

Библиографические ссылки

1. Introduction to aluminium electrolysis Understanding the Hall-Heroult process / K. Grjotheim, H. Kvande (ed.). Dusseldorf : Aluminium-Verlag, 1993.
2. Минцис М. Я., Поляков П. В., Сиразутдинов Г. А. Электрометаллургия алюминия. Новосибирск : Наука, 2001.
3. Grjotheim K., Welch B. J. Aluminium Smelter Technology. Dusseldorf : Aluminium-Verl. GmbH, 1980.
4. Светуных С. Г., Светуных И. С. Методы социально-экономического прогнозирования : учебник для вузов : в 2 т. Т. 1. СПб. : Изд-во СПбГУЭФ, 2009.
5. Ивченко Г. И., Медведев Ю. И. Введение в математическую статистику : учебник. М. : Изд-во ЛКИ, 2010.
6. Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. М. : Финансы и статистика, 2002.
7. Analysis of a potroom performance drift, from a multivariate point of view / J. Tessier, C. Duchesne, G. P. Tarcy et al. // Light Metals. 2008. P. 319–324.
8. Multivariate statistical process monitoring of reduction cell / J. Tessier, T. G. Zwirz, G. P. Tarcy, R. A. Manzini // Light Metals. 2009. P. 305–310.
9. Detecting abnormalities in aluminium reduction cells based on process events using multi-way principal component analysis (MPCA) / N. A. A. Majid, B. R. Young, M. P. Taylor, J. J. J. Chen // Light Metals. 2009. P. 589–593.
10. Multivariate monitoring of the prebaked anode manufacturing process and anode quality / J. Lauzon-Gauthier, C. Duchesne, J. Tessier et al. // Light Metals. 2011. P. 967–972.
11. A Taxonomy for Aluminium Process Fault Detection and Diagnosis / N. A. A. Majid, M. P. Taylor, J. J. J. Chen, B. R. Young // Proc. of the 10th Australasian Aluminium Smelting Technology Conf. Launceston. 2011.
12. Multiblock monitoring of aluminium reduction cells performance / J. Tessier, C. Duchesne, G. P. Tarcy // Light Metals. 2011. P. 407–412.

T. V. Piskazhova, P. V. Polyakov, N. A. Sharypov

USAGE OF METHODS OF MULTI-DIMENSIONAL DATA PROCESSING FOR ANALYSIS OF ENERGY CONSUMPTION OF ALUMINUM REDUCTION CELLS

The authors present a review of methods of application of multi-dimensional statistical analysis of database of aluminum industry. Estimation of effectiveness of devices and quality of products output is made. Application of major components for the analysis of energy consumption of two potrooms of an aluminum smelter is considered.

Keywords: multi-dimensional statistical analysis, major components analysis, aluminum reduction cell, energy consumption.

© Пискажова Т. В., Поляков П. В., Шарыпов Н. А., 2012

УДК 519.6

Е. А. Попов, М. Е. Семенкина, Л. В. Липинский

ПРИНЯТИЕ РЕШЕНИЙ КОЛЛЕКТИВОМ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ*

Предлагается алгоритм генетического программирования для автоматического генерирования искусственных нейронных сетей. Предложена процедура принятия решений коллективом нейронных сетей, основанная на оценке компетентности членов коллектива. Подход апробирован на тестовых и реальных практических задачах.

Ключевые слова: интеллектуальные информационные технологии, алгоритм генетического программирования, коллективное принятие решений.

При математическом моделировании сложных систем и процессов часто используются так называемые интеллектуальные информационные технологии (ИИТ) – генетические алгоритмы (ГА) как метод оптимизации [1], искусственные нейронные сети (НС) как метод аппроксимации или классификации [2], генетическое программирование (ГП) как метод ав-

томатического генерирования математических выражений [3].

Проектирование интеллектуальных информационных технологий само по себе является достаточно сложной интеллектуальной процедурой, включающей, кроме всего прочего, выбор их эффективных структур.

* Работа выполнена при поддержке Министерства образования и науки Российской Федерации (ГК № 11.519.11.4002 и ГК № 16.740.11.0742).

