

**Università degli Studi di Bari “Aldo Moro”**

**Dipartimento di Informatica**

Corso di laurea in Informatica Magistrale

Ingegneria della Conoscenza e Intelligenza nelle Macchine

Anno accademico 2016 / 2017

Sperimentazione su una guida genetica per l’algoritmo A\*

Insegnamento: Intelligenza Computazionale (6 CFU)

Docente: Dott. Corrado Mencar

Studente: Vito Cuccovillo – matricola: 634488

**Sommario**

[Introduzione 3](#_Toc495574538)

[**1.** **Progettazione** 4](#_Toc495574539)

[**1.1.** **Descrizione del sistema** 4](#_Toc495574540)

[**1.2.** **A\* con coda di priorità** 5](#_Toc495574541)

[**1.3.** **Funzione di distanza** 5](#_Toc495574542)

[**2.** **Implementazione** 6](#_Toc495574543)

[**2.1. Algoritmo genetico** 6](#_Toc495574544)

[**2.2. Implementazione del sistema** 7](#_Toc495574545)

[**3.** **Sperimentazione** 8](#_Toc495574546)

[**Conclusioni** 10](#_Toc495574547)

[**Riferimenti** 11](#_Toc495574548)

# Introduzione

Nel processo di individuazione di una soluzione ad un problema si possono impiegare differenti tecniche, ognuna delle quali può portare o meno ad una soluzione ottimale. In questa sperimentazione si analizza l’algoritmo A\*, che consente di individuare una soluzione ottima ad un determinato problema. Tale algoritmo, basandosi su una ricerca informata su un grafo, sceglie il percorso solutivo in base alla “bontà” degli stati in cui si trova. In altri termini preferisce esplorare per primi i percorsi più promettenti, che probabilmente consentiranno di giungere alla soluzione.

Un algoritmo genetico trae ispirazione dall’evoluzione e selezione biologica. Il loro principio consiste nel partire da alcune soluzioni di partenza, ricombinarle, perturbarle e selezionarne quelle più promettenti in modo da generarne sempre di migliori.

La sperimentazione corrente fonde questi due approcci in modo tale da fornire una linea guida all’algoritmo A\* laddove il suo criterio di preferenza non riesca a discriminare uno stato migliore da espandere rispetto ad altri. Dunque la soluzione di un algoritmo genetico fungerà da filo conduttore per sostenere A\* cercando di indirizzarlo verso un percorso solutivo laddove ci fossero situazioni di parità.

