

УДК 004.8, 004.942, 622.279

Ю.А. Келлер**ПРИМЕНЕНИЕ КЛАСТЕРИЗАЦИИ ДАННЫХ НА ОСНОВЕ
САМООРГАНИЗУЮЩИХСЯ КАРТ КОХОНЕНА ПРИ ПОДБОРЕ
СКВАЖИН-КАНДИДАТОВ ДЛЯ МЕТОДОВ УВЕЛИЧЕНИЯ НЕФТЕОТДАЧИ**

Предложена методология предварительной кластеризации промысловой информации на основе самоорганизующихся карт Кохонена в рамках работ по подбору скважин-кандидатов для выполнения методов увеличения нефтеотдачи. Приведены результаты численных расчётов, характеризующие эффективность разработанной методики.

Ключевые слова: методы увеличения нефтеотдачи; нагнетательная скважина; выравнивание профиля приёмности; самоорганизующиеся карты Кохонена; нейронные сети.

При разработке нефтяных месторождений одной из главных проблем на данный момент является преждевременное обводнение добываемой продукции. Уже сейчас это приводит к стабилизации темпов роста извлечения нефти, а в ближайшие 10–20 лет при сохранении выбранной стратегии разработки может привести к его снижению [1]. Поэтому с целью увеличения рентабельного срока эксплуатации нефтяных месторождений для снижения текущей обводненности продукции применяется выравнивание профиля приёмности (ВПП) на нагнетательных скважинах. Для этого в скважину закачивается специальный реагент для уменьшения проницаемости промытых прослоёв пласта, что, в свою очередь, приводит к более равномерному фронту вытеснения нефти водой, вовлечения в разработку непромытых зон пласта, уменьшая тем самым обводненность продукции добываемых скважин.

На данный момент критерии выбора скважин-кандидатов под методы увеличения нефтеотдачи (МУН) проработаны в полной мере, поэтому на первый план выходит корректная оценка ожидаемой дополнительной добычи нефти за счет снижения обводненности продукции в результате проведения МУН. Известны методики по оцениванию технологической эффективности ВПП на основе теории нейронных сетей [2–4].

В данной работе предлагается осуществлять предварительную кластеризацию исходной промысловой информации с помощью самоорганизующихся карт (СОК) Кохонена с последующим построением нейронных сетей по методологии [4] для оценки технологической эффективности ВПП.

1. Алгоритм кластеризации

Особенностью моделирования задач оценки технологической эффективности ВПП является большая степень дифференциации исходной информации, обусловленная расчлененностью характеристик пласта (пористость, проницаемость, выдержанность и др.), различными промысловыми показателями скважин (дебиты, величины забойного давления), степенью эффективности поддержки пластового давления в зоне отборов и др. Таким образом, при попытке создания модели искусственной нейронной сети (ИНС), основанной на несгруппированной исходной информации [Там же], средняя ошибка предсказания дебита нефти добываемой скважины после ВПП составила более 50%. Поэтому применение данной модели на практике вносило существенную долю неопределенности в прогнозируемые характеристики. На данный момент известны успешные опыты внедрения кластеризации

информации при решении нефтегазопромысловых задач в области МУН [5, 6]. С целью улучшения предсказательной способности модели ИНС [4] было принято решение использовать СОК Кохонена для предварительной кластеризации исходной информации.

Задача кластеризации сводится к разбиению множества объектов, представленных векторами $x^i \in R^n$, $i = 1, 2, \dots, N$, на группы, называемые кластерами. Объекты формируются в кластеры таким образом, чтобы те из них, что попали в один и тот же кластер, были максимально похожи друг на друга и одновременно максимально отличными от объектов из других кластеров [7]. До внедрения систем искусственного интеллекта предпочтение отдавалось классическим методам кластеризации (древовидная кластеризация, метод k -средних и т.д.). В данной группе методов предполагается, что наиболее типичный представитель кластера (ядро кластера) известен, и задача кластеризации решается путём «стягивания» к ядрам кластера других векторов [8. С. 241]. Однако в последние годы бурно развивающиеся альтернативные классическим методам кластеризации стали СОК Кохонена. Они применяются при моделировании, прогнозировании, поиске закономерностей в больших массивах данных, выявлении кластеров [9, 10]. Важное преимущество новой методологии кластеризации заключается в особенности сети Кохонена от других ИНС: данный тип нейронных сетей способен к неконтролируемому обучению, т.е. обучающая выборка состоит лишь из значений входных переменных, в процессе обучения нет сопоставления выходов нейронов с эталонными значениями. Схематично структура сети Кохонена представлена на рис. 1.

