

18. Кожухов, А. М. Оценка координат астероида на дискретном изображении [Текст] / А. М. Кожухов // Радиоэлектроника и молодежь в XXI веке: 15-й Юбилей. Междунар. молодеж. форум, 18–20 апреля 2011 г. : сб. матер. форума. Т. 3. – Харьков, 2011. – С. 213–214.
19. Закс, Ш. Теория статистических выводов [Текст]: пер. с англ. / Ш. Закс. – М.: Мир, 1975. – 776 с.
20. Бахвалов, Н. С. Численные методы [Текст] / Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков. – М.: Лаборатория базовых знаний, 2003. – 632 с.
21. Демидович, Б. П. Основы вычислительной математики. [Текст] / Б. П. Демидович, И. А. Марон. – М.: Наука, 1970. – 664 с.
22. Minor Planet Center. MPC/MPO/MPS Archive [Electronic resource] / Available at : http://www.minorplanetcenter.org/iau/ECS/MPCArchive/MPCArchive_TBL.html
23. Minor Planet Center. Numbered-Residuals Statistics For Observatory Codes [Electronic resource] / Available at: <http://www.minorplanetcenter.net/iau/special/residuals2.txt>
24. Вавилова, И. Б. Астроинформационный ресурс Украинской виртуальной обсерватории (УкрВО): объединенный архив данных наблюдений, научные задачи и программное обеспечение [Текст] / И. Б. Вавилова, Л. К. Пакуляк, В. Е. Саваневич и др. // Кинематика и физика небесных тел. – 2012. – Т. 28, № 2. – С. 59–80.

Запропоновано метод формування навчальних вибірок, який дозволяє охарактеризувати індивідуальну інформативність зразків відносно центрів і меж інтервалів ознак. Це дозволяє автоматизувати аналіз вибірки і її поділ на підвибірки, і, як наслідок, скоротити розмірність навчальних даних. Розроблено програмне забезпечення, що реалізує запропоновані показники та проведені експерименти з дослідження їхніх властивостей

Ключові слова: вибірка, відбір екземплярів, редукація даних, нейронна мережа, скорочення розмірності

Предложен метод формирования обучающих выборок, позволяющий охарактеризовать индивидуальную информативность экземпляров относительно центров и границ интервалов признаков. Это позволяет автоматизировать анализ выборки и её разделение на подвыборки, и, как следствие, сократить размерность обучающих данных. Разработано программное обеспечение, реализующее предложенные показатели, и проведены эксперименты по исследованию их свойств

Ключевые слова: выборка, отбор экземпляров, редукация данных, нейронная сеть, сокращение размерности

УДК 004.93

DOI: 10.15587/1729-4061.2014.28027

РАЗРАБОТКА МЕТОДА ЦЕНЗУРИРОВАНИЯ ВЫБОРОК ДЛЯ ПОСТРОЕНИЯ НЕЙРОМОДЕЛЕЙ

С. А. Субботин

Доктор технических наук, профессор
Кафедра программных средств
Запорожский национальный
технический университет
ул. Жуковского, 64,
г. Запорожье, Украина, 69063
E-mail: subbotin@zntu.edu.ua

1. Введение

Нейросетевые методы [1], являясь одним из наиболее мощных инструментальных средств вычислительного интеллекта, получили широкое распространение в задачах автоматизации принятия решений, где необходимо синтезировать модели принятия решений на основе набора прецедентов (экземпляров) – наблюдений за состоянием объекта или процесса в некоторый момент времени.

Как правило, процесс построения нейромодели является длительным по времени и высоко итеративным вследствие того, что время обучения и точность получаемой нейромодели существенно зависят от размерности и качества используемой обучающей выборки. Поэтому для повышения скорости построения и качества нейромодели по прецедентам необходимо сократить размерность

выборки, обеспечив при этом сохранение основных её свойств.

