

Referee Report

Titolo del Paper

Fast Shortest-path Distance Queries on Road Networks by Pruned Highway Labeling

Autori del Paper

Takuya Akiba, Yoichi Iwata, Ken-ichi Kawarabayashi, Yuki Kawata

Sommario

L'articolo proposto tratta di un metodo di labeling innovativo per il calcolo di percorsi minimi e query di distanza tra due nodi distinti di grafi di road networks. Gli autori hanno utilizzato risultati dalla letteratura trattante l'argomento, in particolare: le analisi teoriche riportate da Thorup in *Compact oracles for reachability and approximate distances in planar digraphs* da cui ricavano le basi sulla decomposizione del grafo in input attraverso percorsi minimi, necessario per memorizzare la distanza dai vertici ai percorsi più brevi (shortest path); l'idea di *highway*, cruciale nel loro lavoro e, infine, al concetto di *pruned labeling*, proposto da loro in un articolo precedente. Tutto ciò porta all'introduzione di un framework per il labeling del grafo, chiamato dagli autori *Highway-based Labeling Framework*, e di una tecnica per fare un pre-processing veloce, chiamata *Pruned Highway Labeling Framework*. Attraverso queste proposte gli autori hanno raggiunto buoni risultati sia in termini di velocità di risposta alle query che alla velocità del pre-processing, vero e proprio miglioramento introdotto in questo articolo. Come ulteriore nota positiva, è messo in risalto che il framework proposto ha un'occupazione di spazio che è paragonabile agli altri modelli se non migliore.

Originalità

Con questo articolo gli autori non hanno proposto niente di nuovo rispetto agli altri algoritmi di labeling dello stato dell'arte, bensì sono riusciti a migliorare i precedenti framework sia per quanto riguarda il tempo di pre-processing, sia per la risposta alle query che per quanto riguarda lo spazio occupato. L'innovazione descritta nel paper risiede nell'utilizzo delle *highway*, concetto che, nei metodi già affermati, non veniva utilizzato a pieno. In particolare viene sottolineato come da alcuni nodi passino la maggior parte degli shortest path e come, questi nodi, siano appunto le highway. Per valutare, infatti, uno shortest path viene utilizzata la *speed* di un arco del grafo (intesa come rapporto tra la lunghezza dell'arco e il tempo di percorrenza di quest'ultimo) e, è intuibile notare che, molto frequentemente tali strade vengono percorse. Questa intuizione permette di attuare delle tecniche per velocizzare la fase di costruzione delle label.

Correttezza Matematica

L'articolo presenta al suo interno due dimostrazioni: una riguardante il *Lemma 3.1* con il quale viene dimostrato l'esistenza di due triple necessarie per calcolare la minima distanza e il *Teorema 4.1* con cui si dimostra la correttezza del metodo di pruning proposto dagli autori stessi. Dal punto di vista matematico i teoremi e le dimostrazioni sono ben esposti dagli autori, facendo vedere i passaggi e i ragionamenti utilizzati per arrivare a ciò che volevano dimostrare. Le uniche cose che si possono sottolineare sono: la mancanza della citazione alla metodologia di dimostrazione utilizzata (sia il teorema che il lemma, infatti, sono dimostrati per assurdo) e la mancanza di citazione a lemmi, definizioni o corollari che vengono utilizzati nella dimostrazione. Come,

ad esempio, nel *Lemma 3.1* viene usata la definizione di *Query* fatta poche righe precedenti, oppure nel *Teorema 4.1* viene usato il risultato del *Lemma 4.1* descritto a inizio paragrafo 4.

Timeliness

Gli autori hanno sperimentato il framework proposto e l'hanno messo a confronto con diversi sistemi dello stato dell'arte. In particolare hanno mostrato notevoli performance in pre-processing time, space usage e query time. Si noti però che questo framework proposto è stato notevolmente messo a confronto con *Hub-based Labels*, il loro precedente paper. I miglioramenti riportati non sono molto soddisfacenti in quanto, sia per quanto riguarda il pre-processing time che il query time, l'*Hub-based Labels* è leggermente migliore, o comunque sia i risultati riportati sono paragonabili. Il vantaggio del framework presentato sta nel fatto che lo spazio di memorizzazione è molto minore e, per tanto, si può considerare a tutti gli effetti un miglioramento e che nel framework proposto è possibile rispondere a query a cui il precedente non era in grado di dare una stima senza fare una ricerca vera e proprio nel grafo totale (cioè nel caso in cui due nodi sorgente/destinazione non avevano nodi in comune nelle loro label).