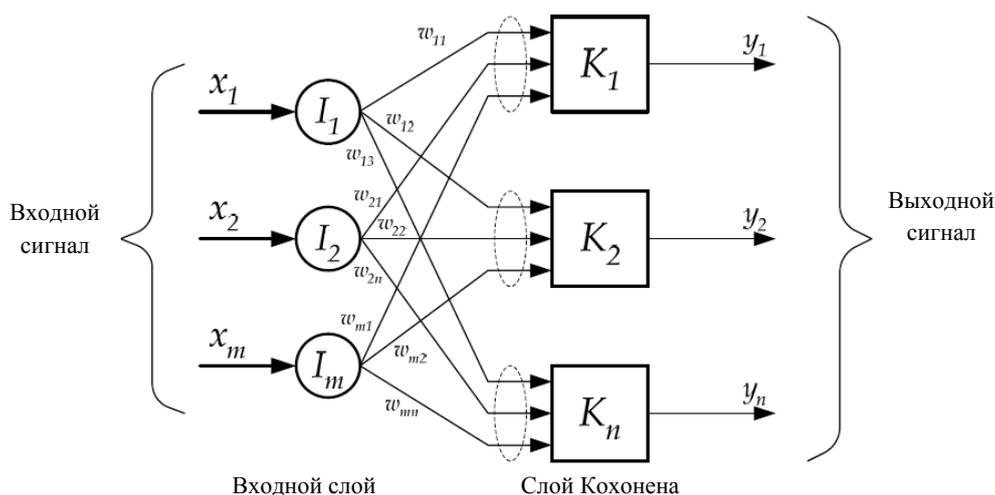


Рис. 1. Структура самоорганизующейся сети Кохонена:

x_1, x_2, \dots, x_m – множество входных характеристик; ω_j – весовые коэффициенты сети; y_1, y_2, \dots, y_n – множество выходных характеристик, соответствующих n кластерам

Сам алгоритм обучения получил название «победитель забирает всё», который представляет последовательность следующих шагов [11]:

1. Выбирается шаг обучения $\eta > 0$ (обычно от 0,1 до 0,7).
2. Происходит инициализация весов сети случайными числами.
3. Осуществляется настройка весов. Для каждого вектора производится:

– расчёт выходов нейронов слоя Кохонена. Выбирается нейрон-победитель r с наибольшим значением на выходе;

– корректировка весов победившего нейрона:

$$\omega_r = \omega_r + \eta(x - \omega_r);$$

– нормировка вектора весов нейрона-победителя:

$$\omega_r = \frac{\omega_r}{\|\omega_r\|}.$$

Пункты 1–3 представляют собой серию последовательных приближений. Во время обучения сети на вход подаются исходные характеристики и сеть настраивается под определенные закономерности, обнаруженные в выборке. На начальном этапе процедура обучения начинается со случайно выбранного выходного расположения центров. Далее на вход сети подаются обучающие примеры из выборки и находится наиболее похожий нейрон, для которого скалярное произведение весов ω_{ij} и обучающего примера минимально. Система назначает этот нейрон победителем, и он объявляется центральным для настройки весов всех близкорасположенных нейронов. В отличие от других нейронных сетей, суть обучения сводится к настройке весов таким образом, чтобы достичь наилучшего совпадения с поданными на вход примерами. После того как на вход сети подано достаточное количество обучающих примеров, сеть имеет возможность выделить группы и система весов генерируется так, чтобы близкие с точки зрения топологии узлы были чувствительны к идентичным входным сигналам. В итоге центр кластера фиксируется так, чтобы наилучшим образом группировать примеры, для которых такой нейрон выигрывает. Поэтому завершением обучения СОК Кохонена будет задание меры соседства нейронов, окружающих победивший нейрон. СОК формирует такую топологическую карту, где идентичные примеры представляют кластеры, близко расположенные друг к другу. В итоге СОК Кохонена разбивает входные примеры на кластеры, схожие по совокупности своих характеристик, и позволяет визуальным образом представить их в плоскости нейронов. Отсюда следует еще одно важное свойство сетей Кохонена – они позволяют трансформировать N -мерное пространство входных признаков в двумерное. Поэтому задача инженерного анализа может быть сведена к простому визуальному наблюдению кластеров на топологической карте с выдачей соответствующих рекомендаций.

