Прогнозирование транспортной загруженности дорог

07, июль 2012

DOI: 10.7463/0712.0418480

Вялых К. М., Павлов Ю. Н., Степанова М. Б.

УДК. 004.021

Россия, МГТУ им. Н.Э. Баумана

kostyoskol@mail.ru

pavlov@bmstu.ru

mari-na-bzzz@yandex.ru

Введение

Загруженность автомобильных дорог стала в настоящее время глобальной проблемой. Особенно остро она ощущается в мегаполисах.

Данная статья посвящена разработке системы прогнозирования дорожной обстановки, которая будет предоставлять информацию о характерной загруженности дорог, наиболее вероятной дорожной ситуации в текущий момент времени, а также строить прогноз изменения дорожной ситуации.

В разрабатываемой системе прогнозирования в качестве источника первичной информации используются данные о передвижении автомобилей, полученные от портативных спутниковых навигаторов. Информация включает данные о маршруте следования автомобиля и времени пересечения им узлов на графе дорог.

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 на основе известной предыстории и данных о текущей ситуации, используя информацию с датчиков, установленных под дорожным полотном или у обочины. Поэтому таким системам доступна полная информация о числе и скорости движения автомобилей на каждом участке трассы. Одной из самых известных разработок является система *Traffic Prediction Tool (TPT)* корпорации IBM [2].

Разрабатываемая авторами статьи система, в отличие от существующих зарубежных аналогов, строит прогноз изменения дорожной ситуации на основе зашумленных и неполных данных о скорости движения автотранспорта. Неполнота данных обусловлена тем, что мы знаем дорожную ситуацию только на тех дорогах, где находятся автомобили, оборудованные портативными спутниковыми навигаторами. А зашумленность обусловлена ошибками измерения местоположения и скорости движения этих автомобилей. В отличие от известных российских разработок, при построении прогноза учитываются не только исторические данные о скоростях, но и ситуация, сложившаяся в городе в данный момент.

Система использует ряд математических моделей, каждая из которых имеет большую или меньшую предсказательную способность при определенных условиях. Ниже описаны модели и критерии, на основе которых происходит выбор наиболее эффективной модели для сложившейся ситуации.

Научная новизна работы заключается в разработке алгоритма прогнозирования дорожной ситуации, который показал себя лучше аналогов в условиях неполноты и зашумленности исходных данных.

1. Количественные характеристики дорожных ситуаций

Дорожной ситуацией будем называть совокупность местоположений, скоростей и габаритных размеров всех транспортных средств на рассматриваемой дороге. Однако используемый тип датчиков позволяет получить информацию только о средней скорости движения. Дорожную ситуацию будем характеризовать метрикой – непрерывно дифференцируемой функцией $M(V)$, зависящей от скорости движения транспортного потока V и выбранной так, чтобы разность значений метрики, соответствующих двум ситуациям на дороге, несла информацию о различии этих ситуаций. Метрика должна обладать свойствами.

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

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

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 .

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, то есть

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

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

Перечисленные требования оставляют некоторую свободу в выборе конкретного вида функции $M(V)$, однако среди основных элементарных функций им удовлетворяет только логарифмическая метрика вида $M(V) = \ln V$.

Меру различия J некоторых дорожных ситуаций 1 и 2, относящихся к группе дорог и отстоящих по времени на интервал Δt , будем оценивать средним квадратом разности их метрик:

$$J = \frac{1}{N} \sum_{i=1}^I \sum_{t=t_{beg}}^{t_{end}} (M(V_1^i(t)) - M(V_2^i(t - \Delta t)))^2 = \frac{1}{N} \sum_{i=1}^I \sum_{t=t_{beg}}^{t_{end}} (\ln V_1^i(t) - \ln V_2^i(t - \Delta t))^2,$$

где N – число измерений, I – число рассматриваемых дорог, t_{beg} и t_{end} – начало и конец временного интервала, на котором производятся измерения. Время t меняется дискретно с постоянным шагом дискретизации.

