

УДК 338.27:004.032.26(470.55) + 332.85(470.55)

## **ПРОГНОЗИРОВАНИЕ ЦЕН НА РЫНКЕ НОВОСТРОЕК г. ЧЕЛЯБИНСКА С ПОМОЩЬЮ МОДЕЛЕЙ НЕЙРОННЫХ СЕТЕЙ**

*И.В. Малев, М.А. Лузина, А.С. Головина*

Рассматривается возможность использования моделей нейронной сети для прогнозирования цен на рынке недвижимости г. Челябинска. Раскрывается порядок расчета и состав обрабатываемых данных в прогностической нейронной сети.

Ключевые слова: прогнозирование цен, рынок жилья г. Челябинска, программный комплекс *STATISTICA*, нейронные сети.

В настоящее время, когда Россия переживает последствия экономических санкций, особую важность приобрела проблема прогнозирования основных параметров рынка. Это обусловлено тем, что в периоды системных кризисов, охватывающих все рынки, значимость и эффективность принимаемых решений определяется их будущими результатами, которые невозможно оценить, не предвидев, каким это будущее станет.

Безусловно, любое будущее событие является весьма неопределенным. Поэтому не существует способов, позволяющих точно «распознать» его последствия. Вместе с тем известно множество приемов, методов и соответствующих средств, позволяющих обнаружить тенденции, логику развития тех или иных процессов и на основе сопоставления с прошлым опытом предсказать характеристики наиболее вероятного развития событий.

Приемлемую эффективность прогнозирования можно обеспечить, изучая процессы изменения цен на рынке, выявляя устойчивые тенденции, которые проявлялись в прошлом и относительно которых имеются основания предполагать, что они сохранятся и в будущем. В качестве основного инструментария для этого обычно используются модели и статистические методы анализа временных рядов, развитые в рамках традиционной методологии случайных процессов.

Для построения прогнозов нами был применен инструментарий нейросетевого прогнозирования. Его главная особенность – нейронная сеть способна самостоятельно обучаться и действовать на основании предыдущего опыта, с каждым разом делая все меньше ошибок.

Нейросеть имитирует не только деятельность, но и структуру нервной системы человека. Такая сеть состоит из большого числа отдельных вычислительных элементов («нейронов»). В большинстве случаев каждый «нейрон» относится к определённому слою сети. Входные данные последовательно проходят обработку на всех слоях сети. Параметры каждого «нейрона» могут изменяться в зависимости от результатов, полученных на предыдущих наборах входных данных, изменяя таким образом и порядок работы всей системы.

В качестве объекта прогнозирования нами были выбраны цены на новостройки г. Челябинска. Было взято 60 значений в течение пяти лет, с 01.01.2013г. по 01.12.2017г. с интервалом в 1 месяц. Данные о ценах за кв.м собирались на сайте [gosrealt.ru](http://gosrealt.ru) в разделе «Цены на недвижимость в Челябинске» [2]. Временной ряд цен представлен на рис. 1.



Рис. 1. Динамика временного ряда цен на недвижимость

График имеет явно убывающий тренд (впоследствии это окажется логистическая функция) и явную периодичность. Чтобы точнее определить периодичность, можно воспользоваться аналитическим методом – спектральным анализом Фурье и провести расчеты в программном комплексе *STATISTICA* [3]. На рис. 2 представлены результаты анализа. На периодограмме наибольший пик соответствует периоду колебаний цен на недвижимость в течение года. В данном случае он составляет 10 месяцев.

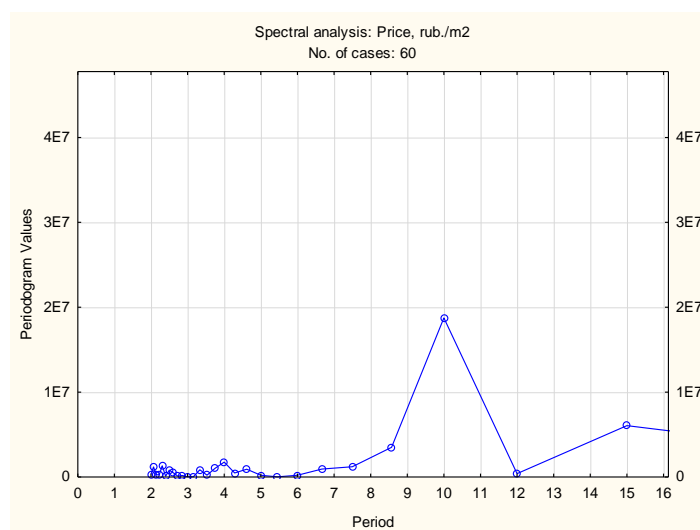


Рис. 2. Спектральный анализ Фурье

Дальнейшая задача прогнозирования будет проходить в три этапа. На первом – мы построим нейросетевую модель, которая будет прогнозировать этот временной ряд. На втором – оценим качество этой модели и на третьем – построим прогноз на более длительный период.

