

УДК 519.7:004
ББК 22.17

МАССОВАЯ ОЦЕНКА МНОГОПАРАМЕТРИЧЕСКИХ ОБЪЕКТОВ ПРИ ДИАПАЗОННОМ ЗАДАНИИ ЗАВИСИМОЙ ПЕРЕМЕННОЙ

Корноушенко Е. К.¹

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)*

Массовая оценка объектов в рамках статистического подхода проводится путем построения различных регрессионных моделей и использования итерационных процедур для нахождения коэффициентов этих моделей. Практическое применение таких моделей может быть сопряжено с определенными трудностями (прежде всего из-за слабой помехоустойчивости и моделей, и процедур). В данной работе рассматривается переход от непрерывной зависимой переменной регрессионной модели к дискретной (диапазонной), что не требует нахождения коэффициентов модели и повышает её помехоустойчивость. При этом вместо итерационных процедур можно применять любой алгоритм классификации. В работе предлагается несложный алгоритм классификации, обладающий рядом важных особенностей. Работа алгоритма показана на практическом примере массовой оценки квартир в различных районах г. Сочи.

Ключевые слова: регрессионная модель, массовая оценка, диапазонная целевая переменная, классификация.

1. Введение

Массовая оценка объектов используется в тех случаях, когда большое количество однотипных объектов с аналогичными признаками необходимо оценить с использованием того или иного критерия, а индивидуальный подход к анализу каждого объекта нецелесообразен (или невозможен) по экономическим

¹ Евгений Константинович Корноушенко, доктор технических наук, главный научный сотрудник (ekorno@mail.ru).

соображениям. Подобные ситуации возникают во многих практических задачах (маркетинговых, финансовых, экономических и т.д.). Типичнейшим примером в этом плане является массовая (кадастровая) оценка объектов недвижимости: каждый объект характеризуется совокупностью описывающих его признаков, а в качестве критерия оценки используется тот или иной стоимостный показатель (удельная или полная стоимость объекта, арендная плата и т.п.). В принципе, для оценки объектов недвижимости существует несколько подходов. При статистическом подходе используются регрессионные модели оценивания, что предусматривает наличие исходных выборок с рыночными данными на каждой оцениваемой территории. По каждой такой выборке строится регрессионная модель оценки того или иного типа, которая далее используется для оценки объектов на данной территории. Для нахождения коэффициентов модели применяется метод наименьших квадратов (линейный или нелинейный в зависимости от типа модели), а также итерационные методы Ньютона и Левенберга—Маркуардта. Вопросы построения и анализа регрессионных моделей достаточно полно освещены в литературе.

Несмотря на отработанность технологий построения регрессионных моделей, доступных в различных статистических пакетах (SPSS, SAS, Matlab, NCSS и др.), регрессионным моделям присущ весьма важный недостаток: зависимость коэффициентов модели (и, соответственно, качества оценки) от возмущений, обусловленных

а) техническими причинами: неудовлетворительным качеством исходных данных, наличием помех, неполнотой рыночных данных и т.п.;

б) аналитическими причинами: отличием отображения значений признаков объектов в значения зависимой переменной, реализуемого построенной моделью, от аналогичного (неизвестного) отображения на оцениваемых объектах (такое отличие называется «смещением» (bias) выборки). Подобное отличие, неизбежно присутствующее в практических ситуациях, подробно исследовано в работах [7, 8]; Для адаптации к «смещению» выборки при проведении массовой оценки в работе [3] в рамках регрессионного подхода предложен многомодельный подход

к массовой оценке с использованием нескольких регрессионных моделей. При этом вместо одного «эталонного» отображения, с которым сравниваются отображения оцениваемых на данной территории объектов, появляется несколько «эталонных» отображений. Множество оцениваемых объектов путем предварительной классификации разбивается на подмножества, на каждом из которых «работает» соответствующая регрессионная модель оценки, что позволяет полнее использовать доступную рыночную информацию;

в) экономическими причинами: в изменяющейся экономической ситуации стоимость объектов недвижимости зависит не только от «физических» ценообразующих факторов, но и от изменяющейся экономической ситуации, влияющей на состояние рынка недвижимости. В работах Л.Н. Ясницкого и его коллег [1, 6, 14] показано, как изменения за десятилетие (2005–2015 гг.) таких экономических факторов как цена на нефть, темп роста ВВП, соотношение рубль/доллар и др. влияли на стоимость недвижимости. Подобные факты говорят о том, что для повышения точности оценки в коэффициенты регрессионной модели целесообразно вводить функциональные зависимости влияния того или иного экономического фактора на стоимость оцениваемых объектов. В отличие от обычных («статических») регрессионных моделей модели с такими зависимостями предложено называть «динамическими».

