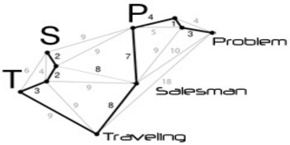
Μια υπολογιστική μελέτη ευρετικών μεθόδων αρχικοποίησης διαδρομών για το πρόβλημα του πλανόδιου πωλητή

Λαζαρίδης Αλέξανδρος

Πανεπιστήμιο Μακεδονίας, ΠΜΣ Εφαρμοσμένης Πληροφορικής Συστήματα Υπολογιστών

Ιούνιος 23, 2015



Περιεχόμενα

- Εισαγωγή
  - Συνεισφορά της εργασίας
- Βιβλιογραφική επισκόπηση
  - Περιγραφή προβλήματος
  - Μαθηματική μοντελοποίηση
- Construction heuristics
- Περιγραφή της TSPLIB και των National TSPs
- Υπολογιστική μελέτη
- Στατιστική ανάλυση
- Επίλογος

Συνεισφορά της εργασίας

- Η συνεισφορά της παρούσας διπλωματικής θέσης είναι η εξής:
  - Ανάλυση του TSP.
  - Εκτεταμένη περιγραφή των Construction heuristics.
  - Εκτέλεση των μετροπροβλημάτων National TSPs και της TSPLIB βιβλιοθήκης.
  - Στατιστική ανάλυση των ευρετικών μεθόδων ως προς τη χρόνο εκτέλεσης και τη ποιότητα λύσης.

Περιγραφή προβλήματος

- Το TSP:

1. περιγράφεται ως το πρόβλημα εύρεσης της συντομότερης περιοδείας ενός πωλητή (salesman) που αρχίζει και τελειώνει σε ένα σημείο εφόσον προηγουμένως έχει επισκεφτεί έναν αριθμό n πόλεων (κόμβων) ακριβώς μια φορά, (Hamiltonian circuit).
2. ορίζεται σε ένα πλήρη γράφο $G = (V, A)$.
3. είναι ένα κλασικό NP hard πρόβλημα.
4. έχει πολλαπλές παραλλαγές (mTSP, MaxTSP, TDTSP)
5. βρίσκει εφαρμογή σε πολλούς τομείς (τεχνολογία, εφοδαστική αλυσίδα κ.ά).

Μαθηματική Μοντελοποίηση του TSP

- Ορίζουμε με i τους κόμβους του γράφου.
- Κάθε ζεύγος (i,j) με $i \neq j$, αντιστοιχεί σε μια ακμή ή ένα τόξο του γράφου.
  1. Συμμετρικό: $c_{ij} = c_{ji}$
  2. Ασύμμετρο: $c_{ij} \neq c_{ji}$

$$\text{(STSP)minimize} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\sum_{i<k} x_{ik} + \sum_{j>k} x_{kj} = 2, \quad k \in V, \quad i = 1, \ldots, n$$

$$\sum_{j \in S} \sum_{i,j \in \bar{S}} x_{ij} \geq 1, (S \subset V, S \neq \emptyset)$$

$$\text{(ATSP)minimize} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\sum_{i=0}^{n} x_{ij} = 1, i = 1, \ldots, n$$

$$\sum_{j=0}^{n} x_{ij} = 1, j = 1, \ldots, n$$

$$\sum_{j \in S} \sum_{i,j \in \bar{S}} x_{ij} \geq 1, (S \subset V, S \neq \emptyset)$$

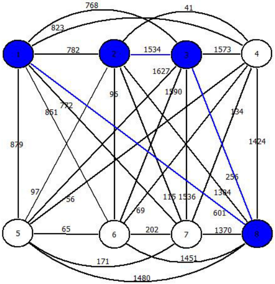
$$x_{ij} \in \{0, 1\}, \ \forall i, \ \forall j, \ i \neq j$$

Nearest Neighbor(NN)

- **Τι είναι μέθοδοι αρχικοποίησης**;
- Είναι αλγόριθμοι που μπορούν να κατασκευάσουν μια περιοδεία TSP από το μηδέν, καθώς δίνουν μόνο αρχική λύση.
- Η πιο δημοφιλής ευρετική μέθοδος επίλυσης του TSP.
- Χρονική πολυπλοκότητα: $T(n) = O(n^2)$
- Ο (NN) παράγει μία διαδρομή 25% μεγαλύτερη από την πλέον σύντομη για ένα καθορισμένο πλήθος πόλεων, τυχαία τοποθετημένων σε ένα επίπεδο, (Johnson, D.S. and McGeoch, L.A.(1995). The Traveling Salesman Problem: A Case Study in Local Optimization, November 20, 1995).