Сокращение размерности выборки возможно обеспечить путем выделения подмножества наиболее информативных признаков из исходного набора признаков, характеризующих экземпляры выборки, а также путем выделения подмножества наиболее ценных (информативных) экземпляров, позволяющих аппроксимировать межклассовые границы. Раздельная реализация данных процедур на основе различных методов приводит к существенным затратам времени, поскольку при таком подходе информация о выборке, получаемая в результате промежуточных вычислений в одной процедуре, как правило, не используется в другой. Поэтому целесообразно разработать метод сокращения размерности выборки за счет объединения процедур отбора признаков и экземпляров и использующий один набор промежуточных расчетов для обеих

процедур. Это позволит сократить время расчетов, а также за счет единообразного подхода к оценке значимости признаков и экземпляров позволит обеспечить более высокое качество формируемых выборок.

2. Анализ литературных данных и постановка проблемы

Пусть задана исходная выборка в виде набора прецедентов $\langle x, y \rangle$, где $x = \{x^s\}$, x^s – s -й экземпляр выборки, $x = \{x_j\}$, $x^s = \{x_j^s\}$, $x_j = \{x_j^s\}$, x_j^s – значение j -го диагностического признака x_j , характеризующее экземпляр x^s , $y = \{y^s\}$, y^s – значение выходного признака, сопоставленное экземпляру x^s , $s = 1, 2, \dots, S$, S – число экземпляров в исходной выборке, $j = 1, 2, \dots, N$, N – число диагностических признаков, характеризующих выборку.

Для заданной выборки прецедентов $\langle x, y \rangle$ задача синтеза нейромодели может быть представлена как задача нахождения

$$\langle F, w \rangle: y^{s*} = F(w, x^s), f(F, w, \langle x, y \rangle) \rightarrow \text{opt},$$

где структура модели F на практике, как правило, задается пользователем, а набор управляемых параметров w настраивается на основе обучающей выборки, f – пользовательский критерий, характеризующий качество аргумента относительно решаемой задачи, opt – оптимальное (желаемое или приемлемое) значение функционала f для решаемой задачи, y^{s*} – расчетное значение выходного признака для s -го экземпляра на выходе нейромодели.

Задача формирования подвыборки заданной выборки $\langle x, y \rangle$ заключается в нахождении такого набора $\langle x', y' \rangle$: $x' \in \{x^s\}$, $y' = \{y^s | x^s \in x'\}$, $S' < S$, $N' = N$, при котором $f(\langle x', y' \rangle, \langle x, y \rangle) \rightarrow \text{opt}$, где N' – число признаков в подвыборке, S' – число экземпляров в подвыборке.

Методы формирования выборок для построения моделей по прецедентам в [2, 3] разделяют на методы выбора прототипа (prototype selection methods) и методы построения прототипа (prototype construction methods). Здесь под прототипом подразумевается выделяемая подвыборка относительно исходной выборки.

Методы выбора прототипа не модифицируют, а только отбирают наиболее важные экземпляры из исходной выборки. Данные методы в зависимости от стратегии формирования решений разделяют на методы с добавлением (incremental methods) [4–6] (последовательно добавляют экземпляры из исходной выборки в формируемую подвыборку) и методы с удалением (decremental methods) [7, 8] (последовательно удаляют экземпляры из исходной выборки, получая в итоге подвыборку). Также отдельно выделяют методы фильтрации шума (noise filtering methods) [9, 10] (удаляют экземпляры, метки классов которых не совпадают с метками большинства соседних экземпляров), методы конденсации (condensation methods) [11, 12] (добавляют экземпляры из исходной выборки в формируемую подвыборку, если они несут новую информацию, но не добавляют, если они имеют те же метки классов, что и соседние с ними экземпляры) и методы на основе стохастического поиска [11, 13, 14] (случайным образом формируют подвыборку из исходной выборки, возможно перебирая некоторое множество вариантов

и отбирая наилучший из них). Общими недостатками данных методов являются высокая итеративность и длительность поиска, а также неопределенность выбора критериев качества получаемой подвыборки.

