

ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В КАДАСТРОВОЙ ОЦЕНКЕ НЕДВИЖИМОСТИ

PROSPECTS FOR THE APPLICATION OF MACHINE LEARNING METHODS IN THE CADASTRAL VALUATION OF REAL ESTATE

**O. Lepikhina
T. Baltyzhakova
I. Raguzin**

Summary. The article considers the possibilities of applying methods of machine learning in the field of cadastral evaluation. Methods applied nowadays for calculation of cadastral value do not allow to achieve steadily good results, therefore it is necessary to pass to more non-standard methods, such as methods of machine learning. The researches devoted to application of methods of machine learning at various stages of cadastral value estimation are analyzed. The article also considers in detail the cases of successful application of methods of machine learning for the evaluation of real estate in foreign countries.

Keywords: real estate, cadastral valuation, cadastral value, machine learning, regression analysis.

Лепихина Ольга Юрьевна

К.т.н., доцент, Санкт-Петербургский горный университет

Olgalepikhina1984@gmail.com

Балтыжакова Татьяна Игоревна

К.т.н., ассистент, Санкт-Петербургский горный университет

tatiyana.baltyzhakova@gmail.com

Рагузин Иван Игоревич

Санкт-Петербургский горный университет;

Аналитик, Городское управление кадастровой оценки,

Санкт-Петербург

Raguzin.iv@yandex.ru

Аннотация. В статье рассмотрены возможности применения методов машинного обучения в сфере кадастровой оценки. Применяемые в настоящее время методики для расчета кадастровой стоимости не позволяют добиться устойчиво хороших результатов, поэтому необходим переход к более нестандартным методам, таким как методы машинного обучения. Проанализированы исследования, посвященные применению методов машинного обучения на различных этапах оценки кадастровой стоимости. Также в статье подробно рассмотрены случаи успешного применения методов машинного обучения для оценки недвижимости в зарубежных странах.

Ключевые слова: недвижимость, кадастровая оценка, кадастровая стоимость, машинное обучение, регрессионный анализ.

Современная кадастровая система Российской Федерации является многоцелевой, то есть эта система выполняет целый ряд функций, одной из которых является фискальная. Фискальная функция кадастра выражается, прежде всего, в определении порядка налогообложения недвижимости и расчете величины налогов, а также налогооблагаемой базы. В настоящее время для определения величины налога на недвижимость используется кадастровая стоимость, которая является стоимостью объекта недвижимости, рассчитанной в соответствии с методикой государственной кадастровой оценки.

Согласно существующей методике наиболее широко применяемым является сравнительный подход, в рамках которого стоимость определяется методом регрессионного анализа. Безусловно предусмотрено и применение других подходов и методов в зависимости от оцениваемого сегмента, но именно регрессионный подход применяется для тех сегментов, которые являются наиболее социально значимыми с точки зрения налогообложения, так как налог на недвижимость в этих сегментах выплачивают граждане. К таким сегментам относятся, прежде всего, сегмент многоэтажной и среднеэтажной застрой-

ки, индивидуальной жилой застройки и земли для ведения садоводства, огородничества и дачного хозяйства.

Также необходимо отметить, что кроме численных методов при определении кадастровой стоимости широко применяются экспертные методы, что повышает степень неоднозначности результатов кадастровой оценки. Экспертные методы применяются, как правило, на подготовительных этапах оценки: группировка объектов, оценочное зонирование и выбор значимых ценообразующих факторов.

Количество оспариваний величины кадастровой стоимости недвижимости [3, 4], а также большое количество исследований, посвященных проблематике определения кадастровой стоимости, показывают, что необходимо применение новых и нестандартных подходов в этой сфере.

В последние годы широкое распространение во многих сферах науки и экономики получили методы машинного обучения, которые помогают обнаруживать неочевидные и объективные закономерности в имеющихся данных.

Всплеск интереса к подобным методам связан не только с их способностью обнаруживать ранее не выявленные закономерности, но и с тем, что в настоящее время накоплены значительные массивы доступной или легко поддающейся автоматизированному сбору информации. Так, например, информация о местоположении объекта недвижимости и его окружении может быть получена из общедоступных баз данных: OpenStreetMap, Google Карты, Яндекс.Карты, 2ГИС и прочие. Сведения о предложениях на рынке недвижимости могут быть собраны на основе открытых данных агентств недвижимости с помощью веб-скраперов — скриптов и программ, которые на основе массива ссылок агрегируют информацию по каждому элементу этого массива.

