

PERCORSI MINIMI PSICOLOGICI: UN' IMPLEMENTAZIONE CON GIS AD INTELLIGENZA ARTIFICIALE

Vincenzo BARRILE , Francesco COTRONEO

Università degli Studi Mediterranea di Reggio Calabria - DIMET,
barrile@ing.unirc.it, francesco.cotroneo@unirc.it

Riassunto

Alla conferenza ASITA 2004 gli autori di questo stesso articolo, presentarono un lavoro dal titolo "ALGORITMI AVANZATI SUI GEODATABASE", nel quale veniva definita e affrontata per la prima volta la problematica dei Percorsi Minimi Psicologici, ovvero di quei percorsi che non indicano il raggiungimento di un determinato target a partire dall'ottimizzazione di parametri come il tempo o le distanze euclidee (come avviene in un classico problema di logistica), ma percorsi il cui contesto di instradamento appaia ergonomico alla psiche umana. Si vuole, in altri termini, implementare uno strumento capace di "disegnare" sulla mappa virtuale di un applicativo GIS, un cammino che possa essere eseguito dall'utente del sistema informativo senza che esso implichi un disagio psicologico per l'utente medesimo. E' da sottolineare, inoltre, come esistano molti ambiti nei quali vengono utilizzati GIS con funzionalità di routing e per i quali è strettamente necessario disporre di un simile strumento decisionale.

La soluzione di tale problema è stata trovata nell'implementazione di un software che utilizza tecniche di Intelligenza Artificiale. Nel caso in esame, l'algoritmo utilizzato per creare l'applicativo fa uso di particolari reti neurali, le quali usano per la fase di addestramento un training set che da una parte è popolato con target forniti da una popolazione di fruitori, scelti in modo casuale del sistema, dall'altra dai domini di contorno agli instradamenti che vengono resi disponibili dal GIS.

L'applicativo di cui prima, è in grado di tracciare sui layers di un GIS percorsi validi, "Psicologicamente Ottimizzati", ovvero più vicini possibili ai percorsi che avrebbero definito gli individui, le cui indicazioni sono state utilizzate nella fase di training. Ma cosa ancora più importante, e che questo risultato lo si ha anche con una cartografia completamente diversa da quella utilizzata per l'apprendimento, ma senza dover ripetere nuovamente tale operazione.

Abstract

During 2004 ASITA symposium we presented a work titled "Advanced geodatabase algorithms", in which we first came to define and discuss problems involving Minimal Psychological Paths, that's to say paths not addressing the achievement of any target starting from optimization of some parameters like time or euclidean distances (as it's made in the case of a well-known logistic problem), but paths whose contest of routing is *ergonomic* to human psyche. By other words, we want to implement an instrument able to "draw" the virtual map of a GIS application with a path users may run without any psychological unease for themselves. Moreover, we have to remark there exist many areas of expertise in which we use GISs with routing capabilities making necessary the employment of such a decisional instrument. We found a solution to the problem by implementing a software which involves AI techniques. In the case of interest, our algorithm exploits specific neural networks, trained by an input-data set made available through a sample population of users whose choice is randomly performed by both system decisional rules and suitable routing contour conditions accessible from GIS. The application we are speaking about can draw GIS layers having valid and "psychological optimized" paths, that's to say paths being as near as possible to the real ones that would have been run by users making their own suggestions available for training. More important to say, the same result can be obtained also referring to a cartography quite different from the one employed for training, and this without any further need of repeating such an operation.

