Implementazione di un dimostratore di teoremi per risoluzione

Enrico Scapin vr353597

30 gennaio 2012

1 Introduzione

Uno degli obiettivi del ragionamento automatico consiste nel cercare di costruire *Dimostratori Automatici di Teoremi* al fine di ottenere meccanismi automatizzabili per asserire, a partire da un insieme di assunzioni H, la validità o meno di una determinata congettura φ. Più formalmente:

$$H \models \varphi$$

I dimostratori automatici tipicamente procedono in maniera refutazionale in quanto ogni formula è valida se e solo se la sua negazione è insoddisfacibile. Quindi, riferendoci al sequente sopra:

$$H \models \varphi \iff H \cup \{\neg\varphi\} \text{ è insoddisfacibile}$$

Da questa considerazione possiamo quindi cercare di costruire una procedura che, se dimostra l'insoddisfacibilità di $H \cup \{\neg\varphi\}$, allora $H \models \varphi$ è valido, altrimenti, se ne dimostra la sua soddisfacibilità, il modello che lo soddisfa costituirà il controesempio alla validità di $H \models \varphi$.

Siamo interessati a dimostrare formule espresse in Logica del Primo Ordine[1] in quanto, al contrario della logica proposizionale, essa è sufficientemente espressiva da poter modellare una buona parte della nostra conoscenza. Utilizzando questo linguaggio però il problema della validità non è più decidibile bensì semidecidibile: infatti se l'insieme $H \cup \{\neg\varphi\}$ è insoddisfacibile allora la procedura terminerà sempre con la risposta corretta, mentre se esso è soddisfacibile non è detto che la procedura termini.

La semidecidibilità deriva direttamente dal *Teorema di Herbrand* che afferma che un insieme finito S di clausole in *FOL* è soddisfacibile se e solo se esiste un insieme finito S' di istanze ground (clausole in cui tutte le variabili sono istanziate ad una qualche costante) di clausole di S tale che S' è soddisfacibile. Quindi per dimostrare la soddisfacibilità di S è necessario generare tutti gli insiemi S' di istanze ground e dimostrarne la soddisfacibilità ma, se in S si quantifica su insiemi infiniti, allora la cardinalità degli insiemi S' da generare è anch'essa infinita.

Questo dimostratore prende in ingresso un insieme di clausole scritte in Forma Normale Congiunta[2] e definite da una sintassi standard compatibile con frammento CNF senza uguaglianza della libreria *TPTP*[3]. Si è quindi implementata una procedura di semi-decisione che, basata su un sistema di cinque regole di inferenza (di cui due di espansione e tre di contrazione), implementa un piano di ricerca denominato *Ciclo della Clausola Data* (Given clause loop) che è uno standard alla base di molti dimostratori di insieme di formule in logica al primo ordine come Otter, E, Vampire, Gandalf etc.

2 Scelte progettuali

L'elaborato è stato implementato utilizzando il linguaggio *Java* in quanto il livello di astrazione è tale da permettere al programmatore di concentrarsi principalmente sulla progettazione dell'algoritmo ed, in particolare, su quali strutture dati sia meglio utilizzare. Inoltre la scelta di Java consente l'utilizzo di *JavaCC*, uno strumento che ha permesso di effettuare sia il parsing delle formule sia quello degli argomenti che vengono passati da riga di comando.

La scelta di tale linguaggio ha inoltre permesso un minimo di progettazione orientata agli oggetti soprattutto delle cosidette classi bean[4] e, in particolare, delle classi che definisco i termini (funzioni, variabili e costanti) e i letterali.

[1] *FOL*, First Order Logic

[2] *CNF*, Conjunctive Normal Form

[3] http://www.cs.miami.edu/ tptp/

[4] così si denotano le classi che logicamente contengono le informazioni da manipolare

