DOI: 10.18721/JPM.10206
УДК 004.032.26

ПРИМЕНЕНИЕ АЛГОРИТМОВ КЛАСТЕРИЗАЦИИ К ФОРМИРОВАНИЮ ПРЕДСТАВИТЕЛЬСКОЙ ВЫБОРКИ ДЛЯ ОБУЧЕНИЯ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА

А.А. Пастухов, А.А. Прокофьев

Национальный исследовательский университет «МИЭТ», Москва, Российская Федерация

Рассмотрен вопрос эффективного формирования представительской выборки для обучения нейронной сети многослойный персептрон. Предложен подход, основанный на применении кластеризации, позволяющий увеличить энтропию обучающего множества. Рассмотрены различные алгоритмы кластеризации для формирования педставительской выборки. На базе алгоритмов проведена кластеризация факторных пространств различной размерности и сформированы представительские выборки. Синтезирована и обучена нейронная сеть многослойный персептрон на множествах, сформированных с использованием и без использования кластеризации. Проведен сравнительный анализ эффективности алгоритмов кластеризации применительно к задаче формирования представительской выборки.

Ключевые слова: нейронная сеть; алгоритм кластеризации; представительская выборка; многослойный персептрон

Ссылка при цитировании: Пастухов А.А., Прокофьев А.А. Применение алгоритмов кластеризации к формированию представительской выборки для обучения многослойного персептрона // Научно-технические ведомости СПБГПУ. Физико-математические науки. Т. 10. № 2. С. 58–68. DOI: 10.18721/JPM.10206

CLUSTERING ALGORITHMS APPLICATION TO THE REPRESENTATIVE SAMPLE FORMATION IN THE TRAINING OF THE MULTILAYER PERCEPTRON

A.A. Pastukhov, A.A. Prokofiev

National Research University of Electronic Technology, Moscow, Russian Federation

In this paper, we have considered the problem of effective forming the representative sample for training the neural network of the multilayer perceptron (MLP) type. An approach based on the use of clustering that allowed to increase the entropy of the training set was put forward. Various clustering algorithms were examined in order to form the representative sample. The algorithm-based clustering of factor spaces of various dimensions was carried out, and a representative sample was formed. To verify our approach we synthesized the MLP neural network and trained it. The training technique was performed with the sets formed both with and without clustering. A comparative analysis of the effectiveness of clustering algorithms was carried out in relation to the problem of representative sample formation.

Key words: neural network; clustering algorithm; representative sample; multilayer perceptron

Citation: A.A. Pastukhov, A.A. Prokofiev. Clustering algorithms application to the representative sample formation in the training of the multilayer perceptron, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 10(2) (2017) 58–68. DOI: 10.18721/JPM.10206

Введение

Для обучения нейронной сети типа «многослойный персептрон» (MLP) необходим этап предобработки данных еще до применения алгоритма обратного распространения ошибки. В большинстве опубликованных работ по применению нейронных сетей методика предобработки сводится к нормализации, масштабированию, а также начальной инициализации весов.

Данные действия, несомненно, необходимы, но их нельзя считать достаточными. При небольшой размерности факторного пространства следует учитывать специфику распределения исходных данных для эффективного обучения нейронной сети. При большом же количестве факторов эта задача существенно усложняется. В таком случае целесообразно применять кластеризацию для формирования обучающего множества из примеров признаков, наиболее уникальных по совокупности.

Существует большое количество алгоритмов кластеризации, но все их можно условно разделить на две группы: четкие и нечеткие. В свою очередь среди четких методов можно выделить две подгруппы: иерархические и неиерархические методы [1].

Отдельным классом следует также выделить алгоритмы кластеризации на основе нейронных сетей, которые нашли широкое применение в различных областях.

Так например, в работе [2] исследуется вопрос кластеризации данных на основе марковского алгоритма, а также самоорганизующихся растущих нейронных сетей. В работе [3] проводится сравнение алгоритмов кластеризации *k*-средних (*k*-means) и по плотности (DBSCAN – Density-Based Clustering of Applications with Noise) на случайной выборке, а также приводятся оценки эффективности этих алгоритмов на основании индекса Дэвис – Боулдина. Работа [4] посвящена исследованию кластеризации текстовых документов для создания автоматических рубрикаторов с использованием метрики Евклида – Махаланобиса, а в работе [5] изучены различные алгоритмы нейросетевой ассоциативной памяти для создания памяти антропоморфного робота.

В данной работе исследованы три метода четкой и один метод нечеткой кластеризации в задаче формирования представительской выборки для обучения нейронной сети и проанализированы эффективности этих алгоритмов с точек зрения прироста энтропии обучающего множества и повышения качества обучения нейронной сети типа MLP. Кроме того, проведен анализ изменений энтропии обучающего множества и среднеквадратичной ошибки обучения при использовании указанных алгоритмов.

