

ОЦЕНКА ЭФФЕКТИВНОСТИ И СРАВНЕНИЕ МОДЕЛЕЙ БИЗНЕС-АНАЛИТИЧЕСКИХ ПРОЕКТОВ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИЙ DATA MINING

УДК 004.051 (004.052.2)

Алла Алексеевна Попкова,
к.т.н., доц. каф. Информационных систем
и технологий, Норильский индустриальный
институт (НИИ)
Тел.: (3919) 22-45-36
Эл. почта: infalla@bk.ru

В статье излагаются методические основы сравнения разнотипных моделей, реализующих решение поставленной научной или инженерной задачи многомерной обработки данных. Основное внимание сосредоточено на оценке эффективности и точности классификационных и регрессионных моделей. Выделены специальные методики сравнения такие, как Lift- и Profit-кривые, Gain-диаграммы, ROC-анализ, матрицы классификаций и издержек.

Ключевые слова: системы поддержки принятия решений (СППР), бизнес-аналитика, технология Data Mining, оценка адекватности моделей, оценка эффективности классификационных моделей, Lift-кривые, Profit-кривые, Gain-диаграммы, ROC-анализ, матрица классификаций, матрица издержек.

Alla A. Popkova,
PhD in Technical Sciences, Associate
Professor, the Department of Information
Systems and Technologies, Norilsk industrial
institute (NII)
Tel.: (3919) 22-45-36
E-mail: infalla@bk.ru

ASSESSMENT OF EFFICIENCY AND COMPARISON OF MODELS OF BUSINESS ANALYTICAL PROJECTS WITH APPLICATION OF THE DATAMINING TECHNOLOGIES

The methodical bases of comparison of polytypic models which solve the scientific or engineering task of multidimensional data processing are stated in the article. The main attention is paid to an assessment of efficiency and accuracy of classification and regression models. Special methods of comparison such as Lift- and Profit- curves, Gain- charts, ROC- analysis, matrixes of classifications and costs are marked out.

Keywords: Decision Support Systems (DSS), Business Analytics, Data Mining technology, assessment of models adequacy, assessment of efficiency of classification models, Lift-curves, Profit-curves, Gain-charts, ROC-analysis, matrixes of classifications, costs classification.

1. Введение

Бизнес-аналитические проекты с применением технологий Data Mining, объединяющие методы оценивания эффективности и сравнения различных моделей и получивших большое распространение в бизнесе, экономике, маркетинге, медицине, науке и технике, являются дорогостоящими и трудоёмкими разработками [1]. Достаточно очевидно, что построенную модель нельзя применять на практике до тех пор, пока не будут проверены её эффективность и точность и, не будет доказано, что данная модель является наилучшей по сравнению с остальными предлагаемыми разнотипными моделями СППР. Существующие технологии Data Mining являются малоизвестными и сложными в их применении для оценки адекватности построенных моделей СППР и на сегодняшний день предлагаемые методики (Lift-, Profit-кривые и др.) получили широкое применение в случае бинарной классификации. Рассмотрим методики оценки эффективности и сравнения классификационных моделей для случая 4-арной классификации на примере модулей, задаваемых совокупностью баз продукционных правил Y_1 , Y_2 и Y_3 гибридной нейро-нечёткой сети оценки эффективности управления информационно-библиотечным фондом вуза (рис. 1) [2].

С целью корректной настройки параметров полученных моделей было проведено обучение трёх модулей типа ANFIS с помощью метода, который дал наименьшую ошибку классификации (табл. 1). Выполним оценку адекватности полученных моделей с помощью технологий Data Mining.

2. Оценка адекватности гибридной нейро-нечёткой сети

Ошибки, допускаемые моделью при её практическом применении на новых данных, могут приводить к ложным выводам, что влечёт материальные и, следовательно, финансовые издержки ошибки классификации от неправильно принятых решений. При построении модели необходимо минимизировать вероятность появления ошибок, которые вызывают наибольшие издержки классификации. Такая методика называется классификацией с учётом издержек (cost-sensitive classification).

Рассмотрим модель на основе базы правил (БП) Y_1 , выходом которой является интегральный показатель – «Уровень материально-технической базы библиотеки». В качестве положительного исхода предсказания выбрано значение *Высокий*, а в качестве отрицательного – *Низкий*. Дополнительно определим классы *Нормальный (Скорее высокий)* и *Средний (Скорее низкий)* (табл. 2).