Nearest Neighbor(NN) - Βήματα εκτέλεσης

1. Επιλέγεται τυχαίος κόμβος i, ως αρχικός κόμβος: **1**.
2. Εντοπίζεται ο πλησιέστερος γειτονικός κόμβος k του i που δεν βρίσκεται στη περιοδεία: **8**.
3. Εισάγεται το k στο τέλος της περιοδείας: T=\{**1**,**8**\}



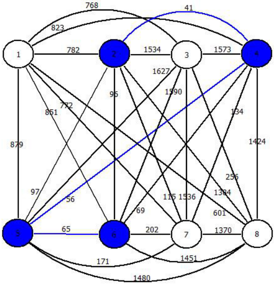
4. Επαναλαμβάνεται το Βήμα 2, μέχρι να εισαχθούν όλοι οι κόμβοι στη περιοδεία:**3** (8), T=\{**1**,**8**,**3**\}
5. Επιστροφή στον αρχικό κόμβο, σχηματίζοντας Hamiltonian circuit.

Double nearest neighbor(DNN)

- Είναι **παραλλαγή** του (NN).
- Χρονική πολυπλοκότητα: $T(n) = O(n^2)$
- Κεντρική ιδέα: οι κόμβοι δίνονται να προσθέσουν και τα δύο άκρα της περιοδείας.

Double nearest neighbor(DNN) - Βήματα εκτέλεσης

1. Εντοπίζεται η συντομότερη ακμή και επιλέγεται ένας από τους δύο κόμβους της ακμής ως αρχικός κόμβος: **[4,2] (4)**, T={**4**}
2. Εντοπίζεται κόμβος που είναι πιο κοντά στο αντίστοιχο τελικό άκρο της διαδρομής: **2 (4)**, T={**4,2**}
3. Επαναλαμβάνεται το Βήμα 2: **5** για λογαριασμό του κόμβου **4**, T={**5,4,2**}



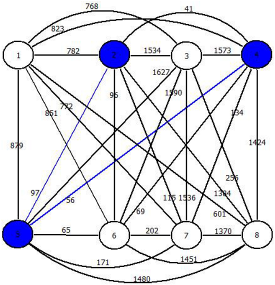
4. Επιστροφή στον αρχικό κόμβο, σχηματίζοντας Hamiltonian circuit.

Nearest Insertion(NI)

- Η πρώτη μέθοδος της ομάδας εισαγωγής. (Rosenkrantz, S. and Lewis An Analysis of Several Heuristics for the Traveling Salesman Problem SIAM Journal on Computing, 1977, 6(3), 563 - 581)
- Χρονική πολυπλοκότητα: $T(n) = O(n^2)$
- Οι παραλλαγές του NI είναι οι ευρετικές μέθοδοι Farthest Insertion, Cheapest Insertion και Random Insertion.
- Οι μέθοδοι εισαγωγής διακρίνονται από δύο τμήματα:
    1. την επιλογή του επόμενου κόμβου.
    2. την εισαγωγή του κόμβου που επιλέχθηκε στη υπάρχουσα περιοδεία.
- Η εκχώρηση κάθε κόμβου k γίνεται με τον υπολογισμό της διαφοράς κόστους: $\Delta f = c_{ik} + c_{kj} - c_{ij}$

Nearest Insertion(NI) - Βήματα εκτέλεσης

1. Επιλέγεται τυχαίος κόμβος i, ως αρχικός κόμβος: **4**.
2. Εντοπίζεται κόμβος j πιο κοντινός σε i, έτσι ώστε το κόστος c_{ij} να ελαχιστοποιείται: **2 (4)**, T=\{**4,2**\}.
3. Εντοπίζεται κόμβος k που να μην ανήκει στη περιοδεία και να είναι πιο κοντά σε οποιοδήποτε από τους κόμβους αυτής, $d(k, T) = min$ $d(i, T))$: **5**



Nearest Insertion(NI) - Βήματα εκτέλεσης

1. Εντοπίζεται ακμή $[i, j]$ της περιοδείας και εισάγεται ο k, έτσι ώστε η Δf να ελαχιστοποιείται: **[4,2] και [2,4]**, T=**{4,5,2}**.
2. Επαναλαμβάνεται το βήμα 3 μέχρι να σχηματιστεί ένα Hamiltonian circuit.

