ДОКЛАДЫ АН ВШ РФ
2015 июль–сентябрь 3(28)
ТЕХНИЧЕСКИЕ НАУКИ

УДК 519.242.5

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ ДЛЯ ОЦЕНКИ МАТРИЦЫ ТРАНСПОРТНЫХ КОРРЕСПОНДЕНЦИЙ

В.И. Хабаров[1], А.А. Теселкин[2], К.П. Косолапов[1]
[1]*Сибирский государственный университет путей сообщения*
[2]*Новосибирский государственный технический университет*

Задача планирования наблюдений для оценки транспортных корреспонденций рассматривается как задача распределения ресурса на узлах транспортной сети. Транспортная сеть представляется как граф, вершины которого ассоциированы с узлами транспортной сети, а дуги – с возможными путями сообщений. Задача распределения ресурса для наблюдения решается с применением методов оптимального планирования эксперимента. Используются минимаксные D-оптимальные планы. Модель наблюдения предполагает фиксацию количества переходов транспортных средств из одной вершины транспортного графа в другую. Данная модель описывается цепью Маркова с дискретным временем. Матрица переходных вероятностей цепи оценивается на основе наблюдений за цепью в дискретные моменты времени. Для оценки переходных вероятностей используется метод максимального правдоподобия в предположении, что марковская цепь стационарна. Строится информационная матрица Фишера для предлагаемой модели наблюдений. Решение задачи планирования приводится в общем аналитическом виде. Предлагается интерпретация для задачи подсчета интенсивности транспорта и некоторые рекомендации для использования результатов в практических целях.

Ключевые слова: транспортная сеть, матрица корреспонденции, марковские цепи, задача планирования эксперимента, оценка матриц переходных вероятностей по агрегированным данным.

DOI: 10.17212/1727-2769-2015-3-109-116

Введение

Матрица корреспонденции является фундаментальной характеристикой транспортной сети, поскольку она определяет общий объем транспортного потока. Поэтому задача оценки матрицы корреспонденции является важнейшей задачей для исследования транспортной сети и транспортных потоков [1]. Один из методов оценки матрицы корреспонденций заключается в восстановлении матрицы корреспонденции на основе наблюдений за транспортным потоком [2]. В качестве наблюдений можно рассматривать интенсивности транспортных потоков на участках сети. Для качественной реализации методов оценки необходимо, чтобы наблюдения были максимально информативны. Возникает задача оптимального планирования наблюдений.

Таким образом, в работе рассматривается задача планирования наблюдений за транспортными потоками для оценки транспортных корреспонденций.

1. Марковская модель транспортных корреспонденций

Имеется граф транспортной сети, или транспортный граф $G(V,E)$, состоящий из m узлов. Данный граф описывается стационарной неприводимой апериодической марковской цепью с дискретным временем и с матрицей переходных вероятностей $P=\{p_{ij}\}$, $(i,j=1,\ldots,m)$, причем каждое состояние цепи ассоциировано с некоторой вершиной графа G. Рассматривается задача подсчета трафика для автотранспорта на транспортной сети G.

С точки зрения теории марковских цепей рассматривается задача оценивания переходных вероятностей на основе наблюдений за цепью в моменты времени $t = \{0, 1, \ldots, T\}$ на основе статистики

$$n_{ij} = \sum_{t=0}^{T} n_{ij}(t),$$

где $n_{ij}(t)$ – количество переходов цепи из состояния i в состояние j в момент времени t. Пусть

$$n_i = \sum_{j=1}^{m} n_{ij}$$

– общее число переходов цепи за время T в состоянии i,

$$N = \sum_{i=1}^{m} n_i$$

– общее число переходов за время T.

Далее будет полезна следующая интерпретация для данной модели наблюдения. Рассматриваются некоторые микрообъекты, переходящие из одного состояния в другое в моменты времени $t = \{0, 1, \ldots, T\}$. Имеются наблюдатели, которые ассоциированы с состояниями цепи. Каждый наблюдатель в состоянии с номером i в некоторый момент времени t фиксирует $n_{ij}(t)$ – количество переходов из состояния с номером i в момент t в состояние с номером j в момент $t + 1$. Для задачи подсчета трафика наблюдатель, находящийся в узле i, в некоторые дискретные моменты времени $t = \{0, 1, \ldots, T\}$ фиксирует перераспределение автотранспорта.

