

УДОСКОНАЛЕННЯ МАТЕМАТИЧНОЇ МОДЕЛІ ОЦІНЮВАННЯ РЕЗЮМЕ КАНДИДАТА В ІТ-РЕКРУТИНГУ

Ткаченко А.І., студент гр. 6151м¹, Фаріонова Т.А. к.т.н., доцент²

НУК імені адмірала Макарова

Україна, Миколаїв

¹ include.free@gmail.com, ² tetyana.farionova@nuos.edu.ua

Анотація. Удосконалено математичну модель оцінювання резюме кандидата в ІТ-рекрутингу на основі моделей бінарної регресії із використанням пробіт та логіт моделей. На основі розрахованих даних отримано вагові коефіцієнти факторів із застосуванням різних моделей оцінки та здійснене їх порівняння.

Ключові слова: оцінювання резюме, ІТ-рекрутинг, пробіт-модель, логіт-модель.

Вступна частина. Підбір персоналу — це система цілеспрямованих дій щодо залучення до роботи кандидатів, які мають якості, необхідні для досягнення цілей, поставлених організацією. Пріоритетним завданням рекрутингової компанії є підвищення ефективності оцінки кандидатів. Але на практиці агентства покладаються лише на досвід і професіоналізм співробітників HR-відділу, що займаються безпосередньо пошуком і відбором працівників, при цьому будь-які об'єктивні обґрунтування, що використовують математичні моделі, які описують даний процес, просто відсутні, але ж це один із суттєвих резервів поліпшення роботи в даному напрямку. Необхідність створення теоретичного підходу до кількісної оцінки процесу підбору персоналу обумовлює актуальність і затребуваність дослідження. Як правило, кількісні математичні моделі оцінювання резюме в ІТ-рекрутингу не застосовуються, оцінка кандидата виконується особисто рекрутером, що не може виключити суб'єктивність з процесу прийняття рішення.

Ціллю роботи є підвищення ефективності оцінювання резюме кандидата в ІТ-рекрутингу шляхом вдосконалення математичної моделі на основі бінарної регресії із використанням пробіт та логіт підходу.

Основна частина

Бінарна регресія – це залежність ендогенної змінної, приймаючої всього два значення - 0 і 1, від набору факторів. Існують обмеження, пов'язані з використанням звичайної лінійної регресії, так як прогнозоване значення має потрапляти в відрізок [0; 1]. Дана проблема вирішується з використанням інтегральних функцій розподілу. Найчастіше використовуються функції нормального розподілу (пробіт) і логістичного розподілу (логіт). Від вибору функції розподілу безпосередньо залежить відповідність прогнозів, отриманих за допомогою моделі, реальним даним.

Припускаючи, що залежна змінна Y , яка представляє собою можливість або неможливість взяти на роботу кандидата (або, у випадку з рекрутинговим агентством - рекомендувати його клієнту), приймає тільки два значення: $\{0; 1\}$, ймовірність того, що вона прийме відповідне значення можна виразити як функції деяких факторів (1) і (2):

$$Prob(Y = 1|x) = F(x^T \beta), \quad (1)$$

$$Prob(Y = 0|x) = 1 - F(x^T \beta). \quad (2)$$

Набір параметрів β відображає вплив зміни кожного фактора на кінцеву ймовірність. Завдання полягає в тому, щоб підібрати адекватну функцію в правій частині рівняння.

Широку популярність придбала так звана пробіт-модель, що використовує функцію (3) стандартного нормального розподілу:

$$Prob(Y = y|x) = \int_{-\infty}^{x^T \beta} \phi(t) dt = \Phi(x^T \beta). \quad (3)$$

Так само популярна логіт-модель, що використовує функцію (4) логістичного розподілу:

$$Prob(Y = y|x) = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)}. \quad (4)$$

Оцінка параметрів β здійснюється методом максимального правдоподібності. Кожне спостереження є схемою Бернуллі, тому функція правдоподібності (5) постає у вигляді:

$$Prob(Y_1 = y_1, \dots, Y_n = y_n | X) = \prod_{y_i=0} [1 - F(x_i^T \beta)] \prod_{y_i=1} F(x_i^T \beta). \quad (5)$$

Оскільки моделі бінарного вибору є нелінійними моделями, то скористатися звичайним коефіцієнтом детермінації ми не можемо. Альтернативним підходом до побудови аналога коефіцієнта R^2 є порівняння максимумів функції правдоподібності для довгої і короткої моделі.

Позначимо максимум функції правдоподібності для довгої моделі і короткої моделі відповідно L_{full} і L_{null} . При цьому ми справедливо можемо використовувати логарифми цих значень, так як $L_{null} \leq L_{full} \leq 1$, отже $LnL_{null} \leq LnL_{full} \leq 1$.

