

V. P. Kulagin¹, Prof., e-mail: vkulagin@hse.ru,
V. Ya. Tsvetkov², Prof., e-mail: svj2@mail.ru, M. P. Lapchinskaya³
¹Higher School of Economics (National Research University)

²Moscow State Technical University of Radio Engineering, Electronics and Automation

³Moscow State University of Geodesy and Cartography

Identification of Ships in Heavy Seas

Identification is one of the main tasks of remote methods and tools that are installed on aircraft and space platforms. The article describes a method of description and analysis of the sea surface, characterized by a strong emotion. This will involve the application of automated image recognition using characteristic features, such as the dimensions of the basic elements of texture, which are characterized by vagueness of linear size. The method is based on a method of fuzzy sets. The method allows to obtain a model of the sea surface with strong agitation, which allow us to identify marine objects in digital images. On the image surface model, which is subtracted from the original image. The remaining "clean" the image identification of the marine facility is not a rough surface.

Keywords: identification, modeling, rough sea surface, images, fuzzy sets

References

1. Barmin I. V., Saviny'kh V. P., Tsvetkov V. Ya., Zatiagalova V. V. Monitoring zagriaznenii' moria sudami po danny'm distantsionnogo zondirovaniia. *Morskoi' sbornik*. 2013. V. 1998, N. 9. P. 41–49.
2. Pavly'gin E'. D., Sosnin P. I. Mnogoagentnoe modelirovanie i vizualizatsiia okruzhaiushchei' obstanovki morskogo sudna. *Avtomatizatsiia protsessov upravleniia*. 2010. N. 2. P. 3–12.
3. Skjetne R., Smogeli O. N., Fossen T. I. A Nonlinear Ship Manoeuvring Model: Identification and adaptive control with experiments for a model ship. *Modeling, identification ad control*. 2004. V. 25, N. 1. P. 3–27.
4. Gorobets V. N., Gutnik V. G., Zotov S. M., Kivva F. V., Shapiro A. A. Matematicheskaia model' radiolokatsionnogo obraza korablia na morskome volnenii. *Radiofizika ta elektronika*, 2011. V. 2 (16). N. 4. P. 60–65.
5. Skjetne R., Smogeli O. N., Fossen T. I. A Nonlinear Ship Manoeuvring Model: Identification and adaptive control with experiments for a model ship. *Modeling, identification ad control*. 2004. V. 25, N. 1. P. 3–27.
6. Chang S. J. Vessel identification and monitoring systems for maritime security. *Security Technology*, 2003. Proceedings. IEEE 37th Annual 2003 International Carnahan Conference on. IEEE, 2003. P. 66–70.
7. Fossen T. I., Sagatun S. I., Sorensen A. J. Identification of dynamically positioned ships. *Control Engineering Practice*. 1996. V. 4. N. 3. P. 369–376.
8. Buckley G. W. et al. Control system for a marine vessel: Pat. 6273771 USA. 2001.
9. Clifford P. J., Hart N. R., Meulman C. B. Satellite system for vessel identification: Pat.7483672 USA. 2009.
10. Tsvetkov V. Ya. Information Situation and Information Position as a Management Tool. *European Researcher*. 2012. V. 36, N. 12-1. P. 2166–2170.
11. Tsvetkov V. Ya. Semantic environment of information units. *European Researcher*. 2014. V. 76, N. 6-1. P. 1059–1065.
12. Lapchinskaya M. P. Algoritm klassifikatsii odnorodny'kh tekstur zemnoi' poverkhnosti, osnovanny'i na postroenii funktsii' pri nadlezhnosti. *Mezhdunarodny'i' nauchno-tekhnicheskii' i proizvodstvenny'i' zhurnal "NAUKI O ZEMLE"*. 2012. N. 4. P. 42–53.
13. Kofman A. *Vvedenie v teoriu nechetkikh mnozhestv*. M.: Radio i sviaz', 1982. 432 p.
14. Lapchinskaya M. P. Metod tsifrovogo muara. *Geodeziiz i ae'rofotos'emka*. 1997. N. 5. P. 86–113.
15. Saviny'kh V. P. Sistema polucheniiia koordinatno-vremennoi' informatcii dlia resheniia zadach monitoringa. *Mezhdunarodny'i' nauchno-tekhnicheskii' i proizvodstvenny'i' zhurnal "Nauki o Zemle"*. 2012. Iss. 03. P. 5–10.
16. Tsvetkov V. Ya., Lobanov A. A., Matchin V. T., Zhelezniakov V. A. Obnovlenie bankov danny'kh prostranstvennoi' informatcii. *Infomatizatsiia obrazovaniia i nauki*. 2015. N. 1(25). P. 128–136.