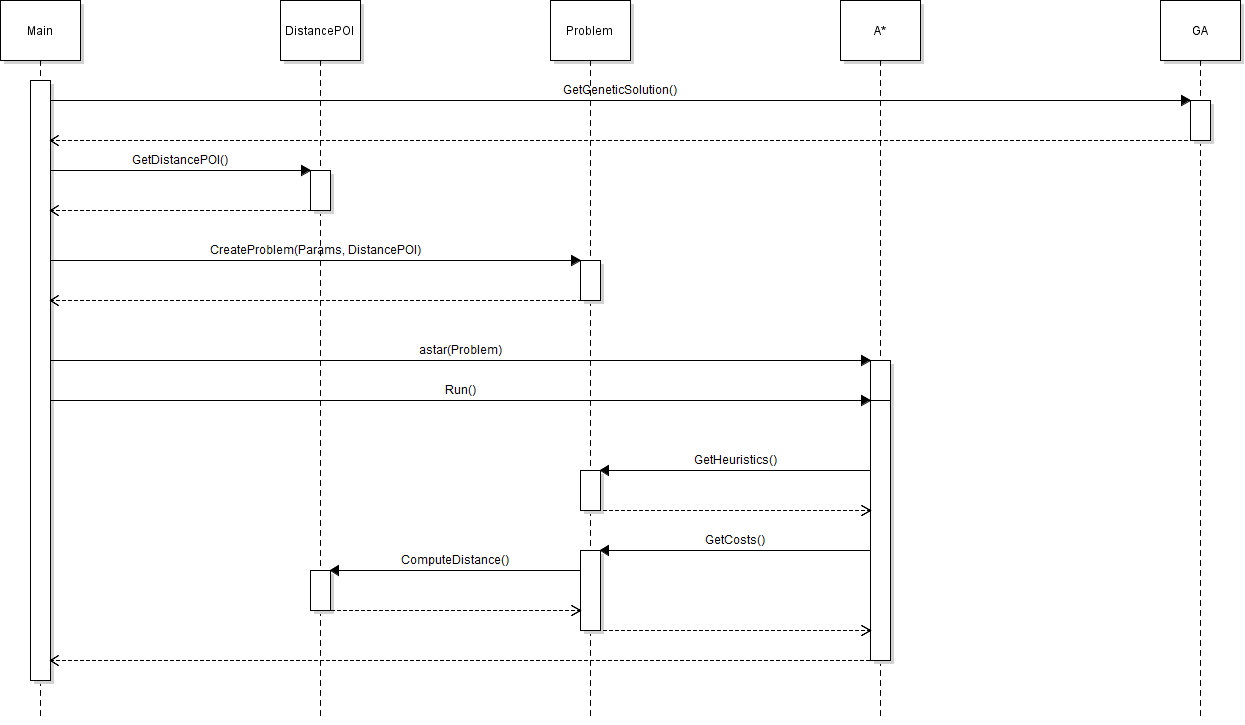
L’obiettivo finale di questo esperimento è quello di verificare se l’efficienza generale di A\* può essere statisticamente migliorata con l’intervento di un algoritmo genetico.

1. **Progettazione**

In questo paragrafo si fornisce una descrizione concettuale degli elementi che compongono il sistema, unita ad una spiegazione di come l’algoritmo A\* funziona a seguito dell’adozione di una coda con priorità con diversi livelli.

* 1. **Descrizione del sistema**

Da un punto di vista concettuale il sistema è organizzato in diversi settori, ognuno dei quali ingloba determinate caratteristiche. È necessario dunque fornire una descrizione del modello adottato, al fine di poterne comprendere i flussi.



**Figure 1 Diagramma di sequenza del flusso eseguito dal sistema**

Nel precedente diagramma di sequenza viene fornita una visione complessiva dell’esecuzione tipica. In particolare il modulo “Main” invoca un algoritmo genetico il quale produce una soluzione molto probabilmente subottimale. Tale soluzione viene incapsulata in un modulo “DistancePOI”, il cui compito è quello di misurare la distanza fra la soluzione genetica fornita e una qualsiasi altra soluzione. In questo caso il concetto di distanza è abbastanza generico, in quanto è possibile fornire a questo modulo una qualsiasi misura di distanza, che però possa riflettere l’effettiva distanza fra le due soluzioni. In seguito, il modulo “Problem” contiene la logica e le specifiche del particolare problema di cui si vuole trovare una soluzione. “Problem” possiede un riferimento al modulo “DistancePOI” che gli consentirà di fornire, oltre alle misure di costo tipiche del problema, anche un costo dato dalla distanza della soluzione corrente da quella genetica generata in precedenza.

A questo punto è possibile avviare la ricerca vera e propria invocando il modulo “A\*” il quale partendo da una situazione iniziale, esplorare il grafo degli stati del problema. Per far ciò esso necessita del calcolo della cosiddetta funzione di valutazione euristica dove *g* rappresenta il costo sostenuto per transitare dallo stato iniziale allo stato corrente *x*, *h* invece il costo stimato per raggiungere uno stato obiettivo partendo dallo stato corrente *x*.

Il modulo “A\*” dunque per poter calcolare il valore effettivo della funzione di valutazione euristica richiede al modulo “Problem” gli effettivi costi. “Problem” fornisce ad “A\*” una serie di valori euristici corrispondenti ad altrettante funzioni euristiche tipiche del problema, ed una tupla di valori che rappresentano costi, uno dei quali è calcolato dal modulo “DistancePOI”. In sintesi, il modulo “A\*” potrà servirsi di informazione non soltanto proveniente dal problema in sé, ma avere un’ulteriore stima del costo data dalla distanza fra la soluzione genetica e lo stato corrente dell’esplorazione.