При этом СОК Кохонена не лишены недостатка классических методов кластеризации: перед проведением обучения сети необходимо выбрать число кластеров, на которое будет разбиваться исходное множество.

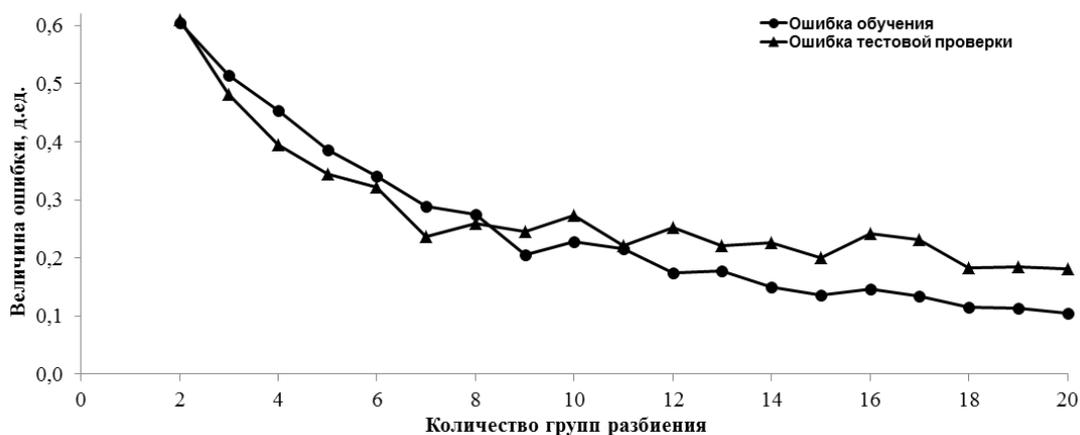


Рис. 2. Зависимость величины ошибки обучения от числа кластеров

Исходя из собственного опыта проведения работ по подбору физико-химическим МУН на нефтяных месторождениях РФ, а также из опыта работы других авторов [12], был пересмотрен набор исходных параметров, предложенных в работе [4]. Исключены из модели ИНС параметры: источник обводнения скважины, вязкость нефти, толщина обрабатываемого слоя коллектора; добавлены в модель ИНС: потенциал динамического уровня и средняя компенсация на участке. Основной целью при пересмотре входного набора параметров было избавиться от тех, что несут в себе долю неопределенности либо определение которых с необходимой степенью достоверности невозможно. Построение сети Кохонена осуществлялось в программном обеспечении Statistica. В качестве входной информации для обучения СОК использовался опыт проведения ВПП на одном из нефтяных месторождений Поволжского региона в период с 2007 по 2011 г. В обучающую выборку попали промысловые пока-

затели по более чем 100 скважинам. 90% исходной информации использовалось для обучения сети Кохонена, 10% – для контрольной проверки работоспособности созданной модели. Выбор оптимального числа кластеров для разбиения осуществлялся следующим образом: поочередно увеличивая число кластеров, на которые будет разбиваться исходная выборка, оценивались величина ошибки обучения модели и величина ошибки кластеризации для тестовой выборки. На оси абсцисс отмечалось число кластеров, на оси ординат – величина ошибки, соответствующая данному числу кластеров. Результаты представлены на рис. 2.

Из рис. 2 видно, что при достижении определенного числа кластеров (количества групп разбиения) величина ошибки значительно не уменьшается и дальнейшее увеличение числа кластеров приводит лишь к увеличению времени обучения без существенного улучшения точности обучения. Поэтому было принято решение выделить четыре группы (кластера) для кластеризации исходной информации.

