
МАТЕМАТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ ЭКОНОМИКИ

УДК 519.86

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОЛУЧЕНИЯ ПРОГНОЗНОЙ ОЦЕНКИ НЕЧЕТКОЙ ПОЛЕЗНОСТИ УЧАСТИЯ БЕЗРАБОТНЫХ В ПРОГРАММАХ ПРОФЕССИОНАЛЬНОГО ОБУЧЕНИЯ И ПЕРЕПОДГОТОВКИ*

Т. В. Азарнова, И. Н. Щепина, И. П. Половинкин, А. С. Демидова

Воронежский государственный университет

Поступила в редакцию 26 июня 2017 г.

Аннотация: оценка эффективности программ профессионального обучения и переподготовки безработных является достаточно сложной для исследования проблемой. Сложность связана с многогранностью самого понятия «эффективность», с ограниченными возможностями получения достоверных данных о безработных после прохождения ими программы, с невозможностью отделить эффект прохождения программы от действия других факторов, оказывающих влияние на положение безработного на рынке труда. Социальная и экономическая эффективность может рассматриваться как с позиции государства, реализующего программы профессионального обучения и переподготовки на рынке труда, так и с позиции безработных, участвующих в программах. В статье анализируется возможность применения методов машинного обучения для получения прогнозных оценок нечеткой полезности участия безработных в программах обучения и переподготовки. В качестве критерия эффективности используется многокритериальная нечеткая лингвистическая полезность, которая вычисляется на основе аппарата специальных лингвистических лотерей и позволяет выявить субъективную полезность программы для конкретного безработного. Методы машинного обучения, используя в качестве обучающего множества (примеров) многомерную выборку респондентов с известной лингвистической полезностью, обучают машину (программу) определять нечеткую полезность для респондентов – кандидатов на участие в программах.

Ключевые слова: программы профессионального обучения и переподготовки безработных, нечеткая лингвистическая полезность программ, методы машинного обучения.

Abstract: evaluation of the effectiveness of vocational training programs and retraining of the unemployed is quite difficult to investigate the problem. The complexity is connected with the versatility of the concept of efficiency itself, with limited opportunities for obtaining reliable data on the unemployed after passing the program, with the inability to separate the effect of passing the program from the effects of other factors affecting the unemployed's position in the labor market. Social and economic efficiency can be viewed both from the position of the state implementing vocational training and retraining programs on the labor market, and from the position of the unemployed participating in the programs. The article analyzes the possibility of applying machine learning methods to obtain a prognostic estimate of the fuzzy utility of participation of unemployed in training and retraining programs. As a criterion of effectiveness, multi-criteria fuzzy linguistic usefulness is used, which is calculated on the basis of the apparatus of special linguistic lotteries and allows to reveal the subjective usefulness of the program for a specific unemployed person. Machine learning methods, using as a learning set (examples) a multidimensional sample of respondents with a known linguistic utility, teach the machine (program) to determine fuzzy utility for respondents – candidates for participation in programs.

Key words: vocational training and retraining programs for the unemployed, fuzzy linguistic utility programs, methods of machine learning.

Общей проблеме оценки эффективности активных программ на рынке труда посвящено достаточно много работ российских и зарубежных авторов: Дж. Хекмана, П. Тодда, Х. Ишимара, Х. Леманна, Дж. Хоули, К. Шмита, Д. Карда, Дж. Ключе, А. Ве-

бера, М. К. Абдразаковой, С. Ю. Рощина, Л. И. Задорожной, А. П. Мартыновой, Т. Четверниной, Ф. Т. Прокопова, Т. Ю. Базыленко и др. [1–5]. Программы профессионального обучения и переподготовки занимают особое место среди мер активной политики регулирования спроса и предложения на рынке труда. Исследованию данных программ, описанию методик оценки их эффективности и разработки технологий ее повышения посвящены

* Статья выполнена при финансовой поддержке гранта РФФИ (номер 16-06-00535 А).

© Азарнова Т. В., Щепина И. Н., Половинкин И. П., Демидова А. С., 2017

работы Е. А. Коломака, Д. В. Зверева, Д. В. Манушина, Б. В. Кузнецова, А. П. Мартынова, С. Ю. Рощина, М. К. Абдразаковой, Н. П. Пашина, А. В. Карпушкиной, Е. Ю. Смирновой, О. И. Зеленова, И. С. Змановского, П. П. Ляха, В. В. Манаева, Л. Н. Рудневой [3–6]. Программы профессионального обучения и переподготовки направлены на заполнение базы вакансий; сокращение времени, в течение которого безработные или высвобождающийся кадровый резерв заняты поиском работы; повышение у участников программ профессиональной мобильности и развития конкурентоспособных преимуществ на рынке труда. В современной динамично развивающейся инновационной экономике с коротким периодом равновесного состояния рынка труда программы должны быть ориентированы не только на решение текущих проблем, но и на перспективу, учитывать прогнозируемые прогрессивные изменения в структуре занятости и охватывать не только безработных, но и занятых, находящихся под риском увольнения, и людей, впервые выходящих на рынок труда. Рассматриваются разные критерии эффективности (социальные, экономические), разные методики формирования базы данных, на которой проводится оценка эффективности. В работе А. В. Косаревой [1; 2] проведена детальная структуризация подходов к оценке эффективности программ на рынке труда. Подходы отличаются субъектом, с позиции которого проводится оценка. Оценка может осуществляться на уровне региона, отрасли, предприятия, индивида, участвующего в программах. Показатели эффективности отражают степень достижения цели рассматриваемого субъекта, в качестве метрик эффективности используются сроки, качество, новизна, экономичность, результативность, удовлетворенность субъекта. В данной статье рассматривается оценка с позиции субъекта, прошедшего в ретроспективном периоде программы профессионального обучения или переподготовки. Удовлетворенность субъекта оценивается как многокритериальная нечеткая полезность. Преимущество использования нечеткой лингвистической полезности заключается в возможности выявить полезность на основе специального аппарата лингвистических лотерей и лингвистических предпочтений.

Данная статья продолжает исследования авторов, направленные на оценку эффективности программ с позиции субъектов, прошедших в ретроспективном периоде программы профессионального обучения или переподготовки [7; 8]. Удовлетворенность субъектов оценивается как многокри-

териальная нечеткая полезность. Преимущество использования нечеткой лингвистической полезности заключается в возможности выявить полезность на основе специального аппарата лингвистических лотерей и лингвистических предпочтений.

Нечеткая лингвистическая полезность оценивает результат процесса профессионального обучения или переподготовки для каждого конкретного безработного. Наряду с данной результирующей характеристикой институты рынка труда, реализующие программы, имеют в своей информационной базе широкий спектр персональных данных: социальные, демографические, профессионально-квалификационные, причины увольнения, вид трудоустройства и т. д.

