A. J. Hoffman
S. Winograd

Finding All Shortest Distances in a Directed Network

Abstract: A new method is given for finding all shortest distances in a directed network. The amount of work (in performing additions, subtractions, and comparisons) is slightly more than half of that required in the best of previous methods.

Introduction

Let $D = (d_{ij})$ be a real square matrix of order n with 0 diagonal. We shall think of each of the numbers d_{ij} as respresenting the "length" of a link from vertex i to vertex j in a directed network. While we do not assume that all d_{ij} are nonnegative, we do assume that, if σ is any permutation of $N = \{1, \cdots, n\}$, then $\sum_i d_{i\sigma i_1} \geq 0$. This is equivalent to the customary assumption that the sum of the lengths around any cycle is nonnegative, an assumption generally made in shortest-distance problems.

Our problem is to calculate all "shortest distances" from i to j for all $i \neq j$. More formally, define a path P from i to j as an ordered sequence of distinct vertices $i = i_0, i_1, \cdots, i_k = j$, and define its length $L(P)$ by $L(P) = \sum_{r=0}^{k-1} d_{i_r, i_{r+1}}$. Our problem is to calculate a square matrix $E = (e_{ij})$ of order n such that $e_{ij} = \min_P L(P)$, where P ranges over all paths from i to j.

To our knowledge, the most efficient method in the literature is due to Floyd [1] and Warshall [2], who showed that E can be calculated in n^3 additions and n^3 comparisons. (Here and elsewhere we suppress terms of lower order unless they are needed in the course of an argument.) The purpose of this paper is to announce an improved method.

- *Theorem*

If D is the matrix of link lengths, E the matrix of shortest distances of a directed network on n vertices, and if $\epsilon > 0$ is given, then E can be calculated from D in $(2 + \epsilon)n^{5/2}$ addition-subtractions and n^3 comparisons.

Proof

The proof of the theorem will consist in producing an algorithm and showing that it has the stated properties. Our algorithm borrows much from Shimbel [3], as well as from [1] and [2], but has two special features which we now outline briefly.

Let A be a $p \times q$ matrix, B a $q \times r$ matrix, and define $A \circ B = C = (c_{ij})$ to be the $p \times r$ matrix given by

$$c_{ij} = \min_k (a_{ik} + b_{kj}) .$$

A straightforward approach to calculating C would require pqr additions and $pr(q-1)$ comparisons. Our method, discussed in the following section, requires $pr(q-1)$ comparisons also, but fewer than $(q - 1/2)\sqrt{2pr(p+r)} + pr$ addition-subtractions.

The second special feature is that we suitably partition the vertices of our network into subsets of proper size and proceed to calculate E by a sequence of operations of the form $A \circ B$ and solutions of shortest-distance problems on the subsets. This part is a direct generalization of [1] and [2] in which the subsets consist of exactly one vertex. Hu [4] has also described a partitioning of D to take advantage of sparseness and geography, which is a different matter. Presumably our method could be modified to take similar advantages, but we do not pursue this point.

Pseudomultiplication of matrices

• *Lemma*

Let A and B be matrices of dimension $p \times q$ and $q \times r$ respectively. Define $(A \circ B)_{ij} = \min_k (a_{ik} + b_{kj})$. Then $A \circ B$ can be calculated in $pr(q-1)$ comparisons and fewer than $(q - 1/2)\sqrt{2pr(p+r)} + pr$ additions.

Proof

Define, for any collection $M^1, M^2, \cdots, M^k$ of matrices of the same dimension,

$$M = \min(M^1, \cdots, M^k) = (m_{ij}) = \min(M^1_{ij}, \cdots, M^k_{ij}).$$

Partition the *columns* of A into nonempty subsets $S_1, \cdots, S_k$ of size $d_1, \cdots, d_k$ respectively. Partition the *rows* of B conformally. Let A_i, $i = 1, \cdots, k$, be the submatrix of A consisting of the columns in S_i. Let B'_i be the corresponding submatrix of rows of B. Clearly