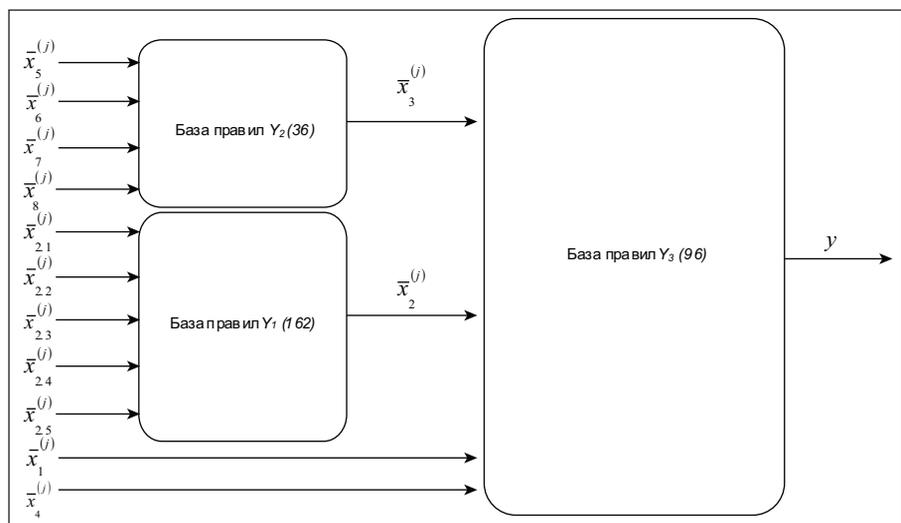


Рис. 1. Структура гибридной нейро-нечёткой сети типа ANFIS

Таблица 1.

Экспериментальные результаты исследования методов обучения модулей нейро-нечёткой сети

Модуль нейро-нечёткой сети	Название алгоритма обучения	Количество итераций обучения (среднее значение)	Количество дополнительных переменных для организации вычислительного процесса (среднее значение)	Величина ошибки обучения E (среднее значение)	Скорость сходимости алгоритма t (среднее время обучения, час.)	Наличие срывов обучения
Модуль нейро-нечёткой сети на основе базы правил Y ₁	Алгоритм конкурентного обучения (competitive learning)	До 100000	64	0,00066	0,21	очень редко
	Имитация отжига	До 100000	142	0.02313	3,4	редко
	Градиентный алгоритм	До 100000	92	0,00060	0,20	часто
	Генетический алгоритм	До 100000	42	0.00060	1,5	очень редко
Модуль нейро-нечёткой сети на основе базы правил Y ₂	Генетический алгоритм	до 500000	42	0,00060	2,4	очень редко
	Алгоритм конкурентного обучения	до 500000	64	0,00160	2,2	редко
	Имитация отжига	до 500000	156	0,00240	4,4	редко
	Градиентный алгоритм	до 500000	86	0,00058	2,2	редко
Модуль нейро-нечёткой сети на основе базы правил Y ₃	Генетический алгоритм	до 500000	42	0,00055	3,1	очень редко
	Градиентный алгоритм	до 500000	71	0,00054	2,0	очень редко
	Алгоритм конкурентного обучения	до 500000	64	0,00068	3,6	редко
	Имитация отжига	до 500000	124	0.02111	6,7	редко

Таблица 2.

Матрица классификации

		Предсказанный класс (модель)			
		Низкий	Средний	Нормальный	Высокий
Фактический класс	Низкий	TN	FM	FR	FP
	Средний	MN	TM	MR	MP
	Нормальный	RN	RM	TR	RP
	Высокий	FN	PM	PR	TP

Таблица 3.

Распределение откликов по результатам наблюдений

Модуль нейро-нечёткой сети на основе базы правил Y ₁ (общее количество наблюдений – 15 000)					
Название алгоритма обучения	Значение предсказания				
Алгоритм конкурентного обучения (competitive learning)	Низкий	1185	480	495	465
	Средний	540	1215	525	510
	Нормальный	570	555	1860	555
	Высокий	1800	1170	1185	1890
Имитация отжига	Низкий	1110	615	585	585
	Средний	585	1200	585	555
	Нормальный	525	615	1560	675
	Высокий	2250	870	945	1740
Градиентный алгоритм	Низкий	660	870	825	735
	Средний	495	840	885	855
	Нормальный	510	525	1166	930
	Высокий	2730	870	887	1218
Генетический алгоритм	Низкий	315	1185	1005	1095
	Средний	495	450	1080	1065
	Нормальный	510	900	465	750
	Высокий	3225	930	1065	465

Используем классификацию (для 4-х классов) для отдельного предсказания, которое имеет шестнадцать исходов (табл. 3):

- 1) Истинноположительный (true positive, TP) исход будет, когда фактическое целевое значение для данного примера *Высокий* и модель на выходе выдаст *Высокий* (предсказанный класс).
- 2) Истиннонормальный (true rate, TR) исход, когда фактическое целевое значение для данного наблюдения *Нормальный* и модель на выходе выдаст *Нормальный* (предсказанный класс).
- 3) Истинносредний (true medium, TM) исход будет, когда фактическое целевое значение для данного примера *Средний* и модель на выходе выдаст *Средний*.
- 4) Истиннотрицательный (true negative, TN) исход, когда фактическое целевое значение для данного наблюдения *Низкий* и модель на выходе даёт *Низкий*.
- 5) Ложноположительный (false positive, FP), когда целевой класс для наблюдения *Низкий*, а модель для него сформирует выход *Высокий* (**ошибка II рода**).

