Уфа : УГАТУ, 2013 Вестник УГАТУ Т. 17, № 2 (55). С. 50–52

УДК 519.178

Оценка достоверности и аппроксимация информации, представленной взвешенным графом

В. П. Житников[1], А. Р. Ураков[2], Т. В. Тимеряев[3]

[1,2,3]urakov@ufanet.ru

[1,2,3] ФГБОУ ВПО «Уфимский государственный авиационный технический университет» (УГАТУ)

Поступило в редакцию 11.06.2012

Аннотация. Рассмотрены проблемы аппроксимации одного взвешенного графа другим взвешенным графом, но с меньшим числом узлов. Задача осложнена тем, что исходная информация представлена не в виде графа, а в виде множеств объектов и связей между ними, причем исходные данные могут содержать ошибки, которые необходимо по мере возможности выявить и исключить из аппроксимирующего графа.

Ключевые слова. *Аппроксимация графа; оценка достоверности графа; матрица достижимости*

Задача управления транспортом при большом числе перемещений по обслуживаемой территории сталкивается с проблемами формирования маршрутов и оценки их длин в динамической, быстроизменяющейся системе. При этом допускается достаточно большая погрешность измерения маршрутов, если это позволит получить желаемую (достаточно высокую) скорость вычислений. Другими словами, требуется аппроксимация одной модели другой, менее точной, но более простой.

Рассмотрим эту задачу в следующем виде. Задано множество объектов N мощностью n. Определенные пары объектов i, j связаны между собой некоторым числом (расстоянием) $x_{i,j}$, причем $x_{i,j} = x_{j,i}$. Требуется найти: а) новое множество M мощностью m (при том, что $m << n$), в котором каждая пара связана между собой числом $y_{i,j}$; б) функцию F, преобразующую N в M таким образом, чтобы значения $x_{i,j}^*$, полученные для объектов N как $x = f(y)$, имели приемлемое приближение с исходными $x_{i,j}$, т. е. $| x_{i,j} - x_{i,j}^*| < \varepsilon$ для всех i, j, где ε – желаемое приближение.

АППРОКСИМАЦИЯ ГРАФА

При постановке задачи множества объектов $\{N\}$ и расстояний $\{x\}$ не заданы виде графа, однако $\{x\}$ можно свести к матрице достижимости A по следующим правилам.

- Если в множестве x элемент $x_{i,j}$ единственный, то $a_{i,j} = x_{i,j}$, иначе $a_{i,j} = \min\{x_{i,j}\}$
- Для всех пар i, j: $a_{i,j} := a_{j,i} := \min(a_{i,j}, a_{j,i})$.
- Если $x_{i,j}$ отсутствует, то $a_{i,j}$ следует найти либо применяя для A алгоритм Дейкстры [1] (поиска кратчайшего маршрута [2]), либо используя следующий специальный способ, в котором будут изменяться не только те $a_{i,j}$, для которых отсутствуют $x_{i,j}$.

Специальный способ.

1. Для всех отсутствующих $x_{i,j}$ положим $a_{i,j} = \infty$.
2. Для каждой пары объектов i, j найдем такие k, для которых $a_{i,k} \neq \infty$ и $a_{k,j} \neq \infty$, и получим суммы: $s_{i,j,k} := a_{i,k} + a_{k,j}$.
3. Присвоим $a_{ij} := \min\limits_{k}(s_{ijk}, a_{ij})$
4. Пересчитаем все суммы, в которых участвует $a_{i,j}$, т.е. $s_{t,i,j}$, $s_{i,t,j}$ где t – все объекты, кроме i и j.
5. Повторим пункты 2–4 до тех пор, пока существуют $a_{i,j} = \infty$.

Конец специального способа.

Аппроксимацию можно проводить либо не задавая дополнительных условий для M, либо из условия, что M является подмножеством N. Используем второй способ, аппроксимацию для которого можно выполнить по алгоритму.

1. Для каждой пары объектов рассчитаем:

$$r_{ij} = \sum_{k=1}^{N-2} (a_{ik} - a_{kj})^2,\ k \neq i, k \neq j$$

или $r_{ij} = \min\limits_{k=1,N-2} (a_{ik}, a_{kj}),\ k \neq i, k \neq j.$

2. В качестве первой пары элементов множества M (M_1 и M_2) выберем N_i и N_j для которых $r_{i,j}$ – максимально.

3. Для каждого элемента k принадлежащего N, но не принадлежащего М, найдем:

$$r_k = \min_i(a_{ki}),\ \text{где}\ i \in M$$

4. Элемент k, для которого r_k максимально, добавим в множество M.

5. Пункты 3–4 будем повторять до тех пор, пока мощность M не достигнет заданной величины m.

