

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Институт экономики и менеджмента

МАТЕРИАЛЫ
региональной научно-практической
конференции студентов и молодых ученых
«Экономика глазами молодых»,
посвященной 120-летию экономического
образования в Сибири

Томск, 20–21 апреля 2018 г.

Под общей редакцией профессора
Д.М. Хлопцова

Томск
Издательский Дом Томского государственного университета
2018

3. Зенков А.В. Методы оптимальных решений : учеб. пособие. М. : Юрайт, 2018. 201 с.
4. Колемаев В.А. Математическая экономика. М. : ЮНИТИ-ДАНА, 2005. 295 с.
5. Кофман А., Анри-Лабурдер А. Методы и модели исследования операций: целочисленное программирование / пер. с фр. Б.Т. Вавилова и др.; под ред. Н.П. Бусленко. М. : Мир, 1977. 432 с.
6. Исследование операций в экономике : учебник / отв. ред. Н.Ш. Кремер. М. : Юрайт, 2017. 438 с. (Бакалавр. Академический курс).
7. Мадера А.Г. Моделирование и принятие решений в менеджменте: руководство для будущих топ-менеджеров. Изд. стер. М. : Изд-во ЛКИ, 2015. 684 с.
8. Рюмкин В.И. Лекции по исследованию операций в экономике : [учеб. пособие] / Том. гос. ун-т, Прокопьевский филиал ТГУ. Томск : Изд-во НТЛ, 2007. 290 с.
9. Таха Х.А. Введение в исследование операций : [пер. с англ.]. М. [и др.] : Вильямс, 2016. 912 с.
10. Экономико-математические методы и модели : учеб. пособие для студентов экономических специальностей вузов / Н.И. Холод, А.В. Кузнецов, Я.Н. Жихар ; под общ. ред. А.В. Кузнецова. 2-е изд. Минск : Белорус. гос. экон. ун-т, 2000. 411 с.

Применение нейронных сетей в кредитном скоринге

И.С. Дуля, студент, 2 курс, ИЭМ ТГУ

Научный руководитель: к.т.н., доцент А.Л. Богданов
E-mail: idulya7@gmail.com

Введение

Перед инвесторами и крупными финансовыми организациями встают задачи, решение которых не может быть получено путем применения известных зависимостей и законов. Одним из примеров таких задач является выявление надежных заёмщиков, которые с высокой долей вероятности выполняют условия договора. Банк, имея модель, способную предсказать надежность сделки, может значительно повысить отдачу от своих вложений и минимизировать риск финансовых потерь. Одним из инструментов, предназначенных для решения подобных задач, являются искусственные нейронные сети.

Постановка задачи. Объектом данного исследования является решение задачи кредитного скоринга с помощью аппарата нейронных сетей,

целью – построение модели, способной выявить заёмщиков, которые с высокой долей вероятности выполнят условия договора. Под *кредитным скорингом* (*credit scoring*) понимают оценку вероятности банкротства потенциального заемщика при рассмотрении возможности его кредитования [1]. В ходе решения данной задачи были произведены поиск, отбор, загрузка, очистка и подготовка исходных данных, построение и обучение нейронной сети, выполнена оценка качества полученного решения.

Данные, на основе которых была построена модель, были получены с платформы p2p-кредитования Lending Club [2] – места встречи кредиторов и заемщиков, являющихся физическими лицами. Такой выбор был обусловлен тем, что данные находятся в свободном доступе, в отличие от банковских данных.

Очистка и преобработка данных

В качестве исходных данных были взяты статистические данные компании Lending Club по ссудам, выданным за 2014 год. Статистическая совокупность содержала 235629 записей о выданных ссудах. Каждая запись содержала 135 признаков (134 входных и один целевой – оценка заемщика). С описанием признаков можно ознакомиться по ссылке https://github.com/dulyaivan/credit_scoring/description.pdf.