Здесь уместно отметить специфику транспортных сетей и задач, связанных с оценкой матриц переходных вероятностей для этих сетей. Транспортный граф с известной структурой дает априорную информацию о том, какие переходы в каком состоянии разрешены. Для учета этой информации далее будет использоваться понятие «емкость состояния». Под емкостью состояния $s_i, i = 1, \ldots, m$, будем понимать количество выходящих дуг m_i из вершины графа сети с номером i. Очевидно, что

$$\sum_{j=1}^{m_i} p_{ij} = 1,\ p_{ij} > 0,\ i = 1, \ldots, m. \tag{1}$$

Это избавит в дальнейшем от необходимости оговаривать условия использования выражений типа p_{ij}^{-1} или $\log p_{ij}$.

В качестве метода оценивания матрицы $P = \{p_{ij}\},\ (i, j = 1, \ldots, m)$ рассмотрим *метод максимального правдоподобия*. Функция правдоподобия в этом случае есть [3–6]

$$L(n,P) = \prod_{i=1}^{m}\prod_{j=1}^{m_i} p_{ij}^{\,n_{ij}} = \prod_{i=1}^{m} p_{i1}^{n_{i1}}\, p_{i2}^{n_{i2}} \ldots p_{im_i-1}^{n_{im_i-1}} \left(1 - \sum_{j=1}^{m_i-1} p_{ij}\right)^{n_i - \sum_{j=1}^{m_i-1} n_{ij}}. \tag{2}$$

Логарифмическая функция правдоподобия, соответственно, будет иметь вид

$$\log L(n,P) = \sum_{i=1}^{m}\sum_{j=1}^{m_i-1} n_{ij}\log p_{ij} + \left(n_i - \sum_{j=1}^{m_i-1} n_{ij}\right)\log\left(1-\sum_{j=1}^{m_i-1} p_{ij}\right). \tag{3}$$

Оценка максимального правдоподобия (ОМП) является решением уравнения

$$\frac{\partial}{\partial p_{ij}}\log L(n,P) = \frac{n_{ij}}{p_{ij}} - \frac{n_i - \sum_{j=1}^{m_i-1} n_{ij}}{1-\sum_{j=1}^{m_i-1} p_{ij}} = 0,\ \ i=1,\ldots,m;\, j=1,\ldots,m_i-1, \tag{4}$$

при условии (1). Решением (4) относительно p_{ij} являются оценки вида

$$p_{ij}^* = \frac{n_{ij}}{n_i}, i=1,\ldots,m;\, j=1,\ldots,m_i. \tag{5}$$

Такая оценка состоятельна и асимптотически нормальна [3–6].

2. Планирование наблюдений

Задачу планирования экспериментов для модели наблюдений, рассмотренной выше, можно интерпретировать следующим образом. Наблюдатели фиксируют переходы микрообъектов, находясь в состояниях $\{s_1,\ldots,s_m\}$ в моменты времени $t=\{0,1,\ldots,T\}$. На весь эксперимент отведен ресурс в N наблюдений. Каждый наблюдатель получает часть этого ресурса $\{n_i,\ i=1,..,m\}$ таким образом, что $\sum_{i=1}^{m} n_i = N$. Требуется найти распределение $n=\{n_i, i=1,..,m\}$, максимизирующее некоторый функционал от информационной матрицы Фишера

$$I(P,n) = -E\left(\frac{\partial^2 \log L(n,P)}{\partial P^2}\right). \tag{6}$$

Известно [7], что D-оптимальный план минимизирует обобщенную дисперсию оценок параметров, полученных на основе уравнения (4). Следует, однако, ожидать (смотри далее), что информационная матрица Фишера будет зависеть от истинных значений параметров P. Для исключения этой зависимости далее будем рассматривать минимаксные D-оптимальные планы, для которых

$$n^* = Arg\max_n \min_P \log\det I(P,n). \tag{7}$$

Рассмотрим структуру информационной матрицы Фишера более детально [9]. Для этого сначала необходимо найти элементы матрицы вторых производных:

$$\frac{\partial^2 \log L(n,P)}{\partial p_{ik}\partial p_{il}} = \begin{cases} -\dfrac{n_{ik}}{p_{ik}^2} - \dfrac{n_{im_i}}{p_{im_i}^2},\ k=l, \\[2ex] -\dfrac{n_{im_i}}{p_{im_i}^2},\ k\neq l, \end{cases} \qquad i=1,\ldots,m;\ k,l=1,\ldots,m_i. \tag{8}$$