Методы построения прототипа на основе исходной выборки создают искусственные экземпляры, позволяющие описывать исходную выборку. Среди данных методов выделяют методы на основе кластер-анализа (cluster analysis based methods) [15, 16] (заменяют исходную выборку центрами её кластеров), методы преобразования сжатием (data squashing methods) [17] (заменяют экземпляры исходной выборки на получаемые на их основе искусственные прототипы, имеющие веса) и нейросетевые методы (neural network based methods) [18, 19] (на основе исходной выборки обучают нейронную сеть, из которой затем извлекают центры кластеров как экземпляры формируемой подвыборки). Общими недостатками данных методов является высокая итеративность и значительное время работы, а также неопределенность в задании параметров метода. Методы на основе кластер-анализа характеризуются такими недостатками, как неопределенность выбора числа кластеров, метрики и начальных параметров метода кластеризации и обучения. Методы преобразования сжатием формируют прототипы, которые сложно интерпретировать. Нейросетевые методы имеют такие недостатки, как сложность извлечения прототипов из нейромоделей, отсутствие гарантий получения приемлемой нейромодели в результате обучения, вариативность нейромоделей, влекущая за собой нестационарность выделяемых прототипов, ориентированность на конкретную модель, неопределенность задания начальных параметров модели и методов обучения.

Также выделяют комбинированные методы (combined methods) [3], которые сочетают формирование и отбор прототипов. Комбинированные методы имеют недостатки как методов выделения, так и методов построения прототипа.

3. Цель и задачи исследования

Известные методы формирования выборок, рассмотренные в предыдущем разделе, характеризуются высокой итеративностью и низкой скоростью работы, а также неопределенностью критериев оценки качества получаемой подвыборки.

Целью данной работы являлось повышение скорости процесса формирования и качества выделяемых обучающих выборок для построения нейромоделей по прецедентам.

Поскольку методы построения прототипа и связанные с ними комбинированные методы являются более медленными, чем методы выбора прототипа, то последние целесообразно выбрать как базис для решения задачи формирования выборок.

Для устранения недостатков, присущих рассмотренным выше методам, целесообразно формировать выборку без итеративного перебора комбинаций экземпляров – путем цензурирования исходной выборки, выбирая из неё определенную долю высокоценных экземпляров каждого класса, что позволит существенно сократить затраты времени на перебор решений. При этом предлагается определить показатели, по-

зволяющие оценивать индивидуальную информативность экземпляров с учетом их положения как относительно межклассовых границ, так и относительно центров псевдокластеров, что позволит выделять выборку неслучайным образом, оценивать и гарантировать высокое качество отбираемых подвыборок.

4. Метод цензурирования выборки

Разобьем пространство признаков на прямоугольные области, ограничив диапазон значений каждого признака минимальным и максимальным его значениями. Тогда проекции разбиения на ось признака позволят выделить интервалы значений признака для каждого из прямоугольных блоков. Интервалы могут формироваться как проекции кластеров, либо как регулярная решетка [20], либо на основе границ классов в одномерных проекциях выборки на оси признаков [21].

Тогда каждый такой интервал можно считать термом и оценить его значимость для принятия решений об отнесении экземпляра к кластеру с помощью веса k -го термина j -го признака для s -го экземпляра x^s относительно описания центра соответствующего интервала по формуле (1):

$$w_{C^s_{jk}} = \exp(-0.5(r_{jk} - l_{jk} - x^s_j)^2), \tag{1}$$

а также веса k -го термина j -го признака для s -го экземпляра x^s относительно описания межкластерных границ, определяемого по формуле (2):

$$w_{B^s_{jk}} = \exp(-(\min((r_{jk} - x^s_j), (x^s_j - l_{jk}))^2). \tag{2}$$

Тогда общую значимость термина k -го термина j -го признака для s -го экземпляра x^s относительно описания межкластерных границ можно оценить с помощью веса, определяемого по формуле:

$$w^s_{jk} = \max\{w_{C^s_{jk}}, w_{B^s_{jk}}\}.$$

Определив для каждого s -го экземпляра значимости термов, можно также определить веса термов для всей выборки:

$$w_{jk} = \frac{S_{jk}}{SK_{jk}}.$$

Зная оценки значимости термов, определим оценки информативности признаков по формуле (3):

$$w_j = \max_k \{w_{jk}\}, \tag{3}$$

либо по формуле (4):

$$w_j = \frac{1}{k_j} \sum_{k=1}^{k_j} w_{jk}. \tag{4}$$

Также возможно использовать индивидуальные оценки информативности признаков в диапазоне от нуля до единицы, определяемые с помощью показателей [21].

На основе оценок значимости термов и признаков можно определить оценки информативности для каждого s -го экземпляра выборки по формуле (5):

$$I(x^s) = \sum_{j=1}^N \left(w_j \sum_{k=1}^{k_j} w_{jk} w^s_{jk} \right) / \sum_{j=1}^N \left(w_j \sum_{k=1}^{k_j} w_{jk} \max_p \{w^p_{jk}\} \right), \tag{5}$$

Предложенные показатели (5) дают оценку индивидуальной информативности экземпляра x^s относительно исходной выборки в диапазоне от нуля до единицы. Чем больше будет значение соответствующих показателей, тем более ценным будет экземпляр и наоборот.

При необходимости оценки (5) их можно дополнительно нормировать так, чтобы они давали не абсолютное, а относительное значение значимости экземпляра в выборке (6):

$$I(x^s) = \frac{I(x^s) - \min_p \{I(x^p)\}}{\max_p \{I(x^p)\} - \min_p \{I(x^p)\}}. \tag{6}$$

В этом случае экземпляр с максимальной индивидуальной информативностью получит оценку, равную единице, а экземпляр с минимальной информативностью – оценку, равную нулю. Применение (6) может быть полезно в случае, когда необходимо упростить выбор порога для разделения выборки по соответствующему показателю информативности.

Предложенные показатели оценивания индивидуальной значимости экземпляров возможно использовать в процессе формирования подвыборок из заданной исходной выборки следующим образом: формировать обучающую подвыборку из не более чем S'/K экземпляров каждого класса исходной выборки с наибольшими значениями оценок индивидуальных информативностей.

5. Эксперименты по формированию выборок для решения задачи автоматизации диагностики хронического обструктивного бронхита

Для проведения экспериментов была разработана компьютерная программа, реализующая предложенный метод, являющаяся дополнением "Автоматизированной системы синтеза нейросетевых и нейро-нечётких моделей для неразрушающей диагностики и классификации образцов по признакам" (свидетельство о регистрации авторского права № 35431 от 21.10.2010).

Разработанное математическое обеспечение исследовалось при решении задачи диагностирования хронического обструктивного бронхита по экспериментально полученным данным клинических лабораторных исследований 205 пациентов [22], состояние которых характеризовалось 28 диагностическими признаками: x_1 – пол, x_2 – возраст, x_3 – содержание лейкоцитов, в том числе: x_4 – палочкоядерных, x_5 – сегментоядерных, x_6 – эозинофилов, x_7 – моноцитов, x_8 – лимфоцитов, x_9 – фагоцитарное число нейтрофилов, x_{10} – фагоцитарное число моноцитов, x_{11} – фагоцитарный индекс нейтрофилов, x_{12} – фагоцитарный индекс моноцитов,

x_{13} – индекс завершенности фагоцитоза нейтрофилов, x_{14} – индекс завершенности фагоцитоза моноцитов, x_{15} – НСТ-тест, отражающий способность нейтрофилов генерировать активные формы кислорода; x_{16} – процент НСТ положительных нейтрофилов, x_{17} – НСТ-стимулированный – отражает резервные возможности нейтрофилов генерировать активные формы кислорода в условиях их дополнительно стимулирования, x_{18} – процент НСТ-положительных нейтрофилов, реагирующих на стимулирование; x_{19} – количество тромбоцитов, x_{20} – количество зрелых тромбоцитов, x_{21} – количество активных тромбоцитов, x_{22} – уровень комплемента (C_{H50}), x_{23} – уровень циркулирующих иммунных комплексов (ЦИК), осаждаемых 3 % полиэтиленгликолем (ПЭГ), x_{24} – уровень ЦИК, осаждаемых 4 % ПЭГ, x_{25} – размер ЦИК, x_{26} – активность миелопероксидазы нейтрофилов, x_{27} – активность катионных белков нейтрофилов, x_{28} – уровень сосудистого фактора Виллебранда.