Матрица классификации, содержащая значения показателей оценки эффективности и сравнения моделей 4-арной классификации

Модуль нейро-нечёткой сети на основе базы правил Y_1 (общее количество наблюдений в обучающем множестве – 15 000)						
Название алгоритма обучения	Название показателя					
	Общий показатель успеха (k_{OSR})	Общий показатель ошибки (k_{OVR})	Чувствительность ($k_{Se}^{(0)}$)	Специфичность ($k_{Sp}^{(0)}$)	Чувствительность ($k_{Se}^{(1)}$)	Специфичность ($k_{Sp}^{(1)}$)
Алгоритм конкурентного обучения	0,41	0,59	0,31	0,45	0,42	0,54
Имитация отжига	0,37	0,63	0,30	0,38	0,41	0,46
Градиентный алгоритм	0,26	0,74	0,21	0,21	0,35	0,30
Генетический алгоритм	0,11	0,89	0,08	0,09	0,19	0,14

- 6) Ложнонормальный (false rate, FR) исход, когда фактический целевой класс для данного наблюдения *Низкий*, а модель на выходе выдаст *Нормальный*.
- 7) Ложносредний (false medium, FM) исход, когда фактический целевой класс для данного наблюдения *Низкий*, а модель на выходе выдаст *Средний*.
- 8) Ложноотрицательный (false negative, FN), когда целевая переменная принимает значение *Высокий*, а на выходе модель выдаст *Низкий* (ошибка I рода).
- 9) Средневысокий (medium positive, MP) исход, когда фактическая целевая переменная принимает значение *Средний*, а на выходе модель выдаст *Высокий*.
- 10) Средненормальный (medium rate, MR) исход, когда фактическая целевая переменная принимает значение *Средний*, а на выходе модель выдаст *Нормальный*.
- 11) Среднегегативный (medium negative, MN) исход, когда фактическая целевая переменная принимает значение *Средний*, а на выходе модель выдаст *Низкий*.
- 12) Нормальновысокий (rate positive, RP) исход, когда фактическая целевая переменная принимает значение *Нормальный*, а на выходе модель выдаст *Высокий*.
- 13) Нормальносредний (rate medium, RM) исход, когда фактическая целевая переменная принимает значение *Нормальный*, а на выходе модель выдаст *Средний*.
- 14) Нормальногегативный (rate negative, RN) исход, когда фактическая целевая переменная принимает значение *Нормальный*, а на выходе модель выдаст *Низкий*.
- 15) Позитивнонормальный (positive rate, PR) исход, когда фактическая

целевая переменная принимает значение *Высокий*, а на выходе модель выдаст *Нормальный*.

- 16) Позитивносредний (positive medium, PM) исход, когда фактическая целевая переменная принимает значение *Высокий*, а на выходе модель выдаст *Средний*.

Введём в рассмотрение четыре показателя, позволяющие оценить точность 4-арной классификации модели.

Общий показатель успеха (overall success rate, OSR) – это число правильно классифицированных наблюдений, отнесённое к общему числу наблюдений (доля положительных примеров):

$$k_{OSR} = \frac{TP + TN + TR + TM}{\sum_{i=1}^{n \times n} k_i}, \quad (1)$$

где k_i – значение i -того предсказания, n – количество классов классификатора ($n = 4$).

Общий показатель ошибки (overall error rate, OVR) – это число ложно классифицированных наблюдений, отнесённое к общему числу наблюдений:

$$k_{OVR} = \frac{FP + FN + FR + FM + MR + MP + MN + PR + PM + RN + RM + RP}{\sum_{i=1}^{n \times n} k_i}. \quad (2)$$

Чувствительность определяется как отношение числа истинноположительных наблюдений к числу фактически положительных наблюдений:

$$k_{Se}^{(0)} = \frac{TP}{TP + FN + PM + PR}. \quad (3)$$

Дополнительно введём определение чувствительности для класса *Скорее высокий* (*Нормальный*), которое определяется как отношение числа *скорееположительных* наблюдений к числу фактически положительных наблюдений для данного класса:

$$k_{Se}^{(1)} = \frac{TM}{TM + MN + RN + RM}. \quad (4)$$

Специфичность определяется как отношение числа истинноотрицательных наблюдений к числу фактически отрицательных наблюдений:

$$k_{Sp}^{(0)} = \frac{TN}{TN + FP + FM + FR}. \quad (5)$$

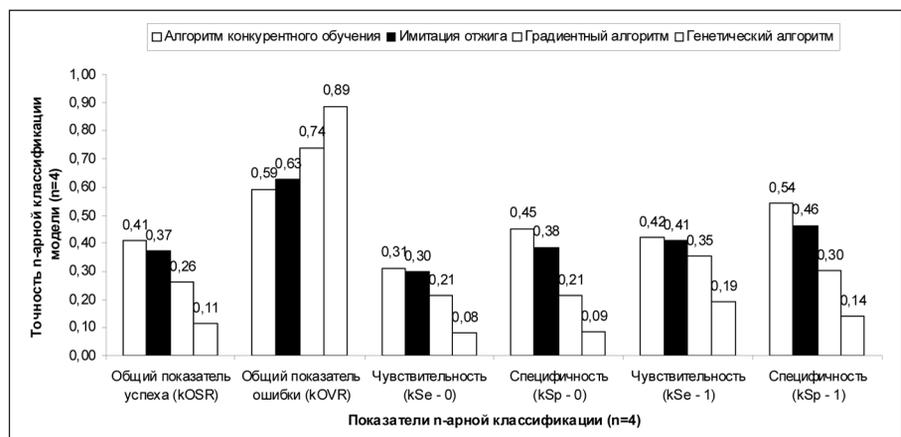


Рис. 2. Точность 4-арной классификации для модуля на основе БП Y_1

Таблица 5.

Матрица издержек для модуля на основе базы правил Y_1

		Предсказанный класс (модель)			
		Низкий	Средний	Нормальный	Высокий
Фактический класс	Низкий	0	0,7	0,8	1
	Средний	0,1	0	0,7	0,9
	Нормальный	0,2	0,1	0	0,7
	Высокий	1	0,2	0,1	0

Таблица 6.

Издержки обученной модели нейро-нечеткой сети

Название алгоритма обучения: <i>Алгоритм конкурентного обучения (competitive learning)</i>		
Тип ошибочной классификации	Вероятность появления класса	Издержки (costs)
Ложноположительные (FP)	0,031	0,031
Ложнонормальные (FR)	0,033	0,026
Ложносредние (FM)	0,032	0,022
Ложноотрицательные (FN)	0,120	0,120
Средненормальный (MR)	0,035	0,025
Средневысокий (MP)	0,034	0,031
Нормальновысокий (RP)	0,037	0,026
Среднеотрицательный (MN)	0,036	0,004
Позитивнонормальный (PR)	0,079	0,008
Позитивносредний (PM)	0,078	0,016
Нормальноотрицательный (RN)	0,038	0,008
Нормальносредний (RM)	0,037	0,004
Полные издержки C_{total}		0,319

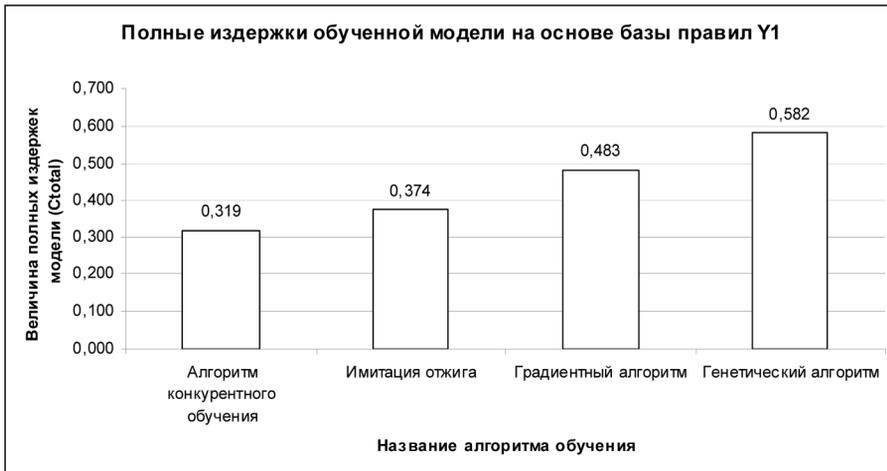


Рис. 3. Величина полных издержек обученной модели на основе БП Y_1

Аналогично, введём определение специфичности для класса *Скорее низкий (Средний)*, которое определяется как отношение числа *скорееотрицательных* наблюдений к числу фактически отрицательных наблюдений для данного класса:

$$k_{Sp}^{(1)} = \frac{TR}{TR + MR + MP + RP} \quad (6)$$

С учётом полученных результатов (табл. 3) выполним расчёт рассмотренных показателей оценки эффективности и сравнения моделей для случая $n = 4$ классификации (1-6) и сведём данные в таблицу 4 в виде матрицы классифи-

кации для модуля нейро-нечёткой сети (см. рис. 1) на основе БП Y_1 .

Применение алгоритма конкурентного обучения модуля нейро-нечёткой сети на основе базы правил Y_1 (рис. 2) даёт наибольшее значение общего показателя успеха (k_{OSR}), наибольшую чувствительность (k_{Se}) и специфичность (k_{Sp}), но при этом общий показатель ошибки (k_{OVR}) является самым минимальным.