Далее согласно (6)

$$I(P,n)=E\begin{bmatrix} A_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & A_m \end{bmatrix}, \tag{9}$$

где с учетом (8)

$$A_i=\begin{bmatrix} \dfrac{n_{i1}}{p_{i1}^2}+\dfrac{n_{im_i}}{p_{im_i}^2} & \cdots & \dfrac{n_{im_i}}{p_{im_i}^2} \\ \vdots & \ddots & \vdots \\ \dfrac{n_{im_i}}{p_{im_i}^2} & \cdots & \dfrac{n_{im_i-1}}{p_{im_i-1}^2}+\dfrac{n_{im_i}}{p_{im_i}^2} \end{bmatrix} \tag{10}$$

для всех $i=1,...,m$. Выше предполагалось, что

$$p_{im_i}=1-\sum_{j=1}^{m_i-1}p_{ij},$$

$$n_{im_i}=n_i-\sum_{j=1}^{m_i-1}n_{ij}.$$

Для мультиномиального распределения известно (см. [8]), что $E[n_{ij}]=n_i p_{ij}$ $\forall i=1,\ldots,m;\ j=1,...,m_i$, отсюда следует, что

$$E[A_i]=n_i\begin{bmatrix} \dfrac{1}{p_{i1}}+\dfrac{1}{p_{im_i}} & \cdots & \dfrac{1}{p_{im_i}} \\ \vdots & \ddots & \vdots \\ \dfrac{1}{p_{im_i}} & \cdots & \dfrac{1}{p_{im_i-1}}+\dfrac{1}{p_{im_i}} \end{bmatrix}=$$
$$=n_i\left(\begin{bmatrix} \dfrac{1}{p_{i1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \dfrac{1}{p_{im_i-1}} \end{bmatrix}+\begin{bmatrix} \dfrac{1}{p_{im_i}} & \cdots & \dfrac{1}{p_{im_i}} \\ \vdots & \ddots & \vdots \\ \dfrac{1}{p_{im_i}} & \cdots & \dfrac{1}{p_{im_i}} \end{bmatrix}\right)\forall i=1,\ldots,m. \tag{11}$$

Далее необходимо найти выражение для $\log\det I(P,n)$. Учитывая блочно-диагональную структуру матрицы (9), имеем:

$$\log\det I(P,n)=\log\prod_{i=1}^{m}\det E\left[A_i\right]=\sum_{i=1}^{m}\log\det E\left[A_i\right]. \tag{12}$$

Воспользуемся результатом [10] для вычисления определителя матрицы, которая представима в форме (11), известной как диагональное разложение. Тогда

$$\det E[A_i] = n_i^{m_i-1}\left(\prod_{j=1}^{m_i-1}\frac{1}{p_{ij}} + \sum_{k=1}^{m-1}\prod_{j\neq k}^{m_i-1}\frac{1}{p_{im}}\frac{1}{p_{ij}}\right) = n_i^{m_i-1}B_i, \tag{13}$$

где

$$B_i = \prod_{j=1}^{m_i-1}\frac{1}{p_{ij}} + \sum_{k=1}^{m-1}\prod_{j\neq k}^{m_i-1}\frac{1}{p_{im}}\frac{1}{p_{ij}}.$$

Последнее выражение дает основание утверждать, что

$$\log\det I(P,n) = (m_i-1)\sum_{i=1}^{m}\log n_i + \sum_{i=1}^{m}\log B_i . \tag{14}$$

Возвращаясь к экстремальной задаче (7), найдем

$$\min_P \log\det I(P,n) = \min_P\left((m_i-1)\sum_{i=1}^{m}\log n_i + \sum_{i=1}^{m}\log B_i\right). \tag{15}$$

Учитывая выпуклость логарифмической функции, а также независимость первого слагаемого в (14) от матрицы P, можно утверждать, что решением экстремальной задачи (15) будут стохастические матрицы P размерности m, с равными элементами в строках.