Обработка ретроспективной информационной базы по участникам программы (или программ определенного типа), представленной парами вида $(a_i = (a_{i1}, a_{i2}, \dots, a_{ip}), p_i)$, $i = 1, \dots, n$ ($a_i = (a_{i1}, a_{i2}, \dots, a_{ip})$ – набор персональных данных о безработном, p_i – нечеткая полезность участия в программе), современными методами машинного обучения может позволить сформировать схемы предсказания нечеткой полезности аналогичных программ для потенциальных кандидатов на участие. В методах машинного обучения ретроспективная информационная база играет роль обучающей выборки. В статье анализируются общие аспекты применения методов машинного обучения к решению рассматриваемой задачи.

Построение нечеткой лингвистической полезности участия в программах профессионального обучения и переподготовки безработных

Многокритериальная полезность участия в программах профессионального обучения и переподготовки моделируется в работе в виде нечеткой лингвистической переменной [9] $V = \langle V, T_V, U_V \rangle$, где $T_V = \{\text{низкая, ниже среднего, средняя, выше среднего, высокая}\}$ – терм-множество значений переменной, U_V – базовое множество задания нечетких термов. В основе моделирования лежит предположение о том, что структуру полезности участия в программе можно представить в виде критериального дерева (вершинами дерева являются лингвистические критерии оценки полезности), допускающего ветвление не более чем на три взаимно независимые по полезности критерия. В качестве критериев может рассматриваться повышение материального статуса, мобильности, профессионализма, самооценки и т. д. Линг-

вистическая оценка многокритериальной полезности осуществляется путем агрегирования (декомпозиции по полезности) с нижнего уровня до корневой вершины дерева [9]. На каждом этапе ветви текущего узла дерева, представленные двумя или тремя взаимно независимыми по полезности лингвистическими критериями, агрегируются путем вычисления комплексной лингвистической полезности по данным критериям. Опишем процедуру агрегирования по трем независимым по полезности лингвистическим критериям $\langle X, T_X, U_X \rangle, \langle Y, T_Y, U_Y \rangle, \langle Z, T_Z, U_Z \rangle$ с терм-множествами $T_X = \{X_i, i = 1, \dots, n\}, T_Y = \{Y_i, i = 1, \dots, m\}, T_Z = \{Z_i, i = 1, \dots, l\}$. Вначале агрегирования вычисляется лингвистическая полезность по каждому отдельному критерию (если критерий является тупиковой вершиной дерева и его полезность не вычислена в процессе агрегирования предыдущих уровней).

При построении нечеткой функции полезности по отдельным критериям предполагается выполнение ряда специальных аксиом, связанных с системой предпочтений респондента [9; 10]. Данные аксиомы формулируются в терминах лингвистических лотерей. В общем случае лингвистическую лотерею по значениям $\{X_1, X_2, \dots, X_r\}$ случайной величины X представляют вектором $L = (P_1, X_1; P_2, X_2; \dots; P_r, X_r)$ (простая лотерея) или вектором $L = (P_1, L_1; P_2, L_2; \dots; P_r, L_r)$, где L_i лотереи вида $L = (P_1, X_1; P_2, X_2; \dots; P_r, X_r), \{P_1, P_2, \dots, P_r\}$ значения лингвистической вероятности $\langle P, T_P, U_P \rangle$ с терм-множеством $\{\text{«низкая»}, \text{«маленькая»}, \text{«средняя»}, \text{«большая»}, \text{«высокая»}\}$, заданным на базовом множестве $[0, 1]$. Для построения нечеткой полезности используются лотереи $L = (P_1, X_1; P_2, X_2)$ с двумя исходами, при формировании которых считается, что компоненты P_1, X_1, X_2 известны и заданы своими функциями принадлежности $\mu_{P_1}(p), \mu_{X_1}(x), \mu_{X_2}(x)$, а функция принадлежности $\mu_{P_2}(p)$ для P_2 определяется по следующему алгоритму [9]:

1) проводится равномерная дискретизация базового множества $[0, 1]$ конечным множеством $\{p^j\}$;

2) фиксируется значение $p_2 = p_2^*$ и анализируется допустимое множество распределения $\{(p(x_i), x_i)\}$

при выполнении условий $\sum_{i=1}^n p(x_i) \mu_{X_2}(x_i) = p_2^*$; $\sum_{i=1}^n p(x_i) = 1$. Если допустимое множество пусто, то осуществляется переход к шагу 3; если состоит из одной точки – к шагу 4; в противном случае – к шагу 5;

3) полагается $\mu_{P_2}(p_2^*) = 0$ и осуществляется переход к шагу 8;

4) вычисляется значение $\sum_{i=1}^n p(x_i) \mu_{X_1}(x_i) = p_1$, полагается $\mu_{P_2}(p_2^*) = \mu_{P_1}(p_1)$ и осуществляется переход к шагу 8;

5) при $p_2 = p_2^*$ определяется нижняя граница p_1^1 области изменения p_1 путем решения задачи:

$$\min_{p(x_i)} \sum_{i=1}^n p(x_i) \mu_{X_1}(x_i) = p_1^1,$$

$$\sum_{i=1}^n p(x_i) \mu_{X_2}(x_i) = p_2^*; \sum_{i=1}^n p(x_i) = 1;$$

6) при $p_2 = p_2^*$ определяется верхняя граница изменения p_1^2 области изменения p_1 путем решения задачи:

$$\max_{p(x_i)} \sum_{i=1}^n p(x_i) \mu_{X_1}(x_i) = p_1^2,$$

$$\sum_{i=1}^n p(x_i) \mu_{X_2}(x_i) = p_2^*; \sum_{i=1}^n p(x_i) = 1;$$

7) полагается $\mu_{P_2}(p_2^*) = \max_{p_1 \in [p_1^1, p_1^2]} \mu_{P_1}(p_1)$;

8) если множество $\{p^j\}$ просмотрено, то осуществляется переход к шагу 9, иначе – к шагу 2;

9) построенный терм P_2 аппроксимируется стандартными термами.

Используя введенное понятие лингвистической лотереи, сформулируем основные аксиомы построения функции полезности [10]:

1. Любые два значения X_i и X_j критерия сравнимы с $X_i S_{ij} X_j$, где S_{ij} – один из термов лингвистической переменной предпочтения $\langle S, T_S, U_S \rangle$ с функцией принадлежности $\mu_{S_{ij}}(s)$ (S_0 – «отсутствует предпочтение», S_e – «приблизительно эквивалентны», S_n – «несколько предпочтительнее», S_z – «значительно предпочтительнее»).

$$2. S_{ij} \geq \max_k \min \{S_{ik}, S_{kj}\}.$$

3. Лотерея вида $L = (P_1, L_1; P_2, L_2; \dots; P_r, L_r)$, где $L_j = (P_1^j, X_1; P_2^j, X_2)$, причем $X_2 S_{21} X_1, L_i S_{i1} X_1, X_2 S_{2i} L_i$ эквивалентна лотерее $L = (P_1^j, X_1; P_2, X_2)$, в которой $P_1^j = P(L_1 X_1 + \dots + L_r X_1), P_2^j = P(L_1 X_2 + \dots + L_r X_2)$.