При оптимизации прогнозирующих моделей необходимо сравнивать прогнозируемую и фактическую ситуацию за один и тот же интервал времени, то есть исследовать величину

$$J = \frac{1}{N} \sum_{i=1}^I \sum_{t=t_{beg}}^{t_{end}} (M(V_r^i(t)) - M(V_p^i(t)))^2 = \frac{1}{N} \sum_{i=1}^I \sum_{t=t_{beg}}^{t_{end}} (\ln V_r^i(t) - \ln V_p^i(t))^2,$$

где V_r^i – реальная скорость на дороге i ; V_p^i – прогнозная скорость на дороге i .

2. Взаимосвязь между дорогами

Заторы на одних улицах провоцируют заторы на других. При этом от структуры развязок и схемы организации движения будет зависеть и скорость транспортного потока, и величина взаимовлияния. Для повышения точности прогноза нужно учитывать эти связи между дорогами.

Поскольку на скорость движения потока влияют такие факторы, 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

$$U = V - V_{char},$$

где U – отклонение скорости, V – текущая скорость, V_{char} – характерная скорость (п. 3).

Примем, что взаимное влияние отклонений скорости движения транспорта на двух дорогах c и d имеет линейный закон. Любую ситуацию, отвечающую этому линейному закону, будем называть равновесной. Всякое отклонение скорости от равновесной скорости, связанное с изменением дорожной ситуации, на одной из дорог распространится на другие дороги с некоторым запаздыванием τ , тем самым обеспечив восстановление равновесия, но уже при иных величинах скоростей. Таким образом, математическая модель взаимосвязи имеет вид

$$U_c(t) = a U_d(t - \tau) + \varepsilon(t),$$

где a – коэффициент взаимосвязи, ε – остаточное отклонение скорости.

Всякое отклонение скорости от равновесной скорости, не связанное с изменением дорожной ситуации, не будет влиять на скорость движения на других дорогах, однако с течением времени будет убывать и тем самым восстанавливать равновесие при тех же величинах скоростей, которые были изначально. Если принять, что скорость затухания случайного отклонения скорости пропорциональна самому отклонению, то получим выражение

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

Физически осмысленным является лишь случай $b < 0$, поскольку в противном случае уравнение описывает неограниченное возрастание отклонения ε . Если в расчёте получено положительное значение b , то это может быть объяснено недостаточностью накопленных статистических данных.

Силу взаимосвязи $\mu_{c,d}$ между дорогой c и дорогой d оцениваем отношением коэффициента затухания b к стандартной ошибке определения этого коэффициента SE_b . Стандартная ошибка определения коэффициента линейной регрессии вычисляется как отношение среднеквадратичного отклонения остатка к среднеквадратичному отклонению фактора, следовательно

$$\mu_{c,d} = \left| \frac{b}{SE_b} \right| = \left| \frac{b \sigma(\varepsilon(t))}{\sigma(\delta(t))} \right|, t \in [t_{beg}, t_{end}],$$

где $\sigma(\cdot)$ – среднеквадратичное отклонение.

Для каждого значения τ из определенного интервала рассчитывается значение величины $\mu_{c,d}$. Затем отыскивается такая величина τ^* , при которой взаимосвязь $\mu_{c,d}$ максимальна. Если $\tau^* < 0$, то дорога c влияет на дорогу d , при $\tau^* > 0$ дорога d влияет на дорогу c . В случае τ^* , близкого к нулю, имеет место как влияние c на d , так и влияние d на c .

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 те отсчеты, в которых скорости на обеих дорогах ниже определенного порога.

3. Прогнозирование на основе характерной скорости движения

В основе предлагаемой модели прогнозирования лежит расчет медианной скорости движения с учетом цикличности действия внешних факторов на дорожную обстановку по времени суток, дням недели и времени года. Эту скорость будем называть характерной скоростью.

