

Interazione Uomo Robot

l'elaborazione del linguaggio naturale

Alcuni aspetti delle ultime ricerche in questo settore della HRI presentate in:

- 1st Annual Conference on Human-Robot Interaction (March 2-4, 2006 Salt Lake City, Utah, USA)
- 2nd Annual Conference on Human-Robot Interaction (March 9-11, 2007 Washington DC, USA)

Ugo Moschini

Sommario

1. Introduzione

COME MIGLIORARE IL DIALOGO FRA UOMO E ROBOT ?

2. Come rendere un robot capace di dialogare?
3. Un robot deve comunicare in modo differente da utente a utente
4. Dialoghi di clarification: un robot deve poter chiedere chiarimenti
5. E' importante che un robot capisca lo stato d'animo di chi parla

MULTIMODALITA': IL DIALOGO E L'AMBIENTE

6. Un robot deve avere una prospettiva dell'ambiente
 7. Un robot elimina le ambiguità usando visione, audio e linguaggio naturale
 8. Un caso di studio recente: un'architettura per il training di un robot
9. Bibliografia

1.Introduzione

Il campo dell' interazione uomo – robot (HRI) si sta sviluppando tanto più rapidamente quanto più i robot si rivelano capaci di agire insieme agli esseri umani in ambienti propri di questi ultimi.

Perché i robot siano accettati sempre di più in contesti domestici o in luoghi di lavoro e perché siano in grado di apprendere ed eseguire compiti correttamente, è necessario che le persone non esperte siano in grado di interagire con loro in maniera il più possibile naturale.

E' necessario che i robot raggiungano pertanto un livello di complessità tale che gli umani li considerino come veri e propri "aiutanti" per le loro necessità.

Queste capacità complesse includono l'abilità di riconoscere la presenza e l'attività di persone vicine, possedere una nozione spaziale e semantica dell'ambiente condiviso con le persone e essere in grado di capire (un sottoinsieme di) i linguaggi naturali .

Dotando i robot di meccanismi per comportarsi e interagire socialmente con le persone attorno, è auspicabile che saranno più facilmente accettati dai non addetti ai lavori come parte della loro vita quotidiana normale.

La sfida cruciale per la HRI è poter realizzare un robot a cui si possa con facilità insegnare dei compiti da eseguire o delle semplici nozioni, nello stesso modo in cui un umano potrebbe insegnarlo ad un'altra persona.

Uno dei metodi più importanti e il più immediato a disposizione degli umani per comunicare fra di loro è il linguaggio parlato. Il suo utilizzo isolato, non è sufficiente affinché un robot comprenda veramente una frase, ma la macchina deve affidarsi ad altre osservazioni, come la vista e una più fine capacità uditiva: deve avere una percezione dell'ambiente. In tal senso, si parla di multimodalità.

In questo seminario, l'attenzione è focalizzata sugli aspetti riguardanti l'interazione uomo-robot attraverso il linguaggio naturale. Sono trascurati discorsi generali o introduttivi sulla HRI, presenti in seminari precedenti.

L'analisi è rivolta ai principali problemi e sviluppi futuri in questo campo scientifico, basandosi sulle ultime ricerche in questo campo, proposte nelle ultime due conferenze internazionali sull' HRI del 2006 e del 2007.

2.Come può dialogare un robot?

I gestori di linguaggio (dialog manager) parlato sono quelli che permettono l'interazione uomo-robot più naturale. Diversi problemi, purtroppo, rendono difficile comprendere il reale intento dell'utente.

Prima di tutto, la tecnologia attuale di riconoscimento vocale è affetta da disturbi dipendenti in parte dalla rumorosità intrinseca di numerosi ambienti. Inoltre, la sua qualità varia molto fra utenti diversi.

Inoltre, anche assumendo un perfetto riconoscimento vocale, resta il problema di determinare l'intento dell'utente di fronte ai noti problemi dell'ambiguità linguistica.

Esiste una classe di algoritmi definiti Intelligent Planning Algorithms che dimostrano un notevole grado di robustezza rispetto alla incertezza inerente alla normale interazione umana.

La classe principale di questi algoritmi è quella degli algoritmi POMDP (Partially Observable Markov Decision Process).

Sono usati per scegliere azioni da compiere, quando certi aspetti dell'ambiente in cui si opera sono ignoti. Per fare un esempio, gli scacchi sono un ambiente completamente osservabile, il poker no: le carte altrui rimangono nascoste.

Un POMDP è una tupla comprendente:

- spazio degli stati
- spazio delle azioni
- spazio delle osservazioni
- una probabilità che l'essere in uno stato in un certo momento conduca a un altro preciso stato dopo una certa azione
- il vantaggio ottenuto da una transizione fra due stati

Lo scopo è massimizzare i vantaggi che si ottengono nelle transizioni fra stati.

Un POMDP applicato a sistemi di pianificazione di dialoghi, richiede quindi un accurato modello dell'utente (i differenti stati dell'utente, cosa un utente potrebbe dire, ...).

Solitamente, un POMDP è specificato usando un grande modello probabilistico con molti parametri: questi parametri sono difficili da definire partendo dalla conoscenza del dominio: l'estrazione di questa mole di dati tramite un procedimento a priori è costosissima.

Esistono nuovi approcci a questo tipo di algoritmi caratterizzati dal fatto che il modello dell'utente è costruito contemporaneamente alla gestione del dialogo. Esistono sofisticate tecniche statistiche che sfruttano un intelligente applicazione della regola di Bayes, che permettono al dialog manager di ripianificare la politica corrente di dialogo, tenendo presenti le recenti interazioni avute con l'utente.

