

УДК 005.51

Количественное моделирование процесса подбора персонала

ЗИНЧЕНКО АЛЕКСЕЙ АЛЕКСЕЕВИЧ, аспирант кафедры «Прикладная математика» Финансового университета
E-mail: a_zinchenko@list.ru

Аннотация. В статье представлен количественный подход к моделированию процесса подбора и оценки персонала. В условиях, когда необходимо собрать большой штат сотрудников в сжатые сроки, применение математического аппарата способно существенно ускорить процесс подбора. Самым сложным и трудозатратным, а также неоднозначным по результату является этап, на котором непосредственно принимается решение о найме сотрудника. Проведение собеседований и просмотр резюме сотрудниками кадровых служб всегда связаны с человеческой субъективностью. Кроме того, не так-то просто найти кандидата, обладающего достаточным профессиональным опытом, удержать его, замотивировать и т.д. Современные труды, посвященные подбору персонала, описывают процесс принятия решения о найме с позиции оценки менеджером профессиональных и личностных качеств будущего работника. Автором данной статьи предложен метод оценки вероятности прохождения испытательного срока на основе информации, предоставленной в резюме. Для решения данной задачи использовались модели бинарного выбора и искусственная нейронная сеть, был проведен сравнительный анализ данных моделей. Преимущества каждого из подходов могут проявляться в зависимости от ситуации. Так, например, коэффициенты бинарной регрессии позволяют узнать, как именно отдельные факторы влияют на конечную вероятность прохождения испытательного срока. Такая информация может быть полезна в ситуации, когда принимается решение о найме сотрудника на управленческую должность и необходим глубокий анализ. С другой стороны, бинарная регрессия не лишена недостатков: в некоторых случаях модель бинарного выбора может давать неуверенную оценку кандидатам.

Данная проблема отсутствует при применении искусственной нейронной сети (ИНС) к задаче оценки кандидата на должность: ИНС осуществляет однозначную классификацию. Однако коэффициенты синаптических связей в искусственной нейронной сети не поддаются интерпретации, что делает невозможным определение влияния отдельных факторов на результат, предоставляемый моделью.

Ключевые слова: подбор персонала, бинарная регрессия, искусственная нейронная сеть, управление персоналом.

Quantitative modeling of the recruitment process

ZINCHENKO ALEKSEY A., post-graduate of „Applied mathematics” Chair, Finance University
E-mail: a_zinchenko@list.ru

Abstract. The article presents a quantitative approach to modeling the process of recruitment. In circumstances where it is necessary to collect a large staff in a short time, the use of mathematical apparatus can significantly accelerate this process. The most complex and labor-intensive, but also ambiguous on the result, is a step in which the direct decision on hiring is made. Interviewing and resume viewing are always associated with human subjectivity. In addition, it is not so easy to find the recruiter with sufficient professional experience, hold him, motivate, etc.

Modern works dedicated to recruiting, describe the process of making a decision on hiring from the point of view of a manager, who assesses professional and personal qualities of the future employee. The author proposed a method for estimating the probability of passing the probation, based on the information provided in the CV. To solve this problem, author used a binary choice model and artificial neural network and made a comparative analysis of these models. Advantages of each approach may occur depending on the situation. For example, coefficients in binary regression let you know exactly how certain factors affect the probability of passing the probation period. Such information can be useful

in a situation where the decision on hiring employees in managerial positions is made and deep analysis is needed. On the other hand, binary regression has some drawbacks: in some cases the binary choice model can produce uncertain evaluation of candidates.

This problem does not occur when using the ANN to the task of evaluating candidates for the position: ANN provides an unambiguous classification. However, synaptic weights in the artificial neural network cannot be interpreted, making it impossible to determine the effect of individual factors on the results provided by the model.

Keywords: recruitment, binary regression, artificial neural network, personnel management.

В конце 1990-х годов руководители крупнейших компаний пришли к выводу о том, что спрос на высококвалифицированные кадры явно стал превышать предложение. Подбор персонала стали выделять как отдельный бизнес-процесс, а следовательно, и как дисциплину для изучения. В абсолютном большинстве трудов, посвященных процессу подбора персонала, этап, связанный непосредственно с оценкой кандидата и принятием решения о найме, освящен только с позиции работы менеджера. В настоящее время существуют программные продукты, призванные облегчить работу кадровых менеджеров, действующие по принципу контекстного поиска по базе резюме.

