

# ProgramD\_TV!\*

## Una chatbot estesa

Emiliano Caponi, Irene Fabrini, Andrea Vanzo

Ingegneria Informatica Laurea Magistrale Università di Tor Vergata<sup>1</sup>

**Abstract.** Le Chatbots sono applicazioni che simulano il dialogo intelligente con esseri umani.

Il paradigma di realizzazione più utilizzato è l'approccio Stimolo-Risposta (S/R). Lo Stimolo è un insieme di frasi attese dalla chatbot, la Risposta sono le frasi restituite all'utente. Spesso tale approccio limita le potenzialità dell'interazione: la chatbot risponde solo quando la frase sottomessa dall'utente corrisponde ad una delle frasi attese. L'obiettivo del lavoro è il "rilassamento" di S/R attraverso due approcci: la semantica relazionale (relazione semantica tra parole) e la semantica distribuzionale (correlazione statistica tra parole). Attraverso tali meccanismi, la chatbot può rispondere anche in presenza di forte similarità tra la frase di input ed una frase attesa, mantenendo ancora coerenti le risposte. L'adozione di questi approcci permette, quindi, di migliorare le prestazioni base della chatbot.

**Keywords:** NLP, dialogo, semantica

## 1 Introduzione

Tra le applicazioni più diffuse dell'elaborazione del linguaggio naturale ci sono l'estrazione e la manipolazione di informazioni.

Tra le altre, ci siamo interessati all'analisi e sviluppo delle Chatbots che hanno il compito di simulare il dialogo e l'interazione tra un essere umano ed un agente. Rendere più efficaci tali sistemi ne permetterebbe l'inserimento in applicazioni industriali di diverse tipologie. Uno dei paradigmi più utilizzati per realizzare una Chatbot è lo Stimolo-Risposta (S/R). Lo Stimolo è rappresentato dall'insieme delle frasi che sono attese dalla chatbot, mentre la Risposta è l'insieme delle frasi che la chatbot ritorna al partner della comunicazione. Negli anni si sono sviluppati diversi linguaggi e piattaforme utili allo sviluppo delle chatbot. Le tecnologie prese in esame in questo lavoro sono l'AIML<sup>2</sup> (Artificial Intelligence Markup Language) e ProgramD<sup>3</sup>. L'AIML si presenta come estensione del linguaggio XML e la sua

---

\* Estensione di ProgramD da parte di studenti di Tor Vergata (TV)

<sup>1</sup> Corso di Intelligenza Artificiale tenuto dalla Prof. M.T. Pazienza, definizione del tema progettuale da parte del prof. F. M. Zanzotto, tutoraggio del dot. A. Turbati

<sup>2</sup> <http://www.alicebot.org/TR/2001/WD-aiml-1.0.1-20011018-005.html>

<sup>3</sup> [http://aitools.org/Program\\_D](http://aitools.org/Program_D)

caratteristica è proprio la codifica del paradigma (S/R). Gli stimoli sono elementi Pattern, mentre le risposte corrispondono ad elementi Template. ProgramD è la piattaforma utilizzata nel lavoro che permette l'implementazione degli standard AIML attraverso il linguaggio Java. Talvolta il paradigma S/R utilizzato da AIML-ProgramD può essere limitante nell'interazione tra utente e chatbot a causa della rigidità del paradigma stesso: la chatbot risponde solo nel caso di perfetta corrispondenza (matching) tra la frase di input ed uno degli elementi Pattern AIML. Data la variabilità lessicale dei linguaggi naturali, un concetto è espresso attraverso diverse forme frasali, che la chatbot non sempre è in grado di riconoscere. Per superare tale limitazione, si è cercato di identificare un approccio alternativo che è stato oggetto della nostra attività progettuale.

L'idea di base è perciò quella di sostituire il semplice pattern-matching con un approccio basato sulla valutazione della similarità tra termini presenti nello stimolo e quelli presenti nella risposta. In questo modo, il successo di una interrogazione non sarà dipendente dalle forme superficiali, ma guidato dal valore di similarità.

## 2 Metodologie utilizzate

L'obiettivo generale del lavoro è quello di estendere ProgramD attraverso l'utilizzo sia della semantica su dizionari strutturati che della semantica distribuzionale con l'intento di estendere il paradigma S/R implementato da ProgramD.

### 2.1 Semantica relazionale su dizionari strutturati: EuroWordNet

Nella semantica relazionale: due parole  $w_1$  e  $w_2$  sono correlate semanticamente se sono in una qualsiasi relazione tra: iperonimia, sinonimia, antonimia... Tali relazioni sono implementate in risorse quali dizionari strutturati o thesauri. Nel progetto si utilizza EuroWordNet<sup>4</sup> per l'italiano che è un database multilingue con reti di parole (wordnet) per alcune lingue europee. Gli EuroWordNet sono strutturati nello stesso modo del WordNet per l'inglese (Princeton WordNet)<sup>5</sup>.