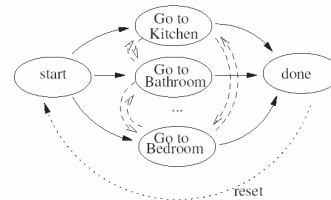
Questo processo aggiuntivo presenta notevoli vantaggi e un costo computazionale che rimane relativamente basso, non inducendo pause importanti in un normale dialogo fra uomo e robot.

Nell'interazione uomo-robot, raccogliere dati sufficienti per costruire un modello accurato è costoso: richiederebbe molto tempo da parte di volontari umani.

Specificare un modello dall'osservazione è altrettanto difficile. Ad esempio: con che probabilità un umano chiede del "tempo", mentre intende riferirsi al "meteo"?

L'idea innovativa è stata partire da un modello relativamente semplice di come un utente può comportarsi, con aggiunta la possibilità di gestire le occasioni di incertezza sul momento (parole riconosciute male, più valori semantici di una stessa parola).

Questo nuovo approccio, si è rivelato utile in campo medico o di assistenza, in quanto prevede un robusto controllo sul vantaggio derivante da certe azioni: fatto indispensabile per non creare situazioni pericolose, soprattutto in campo medico.



Un piccolo esempio di come si può rappresentare in modo semplificato un dialogo POMDP, in un mondo dove è richiesto a un robot (ad esempio, una sedia a rotelle robotizzata) di spostarsi fra varie stanze. Le linee continue rappresentano le transizioni più probabili. Si assume come improbabile (linee tratteggiate) che l'utente cambi l'obiettivo prima che il robot raggiunga una stanza che gli era stata precedentemente ordinata.

E' interessante vedere come questo sistema è stato implementato su una sedia a rotelle robotizzata, con la possibilità di essere comandata a voce. Le specifiche del dialogo non riportavano come gestire le parole "deck" e "elevator".

Nel primo caso, il robot che non è in grado di apprendere una volta decisa la politica di dialogo è confuso e chiede una seconda volta una conferma prima di procedere all'azione relativa.

Nel secondo caso, il robot realizza che "elevator" si riferisce a più luoghi dell'ambiente, ma che l'utente fa più uso solitamente di uno.



Di seguito sono riportate parti del dialogo fra un utente e la sedia.

Without learning
 USER: Take me to the deck between the two towers.
The user says "deck" to refer to the parking lot. The system does not know this yet.
 USER: Where did you want to go?
 ROBOT: Yes, I want to go to central elevator that goes to the deck
 ROBOT: Where did you want to go?
 USER: The middle deck, between Gates and Dreyfoos
 ROBOT: Do you want to go to the Gates Tower?
 USER: No.
 ROBOT: Going to the Dreyfoos Tower.
 USER: No! I want to go to the deck!
The system still cannot associate "deck" with any destination.
 ROBOT: Doing nothing.
 USER: Take me to the deck.
 ROBOT: Where did you want to go?
 USER: The parking lot.
 ROBOT: Where did you want to go?
 USER: I said the deck.
 ROBOT: Where did you want to go?
 USER: Yes, the parking deck.
 ROBOT: Going to the parking lot.
After learning:
 USER: Take me to the deck between the two towers.
 ROBOT: Where did you want to go?
 USER: Yes, I want to go to central elevator that goes to the deck
 ROBOT: Do you want to go to parking lot?
 USER: Yes.
 ROBOT: Going to parking lot.

Primo caso.

Without learning
 USER: Take me to the elevator.
The system does not know what 'elevator' means.
 ROBOT: Where did you want to go?
 USER: The Gates elevator please.
 ROBOT: Do you want to go to the Gates Tower?
 USER: Yes.
 ROBOT: Going to the Gates Tower.
After learning:
 USER: Take me to the elevator.
 ROBOT: Do you want to go to the Gates Tower?
 USER: Yes.
 ROBOT: Going to the Gates Tower.

Secondo caso.

3. Un robot deve comunicare in modo differente da utente a utente

L'interazione uomo robot può essere migliorata progettando robot in grado di gestire un dialogo che si adatti al tipo di utente.

Per una gran parte di ruoli sociali, l'uso del linguaggio naturale è particolarmente appropriato. Si fa riferimento, per esempio, a robot receptionist, guide di musei, ..., che devono essere in grado di rispondere a domande.

Nel caso di un robot receptionist, ci si auspica un diverso tipo di interazione di fronte a una domanda del tipo "dov'è l'ufficio di qualcuno", a seconda che l'utente sia un visitatore esterno rispetto a un impiegato dello stabile.

Il problema è quindi che un robot si trova a dover dialogare con persone non omogenee fra loro, ma ognuna avrà propri interessi e un proprio background culturale. Una mancata attenzione a questo problema porta di fatto a insofferenza da parte dell'utente e al mancato raggiungimento dello scopo per cui un robot è stato progettato.

E' opportuno che il robot, così come le persone, possa far conto su un background di conoscenza condivisa e pensieri condivisi, fra lui e gli utenti.

Per fare ciò, sono usati due approcci:

1. Il robot si basa su caratteristiche come l'età, il sesso, l'appartenenza a qualche gruppo, inteso in senso lato;
2. Il robot riceve direttamente dall'utente un feedback, sotto forma ad esempio di brevi test o domande, dal quale comprende in che modo dialogare con lui.

Come prova di tutto questo, è stato effettuato un esperimento, con un robot "capo-chef", chiamato Pearl, progettato in modo da avere funzioni di assistente per i cuochi esperti e di vero e proprio insegnante per i cuochi principianti.



Figure 1. Pearl as a robot chef.

Gli utenti che interagivano con Pearl, dovevano preliminarmente compilare un breve test con domande di argomento culinario, attraverso le quali il robot definiva il livello di competenza dell'utente.