Класс алгоритмов четкой кластеризации представлен в данной работе наиболее распространенным и простым в реализации алгоритмом *k*-means [3], самоорганизующимися картами Кохонена [6] (карты рассмотрены ранее в работе [7]), а также алгоритмом кластеризации, основанном на построении иерархического дерева кластеров [8].

Среди класса нечетких методов следует выделить алгоритм нечетких *c*-средних (*c*-means) [1] – базовый для большого количества других алгоритмов данного класса, который имеет множество программных реализаций (например, FCM-алгоритм (Fuzzy *C*-Means), реализованный в пакете MatLab).

Постановка задачи

Как известно, обучение нейронной сети производится на трех подмножествах факторного пространства: обучающем, проверочном и тестовом. Вместе они формируют представительскую выборку для обучения нейронной сети. Обучающее множество используется для настройки свободных параметров сети, проверочное – для контроля эффективности переобучения, тестовое – для независимого тестирования уже

обученной нейронной сети. Одним из элементов качественного обучения является формирование обучающего множества из элементов факторного пространства, наиболее уникальных по совокупности признаков. Как было показано в работе [7], это достижимо применением кластеризации факторного пространства и выбора для обучающего множества представителей из каждого кластера.

Для достижения этой цели необходимо выбрать наиболее подходящий алгоритм кластеризации, удовлетворяющий определенным критериям.

Постановка задачи. *Пусть*

$$X = \{X^1, ..., X^M, Y^1, ..., Y^M\}$$

– факторное пространство,
где $X^i = \{x_1, x_2, x_3, x_4\}$, $Y^i = \{y(X^i)\}$; M *– количество векторов в факторном пространстве.*

Требуется выделить среди рассматриваемых тот алгоритм кластеризации, который позволяет найти разбиение факторного пространства на три множества (T *– обучающее,* V *– проверочное и* E *– тестовое*) *и для которого выполняются условия:*

$$H_0(T) < H(T) \le H_{\max}(T), \qquad (1)$$

$$S_T \text{ принимает минимальное значение.} \qquad (2)$$

Здесь $H(T)$ – энтропия обучающего множества с использованием кластеризации; $H_0(T)$ – энтропия обучающего множества для случайного разбиения факторного пространства на представительскую выборку; $H_{\max}(T) = \log_2 N_t$ – максимальная энтропия этого множества (N_t – размер обучающего множества, составляющего 80 % от факторного пространства), S_T – среднеквадратичная ошибка обучающего множества.

Исследование алгоритмов кластеризации

Как было отмечено выше, исследование проведено на четырех алгоритмах кластеризации: *k*-means, *c*-means, самоорганизующиеся карты Кохонена и иерархический метод. Во всех экспериментах количество кластеров было выбрано равным 80 % от объема факторного пространства.

На первом этапе исследования оценивался прирост энтропии обучающего множества, которое состоит из данных, сформированных с применением алгоритмов, указанных выше. Расчет энтропии производился по формуле Шеннона [9]:

$$H(x) = -\sum_{i=1}^{n} p_i \log_2 p_i, \qquad (3)$$

где p_i – вероятность выбора элемента из кластера.

На втором этапе нейронная сеть типа многослойный персептрон с архитектурой 4-4-1 была обучена на данных, сформированных с применением алгоритмов кластеризации.

Факторное пространство включает четыре входных параметра, сформированных случайным образом, и один выходной параметр – отклик.

Связь между входными и выходными параметрами задается нелинейной функцией

$$y = e^{x_1} + e^{x_2} + 2e^{x_3} - 3e^{x_4},$$

где x_1, x_2, x_3, x_4 – соответствующие входные параметры, y – выходной параметр.

В вектор выходных сигналов внесен шум, который описан случайной величиной, распределенной по нормальному закону с дисперсией 0,02. Обучение нейронной сети производилось для функций Neural Network Toolbox (далее NNtoolbox) пакета MatLab. Параметры обучения приведены в табл. 1.

На рис. 1 приведены результаты обучения нейронной сети на данных, сформированных случайным образом (без применения кластеризации), с целью оценки эффективности (разница между среднеквадратичной ошибкой обучающего/проверочного и тестового множеств) алгоритмов.

Здесь и далее приведены результаты обучения нейронной сети на факторном пространстве, включающем 200 обучающих векторов, что не нарушает общности, так как далее будет показано, что прирост энтропии обучающего множества не зависит от количества составляющих его элементов.

