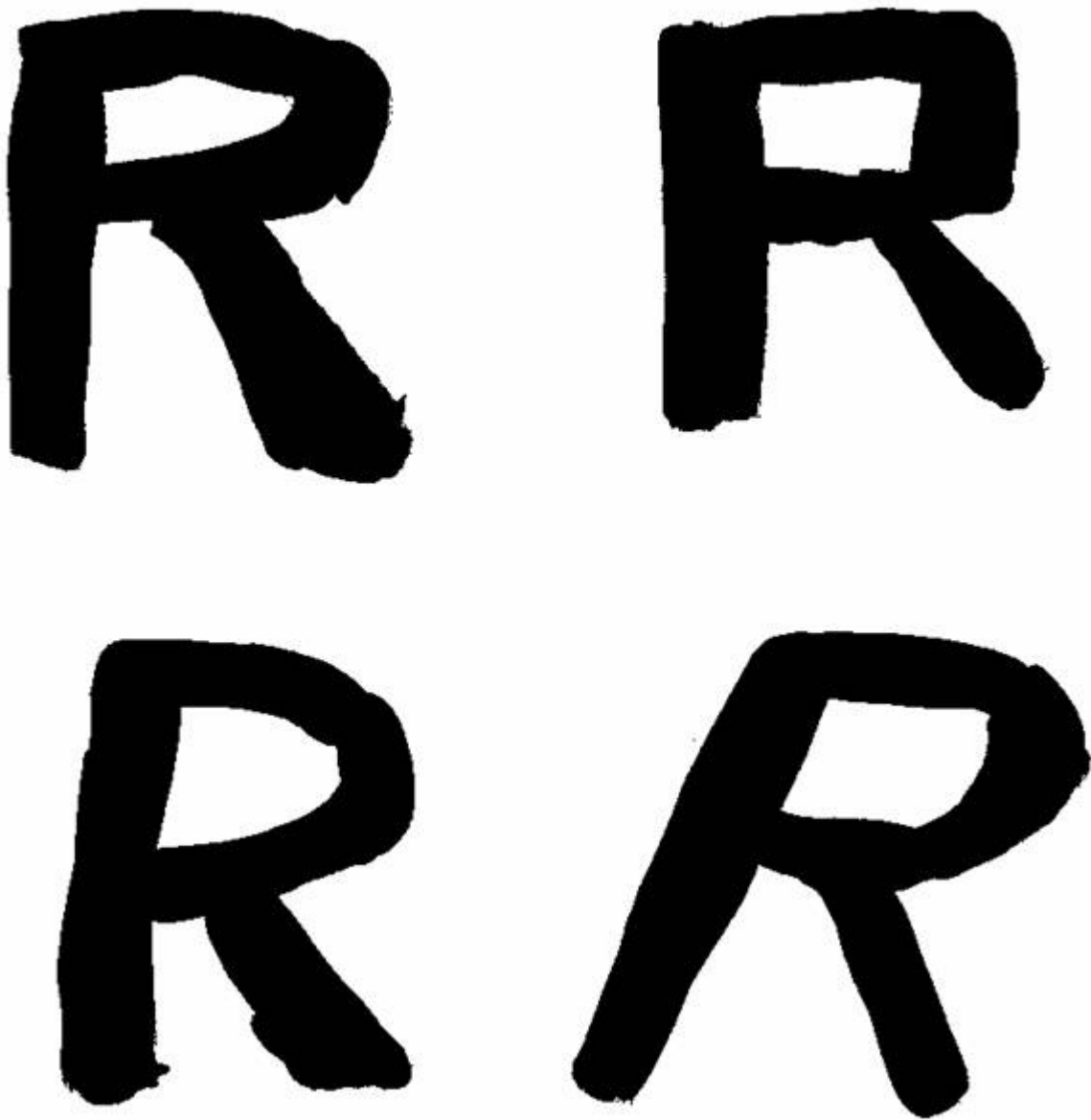


Figura 29

Per brevità di essa viene studiata solo un'area quadrata di 70x70 pixels, che può venire scelta manualmente. La sposto in modo da isolare la R in basso a sinistra di fig.29. Ho lasciato apposta dei disturbi, come la gamba della R di fianco, per complicare la vita al sistema durante la memorizzazione. La luminosità dei pixels è molto varia e dipende dal modo in cui la carta, non distesa bene, ha riflesso la luce

.....