Which of these objects is called a saucepan?



Un esempio di un semplice test tramite il quale Pearl può dedurre il grado di competenza di un utente.

L' esperimento si divideva a sua volta in due fasi: nella prima, il robot dialogava in modo non adattivo, nella seconda in modo adattivo.

Il risultato è stato che, ovviamente, la carenza di spiegazioni durante la prima fase era dannosa per i cuochi principianti, mentre non aveva impatto per i cuochi esperti. D'altro canto, l'uso di molteplici spiegazioni nei confronti dei cuochi principianti, non danneggiava il rendimento di quelli esperti. E ancora, quando il gruppo di cuochi era sotto pressione e indaffarato e con necessità di collaborare l'un l'altro, l'uso di linguaggio adattivo a seconda dell'utente migliorava molto il rendimento.

Questi risultati suggeriscono che il poter adattare il dialogo nella interazione uomo-robot ha conseguenze importanti sul raggiungimento dell'obiettivo comune e sulla coesione sociale.

Esempio di dialogo con un utente cuoco esperto.

Condition	Robot Dialogue
Names Only	Next you want a sharp paring knife. Find the paring knife.
Names Plus Description	Next you want a sharp paring knife. Find the paring knife. It's usually the smallest knife in a set. It has a short, pointed blade that is good for peeling fruit. The blade is smooth, not jagged.

Esempio di dialogo con un cuoco apprendista, più ricco di dettagli.

Sviluppi futuri

Il dialogo adattivo in linguaggio naturale sarà una sfida nel prossimo futuro, su due dimensioni principali:

1. Nel fare da guida alle persone, a seconda della domanda, un robot potrebbe riferirsi a una sua rappresentazione di rete sociale costruita nel tempo: affiancare cioè alle caratteristiche fisiche di un utente e all' interazione con quest'ultimo un contesto sociale;
2. Affiancare a quanto detto finora, delle capacità che permettano all'utente di entrare in maggior confidenza con il robot: il robot può avere una mimica adatta alla situazione o che si adatti a quella dell'utente, che sia in grado magari di adattare il tono della voce o la velocità delle parole a seconda che individui più o meno difficoltà nella comprensione da parte dell'utente.

4.Dialoghi di clarification: un robot deve poter chiedere chiarimenti

In questa sezione ci si riferisce sull'interazione uomo robot focalizzandosi sui problemi relativi all'organizzazione spaziale e alla localizzazione.

Un umano e un robot interagiscono per stabilire una corrispondenza fra la percezione dell'ambiente sua e la rappresentazione che ne ha il robot al suo interno e che può apprendere tramite i suoi sensori.

Uomo e robot hanno due approcci diversi:

1. Un robot ragiona in termini di spazi metrici;
2. Un umano ha una visione più topologica dello spazio.



Creare un collegamento fra questi due aspetti è essenziale per una migliore interazione uomo-robot.

Le tecniche oggi esistenti implementano un modello di dialogo di tipo master/slave: l'umano parla e il robot ascolta. E' un approccio corretto, ma solo nel caso in cui l'unico scopo sia far conoscere al robot i nomi di locazioni differenti dello spazio.

Ci sono situazioni che dovrebbero permettere una interazione più flessibile, permettendo anche al robot di prendere l'iniziativa nel dialogo:

1. Incertezza nel classificare automaticamente: ad esempio, dove conduce una porta;
2. Inconsistenza fra classificazione e descrizione: la classificazione semantica data a una locazione da un robot può essere differente da quella fornita dall'utente;
3. Fallimenti nella percezione: non sempre si riesce a classificare una nuova localizzazione, ad esempio quando il robot entra in nuove aree o quando i suoi sensori non sono sufficienti per notare alcuni ostacoli o ancora errori sulle distanze,

Uno studio recente fornisce un approccio per rendere un robot capace di "intavolare una discussione" per ottenere un chiarimento per le situazioni dette precedentemente.

I sistemi di dialogo usano meccanismi di chiarimento (clarification) per risolvere i problemi di comunicazione, quando non si comprende cosa vuol significare chi sta parlando. Ciò è da attribuirsi a mancanza di percezione dell'oggetto a cui ci si riferisce, ambiguità fra più significati di una parte del discorso o conflitti nella corrispondenza fra le parole dell' umano e la conoscenza interna del robot.

E' stata realizzata una implementazione per realizzare questi meccanismi di clarification, seguendo un'idea di Ginzburg e Larsson. Tale idea si fonda sulla realizzazione di una gestione del dialogo basata su stati che tengano traccia via via delle domande o dei problemi che devono ancora essere trattati nel procedere di un dialogo.

Una domanda di clarification è costituita due parametri: l'argomento della domanda, ad esempio una locazione, e uno stato in cui questa può essere (ad, esempio: se c'è una PORTA(locazione) QUI ACCANTO(stato)).

Queste domande così implementate risiedono in strutture dati e sarà il sottosistema di comunicazione che, a seconda della sua implementazione, stabilisce una comunicazione con l'utente per poter risolvere un problema di clarification.

Il robot, pertanto, è in grado di iniziare spontaneamente una discussione, pianificando una nuova comunicazione. Genera stringhe per esprimere questo concetto e le enuncia, aggiungendo questo nuovo contenuto al contesto del dialogo.

Ciò è molto importante: quando riceve la risposta da parte dell'utente, il robot la parse e cerca di attribuirle un significato nel contesto corrente che si è creato: si stabilisce di fatto un vero e proprio contesto retorico fra la domanda e la risposta, che aiuta nel risolvere i problemi di clarification.