В рамках цього підходу, відомі наступні показники якості для моделей бінарного вибору: псевдо- R^2 (6) і коефіцієнт детермінації Макфаддена (7):

$$pseudoR^2 = 1 - \frac{1}{1 + \frac{2(LnL_{null} \leq LnL_{full})}{n}} \quad (6)$$

$$mcFaddenR^2 = 1 - \frac{LnL_{full}}{LnL_{null}}. \quad (7)$$

Коефіцієнт Макфаддена часто називають індексом відносної правдоподібності (Likelihood Ratio Index, LRI).

Обидва цих показники змінюються в межах від 0 до 1. Якщо для довгої моделі $\beta_2 = \dots = \beta_m = 0$, то $L_{null} = L_{full}$, і обидва коефіцієнта детермінації дорівнюють нулю. Коефіцієнт Макфаддена буде дорівнює одиниці, якщо, $LnL_{full} = 0$, тобто $L_{full} = 1$. У цьому випадку модель дає точний прогноз, таким чином, що $\hat{y}_i = y_i$ для всіх $i = 1, \dots, n$. При цьому, для пробіт та логіт моделей виявляється неможливим довести до кінця ітераційну процедуру оцінювання вектора параметрів β , оскільки абсолютна величина $x_i^T \hat{\beta}$ різко зростає. Відбувається це тому, що у моделей бінарного вибору при кінцевих значеннях $x_i^T \hat{\beta}$ виконуються суворі нерівності $0 < F(x_i^T \hat{\beta}) < 1$, отже, функція правдоподібності не може досягти одиниці.

Перевірка значущості групи регресорів моделей бінарного вибору здійснюється за допомогою критерію відношення правдоподібності.

Визначимо повний набір факторів моделі: стать кандидата; вік; наявність вищої освіти; профіль; стаж працівника (років); кількість організацій, в яких працював кандидат; перераховані обов'язки (кількість); знання англійської мови; інші іноземні мови, зазначені в резюме; рівень знання комп'ютера (3-бальна шкала, 0 — не вказано, 1 — знання MS Office, 2 — знання спеціалізованих пакетів аналізу Statistica, SAP і т. Д., 3 — навички програмування); рівень запитуваної заробітної плати («0», якщо не вказано).

Дані для статистики були взяті з резюме працівників, рекомендованих великим кадровим агентством своїм клієнтам. Бінарна залежна змінна приймає значення «1» в тому випадку, якщо людина продовжила роботу в організації після закінчення випробувального терміну, «0» - в протилежному випадку.

Оцінені коефіцієнти при змінних для пробіт- та логіт-моделі наведено таблиці 1.

Таблиця 1 — Коефіцієнти при змінних для пробіт- та логіт-моделі

	Коефіцієнти при змінних	
	логіт-модель	пробіт-модель
β_0	-16,867	-7,193
Стать кандидата (x1)	0,956	0,768
Вік (x2)	-0,094	-0,0502
Наявність вищої освіти (x3)	9,472	3,870
Стаж роботи (років) (x4)	0,436	0,296
Профіль (x5)	-1,8603	-1,051
Кількість організацій, в яких працював кандидат (x6)	-0,588	-0,296
Знання англійської мови (x7)	9,859	4,140
Інші іноземні мови, вказані в резюме (x8)	0,937	0,687
Рівень знання комп'ютера (x9)	0,524	0,360
Рівень запитуваної заробітної плати (x10)	-0,00001	-0,0000062
Перераховані обов'язки (кількість) (x11)	-0,009	-0,001

В результаті розрахунків були отримані показники якості обох моделей:

- для пробіт-моделі: $pseudoR^2 = 0.42$ $LR = 23,33 > \chi^2(11) = 19,68$; при рівні значущості 95%

- для логіт-моделі: $pseudoR^2 = 0.43$ $LR = 24,046 > \chi^2(11) = 19,68$; при рівні значущості 95%

Значення LR статистики показує, що всі параметри є значущими. Оскільки функція розподілу завжди є монотонно зростаючою, то за знаком параметра можна судити про те, який внесок, негативний або позитивний, він вносить в кінцеву ймовірність. В нашому випадку, оскільки значення всіх факторів є невід'ємними, можна стверджувати, що знак коефіцієнта перед ним можна інтерпретувати буквально. Так, наприклад, знання іноземних мов робить позитивний вплив на кінцеву ймовірність, що цілком очевидно. Настільки ж зрозуміло і негативний вплив кількості організацій, в яких працював кандидат. Так люди, які звільнилися протягом випробувального терміну, могли змінити безліч організацій за два-три роки. Ті, що пройшли випробувальний термін, навпаки, вважали за краще не змінювати настільки часто місце роботи. Варто зазначити, що абсолютна величина коефіцієнтів за факторами жодним чином не говорить про значимість цих факторів в регресії, а лише масштабує їх значення. Виключення будь-якого параметра з моделі призводить до зменшення величини $pseudoR^2$, що говорить про необхідність присутності кожного фактора в регресії. В остаточному підсумку, саме поєднання всіх факторів робить вирішальний вплив на результат, що передбачається моделлю.

На рисунку 1 представлені гістограми передбачених ймовірностей: а) для пробіт-моделі б) для логіт-моделі.