Так, при проектировании ГА необходимо производить выбор его оптимальных настроек, при создании НС – выбирать структуру нейросети (количество слоев, количество нейронов на слое, типы функций активации, наличие связей между нейронами), для построения математических моделей в ГП – выбирать структуру функциональной зависимости. Сложность проектирования ИИТ препятствует действительно массовому их внедрению в производство и другие сферы деятельности человека. Автоматизация проектирования ИИТ смогла бы решить часть трудностей, возникающих при их разработке. Но для автоматизации проектирования ИИТ должны использоваться оптимизационные процедуры, позволяющие осуществлять комбинаторный поиск на сложных структурах. Так как обычные методы математического программирования здесь не работают, применяют методы эволюционного поиска. Известно, что эволюционные алгоритмы требуют значительных вычислительных ресурсов, особенно при их использовании для решения вычислительно сложных задач, т. е. задач, требующих большого количества ресурсов независимо от алгоритма, используемого для их решения. Однако в большинстве практических случаев эта проблема может быть решена за счет использования параллельных и распределенных вычислительных систем [4; 5]. Имея возможность автоматически генерировать алгоритмическое ядро ИИТ, можно расширить класс используемых подходов, в частности – перейти к автоматическому формированию коллективов, состоящих из различных ИИТ, что позволит повысить эффективность получаемых систем анализа данных.

Повышение эффективности разработки и использования ГП, ИНС и ГА даст возможность решать более сложные практические задачи интеллектуального анализа данных. В этой связи разработка и исследование новых алгоритмических схем для автоматизации проектирования ИИТ является актуальной научной задачей.

Алгоритм генетического программирования для автоматического генерирования искусственных нейросетей. Стандартный алгоритм ГП применяется для задач символьной регрессии, т. е. задач одновременного определения точной функциональной зависимости и соответствующих численных ко-

эффициентов. Такие задачи отличаются от обычных регрессионных задач, в которых форма функциональной зависимости (линейная, квадратичная, колоколообразная, ряд Фурье и т. п.) определяется заранее. Модификация стандартного алгоритма ГП позволяет применить его для проектирования произвольных структур нейронных сетей.

При решении задачи символьной регрессии методом генетического программирования обычно используется набор из элементарных математических функций и логических выражений. Для настройки структуры НС при помощи алгоритма ГП в терминальное множество были включены 16 функций активации, а также входные нейроны. При генерации нейросетей алгоритм ГП имеет дело не с числами, а с нейронами, поэтому и допустимые операции из функционального множества специфические:

- постановка нейронов (блоков нейронов) в один слой, являющаяся ассоциативной (обозначим «+»);
- постановка нейронов (блоков нейронов) в последовательные слои, так что нейрон (блок нейронов), пришедший из левой ветви дерева, предшествует нейрону (блоку нейронов), пришедшему из правой ветви дерева; такая постановка не является ассоциативной (обозначим «<»).

При выполнении первой операции новых связей между нейронами не появляется, при выполнении второй операции выходы нейронов из левой ветви подаются на вход нейронам из правой ветви. Важно учитывать, что входные нейроны не должны принимать на вход результаты вычисления сети или ее части, но хотя бы один из них должен подавать сигнал на вход. Поэтому было принято следующее решение: в левой ветви могут присутствовать только входные нейроны и первая операция, а в вершине дерева должна находиться вторая операция. Сборка дерева начинается справа. Пример преобразования дерева в нейронную сеть см. на рис. 1.

Операции рекомбинации и мутации были модифицированы в соответствии с ограничениями, наложенными на вид дерева.

Для настройки весовых коэффициентов полученной нейронной сети применялся модифицированный генетический алгоритм [6] с последующим локальным спуском.

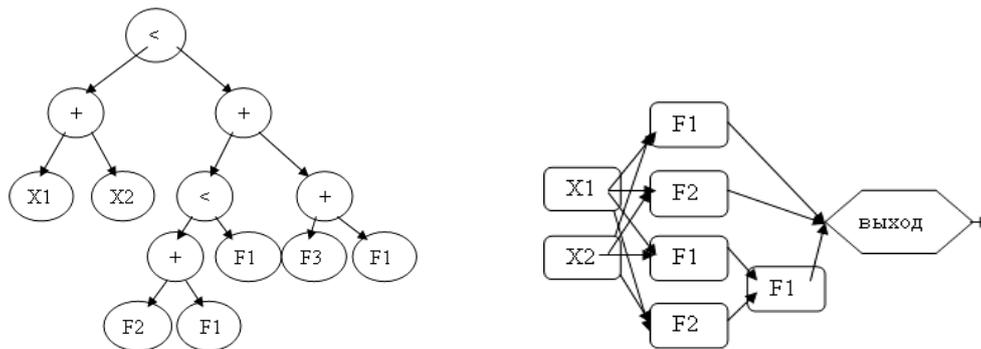


Рис. 1. Пример дерева, сгенерированного представленным алгоритмом генетического программирования, и соответствующей ему нейронной сети

Работоспособность предложенного подхода проверялась на репрезентативном наборе тестовых задач (аппроксимация, прогнозирование временных рядов). Проводилось усреднение стандартной среднеквадратической ошибки нейронной сети по 20 прогонам. Терминальное множество включало в себя 16 функций активации и входные переменные.