Экстраполяция характерной скорости может использоваться как самостоятельный инструмент прогнозирования. Однако в этом случае никак не учитывается дорожная ситуация, сложившаяся в городе на момент построения прогноза, поэтому точность такого прогноза невелика (п. 6). Вычисление характерной скорости может выступать и в качестве составной части более сложных моделей прогнозирования, рассмотренных ниже.

4. Прогнозирование на основе характерной скорости движения с учетом отклонений

При построении прогноза было бы некорректно рассматривать пару взаимосвязанных дорог изолированно от дорожной сети всего города. Между дорогами существуют множественные связи, и их учёт необходим для корректного предсказания дорожной обстановки. На одних дорогах текущая скорость может быть выше характерной, а на других ниже, но из-за влияния дорог друг на друга транспортный поток будет перераспределяться: скорости на недогруженных дорогах будут падать, а на перегруженных расти.

Обобщая линейный закон взаимосвязи отклонений скоростей от характерной скорости (п. 2) на случай нескольких дорог $c, d, \dots, y, z$, можем записать его в виде

$$U_c(t) = a_{c,d} U_d(t - \tau_{c,d}) + a_{c,e} U_e(t - \tau_{c,e}) + \dots + a_{c,z} U_z(t - \tau_{c,z}) + \varepsilon_c(t),$$

$$U_d(t) = a_{d,c} U_c(t - \tau_{d,c}) + a_{d,e} U_e(t - \tau_{d,e}) + \dots + a_{d,z} U_z(t - \tau_{d,z}) + \varepsilon_d(t),$$

...

$$U_z(t) = a_{z,c} U_c(t - \tau_{z,c}) + a_{z,d} U_d(t - \tau_{z,d}) + \dots + a_{z,y} U_y(t - \tau_{z,y}) + \varepsilon_z(t).$$

В построенную многомерную модель включаются лишь те дороги, для которых характеристика взаимосвязи μ оказывается выше определённого порога.

Поиск коэффициентов взаимосвязи a_{ij} осуществляется с помощью метода главных компонент [3], использующего сингулярное разложение матрицы данных.

На вход алгоритма подается матрица, содержащая отклонения скоростей от характерной скорости для всех дорог, включённых в модель, с соответствующими смещениями по времени τ :

$$\mathbf{U} = \begin{bmatrix} U_c(0) & U_c(1) & U_c(2) & \dots & U_c(t) \\ & \vdots & & \ddots & \vdots \\ U_z(0) & U_z(1) & U_z(2) & \dots & U_z(t) \end{bmatrix}.$$

В случае отсутствия данных о скоростях для некоторого момента времени, скорость принимается равной характерной и, следовательно, отклонение равным нулю.

Сингулярное разложение позволяет представить матрицу $\mathbf{U}$ в виде

$$\mathbf{U} = \mathbf{LSP}^T,$$

где $\mathbf{L}$ – матрица собственных векторов, $\mathbf{S}$ – матрица квадратов собственных чисел, $\mathbf{P}$ – матрица поворота. Величины, обратные к собственным числам матрицы $\mathbf{U}$, характеризуют рассеяние данных в направлениях, задаваемых матрицей поворота. Благодаря этому они могут использоваться для выявления главных и второстепенных факторов, определяющих рассеяние данных. Это, в свою очередь, позволяет выделить среди коэффициентов взаимосвязи дорог a_{ij} лишь те, которые существенны для построения прогноза.

Число главных компонент определяется экспериментально. Собственные числа, соответствующие остальным компонентам, в матрице $\mathbf{S}$ обнуляются. После перемножения матриц $\mathbf{L}$, $\mathbf{S}$ и $\mathbf{P}^T$, содержащих только главные компоненты, получим матрицу $\mathbf{U}_{bal}$. В ней содержатся отклонения скорости, «сбалансированные» относительно предположения о линейном законе взаимосвязи между дорогами.