В задаче используется одна переменная – она является и входной, и выходной. Зададим условие – мы не будем использовать последние 10 наблюдений (т.е. данные за последние 10 мес.) с целью оценки аналитической модели и её поведения на независимых данных (кросс-проверка). Всего у нас 60 наблюдений, поэтому будем использовать первые 50.

Стратегию обучения выберем автоматизированную. Зададим контрольную подвыборку в 30 %, которая будет использоваться для остановки процесса обучения нейронной сети. Когда ошибка на контрольной выборке начинает возрастать, это послужит признаком того, что наша модель переобучается. При этом размер окна выбирается не меньше, чем период ряда (в нашем случае 10 месяцев).

Размер окна – это количество наблюдений, которые являются входными в нейронную сеть. Размер окна может быть и больше этого периода, но в этом случае модель только усложняется и остается меньше наблюдений для обучения.

Укажем минимальные настройки для последующего обучения: для персептронов интервал для скрытых нейронов в диапазоне от 2 до 8. Количество сетей для обучения – 20.

В качестве функции ошибки используется сумма квадратов, т. к. временной ряд по сути – это задача регрессии, а в задаче регрессии обычно используется сумма квадратов.

Задаем основные функции активации для скрытых нейронов: тождественная, логистическая, гиперболическая. На выходе, как правило, используется одна (тождественная функция), как и в задачах регрессии.

После запуска режима обучения переходим к анализу результатов. Первое, на что мы обращаем внимание, это на описание сетей. В табл. в столбце «Net. name» указана архитектура сети: например, строка один (MLP 10-8-1) – многослойный персептрон, 10 входов, 8 скрытых нейронов и один выход. Производительность (0,978244) характеризуется величиной корреляции между исходным рядом и предсказанным на каждой подвыборке. Чем выше это значение, чем ближе к 1, тем лучше. При этом нужно обращать внимание, чтобы корреляция на обучаемых контрольных подвыборках была приблизительно на одном уровне, чтобы не было резких отличий между ними. Тогда можно говорить о том, что модель равномерно работает как на обучении, так и на контроле.

Выберем наилучшую суть. Для этого смотрим на производительность сетей (столбец «Training perf.») – лучшая будет иметь максимальную производительность (строка 1).

Таблица

Параметры подобранных сетей

Index	Net. name	Training perf.	Test perf.	Validation perf.	Training error	Test error
1	MLP 10-8-1	0,978244	0,814576		64845,65	427202,6
2	MLP 10-5-1	0,976976	0,813381		68611,60	471654,8
3	MLP 10-4-1	0,977331	0,827599		67503,72	388506,8
4	MLP 10-8-1	0,978243	0,811812		64770,34	438982,8
5	MLP 10-8-1	0,974447	0,830757		76156,21	374220,6

Далее оцениваем распределение остатков – лучшая сеть будет иметь нормальное распределение остатков (рис. 3).

