Транспортная теория

December 2021

1 Введение

1.1 Применение в реальной жизни

Транспортные системы городов и их ареалы являются одним из важных факторов, который влияет на социально-экономическое развитие страны. Совершенствование транспортной сети повышает качество жизни горожан, увеличивает рост занятости, укрепляет бюджет города, развивает бизнес и привлекает инвестиции.
Дороги являются "каркасом"страны (города), с помощью которого соединяются между собой все части страны (города). Поэтому в идеальной модели рыночной экономики улично-дорожные сети, в частности и городские, должны обеспечивать высокую мобильность людей и беспрепятственную перевозку товаров, так как состояние транспортной сети напрямую влияет на состояние экономики в целом.

Для определения потоков и загрузки элементов сети сегодня существует много математических моделей. Но основной проблемой в большинстве данных моделей является необходимость информации о передвигающихся по городу индивидуумах. Эта информация представляется в виде матрицы корреспонденций.

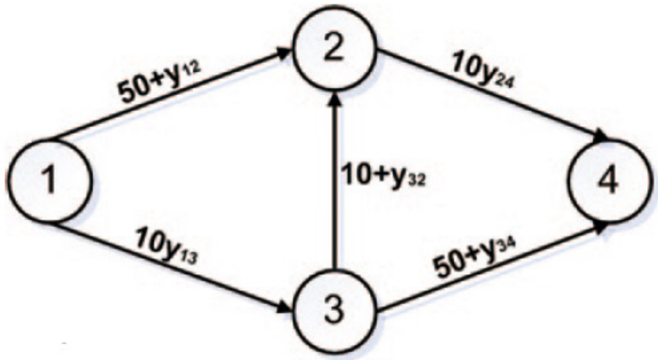
Матрица корреспонденций - информация о том, откуда, сколько и куда людей направляются в единицу времени.

Если матрица корреспонденций построена, то на её основе можно:

- составить наиболее точное расписание движения общественного транспорта
- определить загрузки элементов улично-дорожной сети
- определить главные пассажирообразующие пункты
- оценить количество перевозимых пассажиров по типам пассажиров, по видам транспорта, маршрутам и направлениям
- оценить интенсивность пассажиропотоков между различными пунктами

1.2 Математические понятия

Равновесный принцип Нэша–Вардропа. Для объяснения ниже будет приведен пример Рассмотри следующую транспортную систему:



Из 1 в 4 в единицу времени выезжает 6 автомобилей.

Буквой y обозначается количество автомобилей, проезжающих по данному ребру в единицу времени (поток), формулы над ребрами — веса ребер (время проезда ребра в минутах в зависимости от величины потока автомобилей по ребру y).

f_e - поток по ребру e, где $e \in E = (12), (24), (13), (34)$ - множество всех ребер.

Обозначим за y_{124}, y_{134} - потоки на путях 124 и 134 соответсвенно. Аналогично ребрам, вводится множество путей $P = (124), (134)$.

y_p - поток по пути $p \in P$. В данном примере

$$y_{124} = f_{12} = f_{24}, y_{134} = f_{13} = f_{34}$$

В лекциях присутствует следующее понятие: d_{14} - **корреспонденция** - сколько людей в единицу времени хочет перемещаться из 1 в 4.

$y_{124} + y_{134} = d = 6$

Итого, можем записать следующее равенство:

$$f = \Theta x$$

Потоки по ребрам связаны с потоками по путям.

Условие равновесия: все используемые пути должны иметь одинаковую «длину». Всего путей, ведущих из 1 в 4 три: 1–3–4, 1–3–2–4, 1–2–4.