Проведенная в РФ в последние годы регрессионная (кадастровая) оценка объектов недвижимости вызвала массу нареканий, что обусловлено прежде всего слабой развитостью рынков недвижимости в регионах РФ (и, соответственно, плохим качеством используемых регрессионных моделей оценки). В качестве одного из выходов из такой ситуации управляющими органами РФ рассматривается переход от индивидуальных оценок объектов к формированию так называемых ценовых зон¹. Оценка объектов с точностью до попадания в ценовую зону существенно повышает помехоустойчивость оценивания.

¹ Важность такого перехода отражена в постановлении Минэкономразвития РФ (приказ от 7 июня 2016 г. №358).

Методологические вопросы организации и анализа ценовых зон рассмотрены, в частности, в [5]. В первом приближении ценовая зона представляет собой соответствующий участок карты территории с нанесенными на нее границами зоны по географическому признаку. Однако при более детальном анализе распределения цен внутри ценовых зон могут быть обнаружены факты недопустимого расхождения реальных цен объектов с граничными ценами для рассматриваемой «географической» зоны. Важность такого несоответствия весьма критична для (индивидуального) жилищного строительства и личных подсобных хозяйств. По этой причине приходится применять дополнительные меры для уменьшения разброса реальных цен внутри ценовой зоны: выделять внутри зоны особые участки, полнее учитывать рыночную информацию путем включения в регрессионные модели оценки дополнительных факторов влияния (см. выше), усложнять методологию построения моделей оценки (вплоть до использования нейросетевых технологий [6]). К сожалению, такие меры, требующие определенной квалификации оценщиков и соответствующей вычислительной базы, могут быть эффективно реализованы лишь на хорошо развитых рынках недвижимости.

В данной работе для уменьшения количества объектов с ценовыми «выбросами» внутри ценовой зоны предлагается ценовой диапазон каждой зоны определять по рыночной информации, собранной на рассматриваемом рынке недвижимости. Переход к дискретной (категориальной) зависимой переменной при проведении массовой оценки предусматривает замену регрессионных моделей каким-либо алгоритмом (алгоритмами) классификации. Практическая доступность такого подхода определяется, в частности, вычислительной сложностью используемого алгоритма классификации. В данной работе предлагается алгоритм классификации, описанный в [4] и характеризующийся рядом практически важных особенностей. Прежде всего, вместо операции сравнения векторов значений признаков объектов и использования понятия «ближайшей окрестности», на котором базируется большинство известных алгоритмов классификации, в данном алгоритме производится независимое попарное сравнение соответствующих значений

признаков сравниваемых векторов, что не требует предварительной нормализации этих векторов (и упрощает процедуру сравнения качественных признаков). Для значения каждого признака вводится понятие «допустимой близости» других значений этого признака. При этом вместо ближайшей (по выбранной метрике) окрестности классифицируемого объекта исследуется найденная для него совокупность объектов, в которой каждый из объектов «допустимо близок» к классифицируемому объекту по некоторому признаку (признакам). Такая совокупность больше, чем ближайшая окрестность классифицируемого объекта, и имеет больше «возможностей» для правильной классификации.

В качестве примера в рамках данного подхода с использованием диапазонов зависимой переменной рассматривается массовая оценка удельной стоимости квартир (данные 2013 г.) в многоквартирных домах в разных районах г. Сочи¹. Полученная точность отнесения квартир к назначенным четырем ценовым зонам (точность классификации) равна 87,3%, что является неплохим результатом в шкале оценок алгоритмов классификации и свидетельствует о практических перспективах данного подхода.