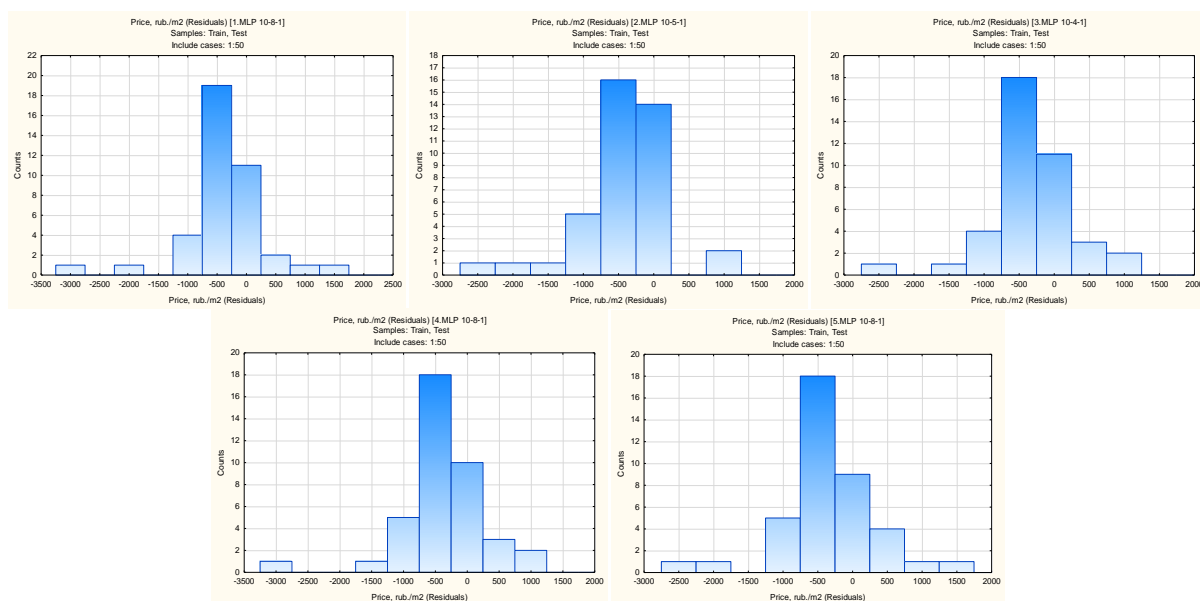


Рис. 3. Архитектура сетей: MLP 10-8-1, MLP 10-5-1, MLP 10-4-1, MLP 10-8-1, MLP 10-8-1

На рис. 3 построены гистограммы распределения остатков для обучающей и контрольной подвыборки. Показателем хорошего качества построенных моделей является приблизительно нормальное распределение остатков. Мы можем отсеять те модели, у которых распределение сильно отличается от нормального. По данным графикам можно сделать вывод, что все модели показывают вполне адекватное распределение.

Отбор сетей можно сделать в том числе и на основе данных о производительности, т.е. убирать те сети, которые имеют очень низкую производительность, тем самым отсеять наиболее неадекватные модели. В нашем случае по этому параметру нет явных аутсайдеров – все сети примерно одинаковые.

Также оценить качество подбора можно на основе диаграммы рассеивания целевых и выходных переменных (рис. 4).



Рис. 4. Диаграмма рассеивания целевых и выходных переменных

Видим, что не все точки ложатся на прямую. Наблюдаются явные выбросы значений. Это говорит об удовлетворительном качестве некоторых построенных нейронных сетей. Идеальное качество было бы, если все точки на графике ложились на прямую.

И, самое главное, отбор сетей мы можем сделать на основе визуального анализа – оценке того, на сколько хорошо наши построенные сети прогнозируют исходный ряд. После этого уже окончательно можно выбрать наилучшую сеть и сделать прогноз цен на будущее.

Посмотрим насколько хорошо визуально наши построенные модели прогнозируют исходный ряд (рис. 5).

На графике видим исходный ряд, промаркированный синим цветом, и предсказанные ряды для каждой модели. В этом случае каждое значение предсказанного ряда строится по предыдущим значениям исходного ряда. Т.е. в этом случае мы не можем сделать прогноз и продлить ряд дальше вправо. Для этой цели используется функция программы *STATISTICA* «проекция временного ряда». В этом случае каждое значение предсказанного временного ряда будет строиться по предыдущим значениям этого же самого временного ряда.

Чтобы визуально посмотреть на то, как проекции ложатся на весь ряд, построим график, начиная с первого наблюдения и заканчивая шестидесятым.

На рис. 6 мы видим тот же самый ряд плюс дополнительную проекцию временного ряда на 10 месяцев вперед. Видим, что есть модели, которые не очень хорошо прогнозируют.

Идет равномерный спад в модели номер пять (5. MLP 10-8-1). Есть модели, которые не улавливают зависимость. Это соответствует модели под

номером два и три (2. MLP 10-5-1, 3. MLP 10-4-1). Эти модели удалим. Наиболее оптимальной в данном случае будет модель номер один (MLP 10-8-1) и четыре (MLP 10-8-1). Оставим первую модель в качестве наиболее удавшейся. Далее будем смотреть и анализировать ее качество, удалив все остальные.