Легко проверить, что если эти 6 автомобилей в одинаковых пропорциях распределятся по всем этим путям (по 2 автомобиля в единицу времени на каждый путь), то время прохождения каждого пути («длина» пути) будет равна 92 минутам $(10 \cdot (2 + 2) + (50 + 2) = 10 \cdot (2 + 2) + (10 + 2) + 10$

$\cdot\ (2+2) = (50+2) + 10 \cdot (2+2))$. Это и будет единственным равновесием (Нэша—Вардропа) в транспортной сети. То есть от такой конфигурации никому из водителей не выгодно отклоняться при условии, что остальные водители не меняют свой выбор.

Транспортная система В работе рассматривается замкнутая транспортная система, описываемая графом $G =< V, E >$, где V- множество вершин ($|V| = n$), а E - множество ребер ($|E| = m$)

Обозначим через $\tau_e(f_e)$ - функцию затрат (например, временных) на проезд по ребру e, если поток автомобилей на этом участке f_e.

$\tau_e(f_e) = \bar{t}_e \cdot (1 + k(\frac{f_e}{\overline{f}_e}))^{\frac{1}{\mu}}$, где

- $\bar{t}_e$ - время проезда ребра e по свободной дороге
- $\overline{f}_e$ - максимальная пропускная способность ребра [авт/час]
- k - какой-то небольшой коэффициент
- $\frac{1}{\mu} = 4$ (обычно берут 4) - показатель степени

Вообще, данная функция относится к классу функций BPR. Также допускается $\mu- > 0+$ - моедль стабильной динамики.

Введем t_e - временные затраты на прохождение ребра e. $t_e = \tau_e(f_e)$. По затратам $t = \{t_e\}_{e \in E}$ можно определить затраты на перемещение из источника i в сток j по кратчайшему пути:

$$T_{ij} = \min_{p \in P_{ij}} \sum_{e \in E} \delta_{ep} t_e$$

, где:

- p - путь (без самопересечений, циклов) на графе (набор ребер)
- P_{ij} - множество всевозможный путей на графе, стартующих в i и заканчивающих в j
- δ_{ij} - дельта функция, показывающая, принадлежит ли ребро e пути p
- **Корреспонденция** .Часть вершин $O \subset V(origin)$ являются источниками корреспонденций, а часть стоками корреспонденций $D \subset V(destination)$. В примере выше, мы рассмотрели корреспонденцию, но в общем случае, корреспонденций может быть несколько, обозначаются они следующим образом $d_{ij} = (i, j) \in (OD)$. Если говорить более точно, то вводится множество пар (источник, сток) корреспонденций $OD \subset V * V$. Не ограничивая общности, будем далее считать, что $\sum_{(i,j) \in (OD)} d_{(i,j)} = 1$. В общем случае корреспонденции – не известны! Однако известны

(заданы) характеристики источников и стоков корреспонденций. То есть известны величины $\{l_i\}_{i\in O}, \{w_j\}_{j\in D}$

$$\sum_{j:(i,j)\in(OD)} d_{ij} = l_i, \sum_{i:(i,j)\in(OD)} d_{ij} = w_j \quad (1)$$

.

Кроме того "сколько вышло, столько и пришло"

$$\sum_{i\in O} l_i = \sum_{j\ inD} w_j = 1$$

.

Условие (1) для краткости будем записывать $d \in (l, w)$

Кроме этого, в примере у нас присутствовала матрица Θ. Эта матрица имеет следующий вид: $\Theta = ||\delta_{ep}||_{e\in E, p\in P}$

1.3 Полезные уравнения:

$$1. f_e = \sum_{p\in P} \delta_p x_p$$

- поток на ребре e равен сумме потоков по тем путям, в которые входит это ребро.

$$2. \sum_{p\in P_w} x_p = d_w, w \in W$$

Затраты на пути p:
$T_p = \sum_{e\in E} \delta_{ep}\tau_e(f_e(x))$ - суммируем по ребрам, которые входят в данный путь.

Что означает, что $x*$ - равновесие?