В работе сотрудников кадрового агентства и кадровых отделов самих организаций используются одни и те же принципы. Оценка потенциального работника главным образом проводится во время собеседования сотрудниками отдела персонала, однако из резюме также можно почерпнуть достаточное количество информации о кандидате. В то же время любое кадровое агентство располагает базой резюме, что позволяет осуществить количественный подход к процессу моделирования подбора персонала.

Очевидно, что оценка кандидата представляет собой задачу бинарного выбора, или бинарной классификации. В качестве статистики для построения моделей будем использовать резюме работников, рекомендованных кадровым агентством своим клиентам. Работники делятся на две группы: те, кто прошел испытательный срок у организации — клиента агентства, и те, кто его не прошел.

Первый подход, который предложен в статье: использование бинарной регрессии для решения поставленной задачи. Преимуществом таких моделей является возможность выявить влияние отдельных факторов на вероятность прохождения испытательного срока.

Бинарная регрессия представляет зависимость эндогенной переменной, принимающей всего

два значения — 0 и 1, от набора факторов. Обычная линейная регрессия для таких переменных не применима, так как она допускает и отрицательные значения, и значения выше 1. Поэтому обычно используются некоторые интегральные функции распределения. Чаще всего используются функции нормального распределения (пробит), логистического распределения (логит) и распределения Гомперца (гомпит) [1]. От выбора функции распределения напрямую зависит соответствие прогнозов, полученных с помощью модели, реальным данным.

Предполагая, что зависимая переменная Y , которая представляет собой возможность или невозможность взять на работу кандидата (или в случае с рекрутинговым агентством — рекомендовать его клиенту), принимает только два значения: $\{0; 1\}$, вероятность того, что она примет соответствующее значение можно выразить как функцию некоторых факторов:

$$\begin{aligned} \text{Prob}(Y = 1 | x) &= F(\bar{x}^T \beta); \\ \text{Prob}(Y = 0 | x) &= 1 - F(\bar{x}^T \beta), \end{aligned}$$

где \bar{x} — вектор, содержащий значения факторов, а набор параметров β отражает влияние изменения каждого фактора на конечную вероятность. Оценка параметров β осуществляется методом максимального правдоподобия.

Для оценки статистической значимости группы регрессоров модели используется так называемая статистика отношения правдоподобия [2], рассчитываемая по формуле

$$LR = 2(L_{full} - L_{null}),$$

где L_{full} и L_{null} — значения логарифмической функции правдоподобия для оцененной модели, включающей все регрессоры, и для модели, состоящей

из одной константы. Значение LR должно превышать критическое значение Хи-квадрат с количеством степеней свободы, равным количеству факторов в модели.

Другим показателем качества модели является псевдокоэффициент детерминации. Существует несколько методик его расчета, в данной статье будет рассмотрен коэффициент детерминации МакФаддена [3], который рассчитывается по формуле

$$R_{pseudo}^2 = 1 - \frac{L_{full}}{L_{null}}.$$

С помощью псевдокоэффициента детерминации также можно оценивать необходимость включения каждого регрессора в модель. Для этого регрессоры последовательно включаются в модель и отслеживается изменение R_{pseudo}^2 . Коэффициент детерминации МакФаддена не имеет абсолютной интерпретации, как классический R^2 . С его помощью мы можем лишь сравнивать различные спецификации модели.

С помощью псевдокоэффициента детерминации также можно оценивать необходимость включения каждого регрессора в модель. Для этого регрессоры последовательно включаются в модель и отслеживается изменение R_{pseudo}^2 . Коэффициент детерминации МакФаддена не имеет абсолютной интерпретации как классический R^2 . С его помощью можно лишь сравнивать различные спецификации модели.

В статье представлены результаты оценки модели, где в качестве функции $F(\bar{x}^T \beta)$ была выбрана функция логистического распределения. Данная модель показала лучший результат по сравнению с пробит- и гомпит-моделями. В качестве экзогенных переменных были выбраны данные из резюме, указанные выше.

Логит-модель (см. таблицу):

$$Prob(Y = 1|x) = \frac{e^{\bar{x}}}{1 + e^{\bar{x}}},$$

где \bar{x} — вектор значений факторов.

$$R_{pseudo}^2 = 0,43;$$

$LR = 24,046 > \chi^2(11) = 19,68$ при уровне значимости 95%.