Для успешного применения методов машинного обучения необходимо, чтобы в исходных данных существовала некая скрытая закономерность или скрытый паттерн, который нельзя непосредственно сформулировать математически. Следует учитывать, что большое значение при применении методов машинного обучения несет интерпретация результатов, которая требует хорошего знания предметной области.

Все методы машинного обучения делятся на два вида: обучение с учителем (*supervised learning*) и обучение без учителя (*unsupervised learning*), кроме того, в последнее время в отдельные области стали выделять методы обучения с подкреплением (*reinforcement learning*) и нейронные сети.

Главное различие между методами обучения с учителем и методами обучения без учителя в наличии зависимой переменной или известных категорий принадлежности объектов. Например, если заранее известны классы и их признаки, то решается задача классификации — распределения объектов на группы (классы) по ряду признаков. В том случае, если имеются только объекты и их признаки, то встает задача кластеризации, то есть поиск групп схожих по своим характеристикам объектов.

В рамках различных исследований, посвященных повышению качества результатов кадастровой оценки, уже рассматривалось применение различных методов машинного обучения на отдельных этапах оценки.

Например, анализ рынка недвижимости для целей кадастровой оценки, выполняемый для группировки объектов недвижимости, может выполняться с применением методов кластеризации [2].

Вместо экспертного подхода для определения значимых ценообразующих факторов и выявления взаимосвязи между факторами возможно применение более объективных численных методов, например, методов факторного анализа [1, 5].

Также может быть улучшено качество самих моделей оценки кадастровой стоимости благодаря применению нейронных сетей [6, 7].

Зарубежный опыт в сфере оценки недвижимости также показывает необходимость применения нового, основанного на данных, подхода [8, 9].

Эта необходимость основана, прежде всего, на том, что при оценке недвижимости все больше возрастает значимость новых, нетрадиционных данных, таких как, например, количество кафе в окрестностях объекта, степень энергосбережения объекта, простота перемещения внутри офисного здания, общая оценка близлежащих сервисов и прочее [8]. Эти данные помогают выявить так называемые «гиперлокальные» закономерности, то есть закономерности, оказывающие влияние на конкретное сообщество, в противовес закономерностям, выделяемым на уровне города в целом, с помощью традиционных методов.

В качестве проверки эффективности основанного на данных подхода в исследовании [8] была рассчитана величина арендной платы за жилье в Сиэтле на 3-летний период. Для построения модели была создана база данных, сочетающая в себе как сведения о традиционных экономических факторах, влияющих на стоимость, так и новые нетрадиционные характеристики. К этим данным были применены методы машинного обучения для расчета. В результате величина арендной платы была определена с точностью до 90%, что намного превосходит точность традиционных подходов и методов.

Однако подобные методы применяются не только в Америке, так в Сингапуре модель, построенная методами машинного обучения, позволяет определять стоимость недвижимости с точностью 87% [9]. Для построения модели были собраны сведения о продажах квартир за последние 4 года и местоположение точек притяжения (кафе, рестораны, отели, остановки общественного транспорта и т.п.) из открытой карты OpenStreetMap. Одними из значимых ценообразующих факторов стали ранее не рассматриваемые, например, близость к почтовому отделению, близость к отелям, банкоматам и основным городским магистралям.

Точность модели более чем на половину обоснована фактором близости к почтовому отделению, который характеризует общую развитость инфраструктуры в районе. Еще 40% точности добавляется за счет включения в модель расстояния до основных точек притяжения.

Следует отметить, что в данном случае высокой точности модели удалось добиться только при помощи добавления одного дополнительного набора данных, получен-

ных только из одного источника. Тем не менее, в данной модели есть возможность ее улучшения, за счет добавления к данным о квартирах сведений о высоте здания, этаже размещения, годе постройки и даже инженерных коммуникациях. Сбор этих данных намного сложнее, чем тех, которые уже были использованы, но может привести к значительному улучшению точности модели [8].

Ценность моделей, построенных в Сиэтле и в Сингапуре не просто в точном определении стоимости существующей недвижимости, но и еще и в применении таких моделей для принятия инвестиционных и девелоперских решений.