Перед построением модели, была проведена подготовка данных. Было удалено 29 признаков, которые не содержали ни одного значения, также были удалены признаки, которыми не может располагать инвестор на момент рассмотрения заявки. С описательной статистикой оставшихся признаков можно ознакомиться по ссылке https://github.com/dulyaivan/credit_scoring/blob/master/statistics.pdf. В описании для каждого признака указаны: тип переменной, количество пропусков, доля пропусков, количество уникальных элементов, доля уникальных элементов, среднее значение, медиана, стандартное отклонение, максимум и минимум. Признаки, в которых доля пропусков составляла более 70%, и атрибутивные признаки, количество уникальных значений которых превышает 300, также были удалены из набора данных, т.к. работа с ними приводит к существенному увеличению объема данных и усложнению процесса обучения нейронной сети. Также были удалены признак `sub_grade` (подкласс кредитоспособности), который является уточнением к признаку `grade`, и признак `loan_status` (статус ссуды), имеющий прямую связь с целевым признаком, а интересующий признак `grade` (класс кредитоспособности) был отделен от общей совокупности.

В результате были получены матрица входных переменных и вектор-столбец значений выходного признака.

Затем все значения признаков были переведены в числовой формат, атрибутивные признаки были переведены в булево пространство, причем отсутствие данных интерпретировалось как отдельное значение признака. Целевой признак `grade`, принимающий значения A, B, C, D, E, F и G, был представлен в числовом формате. Пропуски в количественных признаках были заполнены соответствующим средним значением. На заключительном этапе подготовки данных подверглись процедуре стандартизации.

Построение модели

Для решения задачи классификации заемщиков на классы по платежеспособности (A, B, C, D, E, F, G) был использован многослойный полносвязный перцептрон с двумя скрытыми слоями, обучаемый методом обратного распространения ошибки. Такой выбор обусловлен тем, что описанная задача является классификационной, а также полносвязный перцептрон и алгоритм обратного распространения ошибки относительно просты в использовании. В качестве активационных функций на выходном слое была использована логистическая функция `Softmax`. Функция `Softmax` представляет собой обобщение логистической функции для многомерного случая. На всех остальных слоях активационная функция была задана как гиперболический тангенс. Листинг описанного многослойного перцептрона представлен на репозитории GitHub: https://github.com/dulyaivan/credit_scoring/blob/master/mlp.py.

При построении сети необходимо определить параметры сети (скорость обучения нейронной сети, количество эпох обучения, размер батчи (`batch`), значение параметра регуляризации), выбрать количество скрытых слоев и количество нейронов в них, произвести настройку весов нейронов. Для рассматриваемой задачи решено было использовать сеть с двумя скрытыми слоями. Количество нейронов входного и выходного слоев определяется количеством входных признаков и количеством возможных значений целевого признака (классов) соответственно. Выбор числа нейронов в скрытых слоях является отдельной задачей. Для её решения в работе был использован метод перебора. Количество нейронов в первом и втором скрытых слоях варьировались в диапазоне от 85 до 115 с шагом 3. На рис. 1 представлена тепловая карта `cross-validation` точности каждой конфигурации сети, где строка – количество нейронов в первом слое,

столбец – во втором. Согласно данным карты были взяты 112 нейронов в первом скрытом слое и 85 нейронов – во втором.

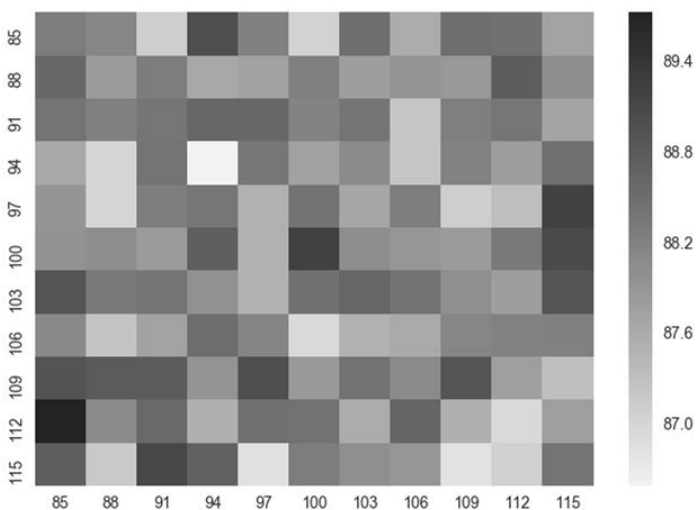


Рис. 1. Тепловая карта с cross-validation точностью различных конфигураций (OSR)