Процедура обучения производилась десять раз для каждого случая. Начальная инициализация весов производилась для

Таблица 1

Параметры обучения нейронной сети

Название параметра	Авторы параметра или его назначение	Наименование параметра в MatLab
Алгоритм оптимизации	Левенберга – Марквардта	TRAINLM
Функция адаптации	Градиентный спуск	LEARNGD
Критерий оптимизации	Среднеквадратичная ошибка	MSE
Начальная инициализация свободных параметров	Нгуен – Видроу	INITNW

функций NNtoolbox по алгоритму Нгуена – Видроу [10] при каждой процедуре обучения. В качестве результата приводится наиболее удачная попытка обучения с точки зрения минимума среднеквадратичной ошибки обучения.

Среднеквадратичная ошибка обучения вычисляется по следующей формуле:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2, \qquad (4)$$

где x_i, y_i – фактический и ожидаемый результаты обучения на обучающем векторе i, соответственно.

В табл. 2 и 3 приведены результаты расчетов энтропии с применением алгоритма SOM (Вариант 1, $H(T) = 0$), а также времени обучения нейронной сети T_1 для случая обучения на данных, сформированных без использования кластеризации ($T_2 = 0$).

Среднеквадратичная ошибка обучения для данного случая составляет 0,31462.

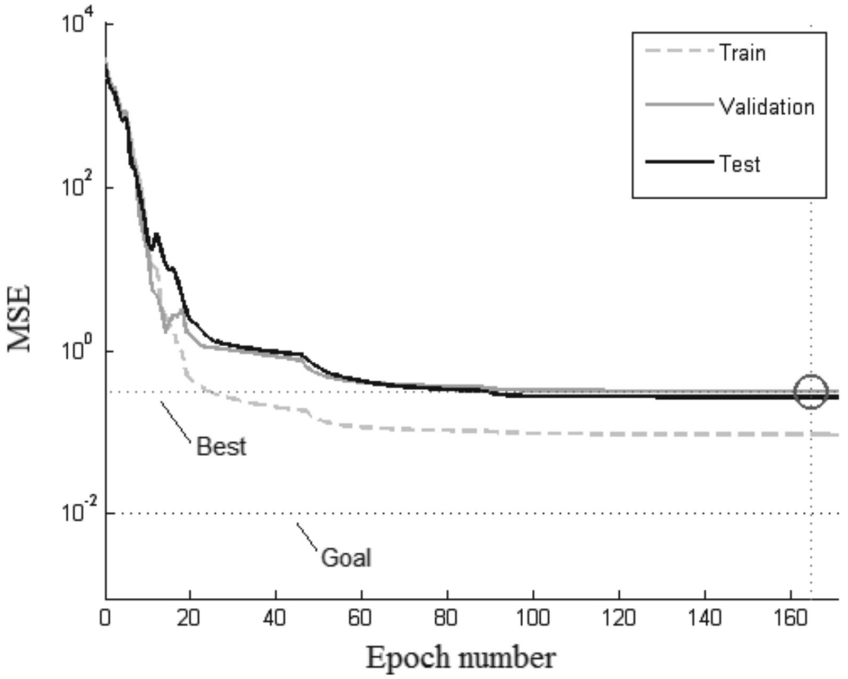

Рис. 1. Результаты обучения нейронной сети типа MLP на данных, сформированных без использования кластеризации:
MSE – среднеквадратичная ошибка обучения; Epoch number – текущая эпоха; представлено поведение ошибок для обучающего (Train), проверочного (Validation) и тестового (Test) множеств; Goal, Best – целевое и наилучшее значения ошибки, последнее достигнуто для проверочного множества

Таблица 2

Результаты расчета энтропии для факторных пространств различной размерности при двух вариантах применения алгоритма SOM

N	Энтропия, бит		
	Варианты 1 и 2		Вариант 2
	$H_{max}(T)$	$H_0(T)$	$H(T)$
100	6,32	6,05	6,19
200	7,32	7,02	7,15
300	7,91	7,67	7,79
400	8,32	8,07	8,17
500	8,64	8,38	8,50
600	8,91	8,66	8,77
700	9,13	8,86	8,97
800	9,32	9,05	9,17
900	9,49	9,16	9,31
1000	9,64	9,29	9,45

Обозначения: N – количество элементов факторного пространства; H_{max} (T) – максимальная величина энтропии обучающего множества; $H(T)$, $H_0(T)$ – величины энтропии этого множества с использованием кластеризации (Вариант 2) и для случайного разбиения факторного пространства на представительскую выборку (Вариант 1, когда $H(T) = 0$), соответственно.

Самоорганизующиеся карты Кохонена (SOM). Указанные карты [6, 11] представляют класс нейронных сетей с обучением без учителя. Данный класс относится к алгоритмам неиерархической кластеризации.