### 2.2 Semantica distribuzionale

La semantica distribuzionale considera come due parole  $w_1$  e  $w_2$  co-occorrono all'interno della stessa frase o di una finestra prestabilita in fase di progetto, considerando la loro distribuzione. In particolare, si sfrutta la Latent Semantic Analysis (LSA) che permette di estrarre e rappresentare il significato delle parole statisticamente, da un vasto insieme di documenti. LSA analizza le relazioni tra un insieme di documenti e i termini che essi contengono producendo un insieme di concetti (o topic) relativi ai documenti e ai termini. LSA assume che parole che sono

---

<sup>4</sup> <http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=820/vers=ita>

<sup>5</sup> G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, «Introduction to WordNet: An Online Lexical Database,» *International Journal Lexicography*, pp. 235-244, 1990.

vicine di significato occorreranno vicine nel testo (co-occorrenza). La matrice di partenza è una matrice termini per documenti, estratta da un corpus. Attraverso una trasformazione matriciale (SVD – Singular Value Decomposition, produce una approssimazione della matrice di partenza con rango minore), si estraggono nuove dimensioni (topic) che catturano meglio il contenuto informativo della matrice di partenza.

### 3 Descrizione ProgramD\_TV

Gli elementi essenziali di ProgramD sono:

- 1) **Core**: modulo che permette di configurare tutti gli elementi per l'esecuzione della ChatBot;
- 2) **Parser**: effettua l'analisi dei file AIML ed XML;
- 3) **Graph**: contiene elementi quali Nodemapper e Graphmaster che hanno il compito di implementare il pattern S/R;
- 4) **Multiplexor**: ha il compito di gestire i flussi di input e di output;
- 5) **Shell**: l'interfaccia grafica che permette l'interazione Utente/Chatbot.

Nella figura 1 si può vedere una schematizzazione della struttura associata a ProgramD.

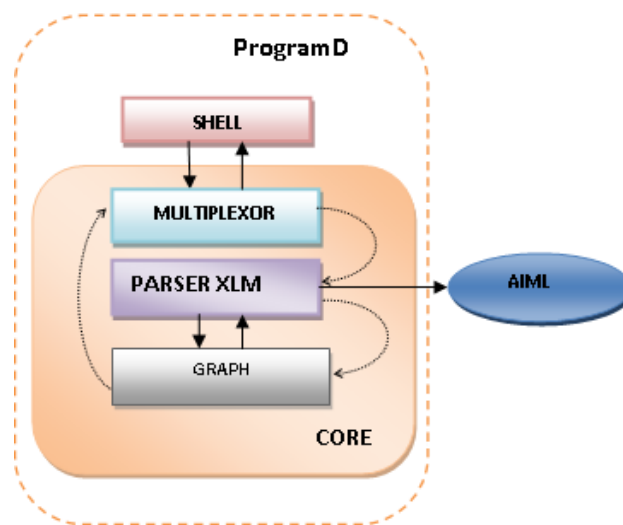


Fig. 1. Overview architettura ProgramD

Il modulo preso in esame è il Graph ed in particolare l'elemento GraphMaster, che ha il compito di costruire l'albero delle frasi attese dalla Chatbot e gestire il meccanismo di matching responsabile della navigazione. La costruzione dell'albero è un'operazione semplice: a partire da un elemento radice fittizio, le parole presenti nel file AIML corrispondono ai nodi dell'albero mentre i Template sono gli elementi foglia. Un elemento Pattern è un percorso tra l'elemento radice ed un nodo che precede gli elementi foglia. L'operazione di matching è anch'essa molto intuitiva. Data una frase di input composta da  $N$  parole:  $w_1, w_2, \dots, w_N$  la condizione di matching è soddisfatta se sarà presente in un qualsiasi percorso dell'albero la sequenza  $w_1, w_2, \dots, w_N$ . Questa breve descrizione sulla operazione di matching mostra i limiti del paradigma S/R cioè deve essere presente nell'albero la sequenza  $w_1, w_2, \dots, w_N$  senza poter ragionare in termini di similarità semantica tra la parola  $w_i$  dell'input e  $w_n$  presente nel nodo. Per superare questo limite è stato arricchito ognuno dei nodi del Graphmaster con i synset associati alla parola  $w_n$  presente indipendentemente dalla categoria sintattica del termine in esame. Questo semplice approccio rappresenta la BASELINE del lavoro. Per fare ciò è stata implementata una libreria che, sfruttando le interfacce e la libreria JMWNL<sup>6</sup> (Pazienza M.T., A. Stellato, A. Tudorache, 2008), utilizza il lessico strutturato di EuroWordNet (Pazienza M.T., A. Stellato, A. Tudorache, 2008). Ogni termine appartenente al pattern AIML analizzato da ProgramD viene espanso mediante EuroWordNet andando così a formare un insieme di termini sinonimi piuttosto che un termine singolo. In questo modo, durante la fase di matching sarà possibile catturare anche frasi che contengano termini sinonimi a quelli già presenti nel file AIML.

