

УДК 519.652

МЕТОД ПРОГНОЗИРОВАНИЯ ДОРОЖНОЙ СИТУАЦИИ В УСЛОВИЯХ НЕПОЛНОТЫ И ЗАШУМЛЕННОСТИ ДАННЫХ

К. М. Вялых¹,

аспирант

Московский государственный технический университет им. Н. Э. Баумана

Рассмотрен общий подход к прогнозированию загруженности городских транспортных магистралей, приведены предварительные результаты экспериментальных исследований. Описан алгоритм прогнозирования дорожной ситуации в условиях неполноты и зашумленности сигнала. Для оценки качества прогнозирования системы производилось сравнение с предсказанием экстраполяцией средней скорости.

Ключевые слова — дорожное движение, прогнозирование заторов, анализ временных рядов, метод главных компонент.

Введение

Сейчас в России основным источником информации о загруженности дорог являются системы, которые отображают текущую дорожную обстановку и характерную загрузку дорог [1].

За рубежом существует ряд систем, которые предоставляют прогноз изменения дорожной обстановки в будущем на основе известной предыстории и данных о текущей ситуации. Однако они используют информацию с датчиков, установленных под дорожным полотном или у обочины. Поэтому таким системам доступна полная информация о количестве и скорости движения автомобилей на каждом участке трассы. Одной из самых известных разработок является система Traffic Prediction Tool (TRT) корпорации IBM [2].

Разрабатываемая система, в отличие от ныне существующих зарубежных аналогов, строит прогноз изменения дорожной ситуации на основе неполных данных о скорости движения автотранспорта. От российских разработок отличие заключается в учете при построении прогноза не только исторических данных, но и ситуации, сложившейся в городе в данный момент.

Система использует ряд математических моделей, каждая из которых имеет большую или меньшую предсказательную способность при определенных условиях.

¹Научный руководитель — доктор технических наук, профессор кафедры ИУ-3 Ю. Н. Павлов.

Методика прогнозирования дорожных ситуаций

Критерии качества предсказания

При выборе наилучшей математической модели прогнозирования дорожной обстановки необходимо задаться некоторой численной характеристикой качества прогноза. Тем не менее, в рамках решаемой задачи невозможно предложить единственно верный критерий точности предсказания ввиду сложности процесса и неоднородности требований пользователей к системе. Ни средняя абсолютная ошибка, ни средняя относительная ошибка, ни среднеквадратическая ошибка прогнозирования скорости не совпадают с интуитивным представлением о погрешности прогноза, что легко показать на следующем примере. Пусть при реальной скорости транспортного потока 80 км/ч две сравниваемые математические модели получили прогнозную скорость 140 и 20 км/ч соответственно. Для обеих моделей абсолютные ошибки прогнозирования равны 60 км/ч, относительные ошибки равны 75 %, а среднеквадратические ошибки равны 60 км/ч. Однако, с точки зрения пользователя, эти результаты не являются равноценными, поскольку применение второй модели приводит к принципиально неверному выводу о наличии затора.

Для решения этой проблемы введем метрику $M(V)$, которая является некой функцией от скорости движения транспортного потока V . Абсолютная ошибка прогнозирования этой метрики несет информацию об ошибке прогнозирования

ситуации. Как было сказано выше, эта задача принципиально не имеет оптимального решения, поэтому будем определять способ ее расчета, опираясь на ряд естественных требований.

Требование 1. Так как водителей интересует относительная погрешность прогнозирования времени движения T , вычисляемая по соотношению

$$\frac{T_p - T_r}{T_r} = \frac{T_p}{T_r} - 1 = \frac{V_p}{V_r} - 1,$$

где индексами p и r обозначены соответственно прогнозируемые и реальные величины, то разность метрик $M(V_p) - M(V_r)$ должна быть пропорциональна отношению V_p/V_r .