Каждому экземпляру выборки было сопоставлено значение целевого параметра – диагноза соответствующего пациента (1 – болен, 0 – не болен). Фрагмент исходной выборки представлен в табл. 1.

На основе исходной выборки были получены оценки информативности экземпляров, с помощью которых были выделены подвыборки исходной выборки, составлявшие 25 %, 50 %, 75 % и 100 % (для контроля) экземпляров с наибольшими значениями информативностей в каждом классе отдельно.

Далее для каждой подвыборки строились нейромодели на основе двухслойной нейросети прямого распространения сигнала, обучавшейся с помощью метода Левенберга-Марквардта [1]. Число входов сети определялось числом признаков N . Второй слой сети содержал один нейрон, выход которого являлся выходом сети. Число нейронов скрытого слоя составляло три. В качестве целевой функции использовалась среднеквадратическая ошибка. Параметры метода обучения задавались следующим образом: шаг обучения – 0,01, допустимое число итераций обучения (эпох) – 1000, приемлемое целевое значение функции ошибки сети – 10^{-6} .

По завершению процесса обучения нейромодели фиксировались его финальные характеристики: время обучения $t_{тр}$ и число затраченных итераций на обучение $er_{тр}$. После обучения каждая модель тестировалась отдельно на обучающей и на всей исходной выборках, для каждой из которых определялись ошибки $E_{тр}$ и E_{all} , соответственно. Ошибки представляли собой число экземпляров соответствующей выборки, для которых расчетное и фактическое значения целевого признака не совпали.

После отбора экземпляров рассчитанные значения весов признаков $\{w_j\}$ использовались как индивидуальные оценки информативности признаков. Для каждой отобранной подвыборки выделялось подмножество из 25 %, 50 % и 75 % признаков с наивысшими индивидуальными информативностями. Для каждой такой подвыборки строилась нейромодель, аналогичным рассмотренному выше способом.

Таблица 1

Фрагмент исходной выборки данных

Признак	Номер экземпляра										
	1	2	3	4	5	6	7	8	9	10	11
x_1	1	1	1	2	2	1	2	2	1	2	1
x_2	43	34	36	63	41	38	62	19	59	42	44
x_3	4,5	4,2	5,8	5,4	4,7	4,6	4,5	10	4,9	6,1	9
x_4	3	2	5	1	4	4	6	7	5	6	5
x_5	58	58	45	49	57	40	48	3	6	4	70
x_6	1	1	0	0	1	9	2	37	60	54	0
x_7	2	1	4	9	0	4	0	2	8	2	9
x_8	36	38	46	41	38	43	44	0	0	2	18
x_9	48	28	16	14	40	48	16	58	26	38	12
x_{10}	16	20	10	10	0	38	24	28	60	40	20
x_{11}	58	106	80	58	134	208	88	199	70	100	43
x_{12}	48	60	30	60	210	400	226	156	132	85	72
x_{13}	142	168	236	176	354	236	320	92	92	110	294
x_{14}	312	80	36	48	24	101	120	60	124	140	60
x_{15}	91	12	40	86	44	2	43	34	56	140	68
x_{16}	39	6	1	31	17	2	26	0,28	0,06	0,94	26
x_{17}	140	19	11	110	55	21	78	130	6	47	71
x_{18}	47	6	3	36	21	18	34	0,41	0,09	1,4	27
x_{19}	195	175	165	320	130	140	260	216	290	244	300
x_{20}	73	82	70	89	40	91	88	88	87	78	75
x_{21}	27	18	30	11	60	9	12	69	80	90	25
x_{22}	50	33	79	16	14	10	16	31	20	10	18
x_{23}	8	7	9	180	90	280	120	90	110	280	100
x_{24}	18	14	20	230	130	290	130	40	340	80	150
x_{25}	2	2	1	3	2	1	1	1,1	3,5	2,4	2
x_{26}	327	283	124	184	220	94	146	71	176	191	184
x_{27}	136	197	95	124	189	159	42	292	193	238	124
x_{28}	124	71	87	98	164,45	87,65	197,16	184	176	185	144,71
y	0	0	0	1	1	0	1	1	1	1	1