* 1. **A\* con coda di priorità**

La peculiarità di questo approccio, rispetto ad A\* tradizionale, consiste nel fatto che tale algoritmo può godere di diversi livelli di priorità tramite i quali poter indirizzare il processo di visita degli stati del problema. Fornendo ad A\* non più un singolo valore con cui scegliere la prossima configurazione da generare ma una tupla di valori, è possibile utilizzare ulteriori criteri in caso di situazioni di assoluta parità dei valori della funzione *f\** al livello cosiddetto n-esimo, procedendo alla verifica del livello (n+1)-esimo. In sintesi, l’utilizzo di *n* livelli di priorità consente di ridurre al minimo la scelta arbitraria di A\* nel caso di situazioni di parità. Nella tabella seguente è mostrato un esempio del criterio di preferenza di A\* tradizionale, confrontato con A\* dotato di coda con priorità.

|  |  |
| --- | --- |
| **OPEN A\* tradizionale** | **OPEN A\* coda priorità** |
| A – [2] | A – [2,4] |
| B – [3] | B – [2,5] |
| AC – [ |  |
|  |  |

* 1. **Funzione di distanza**

Un aspetto di fondamentale importanza è appunto la funzione di distanza che verrà appunto utilizzata per fornire una discriminante nei livelli successivi. In questo progetto è stata introdotta una funzione che implementa la *distanza di Levenshtein*. Questa distanza misura la differenza fra due stringhe fornite in input in base al numero di:

* sostituzione di caratteri
* aggiunta di caratteri
* rimozione di caratteri

da effettuare sulla seconda stringa al fine di trasformarla nella prima. Nella fattispecie, questa funzione di distanza ha complessità quadratica, dunque è opportuno scegliere le funzioni sia in base alla loro complessità, sia in base al loro grado di rappresentare l’effettiva distanza che intercorre fra le soluzioni.

1. **Implementazione**

Il sistema descritto in progettazione è stato implementato nel linguaggio Python versione 3.6.2 ed è suddiviso in diversi package in base alle funzionalità.

## **2.1. Algoritmo genetico**

Come già indicato, il sistema si serve di un algoritmo genetico come guida all’esplorazione. Fra le diverse librerie Python esistenti, è stata utilizzata la libreria open source “Inspyred” nella versione 1.0.1.

Al fine di poter avviare l’algoritmo genetico con suddetta libreria è necessario fornire esplicitamente le funzioni di generazione e valutazione degli individui. Queste funzioni, una volta avviato il tutto, verranno richiamate in automatico dal framework di inspyred.

Aspetto fondamentale è quello relativo alla rappresentazione degli individui, in quanto è necessaria una rappresentazione a lunghezza fissa affinché gli operatori genetici di crossover e mutation integrati nella libreria possano funzionare. Nel caso fosse impossibile avere una rappresentazione degli individui a lunghezza fissa, è possibile fornire delle funzioni “custom” di incrocio e mutazione che siano in grado di gestire questa variabilità.

Prima dell’avvio, è possibile impostare delle ulteriori caratteristiche:

* *observer:* consente di mostrare un plot, una lista dettagliata oppure statistiche relative alla computazione genetica in corso. L’utente può scegliere il tipo di output che preferisce
* *terminator:* indica quando l’evoluzione deve terminare, impostando ad esempio dei vincoli di tempo, di fitness media, di non ulteriore miglioramento etc.