2. Результаты моделирования

Моделирование осуществлялось следующим образом: рассматривались два варианта проведения моделирования для предсказания эффективности проведения ВПП. В первом варианте предварительная кластеризация исходной информации не осуществлялась и ИНС создавалась на всем объеме исходной информации. Во втором варианте на первом этапе вся исходная информация была разделена на четыре кластера и в рамках каждого кластера создавалась своя ИНС (более подробно процедура создания ИНС представлена в [4]). В качестве выходного (предсказываемого) параметра модели ИНС рассматривалась обводненность добывающей скважины после ВПП. Выбор данной характеристики был обусловлен тем, что успешность проведения ВПП на месторождении характеризуется снижением обводненности продукции добывающих скважин (как по отдельным скважинам, так и в целом по участку). Под обводненностью продукции понимается отношение добываемой воды в объеме добываемого флюида. Например, обводненность в 90% на добывающей скважине означает, что из 1000 м³ добываемой жидкости в сутки только 100 м³ составляет нефть, а остальные 900 м³ – вода. Поэтому суммарное снижение обводненности продукции после ВПП позволяет получить дополнительную добычу нефти.

Далее для каждого из двух вариантов оценивалось среднее арифметическое расхождения предсказываемой и фактической обводненности после ВПП. Кросс-плот распределения расчетных и фактических обводненностей представлен на рис. 3, 4. Рис. 3 соответствует первому варианту моделирования (без предварительной кластеризации информации), рис. 4 – второму варианту моделирования (с предварительной кластеризацией информации).

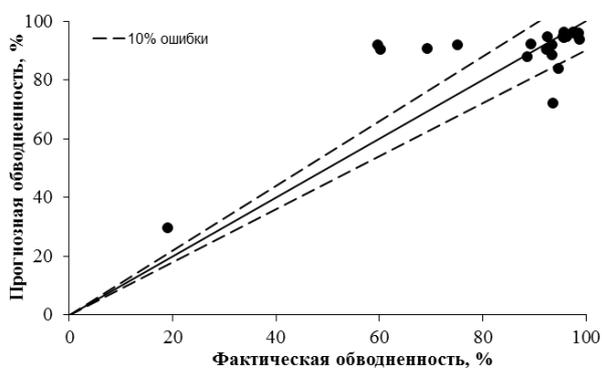


Рис. 3. Сравнение фактических и расчетных значений обводненности после ВПП без предварительной кластеризации информации

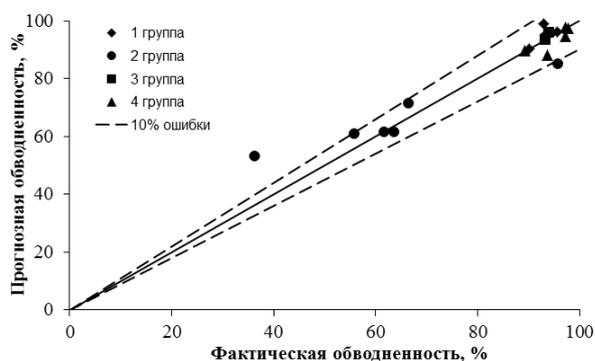


Рис. 4. Сравнение фактических и расчетных значений обводненности после ВПП с предварительной кластеризацией информации

Средняя ошибка предсказания обводненности после ВПП без предварительной кластеризации составила 10,6%, а с предварительной кластеризацией – 6,5%, что почти в два раза меньше. Данный результат моделирования подтверждает, что при попытке использования всего объема выборки для обучения ИНС результат предсказания может быть неудовлетворительным в связи с большой дифференциацией исходной информации. Группировка такой информации по кластерам, которые группируются по степени схожести характеристик, помогает строить более точные с точки зрения предсказания модели ИНС. Важно отметить, что и в первом и во втором случае имеются отдельные скважины, ошибка предсказания обводненности для которых значительно превышает 10%-ный порог. Это объясняется тем, что на данной скважине после ВПП было произведено другое геологическое мероприятие (ГТМ), спровоцировавшее резкий рост обводненности продукции. Поскольку в обучающей выборке такие прецеденты отсутствовали, то построенная модель ИНС для данной скважины не смогла адекватно предсказать ожидаемую обводненность после ВПП. В дальнейшем планируется доработка методики [4] с включением в исходный набор данных информации о проводимых на участке ГТМ.