При 4-арной классификации четыре типа ошибок – ложноположительные (FP), ложноотрицательные (FN), ложнонормальные (FR) и ложносредние (FM) – будут иметь различные

издержки, а четыре типа корректной классификации – истинноположительные (TP), истинноотрицательные (TN), истиннонормальные (TR) и истинносредние (TM) – будут давать различную прибыль (profit).

В случае четырёх классов издержки ошибок сведём в таблицу 5 в виде матрицы издержек, размером 4×4 , в которой диагональные элементы представляют четыре типа правильной классификации, недиагональные – четыре типа ошибок.

Далее вычислим издержки обученной модели на тестовом множестве (табл. 6), перемножив соответствующие количества ошибочно классифицированных наблюдений (табл. 3) на значения связанных с ними издержек (табл. 5) с учётом вероятности появления соответствующего класса, которая оценивается как отношение числа примеров соответствующего класса к общему числу примеров. На рисунке 3 видно, что использование *алгоритма конкурентного обучения* позволяет получить наименьшую величину полных издержек ($C_{total} = 0,319$) нейро-нечёткой модели на основе БП Y_1 , причём к большим издержкам ошибок классификации ведёт *ложноотрицательный* (FN = 0,120) тип ошибки (табл. 6).

Для нашего случая положительным исходом классификации являются следующие классы: *класс P_0* , связанный с *истинноположительной* классификацией (TP) и *класс P_1* , связанный со *скорееположительными* наблюдениями (TM). Отрицательным исходом классификации являются классы: *класс N_0* , связанный с *истинноотрицательной* классификацией (TN) и *класс N_1* , связанный со *скорееотрицательными* наблюдениями (TR).

Для выше обозначенных классов, связанных с положительным и отрицательным исходом классификации, введём соответствующие формулы расчёта издержек:

$$C_{P_0} = C_{FN} + C_{PM} + C_{PR}, \quad (7)$$

$$C_{P_1} = C_{RN} + C_{RM} + C_{MN}, \quad (8)$$

$$C_{N_0} = C_{FP} + C_{FM} + C_{FR}, \quad (9)$$

$$C_{N_1} = C_{MR} + C_{MP} + C_{RP}, \quad (10)$$

где C_{P_0} , C_{P_1} , C_{N_0} и C_{N_1} – издержки, связанные с ошибкой классификации – отнесения i -того наблюдения к классу P_0 , P_1 , N_0 и N_1 соответственно.

В таблице 7 приведены значения издержек положительной и отрицательной классификации модели,

Таблица 7. и

Издержки положительной и отрицательной классификации

Название алгоритма обучения модели нейро-нечёткой сети на основе базы правил Y_1	Издержки положительной классификации		Издержки отрицательной классификации		Полные издержки
	C_{P_0}	C_{P_1}	C_{N_0}	C_{N_1}	
Алгоритм конкурентного обучения	0,144	0,015	0,080	0,081	0,319
Имитация отжига	0,168	0,015	0,099	0,092	0,374
Градиентный алгоритм	0,200	0,014	0,134	0,136	0,483
Генетический алгоритм	0,235	0,016	0,182	0,149	0,582

$$P_i > \frac{1}{1 - (C_{P_{-1}}/C_{N_{-1}})} \quad (11)$$

На рисунке 5 приведена Lift-диаграмма, которая показывает, что для моделей, обученных с помощью алгоритмов конкурентного обучения и имитация отжига, Lift-кривые проходят выше диагональной прямой и тем самым обеспечивают *лифт*. Наиболее эффективной (с точки зрения **оптимизации издержек ошибок классификации**) является модель, обученная с помощью *алгоритма конкурентного обучения* (рис. 3), причём, объём выборки от 10% до 40% обеспечивает наибольшее значение откликов модели (почти в 1,5 раза). Классификаторы (рис. 5), проходящие ниже диагональной прямой, дают отрицательный лифт и потому применение моделей, обученных с помощью градиентного и генетического алгоритмов, с точки зрения оптимизации издержек бессмысленно.

На рисунке 6 представлено, что модель, обученная с помощью *алгоритма конкурентного обучения*, является наиболее эффективной и при объёме выборки от 15% до 40% обеспечивает наибольшее значение откликов модели.

