

Обоснование шага сетки электронной карты при установлении факта попадания точки в контрольную область

Н.М. Нечитайло¹, В.Л. Панасов²

¹Российский университет транспорта

²Ростовский государственный университет путей сообщения

Аннотация: Рассмотрена задача обоснования шага сетки электронной карты, используемой для установления факта попадания транспортного средства, координаты которого считываются с бортового GPS-датчика, в контрольную область. Предлагаемая система поддержки принятия решений определяет максимально допустимый шаг разметки географической области для последующей обработки информации, при котором объём результирующих данных приближен к минимальному с учётом сохранения времени обработки и допустимого уровня погрешности данных для аналитики. Для решения задачи предлагается применить аппарат регрессионного анализа для определения зависимости ожидаемой ошибки вердикта от шага сетки электронной карты. Для доказательства применимости предлагаемого аппарата использован анализ остаточных дисперсий. Для оценки доверительного интервала ошибочных вердиктов использована формула Лапласа.

Ключевые слова: шаг сетки, доверительная вероятность, доля ошибочных вердиктов, регрессионный анализ, коэффициент детерминации.

Рассмотрена задача обоснования шага сетки электронной карты используемой для установления факта попадания транспортного средства, координаты которого считываются с бортового GPS-датчика, в контрольную область [1-3]. Задача учёта транспортных средств попавших в контролируемую область решается, например, ООО «РТ-Инвест Транспортные Системы» для последующего взимания платы за использование транспортными средствами федеральных дорог. Требуется разработать систему поддержки принятия решений, которая определяет максимально допустимый шаг разметки географической области для последующей обработки информации, при котором объём результирующих данных приближен к минимальному с учётом сохранения времени обработки и допустимого уровня погрешности данных для аналитики [4-6]. Реализация данного подхода позволит снизить издержки на хранение и обеспечить более

эффективное управление дорожной инфраструктурой в условиях растущих объёмов транспортных потоков [7, 8].

Входные данные:

- набор координат с бортовых компьютеров ТС;
- набор данных – список координат угловых точек зон, принадлежность к которым определяется у координат ТС;
- шаг разметки географической области x ;

Выходные данные:

- число входящих в географическую зону ТС;
- вердикт о вхождении в географическую зону каждого конкретного ТС.

Математическая постановка задачи: найти такое значение шага сетки x , при котором доля ошибочных решений $P_{ош}$ не превышает заданного значения P_d :

$$P_{ош} = f(x) \leq P_d, x \rightarrow \max$$

где $p_{ош} = \frac{1}{n} \left(\sum_{i=1}^n \left(\frac{m_{i_{ош}}}{m_i} \right) \right) = \frac{1}{n} \sum_{i=1}^n p_i$ – средняя доля ошибочных решений за исследуемый период;

n – число опытов;

p_i – доля ошибок i -го опыта;

m_i – доля общих решений;

$m_{i_{ош}}$ – доля ошибочных решений);

x – шаг разметки.

Для определения зависимости $P_{ош} = f(x)$ предлагается применить аппарат регрессионного анализа (a_1 – угол наклона функции; a_0 – смещение), организовав «соревнование» моделей при применении на конкретных исходных данных-[9]. Каждый раз предлагается использовать модель с наибольшим коэффициентом детерминации (p_{pi} – значение, рассчитанное по уравнению регрессии): В ходе исследований была предпринята попытка использовать библиотеки, применяемые в рамках искусственного интеллекта

(sklearn). Однако проведённые эксперименты показали, что библиотеки, применяемые в системах ИИ, дают худшие результаты, имея при этом время вычислений кратно превышающее время расчётов при использовании классических методов. Очевидно, это объясняется заложенными в библиотеки ИИ градиентными методами, обладающих низкой вычислительной эффективностью.

$$r^2 = \frac{\sum_{i=1}^n (p_{pi} - \bar{p})^2}{\sum_{i=1}^n (p_i - \bar{p})^2}$$

линейная модель

$$p_{pi} = a_0 + a_1 x_i, i = 1 \dots n$$

$$a_1 = \frac{\sum_{i=1}^n p_i x_i - \sum_{i=1}^n p_i \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$a_0 = \frac{1}{n} \left(\sum_{i=1}^n p_i - a_1 \sum_{i=1}^n x_i \right)$$