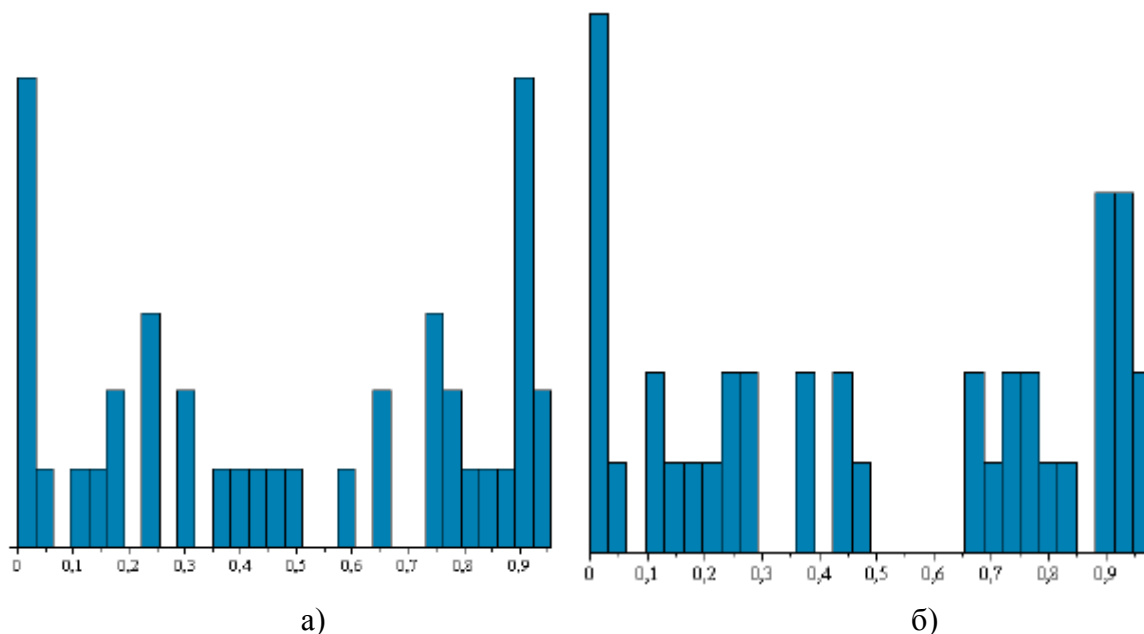


Рисунок 1 – Гістограма передбачених ймовірностей побудованих моделей:
а) пробіт-модель, б) логіт-модель

Гістограма передбачених ймовірностей (для пробіт-моделі), яка представлена на рисунку 1а, вказує на те, що більшість з них близькі або до 0, або до 1. Це означає, що модель рідко дає «невпевнену» оцінку кандидатам. На гістограмі, що представлена на рисунку 2б (логіт-модель), видно, що логістична функція рідше пророкує ймовірність проходження випробувального терміну, близьку до 0,5, що, безумовно, дає більшу впевненість в результаті.

Висновки. 1 Запропоновано використання алгоритму бінарної регресії для побудови моделі оцінювання резюме в ІТ-рекрутингу. 2 Побудовані дві моделі (пробіт і логіт) та проведено розрахунки якості обох, що дозволило визначити перевагу логіт-моделі (має більше значення псевдо- R^2 , в порівнянні з пробіт-моделлю)

ЛІТЕРАТУРА

- [1] Особливості розвитку ринку рекрутингових послуг в Україні Взято з: <http://eir.pstu.edu/bitstream/handle/123456789/12451/c.119-124.pdf?sequence=1>.
- [2] . Оптимизация практики рекрутинга. Взято з <http://ief.org.ua/docs/sr/300.pdf>.
- [3] Кобзарь, А.И. (2006) Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ.
- [4] Орлов, А.И. (2004) Прикладная статистика. М.: Экзамен.
- [5] Коваленко, І.І , Приходько, С.Б, Латанська Л.О. (2011) Сучасні методи статистичного аналізу даних: Навчальний посібник. Миколаїв: НУК.

Tkachenko A.I., Farionova T.A.

Improvement of mathematical model of evaluation of candidate summary in IT-recruiting

A mathematical model for evaluating a candidate's CV in IT recruiting based on binary regression models using probits and logit models has been improved. Based on the calculated data, we obtained the weighting factors of the factors using different estimation models and compared them.

Keywords: resume evaluation, IT recruiting, probit model, logit model.

Ткаченко А.И., Фарионова Т.А.

Усовершенствование математической модели оценивания резюме кандидата в IT-рекрутинге

Усовершенствована математическая модель оценки резюме кандидата в IT - рекрутинга на основе бинарной регрессии с использованием пробит- и логит моделей. На основе рассчитанных данных получены весовые коэффициенты факторов с применением различных моделей оценки и осуществлено их сравнения.

Ключевые слова: оценка резюме, IT-рекрутинг, пробит-модель, логит-модель.