La vera e propria computazione genetica si avvia invocando il metodo *evolve()* che accetta in input:

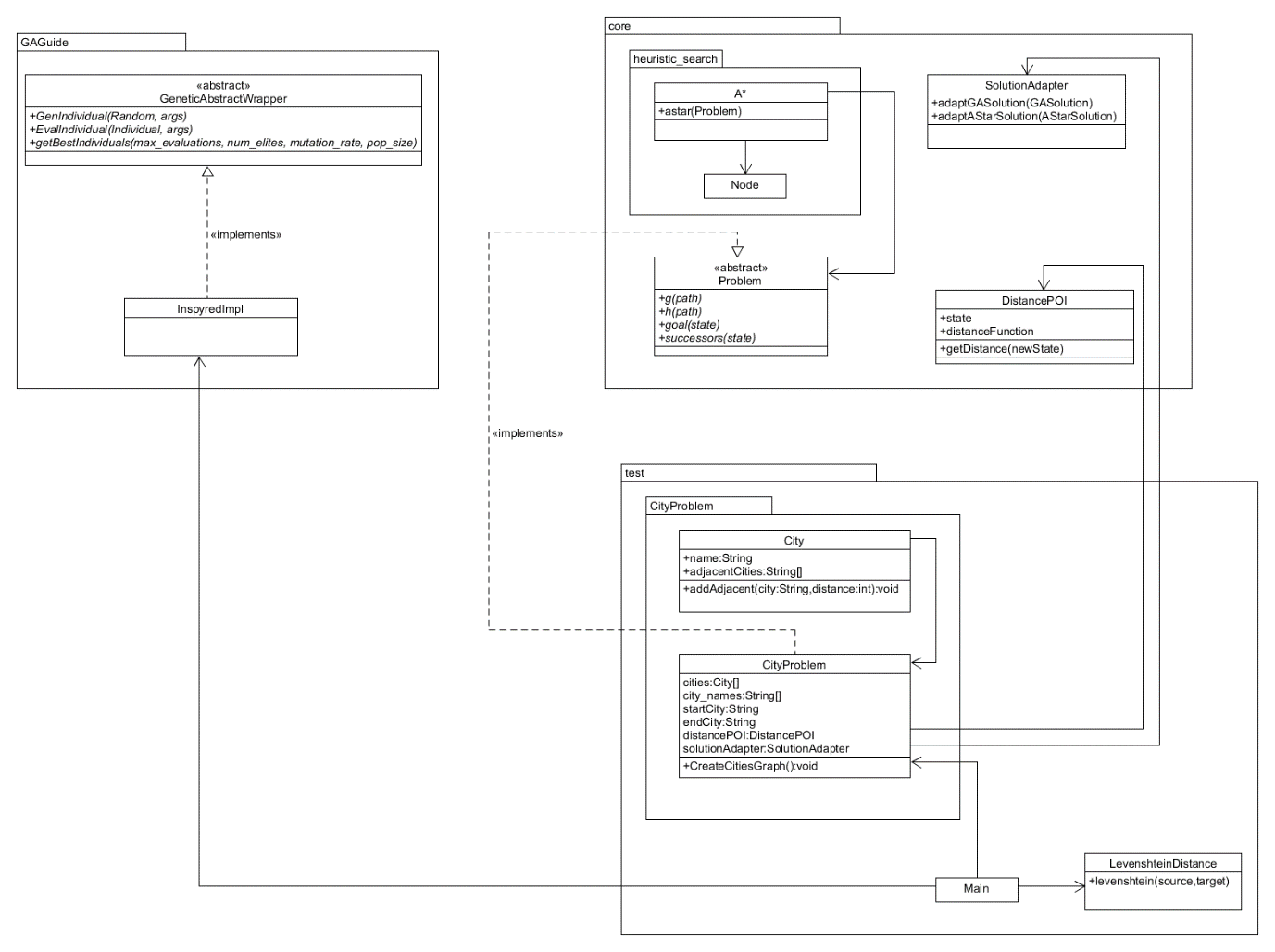
* *generator*: funzione di generazione degli individui
* *evaluator:* funzione di valutazione degli individui (funzione di fitness)
* *max\_evaluations:* numero massimo di valutazioni che l’algoritmo deve effettuare
* *num\_elites:* numero di migliori individui della popolazione che deve sopravvivere
* *mutation\_rate:* la probabilità con la quale si verificano mutazioni, di default 0.1
* *pop\_size:* il numero di individui della popolazione
* *crossover\_rate:* tasso con il quale è effettuato l’incrocio, di default 1
* *num\_crossover\_points:* quanti punti di incrocio utilizzare, di default 1

Al termine della computazione, l’algoritmo restituirà un insieme di individui che corrispondono a soluzioni generate, ordinate in base al valore di *fitness* conseguito. Per gli scopi del progetto corrente si sceglie soltanto la prima soluzione, che sarebbe la migliore generata.