Introduzione

Negli ultimi anni si è assistito ad una evoluzione innovativa molto rapida dei sistemi GIS che ha riguardato soprattutto le tecniche implementative degli stessi ed i paradigmi operativi attraverso i quali strutturare le informazioni geografiche all'interno delle banche dati digitali con l'obiettivo di conservare inalterate le relazioni logico – funzionali che ricorrono tra gli elementi fisici al momento della fasi di rilievo e di popolazione dati, siano essi spaziali che alfanumerici. Tutto ciò ha portato ad avere dei vantaggi enormi negli ambiti classici di utilizzo dei GIS facenti capo alle problematiche del controllo del territorio (forestazione, agricoltura, urbanistica ecc...), ma allo stesso tempo tali innovazioni ne hanno consentito l'utilizzo anche in ambiti non fortemente caratterizzati da contenuto spaziale. Tuttavia gli strumenti di queryng presenti nei GIS non riescono a far fronte a quei casi, riguardanti appunto questi nuovi ambiti, in cui occorre ottenere le relazioni esistenti tra dato cartografico e variabili di ambiente non campionabili in forma chiusa (dati su variazioni ambientali, variabili di tipo socio-economiche, aree tematiche a contorno sfumato ecc...). Ciò è dovuto principalmente alla impossibilità di scomporre tutte le interrogazioni utili in termini di un formalismo riconducibile a logiche deterministico – sequenziale e quindi all'algebra binaria; tale limite accomuna ad oggi la maggior parte dei Software Gis presenti sul mercato. Tra le altre, una tipologia di query intrinsecamente irrisolvibile attraverso una macchina di Turing riguarda il problema dei "Percorsi minimi Psicologici", caso in cui il sistema GIS è chiamato a restituire, a partire da una data mappa e da un punto origine ed una destinazione e dall'utente del sistema stesso, un percorso che oltre a minimizzare la distanza di percorrenza (o altri parametri logistici) appaia "confortevole" agli occhi dell'utente che lo deve percorrere, in funzione di molti parametri difficilmente definibili e quantizzabili (come le reazioni istintuali a certe situazioni) relativi al contesto in cui è collegato il percorso, o come altre proprietà caratterizzanti la psiche umana. Da quanto detto emerge che una qualsivoglia procedura che si utilizzi per ottenere quel tipo di percorso, deve operare su un dominio cui fa parte l'utente stesso che formula la query; questa caratteristica del problema, abbastanza curiosa, sta ad indicare che la soluzione che verrà proposta ha un alto indice di soggettività, il percorso ottenuto può essere accettato non da tutti gli utenti, in quanto ognuno di essi avrà una diversa sensibilità dovuta a fattori come la cultura, il carattere, le esperienze personali.

Alla definizione della problematica dei Percorsi Minimi Psicologici da parte degli autori di questo articolo, seguì, a cura degli stessi, una prima implementazione di un' algoritmo risolvente il problema (ASITA 2004). L'algoritmo, che poteva essere utilizzato in molti ambiti applicativi vista la qualità dei risultati proposti, soffriva di due limitazioni importanti: il risultato era non abbastanza "oggettivo", in quanto dipendente fortemente dall'amministratore del GIS ospitante le mappe cartografiche, ed inoltre si adeguava poco ai cambiamenti (soprattutto quelli nel breve periodo) che potevano interessare quest' ultime.

Di seguito verrà presentato un approccio che è la parte principale di un modulo software per GIS, col quale sarà possibile ottenere buoni risultati che raramente sono afflitti dalle limitazioni cui sopra.

Analisi e rappresentazione del problema

Per poter meglio comprendere le difficoltà inerenti alla definizione di un algoritmo che ha come obiettivo, appunto, quello di generare una soluzione che soddisfi tutte le necessità cui prima si faceva riferimento, ed allo stesso tempo per giustificare le metodologie rappresentative adottate per i dati elaborati, si farà di qui in avanti riferimento alla figura 1. In questa figura sono rappresentati tre situazioni, tre casi, costruiti "ad arte" per sottolineare alcuni aspetti di maggior rilievo descrittivo.