Прогнозная скорость V_p^c для дороги s через время θ рассчитывается с помощью уравнения вида

$$V_p^c(t + \theta) = V_{char}^c(t + \theta) + \beta(\theta) U_{bal}^c(t),$$

где V_{char}^c – характерная скорость на дороге c ; U_{bal}^c – «сбалансированное» отклонение скорости для дороги c ; $\beta(\theta)$ – функция, описывающая влияние «сбалансированного» отклонения скорости от характерной на прогнозную скорость для каждого θ .

Функция $\beta(\theta)$ принимается равной единице при $\theta = 0$ и равной нулю при больших значениях θ . Последнее означает, что на длительных интервалах прогнозирования нивелируется любое начальное отклонение скорости и, следовательно, прогноз строится исходя из одной лишь характерной скорости движения. Для всех промежуточных значений θ величина $\beta(\theta)$ рассчитывается из условия минимизации меры J (п. 1).

5. Прогнозирование на основе значимости и определенности истории

В основе этого метода лежит предположение, что если разделить все дорожные ситуации на группы со схожей историей, то у всех ситуаций внутри группы будет схожее развитие [4]. В качестве метода кластеризации может быть выбран метод k ближайших соседей [5]. В качестве «расстояния» может выступить мера различия двух ситуаций на дороге

$$J_{road} = \sum_{t=t_{beg}}^{t_{end}} (\ln V(t) - \ln V(t - \Delta t))^2.$$

Мера различия ситуаций, вычисленная по множеству дорог, имеющих наибольшее влияние на данную дорогу, равна сумме мер различия для всех дорог этого множества с весами, равными силе взаимосвязи дорог:

$$J_{area} = \frac{1}{l} \sum_{i=1}^l \mu J_{road}^i.$$

Здесь l – число пар отсчетов, в которые известно значение скорости для обеих сравниваемых ситуаций, J_{road}^i – мера различия двух ситуаций на дороге i ; I – число дорог, имеющих наибольшее влияние на данную дорогу.

Мера схожести ситуаций является 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 $[(t - T - n\rho), (t - n\rho)]$, где n – целое число. Кластеризации подлежат только ситуации, для которых выполнен ряд условий: полнота истории, значимость истории и определенность истории.

Полнота истории P оценивается числом измерений и равна

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

где N – число отсчетов в течение времени T . Значимость истории S оценивается числом ситуаций, попавших в группу. Определенность истории D оценивается близостью развития ситуации внутри группы и равна дисперсии величины $J_{area}[(t - T - n\rho), (t - n\rho)]$ по множеству ситуаций в группе.

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

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

6. Результаты численного эксперимента

Для экспериментального исследования точности предсказания скорости транспортных потоков был разработан программный комплекс, реализующий описанные в статье прогнозирующие модели. Исходный код программного комплекса написан на интерпретируемом языке *Python* [6]. Код состоит из 15 функционально замкнутых модулей и соответствующих модульных тестов. Для первоначальной отладки прототипов всех прогнозирующих моделей использовался специализированный язык математического моделирования *R* [7].

В качестве исходных данных была использована информация о скорости движения транспорта по наиболее загруженным дорогам Санкт-Петербурга за сентябрь-октябрь 2010 года по данным компании «СитиГИД» [8].

Абсолютные значения меры погрешности прогноза J (п. 1) лишь опосредованно характеризуют точность предсказания скорости, но позволяют сравнивать применяемые прогнозирующие модели между собой. В табл. 1 приведены значения меры J и соответствующие им оценки относительной погрешности расчёта времени движения δT для прогноза скорости согласно описанным в статье моделям.

Таблица 1.

Погрешности прогнозирования

Модель прогноза	J	$\delta T, \%$
Прогноз на основе средней скорости с учётом цикличности	0,081	33
Прогноз на основе характерной скорости	0,061	28
Прогноз на основе характерной скорости с учетом отклонений	0,041	22
Прогноз на основе значимости и определенности истории	0,034	20

Из табл. 1 видно, что использование характерной (медианной) скорости движения (п. 3) вместо средней позволяет повысить точность прогноза за счёт устранения влияния случайных выбросов скорости на получаемые предсказания. Тем не менее, дальнейшее увеличение достоверности прогноза обеспечивается лишь использованием более сложных математических моделей, учитывающих «сбалансированные» отклонения скорости от характерной (п. 4) либо использующих сравнение дорожной обстановки с аналогичными ситуациями в прошлом (п. 5). Последняя прогнозирующая модель является наиболее точной, однако предъявляет самые жёсткие требования к полноте истории и, следовательно, применима лишь для наиболее загруженных дорог города.

