УДК 51:[656.11+351.811.12]
DOI 10.17513/snt.39759

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ДЛЯ ОБНОВЛЕНИЯ МАТРИЦЫ КОРРЕСПОНДЕНЦИЙ ПРИ ВВЕДЕНИИ НОВЫХ ПУНКТОВ ПРИТЯЖЕНИЯ ПОТОКОВ

Наумова Н.А.

ФГБОУ ВО «Кубанский государственный технологический университет», Краснодар, e-mail: Nataly_Naumova@mail.ru

Изучение динамики транспортных и пешеходных потоков, прогнозирование их характеристик необходимо при формировании и оптимизации транспортной инфраструктуры городов. Основную нагрузку на элементы улично-дорожной сети составляют регулярные по времени потоки транспортных средств между парами районов отправлений – прибытий, информация о которых хранится в матрице корреспонденций. С этой точки зрения актуальной задачей представляется разработка методов определения и динамического обновления матрицы корреспонденций при введении новых потокообразующих элементов. В частности, методов сбора и обработки необходимой информации о потоках, методов идентификации параметров транспортных и пешеходных потоков. Целью данной работы является разработка вероятностно-статистических методов планирования наблюдений для оценки обновления данных о корреспонденциях, вызванных новым слоем спроса. Существуют физические и технологические ограничения при наблюдении за транспортной сетью, которые следует учитывать при планировании эксперимента. Разработан алгоритм для определения оптимального расположения точек наблюдения при уточнении матрицы корреспонденции с учетом энтропии системы. Определено необходимое число измерений для заданной надежности получения параметров распределения транспортного потока. Разработан метод оценки изменений в матрице транспортных корреспонденций, вызванных появлением нового слоя спроса.

Ключевые слова: матрица корреспонденций, математическая модель, транспортный поток, вероятностно-статистические методы

PLANNING THE EXPERIMENT TO UPDATE THE ORIGIN-DESTINATION DEMAND MATRIX WHEN INTRODUCING NEW ITEMS TRAFFIC DEMAND

Naumova N.A.

Kuban State Technological University, Krasnodar, e-mail: Nataly_Naumova@mail.ru

Analysis of the dynamics of growth and pedestrian flows, forecasting The main load on the elements of the road network is met by regular time flows of vehicles between pairs of nearby departures and arrivals, information about which is found in the OD-matrix. From this point of view, the development of methods for determining and dynamically updating the OD-matrix with the introduction of new flow-forming elements seems to be an urgent task. In particular, methods for collecting and processing the necessary information about flows, methods for identifying the parameters of traffic and pedestrian flows must be explored. The purpose of this work is to develop probabilistic-statistical methods for planning observations to evaluate the update of correspondence data caused by a new demand layer. There are physical and technological limitations when observing the transport network, which should be taken into account when planning an experiment. An algorithm has been developed to determine the optimal location of observation points when refining the correspondence matrix, taking into account the entropy of the system. The required number of measurements for a given reliability of obtaining traffic flow distribution parameters is determined. A method has been developed for assessing changes in the matrix of transport correspondence caused by the emergence of a new layer of demand.

Keywords: OD-matrix, mathematical model, traffic flow, probabilistic-statistical methods

Изучение динамики транспортных и пешеходных потоков, прогнозирование их характеристик необходимо при формировании и оптимизации транспортной инфраструктуры городов. Информация о распределении транспортных потоков по улично-дорожной сети важна для управления дорожным движением, в частности с помощью интеллектуальных транспортных систем. Основную нагрузку на элементы улично-дорожной сети составляют регулярные по времени потоки транспортных средств между парами районов отправлений – прибытий, информация о которых хранится в матрице корреспонденций. С этой точки зрения актуальной задачей представляется разработка методов определения и динамического обновления матрицы корреспонденций при введении новых потокообразующих элементов, в частности методов сбора и обработки необходимой информации о потоках, методов идентификации параметров транспортных и пешеходных потоков.