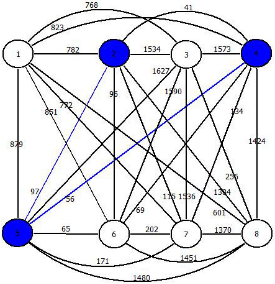
- Παραλλαγή της (NI) είναι η Farthest Insertion(FI).
- Η διαφοροποίηση της από τη (NI), βρίσκεται στο ότι εντοπίζει κόμβο k που δεν ανήκει στην περιοδεία και ταυτόχρονα είναι **πιο μακριά από οποιοδήποτε κόμβο αυτής** (δηλαδή, ισχύει $d(k, T) = max\ d(i, T)$).

Cheapest Insertion(CI)

- Η βασική διαφορά από τον (NI), είναι ότι επιλέγεται ο κόμβος k που δίνει τη μικρότερη αύξηση από τους κόμβους που δεν είναι στη περιοδεία και εισάγεται στην αντίστοιχη ακμή$[i,j]$.
- Χρονική πολυπλοκότητα: $T(n) = O(n^2 logn)$

Cheapest Insertion(CI) - Βήματα εκτέλεσης

1. Επιλέγεται κόμβος i, ως αρχικός κόμβος: 4
2. Εντοπίζεται κόμβος j πιο κοντινός σε i: 2 (4) με περιοδεία T={4,2}
3. Εντοπίζεται ακμή [i,j] της περιοδείας και κόμβος k που να μην ανήκει στην περιοδεία, έτσι ώστε η Δf να ελαχιστοποιείται: επομένως για τον 1 είναι Δf[4,2]=1564, για τον 3: 3066, **για τον 5: 115**, για τον 6: 124, για τον 7: 208 και για τον 8: 2767, T={4,5,2}
4. Επαναλαμβάνεται το βήμα 3 μέχρι να σχηματιστεί ένα Hamiltonian circuit.

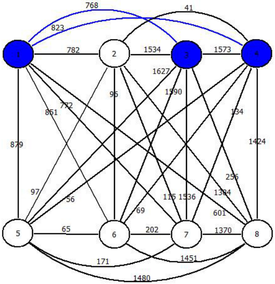


Random Insertion(RI) - Βήματα εκτέλεσης

- **Κεντρική ιδέα**: επιλέγει τυχαία τους κόμβους που θα εισαχθούν στη καλύτερη δυνατή θέση μέσα στη περιοδεία.
- Χρονική πολυπλοκότητα: $O(n^2)$
- **Βήματα εκτέλεσης**
  1. Επίλεγεται με τυχαίο τρόπο ένας κόμβος εκτός της περιοδείας για να μπει στη περιοδεία.
  2. Εντοπίζεται ακμή $[i, j]$ της περιοδείας και εισάγεται το k που επιλέγεται προηγουμένως, έτσι ώστε η αύξηση του μήκους $\Delta f = c_{ik} + c_{kj} - c_{ij}$ να ελαχιστοποιείται. Τροποποιείται η περιοδεία εισάγοντας k μεταξύ i και j.
  3. Επαναλαμβάνεται το βήμα 1 μέχρι να σχηματιστεί έναν Hamiltonian circuit.

Random Insertion(RI)

1. **B1**:Επιλέγεται τυχαία τους κόμβους 3 και 4, T={3,4}.
2. **B1**:Επιλέγεται τυχαία κόμβος: 1.
3. **B2**: Για τον 1 θα υπολογίσουμε τη (Δ.Κ) για την ακμή[3,4]. Επομένως, η περιοδεία θα τροποποιηθεί ως εξής T={3,1,4}.
4. **B3,4**: Επιλέγεται τυχαία ο 8, και θα εισχωρήσει στην ακμή[3,1] καθώς Δf[3,1]=89 έναντι των ακμών [1,4],[4,3].