Результаты тестирования показали, что нейронные сети, созданные алгоритмом генетического программирования, имеют небольшое количество нейронов по сравнению с нейросетями, получаемыми посредством нейроимитаторов, и не являются полностью связными (связей между нейронами мало). При этом данные НС обладают достаточно малой среднеквадратической ошибкой.

Процедура принятия решений коллективом интеллектуальных информационных технологий. Для повышения эффективности применения интеллектуальных информационных технологий предлагается при решении задачи формировать коллектив из нескольких ИИТ. Например, из трех нейронных сетей, полученных алгоритмом генетического программирования для автоматического генерирования нейросетевых моделей, и четырех символьных выражений, полученных алгоритмом генетического программирования для решения задач символьной регрессии.

Для формирования решения коллективом ИИТ возможны несколько подходов. Наиболее часто применяемыми являются метод усреднения [7] и его модификация, учитывающая ошибки каждого из членов коллектива [8]. Решения формируются по следующим схемам.

Схема 1. Выходные значения ИИТ складываются и делятся на число членов коллектива.

Схема 2. Метод, основанный на «оценке компетентности» ИИТ. Для каждого члена коллектива на основании его ошибки, полученной на обучающей выборке, назначается весовой коэффициент, обратно пропорциональный ошибке [9]. Выходное значение ИИТ умножается на весовой коэффициент ИИТ, после чего все полученные значения складываются и делятся на их число.

Заметим, что в данных подходах мнение наиболее компетентного члена коллектива «разбавляется» мнением менее компетентных, что дает повод к сомнению в эффективности этих подходов в конкретных задачах.

Предлагается новая схема учета мнений членов коллектива ИИТ, гарантирующая ошибку не больше той, которую имеет лучшая из ИИТ, участвующих в принятии решений.

Схема 3. Полученный коллектив принимает решение относительно каждой точки, в которой запрашивается решение задачи (выходное значение, вычисленное членами коллектива), следующим образом:

– производится поиск точки из обучающей выборки, которая находится ближе остальных к точке, в которой необходимо получить значение;

– в найденной точке обучающей выборки определяется ИИТ с наименьшей среднеквадратической ошибкой;

– выбранной ИИТ предоставляется право принятия решения относительно значения в интересующей нас точке.

Предложенная процедура была реализована в виде программной системы и проверена на репрезентативном множестве тестовых задач. Было также проведено сравнение трех описанных схем. Выбор состава коллектива (количество ИИТ каждого вида) остается на усмотрение пользователя.

Апробация на практических задачах. Для апробации предложенного подхода были использованы следующие практические задачи:

– моделирование прожиточного минимума в регионах Российской Федерации (3 входа, 1 выход, обучающая выборка – 80 единиц, контрольная – 9);

– моделирование вибрационных характеристик гидротурбины (12 входов, 11 выходов, обучающая выборка – 906 единиц, контрольная – 93 точки);

– моделирование системы подавления шума (1 вход, 1 выход, данные сильно зашумлены, обучающая выборка – 250 единиц, контрольная – 250 точек).

При решении задачи моделирования прожиточного минимума регионов Российской Федерации коллективом интеллектуальных информационных технологий, коллектив состоял из двух нейронных сетей со среднеквадратическими ошибками, равными 0,008 725 7 и 0,008 967, и двух символьных выражений со среднеквадратическими ошибками 0,086 и 0,0305. Коллектив, работающий по первой схеме, допустил среднеквадратическую ошибку, равную 0,016 466, работающий по второй схеме – 0,024 143 1, работающий по третьей схеме – 0,007 748. Только третья схема обеспечила ошибку не ниже, чем у лучшей из ИИТ в отдельности. В этом коллективе худшие нейронная сеть и символьное выражение не внесли вклад в решение задачи. Коллектив, составленный из двух нейронных сетей, работающий по третьей схеме, имеет ту же среднеквадратическую ошибку, что и лучшая из них. Коллектив, составленный только из двух символьных выражений, работающий по третьей схеме, имеет ошибку, равную 0,029 9. Можно сказать, что при решении данной задачи имело бы смысл включить в коллектив только лучшую нейронную сеть и лучшее символьное выражение.

Для апробации на задаче моделирования вибрационных характеристик гидротурбины коллектив интеллектуальных информационных технологий состоял из двух нейронных сетей с ошибками 0,012 917 (рис. 2) и 0,013 444 и двух символьных выражений со среднеквадратическими ошибками 0,017 04 и 0,019 84.