При ее наличии модель может автоматически собирать и анализировать актуальные данные для выявления потенциально ценных, но неиспользуемых участков земли в городе, которые могут быть использованы для застройки. Также с помощью подобных основанных на актуальных данных моделей девелоперы могут понимать, что действительно влияет на стоимость недвижимости и что может ее повысить. Например, в Сингапуре было выявлено, что близость к отелям и ресторанам, является повышающим стоимость квартир фактором [8]. Для других случаев модель может быть «переобучена» на основе актуального для города набора данных, что позволит девелоперам выявить, какие меры могут повысить стоимость их объектов.

Анализ зарубежного опыта безусловно доказывает, что применение современных методов сбора и анализа информации для оценки недвижимости повышает эффективность и точность моделей. Однако необходимо отметить и сложности в применении подобных подходов в нашей стране.

Все подобные методы требуют большого количества высококачественных данных. Зачастую в условиях неустойчившегося и неразвитого рынка недвижимости у оценщиков нет возможности собрать актуальные и реальные значения цен сделок. Несмотря на создание и работу автоматизированной системы мониторинга рынка недвижимости, качество данных в ней очень невысокое с большим количеством фактических ошибок и отклонений от реальной ситуации на рынке. Необходимо отметить, что отсутствие качественных наборов данных и сбор данных значимая проблема, касающаяся не только сферы оценки.

В заключение следует отметить, что переход от традиционного подхода оценки кадастровой стоимости путем построения регрессионных моделей к методам машинного обучения позволит значительно улучшить объективность и точность результатов, но такой переход практически невозможен до тех пор, пока не сформируется стабильный рынок недвижимости, позволяющий собирать необходимое количество качественной информации для моделирования.

ЛИТЕРАТУРА

1. Балтыжакова Т. И. Выявление скрытых факторов кадастровой оценки земель населенных пунктов методом главных компонент // Инженерный вестник Дона. 2015. № 1. URL: ivdon.ru/ru/magazine/archive/n1y2015/2761 (дата обращения: 13.11.19)
2. Быкова Е. Н., Балтыжакова Т. И., Волкова Я. А. Классификация населённых пунктов по уровню развитости рынка земельных участков индивидуального жилого строительства // Известия Томского политехнического университета. Инжиниринг георесурсов. 2018. Т. 329. № 7, с. 17–30
3. Горшков С. И. Оспаривание результатов кадастровой оценки объектов недвижимости // Научные труды Северо-западного института управления. 2016. № 3(20), том 6. С. 246–254.
4. Григорьев В. В. Проблемы определения и оспаривания кадастровой стоимости недвижимости в Российской Федерации // Молодой ученый. 2015. № 21. С. 512–516.
5. Киселев В. А., Беляев В. В., Балтыжакова Т. И. Выявление устойчивых наборов значимых факторов кадастровой стоимости земли малых и средних городов Северо-Запада методом главных компонент // Имущественные отношения в Российской Федерации. 2013. № 2 (137), с. 65–74
6. Киселев В. А., Гринченко С. С. Обоснование выбора оптимальных показателей обучения искусственных нейронных сетей, сформированных для определения кадастровой стоимости земель населенных пунктов // Имущественные отношения в Российской Федерации. 2012. № 11 (134), с. 68–76
7. Ковязин В. Ф., Лепихина О. Ю., Зимин В. П. Разработка прогнозной модели стоимости земель моногородов с учетом экономических факторов деятельности градообразующих предприятий (на примере Мурманской области) // Вестник Московского государственного областного университета. 2018. № 1, с. 51–65.
8. Gabriel Morgan Asaftei, Sudeep Doshi, Aditya Sanghvi, and John Means. Getting Ahead of the Market // Urban land magazine. URL: <https://urbanland.uli.org/economy-markets-trends/getting-ahead-of-the-market/> (дата обращения: 13.11.19)
9. Rodolfo De Guzman. We built a machine learning model that can estimate property values in Singapore. URL: <https://www.techinasia.com/machine-learning-estimate-singapore-property-value> (дата обращения: 13.11.19)

© Лепихина Ольга Юрьевна (OlgaLepikhina1984@gmail.com),

Балтыжакова Татьяна Игоревна (tatiyana.baltyzhakova@gmail.com), Рагузин Иван Игоревич (Raguzin.iv@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»