Таким образом, экстремальную задачу (7) можно свести к более простой экстремальной задаче вида

$$n^* = \operatorname{Arg}\max_{\sum_{i=1}^{m} n_i = N,\, n_i>0}\left(m_i-1\right)\sum_{i=1}^{m}\log n_i. \tag{16}$$

Используя функцию Лагранжа

$$L\ (n,\lambda) = (m_i-1)\sum_{i=1}^{m}\log n_i + \lambda\left(N - \sum_{i=1}^{m} n_i\right),$$

найдем решение (16), решая систему уравнений

$$\begin{cases} \dfrac{\partial L}{\partial n_i} = \dfrac{(m_i-1)}{n_i} - \lambda = 0 \ \ \forall i = 1,\ldots,m, \\ \dfrac{\partial L}{\partial \lambda} = \sum_{i=1}^{m} n_i - N = 0. \end{cases} \tag{17}$$

Решением (17) является

$$n_i^* = \frac{N(m_i-1)}{\sum_{i=1}^{m}(m_i-1)}\ \forall\, i = 1,\ldots,m. \tag{18}$$

Таким образом, план эксперимента (распределение ресурса) можно интерпретировать следующим образом: общий объем наблюдений N перераспределяется между наблюдателями марковской цепи пропорционально $m_i - 1\ (i = 1,\ldots,m)$, где m_i – количество возможных переходов в состоянии s_i.

Интерпретация для транспортной задачи может быть следующей: если дана транспортная сеть с некоторым количеством узлов не меньше, чем m, то следует ограничиться m узлами и N возможных наблюдений за микрообъектами перераспределить согласно (18).

Заключение

В данной работе рассмотрена проблема организации мониторинга транспортных потоков с целью оценки транспортных корреспонденций, которая сводится к задаче распределения ресурса по узлам марковской цепи с дискретным временем. Такая задача была интерпретирована как задача оптимального планирования наблюдений за марковской цепью. В работе поставлена задача планирования экспериментов с использованием минимаксных D-оптимальных планов (7). Получена информационная матрица Фишера (9). Решение задачи планирования получено в аналитическом виде.

Результаты, полученные в данной статье, могут быть использованы в практических целях для оценки матриц корреспонденции, для калибровки транспортных моделей и для проведения системного мониторинга интенсивностей транспорта, что увеличит информативность собираемой информации о транспортных потоках.

ЛИТЕРАТУРА

1. Введение в математическое моделирование транспортных потоков: учебное пособие / А.В. Гасников, С.Л. Кленов, Е.А. Нурминский, Я.А. Холодов, Н.Б. Шамрай, М.Л. Бланк, Е.В. Гасникова, А.А. Замятин, В.А. Малышев, А.В. Колесников, А.М. Райгородский; под ред. А.В. Гасникова. – М.: МФТИ, 2010. – 360 с.
2. **Хабаров В.И., Молодцов Д.О., Хомяков С.В.** Марковская модель транспортных корреспонденций // Доклады ТУСУР. – 2012. – № 1, ч. 1. – С. 113–117.
3. **Ли Ц.-Ч., Джадж Д.Г., Зельнер А.** Оценивание параметров марковских моделей по агрегированным временным рядам. – М.: Статистика, 1977. – 221 с.
4. **Kendall M.G., Stuart A.** The advance theory of statistics. Vol. 2. Inference and Relationship. – London: Carles Griffin and Company Limited, 1961. – 758 p.
5. **Hoel P.G.** A test for Markoff chains // Biometrika. – 1954. – Vol. 41, iss. 3–4. – P. 430–433. – doi: 10.1093/biomet/41.3-4.430.
6. **Кельберт М.Я., Сухов Ю.М.** Вероятность и статистика в примерах и задачах. Т. 2. Марковские цепи как отправная точка теории случайных процессов и их приложения. – М.: МЦНМО, 2010. – 560 с.
7. **Федоров В.В.** Теория оптимального планирования эксперимента. – М.: Наука, 1971. – 312 с.
8. **Хабаров В.И.** Марковская модель процесса последовательного планирования экспериментов // Машинные методы планирования эксперимента и оптимизации многофакторных систем: межвузовский сборник научных трудов. – Новосибирск: Изд-во НЭТИ, 1987. – С. 62–65.
9. Справочник по прикладной статистике. В 2 т. Т. 1: пер. с англ. / под ред. Э. Ллойда, У. Ледермана; пер. с англ. под ред. Ю.Н. Тюрина. – М.: Финансы и статистика, 1989. – 510 с.
10. **Сирл С., Госман У.** Матричная алгебра в экономике / пер. с англ. и науч. ред. Е.М. Четыркина и Р.М. Энтова. – М.: Статистика, 1974. – 368 с.