Самоорганизующиеся карты просты в реализации и позволяют гарантированно распределять данные по заданному числу кластеров после прохождения по слоям карты. Кроме того, алгоритм способен самостоятельно определять центры кластеров благодаря самоорганизации.

Алгоритм обучения SOM сводится к минимизации разности между входами нейронов соответствующего слоя и весовыми коэффициентами выходов этого нейрона:

$$\omega_i(t) = \omega_i(t-1) + \alpha[y_i^{n-1}(t) - \omega_i(t-1)], \quad (5)$$

где y_i^{n-1} – выход нейрона предыдущего слоя, соответствующий входу нейрона текущего слоя; ω_i – весовой коэффициент нейрона i; t – номер эпохи обучения; α – коэффициент скорости обучения (в простейшим случае $\alpha \in [0; 1]$, $\alpha = \text{const}$), $n = 1, 2, ..., N$ – слои карты, $i = 1, 2, ..., M$ – номера нейронов текущего слоя.

Далее приведены результаты расчетов и обучения сети MLP на данных, сформированных с помощью самоорганизующихся карт Кохонена (SOM). Подробный расчет приведен в работе [7].

Результаты расчетов для факторных пространств с количеством векторов от 100 до 1000 приведены в табл. 2 и 3.

Среднеквадратичная ошибка обучения составляет 0,11601. Анализ результатов, представленных в табл. 2 и 3, позволяет сделать вывод, что для любого N выполняются условия (1) и (2). Значение энтропии для случая с использованием кластеризации (Вариант 2) лежит между значениями $H_0(T)$ и $H_{max}(T)$ для всех N. Временные затраты на обучение самоорганизующейся карты Кохонена (см. табл. 3, данные SOM) растут практически линейно. Однако время обучения многослойного персептрона T_2 растет медленнее, чем время обучения самоорганизующейся карты Кохонена T_1.

Результаты обучения нейронной сети с использованием самоорганизующихся карт Кохонена приведены на рис. 2.

Наилучшая производительность нейронной сети (минимальная величина среднеквадратичной ошибки (MSE)) в данном случае составляет 0,11601, что значительно меньше аналогичного значения для случая, когда кластеризация не используется (см. рис. 1). Кроме того, разница между проверочным и тестовым множествами в случае использования кластеризации значительно меньше.

Алгоритм *k*-means. Данный алгоритм [12] также относится к классу неиерархической кластеризации и всегда распределяет данные по указанному количеству кластеров. Таким образом, в случае его применения в обучающее множество включено 80 % векторов, в тестовое – 10 %, в проверочное – оставшиеся 10 % векторов.

Аналогично данным предыдущего раздела, был рассчитан прирост энтропии обучающего множества с использованием кла-

Таблица 3

Зависимость промежутков времени, затраченных при использовании различных алгоритмов, от размера факторного пространства

N	Затраченное время, с				
	T_1	T_2			
		SOM	*k*-means	*c*-means	Hierarchical
100	3	1	0,1	1	0,093
200	5	2	0,1	2	0,101
300	7	4	0,1	3	0,103
400	10	3	0,3	4	0,108
500	15	3	0,4	5	0,111
600	21	4	0,5	6	0,112
700	29	9	0,7	9	0,114
800	36	7	1,0	12	0,117
900	42	8	1,2	16	0,118
1000	53	11	1,5	22	0,120

Обозначения: N – количество элементов факторного пространства; T_1 – время, затраченное на обучение многослойного персептрона на данных соответствующей размерности (архитектура персептрона выбиралась в соответствии с формулой (4) из расчета $\varepsilon = 0{,}2$); T_2 – время, затраченное на кластеризацию с применением текущего алгоритма кластеризации.

Примечания. Исследована работа алгоритмов кластеризации *k*-means, *c*-means, иерархического метода (Hierarchical), а также самоорганизующиеся карты Кохонена (SOM), размер $0{,}8N$, для обучения на данных, сформированных без использования кластеризации (Вариант 1, $T_2 = 0$) и с применением алгоритма кластеризации ($T_2 \neq 0$).