Заключение

Решена задача предварительной кластеризации данных с помощью СОК Кохонена при оценке эффективности проведения ВПП на нефтяном месторождении с помощью математического моделирования. Показано, что применение предварительной кластеризации позволило почти вдвое снизить погрешность предсказания обводненности на добывающих скважинах одного из нефтяных месторождений Поволжского региона.

ЛИТЕРАТУРА

1. *Нефтегазодобывающая и нефтеперерабатывающая промышленность: Тенденции и прогнозы.* URL: <http://vid1.rjan.ru/ig/ratings/oil9.pdf> (дата обращения: 28.05.2014).
2. *Herbas J., Moreno R., Marin T.A., Romero M.F., Coombe D.* Reservoir simulation of non selective of a polymer gel treatment to improve water injection profiles and sweep efficiency in the Lagomar field Western Venezuela (SPE-92025-MS) // SPE international petroleum conference. Puebla, 2004. 10 p.
3. *Ferreira I., Gammiero A., Llamedo M.* Design of a neural network model for predicting well performance after water shutoff treatments using polymer gels (SPE 153908) // SPE Latin American and Caribbean petroleum conference. Mexico, 2012. 9 p.
4. *Келлер Ю.А., Усков А.А.* Применение теории регрессионного анализа и нейронных сетей для оценки эффективности применения потокоотклоняющих технологий на группе месторождений ОАО «Самаранефтегаз» // Инженерная практика. 2013. № 10. С. 56, 57.
5. *Иванов Е.Н., Кононов Ю.М.* Выбор методов увеличения нефтеотдачи на основе аналитической оценки геологической информации // Известия Томского политехнического университета. 2012. Т. 321, № 1. С. 149–154.
6. *Применение метода главных компонент при выборе новых эффективных технологий увеличения нефтеотдачи для условий месторождений ЗАО «Лукойл-Пермь» / И.И. Абызбаев и др.* // Геология, геофизика и разработка нефтяных месторождений. 2000. № 11. С. 18–21.
7. *Сараев П.В.* Нейросетевые методы искусственного интеллекта : учеб. пособие. Липецк : ЛГТУ, 2007. 64 с.
8. *Халафян А.А.* STATISTICA 6. Статистический анализ данных. М. : Бином, 2007. 508 с.
9. *Kohonen T.* Self-Organizing Maps (Third Extended Edition). N.Y., 2001. 501 p.
10. *Манжула В.Г., Федяшов Д.С.* Нейронные сети Кохонена и нечеткие нейронные сети в интеллектуальном анализе данных // Фундаментальные исследования. 2011. № 4. С. 108–114.
11. *Сеньковская И.С., Сараев П.В.* Автоматическая кластеризация в анализе данных на основе самоорганизующихся карт Кохонена // Вестник МГТУ им. Г.И. Носова. 2001. № 2. С. 78, 79.
12. *Жуков Р.Ю., Петраков А.М., Серов А.В.* Применение физико-химических технологий воздействия на завершающей стадии разработки месторождений // Нефтяное хозяйство. 2011. № 1. С. 51–53.

Келлер Юрий Александрович, канд. физ.-мат. наук. E-mail: kua1102@rambler.ru
Томский государственный университет

Поступила в редакцию 27 мая 2014 г.

Keller Yuri A. (Tomsk State University, Tomsk, Russian Federation).

The application of data clustering on basis of the Kohonen self-organizing maps by selecting candidate wells for enhanced oil recovery methods.

Keywords: enhanced oil recovery methods; injector well; improving water injection profile; self-organizing maps; artificial neural networks.