$$A \circ B = \min(A_1 \circ B'_1, \cdots, A_k \circ B'_k).$$

Calculate $A_i \circ B'_i$, $i = 1, \cdots, k$ as follows: Form all $p\, d_i(d_i - 1)/2$ differences $a_{tj} - a_{tk}$, $t = 1, \cdots, p$, $j, k \in S_i$, $j < k$. Similarly, form all $r\, d_i(d_i - 1)/2$ differences $b_{ku} - b_{ju}$, $u = 1, \cdots, s$, $j, k \in S_i$, $j < k$. In order to find the (t,u)th entry of $A_i \circ B'_i$, we observe that $a_{tj} + b_{ju} \le a_{tk} + b_{ku}$ if and only if $a_{tj} - a_{tk} \le b_{ku} - b_{ju}$. Since we have already calculated these differences, it is clear that $(d_i - 1)$ comparisons will yield, for each (t,u), the index l such that $a_{tl} + b_{lu} = \min_{k \in S_i}(a_{tk} + b_{ku})$. Next, for each (t,u), we calculate $a_{tl} + b_{lu}$. Thus we have found $A_i \circ B'_i$ in $[(p+r)/2]d_i(d_i - 1)$ subtractions, pr additions and $pr(d_i - 1)$ comparisons.

It follows that $A \circ B = \min_i\{(A_i \circ B'_i)\}$ can be calculated in

$$[(p+r)/2] \sum d_i^2 - [(p+r)/2]q + prk \tag{1}$$

addition-subtractions (here we have used $\sum d_i = q$), and $pr\left(\sum (d_i - 1) + k - 1\right) = pr(q-1)$ comparisons.

Let us study (1) further. Define m to be the smallest integer not less than $\sqrt{2pr/p+r}$. Thus

$$m = \sqrt{(2pr)/(p+r)} + \theta, \qquad 0 \le \theta < 1. \tag{2}$$

Write $q = am + b$, $0 \le b \le m - 1$.

Case 1. $b = 0$. Choose $k = a$, $d_1 = \cdots = d_k = m$. Then (1) becomes

$$[(p+r)/2]qm - [(p+r)/2]q + pr\, q/m, \tag{3}$$

which is easily seen to be less than the number specified in the lemma. Here we use in our estimates $\sqrt{2pr(p+r)} \le m \le \sqrt{2pr(p+r)} + 1$, and $p + r \le 2pr$ for positive integers p and r.

Case 2. $b \ne 0$. Set $k = a + 1$, $d_1 \cdots = d_a = m$, $d_{a+1} = b$. Then

$$\sum d_i^2 = m^2 a + b^2 = mq + b^2 - mb \le mq + 1 - m,$$

since $1 \le b \le m - 1$. Using this estimate, $\sqrt{2pr(p+r)} \le m \le \sqrt{2pr/(p+r)} + 1$, $k \le q/m + 1$, we obtain the estimate given in the statement of the lemma.

Finding all shortest distances (description and validation of algorithm)

Let N be partitioned into nonempty subsets $S_1 \cup \cdots \cup S_k$ of respective sizes $d_1, \cdots, d_k$. We shall proceed to modify the matrix D by successive steps so that the resulting matrix is E. In our description, the letter D will always stand for the current step of the modification of D. $D[S,T]$ will mean the submatrix of D formed by rows in S, columns in T, $D[S] = D[S,S]$. $\mathscr{E}D[S]$ stands for the shortest distance matrix computed from the submatrix $D[S]$. $\bar{S}$ means the complement of S. The expression $D[S,T] \leftarrow F$ means that in D, $D[S,T]$ gets replaced by F. All other entries of D are unchanged.

a) Let $i = 1$

b) $D[S_i] \leftarrow \mathscr{E}D[S_i]$

c) $D[\bar{S}_i, S_i] \leftarrow D[\bar{S}_i, S_i] \circ D[S_i]$

$D[S_i, \bar{S}_i] \leftarrow D[S_i] \circ D[S_i, \bar{S}_i]$

d) $D[\bar{S}_i] \leftarrow \min\{D[\bar{S}_i], D[\bar{S}_i, S_i] \circ D[S_i, \bar{S}_i]\}$

e) Increase i by 1. If $i = k$, stop. Otherwise, go to b.

After steps a) through e) are completed the first time, d_{ij} equals the shortest distance from i to j in which we are restricted to paths in which all intermediate vertices, if any, belong to S_1. This holds for all i, j. Manifestly, after we have completed a) through e) l times, d_{ij} equals the shortest distance from i to j in which we are restricted to paths where all intermediate vertices, if any, belong to $S_1 \cup \cdots \cup S_l$. Thus, by induction, the algorithm is easily seen to be valid.