Περιγραφή των μετροπροβλημάτων της TSPLIB

- Η TSPLIB είναι μια συλλογή μετροπροβλημάτων, τα οποία είναι διαθέσιμα στην ερευνητική κοινότητα.
- **Τμήμα Προδιαγραφών** (specification part): περιέχει πληροφορίες σχετικά με τη μορφή των αρχείων.
  1. NAME:$< string >$: Προσδιορίζει το όνομα του αρχείου.
  2. TYPE:$< string >$: Καθορίζει το τύπο των δεδομένων(TSP,$ATSP$,$CVRP$).
- **Τμήμα δεδομένων** (data part): περιέχει ρητά δεδομένα.
  1. NODE COORD SECTION: δίνονται οι συντεταγμένες του κόμβου με την εξής μορφή:$< integer >< real >< real >$
  2. DEPOT SECTION: περιλαμβάνει μια λίστα με πιθανές εναλλακτικές αφετηρίες. Η λίστα τερματίζεται με τον ακέραιο -1.

Περιγραφή των National TSPs

- Αποτελείται από 25 TSP.
- Απεικονίζονται 25 διαφορετικές χώρες.
- Κάθε κόμβος αποτελεί μια πόλη μιας χώρας.
- Το κόστος των αποστάσεων μεταξύ των πόλεων καθορίζεται με Ευκλείδεια απόσταση.
- Τα δεδομένα των αποστάσεων και των κόμβων προέρχονται από το National Geospatial-Intelligence Agency. `http://www.math.uwaterloo.ca/tsp/world/countries.html`

Υπολογιστική μελέτη [1/3]

- Μέτρα αξιολόγησης: ποιότητας λύσης και ταχύτητα.
- Υπολογιστικά πειράματα της TSPLIB και τα National TSPs.
- **STSPs**: (102 benchmarks, 6 heuristics)

	Error (%)	CPUtime (sec)
NN	$57,51$	$0,25$
DN	$67,63$	$0,25$
NI	$24,93$	$551,625$
FI	$14,91$	$550,625$
CI	$16,59$	$146,63$
RI	$11,38$	$0,232$

1. Η (RI) είναι η πιο αποτελεσματική και αποδοτική μέθοδος.
2. Ο (NI) και (FI) έχουν μεγαλύτερο χρόνο εκτέλεσης.

Υπολογιστική μελέτη [2/3]

- **ATSPs**: (19 benchmarks, 6 heuristics)

	Error(%)	CPUtime (sec)
NN	$63,89$	$0,001$
DN	$101,5$	$0,004$
NI	$20,81$	$0,74$
FI	$15,83$	$0,74$
CI	$16,58$	$0,23$
RI	$14,17$	$0,004$

1. Ο (FI) είναι πιο αποτελεσματικός έναντι του (NI) και του (CI).
2. Η (RI) είναι η αποτελεσματικότερη ευρετική μέθοδος όλων των μεθόδων, ενώ η (NN) είναι ταχύτερη.

Υπολογιστική μελέτη [3/3]

- **National TSP***s*: (15 benchmarks, 4 heuristics)

	Error(%)	CPUtime (sec)
NN	$52,27$	$3,93$
DN	$48,79$	$4,90$
CI	$20,80$	$4361,83$
RI	$12,24$	$3,18$

1. Η (RI) είναι η αποδοτικότερη και αποτελεσματικότερη ευρετική μέθοδος έναντι των υπολοίπων.
2. Η (CI) έχει το μεγαλύτερο μέσο χρόνο εκτέλεσης.