УДК 004.94

С. В. Лучкова, канд. техн. наук, мл. науч. сотр., Т. О. Перемитина, канд. техн. наук, науч. сотр.,
И. Г. Ященко, канд. геол.-мин. наук, зав. лаб.

Федеральное государственное бюджетное учреждение науки Институт химии нефти СО РАН, Томск

Использование нечеткого моделирования для повышения репрезентативности информации на примере анализа характеристик нефти

Рассмотрена задача восстановления пропущенных значений в выборке о свойствах сернистых нефтей. Описывается алгоритм восстановления данных на основе нечеткого моделирования. Приведены результаты экспериментальных исследований для повышения репрезентативности выборки на основе неполных данных.

Ключевые слова: восстановление данных, нечеткое моделирование, физико-химические свойства нефти

Введение

Для изучения природных объектов анализируются данные, которые имеют сложную структуру и представлены многомерными выборками с пропущенными значениями. Существующая необходимость в анализе неполных данных предполагает их предварительную обработку, такую как восполнение недостающих значений для увеличения репрезентативности выборки. Это необходимо, так как в самом простом решении для анализа таких данных (исключить все записи с пропусками) существуют проблемы с ухудшением достоверности выводов анализа вследствие сокращения размера выборки и, как следствие, проявлением статистических смещений. Однако для первоначального анализа и выявления особенностей данных используются именно полные значения (выборки данных без пропущенных значений), которые позволяют оценить предстоящую работу и выбрать подходящие методы восстановления.

Одним из продуктивных подходов к решению проблемы о пропущенных значениях, касательно нефти, является применение методов нечеткого моделирования, которые позволяют наиболее точно восстановить недостающие данные за счет работы с выборками малой размерности и в условиях статистической неопределенности [1].

Этапы восстановления данных

Процедура восстановления данных предполагает трудоемкий предварительный процесс, который состоит из нескольких этапов:

1. Предварительный анализ исходных данных для оценки их полноты.
2. Формирование полной выборки.
3. Построение модели исследуемых данных по полной выборке.
4. Проверка адекватности построенной модели.
5. Восстановление данных на основе построенной модели.

Рассмотрим этапы подробнее.

Этап "Предварительный анализ" включает в себя не только оценку полноты данных, но также помогает выявить особенности объекта исследования, например, проследить средние значения характеристик, разброс значений характеристик, какие-либо взаимосвязи.

Этап "Формирование полной выборки" заключается в исключении всех записей из выборки, в которых есть пропущенные значения, а также позволяет принять решение о "случайных выбросах" в значениях характеристик, которые не являются характерными для данных и могут оказать значительное влияние на статистические величины. Для этого этапа могут быть использованы различные статистические методы, которые помогут понять исследуемые данные, например, по полной таблице

можно выявить корреляционные связи и главные факторы с помощью факторного анализа [2].

Этап "Построение модели" выполняется с помощью нечеткого моделирования, которое включает в себя два основных подэтапа: идентификацию структуры (НС) с заданной структурой [3], с треугольной функцией принадлежности для построения системы и методом эволюционной стратегии (ЭС) для настройки параметров системы [4, 5].

Идентификация структуры — это определение основных характеристик нечеткой модели (число нечетких правил, число лингвистических термов для входных и выходных переменных). Настройка же параметров НС — это определение неизвестных параметров антецедентов (условная часть — "ЕСЛИ") и консеквентов (заключительной части — "ТО") нечетких правил путем оптимизации работы нечеткой модели по заданному критерию.