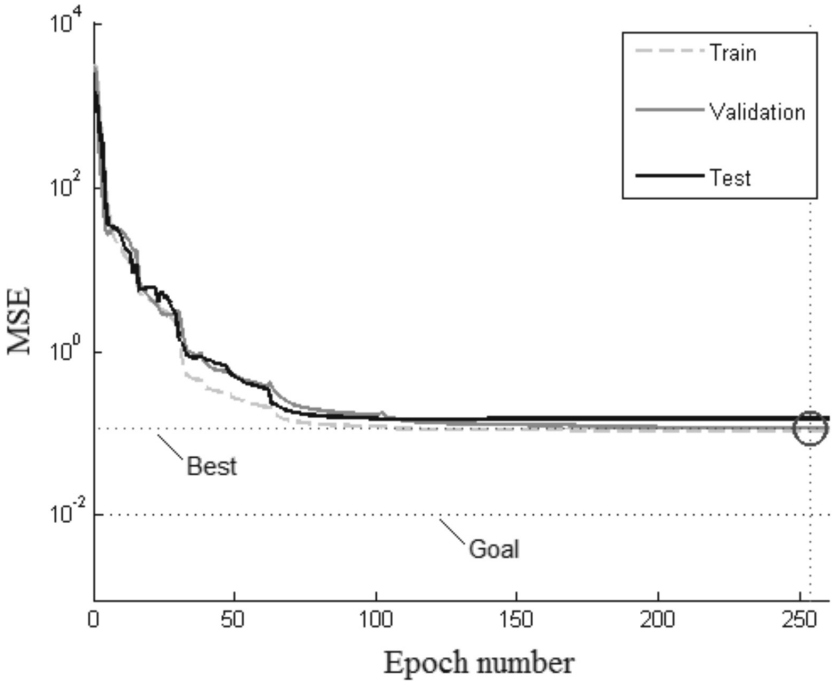

Рис. 2. Результаты обучения нейронной сети MLP на данных, сформированных с использованием алгоритма SOM

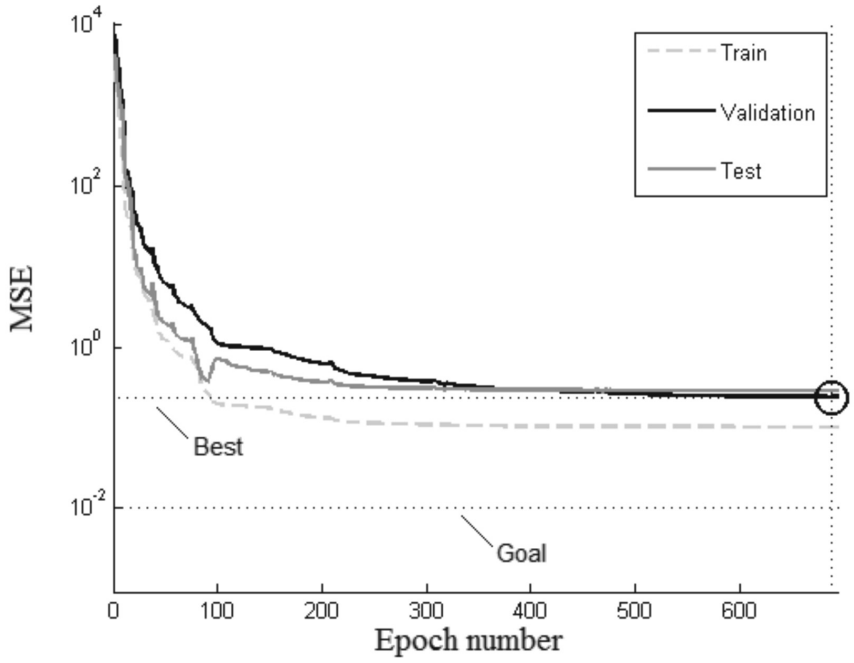

Рис. 3. Результаты обучения нейронной сети MLP на данных, сформированных с использованием алгоритма *k*-means

стеризации. В результате обучения прирост энтропии составил 0,19 бит. Среднеквадратичная ошибка обучения составила 0,23909. Зависимость времени работы алгоритма кластеризации от размера факторного пространства также приведена в табл. 3.

Прирост энтропии, как и в случае применения SOM, практически не меняется в зависимости от объема факторного пространства.

Время обучения нейронной сети оказалось на порядок меньше, чем при использовании SOM. Кроме того, по сравнению с самоорганизующимися картами, наблюдается больший прирост энтропии.

Результаты обучения нейронной сети на множестве, сформированном с применением алгоритма *k*-means, приведены на рис. 3.

Следует констатировать, что, несмотря на большее значения прироста энтропии, нейронная сеть обучилась хуже, чем на данных, сформированных с помощью SOM. Величина среднеквадратичной ошибки, а также разница между ошибками проверочного/тестового и обучающего множеств оказалась больше, чем с применением алгоритма SOM. Однако качество обучения улучшилось, по сравнению со случаем отсутствия кластеризации.

Подводя промежуточный итог вышеизложенному, следует отметить, что для алгоритма *k*-means величина прироста энтропии обучающего множества не является показательной. Это можно объяснить тем, что алгоритм *гарантированно* распределяет данные по указанному количеству кластеров и не справляется с задачей, когда объект в равной степени принадлежит к нескольким кластерам или вообще не принадлежит ни к одному из них.