Замечание 1. В принципе, задача выбора диапазонов для значений зависимой переменной в регрессионной модели может быть решена с использованием соответствующей логистической модели (см., например, [3]), применяемой для классификации объектов. Нецелесообразность практического использования логистической модели при наличии качественных признаков в описаниях объектов обоснована в [12], где указываются осложнения, возникающие при переводе качественных признаков в соответствующие двоичные переменные. Основную трудность при этом представляет этап нахождения коэффициентов логистической модели. Известны различные итеративные процедуры для нахождения коэффициентов [10], и применение этих проце-

¹ На примере этой выборки в [3] рассматривается преимущество многомодельного подхода к массовой оценке по сравнению с традиционным одномодельным подходом.

дур требует определенной подготовки у пользователя и наличия соответствующих вычислительных средств. Все это указывает на практическую нецелесообразность использования логистических моделей в массовой оценке.

2. Классификация объектов при диапазонном задании целевой переменной

2.1. ИСХОДНАЯ ИНФОРМАЦИЯ И ПОСТАНОВКА ЗАДАЧИ

Исходной информацией является обучающая выборка (ОВ) из n объектов и тестовая выборка (ТВ) из p объектов, причем каждый из объектов этих выборок имеет m признаков X_1, X_2, \dots, X_m . Для краткости вектор значений признаков всякого объекта назовем *описанием* этого объекта. Признаки могут быть количественными (непрерывными, с интервалами возможных действительных значений) и качественными (дискретными, с конечными множествами значений – целочисленными, вербальными и т.п.). Качественные признаки могут быть номинальными, когда значения признака несравнимы, или ординальными, когда на значениях признака введено отношение порядка¹. Для всех объектов выборки считаются известными соответствующие значения зависимой переменной, которую далее назовем *целевой* (*target*) переменной. Будем считать, что диапазон значений целевой переменной для объектов ОВ разбит на конечное число непересекающихся поддиапазонов, количество и размеры которых определяются спецификой решаемой задачи и требованиями пользователя (так что вопросы оптимизации вводимых разбиений здесь не рассматриваются). При введении того или иного разбиения желательно выполнение очевидных требований, чтобы каждый из поддиапазонов не был слишком «узким» и на нем наблюдалось заметное изменение целевой переменной. Таким образом, каждый объект ОВ однозначно принадлежит некоторому поддиапазону – блоку введенного на ОВ разбиения $\pi_{ОВ}$, и номер этого блока как *метка* (*label*) приписывается дан-

¹ В данной работе все значения ординального признака рассматриваются как независимые номинальные признаки.

ному объекту. Разбиение $\pi_{ОВ}$ индуцирует аналогичное разбиение $\pi_{ТВ}$ целевой переменной на объектах ТВ путем сохранения одних и тех же значений целевой переменной, определяющих границы блоков в разбиении $\pi_{ОВ}$, для границ блоков разбиения $\pi_{ТВ}$. При этом блоки разбиений $\pi_{ОВ}$ и $\pi_{ТВ}$ с одними и теми же границами имеют одинаковый номер в этих разбиениях, хотя такие блоки могут содержать разное число объектов в ОВ и ТВ.

Рассматриваемая ниже задача классификации состоит в том, чтобы при использовании описаний объектов ОВ и ТВ и известного разбиения $\pi_{ОВ}$ «вложить» каждый объект ТВ в некоторый блок разбиения $\pi_{ТВ}$ таким образом, чтобы метка этого объекта как результат процедуры классификации совпадала с номером блока разбиения $\pi_{ТВ}$, содержащего этот объект.

2.2. СУТЬ ПРЕДЛАГАЕМОГО АЛГОРИТМА КЛАССИФИКАЦИИ

Как уже сказано во введении, отличие используемого ниже алгоритма от известных алгоритмов классификации состоит в том, что вместо использования понятия «ближайшей окрестности» для тестового объекта для каждого значения признаков в тестовом объекте вводится понятие «допустимой близости» значений соответствующих признаков у объектов ОВ. Два действительных значения x_1 и x_2 назовем *допустимо близкими*, если расстояние между ними не больше некоторой заданной величины d (о выборе значения d см. далее). Для каждого значения x каждого признака X объекта Z из ТВ определяется множество $V(x)$ допустимо близких значений признака X у объектов ОВ. При этом объекты ОВ, у которых значения признака X принадлежат множеству $V(x)$, образуют совокупность $W(x)$. Поскольку каждый объект из ОВ имеет однозначно определенную метку, множество меток объектов, входящих в $W(x)$, образует совокупность $C(x)$, что для наглядности можно отобразить как $x \rightarrow V(x) \rightarrow W(x) \rightarrow C(x)$. Подобная процедура повторяется для всех значений всех признаков классифицируемого объекта Z из ТВ, так что объекту Z с описанием (x_1, \dots, x_m) сопоставляется ансамбль меток $\{C(x_1), \dots, C(x_m)\}$, по которому определенным образом строится матрица весов размера $(K \times m)$ для объекта Z ,