Фрагмент результатов проведенных экспериментов представлен в табл. 2.

Из табл. 2 следует, что применение предложенного метода определения значимости экземпляров позволяет на практике выделять из исходной выборки подвыборки меньшего объема, достаточные для построения нейромодели с требуемой точностью, сокращая время на построение моделей.

Как видно из табл. 2, с увеличением числа экземпляров S в формируемой выборке в целом наблюдается рост точности (снижение ошибок как для сформированной обучающей, так и для полной выборок), времени обучения и числа итераций обучения и, наоборот. При этом существенное сокращение объема выборки до 25 % от исходного приводит к понижению точности. Это можно объяснить тем, что в выборку малого объема могут не войти критически важные для описания разделения классов экземпляры.

Таблица 2

Фрагмент экспериментальных результатов

Характеристики размерности выборки данных		t_{tr}, s	n_{tr}	$E_{tr}, \%$	$E_{all}, \%$
S	N				
25 %	25 %	3,6	643	0	2,9
25 %	50 %	5,2	734	0	2,9
25 %	75 %	7,4	768	0	2,9
25 %	100 %	9,1	983	0	2,9
50 %	25 %	5,2	732	0	1,95
50 %	50 %	9,7	893	0	1,95
50 %	75 %	14,3	922	0	1,95
50 %	100 %	18,7	1000	0	1,95
75 %	25 %	7,5	731	0	0,49
75 %	50 %	14,9	720	0	0,49
75 %	75 %	23,4	881	0	0,49
75 %	100 %	28,1	954	0	0,49
100 %	25 %	9,5	793	0	0
100 %	50 %	18,3	931	0	0
100 %	75 %	27,6	982	0	0
100 %	100 %	37,2	1000	0	0

Даже небольшое сокращение объема исходной выборки на 25 % (до 75 % от исходного объема) позволило получить приемлемую точность и сокращение времени обучения более, чем в 1,32 раза. Сокращение объема исходной выборки в два раза (до 50 % от исходного объема) позволило получить выигрыш в скорости в 1,99 раза. Это подтверждает целесообразность применения предлагаемого математического обеспечения при построении нейросетевых моделей по прецедентам.

Достоинство показателей, предложенных в данной работе, заключается в отсутствии необходимости расчета расстояний между экземплярами, а недостаток – в необходимости разбиения пространства признаков. Однако данный недостаток можно рассматривать и как преимущество в случае больших выборок: если использовать простое с вычислительной точки зрения разбиение (например, регулярную решетку) и знать минимальные и максимальные значения каждого признака, то вычислительные затраты предложенных показателей окажутся меньше по сравнению с использованием комплекса [21].

Как видно из табл. 2, существенное уменьшение числа признаков N в сформированной выборке не приводит к снижению точности, следовательно, для рассматриваемой задачи можно выбрать небольшое подмножество признаков, обеспечивающее хорошую точность.