Per valutare tale approccio sono state implementate tre metriche di similarità:

1. **LCH** (Leacock C., M. Chodorow, 1998) trova il più corto path tra due concetti e scala questo valore con la massima lunghezza del path trovata nella gerarchia *is-a* della tassonomia nella quale occorre:

$$sim_{LCH}(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2D} \quad (1)$$

dove  $len(c_1, c_2)$  è la lunghezza del path più breve tra due concetti e  $D$  è la massima profondità della tassonomia.

2. **WUP** (Wu Z. and M. Palmer, 1994) trova la profondità del LCS (Least Common Subsumer) dei concetti e la scala con la somma delle profondità dei concetti individuali (la profondità di un concetto è la sua distanza dal nodo radice):

$$sim_{WUP}(c_1, c_2) = \frac{2N}{N_1 + N_2} \quad (2)$$

dove  $N$ ,  $N_1$  e  $N_2$  sono le distanze tra nodo root e LCS, concetto  $c_1$  e concetto  $c_2$  rispettivamente.

---

<sup>6</sup> forniteci dal gruppo ART di Tor Vergata: <http://art.uniroma2.it/>

3. **PATH** (Pedersen T., S. Patwardhan, J. Michelizzi, 2004) è una metrica di similarità baseline definita come:

$$sim_{WUP}(c_1, c_2) = \frac{1}{path(c_1, c_2)}$$

dove *path* è il percorso più breve tra il concetto  $c_1$  e concetto  $c_2$ .

Con queste tre metriche si calcola la similarità tra la parola in input  $w_i$  e le parole  $w_p$  dei pattern AIML dello stesso livello. Fra tutti i valori che superano un certo valore di threshold, viene scelto quello massimo, che evidenzia quindi un valore maggiore di similarità. Proseguirà, quindi, il matching per il nodo associato al valore scelto. Si è scelto di utilizzare **LCH** come metrica di similarità, fissando il threshold ad 1.8 a fronte dei test eseguiti.

Per utilizzare anche aspetti di semantica distribuzionale si è usata una matrice LSA (Latent Semantic Analysis) formata da 31779 termini per 250 topic, creata utilizzando una finestra di 3 parole per il contesto destro e 3 parole per il contesto sinistro a partire da un corpus di documenti estratti da Wikipedia<sup>7</sup> fornitaci dal gruppo ART<sup>8</sup>. Per calcolare la similarità tra la parola in input  $w_i$  e la parola  $w_p$  dei pattern AIML si è utilizzata la *cosine similarity*:

$$sim(w_i, w_p) = w_i * w_p / |w_i| * |w_p|$$

tra i vettori della matrice LSA corrispondenti a  $w_i$  e  $w_p$ .

Anche in questo caso, tra tutti i valori che superano una certa soglia (0.7), si sceglie quello massimo.

Il nuovo spazio vettoriale (le dimensioni sono i topic ed i vettori i termini) sarà utilizzato per valutare la correlazione semantica tra termini con metriche di similarità tra vettori.

## 4 Testing

Per il testing è stato utilizzato come file di riferimento Religione.aiml<sup>9</sup> in lingua italiana ed appartenente al Package AIML Maria. Questo file consiste di 133 coppie Pattern/Template. Il testing è basato sul confronto tra le risposte della Chatbot con quelle previste da un essere umano e consiste nelle seguenti operazioni:

---

<sup>7</sup> Il corpus è sviluppato dalla comunità WaCky ed è disponibile nel sito web: <http://medialab.di.unipi.it/Project/QA/wikiCoNLL.bz2>

<sup>8</sup> <http://art.uniroma2.it/> Si ringraziano i dottorandi D. Croce e D. Previtali per aver condiviso i dati (Croce D., D. Previtali, 2010).

<sup>9</sup> [http://aitools.org/Free\\_AIML\\_sets#Italian\\_AIML\\_.28Maria.29](http://aitools.org/Free_AIML_sets#Italian_AIML_.28Maria.29)

- 1) costruzione del Testing Set (TS);
- 2) annotazione manuale del TS (Gold Standard);
- 3) confronto tra le risposte della Chatbot e quelle del Gold Standard.