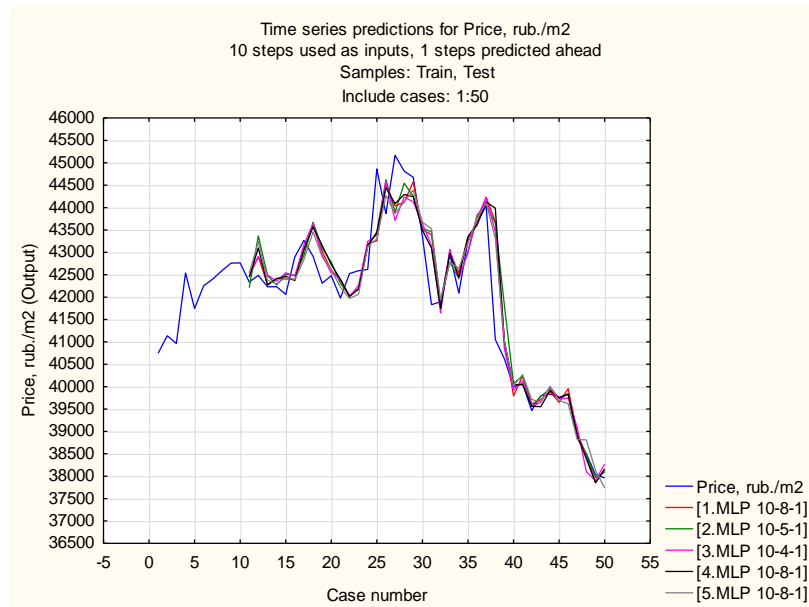


Рис. 5. Линейные графики построенных нейронных сетей

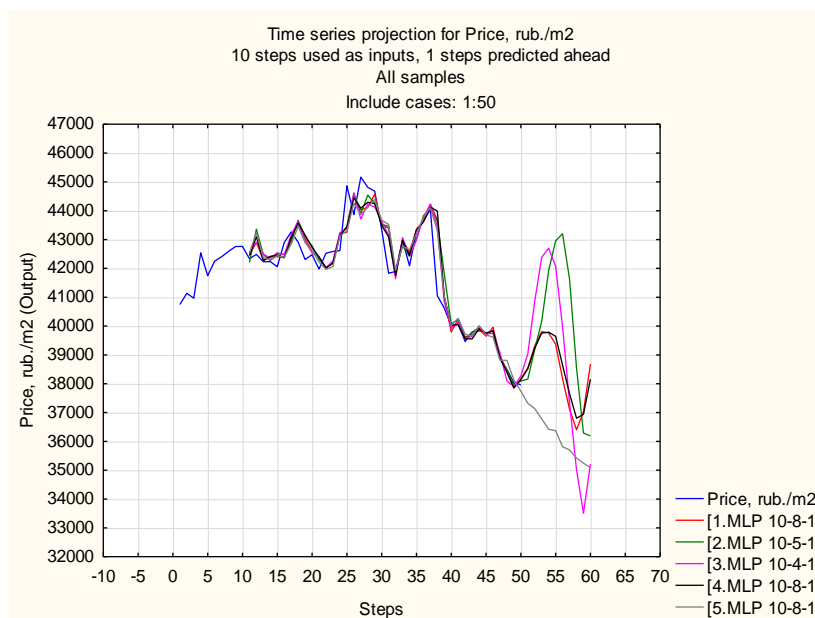


Рис. 6. Прогнозирование на основе построенных моделей

На последнем этапе сравним прогноз построенной сети с тестовой выборкой, которую мы оставили с самого начала (10 последних значений временного ряда).

На рис. 7 представлен график прогнозных значений и исходных данных с кросс-проверочной выборкой. Мы видим некоторое визуальное расхождение графиков, однако рассчитанная относительная ошибка исходных и прогнозных данных составляет всего лишь 0,014 (или 1,4 %), что указывает на очень хорошее качество прогноза, в частности по сравнению с методом экспоненциального сглаживания [1].

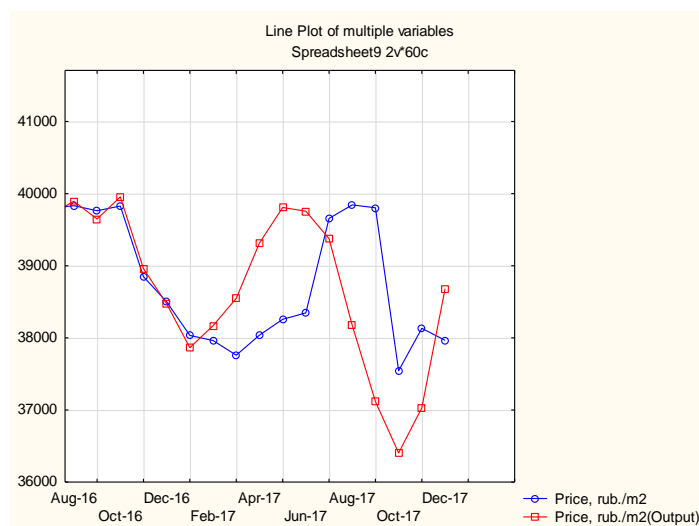


Рис. 7. Прогнозная часть графика временного ряда

Несмотря на небольшое количество испытаний, мы получили вполне адекватный результат. Помимо этого, его можно в дальнейшем улучшить, проведя настройки параметров моделей, – задать различное количество скрытых нейронов (небольшое–большое), попытаться использовать различные функции активации, и главное – протестировать модель на гораздо большем количестве сетей.

#### Библиографический список

1. Малев, И.В. Прогнозирование цен на рынке жилья г. Челябинска / И.В. Малев // Наука ЮУрГУ. Материалы 67-й научной конференции. Секция технических наук. – Челябинск: Издательский центр ЮУрГУ, 2015. – С. 211–217.
2. Недвижимость в России и за рубежом [Электронный ресурс]. – URL: <https://rosrealt.ru>.
3. Электронный учебник по статистике StatSoft: нейронные сети [Электронный ресурс]. – URL: <http://statsoft.ru/home/textbook/modules/stneunet.html/>.

[К содержанию](#)