где K – число классов в разбиении π_{OB} . К строкам этой матрицы применяется нелинейная монотонная свертка (о чем сказано ниже), и номер строки с наибольшим значением такой свертки интерпретируется как метка класса, к которому относится тестовый объект Z из ТВ.

2.3. ЭТАПЫ АЛГОРИТМА КЛАССИФИКАЦИИ

А. Определение понятия допустимой близости значений количественных признаков. Для каждого количественного признака X_k определяются его максимальное X_{kmax} и минимальное X_{kmin} значения по всем объектам ОБ и ТВ. Разность этих значений делится на некоторое выбираемое число h (о выборе значения h см. ниже – п. В). Обозначим $\Delta X_k = X_{kmax} - X_{kmin}$ и $d_k = \Delta X_k/h$. Два значения признака X_k назовем d_k -близкими, если модуль их разности не больше d_k . Выбор значения¹ h существенно влияет на качество алгоритма классификации. В данном алгоритме используется простейший вариант – независимость h от k . Качественные признаки здесь рассматриваются как номинальные, т.е. при сравнении двух значений всякого качественного признака X_q эти значения могут либо не совпадать, либо совпадать (при этом $d_k = 0$ не зависимо от h).

Б. Построение матрицы весов для тестового объекта Z из ТВ. Пусть O_z – описание тестового объекта Z , x_{zj} – численное значение признака X_j (количественного или качественного) в описании O_z , а d_j – заданная величина допустимой близости для значений признака X_j . Совокупность значений признака X_j в описаниях объектов ОБ, таких, что эти значения d_j -близки к значению x_{zj} , назовем d_j -окрестностью значения x_{zj} и обозначим эту совокупность как $V(x_{zj}, d_j)$, а совокупность объектов из ОБ, у которых значение признака X_j принадлежит $V(x_{zj}, d_j)$, – как $W(x_{zj}, d_j)$. Пусть M_{zjv} – число объектов из $W(x_{zj}, d_j)$, входящих в блок C_v разбиения π_{OB} . Сопоставим значению x_{zj} число $I_{zjv} = M_{zjv}/(|W(x_{zj}, d_j)| |C_v|)$, где $|W(x_{zj}, d_j)|$ и $|C_v|$ – соответственно

¹ В принципе, для каждого признака X_k можно выбирать свое h_k , но это ведет к значительному увеличению вычислительной сложности алгоритма.

количество элементов в множествах $W(x_{ij}, d_j)$ и C_ν , $1 \leq \nu \leq K$. Найденное число $I_{zj\nu}$ будем рассматривать как вес метки ν в множестве меток объектов из $W(x_{zj}, d_j)$. Структура числа $I_{zj\nu}$ имеет две особенности:

а) отношение $M_{zj\nu}/|W(x_{zj}, d_j)|$ является, по существу, показателем доверия (*confidence*¹) к утверждению, что всякий объект из ОВ со значением признака X_j , d_j — близким к x_{zj} , относится к классу² C_ν ;

б) показатель доверия для класса C_ν взвешивается с множителем $1/|C_\nu|$. Такое взвешивание показателей доверия для разных меток объектов из $W(x_{zj}, d_j)$ введено для коррекции весов $I_{zj\mu}$ соответствующих меток, относящихся к классам различной мощности³.

