

УДК 533.9.924: 621.785.05

ПРИМЕНЕНИЕ ЭКСПЕРТНЫХ СИСТЕМ ДЛЯ ИССЛЕДОВАНИЯ И ОПТИМИЗАЦИИ ПРОЦЕССОВ ЗАКАЛКИ

д-р техн. наук, проф. С.П. КУНДАС

(Международный государственный экологический университет им. А.Д. Сахарова, Минск),

В.И. КОВАЛЕНКО, Д.Г. ИВАНОВ

(Белорусский государственный университет информатики и радиоэлектроники, Минск)

Предложено использование экспертных систем для исследования и оптимизации процессов термообработки, и прежде всего наиболее широко распространенного метода термообработки - закалки. Показано, что наиболее эффективным в рассматриваемом случае является применение гибридных систем на основе многослойных нейронных сетей, с предварительной классификацией входной информации с помощью экспертных систем.

Введение. Математическое моделирование широко применяется в настоящее время в большинстве отраслей промышленности для решения задач инженерного анализа конструкций и технологических процессов при их автоматизированном проектировании [1].

Использование методов компьютерного моделирования при разработке и совершенствовании процессов термообработки позволяет существенно сократить затраты времени и материальных средств на технологическую подготовку производства деталей [1, 2]. Известно, что для формирования эксплуатационных свойств деталей машин наиболее широкое применение находит процесс закалки, который характеризуется высокими скоростями изменения температуры и большим градиентом температур в технологической зоне, а также многообразием теплофизических процессов и явлений, происходящих в материале. Все это обуславливает сложность и трудоемкость экспериментальных методов исследования и оптимизации процессов закалки при технологической подготовке производства и связано с большим расходом материалов вследствие необходимости использования разрушающих методов контроля. Поэтому разработка современных методов компьютерного моделирования процессов термообработки является актуальной научной и практической задачей.

Разработанные модели и программные комплексы для инженерного анализа процессов закалки решают в основном прямую задачу моделирования, т.е. нахождение свойств деталей после закалки исходя из технологических режимов термообработки деталей [3]. Однако в системах автоматизированного проектирования часто требуется решение и обратной задачи моделирования - определение по требуемым свойствам деталей после закалки необходимых режимов термообработки. Эту задачу достаточно сложно решить с использованием традиционных подходов к моделированию и известных численных методов (метод конечных разностей, метод конечных элементов и т.п.). Как показали проведенные нами исследования, применительно к процессу плазменного напыления покрытий, обратная задача компьютерного моделирования технологических процессов может быть достаточно эффективно решена с применением экспертных систем и нейронных сетей [4].

Структура и принципы построения экспертной системы для инженерного анализа процесса закалки. В настоящее время существует три подхода для построения экспертных систем [5]. Первый подход является традиционным. Он использует методы логических рассуждений и символической обработки информации. Второй подход связан с использованием нейронных сетей. Он опирается на биологические основы естественного интеллекта и позволяет проектировать системы, способные к обучению и самоорганизации. Третий подход является комбинацией первого и второго. Применительно к решению задач проектирования процесса закалки (в том числе и обратных) более эффективным является применение третьего подхода [5]. Это прежде всего связано со сложной динамикой процесса и множеством коррелирующих между собой параметров.

Если использовать для поставленных целей классическую нейронную сеть, то в этом случае появляются проблемы, связанные с тем, что такая сеть не располагает достаточными возможностями для работы с любой выборкой входной информации. Поэтому необходима их предварительная подготовка для последующего использования. Целесообразным является проведение предварительной классификации входной информации, установление определенных закономерностей и зависимостей, что может быть выполнено с помощью экспертной системы.

Разработка экспертной системы обычно включает шесть следующих этапов: идентификацию, концептуализацию, формализацию, выполнение, тестирование, опытную эксплуатацию [5].

На этапе идентификации определяются задачи, которые подлежат решению, выявляются цели разработки. Одной из важных задач настоящей работы является разработка вектора рассуждений гиб-

ридной экспертной системы, с целью её последующей самоорганизации. В дальнейшем система будет модифицироваться и оптимизироваться с целью увеличения достоверности результатов. При этом необходим начальный экспертный анализ информации с целью выявления наиболее критичных аспектов, изменение которых даст желаемый результат с выходов нейронной сети.

На этапе концептуализации проводится содержательный анализ проблемной области, выявляются используемые понятия и их взаимосвязи, определяются методы решения задачи. Проведём концептуализацию, применительно к процессу закалки. В качестве входных данных будет использоваться информация о свойствах детали и ее геометрии, а также технологические режимы термообработки, включая условия теплообмена с охлаждающей средой. Эту информацию можно классифицировать следующим образом:

1. Свойства материалов:

1.1. Теплофизические: плотность, теплоёмкость, теплопроводность.

1.2. Механические: коэффициент Пуассона, модуль Юнга, коэффициент упрочнения, предел текучести.

2. Термокинетическая диаграмма.

3. Изотермическая диаграмма.

4. Параметры охлаждающей среды: коэффициент теплообмена; температура среды.

5. Геометрия детали, заданная в STL-формате.

6. Начальное значение температуры детали.

7. Конечное значение температуры детали.

Анализ этой информации показывает, что она обладает значительной избыточностью и не может быть применима в качестве входов нейронной сети. Необходимо также учитывать, входная информация нейронной сети должна быть нормированной в пределах от 0 до 1.

После анализа и нормализации данных получаем следующую их модификацию:

1. Материал детали.