Существует ещё одна разновидность Lift-диаграмм – Gain-диаграмма, которая применяется для сравнения **точности** моделей. По горизонтальной оси Gain-диаграммы откладывается размер выборки (в процентах), полученного из упорядоченного списка наблюдений. По вертикальной оси фиксируется значение чувствительности модели $k_{se}^{(0)}$ для класса P_0 (или $k_{se}^{(1)}$ для класса P_1). Gain-диаграммы позволяют определить оптимальное

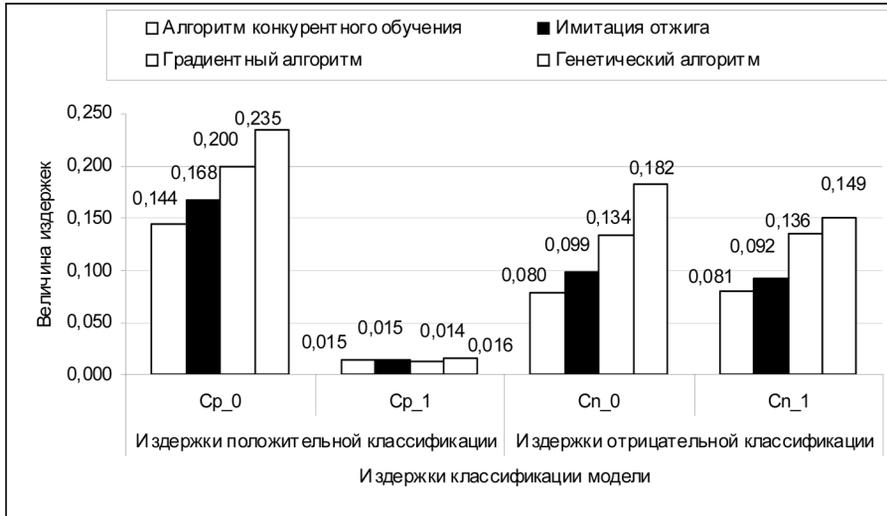


Рис. 4. Издержки положительной и отрицательной классификации

обученной с помощью различных алгоритмов обучения. Наименьшее распределение издержек даёт модель на основе алгоритма конкурентного обучения (рис. 4).

С помощью Lift-кривой определим оптимальный размер выборки нейро-нечёткой модели на основе БП Y_1 , который даст наилучшее соотношение издержек и прибыли и тем самым позволит минимизировать издержки II-рода (класс N_0), или максимизировать выгоду (класс P_0). Для расчёта коэффициента лифта для i -того на-

блюдения вычислим отношение доли положительных примеров в выборке, то есть k_i/K_i , к доле положительных примеров в исходном множестве n/N ,

$$L = \frac{k_i/K_i}{n/N}$$

то есть $L = \frac{k_i/K_i}{n/N}$. Для отнесения i -того наблюдения к классам P_0 и P_1 используем правило Байеса, преобразованное соответственно к следующему виду:

$$P_i > \frac{1}{1 - (C_{P_0}/C_{N_0})}$$

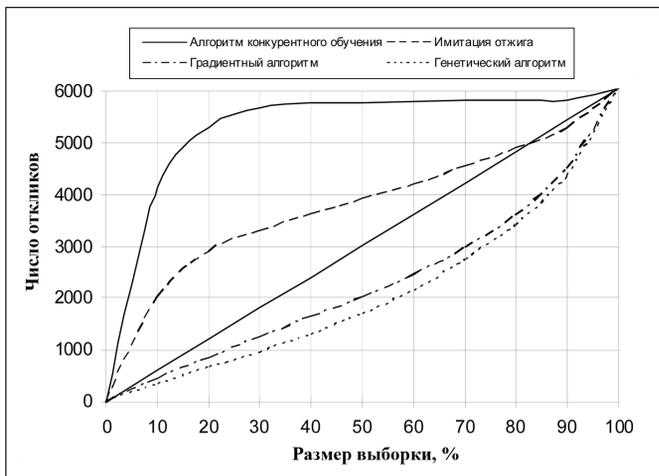


Рис. 5. Lift-диаграмма класса P_0

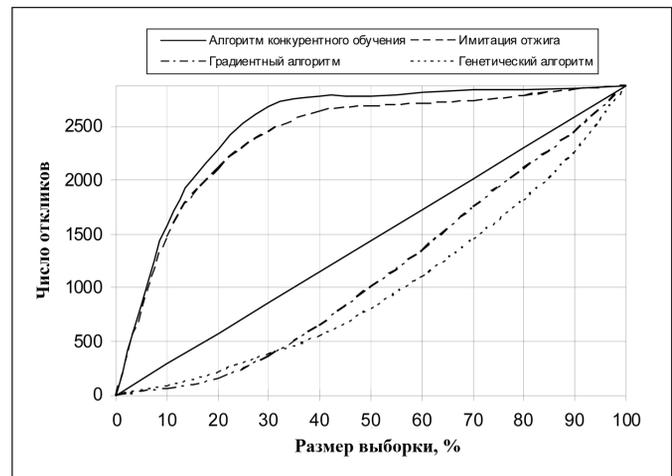


Рис. 6. Lift-диаграмма класса P_1

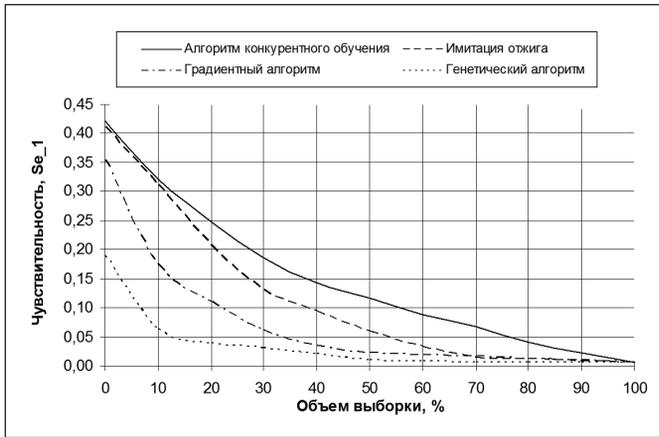


Рис. 7. Gain-диаграмма класса P₀

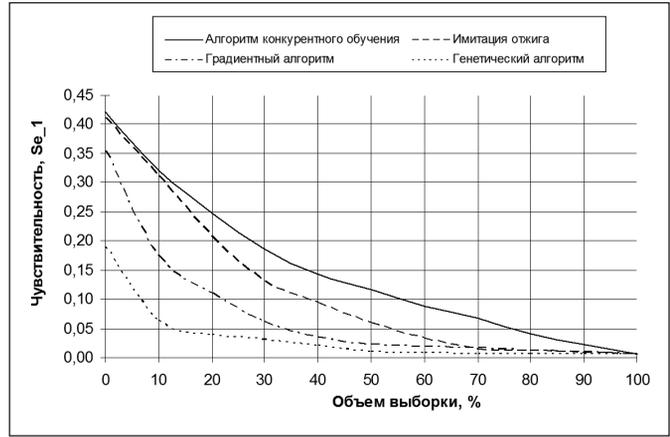


Рис. 8. Gain-диаграмма класса P₁

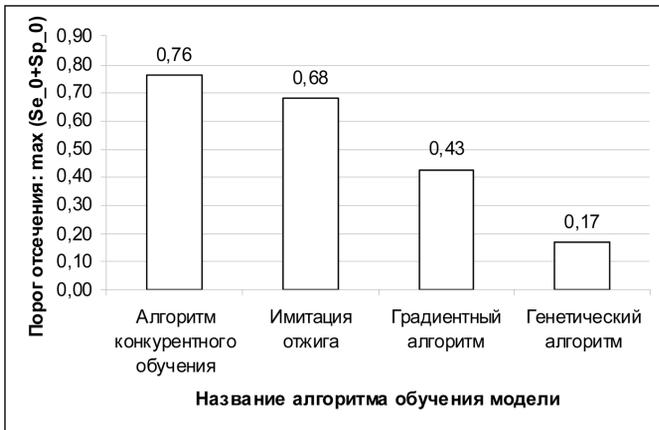


Рис. 9. Максим. порог класса P₀

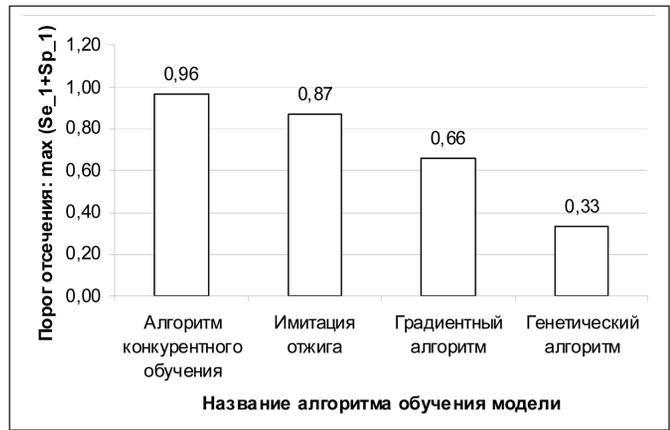


Рис. 10. Максим. порог класса P₁

соотношение между размером выборки и количеством ошибок классификатора, то есть числом случаев ложноотрицательной классификации. На рисунке 7 видно, что для класса P₀ наиболее точной является модель, обученная с помощью алгоритма конкурентного обучения, так как Gain-кривая этой модели проходит выше всех остальных кривых и, например, при выборке размером 20% обеспечивает чувствительность около 0,19, а модель, обученная с помощью

алгоритма имитация отжига, – только 0,10 (что почти в 2 раза меньше чувствительности первой модели). Для класса P₁ наиболее точной также является модель на основе алгоритма конкурентного обучения и, к примеру, при объеме выборке >50% обеспечивает чувствительность в 2 раза больше по сравнению с остальными моделями (рис. 8).