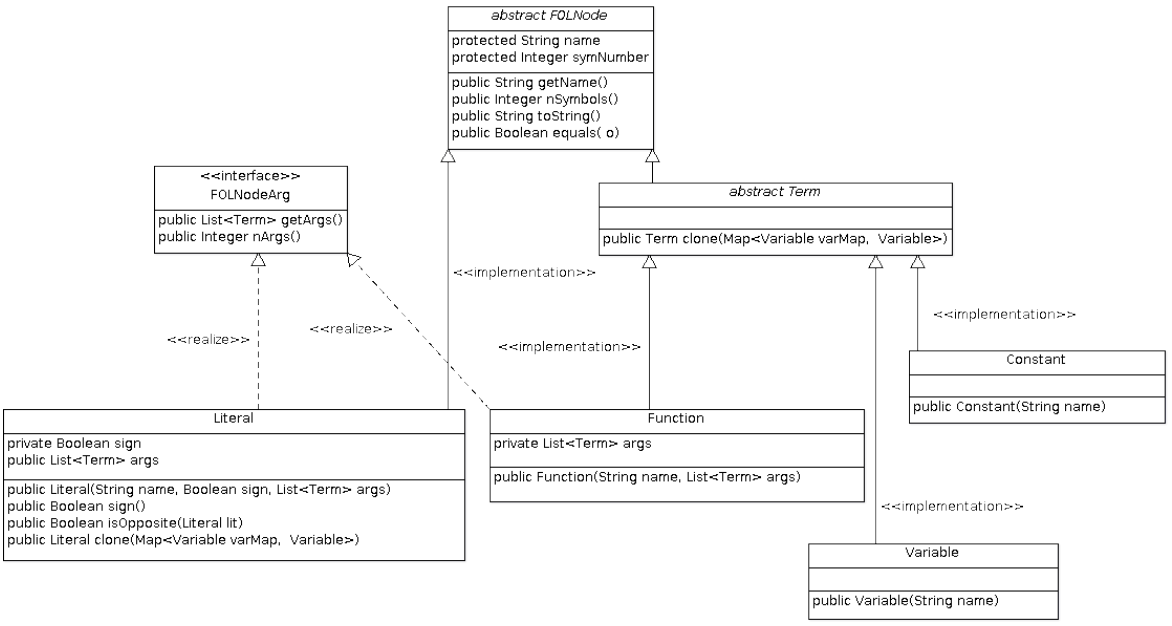

Figura 1: UML delle bean class

Il diagramma UML visibile in Figura 1 rappresenta come è state progettate le classi: ovviamente è molto rilevante la classe astratta `Term`, che rappresenta un qualsiasi oggetto di tipo `Function`, `Variable` o `Constant`, in quanto è spesso necessario riferirsi ad una di esse indistintamente (come ad esempio quando si specificano gli argomenti di un letterale). Il metodo `clone(Map<Variable, Variable>)` presente sia nella classe `Term` sia nella classe `Literal` permette di clonare l'oggetto in questione restituendone un altro con la stessa struttura. Questo metodo si rivela essere molto importante quando è necessario generare una nuova clausola a partire da un'altra: infatti è sufficiente chiamare il metodo `clone` su ogni letterale che ricorsivamente clonerà anche i termini che compongono i suoi argomenti. Quando si crea un nuova clausola è però necessario mantenere inalterato il numero delle variabili in modo tale che tutte le occorrenze di ogni variabile nella vecchia clausola vengano rimpiazzate dalla stessa nuova variabile nella nuova: per far ciò è necessario avere una mappa che mantiene il collegamento tra il vecchio oggetto `Variable` con il nuovo se esso è già stato creato in precedenza in quanto quella determinata variabile è stata già incontrata in precedenza durante la clonazione. Alternativamente se quella variabile non è stata incontrata in precedenza allora è necessario crearne una nuova e inserire la coppia $(vecchia, nuova)$ all'interno della mappa. Per quanto riguarda il metodo `toString` è, anche in questo caso, rilevante l'implementazione nella classe `Variable` in quanto per differenziare le variabili diverse ma con lo stesso nome si è scelto di concatenarci gli ultimi 3 caratteri del loro *hash code* codificato in esadecimale. Da notare infine il metodo `equals` che nel caso delle variabili ritorna `true` se e solo se sono lo stesso oggetto mentre nel caso delle costanti ritorna `true` se esse hanno lo stesso nome (e questo è consistente col fatto che le costanti sono le stesse per tutto l'insieme di clausole). Di conseguenza i metodi `equals` delle funzioni e dei letterali, dopo aver controllato l'uguaglianza dei segni e/o dei nomi, andranno in ricorsione sul termini che compongono i loro argomenti e solo se sono tutti uguali ritorneranno `true`.