Оценка статической матрицы корреспонденций достаточно хорошо изучена. Обычно используется предварительная матрица для получения единственного решения. В целом эти методы могут быть классифицированы следующим образом [1]:

– метод наименьших квадратов (Cascetta E., Nguyen S., 1988; Doblas J., Benitez F.G., 2005) и обобщенный метод наименьших квадратов (Bell M.G.H., 1991; Nie Y.M., Zhang H.M., 2010; Caggiani L., 2013). Как правило, эти методы представляют собой двухуровневые задачи. Верхний уровень должен минимизировать отклонение между наблюдаемым и расчетным количеством трафика за все временные интервалы, а также отклонение между целевым или установленным спросом и расчетными матрицами корреспонденций. Модель нижнего уровня решает статическую задачу пользовательского равновесия;

– методы, основанные на концепции энтропии (например, Van Zuylen H.J., Willumsen L., 1980; Xie C., Kockelman K.M., Waller S.T., 2011). Эти методы максимизируют энтропию с учетом набора ограничений. Концепция энтропии измеряет, насколько более точно отражает реальную ситуацию текущая матрица корреспонденций по сравнению с предшествующей;

– методы максимального правдоподобия (например, Spiess H., 1987; Parry M., Hazelton M.L., 2012). В этом случае предполагается, что элементы исходной матрицы корреспонденций получены из набора случайных величин с заданным распределением вероятностей;

– Байесовские методы (например, Maher M.J., 1983; Perrakis K., Karlis D., Cools M., Janssens D., Vanhoof K., Wets G., 2012; Wei C., Asakura Y., 2013; Castillo E., Menéndez J.M., 2008; Cheng L., Zhu S., Chu Z., Chen J., 2014), которые используют наблюдаемые подсчеты трафика для обновления предполагаемого априорного распределения, апостериорное распределение всех переменных строится на основе теоремы Байеса. Эти методы рассматривают поток трафика как случайную величину.

Оценка зависящего от времени спроса на корреспонденции гораздо сложнее случая статического. Некоторые исследователи прямо распространяют методы статической оценки матрицы корреспонденций на динамический случай с использованием изменяющихся во времени подсчетов данных (например, Cascetta E., Inaudi D., Marquis G., 1993; Tavan H., Mahmassani H.S., 2001; Bierlaire M., Crittin F., 2004).

Однако, несмотря на большое количество исследований [2, 3], изучение методов составления и обновления матрицы корреспонденций остается актуальной задачей, так как в связи с развитием новых технологий появляются новые возможности по сбору данных. Кроме того, для разных моделей транспортной сети и разных транспортных задач требуются различные исходные данные, а также приемлемы различные методы обновления или получения информации.

Целью данной работы является разработка вероятностно-статистических методов планирования наблюдений для оценки обновления данных о корреспонденциях, вызванных новым слоем спроса.

Материалы и методы исследования

Матрица корреспонденций (OD-матрица) содержит информацию о числе транспортных средств, проходящих от «источника» (места отправления) к «стоку» (месту прибытия) в течение времени t . Матрицу пропорциональных отношений (link proportions), определяющую часть трафика на каждой из дуг, относящихся к определенной паре «источник – сток», обозначают как LP-матрицу. Ранее [4] автором был разработан и обоснован метод определения динамической матрицы корреспонденций, построенной с учетом исходных данных, требующихся для расчетов с помощью авторской модели TIMeR_Mod [5]. Модель основана на оценке маршрутов транспортных потоков: устанавливается зависимость между элементами матрицы корреспонденций и стоимостью движения по маршруту. Основная идея заключается в том, чтобы использовать стоимость движения по маршрутам (функцию транспортных затрат) для установления пользовательского равновесия. Для динамического обновления матрицы корреспонденций применяется фильтр Кальмана. С помощью OD-матрицы и LP-матрицы корректируется распределение интенсивности по дугам транспортной сети, которая в авторской модели TIMeR_Mod содержится в матрицах $A_{STREETS}$ и $B_{INTERSECTION}$ [5].

В данной работе ставится задача планирования эксперимента для обновления данных о распределении транспортных потоков при введении новых пунктов притяжения корреспонденций, то есть в случае появления нового слоя спроса.