OMISSIS

.....

Il risultato è in fig. 30b.



fig. 30 a



fig.30 b

.....

OMOSSIS

.....

la meccanizzazione del movimento e della visione periferica, che permette di portare in scala le figure facendole occupare tutto il campo visivo, rendono le figure tutte della stessa dimensione, che io realizzo in un quadrato di 70x70 pixels.

Estrazione delle caratteristiche

I contorni della figura (o il solo contorno principale, o tratti di contorno) vengono trasmessi ad un programma che ricava gli angoli su esso.....

.....

OMISSIS

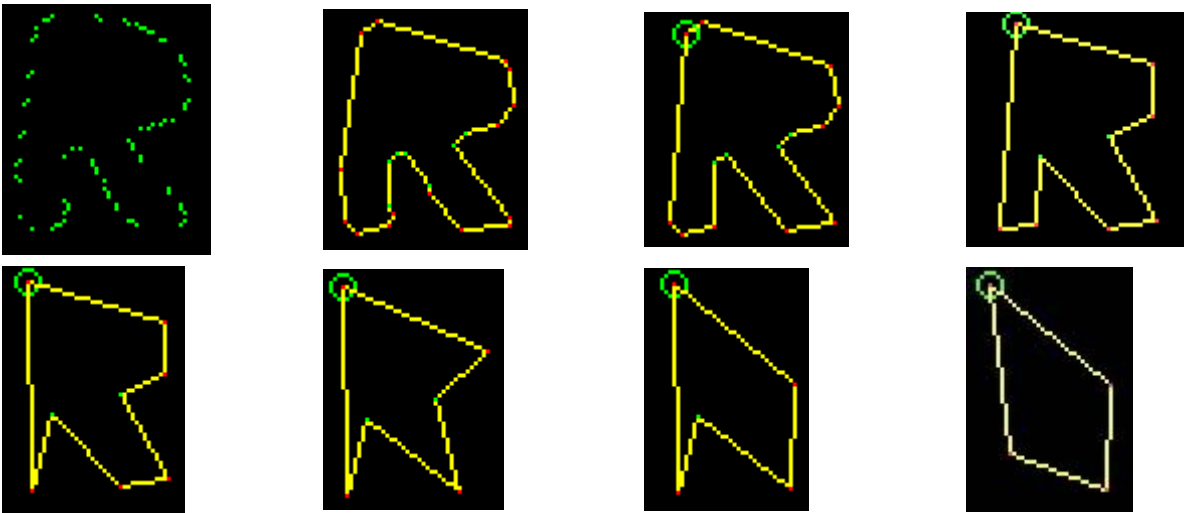
.....

Nelle figg. 34 vi sono alcuni risultati dell'applicazione dell'algorithmo attuale alla forma in fig. 30b.

.....

OMISSIS

.....



Figg. 34

Le figg. 34 mostrano che le contrazioni tolgono i particolari meno importanti alle forme. E' piuttosto interessante la sesta delle figg. 34, le R, tendono a ridursi tutte a quella forma. Solo R disegnate, volutamente, in modo molto sproporzionato non diventano così. Si può dire che le contratte alte si adattano a poche forme ben precise mentre le

contratte basse a molte forme. La finale di ogni contrazione di una linea chiusa è inevitabilmente un triangolo. Una buona codificazione, attraverso memorizzazione delle contratte alte, deve permettere di ottenere un riconoscimento preciso fin nei minimi particolari delle forme, con il rischio che due figure che un uomo vede somigliantesi non lo siano per la macchina che le considererà differenti. Viceversa, la memorizzazione delle contratte basse, fa riconoscere come uguali forme parecchio differenti e fa cadere il computer nella confusione. Supponendo che tutte le R generino la sesta delle figg. 34 occorre che la sua codifica venga memorizzata dal computer in modo automatico, non deve essere l'operatore a dire al computer: "bada la codifica da memorizzare è quella, perché è la *forma essenziale* della R".