Правила нечеткой модели типа сингтон имеют следующий вид:

$$\text{правило } i: \text{ЕСЛИ } x_1 = A_{1i} \text{ И } x_2 = A_{2i} \text{ И } \dots \text{ И } x_m = A_{mi} \\ \text{ТО } y = r_i$$

где A_{ji} — лингвистический терм, которым оценивается переменная x_j , а выход y оценивается действительным числом r_i .

Модель осуществляет отображение $F: \mathfrak{R}^m \rightarrow \mathfrak{R}$, заменяя оператор нечеткой конъюнкции произведением, а оператор агрегации нечетких правил сложением. Отображение F для модели типа сингтон определяется формулой

$$F(\mathbf{x}) = \frac{\sum_{j=1}^m r_j \prod_{i=1}^n \mu_{A_{ij}}(x_i)}{\sum_{j=1}^m \prod_{i=1}^n \mu_{A_{ij}}(x_i)},$$

где $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathfrak{R}^n$ — значение i -го входа; $\mu_{A_{ij}}(x_j)$ — функция принадлежности лингвистического терма A_{ij} ; r_j — значение консеквента в j -м правиле.

Нечеткая система может быть представлена как $y = f(\mathbf{x}, \boldsymbol{\theta})$, где $\boldsymbol{\theta} = \|\theta_1, \dots, \theta_N\|$ — вектор параметров; N — сумма термов по каждому исследуемому параметру; y — скалярный выход системы.

Задача параметрической идентификации — определить неизвестные параметры антецедентов и консеквентов нечетких правил путем оптимизации работы нечеткой системы по заданному критерию.

Параметрическая идентификация рассматривается как процесс оптимизации нечеткой модели, который сводится к нахождению таких параметров нечеткой системы, при которых ошибка вывода была бы минимальной. При этом оценивается качество нечеткого вывода по значениям ошибки вывода — разницы между значениями выходной пере-

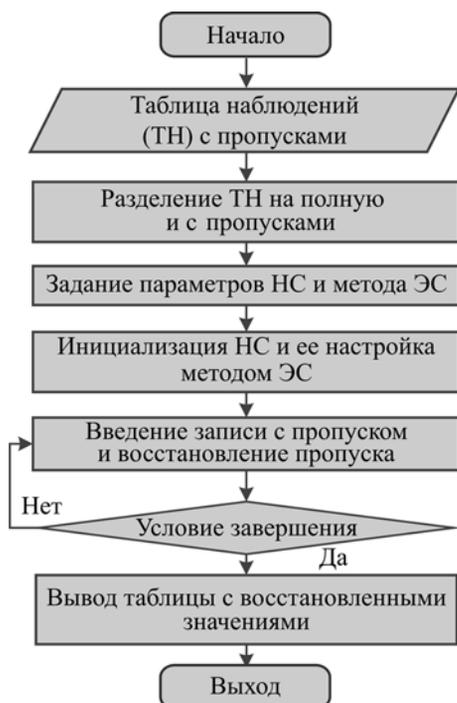


Рис. 1. Алгоритм восстановления данных

менной из таблицы наблюдений $f(x)$ и значениями $F(x)$, полученными нечеткой системой. В работе для оценки результатов вывода НС используется средняя квадратичная ошибка (СКО):

$$E(\theta) = \sqrt{\frac{\sum_i (f(x_i) - F(x_i, \theta))^2}{N}}$$

На рис. 1 представлен алгоритм восстановления данных, основанный на описанной модели.

Далее необходимо проверить "насколько корректно подобрана модель". Так, на **этапе "Проверка адекватности"** полученная модель проверяется с помощью метода скользящего экзамена [6], когда в полную таблицу будут построчно вводиться искусственные пропуски, затем значения будут восстанавливаться согласно предложенному алгоритму восстановления и сравниваться с исходными значениями. Такая проверка позволит сделать выводы о полученной модели и решить, подойдет ли она для дальнейшей обработки данных. Если же модель подходит, наступает следующий **этап — "Восстановление данных"** в полной выборке по имеющейся адекватной модели.