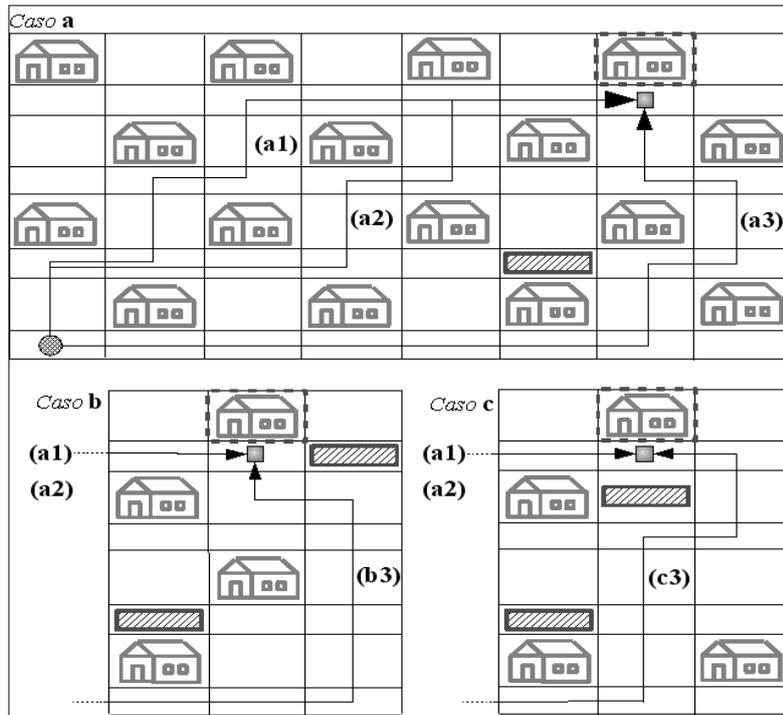


Figura 1

Nel primo caso (Caso *a*) si ha una mappa i cui elementi sono esclusivamente delle abitazioni (disposte in maniera ricorsiva) e delle costruzioni di vario genere accomunate dalla proprietà di essere dislocate in una data zona per un breve lasso temporale, quest'ultime sono rappresentate con dei rettangoli retinati. I casi *b* e *c* rappresentano mappe identiche a quelle del caso *a* modificate esclusivamente nelle ultime tre colonne e visualizzate quest'ultime in figura 1. Nella stessa figura sono descritti tre esempi di possibili percorsi per raggiungere l'abitazione con contorno tratteggiato (destinazione) a partire dalla cella avente un cerchietto retinato (origine); se si assume che in termini di costo di percorrenza ogni cella è identica (con costo unitario), il percorso *a1* e *a2* hanno costo pari a 12, mentre quello di *a3* è 14. In termini di ottimizzazione delle risorse in senso classico, un risolutore di percorsi dovrebbe suggerire al suo utente utilizzatore, uno dei due primi percorsi. Ma in termini di Minimi Percorsi Psicologici può essere conveniente il percorso *a3* qualora l'utente stesso sia restio (a livello cosciente o subcosciente) ad effettuare variazioni repentini di traiettoria in un dato contesto, più che in una data mappa. Questo ultimo concetto è utile venga ribadito in quanto rappresenta il cuore del problema: si vuole evidenziare infatti che lo stesso utente davanti a due mappe identiche così da poter essere perfettamente sovrapponibili e con gli stessi punti origine e destinazione, potrebbe scegliere percorsi completamente diversi nelle due situazioni, questo proprio perché oltre al contenuto spaziale ci possono essere in un caso o nell'altro elementi condizionanti, dipendenti dalla realtà geografica che le due mappe rappresentano. Questa sua reticenza ai percorsi a zig-zag nel contesto in esame, potrebbe fargli scegliere nel caso *b*, addirittura il percorso *b3* che in dati punti ha una distanza in linea d'area dal punto origine, maggiore di quella degli altri percorsi in qualsiasi loro punto. Potremmo comunque osservare, modificando la mappa (ma non il contesto rappresentato), come nel caso *c*, che lo stesso utente adotti quei percorsi prima scartati poiché sono intervenuti cambiamenti che interessano il suo lato comportamentale e caratteriale; si può immaginare, infatti che il percorso *c3* venga rifiutato poiché lo costringe ad adottare una direzione che comporta un momentaneo allontanamento, in linea d'area, dalla destinazione in un tratto prossimo ormai a questa.