Дополнительно, для оценки эффективности моделей используется ещё одна графическая методика – ROC-ана-

лиз. Задачей данного метода является помочь аналитику выбрать точку отсеечения, то есть порога вероятности, разделяющего два класса. В нашем случае необходимо минимизировать ошибки II рода, поэтому необходимо использовать модель с высокой специфичностью (нужно снизить число ложноположительных исходов). Определим порог отсеечения модели, при котором обеспечивается максимальное количество как положительных, так и отрицательных правильно распоз-

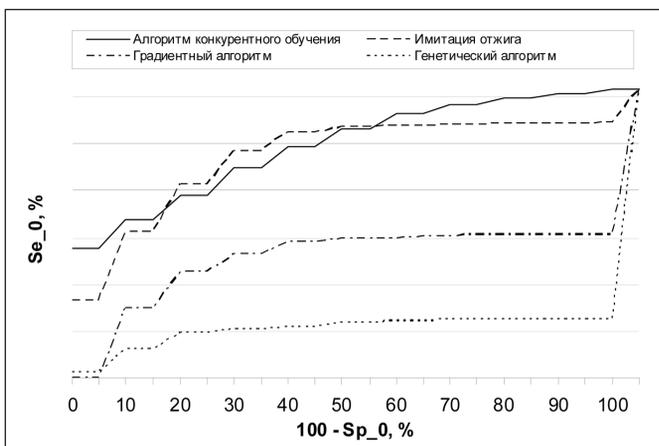


Рис. 11. ROC-кривые класса P₀

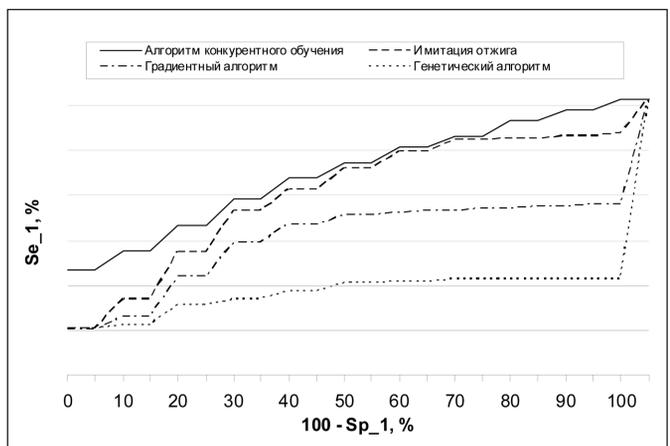


Рис. 12. ROC-кривые класса P₁

нанных примеров, то есть $\max(Se_0 + Sp_0)$ и $\max(Se_1 + Sp_1)$. На рисунках 9 и 10 показано, что модель на основе *алгоритма конкурентного обучения* обеспечивает максимальное значение порога отсекаения соответственно для классов P_0 и P_1 соответственно.

На рисунках 11 и 12 показано, что модели на основе алгоритмов конкурентного обучения и имитация отжига демонстрируют достаточно высокую точность ($> 0,7$), но наиболее лучшей прогностической силой обладает модель на основе *алгоритма конкурентного обучения*, так как площадь под её ROC-кривой превышает площадь под ROC-кривой второй модели. Однако при выборке объёмом на интервале от 20% до 50% наиболее точной является модель «Имитация отжига» (рис. 11), а две остальные модели, обученные с помощью градиентного и генетического алгоритмов, являются «беспольными» классификаторами, так как проходят ниже диагональной прямой (рис. 11 и 12).

3. Заключение

В результате проведенного исследования получено, что модель на основе алгоритма конкурентного обучения модуля нейро-нечеткой сети на основе БП Y_1 является наиболее эффективной, точной и обладает высокой специфичностью, что позволяет существенно снизить количество ошибок II рода. Аналогично рассмотренной методике была выполнена оценка адекватности моделей на основе БП Y_2 , выходом которой является интегральный показатель «Требования Министерства образования и науки», и на основе БП Y_3 , выходом которой является «Показатель эффективности деятельности библиотеки вуза», обученных с помощью генетического и градиентного алгоритмов соответственно.

Литература

1. Паклин, Н. Бизнес-аналитика: от данных к знаниям [Текст] / Н. Паклин, В. Орешков. – СПб.: Изд-во «Питер». – 2009. – 400 с.

2. Куровский, В.Л. Мягкий многомерный анализ информационных ресурсов ВУЗа / В.Л. Куровский, С.Г. Фомичёва, А.А. Попкова. – В кн.: Моделирование развития информационно-телекоммуникационных систем / Под ред. канд. техн. наук, д-ра экон. наук, проф. А. В. Бабкина. – СПб.: Изд-во «Синтез Бук». – 2009. – 384 с.

References

1. Paklin N. Business intelligences: from data to knowledge [Text] / N. Paklin, V. Oreshkov. – SPb.: Publishing house «Piter». – 2009. – 400 p.

2. Kurovsky V.L. Soft multidimensional analysis of information resources of HIGHER SCHOOL EDUCATION / V.L. Kurovsky, S.G. Fomitchyova, A.A. Popkova. – In book: Modeling of development of information and telecommunication systems / Under the editorship of Cand.Tech.Sci., PhD in Economics, Professor, A.V. Babkina. – SPb. : Publishing house "Synthesis Beech". – 2009. – 384 p.