$\forall w \in W, \forall p \in P_w$:

$$X_p^* \cdot (T_p(x^*) - \min_{q\in P_w} T_q(x^*)) = 0 \quad (2)$$

(2) - условие комплиментарности. Также можно считать переформулировкой понятия равновесия по Нэшу.
$p \in P_w$ - пути, отвечающие корреспонденции w

2 Энтропийная модель расчета матрицы корреспонденций

Под энтропийной моделью расчета матрицы корреспонденций $d(T)$ понимается определенный способ вычисления набора корреспонденций $d_{ij\,(i,j)\in W}$

по известной матрице затрат $\{T_{ij}\}_{(i,j)\in W}$. Этот способ заключается в решении задачи энтропийно-линейного программирования, которую можно понимать, как энтропийно-регуляризованную транспортную задачу:

$$\min_{d\in(l,w);d\geq 0} \sum_{(i,j)\in OD} d_{ij}T_{ij} + \gamma \sum_{(i,j)\in OD} d_{ij} ln d_{ij} \quad (3)$$

, где параметр $\gamma > 0$ считается известным.

3 Модели равновесного распределения транспортных потоков по путям:

Матрица корреспонденций ${d_{ij}}_{(i,j)} \in OD$ порождает (вообще говоря, неоднозначно) некий вектор распределения потоков по путям x. Неоднозначность заключается в том, что балансовые ограничения, которые возникают на

$$x \in X(d) : x \geq 0 : \forall (i,j) \in OD -> \sum_{p\in P_{ij}} x_p = d_{ij}$$

как правило, не определяют вектор x однозначно. Вектор x, в свою очередь, пораждает вектор потоков на ребрах, $f = \Theta x$, который, в свою очередь, порождает вектор (временных) затрат на ребрах $t(f) = \tau_e(f_e)_{e\in E}$. На основе последнего вектора уже можно рассчитать матрицу затрат на кратчайших путях $T(t) = \{T_{ij}(t)_{(i,j)\in OD}\}$. Собственно, модель равновесного распределения потоков это формализация принципа Нэша–Вардропа о том, что в равновесии каждый водитель выбирает для себя кратчайший путь. Другими словами, если для заданной корреспонденции $(i, j) \in OD$ известно, что (условие комплиментарности)

$$x_{p'} > 0, p^{'} \in P_{ij} -> T_{ij}(t) = \min_{p\in P_{ij}} \sum_{e\in E} \delta_{ep} t_e = \sum_{e\in E} \delta_{ep'} t_e$$

Решение задачи дает модель вычисления вектора потока на ребрах при заданной матрице корреспонденций $f(d)$

Задача поиска равновесия сводится, таким образом, к поиску такого вектора $x \in X(d)$, который бы порождал такие затраты $T := T(t(f(x)))$, что выполянется условие комплиментарности. В написанном выше виде искать равновесный вектор $x \in X(d)$ представляется сложной задачей, сводящейся к решению системы нелинейных уравнений. Однако, в данном случае (рассматривается потенциальная игра загрузки) можно свести поиск равновесия к решению задачи выпуклой оптимизации

$$\min_{(f,x):f=\Theta x; x\in X(d)} \sum_{e\in E} \int_0^{f_e} \tau_e(z)dz \quad (4)$$

Решение задачи дает модель вычисления вектора потока на ребрах при заданной матрице корреспонденций $f(d)$

4 Двухстадийная модель:

Выше были описаны две модели. В первой (расчёт матрицы корреспонденций) на вход подается матрица затрат T, а на выходе получается матрица корреспонденций $d(T)$. Во второй модели наоборот, на вход подается матрица корреспонденций $d(T)$, а на выходе рассчитывается матрица затрат $T(d) = T(t(f(d)))$.