.....

OMISSIS

.....

Memorizzazione e riconoscimento

Lo scopo che ora mi propongo è mostrare come il computer, in modo automatico, riesca a memorizzare l'essenziale delle forme, le affini, ne tolga i particolari meno importanti e possa poi riconoscerle. Il computer riesce a meccanizzare la somiglianza, cioè a dare lo stesso nome a forme a cui anche una persona darebbe lo stesso nome. Il computer riesce a vedere quali sono le parti simili fra due figure e quali sono le parti differenti. Rende possibile il riconoscimento di un oggetto in parte nascosto e la formazione e la chiusura del contorno in memoria a partire da pezzi di contorno percepiti.

Il tutto ha i limiti della visione topologica, tuttavia la visione sintattica altro non è che la moltiplicazione delle singole percezioni organizzate in una frase.

Uno sconfinamento nella la visone sintattica si ha quando introduco un nuovo algoritmo che è il riconoscimento per differenza. Anche se è cosa da poco riesce a discriminare in modo efficacissimo forme che sarebbero altrimenti confuse e fa intravedere la possibilità di formazione della frase.

Il metodo di memorizzazione e della ricerca dei ricordi è Human-like: ripescare i ricordi senza ripercorrerli tutti e anche come far formulare al computer le frasi, acquisendo qualunque struttura grammaticale. Questa tuttavia è cosa che si può posporre.

Il riconoscimento delle parti simili nei contorni delle figure.

Il programma di riconoscimento parte dalle matrici ottenute dalla contrazione. Opera così. Supponiamo che il computer abbia nulla in memoria e che la forma, la prima che gli si presenti, sia la R in alto a sinistra di fig.28. Ad essa cui diamo nome sia “ERRE”. Il computer, attraverso una serie di programmi, elabora la forma R fino ad ottenerne le contratte. Il computer non ha niente in memoria: esso memorizza la codificazione numerica delle contratte, senza alcuna modifica e il nome della forma.

.....
OMISSIS

.....
Fatto questo mostriamo al computer la seconda lettera R in alto della fig.38. La lettera R della figura 38 è la forma attuale, visibile. Essa deriva dalla seconda lettera R in alto di fig.28. Gli asterischi sul contorno distano 50 pixels (di contorno) fra loro. Quello rosso è il punto di partenza. Si va in senso orario. L’informazione delle contratte di questa forma è condensata nelle linee verticali verdi, nella parte bassa della figura, che rappresentano.....

.....
OMISSIS
.....

Gli asterischi sulla linea di puntini orizzontale gialla al centro della fig.38 permettono di capire a quali angoli della figura attuale corrispondono le barre. Il punto in alto a destra della lettera R, che è marcato dall’asterisco rosso, corrisponde al primo asterisco a sinistra della linea gialla. Siccome la figura è chiusa esso corrisponde anche all’ultimo asterisco. L’angolo in alto a destra è alla fine del perimetro, ovvero un po’ prima dell’inizio del perimetro (che è chiuso) e quindi sta alla fine della linea gialla: esso è rappresentato dalla lunga linea verde verticale a destra.

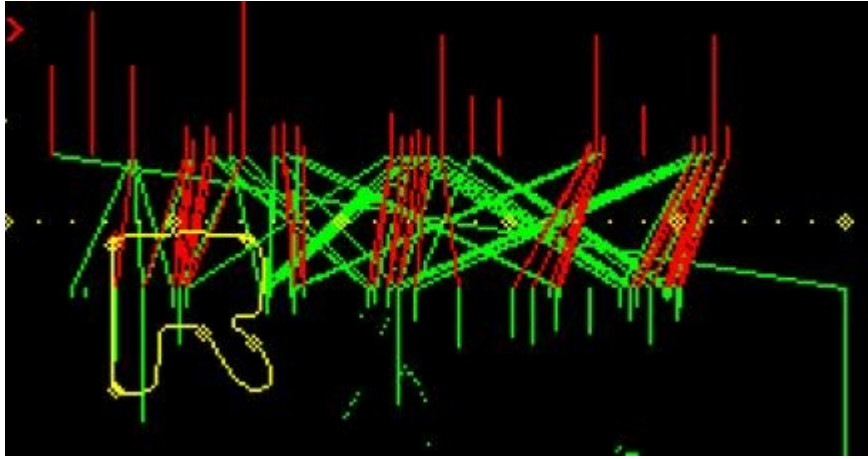


Figura 38

Fin qui per la figura attuale, ovvero la R in alto a sinistra di fig.28.