Требование 2. Так как на разгруженной дороге скорости зависят в большей степени от характера вождения и типа транспортного средства, чем от дорожной ситуации, то чувствительность метрики к скорости должна быть тем меньше, чем больше значение самой скорости, т. е.

$$\lim_{V \rightarrow \infty} \frac{dM(V)}{dV} = 0.$$

Требование 3. Погрешность разности метрик $M(V_p) - M(V_r)$, обусловленная погрешностью измерения времени T , не должна зависеть от длины рассматриваемого участка дороги.

Перечисленным требованиям удовлетворяет метрика вида $M(V) = \ln V$.

При расчете параметров моделей и выборе моделей оптимизируется среднее квадратическое отклонение метрики $M(V)$.

Прогнозирование на основе характерной скорости движения

В основе предлагаемого алгоритма прогноза лежит предположение о стационарности характеристик распределения скорости. В качестве прогнозного значения скорости используется медиана распределения скорости с учетом периодичности по времени суток, дням недели и времени года. Эта модель наиболее проста, но она не позволяет извлечь никакой дополнительной информации из текущей дорожной обстановки и поэтому наследует все недостатки, свойственные существующим системам предсказания загруженности магистралей.

Взаимосвязь между дорогами

Ситуация на одних участках движения влияет на ситуации на других участках движения. Эта связь наиболее очевидна, когда все автомобили с определенного участка могут выехать только на одну дорогу. Затор на этой дороге повлияет на дорожную ситуацию всего участка. Величина этой взаимосвязи зависит от структуры развязки и схемы организации движения.

Поскольку на скорость движения потока влияют внешние факторы, такие как часы пик, праздники и т. д., то чтобы минимизировать это влияние, требуется исследовать отклонения скорости от ее характерной величины для данного времени суток, дня недели и времени года:

$$V' = V_{\text{cur}} - V_{\text{char}},$$

где V' — отклонение скорости; V_{cur} — текущая скорость; V_{char} — характерная скорость.

Рассмотрим пару дорог c и d . Предположим, что взаимное влияние одной дороги на другую имеет линейный закон и распространяется с некой задержкой τ . Если $\tau < 0$, то дорога c влияет на дорогу d . Если $\tau > 0$, то дорога d влияет на дорогу c . Если τ^* близко к нулю, то имеет место как влияние c на d , так и влияние d на c . То есть имеется взаимосвязь вида

$$V'_c(t) = a V'_d(t - \tau) + \varepsilon(t),$$

где V'_c — отклонение скорости для дороги c ; V'_d — отклонение скорости для дороги d ; a — величина взаимосвязи; ε — остаточное отклонение скорости.

Остаточное отклонение $\varepsilon(t)$ состоит из двух составляющих: случайного отклонения, которое затухает при неизменных значениях $V'_c(t)$ и $V'_d(t - \tau)$, и неслучайного отклонения, которое затухает с изменением значений $V'_c(t)$ и $V'_d(t - \tau)$. Априорно разделить эти отклонения не представляется возможным, однако величину взаимосвязи характеризуют скорости затухания этих отклонений. Поэтому вне зависимости от изменения $V'_c(t)$ и $V'_d(t - \tau)$ взаимосвязь характеризуется затуханием величины $\varepsilon(t)$. Если принять, что скорость затухания случайного отклонения скорости пропорциональна самому отклонению, то получим выражение

$$\varepsilon(t) - \varepsilon(t - 1) = b\varepsilon(t - 1) + \delta(t - 1).$$

Сила взаимосвязи между дорогой c и дорогой d , $\mu_{c,d}$, будет оцениваться как отношение коэффициента затухания b к стандартной ошибке определения этого коэффициента SE_b . Стандартная ошибка определения коэффициента линейной регрессии вычисляется как отношение среднее квадратического отклонения остатка к среднее квадратическому отклонению фактора, следовательно:

$$\mu_{c,d} = \left| \frac{b}{SE_b} \right| = \left| \frac{b\sigma(\varepsilon)}{\sigma(\delta)} \right|,$$

где $\sigma(\cdot)$ — среднее квадратическое отклонение.