Под равновесием в двухстадийной транспортной модели понимается такая пара (f, d), что

$$d = d(T(f(t))), f = f(d) \quad (5)$$

то есть (f, d) – есть неподвижная точка описанных двух блоков моделей. Собственно, часто на практике так и ищут равновесие последовательно (друг за другом) прогоняя описанные два блока. Но присутствует проблема, что нет никаких теоретических гарантий, что последовательная прогонка будет сходиться к неподвижной точке. Даже если наблюдается сходимость, то непонятно, насколько эта сходимость может быть быстрой, и лучший ли это способ (простая прогонка) численного решения (5)?

Теорема 1: Задача поиска неподвижной точки (5) сводится к задаче выпуклой оптимизации:

$$\min_{(f,x,d):f=\Theta x; x\in X(d); d\in(l,w); d\geq 0} \sum_{e\in E} \int_0^{f_e} \tau_e(z)dz + \gamma \sum_{(i,j)\in OD} d_{ij} ln\, d_{ij} \quad (6)$$

. Смысл в том, что теперь это задача выпуклой оптимизации, которую можно решать оптимальным по скорости (глобально сходящимися) алгоритмами. **Доказательство присутствует в самой статье, выписывать его нет смысла.**

Заметим, что выписанная задача (6), как задача оптимизации относительно d при «заморожоженных» (f, x), совпадает с задачей (3) и, наоборот, при «замороженном» d задача (6), как задача оптимизации относительно (f, x), совпадает с задачей (4).

Таким образом, если удалось найти такую задачу, решение которой одновременно дает нужные нам связи переменных, описываемые формулой (5), то это и означает, что нам удалось свести поиск неподвижной точки сложного нелинейного отображения (которое не удается выписать аналитически) к явно выписанной задаче оптимизации (6). Ометим, что решение этой выпуклой задачи оптимизации по сложности сопоставимо с решением задачи (4), что будет пояснено в следующем разделе.

5 Переход к двойственной задаче

задача выпуклой оптимизации (6) можно переписать эквивалентным седловым образом, введя двойственные переменные $t = \{te\}_e$, которые имеют естественную интерпретацию вектора потоков на ребрах,

$$\min_{d\in(l,w);d\geq 0}\max_{t_e\in dom\ \sigma_e^*,e\in E}\left\{\sum_{(i,j)\in OD} d_{ij}T_{ij}(t)-\sum_{e\in E}\sigma_e^*(t_e)+\gamma\sum_{(ij)\in OD} d_{ij} ln\ d_{ij}\right\}\quad (**),$$

где $\sigma_e(f_e)=\int_0^{f_e}\tau_e(z)dz$, а $\sigma_e^*(t_e)=max_{f_e\geq 0}\{f_e t_e-\sigma_e(f_e)\}$ - сопряженная функция.

Уравнение $(**)$ удобнее переписать следующим образом:

$$\max_{t_e\in dom\ \sigma_e^*,e\in E}\ \min_{d\in(l,w);\sum_{(i,j)\in OD};d_{ij}=1,d\geq 0}\left\{\sum_{(i,j)\in OD} d_{ij}T_{ij}(t)+\gamma\sum_{(i,j)\in OD} d_{ij} ln\ d_{ij}\right\}-\sum_{e\in E}\sigma_e^*(t_e)$$

Вспомогательную задачу минимизации можно представить через двойственную к ней:

$$\max_{t_e\in dom\sigma_e^*;e\in E;(\lambda,\mu)}\ -\gamma\ ln\left(\sum_{(i,j)\in OD} exp\left(\frac{-T_{ij}(t)+\lambda_i+\mu_j}{\gamma}\right)\right)+<l,\lambda>+<w,\mu>-\sum_{e\in E}\sigma_e^*(t_e)=$$

$$-\min_{t_e\in dom\sigma_e^*;e\in E;(\lambda,\mu}\ \gamma\ ln\left(\sum_{(i,j)\in OD} exp\left(\frac{-T_{ij}(t)+\lambda_i+\mu_j}{\gamma}\right)\right)-<l,\lambda>-<w,\mu>+\sum_{e\in E}\sigma_e^*(t_e)\quad (7)$$