2.1 Parser

Dopo che la stringa è stata letta (inserita tramite riga di comando oppure prelevata da un file), il parser si preoccupa di farne un'analisi lessicale e sintattica per il riconoscimento di formule espresse nell'alfabeto $\Sigma_E \cup \Sigma_{cons}$, ovvero l'unione delle segnature della teoria dell'uguaglianza e delle liste (che ne è un estensione).
Il parser, che si può trovare nella cartella *src/parser*, consiste in un file con estensione `.jj` costituito da un'unità di compilazione java e da una grammatica context-free di tipo *LL(2)* le cui produzioni sono espresse in BNF (*Backus–Naur Form*). La grammatica è formata dalle seguenti categorie sintattiche:

```
Formula ::= Clause (; Clause? )*
Clause  ::= Term = Term | Term != Term | atom(Term) | -atom(Term) |
            (-)? Pred(Term (, Term)*)
Term    ::= cons(Term, Term) | car(Term) | cdr(Term) | FunVar(Term (, Term)*) | FunVar
```

I token `Pred` e `FunVar` sono specificati dalle seguenti espressioni regolari:

`Pred` ::= `([A-Z])`$^+$`([0-9])`*
`FunVar` ::= `([A-Z])`*`([a-z])`$^+$`([a-z,A-Z])`*`([0-9])`*

Considerazioni:

- il punto e virgola denota una congiunzione tra clausole;
- sono stati utilizzati due caratteri diversi per la negazione solamente per ovviare ad un problema di compatibilità con la bash in caso l'input sia inserito da riga di comando;
- i simboli di predicato non interpretati sono espressi tramite lettere maiuscole mentre i simboli di funzione e di variabile con almento una lettera minuscola (opzionalmente possono anche essere seguiti da cifre);
- ogni clausola identifica un unico predicato che è formato da uno o più termini; ogni termine può essere a sua volta formato da sottotermini oppure identificare un simbolo terminale, ovvero una variabile o una costante (nel frammento senza quantificatori si perde la distinzione fra queste due categorie sintattiche).

Per quanto riguarda l'aspetto semantico, per ogni predicato e per ogni termine è stato inserito del codice Java che principalmente si occupa della costruzione del DAG, mantenuto tramite una tavola hash implementata utilizzando la classe `HashMap`. Il grafo è costruito mediante un approccio bottom-up.

- Ogni produzione di ciascun termine riceve uno o più oggetti di tipo `Node` (*src/bean*) che rapprensentano i nodi del grafo dei suoi sottotermini; viene creato un nuovo nodo rappresentante quel termine ed si effettua il matching tra il suo campo *argument* (`LinkedList`) ed il campo *parent* (`HashSet`) dei suoi figli. Ciò avviene inserendo la chiave del nodo nei campi *parent* dei figli e le chiavi dei figli nel campo *argument* del padre.
- Ogni produzione di ciascun predicato riceve anch'esso uno o più oggetti di tipo nodo ma si comporta in maniera diversa a seconda del tipo di predicato:

  a) per quanto riguarda i predicati di uguaglianza, vengono semplicemente inserite le coppie di termini[5] in due insiemi a seconda che il predicato sia positivo o negativo;

  b) per quanto riguarda il prediato `atom`, se è positivo il suo argomento viene inserito nell'insieme (`HashSet`) di tutti i termini che sono argomenti di `atom`, mentre se è negativo si effettua una trasformazione sintattica creando tre nuovi nodi: due variabili fresh ed un termine `cons` che ne diviene il padre; la coppia formata da questo nuovo nodo e dall'argomento del predicato viene poi aggiunta all' insieme dei predicati di uguaglianza;

  c) per quanto riguarda gli altri predicati non interpretati si effettua anche in questo caso una trasformazione sintattica in maniera da convertirli in fuzioni; li si inserisce poi nell'insieme dei predicati di uguagliaza o di disuguaglianza, a seconda che siano positivi o negativi, in coppia con un simbolo speciale che è stato scelto essere "#".

Per quanto riguarda le produzioni delle funzioni e dei predicati in cui il numero di argomenti non è noto a priori, si mantiene una corrispondenza (`HashMap`) tra i simboli letti ed il corrispettivo numero di argomenti. In questo modo, in caso si trovino due simboli uguali con un diverso numero di argomenti, viene sollevata un'eccezione.
Durante l'elaborazione di ogni nodo, vengono aggiornati anche tutti quei dati, come il numero di archi o di clausule di uguaglianza, che saranno poi forniti in output.

2.2 Algoritmo di Nelson-Oppen

Dopo che è stato effettuato il parsing dell'intera formula e creato il DAG, viene eseguito l'algoritmo di chiusura di congruenza di cui ne sono state implementate due versioni selezionabili dall'utente mediante l'eventuale uso dell'opzione `-h`.

La prima versione è la semplice trasposizione in codice Java dell'algoritmo presente al capitolo 9.3 del libro *The Calculus of Computation* di A. R. Bradley e Z. Manna.