Эксперименты также обнаружили, что наиболее индивидуально информативные признаки x_{10} , x_{12} , x_{15} ,

x_{17} , x_{20} – x_{23} , а наименее информативные признаки x_4 – x_6 , x_8 , x_9 , x_{16} , x_{18} , x_{19} , x_{28} . Полученные результаты объясняются тем, что в перманентный воспалительный процесс в легочной ткани вовлечены гранулоциты, присутствуют активация гранулоцитов, мононуклеаров и тромбоцитов, потребление комплемента, рост иммунных комплексов и фактора Виллебранда. Полученная модель позволяет выделить ведущие звенья, поддерживающие воспалительный процесс. Это активация метаболической активности нейтрофилов и моноцитов, включая генерацию активных форм кислорода, тромбоцитопению и активацию тромбоцитов, потребление комплемента и повышенный уровень ЦИК. Следовательно, предложенный метод позволяет выделить информативные признаки периферической крови, отражающие интенсивность воспалительного процесса и его перманентность в бронхо-легочной ткани при хроническом обструктивном бронхите.

6. Выводы

Решена актуальная задача разработки математического обеспечения для автоматизации формирования выборок при построении диагностических и распознающих нейромоделей по прецедентам, позволяющего повысить скорость процесса формирования и качество выделяемых обучающих выборок.

Научная новизна полученных результатов состоит в том, что впервые предложен метод формирования обучающих выборок, который для заданной исходной выборки прецедентов и заданного разбиения пространства признаков определяет веса, характеризующие полезность термов и признаков. С учетом значений весов метод позволяет охарактеризовать индивидуальную абсолютную и относительную информативность экземпляров относительно центров и границ интервалов признаков. Это позволяет автоматизировать анализ выборки и её разделение на подвыборки, и, как следствие, сократить размерность обучающих данных. В свою очередь, это позволяет сократить время и обеспечить приемлемую точность обучения нейромоделей.

Практическая ценность полученных результатов состоит в том, что разработано программное обеспечение, реализующее предложенные показатели, а также проведены эксперименты по исследованию их свойств при решении задачи диагностирования хронического обструктивного бронхита, результаты которых позволяют рекомендовать предложенные показатели для использования на практике, а также определять эффективные условия применения предложенных показателей.

Литература

- Engelbrecht, A. Computational intelligence: an introduction [Text] / A. Engelbrecht. – Sidney : John Wiley & Sons, 2007. – 597 p. doi: 10.1002/9780470512517
- Jankowski, N. Comparison of instance selection algorithms I. Algorithms survey [Text] / N. Jankowski, M. Grochowski // Artificial Intelligence and Soft Computing : 7th International Conference ICAISC-2004, Zakopane, 7–11 June, 2004 : proceedings. – Berlin : Springer, 2004. – P. 598–603. – (Lecture Notes in Computer Science, Vol. 3070. doi:10.1007/978-3-540-24844-6_90
- Reinartz, T. A unifying view on instance selection [Text] / T. Reinartz // Data Mining and Knowledge Discovery. – 2002. – № 6. – P. 191–210. doi:10.1023/A:1014047731786