Обратим внимание, что добавленное по d ограничение $\sum_{(i,j)\in OD} d_{ij}=1$ тавтологично, поскольку следует из $d\in(l,w)$. Тем не менее, удобнее его добавить, чтобы при взятии min получалась равномерно гладкая функция (типа softmax), а не сумма экспонент, имеющая неограниченные константы гладкости. Множители λ и μ являются двойственными множителями (множителями Лагранжа) к ограничениям $d\in(l,w)$, которые заносятся в функционал (ограничения $\sum_{(i,j)\in OD} d_{ij}=1; d\geq 0$ не заносятся в функционал). Заметим, что если (t,λ,μ) – решение задачи (7), то $(t,\lambda+(C_\lambda,\ldots,C_\lambda)^T,\mu+(C_\mu,\ldots,C_\mu)^T)$ - также будет решением, т.е. решение (7) не единственное. Заметим также, что зная (λ,μ), можно посчитать матрицу корреспонденци:

$$d_{ij}(\lambda,\mu)=\frac{exp\left(\frac{-T_{ij}(t)+\lambda_i+\mu_j}{\gamma}\right)}{\sum_{(k,l)\in OD} exp\left(\frac{-T_{kl}(t)+\lambda_k+\mu_l}{\gamma}\right)}\quad (8)$$

Для решения задачи выпуклой оптимизации (но, вообще говоря, не гладкой, поскольку функции $T_{ij}(t)$ – негладкие) можно использовать субградиентные методы. А именно, субградиент (далее обозначаем (супер-)субградиент таким же символом, как и градиент ∇) целевого функционала по t (стоящего под минимумом) (7) можно посчитать по формуле Демьянова–Данскин

$$\sum_{(o,j)\in OD} d_{ij}(\lambda,\mu)\nabla T_{ij}(t)+f=\sum_{(i,j)\in OD} d_{ij}(\lambda,\mu)\nabla T_{ij}(t)+\left(\{\tau_e^{-1}(t_e)\}_{e\in E}\right)^T\quad (9)$$

, где $\tau_e^{-1}-$ обратная функция к τ_e.

Примечательно, что отличие формулы (9) от ее аналога, который можно получить, решая задачу (4) посредством перехода к двойственной задач только в том, что $d_{ij} = d_{ij}(\lambda, \mu)$, где (λ, μ) определяются из решения задач

$$\min_{(\lambda,\mu)} D(t, \lambda, \mu) := \gamma ln \left(\sum_{(i,j)\in OD} exp\left(\frac{-T_{ij}(T) + \lambda_i + \mu_j}{\gamma}\right)\right) - < l, \lambda > - < w, \mu > \quad (10)$$

Важное наблюдение, заключается в том, что решать задачу (7) выгоднее не как задачу оптимизации по переменным (t, λ, μ), а как задачу только по переменной t, в то время как переменные (λ, μ) лишь используются для подсчета субградиента целевого функционала по формуле (9) с $d_{ij}(\lambda, \mu)$ рассчитываемым по формуле (8), в которой

$$(\lambda, \mu) := (\lambda(t), \mu(t)) \in arg \min_{(\lambda,\mu)} D(t, \lambda, \mu) \quad (11)$$

определяются как решение задачи (10). 2). Заметим, что сложность вычисления $\nabla T_{ij}(t)$ оптимальным алгоритмом Дейкстры будет сопоставима со сложностью вычисления (с нужной точностью) матриц $d_{ij}(\lambda(t), \mu(t))$. Таким образом, получается, что сложность вычисления субградиента для двойственной задачи к (4) и для задачи (7) сопоставимы. При этом, свойства гладкости (определяющие скорость сходимости используемых методов) целевого функционала в задаче (7) при переходе от оптимизации в пространстве (t, λ, μ) к оптимизации по переменной t могут только улучшиться (в любом случае не ухудшиться)