La lettera R in alto a destra di fig. 28 è stata trattata in precedenza e ora è in memoria. Essa non può più essere visualizzata. E' come per noi: noi riconosciamo le persone ma, eccezioni a parte, mi riferisco a persone dotate, non le sappiamo disegnare. Comunque anche i pittori vogliono il modello davanti a se, difficilmente lo disegnano a memoria. Una cosa è il riconoscimento, un'altra è il disegno. I segmenti rossi angolati sono i collegamenti "giusti", secondo il programma.

Voglio rimarcare che gli asterischi sulla linea orizzontale gialla in fig. 38 sono relativi alla figura attuale. Per comprendere quali siano gli angoli collegati fra le due figure, attuale e memorizzata si può ragionare così: faccio osservare al computer la R a destra di fig. 28, che diventa figura attuale, oltre che memorizzata.

In fig. 39 ho riportato il risultato di questa operazione, con gli asterischi, in modo che il lettori si capaci circa gli

angoli che sono collegati in fig. 38.

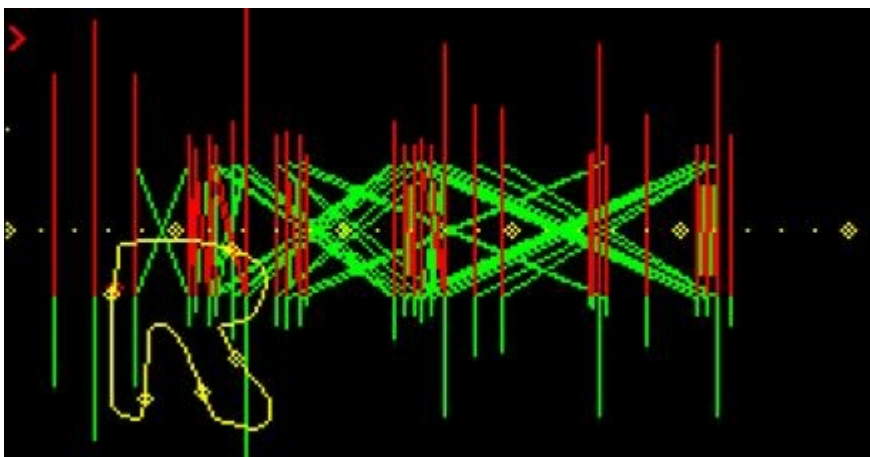


Figura 39

Anche se la fig. 39 ha lo scopo di mostrare gli asterischi sul contorno della prima R di fig. 28 una

sua analisi può essere utile per capire meglio come funziona il programma di riconoscimento. Poiché in memoria vi è la R in alto a sinistra di fig. 28, le linee verticali rosse sono uguali a quelle di fig. 38. Ovviamente alle linee verdi verticali in basso di fig. 39 sono uguali alle linee verticali rosse in alto. Le linee rosse altro non sono che i fasci “giusti” che collegano gli angoli che si corrispondono lungo il contorno, che il programma sceglie. Si vede che anche in questo caso il programma funziona: i fasci rossi verticali indicano che la macchina fa corrispondere le parti uguali della figura.

La meccanizzazione della somiglianza.

Nel paragrafo precedente, in fig. 39, abbiamo visto il confronto fra quanto in memoria, le matrici della R in alto a sinistra di fig. 28, e la figura attuale, la R in alto a destra di fig. 28. Possiamo osservare che vi sono degli angoli che si corrispondono e altri no. Infatti i collegamenti fra gli angoli della figura memorizzata, rappresentati dalle linee verticali rosse e quelli della figura attuale, rappresentati dalle linee verticali verdi, non sono tutti collegati dalle linee oblique rosse.