Объект исследования и постановка эксперимента

В условиях роста объемов добычи углеводородного сырья во всем мире и при существенном истощении запасов качественной нефти становятся актуальными вопросы освоения трудноизвлекае-

мых запасов нефти и изучение физико-химических свойств нефти для возможности прогнозирования ее качества в новых районах нефтедобычи [7]. К трудноизвлекаемым относятся сернистые нефти, для которых, согласно классификации нефтей [8], содержание серы от 3 мас. % и выше. Так, в целях апробации указанных методов восстановления пропущенных значений на основе информации из базы данных Института химии нефти СО РАН [9] была сформирована выборка из 548 образцов сернистой нефти по восьми характеристикам. Описание характеристик представлено в табл. 1.

Первоначальный анализ показал, что из всего массива данных общим объемом 4384 значений (матрица 548×8) пропущено 1221 значение (27,85 % общего количества данных). В табл. 2 приведено число пропусков для каждой из восьми характеристик. Данный предварительный анализ информации позволяет оценить возможности восстановления, исключив записи, дающие при анализе сильные статистические смещения.

Общее соотношение пропусков менее 30 %, что позволяет использовать вышеописанный метод восстановления данных с помощью нечеткого моделирования и ожидать успешного результата. Анализ же неполноты отдельных характеристик, таких как температура пласта и пластовое давление показывает, что данные параметры следует тщательно моделировать, так как практически 65 % данных отсутствует, что может привести к восстановлению данных с большой ошибкой вывода системы.

Таблица 1

Характеристики объекта исследования

№	Характеристики сернистой нефти	Среднее значение
1	Плотность, г/см ³	0,9238
2	Вязкость при 20 °С, мм ² /с	590,3
3	Содержание серы, % мас.	3,87
4	Содержание парафинов, % мас.	3,912
5	Содержание смол, % мас.	23,18
6	Содержание асфальтенов, % мас.	7,42
7	Температура пласта, °С	33,28
8	Пластовое давление, МПа	14,79

Таблица 2

Количество пропусков в характеристиках объекта исследования

№	Характеристики нефти	Число пропусков/(%)
1	Плотность, г/см ³	9/(1,64 %)
2	Вязкость при 20 °С, мм ² /с	191/(34,85 %)
3	Содержание серы, % мас.	0/0
4	Содержание парафинов, % мас.	121/(22,08 %)
5	Содержание смол, % мас.	116/(21,16 %)
6	Содержание асфальтенов, % мас.	79/(14,41 %)
7	Температура пласта, °С	349/(63,68 %)
8	Пластовое давление, МПа	356/(64,96 %)
	Общее соотношение (от 4384)	1221/(27,85 %)

Оптимальные параметры нечеткой системы

Число термов на каждый параметр	Средняя квадратичная ошибка вывода нечеткой системы
(5 3 3 6 3 4 3 3)	0,043547
(7 5 3 8 4 4 3 3)	0,051112
(6 4 3 7 4 5 4 3)	0,037304

Примечание. Число термов представлено для характеристик нефти в порядке следования в табл. 1.

Далее следует сформировать "полную" выборку для моделирования. Ввиду числа пропусков, описанных выше, в сформированной "полной" таблице всего 75 записей без пропусков (из 548), на основе которых будет проведено моделирование системы и проверка адекватности подобранной системы.

Моделирование системы

Нечеткая система с заданной структурой [3] предполагает, что для каждого показателя нефти необходимо подобрать оптимальное разбиение функцией принадлежности, за которое отвечает показатель "терм" в структуре нечеткой системы. Подробно способ подбора параметров был описан в работе [10]. Так, для полной выборки из 75 образцов нефти проведено моделирование для поиска наиболее оптимальных параметров нечеткой системы и метода эволюционной стратегии. После ряда экспериментов были получены результаты, часть из которых представлена в табл. 3.