5.E' importante che un robot capisca lo stato d'animo di chi parla

Riconoscere e rispondere a uno stato d'animo di un utente è importante per un miglior svolgimento di compiti collaborativi fra team misti di umani e robot. Questioni come interpretare correttamente comandi, possono essere risolte mediante il riconoscere espressioni non verbali, come espressioni facciali, certi tipi di movimenti degli occhi o della bocca.

La NASA stessa sta portando avanti studi in questo settore.

Il duplice vantaggio di questo approccio è il seguente:

1. Capire gli stati d'animo abbassa la percentuale di fallimenti e di ambiguità nel comprendere correttamente un certo tipo di comando;
2. Migliora molto le prestazioni e l'efficienza in lavori cooperativi fra uomini e robot.

Uno studio recente ha verificato quantitativamente l'importanza degli stati d'animo e come il rispondere da parte del robot a questi stati d'animo in linguaggio naturale impatti favorevolmente sulle performance.

Si basa sul fatto che molti aspetti di una comunicazione attraverso il linguaggio naturale non possono essere compresi senza tener conto del tono di voce o di altri segnali che indichino uno stato d'animo.

Questo studio ha portato alla definizione di un'architettura dotata di meccanismi che consentano di integrare in un robot l'elaborazione del linguaggio naturale assieme al riconoscimento di stati d'animo dell'utente: l'architettura DIARC.

L'ARCHITETTURA DIARC

E' stata sviluppata un'architettura chiamata DIARC (Distributed Integrated Affect, Reflection, Cognition) che integra meccanismi cognitivi uniti a quelli affettivi.

Tre componenti principali di questa architettura sono:

1. Affective Action Interpreter;
2. Affect Recognition in Spoken Language;
3. Affective Speech Production.

L' **Affective Action Interpreter** è una nuova idea di interprete che è usato per la comprensione del linguaggio naturale, decisione dell'azione e esecuzione della stessa.

Un robot ha al suo interno una serie di scopi sempre presenti (ad esempio, il fatto di "essere gentile", di "non provocare danni") ai quali si aggiungono quelli relativi alle nuove azioni ordinate.

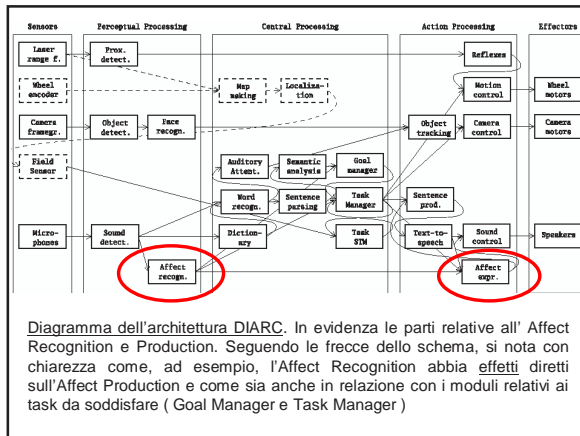
Il robot non esegue i comandi che gli sono ordinati nella sequenza in cui li ha ricevuti. Basa il suo funzionamento su un sistema di priorità fra i compiti, in cui agiscono:

1. i vantaggi relativi a ogni azione
2. il costo dell'azione
3. lo stato d'animo in cui si trova il robot, calcolato da sottoroutine con parametri come il numero di fallimenti nel riconoscere parole, il numero di compiti portati a buon termine o meno e il riconoscimento di determinati stati d'animo nella voce dell'utente, come ad esempio stress,

L' **Affect Detection in Spoken Language** ha come obiettivo il riconoscimento di stati d'animo con la semplice analisi della voce dell'utente. Studi empirici hanno mostrato come ad ogni inflessione nella voce corrisponda una certa variazione dell'intensità di una precisa frequenza. Il robot quindi calcola questa frequenza e a seconda che superi o meno certe soglie, ne deduce lo stato d'animo

dell'umano con cui interloquisce. Il vantaggio di questo approccio basato sulle frequenze è che speaker-independent, e non richiede nessuna attività di training per ogni specifico utente.

Tramite lo stesso principio funziona anche l' **Affective Speech Production**, un sistema adottato per la sintesi del parlato. Ad esempio, per dare al robot una voce del tipo "spaventato", viene modificato il valore e il range della frequenza relativa al suono della voce. Perché abbia un comportamento concitato, è aumentata la velocità tramite la quale sono enunciate le parole.



6. Un robot deve avere una prospettiva dell'ambiente

Nel normale dialogo fra persone, gli umani hanno una percezione dell'ambiente che li circonda che permette loro di far fronte ad una serie di problemi che potrebbero impedire il corretto svolgimento del dialogo: rumori improvvisi che limitano la possibilità di comprensione, spostamenti di posizione da parte di un interlocutore, ...

Un robot deve essere messo in grado di usare la sua conoscenza dell'ambiente, sia a priori sia in seguito a apprendimenti tramite i suoi sensori, per poter predire ciò che l'utente è in grado di comprendere o sentire. Ad esempio, in caso di rumore forte, il robot deve alzare il tono della voce o addirittura sospendere il suo parlato, così come farebbe un interlocutore umano.

In generale, un robot deve adattarsi, come farebbe un umano, a 4 tipi di situazione:

1. Rumori
2. Interruzioni / distrazioni
3. Spostamenti del robot o dell' interlocutore
4. Differenze individuali fra gli interlocutori

E' stata implementata una interfaccia audio-visiva per un vero robot, in grado di gestire le prime due situazioni.

Figure 1. The B21R combines audio and visual interfaces

Il robot aveva il compito di dialogare con militari in un ambiente con rumori tipici di una base militare: linguaggio parlato, rumori di radio e radar, sistemi HVAC (*Heating, Ventilation and Air Conditioning*).

Il robot era dotato di microfoni direzionali e della possibilità di poter alzare o abbassare il tono della voce.