гиперболическая модель

$$p_{pi} = a_0 + a_1 / z_i, i = 1 \dots n$$

$$z_i = 1/x_i$$

$$a_1 = \frac{\sum_{i=1}^n p_i z_i - \sum_{i=1}^n p_i \sum_{i=1}^n z_i}{\sum_{i=1}^n z_i^2 - (\sum_{i=1}^n z_i)^2}$$

$$a_0 = \frac{1}{n} \left(\sum_{i=1}^n p_i - a_1 \sum_{i=1}^n z_i \right)$$

экспоненциальная модель

$$p_{pi} = a_0 e^{a_1 l}, \quad a_0 = e^{b_0}$$

$$a_1 = \frac{\sum_{i=1}^n \ln(p_i) x_i - \sum_{i=1}^n \ln(p_i) \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$b_0 = \frac{1}{n} \left(\sum_{i=1}^n \ln(p_i) - a_1 \sum_{i=1}^n x_i \right)$$

Усреднённые данные по большинству регионов РФ показывают, что x может находиться в пределах 50-800 м, а доля ошибочно принятых решений составляет 4,5-10,5 %.

На рис. 1 представлен график зависимости размеров SQL-файлов для хранения тестовых данных от величины шага.

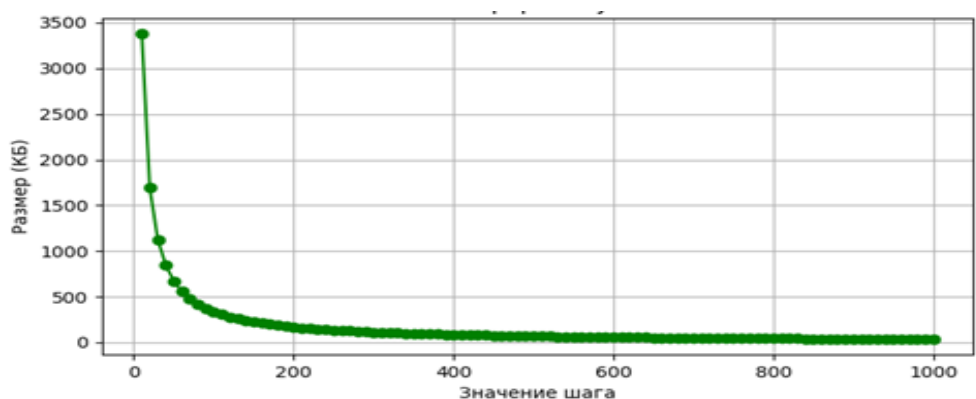


Рисунок 1 – Размеры SQL-файлов тестовых данных и величина шага

Как следствие подобная зависимость наблюдается и в отношении времени обработки информации.

Поскольку применение большинства методов прогнозирования исходит из предположения о гладкости функции связи между параметрами на исследуемом диапазоне данных (вторая производная на этом диапазоне не меняет знак), этот факт следовало установить. Для доказательства обоснованности использования методов регрессионного анализа был применён тест Голдфельда-Квандта – анализ остаточных дисперсий. Тест показал незначительный разброс дисперсий начального и конечного участков исследуемых диапазонов [9, 10].

Чтобы показать выполнение требования $P_{ош} = f(x) \leq P_d$ был проведён эксперимент (28 измерений при $\gamma = 272$ м и при $P_d = 0,6$) на электронной карте одной из областей РФ. Расчёт доверительного интервала проведён по формуле Лапласа (рис. 2). Результат показал, что с вероятностью 0,95 верхняя граница среднего значения доли ошибочных вердиктов не превышает директивно установленного значения.

	$\Delta = \frac{t \sqrt{\sigma^2}}{\sqrt{n}}$
$P_d =$	0,06
Довер. вер-ть =	0,95
$t =$	1,96
$n =$	28
$\Delta =$	0,0006049
$P_{ср} + \Delta =$	0,059998

Рисунок 2 – Расчёт доверительного интервала

Следует отметить, что в среднем уменьшение шага сетки на 2 % даёт сокращение времени обработки информации на 4-6 %, что объясняется квадратичной зависимостью размерности таблиц от шага сетки.