Как видно из табл. 3, из представленных вариантов лучше всего подходит третий, который характеризуется разбиением на термы от 3 до 7. При этом были использованы следующие параметры метода эволюционной стратегии: алгоритм стратегии (20 + 50), 1000 итераций, алгоритм скрещивания — многоточечный, алгоритм селекции — элитарный, вероятность мутации — 0,15.

Результаты проверки построенной модели с помощью метода скользящего экзамена следующие:

- средняя квадратичная ошибка вывода НС — 0,02506;
- точность модели — 96,78 %.

Получив представленный результат, было принято решение, что подобранная модель имеет оптимальные параметры, а следовательно, можно перейти к основной задаче — повышению репрезентативности выборки сернистой нефти, другими словами, с помощью полученной модели необходимо восстановить пропущенные значения в исходной выборке. Результаты средних показателей характеристик нефти до и после восстановления представлены на рис. 2.

Как видно из рис. 2, видимое различие между данными есть только в показателе "Вязкость", значение которого в исследуемой выборке сернистой нефти варьируется от 1,071 до 70077,93 мм²/с, что и объясняет такое расхождение в значениях до и после восстановления. Для более точного восстановления данного параметра необходимо дополнительное исследование и, возможно, привлечение новой информации. Остальные же показатели находятся в допустимом интервале 5 %, что позволяет сделать выводы об успешном восстановлении и возможности дальнейшего использования методов восстановления данных для статистического анализа свойств не только сернистых нефтей, но и других видов трудноизвлекаемых нефтей.

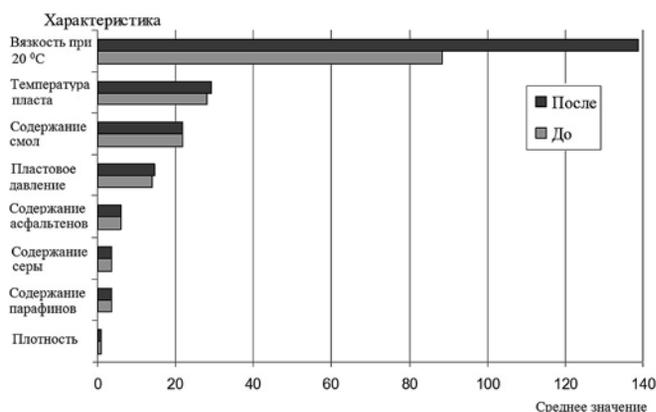


Рис. 2. Средние показатели характеристик сернистой нефти до и после восстановления

Заключение

В работе рассмотрены основные моменты моделирования и восстановления пропущенных данных на основе нечеткого моделирования с использованием информации об основных физико-химических свойствах сернистой нефти. Рассмотрены результаты моделирования объекта исследования в условиях малой размерности выборки. Проведена проверка полученной модели с помощью алгоритма скользящего экзамена, что позволило принять модель и повысить репрезентативность исходной выборки характеристик сернистой нефти. Восстановленная выборка позволит провести последующий анализ характеристик нефти, выявить новые закономерности изменения свойств нефти или подтвердить те, которые были получены в условиях статистической неопределенности.

Список литературы

1. Лучкова С. В., Перемитина Т. О. Применение программного комплекса "Нечеткая система на основе эволюционной стратегии" для задачи импутирования // Информационные технологии. 2013. № 2. С. 47—50.
2. Ходашинский И. А., Гнездилова В. Ю., Дудин П. А., Лавыгина А. В. Основанные на производных и метаэвристические методы идентификации параметров нечетких моделей // Труды VIII международной конференции "Идентификация систем и задачи управления" SICPRO '08 (Москва). 2009. С. 501—528.
3. Bahrami H., Abdechiri M., Meybodi M. R. Imperialist Competitive algorithm with adaptive colonies movement // I. J. Intelligent System and Applications. 2012. N. 2. P. 49—57.

4. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. М.: Горячая линия—Телеком, 2006. 383 с.