Коллектив, работающий по первой схеме, допустил ошибку 0,013 45, работающий по второй схеме – 0,013 643, работающий по третьей схеме – 0,010 034. Коллектив, составленный из двух нейронных сетей, работающий по третьей схеме, имеет среднеквадратическую ошибку 0,011 746. Коллектив, составленный из лучших символьного выражения и нейросети, работающий по третьей схеме, имеет ошибку, равную 0,010 819.

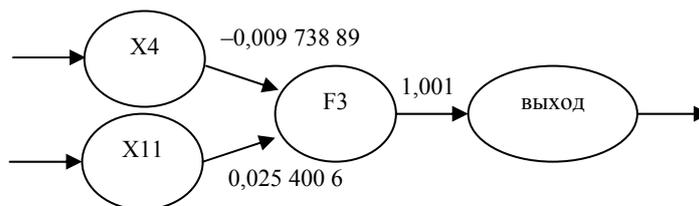


Рис. 2. Нейронная сеть, полученная при решении задачи моделирования вибрационных характеристик гидротурбины

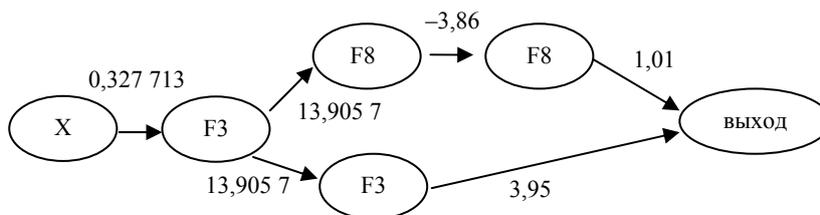


Рис. 3. Нейронная сеть, полученная при решении задачи моделирования системы подавления шума

Коллектив, состоящий из обеих нейросетей и лучшего символического выражения, имеет ошибку ту же, что и коллектив, состоящий из всех четырех ИИТ. Можно сказать, что при решении данной задачи имело бы смысл включить в коллектив только лучшее символическое выражение и обе нейронные сети.

При решении задачи моделирования системы подавления шума коллективом интеллектуальных информационных технологий, коллектив состоял из двух нейронных сетей с ошибками 0,684 407 (рис. 3) и 0,686 113 соответственно, и двух символических выражений со среднеквадратическими ошибками 0,493 897 и 0,493 553 соответственно.

Коллектив, работающий по первой схеме, допустил ошибку 0,532 637, работающий по второй схеме – 0,551 447, работающий по третьей схеме – 0,407 085. Коллектив, составленный только из этих двух нейронных сетей, работающий по третьей схеме, имеет среднеквадратическую ошибку 0,662 504. Коллектив, составленный только из лучших символического выражения и нейросети, работающий по третьей схеме, имеет ошибку, равную 0,420 15.

Коллектив, состоящий из обеих нейросетей и лучшего символического выражения, имеет ошибку 0,481 581. Можно сказать, что при решении данной задачи имело бы смысл включить в коллектив все четыре ИИТ.

Таким образом, разработанная программная система позволяет формировать коллективы автоматически генерируемых интеллектуальных информационных технологий, т. е. дает возможность решать сложные задачи анализа данных. Создание коллектива увеличивает эффективность и надежность применения ИИТ. Работоспособность данного подхода была проверена не только на тестовых, но и на реальных практических задачах из различных областей человеческой деятельности, что позволяет рекомендовать их конечным пользователям в качестве удобного и эф-

фективного инструмента интеллектуального анализа данных.

Библиографические ссылки

1. Eiben A. E., Smith J. E. Introduction to Evolutionary Computing. Springer Verlag Berlin, 2003.
2. Оссовский С. Нейронные сети для обработки информации / пер. с пол. И. Д. Рудинского. М. : Финансы и статистика, 2002.
3. Koza J. R. Genetic Programming: On Programming Computer by Means of Natural Selection and Genetics. Cambridge, MA : The MIT Press, 1992.
4. Ефимов С. Н., Семенкин Е. С., Тынченко В. В. Формирование грид-системы эффективной архитектуры для распределенного решения сложных задач // Системы управления и информ. технологии. 2009. № 2 (36). С. 4–7.
5. Формализация задач выбора эффективного варианта распределенных систем управления / С. Н. Ефимов, И. А. Панфилов, Е. С. Семенкин, В. А. Терсков // Вестник СибГАУ. 2003. Вып. 4. С. 24–31.
6. Семенкин Е. С., Семенкина М. Е. Применение генетического алгоритма с модифицированным оператором равномерной рекомбинации при автоматизированном формировании интеллектуальных информационных технологий // Вестник СибГАУ. 2007. Вып. 3 (16). С. 27–33.
7. Hansen L. K., Salamon P. Neural network ensembles // IEEE Trans. Pattern Analysis and Machine Intelligence. 1990. № 12 (10). P. 993–1001.
8. Jimenez D. Dynamically weighted ensemble neural networks for classification // Proc. IJCNN-98. Anchorage, 1998. P. 753–756.
9. Cooper L. N., Perrone M. P. When networks disagree: ensemble method for neural networks // R. J. Mammone (ed.) // Artificial Neural Networks for Speech and Vision. N. Y. : Chapman & Hall, 1993. P. 126–142.