4. Если $X_i S_{ij} X_j$ и $X_j S_{jk} X_k, S_{ij}, S_{ik} \in \{S_n, S_z\}$, то существует $L = (P_1, X_i; P_2, X_k)$ такая, что $X_j S_e L$.

5. Пусть $L = (P_j, X_j; P_k, X_k)$, тогда если $X_j S_e X'_j$, то $L S_e (P_j, X'_j; P_k, X_k)$.

6. Пусть $L_i = (P_1^i, X_1; P_2^i, X_2), L_j = (P_1^j, X_1; P_2^j, X_2)$ и $X_2 S_{12} X_2, S_{12} \in \{S_n, S_z\}$, то более предпочтительной является лотерея с большей вероятностью исхода X_1 .

Если по некоторому критерию все аксиомы выполнены, то по нему можно построить функцию полезности. В рамках работы рассматривается метод оценки лингвистических отношений и метод анализа выборов [10]. Метод оценки лингвистиче-

ских отношений базируется на предположении, что для одного из исходов X_i респондент может определить нечеткую полезность V_i с функцией $\mu_{V_i}(v)$, и лингвистическое отношение предпочтения S_{ij} между любыми двумя исходами X_i и X_j . Функция принадлежности неизвестной полезности любого исхода X_j восстанавливается по формуле:

$$\mu_{V_j}(v) = \sup_{s \in U_s} \min \{ \mu_{S_{ij}}(s), \mu_{V_i}(s \cdot v) \}.$$

В методе анализа выборов используются лингвистические лотереи. Пусть X_j – исход, для которого требуется определить полезность, L_j – лингвистическая лотерея с такими исходами X^0 и X^* , что исход X^* «несколько предпочтительнее» или «значительно предпочтительнее», чем X^0 и любой другой исход; любой исход «несколько предпочтительнее» или «значительно предпочтительнее», чем X^0 ; полезности исходов X^0 и X^* известны. Респондента просят указать значения P^{j1} и P^{j2} лингвистической вероятности $\langle P, T_p, U_p \rangle$, при которых для него исход X_j становится приблизительно эквивалентным участию в лотерее $L_j = (P^{j1}, X^0; P^{j2}, X^*)$. Для определения полезности исхода X_j вычисляется нечеткая ожидаемая полезность участия в лотерее и V_j восстанавливается по отношению «почти эквивалентности» между лотереей и исходом X_j . Полезность лотереи L_j с исходами X^0 и X^* полагается равной P^{j1} .

После вычисления полезности по каждому критерию осуществляется переход к построению многокритериальной полезности, метод декомпозиции работает в предположении взаимной независимости по полезности выбранных критериев. Критерий K^1 называется независимым по полезности от критерия K^2 [8], если отношение предпочтения между двумя лотереями, в которых исходы отличаются лишь значениями критерия K^1 , не зависит от значения критерия K^2 . Критерии K^1 и K^2 взаимно независимы по полезности, если K^1 не зависит от K^2 и K^2 не зависит от K^1 .

Для взаимно независимых по полезности критериев $\langle X, T_x, U_x \rangle, \langle Y, T_y, U_y \rangle, \langle Z, T_z, U_z \rangle$ нечеткая функция полезности вычисляется по формуле:

$$\begin{aligned} V(Y_j, Z_k, X_i) &= k_y v_y^-(Y_j) + \frac{v_y^+(Y_j) - v_y^-(Y_j)}{v_y^+(Y_0)} k_z v_z^-(Z_k) + \\ &+ \frac{v_y^+(Y_j) - v_y^-(Y_j)}{v_y^+(Y_0)} \frac{v_z^+(Z_k) - v_z^-(Z_k)}{v_z^+(Z_0)} k_x v_x^-(X_i) + \\ &+ \frac{v_y^+(Y_j) - v_y^-(Y_j)}{v_y^+(Y_0)} \frac{v_z^+(Z_k) - v_z^-(Z_k)}{v_z^+(Z_0)} \times \\ &\times \frac{v_x^+(X_i) - v_x^-(X_i)}{v_x^+(X_0)} k_x V_x(X_0), \end{aligned}$$

где $V_y(Y_j), V_z(Z_k), V_x(X_i)$, – нормализованные нечеткие функции полезности с $v_y^-(Y_0) = 0, v_y^+(Y_1) = 1, v_z^-(Z_0) = 0, v_z^+(Z_1) = 1, v_x^-(X_0) = 0, v_x^+(X_1) = 1$ (v^- – левая граница носителя нечеткого множества, v^+ – правая граница носителя нечеткого множества); k_x, k_y, k_z – положительные константы, удовлетворяющие условиям

$$k_z = k_y \frac{v_y^+(Y_0)}{v_z^+(Z_0)}, \quad k_x = k_y \frac{v_y^+(Y_0)}{v_x^+(X_0)}.$$

Левые границы вычисляются по формулам:

$$v_y^-(Y_j) = \frac{v_y^+(Y_j) - v_y^+(Y_0)}{1 - k_y \alpha v_y^+(Y_0)},$$

$$v_z^-(Z_k) = \frac{v_z^+(Z_k) - v_z^+(Z_0)}{1 - k_z \alpha v_z^+(Z_0)},$$

$$v_x^-(X_i) = \frac{v_x^+(X_i) - v_x^+(X_0)}{1 - k_x \alpha v_x^+(X_0)},$$

величина α – по формуле:

$$\alpha = \frac{v_y^+(Y_0) - v_y^+(Y_1) + v_y^-(Y_1)}{k_y v_y^-(Y_1) v_y^+(Y_0)}.$$

Для введенных величин выполняется равенство

$$\frac{v_y^-(Y)}{v_z^-(Z)} = \frac{v_y^+(Y) - v_y^+(Y_0)}{v_z^+(Z) - v_z^+(Z_0)}.$$

Таким образом, для вычисления нечеткой полезности $V(Y_j, Z_k, X_i)$ можно использовать следующую последовательность шагов:

- 1) устанавливаются правые границы $v_y^+(Y_j), v_z^+(Z_k), v_x^+(X_i)$;
- 2) определяется $V_x(X_0)$;
- 3) устанавливается $v_y^-(Y_1)$;
- 4) в соответствии с равенством $\frac{v_y^-(Y)}{v_z^-(Z)} = \frac{v_y^+(Y) - v_y^+(Y_0)}{v_z^+(Z) - v_z^+(Z_0)}$ вычисляется $v_z^-(Z_1)$;