Одной из основных проблем является проблема идентификации, связанная с построением математических моделей динамических систем по экспериментальным данным. Ее качественному решению способствует применение на практике математических методов и новых технологий. Суть вероятностно-статистических методов принятия решений состоит в использовании вероятностных моделей на основе оценивания и проверки гипотез с помощью выборочных характеристик. На практике вероятностные и статистические методы обычно применяются в том случае, когда выводы, основанные на выборочных данных, переносятся на всю совокупность [6]. Однако при этом в каждой конкретной ситуации следует предварительно оценить принципиальную возможность получения достаточно достоверных вероятностных и статистических данных.

В данной работе рассматривается задача обновления данных об OD-матрице при появлении нового пункта притяжения корреспонденций. Задача о распределении транспортных потоков предполагает два основных этапа. На первом этапе на основе исходных данных о потокообразующих факторах строится одна или несколько матриц корреспонденций населения, которые определяют ограничения на общий объем передвижений в сети между источниками и стоками движения. Второй этап предполагает равновесное распределение транспортных потоков согласно первому принципу Вардропа.

Существуют физические и технологические ограничения при наблюдении за транспортной сетью. Реализация отдельных моделей может быть очень затратной. Поэтому при планировании эксперимента следует учитывать вышеуказанные ограничения [7].

Современные технологии позволяют использовать различные типы датчиков для сбора данных о транспортных потоках. Сюда можно отнести GPS, видео, автоматическую идентификацию транспортных средств, сканирование номерных знаков и т.д. По типу собираемых датчиками данных их можно разделить на несколько категорий [1]:

− датчики, подсчитывающие количество транспортных средств по полосам дорожной сети (I категория);

− видеодатчики, которые могут снимать видеоизображения движущихся потоков (II категория);

− идентификаторы транспортного средства, с помощью которых можно получать различные данные о трафике, включая маршрут, задержки на перекрестках, время в пути или по отдельным отрезкам пути (III категория).

Синтезировать данные от всех типов источников для оценки корреспонденций позволяет Байесовский метод. Кроме того, доказано, что данный метод хорошо аппроксимирует наблюдаемые данные [1–3].

В данной статье рассматриваются вероятностно-статистические методы планирования наблюдений для оценки обновления данных о корреспонденциях, вызванных новым слоем спроса, не останавливаясь на их физической реализации.

Результаты исследования и их обсуждение

1. Планирование эксперимента для оценивания матрицы корреспонденций

Для оценивания интенсивности движения по полосам транспортной сети, что может использоваться в Байесовском методе как априорная информация для получения оценок, необходимо выбрать узлы транспортной сети (вершины графа – в модели транспортной сети) для установки датчиков I или II категории. В силу ограниченности ресурсов невозможно установить датчики во всех узлах. Поэтому надо выбрать наиболее информативные. В этом случае обратимся к вероятностно-статистическим методам.

Степень неопределенности системы определяется числом ее состояний и вероятностями пребывания в этих состояниях. Характеристикой априорной вероятности системы является ее энтропия.

Энтропией системы [8] называется сумма произведений вероятностей различных состояний системы на логарифмы этих вероятностей с противоположным знаком:

$$H(X) = -\sum_{i=1}^{n} p_i \log_a p_i \,, \qquad (1)$$

где $a > 0$, $a \neq 1$

Или, учитывая определение математического ожидания,

$$H(X) = -\sum_{i=1}^{n} p_i \log_a p_i = M\left(-\log_a p_i\right). \qquad (2)$$

Энтропия сложной системы (X, Y) равна

$$H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_a p_{ij} = M\left(-\log_a p_{ij}\right). \qquad (3)$$

Условная энтропия системы Y при условии $X = x_i$:

$$H(Y \mid X = x_i) = -\sum_{i=1}^{n} P\left(y_j \mid x_i\right) \cdot \log_a P\left(y_j \mid x_i\right) = M_{x_i}\left(-\log_a P(Y \mid x_i)\right). \qquad (4)$$