L'esperienza è stata un successo in termini di usabilità di questo robot rispetto ai precedenti e degli effetti positivi sulla collaborazione coi militari.

7. Un robot elimina le ambiguità usando visione, audio e linguaggio naturale

Come visto nei capitoli precedenti, creare una interfaccia uomo-robot è un compito che diventa ben presto molto complesso. Le capacità e le funzionalità dell'interfaccia, dipendono fortemente dalla robustezza dei vari sensori e dalle modalità di input. Ad esempio, il riconoscimento degli oggetti è limitato dalle tecnologie oggi a disposizione. E' stato visto che il riconoscimento del parlato in ambienti rumorosi presenta problemi per i sistemi audio di un robot. Comprendere il linguaggio naturale e i dialoghi è un obiettivo spesso circoscritto a una ben limitato dominio semantico e non raggiunto nel caso di frasi ambigue o nuove.

I dialog managers sono limitati dal fatto di essere specifici di un dominio di applicazione.

La tendenza attuale e per gli anni futuri è di integrare diverse modalità, come la visione, l'audizione e il linguaggio naturale per aumentare le potenzialità di ciascuna modalità, presa come isolata.

Un recente esperimento ha voluto combinare input visivi, auditivi e linguistici in un robot, combinandoli fra loro per risolvere alcuni tipi di problemi:

1. Disambiguare i riferimenti (oggetti nell'ambiente)
2. Localizzazione di speaker umani
3. Determinazione della provenienza di un discorso
4. Appropriattezza delle risposte date all' umano.

Il robot adottato per l'esperimento è quello visto nel capitolo precedente, un esemplare di robot mobile B21r, chiamato George.

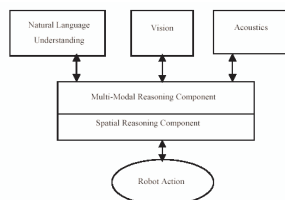


Figure 2. George, a B21r mobile robot

Da B21r ci si aspettava che, affinché fosse largamente efficace, non dovesse solo elaborare gli input dai suoi sensori visivi, auditivi e linguistici, ma che li manipolasse fra loro per risolvere le ambiguità.

Il robot non solo deve discriminare fra oggetti e persone, ma sentire le interazioni verbali, capire chi sta parlando con chi, qual è l'argomento della conversazione e interagire in queste conversazioni.

L'unione delle differenti modalità di input riesce a superare i limiti che avrebbe ognuna presa singolarmente.



L'architettura del sistema.

Nella figura precedente è esposta l'architettura implementata nel robot. Il Multimodal Reasoning Component (MMRC) stabilisce se l'input può essere mappato su una appropriata azione.

Altrimenti, MMRC determina quale elemento manca per poter produrre un'azione e domanda all'utente un'ulteriore informazione verbale, visiva o acustica. A questo punto, MMRC aggiunge questa nuova informazione e riesce a stabilire quale azione eseguire.

In questo studio sono stati integrati:

1. Una tecnologia visiva più sofisticata, in grado di poter riconoscere meglio i gesti di un utente umano
2. Un componente audio migliorato per la localizzazione dei suoni
3. Un nuovo componente di Spatial Reasoning, in grado di riconoscere e localizzare gli oggetti in un modello 3D del mondo
4. Il sistema di riconoscimento del linguaggio naturale, NAUTILUS, che incorpora un dialog manager.

Lo scenario è quello in cui degli umani parlano fra loro e un robot deve interagire con loro, trovando degli oggetti richiesti.

Di seguito, alcune soluzioni evidenziate in questo studio.

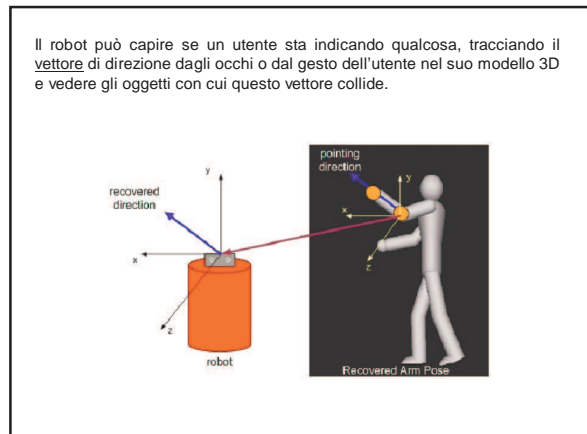
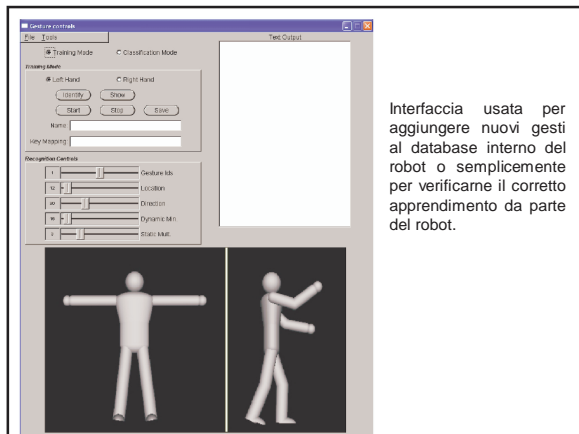
RICONOSCIMENTO DEI GESTI

E' molto importante per un robot riconoscere i gesti generati dalle mani oppure le direzioni in cui guardano gli occhi dell'utente mentre sta parlando.

Il robot è dotato di una visione stereo grazie a due telecamere.

Una nuova idea per risolvere il problema del riconoscimento di gesti è usare i Segmental Hidden Markov Model, già usati per riconoscere il parlato, la scrittura a mano e le forme d'onda. Questo sistema riesce ad adattarsi alle differenti velocità di movimento da utente a utente.