5) вычисляются k_x, k_y, k_z :

$$\begin{aligned} k_y v_y^-(Y_1) + \frac{1 - v_y^-(Y_1)}{v_y^+(Y_0)} k_z v_z^-(Z_1) + \\ + \frac{1 - v_y^-(Y_1)}{v_y^+(Y_0)} \frac{1 - v_z^-(Z_1)}{v_z^+(Z_0)} k_x v_x^-(X_1) = 1; \end{aligned}$$

6) вычисляется α ;

7) вычисляются $v_y^-(Y_j), v_z^-(Z_k), v_x^-(X_i)$:

$$v_y^-(Y_j) = \frac{v_y^+(Y_j) - v_y^+(Y_0)}{1 - k_y \alpha v_y^+(Y_0)},$$

$$v_z^-(Z_k) = \frac{v_z^+(Z_k) - v_z^+(Z_0)}{1 - k_z \alpha v_z^+(Z_0)},$$

$$v_x^-(X_i) = \frac{v_x^+(X_i) - v_x^+(X_0)}{1 - k_x \alpha v_x^+(X_0)}$$

8) вычисляется $V(X_p, Y_j, Z_k)$;

9) построенная нечеткая полезность аппроксимируется одним из базовых термов.

В рамках исследования для вычисления многокритериальной иерархической полезности было разработано комплексное программное обеспечение, реализующее все этапы лингвистической обработки критериев и описанные выше алгоритмы вычисления нечеткой полезности отдельных критериев и многокритериальной полезности в соответствии с иерархическим критериальным деревом. Продемонстрируем работу данного обеспечения на небольшом примере декомпозиции полезности по трем взаимно независимым критериям: мотивация к успеху (низкая, средняя, умеренно высокая, слишком высокая), уровень соответствия ценностей и степень удовлетворенности потребностей (неудовлетворенность, частичное неудовлетворение, удовлетворительное соответствие, полное удовлетворение), нервно-психическая устойчивость (неустойчивость, недостаточная устойчивость, полная устойчивость).

В процессе лингвистической обработки критериев методом «α-срезов» строятся соответствующие им терм-множества (рис. 1–3).

Следующим этапом работы программного обеспечения является построение нечеткой функции полезности для каждого критерия. Рассмотрим пример формирования функции полезности по критерию «Мотивация к успеху» для одного из участников программы (рис. 4). Полезность для значения «Умеренно высокая» полагается равной «Высокая» и устанавливается следующая система предпочтений: полезность термина «Слишком высокая» эквивалентна полезности термина «умеренно высокая»; полезность термина «Умеренно высокая» несколько предпочтительнее полезности термина «Средняя»; полезность термина «Умеренно высокая» значительно предпочтительнее полезности термина «Низкая».

В соответствии с данными предпочтениями получились функции полезности термов критерия «Мотивация к успеху», представленные на рис. 4.

Для построения функции полезности по трем критериям вводится предположение, что для набора полезностей по отдельным критериям «Умеренно высокая», «Удовлетворительное соответствие»,

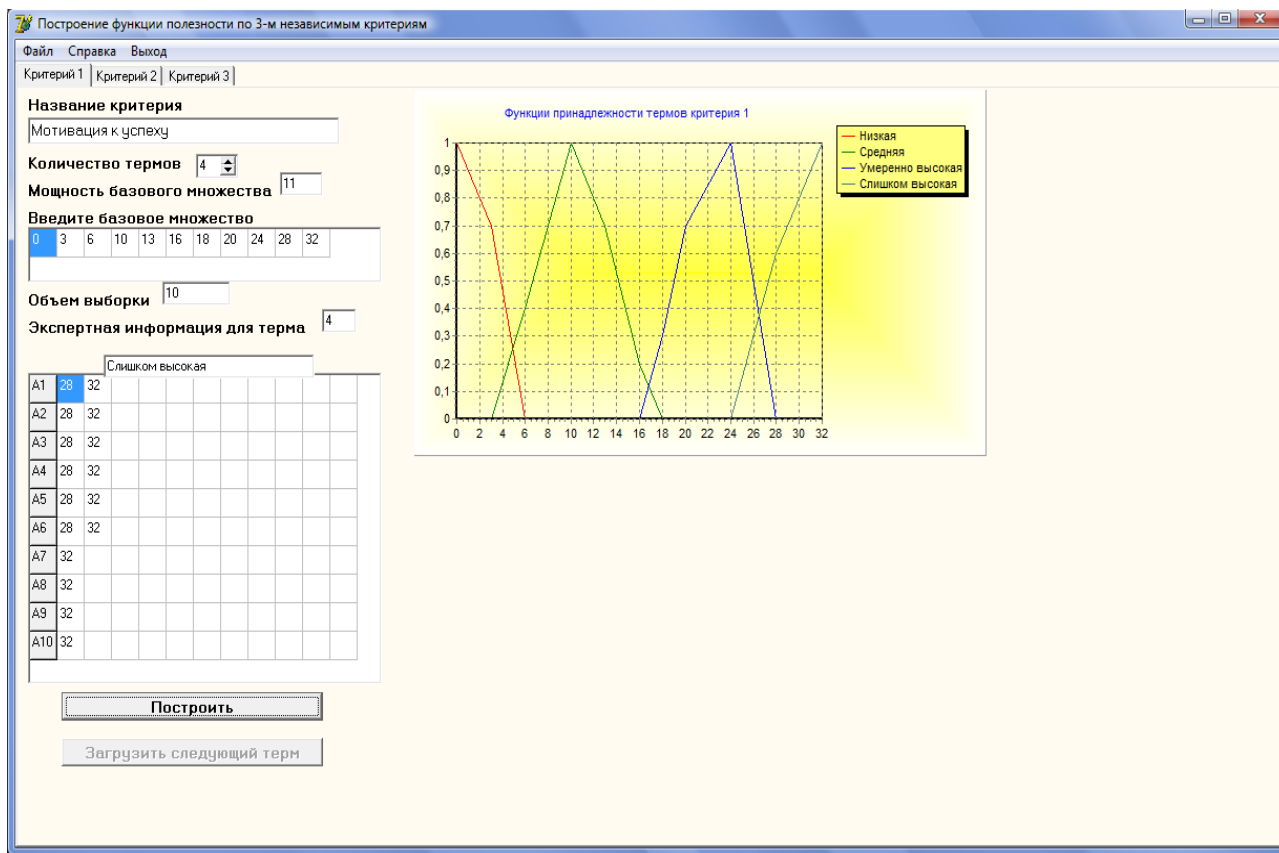


Рис. 1. Построение термов критерия «Мотивация к успеху»