Наибольший интерес с точки зрения расчета прироста энтропии представляет класс алгоритмов нечеткой кластеризации. Результаты такого исследования, проведенного для алгоритма *c*-means, приведены далее.

Алгоритм *c*-means. Данный алгоритм [1] относится к классу алгоритмов нечеткой кластеризации, т. е. определяет принадлежность данных к кластеру с некоторой вероятностью.

В данном случае значение вероятности

p принадлежности обучающего примера к кластеру выражается как

$$p = \frac{p_i^k}{N},$$

где p_i^k – вероятность попадания вектора исходных данных i в кластер k.

В расчетах значения вероятности менее 0,1 считаются малыми и не учитываются.

Прирост энтропии составил 0,36 бит, а среднеквадратичная ошибка обучения равна 0,1141. Разница среднеквадратичной ошибки между обучающим/проверочным и тестовым множествами оказалась значительно ниже, чем в случае, когда кластеризация не используется.

Время работы алгоритма в зависимости от размера факторного пространства приведено в табл. 3. Видно, что оно существенно возрастает по мере увеличения размера факторного пространства.

Результаты обучения нейронной сети MLP приведены на рис. 4.

Следует отметить, что, несмотря на значительный прирост энтропии, по сравнению, например, с самоорганизующимися картами Кохонена, уменьшение среднеквадратичной ошибки обучения не столь существенно и составляет 0,11419. Кроме того, уменьшение величины среднеквадратичной ошибки ничтожно мало, по сравнению с ростом времени работы алгоритма на факторных пространствах большого объема.

Алгоритм на основе построения иерархического дерева кластеров. Данный алгоритм относится к классу иерархической кластеризации [8]. В качестве метрики используется евклидово расстояние между элементами факторного пространства, которое вычисляется по следующей формуле:

$$d(X^1, X^2) = \sqrt{\sum_{i=1}^{n} (X_i^1 - X_i^2)^2}, \qquad (6)$$

где X^1, X^2 – элементы факторного пространства, причем $X_i^1 \in X^1$, $X_i^2 \in X^2$.

Прирост энтропии составил 0,1207 бит. Как и в предыдущих экспериментах, разница среднеквадратичной ошибки между обучающим и тестовым множествами меньше, чем в случае, когда кластеризация не используется.

Время работы алгоритма в зависимости

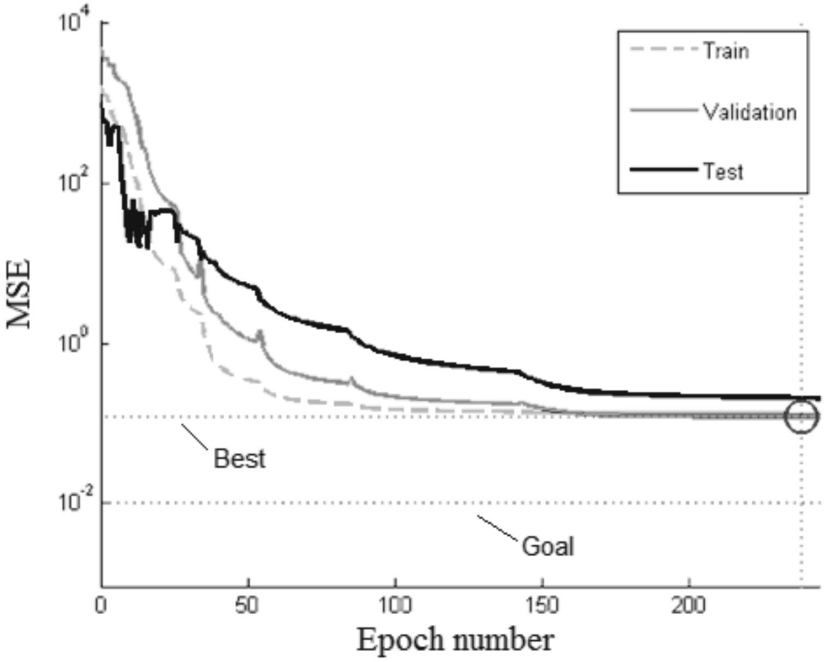

Рис. 4. Результаты обучения нейронной сети MLP на данных, сформированных с использованием алгоритма *c*-means;