4. Hart, P. E. The condensed nearest neighbor rule [Text] / P. E. Hart // IEEE Transactions on Information Theory. – 1968. – Vol. 14. – P. 515–516. doi:10.1109/TIT.1968.1054155
5. Aha, D. W. Instance-based learning algorithms [Text] / D. W. Aha, D. Kibler, M. K. Albert // Machine Learning. – 1991. – № 6. – P. 37–66. doi:10.1023/A:1022689900470
6. Brighton, H. Advances in instance selection for instancebased learning algorithms [Text] / H. Brighton, C. Mellish // Data Mining and Knowledge Discovery. – 2002. – № 6. – P. 153–172. doi:10.1023/A:1014043630878
7. Wilson, D. R. Instance pruning techniques [Text] / D. R. Wilson, T. R. Martinez // Machine Learning : Fourteenth International Conference ICML–1997, Nashville, 8–12 July 1997 : proceedings. – Burlington : Morgan Kaufmann, 1997. – P. 403–411.
8. Kibbler, D. Learning representative exemplars of concepts: an initial case of study [Text] / D. Kibbler, D. W. Aha // Machine Learning : 4th International Workshop, Irvine, 22–25 June 1987 : proceedings. – Burlington : Morgan Kaufmann, 1987. – P. 24–30. doi:10.1016/b978-0-934613-41-5.50006-4
9. Gates, G. The reduced nearest neighbor rule [Text] / G. Gates // IEEE Transactions on Information Theory. – 1972. – Vol. 18, Issue 3. – P. 431–433. doi:10.1109/TIT.1972.1054809
10. Wilson, D. L. Asymptotic properties of nearest neighbor rules using edited data [Text] / D. L. Wilson // IEEE Transactions on Systems, Man, Cybernetics. – 1972. – Vol. 2, Issue 3. – P. 408–421. doi:10.1109/TSMC.1972.4309137
11. Wilson, D. R. Reduction techniques for instancebased learning algorithms [Text] / D. R. Wilson, T. R. Martinez // Machine Learning. – 2000. – Vol. 38, Issue 3. – P. 257–286. doi:10.1023/A:1007626913721
12. Ritter, G. L. An algorithm for a selective nearest neighbor decision rule [Text] / G. L. Ritter, H. B. Woodruff, S. R. Lowry, T. L. Isenhour // IEEE Transactions on Information Theory. – 1975. – Vol. 21, Issue 6. – P. 665–669. doi:10.1109/TIT.1975.1055464
13. Li, X. Data reduction via adaptive sampling [Text] / X. Li // Communications in Information and Systems. – 2002. – Vol. 2, Issue 1. – P. 5–38. doi:10.4310/cis.2002.v2.n1.a3
14. Domingo, C. Adaptive sampling methods for scaling up knowledge discovery algorithms [Text] / C. Domingo, R. Gavaldà, O. Watanabe // Discovery Science : Second International Conference, DS'99 Tokyo, 1999 : proceedings. – Berlin : Springer, 1999. – P. 172–183. doi:10.1007/3-540-46846-3_16
15. Li, B. Support cluster machine [Text] / B. Li, M. Chi, J. Fan, X. Xue // Machine Learning : 24th International Conference, Corvallis, 20–24 June 2007 : proceedings. – New York, 2007. – P. 505–512. doi:10.1145/1273496.1273560
16. Evans, R. Clustering for classification: using standard clustering methods to summarise datasets with minimal loss of classification accuracy [Text] / R. Evans. – Saarbrücken: VDM Verlag, 2008. – 108 p.
17. Madigan, D. Likelihood-based data squashing: a modeling approach to instance construction [Text] / D. Madigan, N. Raghavan, W. DuMouchel, M. Nason, C. Posse, G. Ridgeway // Data Mining and Knowledge Discovery. – 2002. – Vol. 6, Issue 2. – P. 173–190. doi:10.1023/A:1014095614948
18. Kohonen, T. Learning vector quantization [Text] / T. Kohonen // Neural Networks. – 1988. – Vol. 1. – P. 303. doi: 10.1016/0893-6080(88)90334-6
19. Sane, S. S. A Novel supervised instance selection algorithm [Text] / S. S. Sane, A. A. Ghatol // International Journal of Business Intelligence and Data Mining. – 2007. – Vol. 2, Issue 4. – P. 471–495. doi:10.1504/IJBIDM.2007.016384
20. Subbotin, S. The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition [Text] / S. Subbotin // Optical Memory and Neural Networks (Information Optics). – 2013. – Vol. 22, Issue 2. – P. 97–103. doi: 10.3103/s1060992x13020082
21. Subbotin, S. A. Methods of sampling based on exhaustive and evolutionary search [Text] / S. A. Subbotin // Automatic Control and Computer Sciences. – 2013. – Vol. 47, Issue 3. – P. 113–121. doi: 10.3103/s0146411613030073
22. Колісник, Н. В. Моделювання імунопатогенезу хронічного обструктивного бронхіту за допомогою нейромереж [Текст] : зб. матер. / Н. В. Колісник, С. О. Суботин // Сучасні проблеми біології, екології та хімії : II Міжнародна конференція, Запоріжжя : ЗНУ, 2009. – С. 124–125.