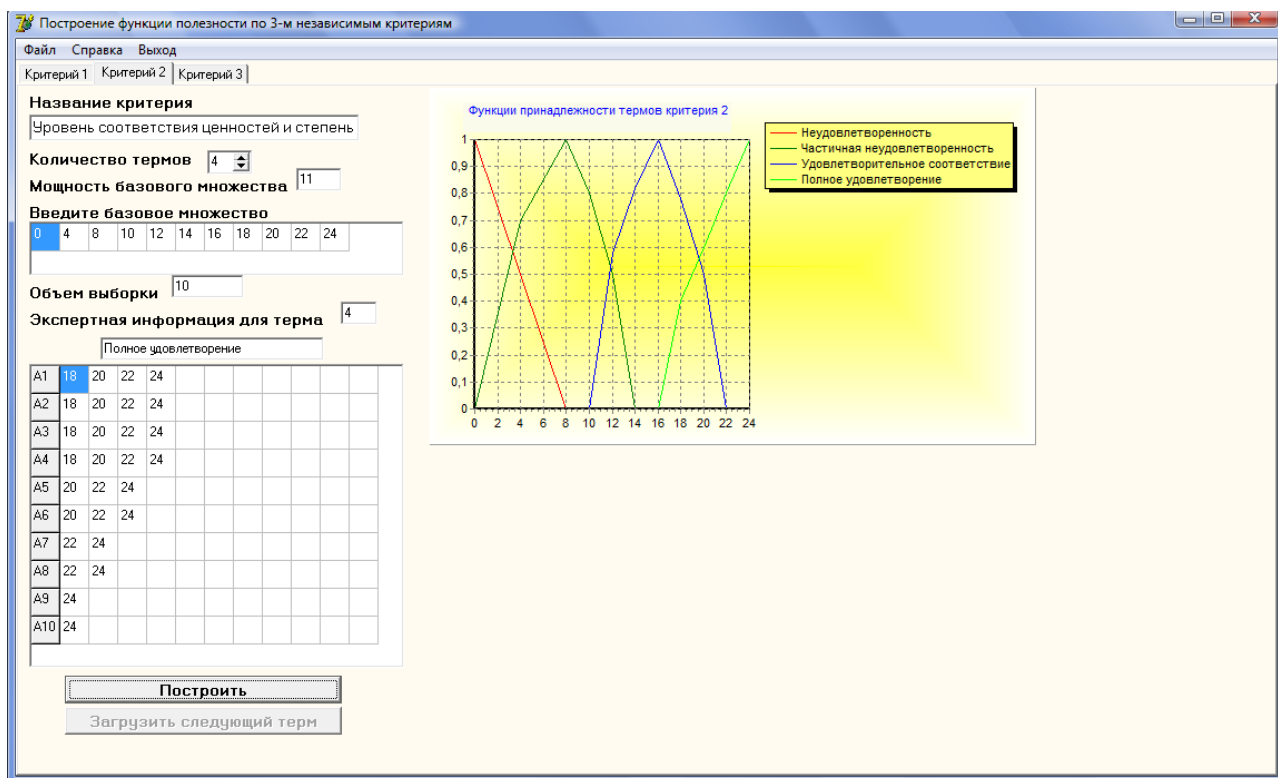


Рис. 2. Построение термов критерия «Уровень соответствия ценностей и степень удовлетворенности потребностями»

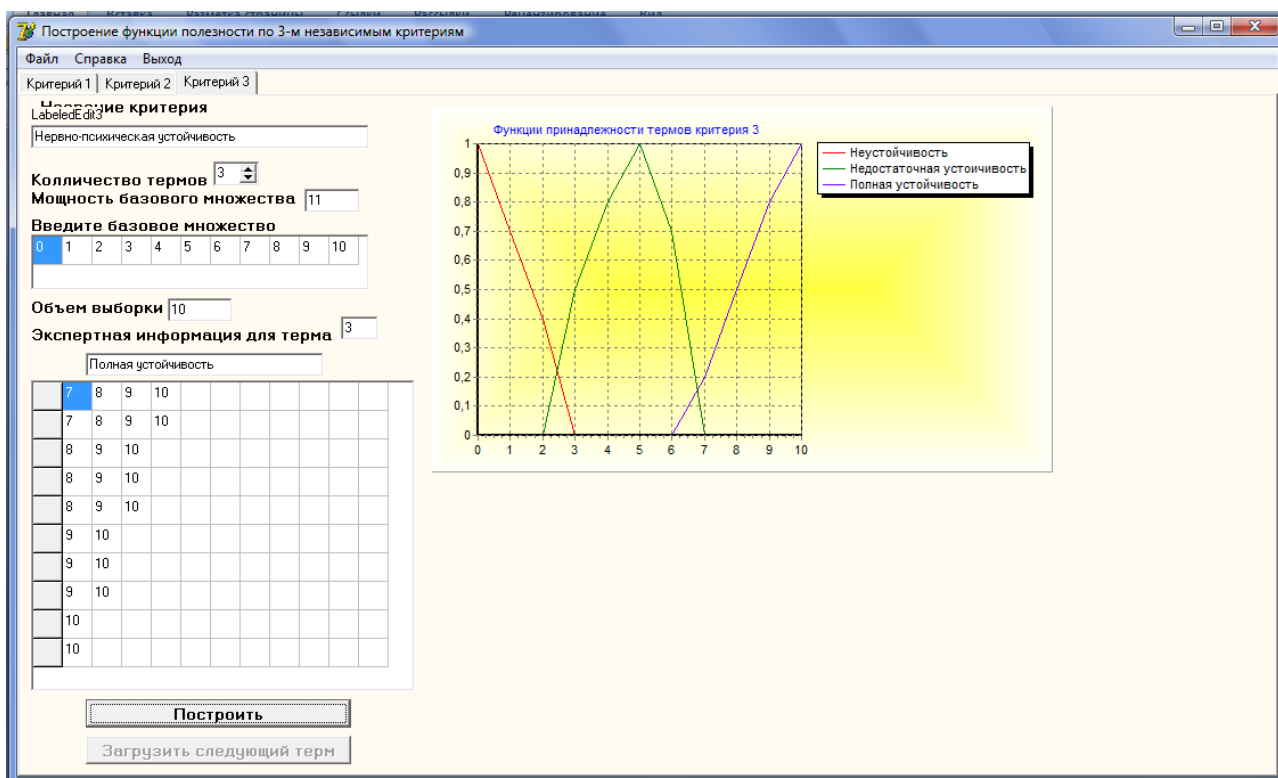


Рис. 3. Построение термов критерия «Нервно-психическая устойчивость»