Значение LR статистики показывает, что все параметры являются значимыми. Поскольку

функция распределения всегда монотонно возрастающая, по знаку параметра можно судить о том, какой вклад (отрицательный или положительный) он вносит в конечную вероятность. В нашем случае, поскольку значения всех факторов являются неотрицательными, можно утверждать, что знак коэффициента перед ним можно интерпретировать буквально. Так, знание иностранных языков оказывает положительное влияние на конечную вероятность, что вполне очевидно. Столь же понятно и отрицательное влияние количества организаций, в которых работал кандидат. Уже на стадии обработки данных для статистики автор заметил тот факт, что люди, которые уволились в течение испытательного срока, могли сменить множество организаций за два-три года. Прошедшие испытательный срок, напротив, предпочитали не менять столь часто место работы.

Гистограмма предсказанных вероятностей (рис. 1) указывает на то, что большинство из них близки либо к 0, либо к 1. Это означает, что модель редко дает «неуверенную» оценку кандидатам.

Выбор порогового значения вероятности для принятия окончательного решения зависит уже от руководителя кадровой службы. Очевидно, что чем выше пороговое значение, тем выше степень уверенности в будущем кандидате.

Коэффициенты при переменных (логит-модель)

β_0	-16,867
Пол кандидата	0,956
Возраст	-0,094
Наличие высшего образования	9,472
Профиль	-1,8603
Стаж работника (лет)	0,436
Количество организаций, в которых работал кандидат	-0,588
Перечисленные обязанности (количество)	-0,009
Знание английского языка	9,859
Другие иностранные языки, указанные в резюме	0,937
Уровень знания компьютера	0,524
Уровень запрашиваемой заработной платы	-0,00001

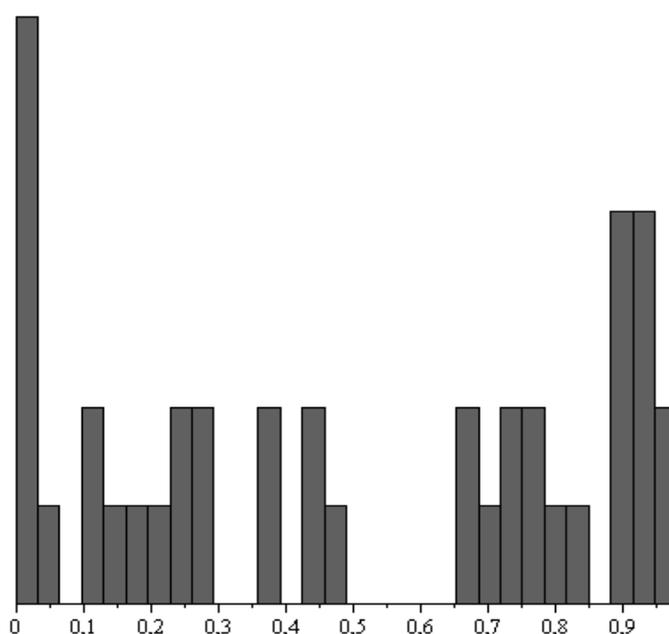


Рис. 1. Гистограмма предсказанных вероятностей, логит-модель

Далее рассмотрим модель, лишенную такого недостатка, как неуверенная оценка кандидатов.

Искусственная нейронная сеть представляет собой математическую модель, построенную по принципу организации и функционирования биологических нейронных сетей. Это система соединенных и взаимодействующих между собой простых моделей — искусственных нейронов. Понятие «нейронная сеть» было формализовано Маккалоком и Питтсом в 1943 г. и с тех пор нейросетевые модели прошли долгий эволюционный путь. В настоящее время ИНС широко применяются для решения задач классификации, распознавания образов, прогнозирования [4].

Преимущество использования искусственных нейронных сетей для задач классификации обусловлено их исключительной способностью моделировать нелинейные зависимости с большим количеством переменных [5]. К тому же в случае с оценкой работника, когда количество классов равно двум (0 — работника не следует рекомендовать клиенту, 1 — работника можно рекомендовать клиенту), использование ИНС не вызовет сложностей, связанных с представлением данных на выходе сети.

Для оценки будущего работника будем использовать многослойную нейронную сеть, реализованную в программном пакете *Deductor*.