Ciò permette di addivenire ad una nuova memorizzazione che affina le caratteristiche della forma, basta seguire lo schema classico delle reti neurali, pertanto

.....
OMISSIS
.....

Questi valori vengono memorizzati e sostituiscono i precedenti. In tal modo si memorizza una forma che contiene le caratteristiche di due R. Continuando a mostrare altre R la memorizzazione si affina ed acquista le caratteristiche comuni alle R.

In base alla funzione con cui sommo le potenze degli angoli “uguali”, alla definizione della funzione con cui si calcola la potenza e alla funzione che mi determina la cancellazione di questi angoli io ho diversi tipi di “intelligenza” della macchina. Sotto larghissime condizioni l’applicazione di queste regole porta alla eliminazione di quasi tutti gli elementi delle contratte poco profonde, che hanno potenza grande ma non si ripetono, perché diversi da figura a figura. Questo non significa che venga cancellata tutta la contratta poco profonda, possono esservi parti della forma che restano costanti in ogni minimo particolare nell’insieme delle figure somigliantesi, quindi la codifica di quelle parti sarà preservata nelle contratte poco profonde, per il resto vuote. Circa il tipo di intelligenza, se scelgo una funzione additiva delle potenze che privilegi le contratte alte io potrò riconoscere memorizzare particolari della figura che altrimenti sarebbero dimenticati, perché eliminati dalla funzione di decremento di memoria. Solo figure estremamente somigliantesi, saranno riconosciute come uguali. Mentre se io scelgo una somma che privilegi le contratte basse non sarà possibile discriminare (con una sola percezione) fra figure molto diverse.

Il primo sistema, con il metodo degli omonimi, spiegato sotto, permette di attribuire lo stesso nome a figure che memorizza come diverse. Permette un riconoscimento rapido ma richiede più spazio in memoria del secondo. A mio avviso esso è usato più dagli animali che dall’uomo. Il secondo metodo richiede, per un buon riconoscimento della figura, una applicazione multipla e quindi la logica e il linguaggio.

Io presuppongo che alla macchina sia dato un insegnamento decente e non si faccia di tutto per

confonderla, tuttavia supponiamo che un buontempone mostri alla macchina una F dicendo ad essa che la forma ha nome ERRE. La macchina richiama la R memorizzata e, fra il codice delle R, che ha memorizzato, e la F troverà molta differenza: gli angoli che si corrisponderanno saranno pochi. In altre parole, fra la potenza totale della figura attuale la F e quella della figura in memoria la R solo una minima potenza risulterà essere coinvolta nella corrispondenza. Quando questa percentuale di corrispondenza è sotto una soglia la macchina avverte per dire che ha rilevato un omonimo o un errore dell'insegnante. Omonime risultano anche per la macchina due lettere R maiuscole di foggia tipografica diversissima. Inoltre omonime risultano due R molto simili se la funzione somma privilegia le contratte poco profonde. Se nonostante l'avvertimento si insistesse con il dire che non si tratta di un omonimo o di un errore dell'insegnante allora la macchina sommerebbe le caratteristiche della F su quelle della R. Un tale insegnamento ripetuto farebbe sì che solo le contratte basse si salverebbero e le due forme non sarebbero distinguibili in futuro.

.....
 OMISSIS

La rilevazione delle differenze.

Consideriamo l'insieme delle lettere A della fig.40 e l'insieme delle lettere R della fig.28. Qual è la differenza fra essi? dove si colloca? Non si proceda oltre, dopo aver generalizzato, nella sua mente, la forma delle A e quella delle R, **si disegni queste due lettere e si segni con la matita la differenza fra le A e le R, sulla lettera R.**

Le righe che seguono dovrebbero dimostrare che il programma dà la stessa risposta.