DESIGN OF EXPERIMENTS FOR TRANSPORT CORRESPONDENCE MATRIX ESTIMATION

Khabarov V.I.[1] , Tesselkin A.A.[2], Kosolapov K.P.[1]
[1]*Siberian Transport University, Novosibirsk, Russian Federation*
[2]*Novosibirsk State Technical University, Novosibirsk, Russian Federation*

The problem of study planning for transport correspondence estimation is considered as a problem of resource distribution at the nodes of the transport network. The transport network is considered as a graph whose vertices are associated with the nodes of the transport network, and edges are associated with possible routes. The resource distribution problem is solved using optimal experimental design methods. Minimax D-optimal designs are used. Recording the number of vehicle transitions from one vertex of the transport graph to another is taken as an observation model. This model is described by the Markov discrete time chain. The transition matrix is estimated using observations of the chain at discrete time points. The maximum likelihood method is used for the estimation of transition probabilities assuming that the Markov chain is steady. The Fisher information matrix for the proposed model is constructed. The solution of the experimental design problem is given in general analytical form. The interpretation for calculating the traffic intensity is proposed and some recommendations for practical applications of the results are given.

Keywords: transport network, correspondence matrix, Markov chains, design of experiments, estimation of correspondence under aggregated data.

DOI: 10.17212/1727-2769-2015-3-109-116

REFERENCES

1. Gasnikov A.V., Klenov S.L., Nurminskiĭ E.A., Kholodov Ya.A., Shamraĭ N.B, Blank M.L., Gasnikova E.V., Zamyatin A.A., Malyshev V.A., Kolesnikov A.V., Raĭgorodskiĭ A.M. *Vvedenie v matematicheskoe modelirovanie transportnykh potokov* [Introduction to the mathematical modelling of traffic flows]. Moscow, MFTI Publ., 2010. 360 p.
2. Khabarov V.I., Molodtsov D.O. Khomyakov S.V. Markovskaya model' transportnykh korrespondentsii [Model Markov chains for transport correspondence]. *Doklady TUSUR – Proceedings of TUSUR*, 2012, no. 1, pp. 113–117.
3. Lee T.C., Judge G.G., Zellner A. *Estimating the parameters of the Markov probability model from aggregate time series data*. Amsterdam, London, Publishing Company, 1970. 254 p. (Russ. ed.: Li Ts.-Ch., Dzhadzh D.G., Zel'ner A. *Otsenivanie parametrov markovskikh modelei po agregirovannym vremennym ryadam*. Translated from English A.D. Kasavin, V.A. Lototskii, A.S. Mandel'. Moscow, Statistika Publ., 1977. 221 p.).
4. Kendall M.G., Stuart A. *The advance theory of statistics*. Vol. 2. *Inference and Relationship*. London, Carles Griffin and Company Limited, 1961. 758 p.
5. Hoel P.G. A test for Markoff chains. *Biometrika*, 1954, Vol. 41, iss. 3–4, pp. 430–433. doi: 10.1093/biomet/41.3-4.430
6. Kel'bert M.Ya., Sukhov Yu.M. *Veroyatnost' i statistika v primerakh i zadachakh*. T. 2. *Markovskie tsepi kak otpravnaya tochka teorii sluchainykh protsessov i ikh prilozheniya* [Probability and Statistics in the examples and problems. Vol. 2. Markov chain as the starting point of the theory of stochastic processes and their applications]. Moscow, MTsNMO Publ., 2010. 560 p.
7. Fedorov V.V. *Teoriya optimal'nogo planirovaniya eksperimenta* [The theory of optimal design of experiments]. Moscow, Nauka Publ., 1971. 312 p.
8. Khabarov V.I. Markovskaya model' protsessa posledovatel'nogo planirovaniya ekspe-rimentov [Markov model of sequential design of experiments]. *Mashinnye metody planirovaniya eksperimenta i optimizatsii mnogofaktornykh system*: mezhvuzovskii sbornik nauchnykh trudov [Machine methods of the expermental design and the multifactor system optimization: interuniversity collection of scientific papers], 1987, pp. 62–65.
9. Ledermann W., ghief ed. *Handbook of applicable mathematics*. Vol. 6, pt. A. *Statistics*. Ed. by E. Lloyd. Chichester, John Wiley & Sons, 1984. 580 p. (Russ. ed.: Lloid E., Lederman U., eds. *Spravochnik po prikladnoi statistike*. V 2 t. T. 1. Translated from English, ed. by Yu.N. Tyurin. Moscow, Finansy i statistika Publ., 1989. 510 p.).