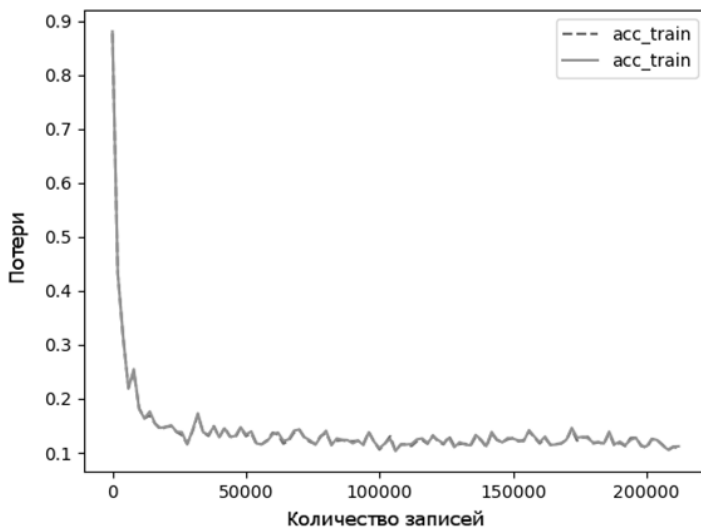


Рис. 2. Потери обучения на тестовом и тренировочном множестве

Для предотвращения переобучения нейронной сети в работе был применен следующий метод регуляризации: к вычисленной величине коррекции Δw_{ij} весового коэффициента прибавлялось текущее значение w_{ij} этого весового коэффициента, умноженное на значение параметра регуляризации, после чего полученное значение становилось новой величиной коррекции Δw_{ij} . Скорость обучения и параметр регуляризации были заданы константой равной 0,005. Это типичное значение для выбранной архитектуры сети, обеспечивающее достаточно высокую скорость обучения сети и хорошую точность настройки весов.

Оценка точности модели

Одним из важнейших этапов построения классификационных моделей является оценка ошибки, ибо именно на данном этапе подтверждается значимость модели и ее эффективность. Для выявления возможного переобучения был проведен анализ динамики потерь на тестовом и обучающем множествах. По результатам анализа (рис. 2) был сделан вывод, что эффект переобучения сети не возникает.

Таблица 1

Матрица классификации

		Предсказанный класс						
		A	B	C	D	E	F	G
Фактический класс	A	3028	0	0	576	0	0	0
	B	1	6059	204	318	5	2	0
	C	1	340	3773	1	219	5	0
	D	72	550	0	5661	0	0	0
	E	1	1	173	0	1794	26	0
	F	0	0	0	0	104	502	0
	G	1	0	0	0	6	140	0

В табл. 1 представлена матрица классификации, полученная в результате использования построенной сети на тестовом множестве. Как видно из таблицы, на тестовом множестве модель смогла определить 20817 из 23563 элементов, то есть точность прогноза составила 88,35%. Большая часть ошибочных классификаций сосредоточена вдоль главной диагонали. То есть, если модель ошибается, то в основном она относит элемент одного класса к соседнему классу, что при многоклассовой классификации не столь критично. Низкие результаты для класса G обусловлены относительно малым количеством элементов выборки. С практической

точки зрения наибольший интерес представляют элементы, расположенные ниже главной диагонали. Точность построенного классификатора, вычисленная методом 10 блочной кросс-валидации равна 88,23%.