We now show inductively that the number of comparisons $f(n)$ required by this algorithm is at most n^3. Examination of a) through e) shows that

$$\begin{aligned} f(n) = {} & \sum f(d_i) + 2\sum (n - d_i)d_i(d_i - 1) \\ & + \sum (n - d_i)(n - d_i)(d_i - 1) \\ & + \sum (n - d_i)(n - d_i). \end{aligned}$$

Assuming inductively that $f(d_i) \le d_i^3$, and using $\sum d_i = n$, we get $f(n) \le n^3$. Note that this does not depend on the magnitudes (d_i).

Count of addition-subtractions in the algorithm

If now we let $f(n)$ be the number of addition-subtractions required, we find from a) through e) that

$$f(n) < \sum f(d_i) + 2 \sum d_i \sqrt{2(n-d_i)nd_i} + 2 \sum (n-d_i)^2 + 2 \sum d_i(n-d_i) \sqrt{n-d_i}. \quad (4)$$

(Here we have suppressed the factor "−1/2" in the lemma.) In order to get an estimate of how n grows, let us tentatively assume $n = a^t$, $d_i = a^{t-1}$, $k = a$. Then we have

$$\begin{aligned} f(a^t) &\leq af(a^{t-1}) + 2a\{a^{t-1}[2(a^t - a^{t-1})a^{2t-1}]^{1/2} \\ &\quad + (a^t - a^{t-1})a^{t-1} + a^{t-1}(a^t - a^{t-1})^{3/2}\} \\ &= af(a^{t-1}) + 2a \cdot a^{(5/2)t}\left\{\frac{1}{a}\left[2\left(1-\frac{1}{a}\right)\frac{1}{a}\right]^{1/2} + \frac{1}{a}\left(1-\frac{1}{a}\right)^{3/2}\right\} + O(a^{2t}) \\ &= af(a^{t-1}) \\ &\quad + 2a^{(5/2)t}\left[\sqrt{1-(1/a)}\sqrt{(2/a)+1-(1/a)}\right] \\ &\quad + O(a^{2t}). \end{aligned} \quad (5)$$

Setting $f(a^t) = Aa^{(5/2)t}$, we find

$$Aa^{(5/2)t} \leq 1/(a^{3/2})\, Aa^{(5/2)t} + 2\left[\sqrt{1-(1/a)}\sqrt{(2/a)+1-(1/a)}\right]a^{(5/2)t}$$

$$A \leq \frac{2\sqrt{1-(1/a)}\sqrt{(2/a)+1-(1/a)}}{1 - 1/(a^{3/2})} = 2 + \epsilon(a),$$

where $\epsilon(a) \to 0$ as $a \to \infty$.

But in order to establish this rigorously, we must proceed more carefully, without assuming that $n = a^t$. Because the details are tedious, we shall confine ourselves to an outline of the algorithm and proof.

First, an integer a is chosen. If $n < a$, the problem is solved by the method of [1] and [2]. If $n \geq a$, write $n = am + b$, $0 \leq b < a$. Then $n = b(m+1) + (a-b)m$. Let $d_1, \cdots, d_b = m+1$, $d_{b+1} = \cdots = d_a = m$. Partition n into subsets of size $d_1, \cdots, d_a$ and apply the algorithm given in this section. To prove that $f(n)$, the number of additions required, is at most $An^{5/2}$ + terms of lower order in n, the strategy is to assume inductively that $f(n) = An^{5/2} + P(a)n^2$, where P is a certain polynomial in a. Then using (4), replace m by $m+1$ throughout. One finds that an auspicious choice of $P(a)$ makes $f(n) \leq f[a(m+1)] \leq A(am)^{5/2} + P(a)(am)^2 \leq An^{3/2} + P(a)n^2$. This choice of $P(a)$ also makes the formula valid if $n \leq a$ and completes the proof.

Acknowledgment

This work was supported in part by the Office of Naval Research under contract numbers N0014-69-C-0023 and N0014-71-C-0112.

References

1. R. W. Floyd, "Algorithm 97, shortest paths," *Comm. of ACM* **5**, 345 (1962).
2. S. Warshall, "A theorem on Boolean matrices," *J. ACM* **9**, 11 (1962).
3. A. Shimbel, "Structure in communication nets," Proc. of the Symposium on Information Networks, (April 1954) 199–203, Polytechnic Institute of Brooklyn, New York (1955).
4. T. C. Hu, "A decomposition algorithm for shortest paths in a network," *Operations Research*, **16**, No. 1, 91–102 (January-February 1968).

Received November 15, 1971

The authors are located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598.