Оцениваемыми параметрами в ИНС являются синаптические веса, а процесс их оценки называется обучением нейронной сети. Настройка нейронной сети включает в себя выбор следующих параметров:

- архитектуры нейронной сети (выбор количества скрытых слоев и количества нейронов в них);
- активационной функции нейронов;
- алгоритма и параметров обучения.

Подбор всех параметров ИНС осуществляется экспериментально, что является основным недостатком их использования, наряду с отсутствием возможности интерпретировать результаты классификации, т.е. установить, какие именно факторы имеют решающее значение в определении класса, к которому принадлежит работник.

Итак, из резюме работников, предоставленных кадровым агентством, автор выделил следующие факторы, используемые для формализации:

- 1) пол кандидата;
- 2) возраст;
- 3) наличие высшего образования;
- 4) профиль;
- 5) стаж работника (лет);
- 6) количество организаций, в которых работал кандидат;
- 7) перечисленные обязанности (количество);
- 8) знание английского языка;

9) другие иностранные языки, указанные в резюме;

10) уровень знания компьютера (3-балльная шкала: 0 — не указан; 1 — знание *MS Office*; 2 — знание специализированных пакетов анализа (*STATISTICA*, *SAP* и т.д.); 3 — навыки программирования);

11) уровень запрашиваемой заработной платы (0 — если не указан).

Далее экспериментальным путем было выявлено, что наилучшим образом классификацию осуществляет ИНС со следующей конфигурацией:

- количество скрытых слоев — 2;
- количество нейронов в первом скрытом слое — 2;
- количество нейронов во втором скрытом слое — 4;
- активационная функция нейронов — гипертангенс;
- алгоритм обучения — алгоритм эластичного распространения ошибки (*resilent propagation*).

Нейронная сеть данной конфигурации осуществляет верную классификацию для всех работников и не дает неуверенных оценок (рис. 2), как это бывает при использовании моделей бинарного выбора (пробит, логит, гомпит).

Одним из недостатков ИНС является то, что они не показывают, как именно влияют отдельные факторы на классификацию. Однако это также можно установить экспериментально, с помощью анализа «что, если», реализованного в программном пакете *Deductor*. При изменении значений факторов можно заметить, какие из них оказывают большее влияние на классификацию кандидатов (рис. 3). В частности, можно утверждать, что в нейронной сети, описанной в статье, такой фактор как «запрашиваемая заработная плата», указываемый в резюме, не имеет влияния на то, к какому классу будет отнесен работник. Проверить данный факт можно, исключив соответствующий фактор из входного слоя сети, при этом качество классификации не меняется.

В случае, когда кадровое агентство осуществляет массовый подбор работников и стремится автоматизировать этот процесс, не принципиально знать, какое влияние конкретные факторы оказывают на классификацию. Если речь идет о подборе высококвалифицированных кадров (например, на управленческие должности), такая информация, безусловно, была бы полезна, так как при этом математическая модель может