La seconda versione è una nuova reingegnerizzazione di tale algortimo tramite l'utilizzo di alcune euristiche atte soprattutto a permettere, non appena viene trovata una contraddizione tra due termini, l'immediata interruzione del flusso di esecuzione restituendo insoddisfacibile.

L'idea è quella di mantenere per ogni nodo un *banned set* composto da tutti quei termini che non possono appartenere alla sua stessa classe di congruenza. Tale insieme viene popolato durante il parsing: ogniqualvolta si incontra un predicato di disuguaglianza $s_i \neq t_i$, viene aggiunto al *banned set* del nodo

[5] Poichè si è reso necessario memorizzare i termini in coppia, è stato creata la classe `TermPair` (*src/bean*) che tramite l'overriding dei metodi `hashCode` ed `equals` permette di mantenere coppie non ordinate nei vari `HashSet`.

che rappresenta s_i il termine t_i e viceversa. In questo modo ogni volta che due termini devono essere uniti in base agli assiomi di congruenza, si controlla se uno è nel *banned set* dell'altro ed, in caso affermativo, si restituisce immediatamente insoddisfacibile.

Questa euristica può essere ulteriormente migliorata sulla base di questa considerazione: quando si effettua l'operazione di `merge` tra due nodi in realtà si stanno unendo le loro classi di congruenza e quindi è utile controllare non solo se i due termini formano un predicato di disuguaglianza, ma anche se nel *banned set* di ciausun nodo vi è il riferimento ad uno dei nodi appartenenti alla stessa classe di congruenza dell'altro. Per implementare ciò, quando si effettua l'operazione di `union`, non si fa l'unione solamente tra i campi `ccpar`, ma anche tra i campi `banned`. In questo modo, ogniqualvolta viene eseguito il metodo `merge` tra due nodi, si controlla se è possibile unire le loro classi di congruenza: si seleziona il *banned set* del rappresentante di uno dei due nodi e, per ogni termine presente in esso, si controlla se la sua classe di congruenza è la stessa classe dell'altro nodo con cui si vuole effettuare la `union` ritornando, in caso positivo, insoddisfacibile.

```
// id1, id2 : nodes' id for which make union
for(String ban: node(find(id1)).getBanned())
                if(find(ban).equals(find(id2))) // CONFLICT
                        return new TermPair(id1, id2);
```

Tale operazione è ripetuta simmetricamente anche per l'altro nodo e non peggiora la complessità dell'algoritmo in quanto è lineare sul numero di nodi dei *banned set*.

Sulla base di questa euristica, che in pratica modifica i *banned set* durante l'applicazione degli assiomi di congruenza, possiamo riformulare l'algoritmo di Nelson-Oppen.

1. Controllare che nessun argomento dei predicati `atom` sia un termine `cons`, in accordo con l'assioma *(atom)* della teoria delle liste. Se ne viene rilevata la presenza restituire UNSAT.
2. Aggiungere ai *banned set* dei nodi che rappresentano gli argomenti del predicato `atom` tutti i termini `cons` (e viceversa) in modo da rispettare l'assioma precedente anche durante la propagazione degli effetti di nuove congruenze.
3. Per ogni nodo `cons` aggiungere al DAG due nuovi nodi `car` e `cdr` il cui archi puntano a quel nodo e poi effettuare la `merge` del `car` con il primo argomento del `cons` e del `cdr` con il secondo. In caso di conflitto durante la `merge`, ritornare UNSAT.
4. Effettuare la `merge` di ogni coppia di termini presente nell'insieme dei predicati di uguaglianza. In caso di conflitto ritornare UNSAT.
5. Ritornare SAT.

Da notare che in realtà l'implementazione dell'algoritmo non ritorna insoddisfacibile bensì la coppia di termini tra i quali si è manifestato il conflitto. Inoltre la percentuale che si può osservare durante l'esecuzione, consiste nella percentuale del numero di clausole di uguaglianza fino a quell'istante computate.

Poichè l'algoritmo opera sui nodi del DAG come su una foresta di insiemi disgiunti, è possibile implementare altre due euristiche nei metodi `find` ed `union` in modo da abbassarne ulteriormente il tempo di esecuzione.