Anche se il problema è sicuramente non riconducibile facilmente ad altri di natura logistica, si ha comunque che il peso complessivo del percorso debba essere in qualche modo sempre considerato dall'algoritmo risolutore, poiché basandosi solo su i criteri prima visti ci sarebbero molti casi in cui l'inefficienza in "senso classico" della soluzione sarebbe così evidente da costituire comunque un disagio psicologico per l'utente. Proprio per questa ragione la struttura dati sulla quale l'algoritmo implementato si appoggia è quello dei grafi non orientati, i quali nell'approccio in

questione vengono rappresentati attraverso l'utilizzo di matrici quadrate che hanno come colonne e come righe i vertici del grafo, e i cui valori possono essere 1 o 0, a seconda se in una data cella il vertice della relativa riga è connesso o meno a quello indicizzante la colonna. Quella della matrice non è una scelta condizionata da ragioni prestazionali al momento della computazione dell'algoritmo, o della quantità di memoria necessari a quest' ultimo, ma è necessaria all'approccio scelto che è di tipo funzionale.

L'algoritmo risolutivo

Come più volte accennato la soluzione a tale problema non può intrinsecamente essere di tipo logico deterministica, ed infatti uno dei passi svolti dall'algoritmo in questione coinvolge funzioni costruite secondo i dettami dell'intelligenza artificiale. In particolar modo si è fatto uso delle Reti Neurali Supervisionate (che tra l'altro sono state modellate a partire dal comportamento dei neuroni biologici). Brevemente il neurone biologico viene simulato attraverso una delle funzioni standard per questo utilizzo (come la sigmoide o la tangente iperbolica), l' uscita di questa funzione con valori appartenenti al campo dei reali, diviene l'ingresso di altri neuroni, previa modulazione attraverso dei pesi anch'essi a valori reali: si viene ad avere così una rete connessa direzionalmente ed a strati, che ha come input N ingressi (ossia il numero dei neuroni del primo strato) ed M uscite (il numero dei neuroni dell'ultimo strato). L'utilizzo delle Reti Neurali Supervisionate concerne lo studio proprio di fenomeni dei quale non si conosce una descrizione sintetica in modo tale da poterne riprodurre i comportamenti, ma si possono osservare le risposte relative a dati imput. In altre parole si può considerare tale fenomeno come una funzione $f(.)$ che associa a determinati N valori di ingresso altri M in uscita, ma di cui non si conosce la descrizione in forma chiusa. Al momento della sua costruzione, con un determinato numero di neuroni, scegliendo una funzione modellante per essi e definendone la topologia per le connessioni, una rete neurale altro non è che una funzione indefinita. A questo punto la Rete Neurale può essere specializzata al fine di simulare il fenomeno di cui prima. Per far questo occorre che la rete sia sottoposta alla cosiddetta fase di "apprendimento" in analogia con quanto avviene per le reti biologiche; tale fase in pratica si risolve nel modificare adeguatamente i pesi della rete stessa. Comunque bisogna ricordare che la fase di apprendimento, diversamente da quanto accade nella mente umana, è nettamente distinta dall' altra fase di esecuzione (al momento dell'utilizzo) ed, inoltre, la precede. La modifica dei pesi avviene minimizzando l'errore Φ che intercorre tra le uscite della funzione $f(.)$ (quella del modello reale) e quelle della Rete Neurale (uscite relative agli stessi ingressi). I valori che si usano nella fase di apprendimento vengono detti *training-set* ed evidentemente vengono rilevati osservando il fenomeno; la procedura di apprendimento si interrompe quando Φ è minore di un certo ξ_{\max} predefinito a priori e comunque molto basso. A questo punto la Rete Neurale è una funzione $g(.)$ che anche se diversa dalla $f(.)$ ci si approssima molto in ogni punto anche in quelli non usati nel training set e può essere utilizzata al suo posto; in altre parole la rete è in grado di estrapolare la legge che governa il fenomeno. Le procedure di minimizzazione dell'errore, comunque, non sempre hanno buon esito, infatti se anche dopo moltissime iterazioni di minimizzazione (detti epoche d'apprendimento) non si riesce a scendere sotto ξ_{\max} si dice che la rete non "converge". Questo può dipendere da vari fattori legati alla costruzione della rete, ma soprattutto dal fatto che il fenomeno è caotico, ovvero non è possibile estrapolare nessuna legge poiché essa tende ad essere intrinsecamente probabilistica.