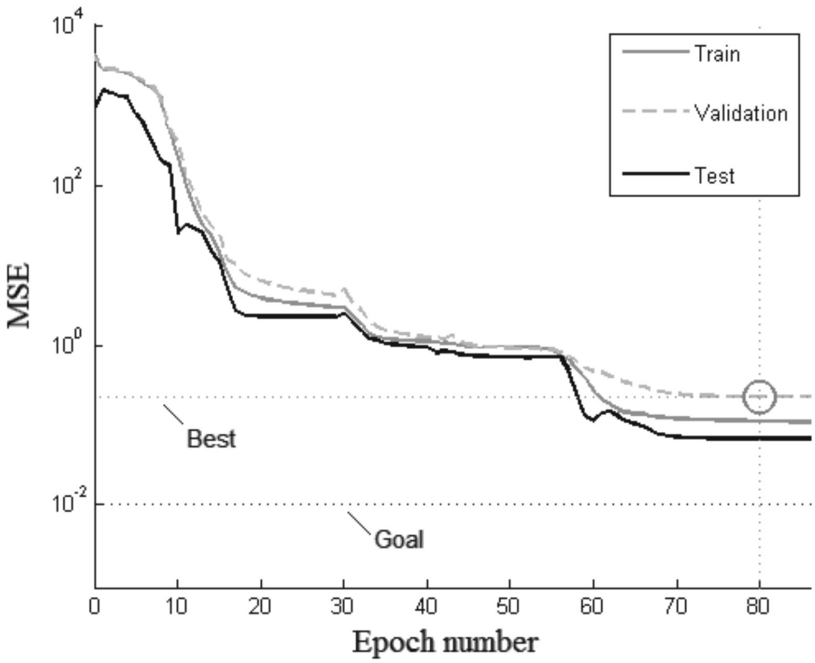

Рис. 5. Результаты обучения нейронной сети MLP на данных, сформированных с использованием алгоритма *c*-means; минимальное значение среднеквадратичной ошибки для проверочного множества показано окружностью

от размера факторного пространства приведено в табл. 3. Видно, что время работы алгоритма по мере увеличения размера факторного пространства возрастает несущественно.

Результаты обучения нейронной сети MLP приведены на рис. 5.

В данном случае показатель прироста энтропии является минимальным среди всех алгоритмов кластеризации. Среднеквадратичная ошибка обучения равна 0,22595, что, однако, несмотря на малое значение энтропии, меньше, чем при использовании алгоритма *k*-means, для которого прирост энтропии оказался максимальным среди всех рассматриваемых методов. Следует отметить, что применение иерархического метода значительно выигрывает у всех рассматриваемых алгоритмов по показателю времени работы (см. табл. 3).

Данные основных результатов исследования для всех алгоритмов кластеризации сведены в табл. 4.

Заключение

Проведенное исследование приводит к заключению, что для всех рассмотренных алгоритмов кластеризации выполняется

Таблица 4

Сравнение результатов проведенных экспериментов

Алгоритм	Прирост энтропии	Среднеквадратичная ошибка
SOM	0,14	0,11601
k-means	0,19	0,23909
c-means	0,36	0,11410
Hierarchical	0,12	0,22595
Без кластеризации	–	0,31462

условие (1), сформулированное в постановке задачи. Во всех экспериментах наблюдается прирост энтропии и уменьшение среднеквадратичной ошибки обучающего множества, а также снижение разницы между среднеквадратичной ошибкой проверочного/обучающего и тестового множеств, что говорит о повышении качества обучения.

С точки зрения выполнения условия (2), наилучший результат получен для алгоритма *c*-means. Однако следует принять во внимание тот итог, что при значительной размерности факторного пространства выигрыш в эффективности существенно ниже, по сравнению с ростом времени проведения кластеризации.

В заключение отметим, что, несмотря на положительный эффект применения алгоритмов кластеризации, эксперименты показали, что энтропия обучающего множества является важным, но не определяющим фактором улучшения качества обучения нейронной сети типа MLP.

СПИСОК ЛИТЕРАТУРЫ

1. **Нейский И.М.** Классификация и сравнение методов кластеризации. // Интеллектуальные технологии и системы. М.: НОК «Claim», 2006. Вып. 8. С. 130–142.

2. **Федоренко Ю.С., Гапанюк Ю.Е.** Кластеризация данных на основе самоорганизующихся растущих нейронных сетей и марковского алгоритма кластеризации // Нейрокомпьютеры: разработка, применение. 2016. № 4. С. 3–13.

3. **Подвальный С.Л., Плотников А.В., Белянин А.М.** Сравнение алгоритмов кластерного анализа на случайном наборе данных // Вестник Воронежского государственного технического университета. 2012. № 5. С. 4–6.

4. **Хачумов М.В.** Задача кластеризации текстовых документов// Информационные технологии и вычислительные системы. 2010. № 2. С. 42–49.

5. **Корягин Е.В.** Разработка модели ассоциативной памяти робота АР-600 для задачи кластеризации и обобщения данных// Нейроинформатика-2015. Сборник научных трудов, 2015. С. 38–47.

6. **Кохонен Т.** Самоорганизующиеся карты. М.: Бином. Лаборатория знаний, 2008. 655 с.