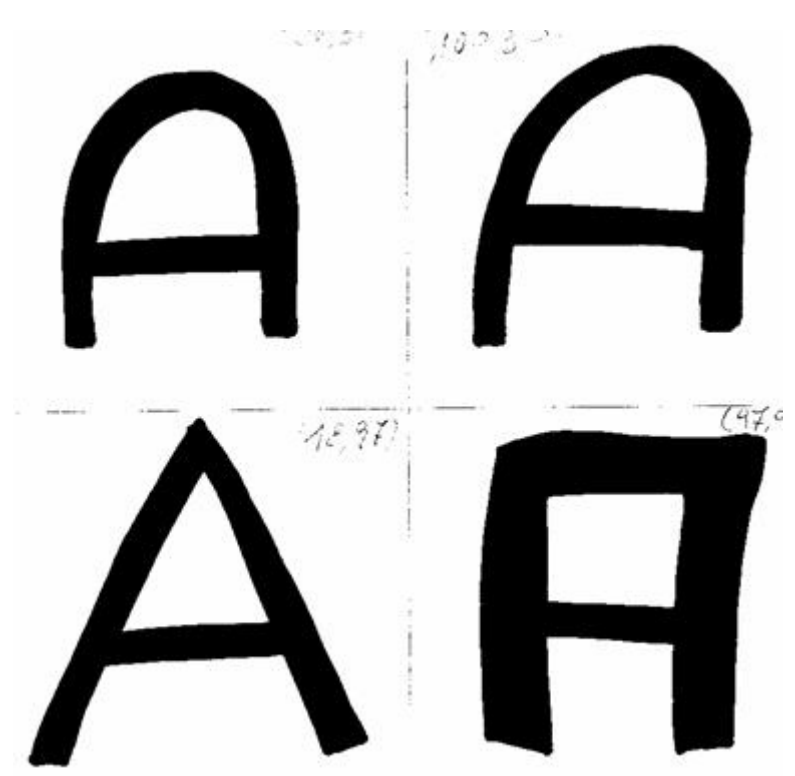


Figura 40

Cancello ogni elemento dalla memoria del computer, mostro al computer una sola A, quella in alto a sinistra e ne registro in RICORDI la sua codifica. Quindi mostro una R, la prima a sinistra e gli chiedo di indicare i particolari diversi fra loro. Ecco il risultato, fig. 41, a prima vista, nessun risultato, direi. Ricordo che le linee verticali rosse sono le caratteristiche della forma A in memoria e quelle verdi le caratteristiche della R attualmente presentata al computer.