La combinazione di più SHMM permette la comprensione di gesti più complessi rispetto a un semplice "Stop!" o "Vai laggiù!". E' possibile insegnare al robot nuovi gesti tramite un programma o tramite la semplice osservazione dell'utente. Vedi la figura seguente.



RICONOSCERE LE PERSONE

Il sistema visivo del robot cerca profili frontali o volti nella scena che si sta rappresentando, tramite un algoritmo di object detection open source, OpenCV (computer vision library). Una volta che crede di aver trovato un volto, apprende il colore della pelle tramite un istogramma HSV. Per diminuire i falsi positivi, si analizzano i pixel a una certa distanza dal robot, per impedire che lo sfondo della stanza possa influire in senso negativo.

Il robot riesce anche a capire la posizione della mano o del braccio in situazioni ingannevoli: ad esempio, se l'utente indossa una maglietta anziché un maglione, sarà visibile una maggior parte del braccio color pelle. Tale situazione è gestita con successo.



Il piccolo quadrato rosso che circonda il volto della persona dimostra il buon funzionamento del riconoscitore di volti. Si noti che l'intera persona è riconosciuta nelle sue parti dal tronco in su (quadrato giallo) e interamente (quadrato violetto).



Nella figura a sx, si vede come il robot distingue i pixel della persona da quelli dello sfondo.

ACUSTICA

Un robot, per sapere chi è che gli sta parlando, chi sta parlando a chi e quali sono le relative posizioni di chi parla e di sé stesso, necessita di capacità acustiche che vanno oltre il semplice riconoscimento del parlato.

Il robot è stato pertanto dotato di una batteria di microfoni connessi l'un l'altro, tramite i quali il robot effettua i compiti suddetti.

Prima che un robot possa localizzare da dove proviene un discorso, deve prima individuare questo tipo di suoni nell'ambiente. Per questo scopo, ogni microfono ricerca particolari frequenze che si sanno essere tipiche del linguaggio parlato e le compara con una soglia dipendente dall'ambiente in cui il robot si trova. Sebbene sia un metodo semplice e anche soggetto a errori, tuttavia riesce nel suo intento affiancato ad esempio alla visione.

Dopo che la presenza di un discorso parlato è stata individuata, il robot può voler conoscere la posizione o l'origine dei suoni del discorso.

Se è presente nell'ambiente una persona soltanto, il sistema di visione può ricondursi a quella con alta probabilità, ma se ci sono più persone, entra in gioco un utilizzo più potente del sistema auditivo del robot.

Le distanze nello spazio sono basate sul principio delle differenze temporali. Dato che la velocità del suono è finita e i microfoni sono fisicamente distanziati fra di loro, il tempo di arrivo di un segnale a ogni singolo microfono differisce leggermente.

Usando questo stratagemma, un robot può capire da che parte proviene un certo suono, ovvero dalla parte di quei microfoni che l'hanno ricevuto per primi.

COMPRENDERE IL LINGUAGGIO NATURALE

Così come le altre modalità, anche il linguaggio naturale è un valido aiuto per disambiguare posizioni di oggetti o altro.

Per permettere una interazione tramite il linguaggio naturale sono presenti:

1. ViaVoice™ mappa gli enunciati parlati come "Prendimi quel cuscino!" in stringhe;
2. NAUTILUS, un sistema per comprendere il linguaggio naturale, parse queste stringhe. Le rappresentazioni di queste stringhe sono poste in determinati contesti e sono ottenute con la combinazione delle informazioni ottenute dal componente visivo del sistema.

Il contesto, combinato con le informazioni semantiche ad esso associate, permette di determinare se un gesto è necessario o meno.

Ad esempio, se l'umano chiede al robot "Vai là", senza fornire un gesto appropriato, il robot richiede informazione ulteriore, come "Dove?".

Se il gesto non è appropriato, il sistema di linguaggio naturale provvede a raggiungere dei chiarimenti e a risolvere le ambiguità.

C'è anche interesse a disambiguare quei gesti che invece non hanno alcun significato, come ad esempio strusciarsi il naso o arricciarsi i capelli durante una conversazione.

Tutte queste interazioni permettono all'utente del robot di descrivergli le posizioni dei vari oggetti e dar loro un nome affinché il robot li apprenda. Ad esempio, pronunciando "Questa è una sedia"+[gesto], il sistema del robot potrebbe chiedere: "Quale?" Al che l'utente umano può dire: "Quello più vicino/più lontano da te/da me", sfruttando la capacità del robot di possedere una propria prospettiva e conoscenza dell'ambiente circostante.

8.Un caso di studio recente: un'architettura per il training di un robot

L'obiettivo principale e il più ambizioso della HRI è il poter interagire facilmente con un robot allo scopo di insegnargli determinate azioni da intraprendere a seguito di comandi (fase di training).

Esiste l'interesse, come detto precedentemente, di ideare algoritmi che permettano a un umano di poter insegnare a un robot in modo naturale e a un robot di interagire tramite un dialogo intelligente con l'uomo per verificare se la comprensione è stata esatta.

A raggiungere questo scopo concorrono:

1. comprensione di comandi vocali
2. osservazione dei gesti dell'uomo
3. imitazione di quei gesti
4. dialogo fra uomo e robot per verificare la bontà di quanto appreso

Ci sono studi recenti compiuti su bracci robotici, in cui un umano insegna dei movimenti e delle manovre esemplificandoli: ad esempio, un robot deve mettere uno sopra l'altro dei blocchi colorati e l'umano gli mostra quale sia la corretta sequenza.