Средняя или полная энтропия системы Y в зависимости от того, какие состояния принимает система X:

$$H(Y \mid X) = \sum_{i=1}^{n} p_i H\left(Y \mid x_i\right) = M_{x_i}\left(-\log_a P(Y \mid X)\right). \quad (5)$$

Если системы X и Y зависимы, то

$$H(X, Y) = H(X) + H(Y \mid X). \quad (6)$$

Если рассмотреть случай независимости систем, то

$$H(X, Y) = H(X) + H(Y). \quad (7)$$

Другой крайний случай – состояние системы X полностью определяет состояние системы Y. Тогда $H(Y \mid X) = 0$, а энтропия системы равна

$$H(X, Y) = H(X). \quad (8)$$

Количество информации I_X измеряется уменьшением энтропии системы, для уточнения которой предназначены сведения. Количество информации, необходимое для полного выяснения состояния системы: $I_X = H(X)$.

Поставим задачу определения оптимального расположения точек наблюдения при составлении или уточнении матрицы корреспонденции с учетом новых пунктов притяжения. Абстрактный транспортный граф – это совокупность вершин (перекрестков, узлов сети) и ребер (часть улично-дорожной сети между соседними перекрестками).

Пусть K – новый пункт притяжения (новый сток заявок), а D_i, $i = 1,2,\ldots,k$ – возможные источники заявок. Требуется определить наиболее информативные вершины для расположения наблюдателей при условии ограниченных ресурсов, то есть возможность установить не более N_0 наблюдателей.

Пусть определены вершины графа, через которые может проходить маршрут. Их количество равно N. Априори произвольная заявка может пройти с равной вероятностью через любую из этих вершин на пути в K, то есть вероятность пребывания в вершине $X = x_i$ равна, $p_i = 1 / N$, $i = 1,2,\ldots,N$. В этом случае (для слоя $\{D\}_i \leftrightarrow K$) можем считать, что состояние системы X {вся улично-дорожная сеть} полностью определяет система Y.

Пусть имеется m_i направлений, по которым можно продолжить движение от вершины $X = x_i$, причем только s_i из них могут привести в пункт K ($s_i < m_i$). Найдем количество информации о системе, которую можно получить отдельно от каждого из состояний $X = x_i$.

$$I_{x_i} = -\log_a P\left(\left(X = x_i\right)\cdot K\right) = -\log_a \frac{s_i}{m_i N} = \log_a\left(N\frac{m_i}{s_i}\right). \quad (9)$$

Заметим, что $(m_i / s_i) > 1$, логарифм – монотонно возрастающая функция при $a > 0$, $a \neq 1$. Поэтому больше информации несет тот узел сети, из которого выходит больше дуг, ведущих в пункт K.

Предлагается следующий ***алгоритм*** для определения оптимального расположения точек наблюдения при уточнении матрицы корреспонденции для слоя $\{D\}_i \leftrightarrow K$:

1) определяем маршруты как перечень вершин графа, ведущие от каждого из источников D_i, $i = 1,2,\ldots,m$ к стоку – пункту K;

2) включаем все вершины из пункта 1) ***алгоритма*** в множество $X = \{x_i \mid i = 1,2,\ldots,N\}$;

3) определяем количество m_i направлений, по которым можно продолжить движение от вершины $X = x_i$, количество s_i тех из них, которые могут привести в пункт K ($s_i < m_i$);

4) найдем количество информации:

$$I_{x_i} = \log_a\left(N\frac{m_i}{s_i}\right), \ i = 1,2,\ldots,N;$$

5) упорядочим по убыванию величины I_{x_i}, выберем N_0 из них.

Так как $I_X = M\left(-\log_a p_i\right)$, то это обеспечит наибольшую информативность.

Следующая задача: оценить необходимое число измерений, которые следует произвести в каждой вершине, чтобы с заданной надежностью получить параметры распределения транспортного потока.

Рассмотрим оценку вероятности выбрать определенное направление из положения $X = x_i$.