## **2.2. Implementazione del sistema**

Al fine di garantire una buona modularizzazione e separazione delle responsabilità, è stato implementato un wrapper per gestire l’algoritmo genetico. Tale wrapper fornisce sostanzialmente un’interfaccia da implementare in base alla specifica libreria di algoritmi genetici utilizzata nel sistema.

L’intero sistema è rappresentato tramite il seguente diagramma delle classi UML:



**Figure 2 Diagramma delle classi impiegate nel sistema**

Nel precedente diagramma è possibile individuare diversi package di cui ***GAGuide*** contenente un wrapper per algoritmi genetici e la relativa implementazione con la libreria *inspyred*.

Il package ***Core***contiene la classe astratta *Problem* che dovrà essere implementata da un problema concreto. Si trovano anche *SolutionAdapter* che si occupa di uniformare le soluzioni di algoritmo genetico e di A\* in modo da renderle confrontabili ed infine *DistancePOI*, inizializzato con la soluzione dell’algoritmo genetico, il quale calcola la distanza fra tale soluzione ed un’altra qualsiasi soluzione. Nella fattispecie la distanza è calcolata tramite una funzione che deve essere fornita dall’utente. Nel sotto-package *heuristic\_search* si trova l’implementazione dell’algoritmo A\* che accetta in input un oggetto di tipo *Problem*.

Il package **Test** contiene la realizzazione del problema utilizzato come esempio, CityProblem, un’implementazione della funzione di distanza di Levenshtein ed infine un modulo Main che funge da entry point per avviare il processo.

1. **Sperimentazione**

L’obiettivo di questa sperimentazione è verificare che l’utilizzo di una guida genetica possa in qualche modo rendere più efficiente l’algoritmo A\*. Tale efficienza è misurata nel numero di nodi espansi prima di giungere ad uno stato obiettivo.

Per svolgere le relative prove del sistema è stato implementato, come già accennato, il cosiddetto CityProblem. Tale problema consiste nel trovare il percorso migliore, costituito da una sequenza di città da attraversare, per giungere ad una città obiettivo partendo da una città indicata. Si dispone a questo scopo di una struttura a grafo utile a rappresentare la topologia delle città, i relativi collegamenti e le distanze che intercorrono fra l’una e l’altra. La rappresentazione dei percorsi, al fine di essere compatibile con i requisiti dell’algoritmo genetico, è di questo tipo:

**[0,1,0,3,2,0,0,5,0,4]**

Questa sequenza di interi rappresenta un individuo generato da un algoritmo genetico. Ogni individuo ha una lunghezza fissa pari al numero delle città presenti nel grafo. Ad ogni posizione nell’individuo corrisponde una precisa città, di cui si indica il non attraversamento con lo zero oppure un intero che rappresenta l’ordine con le quali le città del percorso sono attraversate.

Le città che costituiscono i nodi del grafo sono rappresentate in un’altra sequenza in cui è rilevante la posizione ai fini delle soluzioni genetiche:

**[A,B,C,D,E,F,G,H,I,J]**

Dunque l’individuo rappresentato precedentemente indica un percorso di questo tipo: **B->E->D->J->H**.

Tali rappresentazioni verranno poi adattate al fine di essere confrontabili con le soluzioni prodotte dall’algoritmo A\*.

**Capire se la distanza di Levenshtein può guidare una BFS (breadth first search) in quanto ho impostato euristica = 1. Capire se levenshtein è una buona guida**

**Conclusioni**

**Riferimenti**