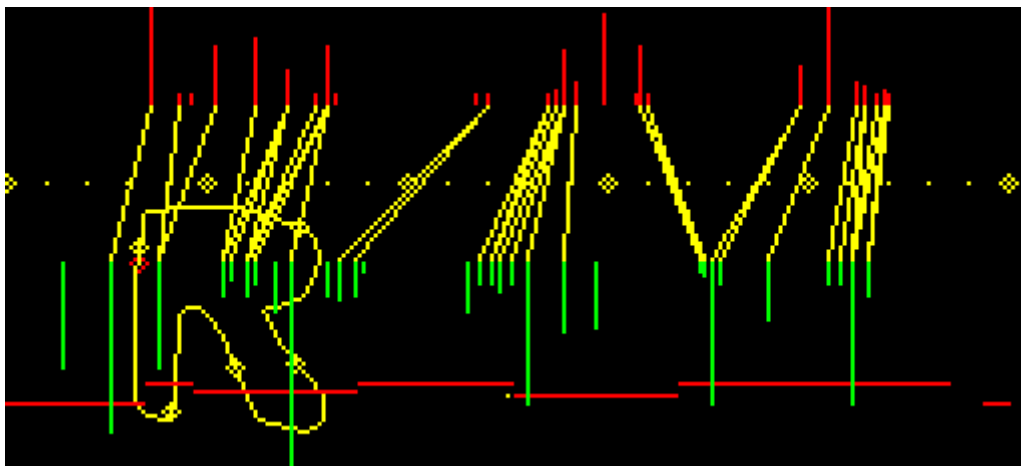


Figura 41

Tuttavia se io “do una mano al computer a capire” quali siano le caratteristiche generali della A, forse le cose andranno in un altro modo. Non che io debba dirgli che cosa deve fare, basta che io gli faccia vedere delle altre A, in modo che lui si faccia un’idea, come farei con un bambino. Allora mostro al computer la seconda A in alto di fig.40, gli spiego che tale forma si chiama A ed esso unirà le caratteristiche con la A di prima, mediando quanto deve mediare, cancellando quando deve cancellare. Alla fine nella memoria del computer avrò le caratteristiche derivanti da due A e di questo ci si rende conto osservando le linee verticali rosse. Ora torno a mostrare al computer la R di prima. Ovviamente le verticali verdi sono immutate e il risultato è questo.

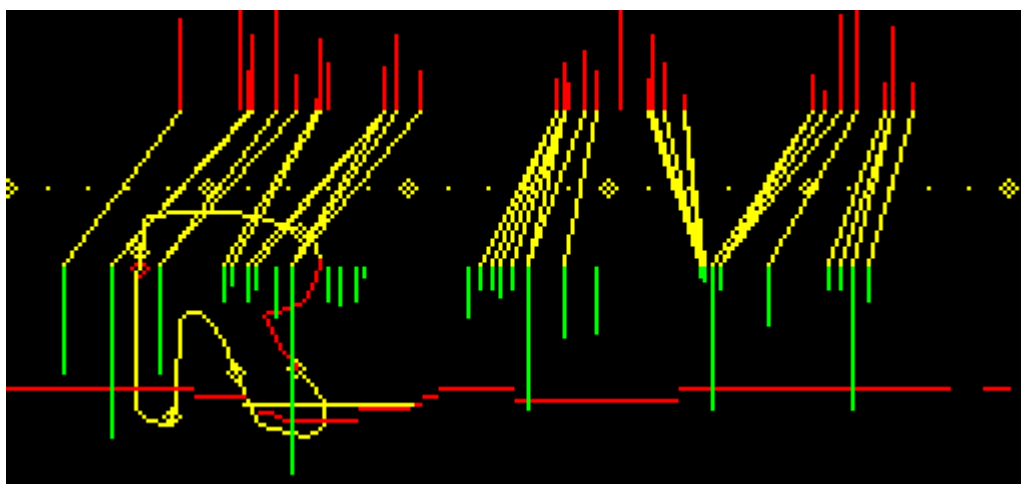


Figura 42

Si noti la parte rossa sulla R, il computer ha individuato la differenza fra le A e le R e tanto non attraverso l’assimilazione di molte forme ma di sole due forme! A ben vedere le caratteristiche della

parte rossa erano già prive di collegamento nella fig. 41, si guardi le linee verdi verticali in corrispondenza della rientranza della R, fra le distanze 50 e 100 i collegamenti sono molto deboli. Tuttavia la potenza dei collegamenti era bassa e il computer non si è fidato a dare un giudizio. In figura 42 l'assenza di collegamento è molto più marcata.

Se io confrontassi la codifica di molte A e di molte R, cioè le due lettere memorizzate avrei la parte rossa ancor più potentemente segnata, anche se basta così.

Si noti che la parte rossa del contorno è una linea, ha delle contratte con tutti gli elementi di esse e può avere un nome, per esempio "rientranza".

Cenno sull'uso delle differenze nel riconoscimento.

Quanto segue sconfinava nel riconoscimento sintattico, è oltre gli obiettivi del momento e pertanto non lo tratterò con la completezza dovuta. Si tratta di una procedura che permette la formazione della catena di percezioni che dà origine alla frase. E' un "ragionamento" che fa la macchina le prime volte che deve distinguere fra certe forme somiglianti per capire quali sono i particolari a cui badare. In seguito impara "a memoria" tali particolari e, di fronte a un riconoscimento analogo e quindi di dubbia certezza, baderà immediatamente a questi particolari, senza più "ragionare". E' ovvio che tali differenze potrebbero essere insegnate da un maestro e la macchina non "faticerebbe". Il richiamo alla memoria, alla sua organizzazione e al modo di recuperare i ricordi entro essa mi impedisce di proseguire il discorso. Riferendoci all'esempio del paragrafo precedente, supponiamo che la forma attuale, quella che sta davanti alla telecamera sia una R e si chieda al computer che nome ha tale forma e supponiamo che dal computer emergano tre risposte:

ERRE78

A.....72

PI.....63.