Aspetti formali

L'intera descrizione del lavoro svolto, è stata fatta in modo adeguato e succinto, anche se tal volta sarebbe servita descrizione un po' più approfondita dei metodi che sono stati presi come base a quelli proposti: come nel paragrafo 2.3 dove è stato soltanto accennato l'*Hub-based Labels*, ulteriore framework di labeling del grafo sulla base del quale viene presentato l'*Highway-based Labeling Framework*. Anche l'esempio riportato nel paragrafo 4.3 è invece troppo poco dettagliato: sarebbe stato più appropriato descrivere in modo semplice e specifico la prima iterazione dell'algoritmo proposto, ovvero riferita al percorso P_1 , invece di presentare in modo veloce ogni iterazione per ogni percorso. Dal punto di vista dell'accessibilità l'articolo è, per la maggior parte, ben comprensibile senza una specifica necessità di conoscenza di nozioni o preparazione preliminare. Alcune parti sono più specifiche dove gli autori si soffermano su dettagli implementativi, quali, ad esempio, come salvare le label in memoria per utilizzare al meglio gli hit della cache oppure la descrizione della decomposizione del grafo di input in shortest path. Altro punto di difficile comprensione potrebbe essere la descrizione dell'algoritmo, paragrafo 4.2. Infatti non è ben chiarito, per il nostro punto di vista, il corretto funzionamento di questo, in particolare la parte di pruning dell'algoritmo Dijkstra. Di contro però, la presentazione dei teoremi e delle dimostrazioni è ben trattata e di facile comprensione. Il paragrafo 5, quello della descrizione sulla decomposizione del grafo e le tecniche/euristiche per diminuire la dimensione delle label, manca di descrizione implementativa. Gli autori, dal nostro punto di vista, danno dettagli ad alto livello delle tecniche utilizzate senza una descrizione di un algoritmo che sarebbe stato senz'altro utile alla comprensione delle stesse. Infine negli esperimenti condotti, c'è da notare il fatto che gli autori sottolineano che il loro lavoro migliora il tempo di pre-processing rispetto agli altri sistemi dello stato dell'arte, fatta eccezione per HL-local. La motivazione che questi danno sta nel fatto che per questo sistema l'algoritmo di calcolo delle label è stato parallelizzato mentre il loro PHL no. A questo punto sorge spontanea la domanda: perché nel loro caso non è stato fatto? Avrebbero potuto, quindi, migliorare questo aspetto o, comunque, dare una motivazione più forte a difesa di questo fatto.

Dal punto di vista stilistico, questo articolo segue il template dei paper contenenti ricerche o approfondimenti su campi scientifici, indicati anche nella bibliografia dell'articolo stesso. Unico appunto che viene da fare è la struttura dell'introduzione: i paragrafi 1.1 e 1.2 si sarebbero potuti unire o scrivere meglio, come il paragrafo 2. Infatti i paragrafi 1.1, 2.1 e 2.3 sarebbero anch'essi potuti essere uniti. Ricapitolando si poteva lasciare nell'introduzione soltanto il paragrafo 1.2, introducendo un po' i concetti trattati nel paragrafo 1.1. Infine nel paragrafo 2 poteva esserci

tutta la parte descritta nei capitoli rimanenti sopra citati dando più fluidità anche a ciò che viene trattato. Un'altra cosa da sottolineare è la mancanza di spazio tra il titolo del paragrafo e l'inizio della trattazione di esso, soprattutto per quanto riguarda i sotto paragrafi. Mentre, per la parte rimanente, compreso posizione delle figure, tabelle e altri elementi grafici, gli autori hanno fatto un buon lavoro, inserendo questi elementi in modo sparso da non appesantire le pagine.

Replicabilità

Gli autori hanno messo a confronto il framework con gli altri modelli dello stato dell'arte e sono stati specificati a dovere il modo in cui questi sono stati testati. Da notare il fatto che, però, non c'è modo di replicare i risultati ottenuti anche che i dataset sono pubblici e utilizzabili, inoltre anche il codice sorgente è stato rilasciato da parte degli autori. I risultati sono replicabili in parte in quanto gli algoritmi che sono presenti nel paper sono implementabili senza problemi, purtroppo però è assente la descrizione per l'algoritmo della decomposizione del grafo e la descrizione letterale presente nel paragrafo 5. Ciò quindi lascia troppa libertà nell'implementazione e non si ha una certezza dell'effettiva efficacia di ciò che si potrebbe implementare. Inoltre non è semplice implementare tutti i trick per la memorizzazione delle label, in quanto la descrizione degli autori su questi aspetti è presente, ma poco dettagliata.

Commenti finali

L'articolo analizzato in questo referee è stato di piacevole lettura e molto interessante per quanto riguarda l'argomento trattato. Gli autori, a nostro avviso, hanno fatto un ottimo lavoro sia per quanto riguarda i miglioramenti proposti rispetto ai precedenti framework sia per l'innovazione introdotta. Sono possibili alcuni miglioramenti, seppur di poco spessore, e, sicuramente, sono possibili numerosi sviluppi futuri, sia per quanto riguarda le tecniche di decomposizione del grafo che di contrazione della dimensione delle label.