7. **Пастухов А.А., Прокофьев А.А.** Применение самоорганизующихся карт Кохонена для формирования представительской выборки при обучении многослойного персептрона // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2016. № 2 (242). С. 95–107.

8. **Кулаичев А.П.** Методы и средства комплексного анализа данных. М.: Инфра-М, 2006.

9. **Шеннон К**. Работы по теории информации и кибернетике. М.: Изд-во иностр. лит-ры, 1963. 830 с.

10. **LeCun Y.** Efficient learning and secondary methods. MA: MIT Press, 1993. 71 p.

11. **Хайкин С.** Нейронные сети. Пер. с англ. М.: ИД «Вильямс», 2008. 1104 с.

12. **Ту Дж., Гонсалес Р.** Принципы распознавания образов. М.: Мир, 1978. С. 109–112.

Статья поступила в редакцию 13.01.2017, принята к публикации 08.03.2017.

СВЕДЕНИЯ ОБ АВТОРАХ

ПАСТУХОВ Алексей Андреевич – *аспирант кафедры высшей математики № 1 Национального исследовательского университета «МИЭТ».*

124498, Российская Федерация, г. Москва, Зеленоград, проезд 4806, д. 5

pastuhov1992@gmail.com

ПРОКОФЬЕВ Александр Александрович – *доктор педагогических наук, кандидат физико-математических наук, заведующий кафедрой высшей математики № 1 Национального исследовательского университета «МИЭТ».*

124498, Российская Федерация, г. Москва, Зеленоград, проезд 4806, д. 5

aaprokof@yandex.ru

REFERENCES

[1] **I.M. Neyskiy,** Klassifikatsiya i sravneniye metodov klasterizatsii [Classification and comparison of clustering procedures], Intellektualnyye tekhnologii i sistemy, Moscow, NOK ‘Claim’.

No. 8 (2006) 130–142.

[2] **Yu.S. Fedorenko, Yu.E. Gapanyuk,** Klasterizatsiya dannykh na osnove samoorganizuyushchikhsya rastushchikh neyronnykh setey i markovskogo algoritma klasterizatsii [Data clustering based on self-organizing growing neural networks and on Markovian clustering algorithm], Neurocomputers: building and application. No. 4 (2016) 3–13.

[3] **S.L. Podvalnyy, A.V. Plotnikov, A.M. Belyanin,** Sravneniye algoritmov klasternogo analiza na sluchaynom nabore dannykh [Comparison of algorithms of cluster analyses based on random data set], Vestnik VGTU. No. 5 (2012).

[4] **M.V. Khachumov,** Zadacha klasterizatsii tekstovykh dokumentov [Clustering of textual documents], Informatsionnyye tekhnologii i vychislitelnyye sistemy. No. 2 (2010) 42–49.

[5] **E.V. Koryagin,** Razrabotka modeli assotsiativnoy pamyati robota AR-600 dlya zadachi klasterizatsii i obobshcheniya dannykh [Model building of the associative memory of AR-600 robot for clustering and data integrating], Neuroinformatics-2015, Collected papers (2015) 38–47.

[6] **T. Kokhonen,** Samoorganizuyushchiyesya karty [Self-organizing maps], Moscow, Binom. Laboratoriya znaniy, 2008.

[7] **A.A. Pastukhov, A.A. Prokofyev,** Kohonen self-organizing map application to representative sample formation in the training of the multilayer perceptron, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. No. 2 (242) (2016) 95–107.

[8] **A.P. Kulaichev,** Metody i sredstva kompleksnogo analiza dannykh [Methods and resources of integrated data analysis], Moscow, Infra-M, 2006.

[9] **K. Shennon,** Raboty po teorii informatsii i kibernetike [Studies in information theory and cybernetics], Moscow, Izd-vo Inostr. Lit-ry, 1963.

[10] **Y. LeCun,** Efficient learning and secondary methods, MA, MIT Press, 1993.

[11] **S. Khaykin,** Neyronnyye seti [Neural networks], 2nd ed., Moscow, ID 'Williams', 2008.

[12] **J. Tu, R. Gonsales**, Printsipy raspoznavaniya obrazov [Pattern recognition principles], Moscow, Mir, 1978, Pp. 109–112.

Received 13.01.2017, accepted 08.03.2017.

THE AUTHORS

PASTUKHOV Aleksey A.
National Research University of Electronic Technology
5 Pass. 4806, Zelenograd, Moscow, 124498, Russian Federation
pastuhov1992@gmail.com

PROKOFIEV Aleksander A.
National Research University of Electronic Technology
5 Pass. 4806, Zelenograd, Moscow, 124498, Russian Federation
aaprokof@yandex.ru