В схеме испытаний случайная величина X принимает значение 1 («успех») с вероятностью p и значение 0 («неудача») с вероятностью $(1-p)$, оптимальной несмещенной оценкой p является относительная частота успешных исходов в n испытаниях:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}. \quad (10)$$

Статистика $n\bar{X}$ имеет биномиальное распределение [9]. Выборочная средняя – асимптотически нормальная величина, у которой $M(\bar{X}) \approx p, \sigma^2(\bar{X}) \approx \frac{p(1-p)}{n}$.

Тогда можно оценить при уровне значимости α (надежности $1-\alpha$) построить асимптотический доверительный интервал:

$$P\left(\frac{|\bar{X}-p|}{\sigma} \le \lambda_\alpha\right) = P\left(\frac{|\bar{X}-p|\sqrt{n}}{\sqrt{p(1-p)}} \le \lambda_\alpha\right) = P\left(\frac{|\bar{X}-p|\sqrt{n}}{\sqrt{\bar{X}(1-\bar{X})}} \le \lambda_\alpha\right) \to 1-\alpha \text{, где } \lambda_\alpha = \Phi^{-1}(\alpha)$$

$$P\left(|\bar{X}-p| \le \lambda_\alpha \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \to 1-\alpha. \quad (11)$$

Следовательно, ошибка при замене вероятности относительной частотой при заданной надежности $1-\alpha$ равна

$$\Delta = \lambda_\alpha \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}. \quad (12)$$

Если требуется, чтобы ошибка не превзошла r% от оценки, то следует решить неравенство

$$\lambda_\alpha \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le r\frac{\bar{X}}{100}. \quad (13)$$

Решив, получим оценку количества опытов для получения необходимой точности оценки:

$$n \ge (\lambda_\alpha)^2 \frac{(1-\bar{X})100^2}{\bar{X} \cdot r^2}. \quad (14)$$

2. Оценивание матрицы корреспонденций

Найдем оценку условной вероятности $P^*(K \mid D_i)$ того, что заявка из пункта D_i выедет в пункт K, с помощью Байесовского подхода [8, 9].

Пусть имеется выборка из распределения Бернулли с параметром p: $\{y_1 y_2, .., y_s\}$, полученная с помощью III категории датчиков.

И пусть про параметр p возможно предположить, будто любые его значения априори одинаково вероятны. Тогда

$$f\left(p/\bar{y}\right) = \frac{f(\bar{y};p)\cdot q(p)}{f(\bar{y})} =$$

$$= \frac{\left(p^{n\bar{y}}(1-p)^{n-n\bar{y}}\right)q(p)}{f(\bar{y})} \quad (15)$$

– условная плотность распределения,

где $\bar{y} = \frac{\sum_{i=1}^{s} y_i}{s}$.

Заметим, что $f(\bar{y}) = const$, а также в силу предположения о равновозможности вероятностей р: $q(p) = q = const$. Данная плотность является плотностью бета-распределения $B(\alpha,\beta)$ с параметрами $\alpha = n\bar{y}+1, \quad \beta = n - n\bar{y}+1$:

$$f\left(p/\bar{y}\right) = \frac{\left(p^{n\bar{y}}(1-p)^{n-n\bar{y}}\right)}{B(n\bar{y}+1; n-n\bar{y}+1)}. \quad (16)$$

Математическое ожидание бета-функции равно

$$M(T/\bar{y}) = \frac{\alpha}{\alpha+\beta} = \frac{n\bar{y}+1}{n+2}. \quad (17)$$

То есть примем

$$p^* = \frac{n\bar{y}+1}{n+2} = P^*(K \mid D_i). \quad (18)$$

Обозначим D_i, i = 1,2,…,k – оценки общего объема заявок в одноименных пунктах D_i.