Altri studi hanno mostrato come un braccio robotico tenti di generalizzare e produrre nuove azioni, basandosi su quelle precedentemente apprese. In questo caso c'è una interazione con l'utente, in cui il robot richiede spiegazioni sulle vere intenzioni di ogni azione dell'operatore così da capire quali azioni sono indispensabili veramente per eseguire un dato compito.

E' stato realizzato un esperimento con un robot che opera in un dominio costituito da una città in miniatura. Per il training del robot è usato il dialogo, a causa del gran numero di informazioni che può essere trasmesso con questa modalità. Se il robot era incapace di eseguire delle azioni, domandava chiarimenti, usando il linguaggio parlato.

Nel training uomo-robot in cui il robot sia un dispositivo di manipolazione semplice, l'attenzione è rivolta a:

1. riconoscimento di oggetti
2. costruzione di piani di azione, in grado di trattare posizioni iniziali differenti di oggetti.

Nel training uomo-robot in cui il robot sia un robot mobile, l'attenzione è rivolta a:

1. mobilità della piattaforma del robot
2. capire dove e come un umano sia in relazione con gli altri oggetti dell'ambiente: c'è una visione più ampia del contesto in cui si opera.

La strada da seguire in ogni caso sembra dotare il robot di sequenze di azioni base e costruire in modo più complesso, gerarchicamente, le altre.

Un articolo presentato a Marzo 2007 all'ultima conferenza sull' HRI presenta un'architettura capace di effettuare il training di un robot.

Il robot osserva un umano che compie un'azione e associa specifici comportamenti che anche lui dovrà ripetere. Sono eseguite più sessioni di training per insegnare correttamente al robot. Dopodiché, se il robot ha problemi, chiede all'umano di osservarlo mentre esegue il compito una prima volta, affinché possa correggere eventuali errori.

Il robot effettua decide dei piani di azione specifici per ogni dominio in cui opera, generati attraverso una osservazione generale dei dati.

L'approccio è quindi combinare una conoscenza *a priori* e informazioni di *contesto* con l'uso del linguaggio naturale parlato. Tale uso permette di eliminare molte ripetizioni di esempio per mostrare come un'azione dev'essere svolta, in quanto i problemi possono essere spiegati, piuttosto che fatti vedere caso per caso.

Vediamo più in dettaglio questo recente studio.

L'architettura che permette di insegnare un'azione al robot si basa sulla combinazione di dimostrazione e descrizione verbale. Se chi insegna e chi apprende hanno una conoscenza condivisa del mondo e degli oggetti di cui si parla, questo approccio è molto efficace.

Per mappare dalle azioni umane (e dalle frasi) alle azioni del robot, il robot ha bisogno di un insieme di comportamenti.

BEHAVIORS

La base del sistema di controllo del robot sono i behaviors (comportamenti). In parole semplici, i behaviors sono funzioni che mappano un insieme di input, compreso le informazioni derivate dai sensori e quelle relative allo stato attuale, in un insieme di azioni.

Internamente, ogni behavior è definito come una macchina a stati finiti, con un singolo stato di partenza e più possibili stati finali, a seconda se gli obiettivi finali sono raggiunti con successo o meno.

Esistono behavior base per costruirne di più complessi. Ad esempio, esistono behavior relativi all'individuazione di una persona nell'ambiente o al muoversi all'interno dell'ambiente stesso.

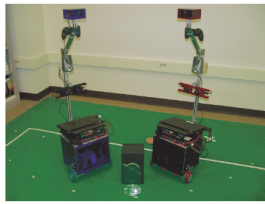
Un elenco di behavior rilevanti per spostarsi e interagire è il seguente:

1. Goto(x,y) / Goto(name): permette al robot di spostarsi da una coordinata all'altra del mondo oppure dirigersi verso una locazione descritta da un nome
2. Say(s) / Tell(s,p) / Ask(s,p): genera discorsi tramite il sintetizzatore del robot
3. Follow(p) / ComeHere(p): il robot individua la persona p e la segue p si dirige verso di lei

Esistono due specifici behavior per il training verso l'apprendimento di azioni da intraprendere per raggiungere uno scopo (*task*) :

1. LearnTask: il robot ascolta ogni frase dell'utente e cerca di tirarne fuori un task
2. FollowLearnTask: permette di iniziare un apprendimento guidato, basato sia sul dialogo che sulla posizione e sui gesti dell'umano nell'ambiente.

Il robot è dotato di una informazione a priori su alcuni aspetti del mondo in cui opera, per poter avere una base di conoscenza a cui appoggiarsi.

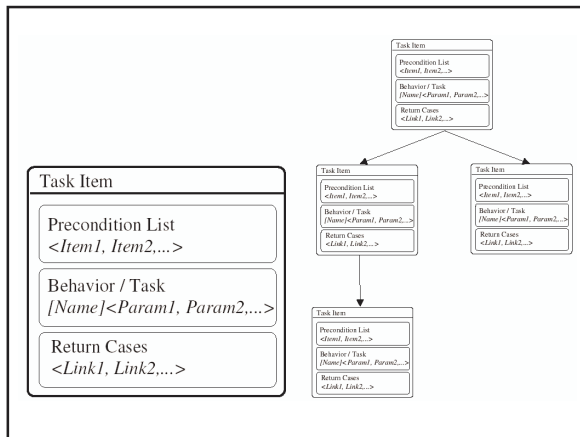


TASK

Un task si viene a configurare, quindi, come un gruppo di behaviors connessi in una struttura a grafo, una rete di behavior o di task, dove gli archi rappresentano transizioni fra un behavior e un altro. In tal modo si viene anche a costituire un utile ordinamento fra behavior necessario per la corretta esecuzione di un task.