Приведенный подход не гарантирует получения оптимального результата, однако для достаточно больших m дает значение ε, допустимое в практическом смысле. Можно надеяться на лучший результат, если на шагах 1 и 2 вместо пары выбирать большее количество элементов, например q. В этом случае вместо матрицы мы получаем массив размерности q. Надежда на улучшение для больших q следует из того, что оптимальный результат для m получается, если $q = m$ (в этом случае этапы алгоритма 3–5 не требуются). Тем не менее, работа с большими q трудна в практическом смысле и требует больших объемов вычислений (степенная сложность алгоритма). Аппроксимирующие множества $\{M\}$ и $\{x\}$ должны быть представлены в виде взвешенного, односвязного и неориентированного графа G – это требуется для дальнейшего использования результатов. Граф можно отыскать, используя решение следующей задачи.

ЗАДАЧА ГЕНЕРАЦИИ ГРАФА НА ОСНОВЕ МАТРИЦЫ ДОСТИЖИМОСТИ

Задан взвешенный, односвязный и неориентированный граф вида $G = (V, E, W)$, где $V(G)$ – множество вершин графа, $E(G)$ – множество ребер, W – весовая функция, определяющая вес $w(e)$ каждого ребра e. В задаче граф изначально представлен как «черный ящик». Известно множество вершин $V(G)$. Множество ребер $E(G)$ и весовая функция $W(V, E)$ неизвестны. Исследование графа заключается в серии K экспериментов следующего вида:

- в некоторую случайную (входную) вершину v_p помещается «исследователь»;
- «исследователь» движется по случайному маршруту, т. е. в каждом узле v_i следующая вершина v_j выбирается случайным образом из числа вершин, связанных с v_i;
- «исследователь» перемещается по ребрам e от вершины v_i к вершине v_j, складывая значения весов пройденных ребер $s_{pj} = s_{pi} + + w_{ij}$;
- в любой вершине «исследователь» может выйти из графа, обозначим эту вершину как v_q;
- после выхода «исследователь» сообщает значение s_{pq}.

Задача заключается в том, чтобы построить граф вида $G' = (V, E', W')$, который в том же эксперименте даст полученные ранее значения s_{pq}, при этом $|E|$ – количество ребер E – в нем будет настолько мало, насколько это возможно.

Так как в дальнейшем граф G' будет использован для поиска кратчайших по весу путей, для G' требуется выполнение следующих условий.

Условие 1. Если в результате k экспериментов мы получаем множество различных по значению $s_{pq}{}^k$, то в графе G' нам требуется обеспечить получение только одного из них $s_{pq} = = \min(s_{pq}{}^k)$.

Условие 2. Если в ходе эксперимента над G для вершин v_i, v_j, v_h выполняется соотношение $s_{ih}+s_{hj} < s_{ij}$, то в графе G' значение s_{ij} следует получить в виде $s_{ij} = s_{ih}+s_{hj}$.

Для простоты рассмотрим случай, когда граф G' односвязный и неориентированный.

Решение состоит из двух этапов.

Этап 1. Подготовка.

В том случае, если в ходе экспериментов не было ни одного прохождения из вершины v_i в вершину v_j, то полагаем $s_{ij} = \infty$. Так как граф G' неориентированный и $a_{ij} = = a_{ji}$, строим исходную матрицу $A = (a_{ij})$ по следующему правилу: $a_{ij} = \min_k\left(s^k_{ij}, s^k_{ji}\right)$, где k – номер эксперимента. Диагональные элементы вида a_{ii} приравниваем ∞. Строим матрицу $R = A$ (т. е. $r_{ij} = a_{ij}$ для всех пар i, j).

Этап 2. Модификация матрицы R

Шаг 1. Для каждого столбца j найдем значение $b_j = \min_{i=1,n} a_{ij}$. Найдем $b = \min_{j=1,n} b_j$. Так как граф G' неориентированный, то у группы столбцов (и не менее чем у двух) b совпадет с b_j, обозначим любые два столбца из группы как l и m, т. е. $b_l = b_m = b$.

Шаг 2. Формируем вектор D, в котором $d_j = = a_{lj} - a_{mj}$.

Шаг 3. Для каждого элемента j строки l в том случае, если $a_{lj} - a_{mj} \geq d_j$, присваиваем $r_{lj} = = \infty$. Для каждого элемента j строки m, в том случае, если $a_{lj} - a_{mj} \leq -d_j$, присваиваем $r_{mj} = \infty$.

Шаг 4. Присваиваем $b_l = b_m = \infty$.

Шаг 5. Повторяем шаги 1–4 до тех пор, пока все b_j не обратятся в ∞ или не останется один элемент b_j, отличный от ∞.

Матрица R есть искомая матрица связности искомого графа G'.

Конец решения.