5. Hoche S., Wrobel S. A Comparative Evaluation of Feature Set Evolution Strategies for Multirelational Boosting // Proc. 13th Int. Conf. on ILP 2003.

6. Загоруйко Н. Г. Методы распознавания и их применение. М.: Сов. радио, 1972. 216 с.

7. Лисовский Н. Н., Халимов Э. М. О классификации трудноизвлекаемых запасов // Вестник ЦКР Роснедра. 2009. № 6. С. 33—35.

8. Ященко И. Г., Полищук Ю. М. Трудноизвлекаемые нефти: физико-химические свойства и закономерности размещения. Томск: В-Спектр. 2014. 154 с.

9. База данных "химия нефти и газа" [Электронный ресурс]. URL: <http://enrit.jpc.tsc.ru/basel.htm> (дата посещения 15.09.2014).

10. Лучкова С. В., Перемитина Т. О., Ященко И. Г. Программный комплекс восстановления пропущенных значений в многомерных данных на основе методов нечеткого моделирования // Программные продукты и системы. 2014. № 1. С. 80—86.

S. V. Luchkova, Junior Researcher, e-mail; sonetta27@gmail.com, T. O. Peremitina, Researcher, I. G. Yashchenko, Head of Laboratory, Institute of Petroleum Chemistry, Tomsk

Usage Fuzzy Modelling for Raising a Sample Representative on the Example the Analysis of Characteristics Oils

In this paper we consider data recovering process of multidimensional data for sulphurous oils sample with missing values. The algorithm of data recovery is described. Experimental researches results are presented. Results of experimental researches for increasing representative data values based on missing data are shown.

Keywords: fuzzy modeling, data recovered, physic-chemical properties of oils

References

1. Luchkova S. V., Peremitina T. O. Primenenie programmnoho kompleksa "Nechetkaja sistema na osnove jevoljucionnoj strategii" dlja zadachi imputirovanija. *Informacionnye tehnologii*. 2013. N. 2. P. 47—50.

2. Hodashinskij I. A., Gnezdilova V. Ju., Dudin P. A., Lavygina A. V. Osnovannye na proizvodnyh i metajevristicheskie metody identifikacii parametrov nechetkih modelej. *Trudy VIII mezhdunarodnoj konferencii "Identifikacija sistem i zadachi upravlenija" SICPRO '08* (Moscow). 2009. P. 501—528.

3. Bahrami H., Abdechiri M., Meybodi M. R. Imperialist Competitive algorithm with adaptive colonies movement. *I. J. Intelligent System and Applications*. 2012. N. 2. P. 49—57.

4. Rutkovskaja D., Pilin'skij M., Rutkovskij L. Nejrornyie seti, geneticheskie algoritmy i nechetkie sistemy. Moscow: Gorjachaja linija—Telekom, 2006. 383 p.

5. Hoche S., Wrobel S. A Comparative Evaluation of Feature Set Evolution Strategies for Multirelational Boosting. *Proc. 13th Int. Conf. on ILP 2003*.

6. Zagorujko N. G. Metody raspoznavanija i ih primenenie. Moscow: "Sov. radio", 1972. 216 p.

7. Lisovskij N. N., Halimov Je. M. O klassifikacii trudnoizvlekaemyh zapasov. *Vestnik CKR Rosnedra*. 2009. N. 6. P. 33—35.

8. Yashchenko I. G., Polishchuk Yu. M. Trudnoizvlekaemye nefiti: fiziko-himicheskie svojstva i zakonomernosti razmeshhenija. Tomsk: V-Spektr, 2014. 154 p.

9. Baza dannyh "himija nefiti i gaza". URL: <http://enrit.ipc.tsc.ru/base1.htm> (data poseshhenija 15.09.2014)

10. Luchkova S. V., Peremitina T. O., Yashchenko I. G. Programmnyj kompleks vosstanovlenija propushhennyh znachenij v mnogomernyh dannyh na osnove metodov nechetkogo modelirovanija. *Programmnye produkty i sistemy*. 2014. N. 1. P. 80—86.