Utilizzando le Reti Neurali, l'algoritmo preposto alla generazione dei Percorsi Minimi Psicologici, necessita anch'esso di una fase d'apprendimento, articolata nei seguenti passi:

1. Scelta della cartografia sulla quale dovrà operare poi l'algoritmo.
2. Rappresentazione della stessa tramite grafo e dunque, nel nostro caso, come matrice archi e nodi.
3. Scelta di diverse unità campione, ovvero di una serie di possibili utenti col quale si fa un vero e proprio rilievo statistico delle preferenze.
4. Scelta di punti origine e destinazione sulla cartografia, adeguati sia per numero che per disposizione (devono essere distribuiti in maniera uniforme).

5. Generazione per ogni coppia di punti origine-destinazione di un insieme di percorsi contenenti oltre al percorso minimo, generato con un algoritmo classico della teoria dei grafi, tutti gli altri percorsi il cui peso totale non superi del 10% quello del percorso ottimo e/o comunque il numero di tale percorsi sia inferiore ad una quantità prefissata .
6. Sottomissione alle unità campione i vari gruppi di percorsi e rappresentazione dei percorsi scelti con l'ausilio delle matrici.
7. Addestramento di una rete neurale, opportunamente modellata, con il training-set generato al punto 6

In relazione al punto 6 si osservi che, differentemente da quanto si potrebbe pensare, per ogni utente e per ogni coppia di punti origine-destinazione non si sono ottenuti dei percorsi di training. Tale scelta progettuale è difatti un grave errore concettuale; infatti nel caso in cui anche due sole unità campione avessero suggerito percorsi diversi, la Rete Neurale, in fase di apprendimento, cercherebbe di modellare quella funzione che dato quel particolare ingresso (coppia di punti) è in grado di fornire due soluzioni (i due percorsi rilevati statisticamente), evidentemente ciò non sarebbe stato possibile in quanto viola la definizione stessa di funzione (si avrebbe una conseguente non convergenza della Rete Neurale nella fase di apprendimento). Questo problema si è risolto utilizzando quindi un solo utente campione per la generazione del training-set, in particolare un utente campione “virtuale” generato a partire dagli utenti “Reali”, nelle modalità riportate nella figura che segue:

	$I_{1,7}$	$I_{8,7}$	$I_{2,9}$	$I_{6,2}$
U_1	1	7	3	1
U_2	1	7	5	2
U_3	1	7	3	1
U_4	2	6	3	9
U_5	2	4	5	6
U_6	3	4	4	9
U_v	1	7	3	1

Figura 2

In pratica dato un insieme di percorsi $I_{i,j}$ relativi alla coppia origine-destinazione i -esima e j -esima, gli utenti U_i ne sceglieranno uno tra questi (nel caso in figura codificato numericamente) in funzione delle loro preferenze, e conseguentemente sarà assegnato all'utente virtuale la scelta maggiormente indicata (in figura nei casi relativi agli insiemi $I_{1,7}$, $I_{8,7}$, $I_{2,9}$) oppure nei casi in cui vi sono più percorsi egualmente indicati (caso $I_{6,2}$) sarà considerata preferibile l'indicazione dell'utente che ha condiviso le altre scelte fatte con il numero di utenti più grande. Il criterio adottato per la definizione dell'utente campione virtuale può sembrare non abbastanza sofisticato per sintetizzare al meglio la grande varietà di situazioni presenti (ad esempio non si indaga sull'esistenza di utenti che abbiano sempre condiviso una scelta con altri), tuttavia in caso di “grande varietà”, ossia qualora gli utenti non siano in gran numero concordi per ogni insieme di percorsi proposti, la Rete Neurale è destinata a non convergere durante la fase di apprendimento in quanto le scelte operate dall'utente virtuale non sarebbero inquadrabili in una logica conforme; in termini matematici ci si troverebbe a voler modellare in maniera predittiva una funzione ad alto contenuto frequenziale e dunque caotica.