Il computer non è sicuro (i numeri non sono percentualmente poco differenti) e passerà al sintattico.

Egli cercherà in memoria gli invarianti delle forme di nome A, ERRE e PI. Quindi provvederà a effettuare le differenze:

ERRE-A; ERRE-PI; A-PI (1)

Per capire che cosa siano queste differenze si badi alla fig. 41 sono quelle contratte, rappresentate dalle canne verticali rosse e verdi, non collegate da linee rosse oblique. L'esempio però non è perfettamente calzante: nella fig. 41 il confronto è fra le contratte derivanti dalla elaborazione di una forma attuale, la R, e quelle di una A memorizzata; in questo caso il confronto è fra la ERRE, la A e la PI memorizzate, prese a due a due.

Chiarito che fin qui le differenze sono fatte fra invarianti memorizzati nel file RICORDI.

Quindi la macchina effettua la differenza fra gli invarianti della R attuale, forma che presento ora computer, che, per chiarezza, indico con ?, e quelli delle forme ERRE, A, PI del file RICORDI cioè:

?-ERRE; ?-A; ?-PI. (2)

?-ERRE è certo un insieme piccolo, di elementi poco significativi.

?-A mostrerà la mancanza della rientranza sulla forma R, che è quella visibile, disegnata sullo schermo. La A è in memoria quindi non può essere disegnata, infatti non è memorizzato il disegno della A ma il codice atto a riconoscerla. Con eccellente approssimazione, si può affermare che ?-A sono gli invarianti relativi alla parte rossa della R in fig. 42, che ora è un esempio perfettamente calzante

?-PI farà colorare di rosso, sulla forma R visibile sullo schermo, la “gambetta” a destra che differenzia appunto R da P.

Quindi da queste tre differenze fra forma attuale e figure memorizzate abbiamo rispettivamente:

- un piccolo insieme di invarianti,
- l'insieme degli invarianti della rientranza, con la loro posizione approssimata e, infine,
- l'insieme degli invarianti della gambetta con la loro posizione approssimata.

A questo punto occorre confrontare ognuna delle (2) con ognuna delle (1).

?-ERRE confrontata con

- A-ERRE non corrisponde,
- PI-ERRE non corrisponde,
- ERRE-ERRE corrisponde;

?-PI confrontata con

- A-PI non corrisponde,
- PI-PI non corrisponde,
- ERRE-PI corrisponde;

?-A confrontata con

- A-A non corrisponde,
- PI-A non corrisponde,
- ERRE-A corrisponde.

?-ERRE avendo pochi elementi con potenza bassissima trova corrispondenza con ERRE-ERRE che è un insieme vuoto. ?-PI trova corrispondenza con ERRE-PI.

?-A trova corrispondenza con ERRE-A.

E' quindi avvalorata l'idea che la forma ? si debba chiamare ERRE perché otteneva già parecchi punti dal riconoscimento topologico, mostra le caratteristiche che differenziano la ERRE dalla PI e la ERRE dalla A e poca differenza con la ERRE memorizzata.

Queste corrispondenze sono quantificate con dei numeri, secondo una funzione che tenga anche conto della percentuale di esattezza della corrispondenza ed andranno ad accrescere la potenza che la ERRE ha avuto nel riconoscimento topologico. Di conseguenza il 78 diventerà, per esempio, 130 quindi la figura sarà inequivocabilmente riconosciuta. Parimenti le mancate corrispondenze porteranno alla diminuzione dei pesi della A e della P, che aumenterà ancora la divergenza fra le potenze dei tre nomi proposti dal riconoscimento topologico.

Questo è un ragionamento che permette la formazione della frase e della catena di percezioni. Nel modello cerebrale che propongo, non si usa questo ragionamento ogni volta. Si memorizzano i suoi risultati e si va verso il riconoscimento sintattico.

[Home](#)