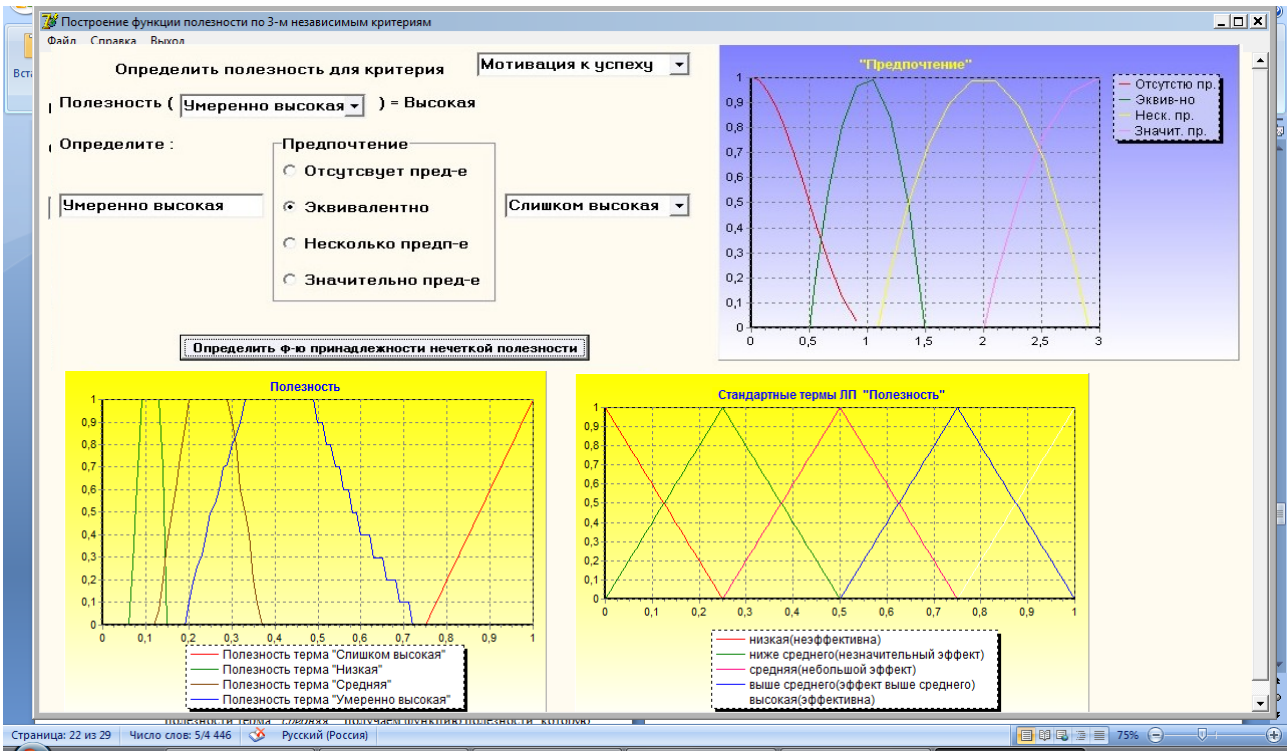


Рис. 4. Функции полезности термов критерия «Мотивация к успеху»

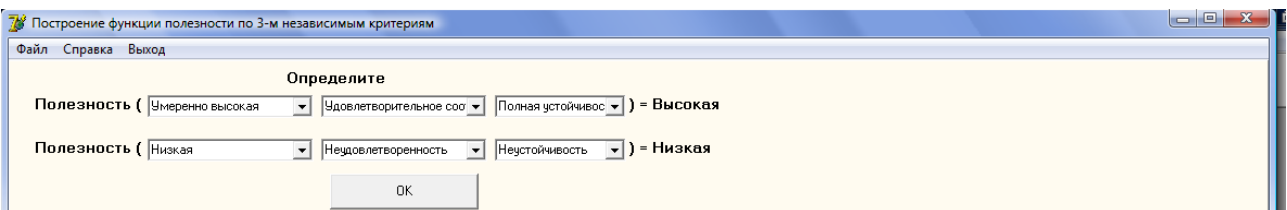


Рис. 5. Описание исходов с «высокой» и «низкой» многокритериальной полезностью

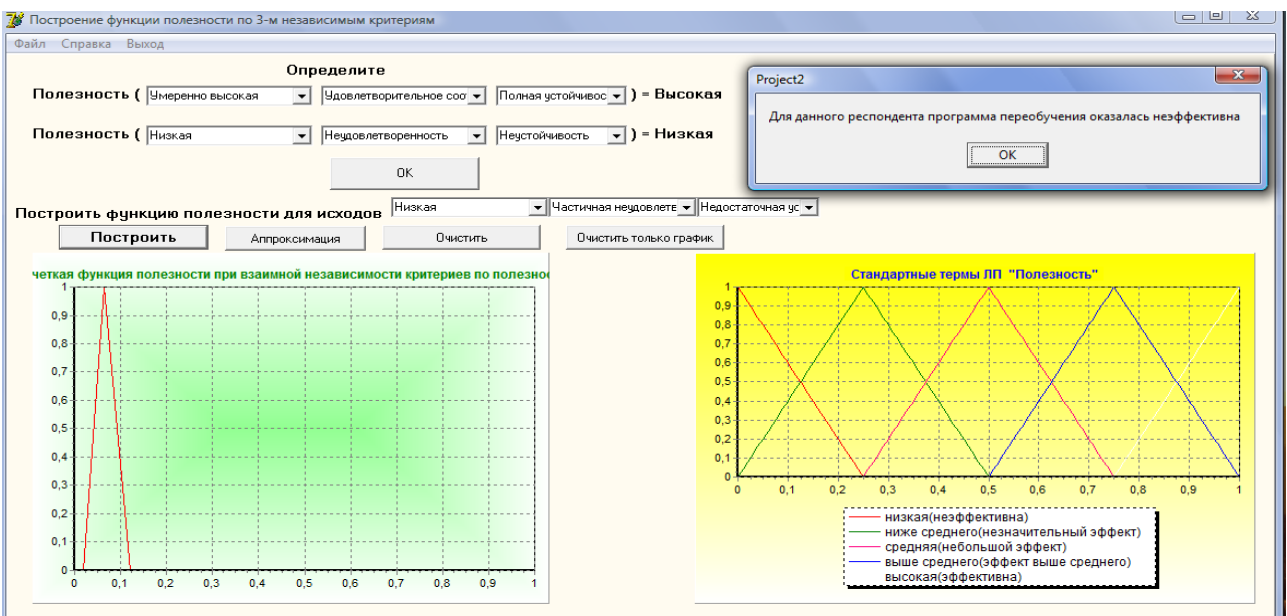


Рис. 6. Результирующий терм многокритериальной полезности

«Полная устойчивость» многокритериальная полезность считается «высокой», а для набора «Низкая», «Неудовлетворенность», «Неустойчивость» – «низкой» (рис. 5).

На рис. 6 приведено вычисление терма многокритериальной полезности для набора значений критериев «Низкая», «Частичная неудовлетворительное», «Недостаточная устойчивость». Аппроксимация результирующего терма одним из базовых термов показывает низкую полезность программы.