Литература

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. - 2-е изд. - М.: "Финансы и статистика", 1986. - 364 с.
2. Дроздов А.А., Миронюк В.П., Цыплаков В.Ю. Повышение эффективности системы двухэтапной транспортировки: на примере управления твердыми муниципальными отходами // Инженерный вестник Дона. - 2012. - №4. URL: ivdon.ru/ru/magazine/issue/107
3. Ерёмин И.Р. Моделирование заполняемости платных парковок: регрессионный анализ с учетом клиентского поведения // Инженерный вестник Дона. – 2025 № 6. URL: ivdon.ru/ru/magazine/archive/n1y2025/9777
4. Иванова О.А., Крупко А.М. Исследование движения лесовозных автопоездов с применением систем GPS-мониторинга // Инженерный вестник Дона. – 2025. № 6. URL: ivdon.ru/ru/magazine/archive/n2y2014/2311
5. Jeffrey M. Wooldridge Introductory Econometrics: A Modern Approach, Seventh Edition // Cengage Learning – 2020. - p. 840.
6. Гарманов В.В., Осипов А.Г., Богданов В.Л. [и др.]; под редакцией В. В. Гарманова Географические информационные системы: учебное пособие / — Санкт-Петербург: СПбГАУ, 2022. — 133 с. — URL: e.lanbook.com/book/325910
7. Препарата Ф. Шеймос М. Вычислительная геометрия - М: Мир, 1989, - -478 с. ISBN: 5-03-001041-6
8. Капралов, Е.Г., Кошкарев А.В., Тикунов В.С.; Под ред. Тикунова В.С. Основы геоинформатики: В 2 кн. Кн. 1: Учеб. пособие для вузов /. — М.: Издательский центр «Академия», 2004. — 352 с., ISBN 5-7695-1443-4

9. Вентцель Е.С. Теория вероятностей: Учеб. для вузов. -М.: Высш.шк., 2001. - 575 с.
10. Кремер, Н.Ш. Теория вероятностей и математическая статистика: учебник для студентов вузов, обучающихся по экономическим специальностям / 3-е изд., перераб. и доп. - М.: ЮНИТИ-ДАНА, 2010. - 551 с. ISBN 978-5-238-01270-4

References

1. Drejper N., Smit G. Prikladnoj regressionny`j analiz [Applied regression analysis]. 2-e izd. M.: "Finansy i statistika", 1986. p. 364.
2. Drozdov A.A., Mironjuk V.P., Cyplakov V. Ju. Inzhenernyj vestnik Dona. 2012. №4. URL: ivdon.ru/ru/magazine/issue/107
3. Erjomin I.R. Inzhenernyj vestnik Dona. 2025. № 6. URL: ivdon.ru/ru/magazine/archive/n1y2025/9777
4. Ivanova O.A., Krupko A.M. Inzhenernyj vestnik Dona. 2025 № 6. URL: ivdon.ru/ru/magazine/archive/n2y2014/2311
5. Jeffrey M. Shhooldridge Introductory Econometrics: A Modern Approach, Seventh Edition. Cengage Learning. 2020. p. 840
6. Garmanov V.V., Osipov A.G., Bogdanov V.L. [i dr.]; pod redakciej Garmanova V.V. Geograficheskie informacionny`e sistemy: uchebnoe posobie [Geographical information systems: a textbook]. Sankt-Peterburg: SPbGAU, 2022. 133 p. URL: e.lanbook.com/book/325910
7. Preparata F., Shejmos M. Vy`chislitel`naya geometriya [Computational geometry]. M: Mir, 1989, p. 478.
8. .Kapralov E.G., Koshkarev A.V., Tikunov V.S.; Pod red. Tikunova V.S. Osnovy` geoinformatiki: V 2 kn. Kn. 1: Ucheb. posobie dlya vuzov [Fundamentals of Geoinformatics: In Book 2, Book 1: Textbook for universities]. M.: Izdatel'skij centr «Akademija», 2004. p. 352. ISBN 5-7695-1443-4



9. Ventcel' E.S. Teoriya veroyatnostej: Ucheb.dlya vuzov [Probability theory: Textbook.for universities]. M.: Vyssh.shk., 2001. p. 575.
- 10.Kremer N.Sh. Teoriya veroyatnostej i matematicheskaya statistika: uchebnik dlya studentov vuzov, obuchayushhixsya po e`konomicheskim special`nostyam [Probability theory and Mathematical Statistics: a textbook for university students studying economics]. 3-e izd., pererab. i dop. M.: JuNITI-DANA, 2010. p. 551. ISBN 978-5-238-01270-4.

Дата поступления: 11.06.2025

Дата публикации: 25.07.2025