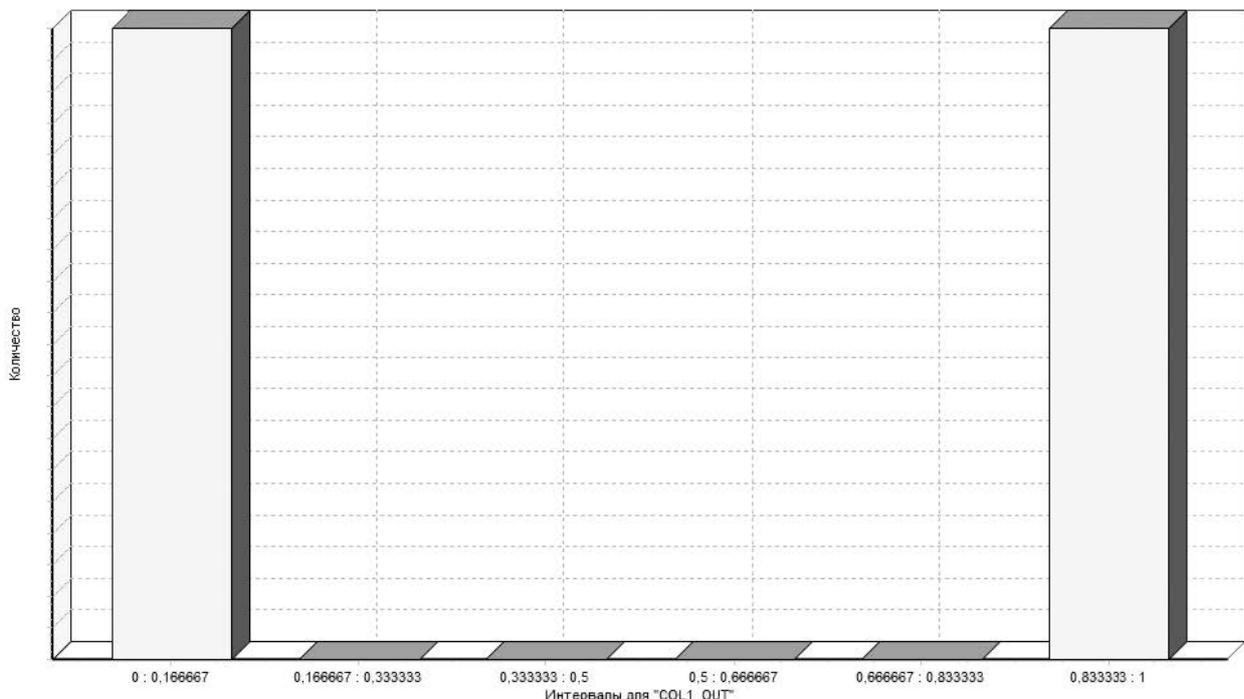


Рис. 2. Гистограмма выходных значений ИНС

Поле	Значение
Входные	
9.0 COL2	1
9.0 COL3	30
9.0 COL4	1
9.0 COL5	2
9.0 COL6	0,75
9.0 COL7	1
9.0 COL8	3
9.0 COL9	1
9.0 COL10	1
9.0 COL11	1
9.0 COL12	25440,04707
Выходные	
9.0 COL1	0

Рис. 3. Анализ «что, если»

выступать лишь как система поддержки принятия решений, в то время как конечное решение о том, можно ли рекомендовать работника клиенту, остается за HR-менеджером.

На сегодняшний день практически во всей литературе, посвященной подбору персонала, авторы делают акцент на применении HR-менеджерами методов, заимствованных преимущественно из психологии. Сотрудникам кадровых служб предлагается исследовать личность кандидата, опираясь на свой профессиональный опыт и интуицию [6].

Результаты, приведенные в исследовании, позволяют утверждать, что информация, которую люди указывают в резюме, несет в себе закономерности, поддающиеся статистическому анализу. Предложенные в исследовании методы позволят повысить объективность принятия решений по кадровым вопросам и помогут частично

автоматизировать процесс подбора и оценки персонала.

Литература

1. Мхитарян В. С., Архипова М. Ю., Сиروتин В. П. Эконометрика. Учебно-методический комплекс. М: центр ЕАОИ, 2008. 144 с.
2. William H. Greene — *Econometric Analysis*, 2003.
3. Hand D.J., Henley W.E. Statistical classification methods in consumer credit // *Journal of the Royal Statistical Society. Series A*. 1997.
4. Барский А. Б. Нейронные сети: распознавание, управление, принятие решений. М.: Финансы и статистика, 2004. 176 с.
5. Хайкин С. Нейронные сети: полный курс, 2-е изд. / Пер. с англ. М.: Вильямс, 2006. 1104 с.
6. Brian Tracy. *Hire and Keep the Best People*. Berrett-Koehler Publishers, 2001.

References

1. Mhitaryan V.S., Arhipova M.Y., Sirotin V.P. *Econometrics. Training package*. Moscow, EAOI Center, 2008, 144 p. (in Russian).
2. William H. Greene — *Econometric Analysis*, 2003 (in English).
3. Hand D.J., Henley W.E. Statistical classification methods in consumer credit // *Journal of the Royal Statistical Society. Series A*, 1997 (in English).
4. Barsky A. B. *Neural networks: cognition, management, decision making*. Moscow, Finances and Statistics, 2004, 176 p. (in Russian).
5. Hikin S. *Neural networks: a full course*, 2nd edition / The translation from English, Moscow, Williams, 2006, 1104 p. (in Russian).
6. Brian Tracy. *Hire and Keep the Best People*. Berrett-Koehler Publishers, 2001 (in English).