E. A. Popov, M. E. Semenkina, L. V. Lipinskiy

DECISION MAKING WITH INTELLIGENT INFORMATION TECHNOLOGY ENSEMBLE

Genetic programming algorithm for neural network automatic design is suggested. Ensemble member competence estimations, based on decision making with intelligent information technologies, is proposed. The approach effectiveness is approved on qualificatory and real-world problems.

Keywords: intelligent information technologies, genetic programming algorithm, group decision making.

© Попов Е. А., Семенкина М. Е., Липинский Л. В., 2012

УДК 517.977.1

А. Н. Рогалев, А. А. Рогалев

ВЫЧИСЛЕНИЕ ПОЛОЖЕНИЯ ИСКУССТВЕННОГО СПУТНИКА ЗЕМЛИ ПРИ ОГРАНИЧЕНИЯХ НА ОШИБКИ ИЗМЕРЕНИЙ

Рассматривается применение метода наименьших квадратов и минимаксного метода для определения положения объекта околоземного пространства при известных ограничениях на ошибки измерений. Исследованы вычислительные схем, используемые для оценки значений траекторий космического объекта при различных условиях, налагаемых на ошибки измерений. Определены характеристики точности данных алгоритмов при различных распределениях ошибок измерений. Приводятся результаты численных расчетов, полученные на модельных примерах.

Ключевые слова: метод наименьших квадратов, минимаксный метод, фильтрация данных.

В задачах оценки параметров объектов астрометрии и космической геодезии необходимо анализировать экспериментальные данные, полученные с помощью измерений [1–4]. Поскольку позиционные наблюдения, как правило, являются косвенными, то результаты (данные) часто представляются алгебраическими моделями с неизвестными коэффициентами, которые являются искомыми параметрами. При оценивании этих параметров в условиях, когда данные наблюдений получены с ошибками, применяются подходы, основанные на методе наименьших квадратов [4–6] и минимаксном методе [4; 7–10].

Для высокоточного определения орбиты по измерениям традиционно используется классический метод наименьших квадратов (МНК), либо какая-то его модификация. Обоснованием для такого решения являются оптимальные свойства оценок параметров, получаемых этим методом. При некоррелированных во времени ошибках измерений, имеющих гауссово распределение, оценка МНК среди всех оценок (как линейных, так и нелинейных) наиболее точна. Однако при коррелированных во времени ошибках измерений, при негауссовых ошибках, а также при неизвестных (частично или полностью) вероятностных характеристиках ошибок измерений традиционные методы могут являться не самыми точными и не давать корректной оценки точности получаемых значений. Для многих задач ошибки единичных измерений объектов в космическом пространстве ограничены сверху определенными константами и имеют неизвестное рас-

пределение, которое хорошо аппроксимируется равномерным законом. Эта постановка не соответствует предположениям, при которых доказаны оптимальные свойства оценок МНК.

Проводится сравнительный анализ нескольких вычислительных схем, построенных на основе применения МНК и минимаксного метода (МНМ) в задаче вычисления положения искусственного спутника Земли при ограничениях на ошибки измерений. Следует отметить, что значимость этих результатов определяется тем, что не всегда ясно, когда эффективен тот или иной подход к решению задачи: классический, при котором характеристики случайных ошибок исходных данных основываются на методе максимального правдоподобия, и неклассический (минимаксный), характеризуемый заданием лишь некоторых множеств, которым принадлежат ошибки или их характеристики. Во втором случае интересно также сравнение возможностей минимаксного и гарантированного подходов [11; 12] к оценке решений управляемых систем.

Искусственный спутник Земли (ИСЗ) – космический аппарат (КА), вращающийся вокруг Земли по геоцентрической орбите, т. е. по эллиптической траектории вокруг Земли. Один из двух фокусов эллипса, по которому движется небесное тело, совпадает с Землей.

Предположим, что движение центра масс КА в околоземном пространстве с достаточной степенью точности и адекватности описывается следующей