Utilizzo dell'algoritmo risolutore ad Intelligenza Artificiale

E' interessante osservare che, da quanto detto relativamente alle proprietà delle Reti Neurali, e delle osservazioni fatte nel secondo paragrafo, l'operazione di apprendimento viene fatta una sola volta e il modello ricavato può essere applicato ovviamente per trovare i percorsi non utilizzati in fase di

training, inoltre, come confermato dalle verifiche sperimentali, svolte sia su mappe fittizie come quella di figura 1, che in situazioni reali, si sono ottenuti ottimi risultati anche variando la cartografia stessa. Ciò che risulta importante per avere dei risultati soddisfacenti, ovvero dei risultati di percorso che sarebbero stati scelti con ogni probabilità dalla maggior parte degli utenti campione nella fase di training set, è che bisogna tener presente che il contesto ambientale della cartografia, più che la sua geometria, che rappresenta il dominio di utilizzo dell'algoritmo, deve essere simile al contesto utilizzato nella fase di addestramento (utilizzando una Rete Neurale addestrata con una mappa catastale di una cittadina di campagna, per trovare Percorsi Minimi Psicologici in una grande città, si avranno dei risultati evidentemente inadeguati, anche per aree che si assomigliano spazialmente).

Il software sviluppato per la determinazione di tali percorsi, proprio per tali motivi, può utilizzare più reti neurali nell'ambito di una stessa mappa (cartografia) per specializzare il più possibile il modello che esso rappresenta.

Più in dettaglio la fase di esecuzione del software prevede i seguenti step:

1. Acquisizione dall'utente del punto origine ed del punto destinazione.
2. Calcolo di un insieme di percorsi contenenti ottimi in senso classico, come avveniva nella fase di addestramento al momento della creazione del training-set.
3. Scelta della Rete Neurale più adeguata tra quelle disponibili
4. Passaggio alla funzione $g(.)$ precedentemente selezionata delle matrici dei percorsi prima individuati.
5. Individuazione del percorso cercato tra quelli d'ingresso, attraverso il confronto tra questi e la matrice in uscita da $g(.)$; verrà scelto il percorso con distanza di Hamming minima; questo passaggio è obbligatorio perché $g(.)$
6. non ritorna un percorso identico ad uno di quelli d'ingresso, ma uno che lo approssima per i motivi visti a proposito delle proprietà funzionali delle Reti Neurali.

Il software sviluppato è stato testato in ambiente GIS con ottimi risultati fintanto che si sono rispettate le modalità di utilizzo viste prima in quanto tali risultati rispettavano le scelte delle unità campione utilizzate. In particolare il software è stato caricato come modulo dell'applicativo GIS ad oggetti Kosmos. Sviluppato dagli stessi autori, Kosmos permette di utilizzare degli operatori non standard, (come overlay union ecc...) ma definiti dall'utente per la creazione dinamica di nuove Features e layer. Ciò permette di utilizzare il modulo risolutore in maniera altamente integrata nel GIS, e di conseguenza si ha la possibilità di creare delle query composte da quelle standard deterministiche con quelle dei Percorsi Minimi Psicologici (si potrebbe fare una query di buffering su un percorso minimo psicologico tra due punti).

Conclusioni

La possibilità di integrare in ambienti GIS uno strumento capace di tracciare dei percorsi psicologicamente accettabili, significa disporre di strumenti per il controllo del territorio e di supporto alla decisione, in grado di offrire una panoramica più dettagliata dei contesti legati a parametri difficilmente sintetizzabili come la qualità della vita, oltre che sugli aspetti aperti ai classici approcci analitici.

Bibliografia

Bondy J.A. e Chvátal V. (1976), "A Method in Graph Theory" *Discrete Mathematics* 15: 111-145.
Ingerman P.Z. (1976), "Algorithm 141: Path Matrix," *Communication of the ACM* 5, 556.
Jhonson, Donald B. (1977), "Efficient Algorithms for Shortest Paths in Sparse Network", *Journal of the ACM* 24, 55.