Анализ возможностей применения методов машинного обучения для прогнозирования эффективности программ

Оценка лингвистической полезности для респондентов, прошедших программы профессионального обучения или переподготовки, в совокупности с набором социально-демографических, профессионально-квалификационных характеристик позволяет создать информационную базу (таблица), которая разбивает респондентов на классы по полезности и создает основу для поиска различий между респондентами с различной полезностью и прогнозирования полезности для новых респондентов – кандидатов на участие в программах. Большим потенциалом для первоначального формирования и предсказания принадлежности к определенному классу обладают методы машинного обучения [11–13]. Среди методов классификации и прогнозирования в машинном обучении широкое применение на практике получили следующие методы: опорных векторов,

дискриминантного анализа, байесовской классификации, решающих правил, случайного леса, градиентного бустинга. Алгоритмы машинного обучения применительно к задачам классификации и прогнозирования можно разделить на статистические и нестатистические. Классические статистические методы предполагают, что наборы исходных данных представляют собой реализацию определенного стационарного стохастического процесса, параметры которого можно оценить на основании имеющейся информационной базы. Правило предсказания строится на основании стохастического процесса с уточненными параметрами. При использовании статистического подхода функция риска для используемого правила предсказания определяется как среднее значение функции потерь, вычисляющей различие между предсказаниями и исходами. Среднее значение определяется в соответствии с найденным распределением исследуемого случайного процесса. В статистических методах машинного обучения также используется стохастическая модель генерации данных, вероятность ошибочной классификации или регрессии называется *ошибкой обобщения*. Отличие от классических методов заключается в том, что распределение наборов информационной базы, как правило, неизвестно, параметры распределения не используются при оценке ошибок классификации или регрессии; заранее неизвестно, какой метод классификации будет построен по обучающей выборке, может быть задан целый класс таких методов, оценки

Т а б л и ц а

Социально-демографические и профессионально-квалификационные характеристики участников программ

Наименование показателя	Описание возможных значений показателя
Профиль образования x_1	{1 – математическое, техническое; 2 – естественное, 3 – гуманитарное}
Возраст x_2	Количество полных лет по паспорту
Опыт работы x_3	Количество лет работы по трудовой книжке
Число месяцев без работы x_4	Число месяцев с момента потери работы до момента прохождения программы
Состояние здоровья x_5	{1 – хорошее; 2 – удовлетворительное; 3 – наличие инвалидности или ограничений в работе}
Способность к обучению x_6	$x_6 =$ баллы от 0 до 100
Приоритеты в работе x_7	{1 – деньги; 2 – карьера; 3 – общение с людьми; 4 – изучение нового; 5 – самореализация}
Оценка стрессоустойчивости x_8	Баллы от 0 до 10
Состав семьи x_9	Количество членов семьи (считая супруга и детей)
Класс по полезности программы	{1 – «программа неэффективна», 2 – «программа незначительно эффективна», 3 – «небольшой эффект программы», 4 – «эффект программы выше среднего», 5 – «программа эффективна»}

ошибки обобщения должны быть равномерными по всем вероятностным распределениям и применяемым методам. Для оценки предсказательной способности используемой технологии классификации применяется теория обобщения, сущность которой заключается в том, что вычисляются оценки вероятности ошибки классификации будущих данных при условии, что обучение проведено на случайной обучающей выборке большого объема и функция классификации согласована с обучающей выборкой. Важнейшим параметром подобной оценки является сложность класса функций классификации и длина выборки. Сложность классов функций измеряется несколькими способами: функцией роста с VC-размерностью; числом покрытия и упаковки с fat-размерностью; средними Радемахера [11].

Нестохастические методы в основном представлены методами предсказания индивидуальных последовательностей, они не используют гипотезы о стохастических механизмах генерации данных. Для оценки качества предсказаний используются тесты, характеризующие рассогласованность между предсказаниями и соответствующими исходами, которые привязаны к задачам, для решения которых будут использоваться предсказания. Наиболее распространенными являются тесты на калибруемость. Алгоритмы должны выдавать предсказания, выдерживающие тесты на калибруемость. Может рассматриваться сравнительная теория предсказания, при использовании которой качество алгоритма предсказания оценивается путем сравнения с предсказаниями определенного набора экспертных методов или экспертов. Качество предсказателя оценивается в

наихудшем случае относительно всех возможных исходов. Рассматривается разность между кумулятивными потерями предсказателя и кумулятивными потерями экспертов. Ошибка алгоритма предсказателя (регрет – regret) определяется как минимальное значение такой разности при произвольной последовательности исходов.

В заключение в качестве примера применения методов машинного обучения к решению задач прогнозирования эффективности программы профессиональной переподготовки приведем прогнозирование средствами дискриминантного анализа [14] в пакете прикладных программ Statistica 10. Рассмотрим 30 респондентов, прошедших программу переподготовки, каждый из которых характеризуется классом по полезности и набором показателей, приведенных в таблице.

На первом этапе работы алгоритма дискриминантного анализа, по коэффициентам дискриминантных функций (рис. 7), оцененным по информационной базе, исследуется значимость различных показателей участников программ с точки зрения различий между классами. Результаты анализа показывают, что значимыми являются практически все показатели (образование; возраст; опыт работы; время, проведенное без работы; приоритеты в работе; стрессоустойчивость).

На втором этапе проводится обучающая классификация, с использованием априорных вероятностей, пропорциональных объемам классов (рис. 8).

Результаты классификации показывают отсутствие несовпадений классов в первоначальной классификации и классификации, полученной с помощью функций Фишера. Это говорит о хорошем

Переменная	Стандартиз. коэффициенты (анкета30) для канонич. переменных			
	Кор. 1	Кор. 2	Кор. 3	Кор. 4
Образование	0,25316	-0,056615	-0,076095	-0,274391
Возраст	0,63677	-0,157951	-0,728909	-0,188337
Опыт работы	0,51359	0,377512	0,066241	-0,400385
Безраб/мес	0,62116	-0,632104	0,335832	-0,257693
здоровье	-0,14775	-0,194504	-0,502314	0,171959
спос-ть/обуч	-0,98146	-0,271019	-0,211773	-0,560634
приоритеты/раб	0,44282	-0,044900	0,060656	0,230179
стресс/уст	-0,78107	-0,784262	-0,270635	0,613094
семья	0,13506	0,257410	-0,443029	0,424761
Соб. зн.	76,37703	9,328405	5,706234	1,674626
Кум. доля	0,82050	0,920709	0,982010	1,000000