Подобным образом сформируем множества $M_{zj\mu}$ для объектов из ОВ, содержащих значения признака X_j , d_j — близкие к значению x_{zj} , но принадлежащих другим классам C_μ , $\mu \neq \nu$, и определим соответствующие числа $I_{zj\mu}$. Сформируем K -столбец весов $H_{zj} = (I_{zj1}, I_{zj2}, \dots, I_{zjK})^T$. Заметим, что все координаты в H_{zj} неотрицательные и не большие единицы. Аналогичным образом рассмотрим другие множества $M_{z_s\mu}$, где x_{z_s} — значение признака X_s в описании O_z тестового объекта Z , и найдем для значений признаков, содержащихся в O_{z_s} соответствующие вектор-столбцы $H_{z_s} = (I_{z_s1}, I_{z_s2}, \dots, I_{z_sK})^T$, $s = 1, \dots, m$. Сформируем матрицу $M_Z = (H_{z1}, H_{z2}, \dots, H_{zm})$, в которой m столбцов (по числу признаков в O_z) и K строк (по числу классов в разбиении $\pi_{ОВ}$). Назовем M_Z матрицей весов объекта Z . Элементы каждой стро-

¹ Понятие «confidence» используется во многих алгоритмах классификации (см., например, [9, 15]), где подчеркивается важность использования данного понятия для повышения точности классификации.

² Другими словами, это означает степень «доверия» к утверждению, что класс C_ν , содержащий некоторые объекты из ОВ, входящие в $W(x_{zj}, d_j)$, содержит и тестовый объект Z в разбиении $\pi_{ТВ}$.

³ Как показывают многочисленные эксперименты, подобное взвешивание показателей «доверия» способствует повышению точности классификации данного алгоритма.

ки $M_{Z\mu}$, $1 \leq \mu \leq K$, матрицы M_Z неотрицательны и не большие единицы. При этом значения ненулевых элементов в каждой её строке $M_{Z\mu}$ пропорциональны значениям $M_{Zs\mu}$, $s = 1, \dots, m$, а распределение ненулевых элементов по строке $M_{Z\mu}$ характеризует распределение метки класса C_μ по признакам объекта Z . Это позволяет рассматривать строку $M_{Z\mu}$ как совокупность локальных классификаторов для объекта Z , «настроенных» по каждому признаку на метку μ . В работах [2, 11] говорится о том, что классифицирующая способность совокупности локальных классификаторов может быть усилена путем введения какой-либо монотонной нелинейной функции агрегирования. Согласно этой рекомендации применим к каждой строке $M_{Z\mu} = (H_{Z1\mu}, H_{Z2\mu}, \dots, H_{Zm\mu})$ следующую монотонную нелинейную функцию агрегирования:

$$(1) \quad E(M_{Z\mu}) = \left(- \sum_{s=1}^m H_{Zs\mu} \log_2(H_{Zs\mu}) \mid H_{Zs\mu} > 0 \right), 1 \leq \mu \leq K.$$

Поскольку значения $H_{Zs\mu}$ не являются в общем случае вероятностными величинами¹, выражение (1) назовем *псевдоэнтропией* строки $M_{Z\mu}$. Результатом применения свертки (1) к каждой строке матрицы M_Z является вектор $G_Z = (E(M_{Z1}), \dots, E(M_{ZK}))$, который назовем *классифицирующим вектором* для объекта Z . Номер координаты вектора G_Z с наибольшим значением принимается в качестве искомой метки, приписываемой объекту Z . Аналогичным образом производится классификация остальных объектов ТВ (а также объектов, требующих оценки).

Точность алгоритма классификации определяется как отношение числа правильно классифицированных объектов к длине ТВ.

В. Выбор величины h , определяющей величину допустимой близости значений количественных признаков. Этот этап является очень важным, поскольку от выбора значений допустимой

¹ Условие нормировки $\sum_{s=1}^m H_{Zs\mu} = 1$ для переменных $H_{Zs\mu}$ может не выполняться.

близости зависит точность классификации предлагаемого алгоритма. Как уже сказано в п. А, величина d_j допустимой близости значений признака X_j определяется как $d_j = \Delta X_j / h$, $j = 1, \dots, m^1$. Независимость h от признаков позволяет определять приемлемое значение h при достижении приемлемой точности классификации в виде простейшей процедуры одномерного поиска. Как показано в приводимом ниже примере, несмотря на простоту нахождения h , результирующая точность классификации может оказаться весьма неплохой. Таким образом, данный алгоритм следует отнести к алгоритмам, в структуре которых принципиальным образом используется обратная связь по точности классификации (*wrapper approach*).