10. Searle S.R., Hausman W.H. *Matrix algebra for business and economics*. New York, Wiley, 1970. 362 p. (Russ. ed.: Sirl S., Gosman U. *Matrichnaya algebra v ekonomike*. Translated from English and sci. eds. E.M. Chetyrkin, R.M. Entov. Moscow, Statistika Publ., 1974. 368 p.).

СВЕДЕНИЯ ОБ АВТОРАХ

Хабаров Валерий Иванович – родился в 1951 году, д-р техн. наук, профессор, член-корреспондент Академии Высшей школы, заведующий кафедрой «Информационные технологии транспорта» Сибирского государственного университета путей сообщения. Область научных интересов: искусственный интеллект, математическое моделирование транспортных потоков, планирование эксперимента. Опубликовано около 120 научных работ. (Адрес: 630049, Россия, Новосибирск, ул. Дуси Ковальчук, 191. Email: khabarov51@mail.ru).

Khabarov Valeriy Ivanovich (b. 1951) – Doctor of Sciences (Eng.), professor, corresponding member of the Russian Higher School Academy of Sciences, Head of IT in Transport Department of Siberian Transport University. His research interests are currently focused on artificial intelligence, mathematical modelling of traffic flows, design of experiments. He is author of about 120 scientific papers. (Address: 191 D. Kovalchuk St., Novosibirsk, 630049, Russian Federation. Email: khabarov51@mail.ru).

Теселкин Александр Александрович – родился в 1992 году, аспирант Новосибирского государственного технического университета. Область научных интересов: математическое моделирование транспортных потоков. Опубликовано 8 научных работ. (Адрес: 630073, Россия, Новосибирск, пр. Карла Маркса, 20. Email: a.tesselkin@gmail.com).

Tesselkin Alexandr Alexandrovich (b. 1992) – postgraduate of Novosibirsk State Technical University. His research interests are currently focused on mathematical modelling of traffic flows. He is author of 8 scientific papers. (Address: 20, Karl Marks Av., Novosibirsk, 630073, Russia. Email: a.tesselkin@gmail.com).

Косолапов Кирилл Павлович – родился в 1991 году, аспирант Сибирского государственного университета путей сообщения. Область научных интересов: математическое моделирование транспортных потоков. Опубликованы 4 научные работы. (Адрес: 630049, Россия, Новосибирск, ул. Дуси Ковальчук, 191. Email: kosolapovkp1@gmail.com).

Kosolapov Kirill Pavlovich (b. 1991) – postgraduate of Siberian Transport University. His research interests are currently focused on mathematical modelling of traffic flows. He is author of 4 scientific papers. (Address: 191 D. Kovalchuk, Novosibirsk, 630049, Russia. Email: kosolapovkp1@gmail.com).

Статья поступила 05 августа 2015 г.
Received August 05, 2015

To Reference:

Khabarov V.I., Tesselkin A.A., Kosolapov K.P. Planirovanie eksperimentov dlya otsenki matritsy transportnykh korrespondentsii [Design of experiments for transport correspondence matrix estimation]. *Doklady Akademii nauk vysshei shkoly Rossiiskoi Federatsii – Proceedings of the Russian higher school Academy of sciences*, 2015, no. 3 (28), pp. 109-116. doi: 10.17212/1727-2769-2015-3-109-116