Заключение

Применение нейронных сетей банками не ограничивается лишь скорингом заявителя, но также используются и для выявления мошенничества, работе с должниками и в маркетинге. Данное свойство обусловлено высокой универсальностью отдельных архитектур нейронных сетей, в том числе многослойного персептрона.

В работе показано применение нейронной сети в решении задачи кредитного скоринга. Как показывают проведенные тесты, многоклассовая классификация заемщиков по рейтингу кредитоспособности может быть эффективно решена простым многослойным персептроном.

Нейронные сети выявляют сложные, нетривиальные функциональные связи между входными и выходными признаками, что позволяет повысить эффективность принимаемых решений, сократить время рассмотрения заявки, снизить влияние человеческого фактора. Автоматизация такой рутинной процедуры как скоринг значительно позволяет направить освободившиеся трудовые и финансовые ресурсы на решение иных задач.

Литература

1. Глинкина Е.В. Кредитный скоринг как инструмент повышения эффективной оценки кредитоспособности // Банковское дело. 2011. № 16.
2. LendingClub : официальный сайт. San Francisco, 2006–2018. URL: <https://www.lendingclub.com> (access date: 10.10.2017).
3. Алешин В.А., Рудаева О.О. Кредитный скоринг как инструмент повышения качества банковского риск-менеджмента в современных условиях // Terra Economicus. 2012. Т. 10, № 2 (3). С. 27–30.
4. Воронцов К.В. Математические методы обучения по прецедентам : курс лекций. М., 2017. 141 с.
5. Дьяков О.А. Особенности применения методов Data Mining в скоринговых решениях для коммерческих банков // Научные записки молодых исследователей. 2017. № 3.
6. Лёзина И.В., Краснов А.Е. Исследование идентифицирующих свойств нечеткого многослойного персептрона // Известия Самарского научного центра Российской академии наук. 2014. № 4.
7. Никаненкова В.В. Кредитный скоринг как инструмент оценки кредитоспособности заемщиков // Вестник Адыгейского государственного университета. Экономика. 2012. № 5.

8. Официальная статистика. Финансы // Федеральная служба государственной статистики (Росстат). М., 1999–2018. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/finance (дата обращения: 01.11.2017).

9. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. СПб. : Питер, 2013. 704 с.

10. Рашка С. Python и машинное обучение / пер. с англ. А.В. Логунова. М. : ДМК Пресс, 2017. 418 с.

Анализ степени концентрации российских отраслевых рынков

А.С. Ерохин, студент, 2 курс, ИЭМ ТГУ

Научный руководитель: к.ф.-м.н., доцент Е.В. Чаусова
E-mail:litno@mail.ru

В реальной экономике практически на каждом рынке существуют фирмы, обладающие рыночной властью – способностью своим объемом выпуска влиять на рыночную цену в собственных интересах. Такие рынки называют несовершенными конкурентными. Выделяют три основные структуры рынка несовершенной конкуренции [3, 4, 8]: монополия (рыночная структура с единственным продавцом товара, не имеющего близких заменителей, для других продавцов вход на данный рынок закрыт), олигополия (рыночная структура с небольшим числом продавцов, конкурирующих между собой, появление новых продавцов затруднено или невозможно), монополистическая конкуренция (много продавцов, но с дифференцированным продуктом).

Одной из основных характеристик структуры рынка является рыночная (экономическая) концентрация. Концентрация рынка определяется наличием или отсутствием лидирующих продавцов на рынке, контролирующих значительную его долю, диктующих правила, обладающих рыночной властью. Высокая концентрация создает благоприятные условия для возникновения монополии, угрожает нормальному функционированию рыночного механизма, ограничивает конкуренцию. Поэтому анализ концентрации отраслевых рынков является одной из наиболее важных и актуальных задач при проведении антимонопольной политики.

Для определения степени концентрации рынка в мировой практике применяют такие количественные показатели, как число фирм на рынке, пороговая доля рынка, индекс концентрации, индекс Херфиндаля-