Вычислительная сложность алгоритма определяется как $O(n^2 m)$. Структура алгоритма содержит лишь циклы, на которых определены соответствующие арифметические вычисления, управляемые условными операторами. Это позволяет реализовать алгоритм в пакете Excel без привлечения более сложных компьютерных средств, что упрощает его практическое использование.

3. Пример: массовая оценка квартир в разных районах г. Сочи при диапазонном задании стоимостей

Исходной информацией является выборка с описаниями квартир в многоквартирных домах, расположенных в разных районах г. Сочи. В качестве ценообразующих признаков выбраны следующие признаки: 1 – площадь квартиры; 2 – район местоположения дома с оцениваемой квартирой; 3 – расстояние от такого дома до делового центра; 4 – расстояние до ж.-д. станции; 5 – расстояние до положительного центра притяжения (рекреации); 6 – расстояние до берега моря. В качестве целевой переменной рассматривается удельная стоимость квартир (дан-

¹ Для качественных признаков, рассматриваемых здесь как номинальные признаки, полагаем $d = 0$.

ные 2013 г.). ОВ содержит 101 квартиру с известной рыночной стоимостью, а ТВ — 401 квартиру.

При назначении диапазонов стоимости для объектов ОВ известные значения удельной стоимости объектов ОВ сортировались по возрастанию, и на результирующей кривой были выбраны 4 диапазона удельной стоимости (руб/кв.м) (ниже после номера диапазона указаны границы удельной стоимости для этого диапазона, а в скобках – число квартир из ОВ, попадающих в этот диапазон): 1 – не более 50 000 (16 квартир); 2 – (50 000, 65 000] (22 квартиры); 3 – (65 000, 80 000] (37 квартир); 4 – более 80 000 (26 квартир).

Работа описанного выше алгоритма классификации начинается с определения величины h (см. пп. А, В) для случая независимости h от признаков. При применении процедуры одномерного поиска используется эвристическое предположение (неоднократно подтверждающееся на практике), что функция, описывающая зависимость точности классификации от h , унимодальна «в большом»¹. В таком случае поведение значений точности классификации в зависимости от h на выбранном интервале указывает направление выбора следующего (суживающегося) интервала. В результате такого поиска было найдено значение $h = 8$, при котором точность классификации, равная 87,3%, считалась приемлемой².

Важной характеристикой всякого алгоритма классификации является его «способность» правильно классифицировать объекты разных классов, несмотря на существенные отличия их мощности. Результаты классификации объектов ТВ и их распределение по диапазонам приведены в таблице 1.

Приведенные характеристики алгоритма свидетельствуют о его практической пригодности для решения задач массовой оценки при задании (или выборе) диапазонов значений целевой переменной.

¹ То есть нарушения монотонности на каждой из ветвей этой унимодальной функции при некоторых h носят локальный характер.

² В принципе, поиск можно продолжить вокруг точки $h = 8$ и т. д.

Таблица 1. Распределение объектов ТВ по диапазонам до и после классификации

Диапазоны	1	2	3	4
До классификации	24	96	228	53
После классификации	0	74	225	51
% заполнения диапазона на ТВ	0	77,1	98,7	96,2

Литература

1. АЛЕКСЕЕВ А.О., ХАРИТОНОВ В.А., ЯСНИЦКИЙ В.Л. *К вопросу об интеллектуальном анализе, массовой оценке и управлении рынком недвижимости регионов России* // Прикладная математика и вопросы управления. – 2017. – №1. – С. 87–99.
2. ВОРОНЦОВ К.В. *Комбинаторный подход к оценке качества обучаемых алгоритмов* // Математические вопросы кибернетики / Под ред. О.Б. Лупанова. – М.: Физматлит, 2004. – Т. 13. – С. 5–36.
3. КОРНОУШЕНКО Е.К. *Многомодельный подход в массовой оценке многопараметрических объектов* // Проблемы управления. – 2014. – №5. – С. 42–49.
4. КОРНОУШЕНКО Е.К. *Алгоритм классификации путем парного сравнения признаков* // Автоматика и телемеханика. – 2017. – №11 (в печати).
5. МИТИН Н.А., НЕЙМАН Е.И., СМОЛЯК А.С., ФИНГЕРТ А.Б. *Применение когнитивных технологий для оценки недвижимости* // Препринт ИПМ им. М.В. Келдыша РАН. – 2015. – №112.
6. ЯСНИЦКИЙ Л.Н., ЯСНИЦКИЙ В.Л. *Разработка и применение комплексных нейросетевых моделей массовой оценки и прогнозирования стоимости жилых объектов на примере рынков недвижимости Екатеринбурга и Перми* // Имущественные отношения в Российской Федерации. – 2017. – №3(186). – С. 68–84.
7. *Biased Sampling and Extrapolation.* – URL: <https://www.ma.utexas.edu/users/mks/statmistakes/biasedsampling.html> (дата обращения: 8.09.2017).