- **Path Compression**: la compressione dei cammini, implementata nel metodo `find_h`, consiste nell'appiattire la stuttura dell'albero che identifica ciascuna classe di congruenza facendo puntare tutti i nodi direttamente al rappresentante; in questo modo le successive chiamate a tale funzione su quei nodi possono essere espletate in tempo costante. Per ottenere ciò si applica un metodo a doppio passaggio: durante il primo passaggio, il metodo risale ricorsivamente il cammino di ricerca per trovare la radice; durante il secondo passaggio, ridiscende il cammino di ricerca per aggiornare i nodi in modo che puntino direttamente al rappresenante.
- **Union by Rank**: l'unione per rango, implementata nel metodo `union_h`, consiste nell'unire due classi di congruenza in modo che il nuovo rappresentante sia quello della classe più grande delle due. Il vantaggio è che i nodi della classe di congruenza che cambia rappresentante avranno una distanza maggiore dalla radice e quindi l'operazione di `find` sarà più onerosa. Per implementare ciò, ogni nodo è provvisto di un campo `rank` che è un limite superiore al numero di archi tra quel nodo e una foglia discendente. Quando due classi di congruenza vengono unite, si trasforma il rappresentante di rango più alevato nel rappresentante di rango più basso, senza però modificare questo campo. Se invece i ranghi dei due rappresentanti sono uguali, se ne trasforma arbitrariamente uno nel padre dell'altro, incrementandone anche il rango.

3 Benchmark

In questa sezione viene riportata l'analisi dell prestazioni di tale algoritmo effettuata su un calcolatore con CPU `Intel Core 2 Duo P8600 2.4GHz`, sistema operativo `Ubuntu 10.10` e `JVM v1.6.0_24`.
Le formule sono state create tramite un generatore automatico, `formulaGenerator`, che fa un ingente uso di una funzione random. Inoltre per automatizzare i test è stato implementato uno script (`generatorTest.sh`) che crea e salva i risultati in un file denominato `testResult.cvs`.

Di seguito sono presentate due tabelle con alcuni dei dati raccolti sulle prestazioni dell'algoritmo: la prima riguarda i test effettuati su formule composte da clausole miste, mentre la seconda su formule contenenti solo clausole di uguaglianza, `-atom` e predicati positivi che sono quindi forzatamente soddisfacibili.

Clausole	Nodi	Archi	Risultato	Time (ms)	Time eur (ms)
64	415	1162	UNSAT	56	43
384	1417	3809	UNSAT	14841	149
512	1747	4926	UNSAT	8407	251
768	2959	8359	UNSAT	40988	422
1152	3581	10563	UNSAT	78023	328
1408	6514	22319	UNSAT	997942	924

Tabella 1: Prestazioni dell'algoritmo su formule con clausole miste

Clausole	Nodi	Archi	Risultato	Time (ms)	Time eur (ms)
64	427	1357	SAT	86	77
512	2018	6111	SAT	21660	1598
384	2381	8450	SAT	22064	2427
1152	4108	11633	SAT	153478	5441
768	4243	15041	SAT	124388	7663
1408	8582	31080	SAT	1184730	34355

Tabella 2: Prestazioni dell'algoritmo su formule soddisfacibili

3.1 Considerazioni

Nelle tabelle sopra riportate si è deciso di ordinare le righe a seconda del numero di archi nel grafo visto che è su questa grandezza che viene calcolata la complessità. Dalla seconda tabella si può notare che non sempre ad un aumento del numero di clausole corrisponde un aumento del numero di archi: infatti il numero di nodi e di archi dipende principalmente dalla quantità di sottotermini presenti nella formula e questa grandezza varia in maniera casuale per ogni formula generata.

Dalle tabelle sopra riportate si può facilmente notare quanto l'algoritmo con euristiche sia più performante dell'altro quando le formule cominciano ad essere consistenti. Questo perché l'algoritmo originale non deve compiere attività di preprocessing (e.g. scorrere ed aggiornare i vari *banned set*) per cui, quando il numero di archi è limitato, riescie ad essere più performante come nel caso del primo test della prima tabella.

Quando invece la grandezza del DAG aumenta, tramite le euristiche si ottengono migliori performance: nei test di formule con clausole miste, poiché il risultato è sempre UNSAT, la differenza tra le prestazioni è dovuta principalmente all'euristica che fa uso dei *banned set* per restituire insoddisfacibile non appena trova un conflitto. Nei test di formule soddisfacibili, invece, quest'euristica è inefficiente in quanto sicuramente non vi sono contraddizioni tra termini o predicati; la differenza di prestazioni rispetto all'algoritmo originale è, in questo caso, data dall'implementazione delle tecniche di unione per rango e compressione dei cammini dei metodi `union` e `find`.