Un task è costituito da

1. una lista di precondizioni che devono essere vere perché il task abbia inizio
2. il nome di un behavior
3. collegamenti ad altri task che verranno eseguiti in seguito al successo/fallimento del task precedente.



TRAINING ATTRAVERSO IL DIALOGO E L'INTERAZIONE

Il training ha inizio quando il robot esegue il behavior LearnTask e si pone in ascolto dell'utente, cercando di costruire la rete di task. Il behavior FollowLearnTask è capace di costruire le strutture a rete viste precedentemente.

LearnTask è invocato quando l'umano dice "When I say <TaskLabel>", dove TaskLabel è il nome assegnato al task.

In seguito è costruita una lista chiamata *actionlist* di azioni enunciate dall'umano, sottoforma di costruito condizionale del tipo if-then-else.

```

>> When I say deliver message
>> If Person1 is present
>>   Give message to Person1
>> Otherwise
>>   If Person2 is present
>>     Give message to Person2
>>   Otherwise
>>     Report message delivery failure
>> Before
>> Before
>> Goto home

```

Esempio di modalità con cui l'utente invoca un LearnTask e lo popola di azioni.

FollowLearnTask è invocato quando l'umano dice al robot "Here is what to do when I say<TaskLabel>".

Oltre alle funzionalità del LearnTask semplice, questo behavior aggiunge la possibilità di tracciare l'umano e far sì che il robot individui e generi task comprendendoli da modalità di comunicazione non verbali.

Quando invocato, questo behavior guida il robot accanto all'umano. Il robot aspetta quindi che l'umano dica qualcosa o che si muova. Il robot dispone di una sua rappresentazione interna dell'ambiente e associa i movimenti dell'umano con gli oggetti dell'ambiente che gli sono vicini nei suoi spostamenti.

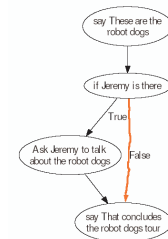
RAFFINAMENTO DEL TASK ATTRAVERSO IL DIALOGO

Una volta che l'umano ha insegnato il task al robot, sono percorsi dei passi di verifica del task stesso.

1. L'umano può chiedere al robot se ha ben compreso il task
2. Essendo complesso per un umano esplorare e descrivere a priori tutte le varie azioni da compiere in ogni ramo del costruito condizionale if-then-else, l'umano descrive solo la condizione vera. Quando si chiede al robot di verificare il task, questi ha in sé solo il ramo per la condizione vera nell'albero delle possibili azioni: il robot allora notificherà tramite il linguaggio parlato all'utente come deve comportarsi negli altri casi.

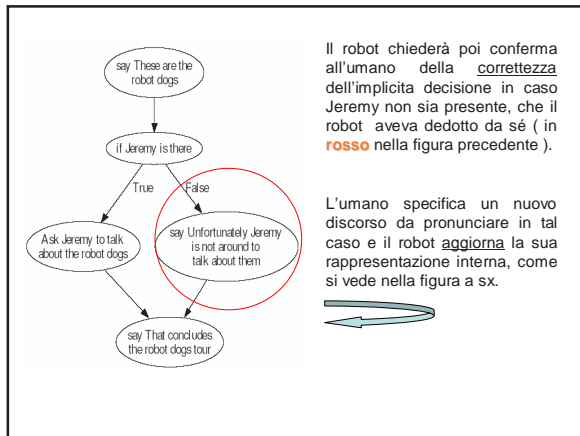
A titolo di esempio, vediamo il caso in cui si chiede al robot il seguente compito: il "Robot dogs tour" task.

Il robot deve condurre i visitatori attraverso varie stanze del laboratorio, dove si trovano i cani-robot, e in ognuna dire allo scienziato che fa da cicerone in ogni stanza (se è presente) che può iniziare la spiegazione.



Vediamo come si svolge la fase di training per questo tipo di task nel semplice caso in cui l'operatore di una data stanza di chiami Jeremy.

La figura a sx, mostra la rappresentazione interna che propria del robot a una richiesta del tipo "Se c'è Jeremy, chiedigli di parlare a proposito dei cani-robot."



9. Bibliografia

1. **Interactive Robot Task Training Through Dialog and Demonstration**
Paul E. Rybski, Kevin Yoon, Jeremy Stolarz, and Manuela M. Veloso
2. **Efficient Model Learning for Dialog Management**
Finale Doshi and Nicholas Roy
3. **Using Vision, Acoustics, and Natural Language for Disambiguation**
Benjamin Fransen, Vlad Morariu, Eric Martinson, Samuel Blisard, Matthew Marge, Scott Thomas, Alan Schultz, and Dennis Perzanowski
4. **Effects of Adaptive Robot Dialogue on Information Exchange and Social Relations**
C. Torrey, A. Powers, M. Marge, S. Fussell, S. Kiesler
5. **Auditory Perspective Taking**
E. Martinson, D. Brock
6. **Using Context and Sensory Data to Learn First and Second Person Pronouns**
K. Gold, B. Scassellati

7. Effects of Adaptive Robot Dialogue on Information Exchange and Social Relations

C. Torrey, A. Powers, M. Marge, S. Fussell, S. Kiesler

8. On Natural Language Dialogue with Assistive Robots

V. Kulyukin

9. The Utility of Affect Expression in Natural Language Interactions in Joint Human - Robot Task

M. Scheutz, P. Schermerhorn, J. Kramer, C. Middendorff

10. Clarification dialogues in human-augmented mapping

G. Kruijff, H. Zender, P. Jenstelt, H. Christensen

11. The Effect of Head-Nod Recognition in Human-Robot Conversation

C. Sidner, C. Lee, L. Morency, C. Forlines

12. How contingent should a communication robot be?

F. Yamaoka, T. Kanda, H. Ishiguro, N. Hagita