Заключение

Сравнение полученных результатов позволяет сделать вывод, что предложенные в статье математические модели прогноза обеспечивают в среднем более высокую точность, чем методы, основанные на экстраполяции средней или характерной скорости движения.

В дальнейшем планируется продолжить разработку новых моделей и протестировать уже существующие модели на большем интервале времени. Также планируется проверить работоспособность системы на данных, полученных в Москве и других городах.

Литература

1. Как работают Яндекс.Пробки. URL:
<http://company.yandex.ru/technologies/yaprobki/> (дата обращения: 01.03.2012)

2. IBM Traffic Prediction Tool. URL: 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., Springer, NY, 2002, - 487 p. - P. 64–68.
4. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989. — 587 с. - С. 282-291.
5. MacQueen J. Some methods for classification and analysis of multivariate observations. In Proc. 5th Berkeley Symp. on Math. Statistics and Probability, 1967, p. 281-297.
6. Python Programming Language – Official Website. URL: <http://www.python.org/> (дата обращения: 01.03.2012).
7. The R Project for Statistical Computing. URL: <http://www.r-project.org/> (дата обращения: 01.03.2012).
8. СитиГИД. Навигационная служба. URL: <http://www.probki.net/news.aspx> (дата обращения: 01.03.2012).

Prediction of traffic situation

07, July 2012

DOI: 10.7463/0712.0418480

Vyalyh K.M., Pavlov Y.N., Stepanova M.B.

Russia, Bauman Moscow State Technical University

kostyoskol@mail.ru

pavlov@bmstu.ru

mari-na-bzzz@yandex.ru

Several prediction models, methods of qualitative assessment of predictions and model selection methodology are presented in this article. Prediction accuracy was evaluated by comparing models with one other. Prediction accuracy of the described models was higher than the prediction of extrapolated average speed. Longtime testing, new models and factors affecting the traffic situation are planned to be considered.

Publications with keywords: [congestion forecasting](#), [time-series analysis](#), [principal component analysis](#), [road traffic](#)

Publications with words: [congestion forecasting](#), [time-series analysis](#), [principal component analysis](#), [road traffic](#)

References

1. *Kak rabotaiut Iandeks.Probki* [How does Yandex.Traffic Jams work]. Available at: <http://company.yandex.ru/technologies/yaprobki/>. Accessed March 01, 2012.
2. *IBM Traffic Prediction Tool*. Available at: http://www.ibm.com/smarterplanet/us/en/traffic_congestion/nextsteps/index.html. Accessed March 01, 2012.
3. Jolliffe I.T. *Principal Component Analysis*. 2nd ed. New York, Springer, 2002. 502 p. (*Springer Series in Statistics*).
4. Aivazian S. A., Bukhshtaber V.M., Eniukov I.S., Meshalkin L.D. *Prikladnaia statistika: Klassifikatsiia i snizhenie razmernosti* [Applied statistics: Classification and dimension reduction]. Moscow, Finansy i statistika Publ., 1989. 587 p.
5. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability*. Vol. 1. Berkeley, Univ. of Calif. Press, 1967, pp. 281-297.

6. *Python Programming Language*. Available at: <http://www.python.org/>. Accessed March 01, 2012.
7. *The R Project for Statistical Computing*. Available at: <http://www.r-project.org/>. Accessed March 01, 2012.
8. *SitiGID. Navigatsionnaia sluzhba* [City guide. Navigation service]. Available at: <http://www.probki.net/news.aspx>. Accessed March 01, 2012.