Рис. 7. Анализ дискриминантной способности переменных

Наблюдение	Классификация наблюдений (анкета30) Неправильные классификации отмечены *					
	Наблюд. Класс.	1 p=,20000	2 p=,20000	3 p=,23333	4 p=,16667	5 p=,20000
1	G 1:1	G 1:1	G 2:2	G 3:3	G 4:4	G 5:5
2	G 1:1	G 1:1	G 2:2	G 3:3	G 4:4	G 5:5
3	G 1:1	G 1:1	G 2:2	G 3:3	G 4:4	G 5:5
4	G 1:1	G 1:1	G 2:2	G 3:3	G 4:4	G 5:5
5	G 1:1	G 1:1	G 2:2	G 3:3	G 4:4	G 5:5
6	G 1:1	G 1:1	G 2:2	G 3:3	G 4:4	G 5:5
7	G 2:2	G 2:2	G 3:3	G 1:1	G 4:4	G 5:5
8	G 2:2	G 2:2	G 3:3	G 1:1	G 4:4	G 5:5
9	G 2:2	G 2:2	G 3:3	G 1:1	G 4:4	G 5:5
10	G 2:2	G 2:2	G 1:1	G 3:3	G 4:4	G 5:5
11	G 2:2	G 2:2	G 3:3	G 1:1	G 4:4	G 5:5
12	G 2:2	G 2:2	G 1:1	G 3:3	G 4:4	G 5:5
13	G 3:3	G 3:3	G 2:2	G 1:1	G 4:4	G 5:5
14	G 3:3	G 3:3	G 2:2	G 1:1	G 4:4	G 5:5
15	G 3:3	G 3:3	G 2:2	G 1:1	G 4:4	G 5:5
16	G 3:3	G 3:3	G 2:2	G 1:1	G 4:4	G 5:5
17	G 3:3	G 3:3	G 2:2	G 1:1	G 4:4	G 5:5
18	G 3:3	G 3:3	G 2:2	G 1:1	G 4:4	G 5:5

Рис. 8. Фрагмент результатов обучающей классификации

качестве классификации с помощью термов многокритериальной нечеткой функции полезности.

С помощью классифицирующих функций была построена прогнозная оценка эффективности программ для следующего респондента: гуманитарное образование (3), 35 лет, опыт работы – 16 лет, 4 месяца без работы, имеет хронические заболевания (3), способность к обучению 70, приоритет в работе – самореализация (5), стрессоустойчивость (4), женат, имеет ребенка (3). По результатам прогноза для данного респондента эффект про-

граммы «выше среднего».

В статье рассмотрены основные аспекты построения многокритериальной нечеткой полезности участия в программах профессионального обучения и переподготовки безработных и предсказания на основе методов машинного обучения эффективности участия в программах новых кандидатов. Применение предложенных технологий на практике может быть реализовано с помощью широкого спектра программных систем машинного обучения.

ЛИТЕРАТУРА

1. Косарева А. В. Сравнительный анализ эффективности антикризисных мер на рынках труда федеральных округов / А. В. Косарева // Научное обозрение. – 2012. – № 2. – С. 584–594.
2. Косарева А. В. Профессиональное обучение как фактор повышения конкурентоспособности лиц, ищущих работу / А. В. Косарева // Каспийский регион : политика, экономика, культура. – 2012. – № 1. – С. 128–134.
3. Лях П. П. Эффективность профессионального обучения безработных / П. П. Лях // Электронное научное издание «Ученые записки ТОГУ». – 2013. – Т. 4. – № 4. – С. 2015–2021.

4. Ананченкова П. И. Профессиональное обучение безработных в системе службы занятости населения / П. И. Ананченкова, М. Ю. Кузнецов // Ученые записки Российской академии предпринимательства. – 2016. – № 47. – С. 176–183.
5. Мартынова А. П. Оценка эффективности государственных программ на рынке труда : эмпирическое исследование спроса и ограничений на оценку / А. П. Мартынова, С. Ю. Рошин : Препринт WP15/2008/04/. – М. : ГУВШЭ, 2008. – 28 с.
6. Зеленова О. И. Эффективность профессионального обучения в системе государственной службы занятости / О. И. Зеленова // Журнал исследований социальной политики. – 2006. – Т. 4. – № 3. – С. 367–380.

7. Азарнова Т. В. Методы оценки полезности участия безработных в активных программах профессионального обучения / Т. В. Азарнова // Вестник Воронеж. гос. техн. ун-та. – 2008. – Т. 4. – № 11. – С. 106–109.

8. Азарнова Т. В. Разработка системы поддержки принятия решений для подбора программ профессионального обучения и переподготовки безработных / Т. В. Азарнова, Е. Л. Кретьнина, А. В. Столбовская // Вестник Воронеж. гос. ун-та. Сер.: Системный анализ и информационные технологии. – 2007. – № 1. – С. 17–25.

9. Борисов А. Н. Обработка нечеткой информации в системах принятия решений / А. Н. Борисов, А. В. Алексеев, Г. В. Меркурьева. – М.: Радио и связь, 1989. – 304 с.

10. Борисов А. Н. Методика оценки функции принадлежности нечеткого множества / А. Н. Борисов,

Я. Я. Осис // Кибернетика и диагностика. – Рига: Риж. политех. ин-т, 1970. – С. 125–134.

11. Вьюгин В. В. Элементы математической теории машинного обучения: учеб. пособие / В. В. Вьюгин. – М.: МФТИ: ИППИ РАН, 2010. – 231 с.

12. Breiman L. Random forests // Machine learning. – 2001. – Vol. 45, no. 1. – P. 5–32.

13. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. – 2016. – 10 с. – Режим доступа: [https://arxiv.org/abs/1603.02754. DOI: 10.1145/2939672.2939785] [chrome-extension://ilhpdfjlmhfdgdbefpinebijmhjijpn/http://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf]

14. Каримов Р. Н. Основы дискриминантного анализа / Р. Н. Каримов. – Саратов: Изд-во СГТУ, 2002. – 108 с.

Воронежский государственный университет

Азарнова Т. В., доктор технических наук, заведующая кафедрой математических методов исследования операций

E-mail: ivdas92@mail.ru

Тел.: 8 (473) 220-82-82

Voronezh State University

Azarnova T. V., Doctor of Engineering, Head of the Mathematical Methods of Operations Research Department

E-mail: ivdas92@mail.ru

Тел.: 8 (473) 220-82-82

Щепина И. Н., доктор экономических наук, доцент кафедры информационных технологий и математических методов в экономике

E-mail: shchepina@mail.ru

Тел.: 8 (473) 228-11-60 + 5116

Shchepina I. N., Doctor of Economic Sciences, Associate Professor of Informational Technology and Mathematical Methods in Economy Department

E-mail: shchepina@mail.ru

Тел.: 8 (473) 228-11-60 + 5116

Половинкин И. П., доктор физико-математических наук, профессор кафедры математического и прикладного анализа

E-mail: polovinkin@yandex.ru

Тел.: 8 (473) 220-83-48

Polovinkin I. P., Doctor of Physics and Mathematics, Professor of the Mathematical and Applied Analysis Department

E-mail: polovinkin@yandex.ru

Тел.: 8 (473) 220-83-48

Демидова А. С., аспирант кафедры математических методов исследования операций

Тел.: 8 (473) 220-82-82

Demidova A. S., Post-graduate Student of the Mathematical Methods of Operations Research Department

Тел.: 8 (473) 220-82-82