ПРОВЕРКА НА ДОСТОВЕРНОСТЬ

В реальности приходится сталкиваться с той проблемой, что исходная информация (множества объектов N и расстояний между ними x) может содержать ошибки. Соответственно, ошибки будет содержать полученный граф G_N. На практике ошибки находятся и исключаются в ходе эксплуатации. Тем не менее, после того как граф G_N будет найден, его достоверность можно повысить на основе следующих соображений. Если при получении матрицы достижимости A, выполняя правило $a_{ij} = \min_k \left(s_{ij}^k, s_{ji}^k\right)$, воспользоваться приведенным выше «специальным способом», следует обратить внимание на п. 3 этого способа. В случаях, когда условие $s_{i,j,k} < a_{i,j}$ выполняется для $a_{i,j} = \min\{x_{i,j}\}$, следует отметить $x_{i,k}$, если $a_{i,k} = \min\{x_{i,k}\}$ и отметить $x_{k,j}$, если $a_{k,j} = \min\{x_{i,j}\}$. Элементы, получившие наибольшее число отметок, следует проверить на достоверность. Практическая транспортная задача строится на основе реальной дорожной сети, которая имеет следующие особенности, исходящие из практического смысла: а) угол между соседними ребрами не может быть малым; б) она может быть представлена как планарный граф. Первое соображение приводит к следующему. С учетом анализа существующей сети выбирается некоторый минимальный допустимый угол между ребрами. В матрице связности из ребер выбираются треугольники. По формуле косинуса по трем ребрам вычисляются углы при каждой из пар ребер. Ребра при малых углах анализируются на предмет достоверности. Полученный граф корректируется по результатам анализа. Полученный граф исследуется на планарность. Если он не планарный, это означает либо ошибку в исходных данных, либо отсутствие узла там, где он необходим. По итогам изучения реальной сети делается соответствующая коррекция графа.

ЗАКЛЮЧЕНИЕ

Представленные алгоритмы позволяют построить аппроксимирующий граф, исключив при этом наиболее явные ошибки в исходных данных.

СПИСОК ЛИТЕРАТУРЫ

1. **Dijkstra E. W.** A note on two problems in connexion with graphs // Numerische Mathematik. 1959. V. 1. P. 269–271.

2. Single-Source Shortest Paths and All-Pairs Shortest Paths / T. H. Cormen [et al.] // Introduction to Algorithms. MIT Press and McGraw-Hill 2001. P. 580–642.

ОБ АВТОРАХ

Житников Владимир Павлович, проф., засл. деятель науки РБ, зав. каф. компьют. математики. Дипл. инженер-физик (МФТИ, 1973). Д-р физ.-мат. наук по механике жидкости, газа и плазмы (Казанск. ун-т, 1993). Иссл. в обл. волновых течений жидкости, электрохим. формообразования, числ.-аналит. методов.

Ураков Айрат Ренатович, доц. той же каф. Дипл. инженер-системотехник (МГТУ, 1993). Канд. физ.-мат. наук по применению выч. техники, мат. моделирования и мат. методов в научных исследованиях (БашГУ, 1997). Иссл. в обл. применения численных методов и верификации данных.

Тимеряев Тимофей Валерьевич, асп. той же каф. Магистр прикл. математики и информатики (УГАТУ, 2011). Иссл. в обл. теории графов.

METADATA

Title: Veracity estimation and approximation of the information provided in the form of weighted graph

Authors: V.P. Zhitnikov, A.R. Urakov, T.V. Timeryaev

Affiliation:
Ufa State Aviation Technical University (UGATU), Russia.

Email: urakov@ufanet.ru.

Langage: Russian.

Source: Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), Vol. 17, No. 2 (55), pp. 50-52, 2013. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: Problems of approximation of a weighted graph by another weighted graph are considered. The problems are complicated by the fact that source information is not in the form of a graph, but in the form of objects sets and relationships between these objects. In addition, the raw data may contain errors that need to be identified and excluded from the resulted graph.

Key words: graph approximation, estimation of graph veracity, graph generation, distance matrix.

References (English Transliteration):

1. Dijkstra E. W. A note on two problems in connexion with graphs / / Numerische Mathematik. 1959. V. 1. P. 269-271.

2. Single-Source Shortest Paths and All-Pairs Shortest Paths / T. H. Cormen [et al.] / / Introduction to Algorithms. MIT Press and McGraw-Hill in 2001. R. 580-642.

About authors:

Zhitnikov, Vladimir Pavlovich, prof., Honored. of Science of Belarus, Head. Department. Compute. mathematics. Dipl. engineer and physicist (MIPT, 1973). Dr. Sci. Science in fluid mechanics, gas and plasma (Kazansk. University Press, 1993). Tests were performed. in the region. wave flows of liquids, electrochemical. forming, Num. analyte. methods.

Urakov, Ayrat Renatovich, Assoc. the same department. Dipl. System Engineer (Bauman, 1993). Candidate. Sci. Science Application calc. equipment mat. modeling and mat. methods in research (BSU, 1997). Tests were performed. in the region. the application of numerical methods and data verification.

Timeryaev, Timothy V., pg. the same department. Master of Applied. Mathematics and Computer Science (USATU, 2011). Tests were performed. in the region. graph theory.