8. BREIMAN L. *Arcing Classifiers* // Annals of Statistics. – 1998. – Vol. 26, No. 3. – P. 801–849. – URL: www.cs.iastate.edu/~honavar/boost4.pdf (дата обращения: 8.09.2017).
9. JING LEI *Classification with Confidence* – URL: www.stat.cmu.edu/~jinglei/conf_class_R2.pdf (дата обращения: 8.09.2017).
10. *The Logistic Regression Analysis in SPSS*. – URL: www.statisticssolutions.com/the-logistic-regression-analysis-in-spss/ (дата обращения: 08.09.2017).
11. SPIRIN N., VORONTSOV K. *Learning to Rank with Nonlinear Monotonic Ensemble*. – URL: www.machinelearning.ru/wiki/images/9/96/Voron11spirin-eng.pdf (дата обращения: 8.09.2017).
12. SRINIVASAN M. *Is logistic regression really a good choice for categorical variables?* – URL: www.quora.com/Is-logistic-regression-really-a-good-choice-for-categorical-variables (дата обращения: 8.09.2017).
13. URALSKY D., MARTIN J.H. *Logistic Regression as a Classifier* // Speech and Language Processing. Draft. – 2015. – 423 p. – URL: <https://web.stanford.edu/~jurafsky/slp3/7.pdf> (дата обращения: 08.09.2017).
14. YASNITSKY L.N., YASNITSKY V.L. *Technique of design of integrated economic and mathematical model of mass appraisal of real estate property by the example of Yekaterinburg housing market* // J. of Applied Economic Sciences. – Winter 2016. – Vol. XI. – Iss. 8 (46). – P. 1519–1530.
15. ZARAGOZA H., D'ALCHE-BUC F. *Confidence Measures for Neural Network Classifiers*. – URL: www.hugo-zaragoza.net/academic/pdf/hugoz_ipmu98.pdf (дата обращения: 8.09.2017).

MASS APPRAISAL OF MULTIPARAMETER OBJECTS WITH RANGE DEPENDENT VARIABLE

Eugeny Kornoushenko, Institute of Control Sciences of RAS, Moscow, Doctor of Science, (Moscow, Profsoyuznaya st., 65, (495) 334-90-00).

Abstract: Statistical mass appraisal of objects is carried out by constructing various regression models and using iterative procedures to find the coefficients of these models. Practical application of such models can be fraught with certain difficulties (first of all, due to weak noise immunity of these models). In this paper, it is proposed in mass appraisal to move from a continuous dependent variable of regression model to a discrete one, which does not require finding the coefficients of the model and increases its noise immunity. Instead of iterative procedures, any classification algorithm can be used in this case. In comparison with the regression model, classification is a more convenient tool in practical applications, because various information processing mechanisms implemented in appropriate algorithms can be used. In connection with the unsatisfactory quality of the cadastral valuation of real estate conducted earlier in the Russian Federation, a course has been chosen by the Ministry of Economic Development to form and evaluate so called price zones. The value of the price range within the zone depends on the properties of the objects entering the zone. In this regard, the choice a discrete dependent variable in models can be practically used in the formation of price zones. In paper, we offer a simple classification algorithm for the formation of price zones. Application of the algorithm is shown by the practical example of a classification of apartments in various districts of Sochi.

Keywords: regression model, mass appraisal, range target variable, classification.

*Статья представлена к публикации
членом редакционной коллегии Д.А. Новиковым.*

*Поступила в редакцию 02.04.2017.
Опубликована 30.09.2017.*