Тогда по формуле Байеса доля заявок, прибывших в пункт K, при условии, что пунктом отправления был D_i:

$$P^*(D_i \mid K) = \frac{P^*(D_i)P^*(K \mid D_i)}{P^*(K)}, \quad (19)$$

где

$$P^*(K) = \sum_{i=1}^{k} P^*(D_i)P^*(K\}D),$$

$$\text{а } P^*(D_i) = \frac{D_i}{\sum_{i=1}^{k} D_i}. \quad (20)$$

При известном предполагаемом объеме прибытий N_K заявок/час в пункте K (возможно получить с помощью датчиков I и II категории), найдем оценку количества $N_{D_i \to K}$ заявок/час, отбывающих из пункта D_i:

$$N_{D_i \to K} = P^*(D_i \mid K) \cdot N_K. \quad (21)$$

Таким образом оцениваются корреспонденции слоев спроса $\{D\}_i \leftrightarrow K$.

Если предварительно известны вероятности предпочтения одного из возможных маршрутов из каждого пункта D_i, то остается перераспределить объемы корреспонденций $N_{D_i \to K}$, пропорционально увеличив интенсивности по соответствующим дугам. Данные об интенсивностях по дугам транспортной сети снимаются в вершинах, отобранных согласно алгоритму в пункте 1. Результат перераспределения объема корреспонденций по улично-дорожной сети в авторской модели TIMeR_Mod отражается в матрицах $A_{STREETS}$ и $B_{INTERSECTION}$.

Заключение

В работе с помощью вероятностно-статистических методов решается задача оптимального планирования эксперимента с целью наблюдения за транспортными потоками на улично-дорожной сети в случае, когда требуется оценить изменение транспортных корреспонденций. Разработан алгоритм для определения оптимального расположения точек наблюдения при оценивании изменений в матрице корреспонденций, вызванных появлением нового слоя спроса $\{D\}_i \leftrightarrow K$. Кроме того, уделено внимание определению количества экспериментов, необходимых для достижения требуемой точности расчетов.

Результаты исследования имеют практическую значимость, так как позволяют оптимизировать временные и трудовые ресурсы при решении задачи обновления матрицы корреспонденций в случае введения в эксплуатацию новых объектов на улично-дорожной сети.

Список литературы

1. Zhu S., Guo Y., Zheng H., Peeta S., Ramadurai G., Wang J. Field Data Based Data Fusion Methodologies to Estimate Dynamic Origin-Destination Demand Matrices from Multiple Sensing and Tracking Technologies . USDOT Region V Regional University Transportation Center Final Report, 31st July 2016. 65 p.

2. Yu H., Zhu S., Yang J., Guo Y., Tang T. A Bayesian Method for Dynamic Origin–Destination Demand Estimation Synthesizing Multiple Sources of Data // Sensors. 2021. Vol. 21 (15). No. 4971. P. 1–20. DOI: 10.3390/s21154971.

3. Теселкин А.А. Байесовский подход к задаче планирования экспериментов для оценки транспортных корреспонденций // Наука. Технологии. Инновации: сборник научных трудов: в 10 ч. Новосибирск: Издательство НГТУ, 2017. С. 245–247.

4. Naumova N.A. Method for estimating an origin-destination matrix for dynamic assignment // International Journal of Control Theory and Applications. 2016. Т. 9, № 30. С. 129–138.

5. Наумова Н.А. Отчет о научно-исследовательской работе в рамках гранта РФФИ № 16-48-230720 Р_А «Разработка методов автоматизированного управления транспортными потоками с учетом динамических изменений в матрице корреспонденций». Краснодар: ФГБОУ ВО «КубГТУ», 2018. 97 с.

6. Петров И.Г. Вероятностно-статистические методы в менеджменте // Человек и общество в цифровой экономике: опыт, проблемы и направления развития: материалы Всероссийской научно-практической конференции, посвященной 100-летию ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации». М., 2018. С. 138–144.

7. Теселкин А.А. Методы планирования и статистического анализа наблюдений для оценки матриц транспортных корреспонденций: дис. … канд. техн. наук. Новосибирск, 2018. 162 с.

8. Вентцель Е.С. Теория вероятностей: учеб. для вузов. 11-е изд. М.: Кнорус, 2018. 664 с.

9. Севастьянов Б.А. Курс теории вероятностей и математической статистики. Серия: Классический учебник МГУ. М., 2022. 256 с.