A methodology of preliminary clustering of field information is suggested on the basis of Kohonen self-organizing maps within the process of selecting candidate wells for enhanced oil recovery methods. The peculiarity of modeling problems of estimation of the technological efficiency of improving water injection profile is the high degree of differentiation of initial information, which is caused by the compartmentalization of formation characteristics (porosity, permeability, lateral continuity etc.), the different well characteristics (rates, bottom hole pressure values), the degree of efficiency of formation pressure maintenance in the production zone etc. Therefore, by creating an artificial neural network model based on ungrouped initial information, the average error of oil-production well rate prediction after improving water injection profile was over 50 per cent. To improve the prediction ability of a neural network, the initial data for training the network have been revised and preliminary grouping initial information by means of the Kohonen self-organizing maps has been done. It has been shown that preliminary grouping improved the prediction efficiency of a neural network nearly twice. The average error in the well water-cut prediction is 6.5 per cent. The suggested method is characterized by the far lesser work-time in comparison with the hydrodynamic modeling.

REFERENCES

1. *Neftegazodobyvayushchaya i neftepererabatyvayushchaya promyshlennost': Tendentsii i prognozy* [Oil and gas extraction and refining industry: Trends and Forecasts]. Available at: <http://vid1.rian.ru/ig/ratings/oil9.pdf>. (Accessed: 28th May 2014).
2. Herbas J., Moreno R., Marin T.A., Romero M.F., Coombe D. Reservoir simulation of non selective of a polymer gel treatment to improve water injection profiles and sweep efficiency in the Lagomar field Western Venezuela (SPE-92025-MS). *Proc. SPE international petroleum conference*. Puebla, 2004. 10 p.
3. Ferreira I., Gammiero A., Llamado M. Design of a neural network model for predicting well performance after water shutoff treatments using polymer gels (SPE 153908). *Proc. SPE Latin American and Caribbean petroleum conference*. Mexico, 2012. 9 p.
4. Keller Yu.A., Uskov A.A. Primenenie teorii regressionnogo analiza i neyronnykh setey dlya otsenki effektivnosti primeneniya potokootklonyayushchikh tekhnologiy na gruppe mestorozhdeniy OAO "Samaraneftegaz" [The application of the theory of regression analysis and neural networks to assess the effectiveness of fluid flow diverting methods on a group of fields "Samaraneftegaz"]. *Inzhenernaya praktika - Oilfield Engineering*, 2013, no. 10, pp. 56-57.
5. Ivanov E.N., Kononov Yu.M. The choice of enhanced oil recovery methods on the basis of analytical assessment of geological and field information. *Izvestiya Tomskogo politekhnicheskogo universiteta – Bulletin of the Tomsk Polytechnic University*, 2012, vol. 321, no. 1, pp. 149-154. (In Russian).
6. Katoshin A.F., Yakimenko G.F., Zotikov V.I., Abyzbaev I.I., Gafurov O.G. Primenenie metoda glavnykh komponent pri vybore novykh effektivnykh tekhnologiy uvelicheniya nefteotdachi dlya usloviy mestorozhdeniy ZAO "Lukoil-Perm" [The application of principal component analysis in the selection of new efficient technologies for increasing oil field conditions JSC "Lukoil-Perm"]. *Geologiya, geofizika i razrabotka neftyanykh mestorozhdeniy*, 2000, no. 11, pp. 18-21.
7. Saraev P.V. *Neyrosetevye metody iskusstvennogo intellekta* [Neural network artificial intelligence techniques]. Lipetsk: LTSU Publ., 2007. 64 p.
8. Khalafyan A.A. *STATISTICA 6. Statisticheskii analiz dannykh* [STATISTICA 6. Statistical analysis]. Moscow: Binom Publ., 2007. 508 p.
9. Kohonen T. *Self-organizing maps*. New York: Springer-Verlag, 2001. 501 p.
10. Manzhula V.G., Fedyashov D.S. Kohonen neural networks and fuzzy neural networks in data mining. *Fundamental'nye issledovaniya – Fundamental research*, 2011, no. 4, pp. 108-114. (In Russian).
11. Senkovskaya I.S., Saraev P.V. Avtomaticheskaya klasterizatsiya v analize dannykh na osnove samoorganizuyushchikhsya kart Kokhonena [Automatic clustering in data analysis based on Kohonen self-organizing maps]. *Vestnik MGTU im. G.I. Nosova*, 2001, no. 2, pp. 78-79.
12. Zhukov R.Yu., Petrakov A.M., Serov A.V. The Application of EOR-methods on the final stages of field development. *Neftyanoe khozyaistvo - Oil Industry*, 2011, no. 1, pp. 51-53. (In Russian).