Nella fase 1 si sono generate 174 domande in lingua italiana sull'argomento religioso. Per la fase 2 è stato utilizzato un approccio di plausibilità della risposta; pertanto si considerano 3 annotatori che selezionano per ogni domanda, le 3 risposte più plausibili. Si fa presente che questo è un approccio restrittivo che può limitare l'identificazione di risposte esatte tra quelle fornite dal sistema. Per gestire l'agreement tra gli annotatori si è utilizzata la K-Fleiss Statistics (Fleiss I. J., 1971), una metrica utile a stabilire quanto le annotazioni sono tra di loro concordanti e molto flessibile sia alla presenza di annotatori multipli che di multi-classificazione. Tuttavia nel testing, la classificazione è stata binaria (Risposte Rilevanti/Non Rilevanti). Nella fase 3 è eseguito il testing, i cui risultati sono visibili nella *tabella 1*.

Come si può vedere nella prima riga sono presenti i 4 approcci differenti utilizzati. I valori all'interno della tabella rappresentano perciò, per ogni approccio, rispettivamente:

- Numero di domande a cui la chatbot non riesce a rispondere diviso il numero totale di domande sottomesse
- Numero di domande a cui la chatbot risponde in modo corretto diviso il numero totale di domande sottomesse
- Numero di domande a cui la chatbot risponde in modo errato diviso il numero totale di domande sottomesse

	ProgramD (senza estensioni)	BASELINE (espansione EuroWordNet)	BASELINE + METRICHE DI SIMILARITÀ	BASELINE + LSA
Non Riponde	145/174	126/174	126/174	114/174
Risposte Corrette	19/174	29/174	29/174	37/174
Risposte Errate	10/174	19/174	19/174	23/174
TOT.	174/174	174/174	174/174	174/174

**Table 1.**

## **5 Analisi dei risultati e conclusioni**

L'obiettivo proposto era quello di migliorare la capacità di riconoscere, da parte di una chatbot, un concetto espresso in modo variabile attraverso diverse forme frasali. Dai dati di testing (*tabella 1*) si evince che si è riusciti ad avere un miglioramento di circa il 6% con l'espansione tramite EuroWordNet (BASELINE). In questo caso, la chatbot risponde a 19 domande in più rispetto a ProgramD senza estensioni, di cui 9 errate e 10 corrette. Con la seconda estensione sono state ottenute le stesse prestazioni della precedente a causa del vincolo temporale: la chatbot deve rispondere entro un certo intervallo di tempo, allo scadere del quale la risposta viene considerata errata. L'ultima modalità usata è quella che dà i risultati migliori, infatti viene incrementata

la percentuale di risposte corrette dell' 11%. In questo caso, l'incremento di decisioni, rispetto a ProgramD semplice, è stato di 31 risposte, di cui 13 errate e 18 corrette. Nel confronto tra i due approcci (BASELINE e BASELINE+LSA), l'introduzione di LSA permette 12 decisioni in più, con 4 errori commessi. L'incremento di risposte errate è dovuto a due fattori:

- il rischio di errore maggiore nel rispondere a più domande
- la creazione del Test Set mediante una annotazione restrittiva (come già analizzato nella sezione 4 *Testing*)

Tali risultati permettono di pensare con fiducia ad applicazioni industriali.

Inoltre un possibile sviluppo futuro potrebbe essere quello di prevedere un approccio che utilizzi la similarità semantica su base ontologica, in modo da poter sfruttare non solo la distanza tra i concetti, ma anche i valori delle proprietà tra loro in comune.

## 6 Referenze

- Croce D., D. Previtali. (2010). Manifold Learning for the Semi-Supervised Induction of FrameNet Predicates: An Empirical Investigation. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics* (p. 7-16). Sweden: Association for Computational Linguistics.
- Fleiss I. J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 378-382.
- Leacock C., M. Chodorow. (1998). Combining local context and WordNet similarity for word sense identification. In C. Fellbaum, *WordNet: An electronic lexical database* (p. 265–283). MIT Press.
- Pazienza M.T., A. Stellato, A. Tudorache. (2008). A Bottom-up Comparative Study of EuroWordNet and WordNet 3.0, Lexical and Semantic Relations. In K. C. Nicoletta Calzolari (Conference Chair) (A cura di), *LREC*. Marrakesh: European Language Resources Association (ELRA).
- Pazienza M.T., A. Stellato, A. Tudorache. (2008). JMWNL: An extensible multilingual library for accessing wordnets in different languages. In K. C. Nicoletta Calzolari (Conference Chair) (A cura di), *LREC*. Marrakesh: European Language Resources Association (ELRA).
- Pedersen T., S. Patwardhan, J. Michelizzi. (2004). WordNet::Similarity - Measuring the Relatedness of Concepts. In *Demonstration Papers at HLT-NAACL 2004* (p. 38-41). Boston, Massachusetts: Association for Computational Linguistics.
- Wu Z. and M. Palmer. (1994). Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics* (p. 133–138). Las Cruces, New Mexico.