Значение силы взаимосвязи между дорогами μ зависит от значения задержки τ . Оптимальное значение τ не выводится аналитически, поэтому определяется перебором. Выбирается такая величина τ^* , при которой взаимосвязь μ максимальна.

Скорость движения при отсутствии заторов зависит не столько от дорожной ситуации, сколько от характера вождения и вида транспортного средства. Поэтому в расчете параметров a и b участвуют только отсчеты, в которых скорости на обеих дорогах ниже определенного порога.

Прогнозирование на основе характерной скорости движения с учетом отклонений

Для прогнозирования дорожной ситуации на дороге c используется информация о ситуации на дорогах с силой взаимосвязи больше определенного порога. Величина взаимосвязи рассчитывается с помощью метода главных компонент [3], использующего сингулярное разложение матрицы отклонений скоростей от характерной скорости V' .

После перемножения компонентов разложения получим матрицу V'_{bal} . В ней содержатся отклонения скорости, «сбалансированные» относительно предположения о линейном законе взаимосвязи между дорогами.

Прогнозная скорость V_{cp} для дороги c через время θ рассчитывается с помощью уравнения вида

$$V_{cp}(t + \theta) = V_{c\ char}(t + \theta) + \beta(\theta)V'_{c\ bal}(t),$$

где V_{cp} — прогнозная скорость на дороге c ; $V_{c\ char}$ — характерная скорость на дороге c ; $V'_{c\ bal}$ — отклонение скорости для дороги c , содержится в строке матрицы V'_{bal} , соответствующей дороге c ; $\beta(\theta)$ — функция, описывающая влияние «сбалансированного» отклонения скорости от характерной на прогнозную скорость для каждого θ .

Прогнозирование на основе прецедентов

В основе метода лежит следующее предположение: если дорожная ситуация в определенном районе подобна ситуации в прошлом, то она получит сходное развитие. Для уменьшения влияния случайных факторов исследуется аналогия не с одной ситуацией-прецедентом, а с группой таких ситуаций. В качестве метода разбиения ситуаций по группам используется метод k ближайших соседей [4]. В качестве «расстояния» выступает мера различия двух ситуаций на дороге

$$J_{road} = \sum_{t=t_{beg}}^{t_{end}} \left(\ln(V_{cur}(t)) - \ln(V_{past}(t)) \right)^2,$$

где V_{cur} — текущая скорость; V_{past} — соответствующая скорость в прошлом.

Мера различия ситуаций вычисляется по множеству дорог с силой взаимосвязи, большей определенного порога:

$$J_{area} = \frac{1}{l} \sum_{k=1}^K \mu J_{road\ k},$$

где l — количество пар отсчетов, в которых известно значение скорости для обеих сравниваемых

ситуаций; $J_{road\ k}$ — мера различия двух ситуаций на дороге k ; K — число дорог, имеющих наибольшее влияние на данную дорогу.

Мера схожести ситуаций считается величиной, обратной мере их различия:

$$W_{area} = \frac{1}{J_{area}}.$$

В случайном процессе $V(t)$ присутствует явная цикличность по времени суток, по дню недели и по времени года. Будем искать схожие ситуации с различным шагом цикличности p .

Ситуации подразделяются на группы с помощью метода k ближайших соседей по метрике J_{area} , рассчитанной на интервале $[t - T - np, t - np]$, где n — целое число. Кластеризации подлежат только ситуации, для которых выполнен ряд условий: полнота истории, значимость истории и определенность истории.

Полнота истории P оценивается количеством измерений:

$$P = \frac{N}{T},$$

где N — количество отсчетов в течение времени T .

Значимость истории S оценивается количеством ситуаций, попавших в группу.

Определенность истории D оценивается близостью развития ситуации внутри группы и равна дисперсии величины $J_{area}[t - T - np, t - np]$ по множеству ситуаций в группе.