Μεθοδολογία στατιστικής ανάλυσης

- Διάκριση των προβλημάτων σε τρεις συλλογές.
- Κάθε συλλογή διακρίνεται από *n* παρατηρήσεις και από δύο μεταβλητές (*Error* και *Cputime*).
- Οι παρατηρήσεις πρέπει να είναι συσχετιζόμενες (correlated) μεταξύ τους.
- One way repeated measures ANOVA (παραμετρικό τεστ) (2 προυποθέσεις)
  1. κανονικότητα: Shapiro - Wilk normality test
  2. σφαιρικότητα: Mauchlys test
  3. Άν απορριφθεί η κανονικότητα και η σφαιρικότητα: Friedman test και Wilcoxon signed rank test
- Η Στατιστική Ανάλυση (Σ.Α) έγινε σε R.

Στατιστική ανάλυση συμμετρικών στιγμιοτύπων [1/3]

- Ως προς τη μεταβλητή *Error*
  1. Κανονικότητα: *pvalue*=2,2*e*-16 < 0,05 (Απόρριψη)
  2. Σφαιρικότητα: *pvalue*=0 < 0,05 (Απόρριψη)
  3. Friedman test: *pvalue*=2.2*e*-16 < 0,05 (σημαντικές διαφορές)
  4. Δίνεται ο πίνακας Wilcoxon signed rank test

	CI	DNN	FI	NI	NN
DNN	5.5e-11	-	-	-	-
FI	0.59	1.1e-09	-	-	-
NI	4.7e-16	0.93	4.4e-16	-	-
NN	2.0e-12	1.00	2.1e-11	1.00	-
RI	7.0e-13	5.8e-16	1.6e-06	< 2e-16	5.1e-16

(Με πανομοιότυπο τρόπο εξετάζουμε και ως προς το *Cputime*)

Στατιστική ανάλυση ασύμμετρων στιγμιοτύπων [2/3]

- Ως προς τη μεταβλητή *Cputime*
  1. Κανονικότητα: $p - value = 1.121e\text{-}06 < 0{,}05$ (Απόρριψη)
  2. Σφαιρικότητα: $pvalue = 1.724758e\text{-}132 < 0{,}05$ (Απόρριψη)
  3. Friedman test: $p - value = 1.16e\text{-}08 < 0{,}05$ (σημαντικές διαφορές)
  4. Δίνεται ο πίνακας Wilcoxon signed rank test

	CI	DNN	FI	NI	NN
DNN	0.207	-	-	-	-
FI	0.134	0.053	-	-	-
NI	0.212	0.084	1.000	-	-
NN	0.207	1.000	0.053	0.084	-
RI	0.207	1.000	0.053	0.084	0.817

(Με πανομοιότυπο τρόπο εξετάζουμε και ως προς το *Error*)

Στατιστική ανάλυση National TSPs [3/3]

- Ως προς τη μεταβλητή *Cputime*
  1. Κανονικότητα: $pvalue = 0.0001027 < 0{,}05$ (Απόρριψη)
  2. Σφαιρικότητα: $pvalue = 3.625195e\text{-}22 < 0{,}05$ (Απόρριψη)
  3. Friedman test: $p - value = 2.031e\text{-}08 < 0{,}05$ (σημαντικές διαφορές)
  4. Δίνεται ο πίνακας Wilcoxon signed rank test

	CI	DNN	NN
DNN	0.0065	-	-
NN	0.0054	0.1969	-
RI	0.0044	0.0044	0.0045

(Με πανομοιότυπο τρόπο εξετάζουμε και ως προς το *Error*)

Επίλογος

- Συμπεράσματα
  1. Η οικογένεια των Insertions εντοπίζει καλές αρχικές λύσεις.
  2. Η (RI) βρίσκει καλύτερες λύσεις σε μεγάλα στιγμιότυπα.
  3. Η Σ.Α μπορεί να δείξει σημαντικές διαφορές μεταξύ των ευρετικών.
- Μελλοντική Έρευνα
  1. να βελτιωθούν προγραμματιστικά οι υπάρχουσες ευρετικές.
  2. να υλοποιηθούν περαιτέρω εκδόσεις των μεθόδων (Fast versions).
  3. να υλοποιηθούν οι υπάρχουσες ευρετικές μέθοδοι με αρχική υποδιαδρομή το κυρτό περίβλημα.
  4. να υλοποιηθούν ευρετικές βελτίωσης γνωστές ως (Improvement Heuristics).

Τέλος παρουσίασης

Ευχαριστώ για την προσοχή σας..