Прогнозная скорость вычисляется как взвешенная сумма экстраполированных характерных скоростей с весами W_{area} :

$$V_p(t + \theta) = \frac{\sum_{t=t_{beg}}^{t_{end}} W_{area}[t - np, t + \theta - np] V_{char}(t + \theta - np)}{\sum_{t=t_{beg}}^{t_{end}} W_{area}[t - np, t + \theta - np]}.$$

Метод k ближайших соседей не гарантирует малой меры различия дорожных ситуаций. Однако благодаря тому, что W_{area} экспоненциально растет при уменьшении J_{area} , влияние существенно отличающихся ситуаций на прогноз будет пренебрежимо мало.

Борьба с неполнотой и зашумленностью данных

Борьба с неполнотой:

- при расчете взаимосвязи между дорогами алгоритму для получения новой информации о взаимосвязи достаточно двух релевантных последовательных измерений на обеих дорогах;
- при прогнозировании на основе характерной скорости движения на небольшом количестве измерений влияние выбросов нивелируется благо-

даря оценке медианой распределения вместо среднего значения;

- при прогнозировании на основе характерной скорости движения с учетом отклонений можно выявлять множественные взаимосвязи при небольшом количестве измерений;

- при прогнозировании на основе прецедентов при расчете меры схожести между ситуациями нет необходимости иметь информацию о ситуации на всех участках области.

Борьба с зашумленностью:

- погрешность разности метрик $M(V_p) - M(V_r)$ не зависит от длины рассматриваемого участка дороги, что позволяет равномерно учитывать ошибку измерения времени;

- благодаря возможности объединить дороги в группы по силе взаимосвязи и рассматривать группы как единое целое, ошибка прогнозирования уменьшается.

Результаты численного эксперимента

Для оценки качества предсказания системы производилось сравнение с предсказанием экстраполяцией средней скорости.

Критерием качества предсказания было выбрано относительное отклонение фактического времени движения автомобиля по маршруту от предсказанного времени. В качестве реального времени движения использовалось время движения автомобиля по маршруту по данным компании «СитиГид» [5] для дорог Санкт-Петербурга за сентябрь—октябрь 2010 года.

При предсказании экстраполяцией средней скорости получено среднее относительное отклонение времени движения автомобиля, равное 37 %. Аналогичная величина при прогнозировании с помощью разрабатываемой системы получилась равной 22 %.

В дальнейшем планируется провести тестирование на большем интервале времени и ввести

в рассмотрение новые модели и факторы, влияющие на дорожную обстановку.

Заключение

Таким образом, разработанные методы прогнозирования имеют точность существенно выше, чем предсказание экстраполяцией средней скорости, которую используют большинство ныне существующих в России систем. Это свидетельствует об эффективности использования данных о сложившейся ситуации при построении краткосрочного прогноза, даже в условиях неполноты и зашумленности этих данных. Результаты данной работы в перспективе могут позволить более рационально использовать имеющиеся транспортные ресурсы, в частности, максимально равномерно распределять транспортный поток по различным магистралям.

Литература

1. Описание системы «Яндекс Пробки». <http://help.yandex.ru/maps/informer.xml> (дата обращения: 01.03.2012).
2. Описание системы Traffic Prediction Tool (TPT). http://www.ibm.com/smarterplanet/us/en/traffic_congestion/nextsteps/index.html (дата обращения: 01.03.2012).
3. Jolliffe I. T. Principal Component Analysis, Series: Springer Series in Statistics. 2nd ed. — NY.: Springer, 2002. — P. 64–68.
4. MacQueen J. Some methods for classification and analysis of multivariate observations // Proc. 5th Berkeley Symp. on Math. Statistics and Probability. 1967. P. 281–297.
5. Описание системы «СитиГид». <http://www.probki.net/news.aspx> (дата обращения: 01.03.2012).