2. Коэффициент теплообмена среды.

3. Температура среды.

4. Температура детали.

Проблемной областью в рассматриваемом случае является классификация геометрии детали. Отдельные авторы [6] используют сходную структуру входного множества, подаваемого на нейронную сеть. В качестве геометрического параметра используется диаметр цилиндра, а геометрия детали представлена в виде цилиндров различных диаметров. Возможности задания произвольных геометрических размеров детали при этом не учитываются.

При классическом представлении геометрии в декартовой системе координат имеет место бесконечное множество решений для каждой координаты точки, и при этом возникают трудности в получении информации о положении точки относительно границ детали. Существуют способы ограничения пространства с помощью специальных граничных плоскостей. Однако в этом случае значительно увеличивается объем входной информации нейронной сети.

Для решения этой проблемы предложена модель представления геометрии описанная следующим алгоритмом:

- деталь и окружающая среда разбиваются на конечные элементы с постоянным шагом;
- определяются плоскости закрепления;
- вводится понятие слоёв (layer). Каждый слой N представляет собой поверхность N -го уровня, отстоящего от предыдущей и последующей поверхности на шаг разбиения.

Получаем в общей сложности четыре вида слоев:

- поверхностный слой;
- слой закрепления 1;
- слой закрепления 2;
- слой закрепления 3.

Наглядное представление модели показано на рис. 1.

Три последних слоя перпендикулярны друг другу и характеризуют направления распространения напряжений и деформаций.

В результате такого подхода геометрические особенности детали описываются следующей информацией:

- номер поверхностного слоя;
- номер слоя закрепления 1;
- номер слоя закрепления 2;
- номер слоя закрепления 3.

Это позволяет идентифицировать все равноудалённые от плоскостей закрепления и поверхности точки детали.

Следующей задачей является определение информации, которая будет представлена в процессе обучения нейронной сети в качестве выходной. При анализе напряженно-деформированного состояния такой информацией будет тензор напряжений и деформаций. Как и для входной информации, ее необхо-

димо нормировать к величине от 0 до 1. Оси X, Y, Z будут характеризоваться слоями закрепления 1, 2, 3 соответственно.

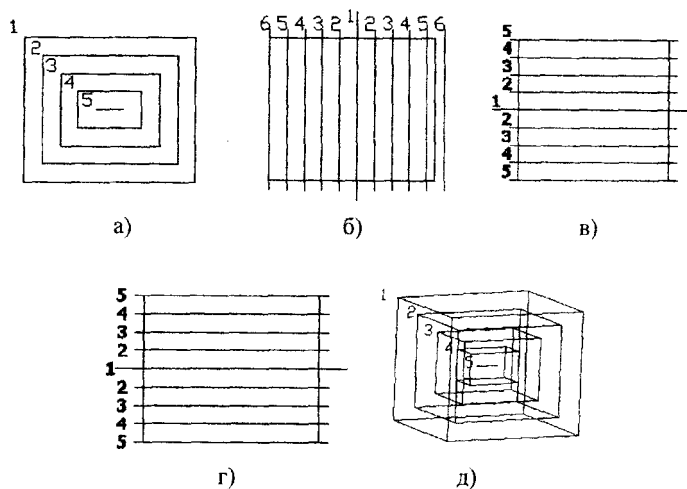


Рис. 1. Представление поверхностного слоя детали (а), (д) и слоёв закрепления по осям X (б), Y (г), Z (в) для нейронной сети

В результате проведенного анализа установлены 8 известных параметров и 12 неизвестных параметров. Эти параметры могут быть представлены как единица информации в базе знаний. Но для нормального функционирования экспертной системы необходимо организовать систему ввода, хранения и выборки информации, а также подготовить инструментальные средства и специализированные правила по работе с приобретаемыми знаниями. Большинство современных инструментальных средств, как правило, используют объектно-ориентированную парадигму, объединённую с парадигмой, ориентированной на правила [5]. Представим объектно-ориентированные модели препроцессора и постпроцессора [7], выполняющие функции инструментальных средств в экспертной системе.

Формализованный в виде класса «Препроцессор» выполняет задачи нормализации, выборки и подготовки входного множества сигналов. Он включает в себя решение таких вопросов, как анализ файла данных; анализ геометрии; экспорт данных; генерация нормируемого входного множества.

Была произведена объектно-ориентированная декомпозиция этого класса, в результате которой круг возложенных на него задач был распределён между следующими подклассами:

1. «Экспортёр» - выполняет экспорт информации из базы знаний, которые будут представлены в нейронной сети в качестве входных. При этом набор информации будет выбираться для каждого участка детали.

2. «Геометрия» - предназначен для работы с геометрическими особенностями детали, включает в себя 3 подкласса: «Экспортёр детали» (выполняет функции выборки основных геометрических особенностей детали); «Конструктор слоёв» (оценивает информацию о геометрии детали и выполняет функцию разбиения детали на слои); «Мешер» (предназначен для разбиения детали и окружающей среды на кубические конечные элементы с постоянным шагом).

3. «Анализатор» - объединяет информацию, полученную из класса «Геометрия» и класса «Экспортёр», и генерирует набор входных параметров для каждого узла разбитой на конечные элементы детали.

4. «Генератор входного множества» - производит выборку входных параметров для каждого узла детали в соответствии с алгоритмом, описанным выше.

Модуль, выполняющий анализ файла данных, представляет собой класс «Постпроцессор». Он представлен в виде следующей иерархии классов:

1. «Импорт результатов» - выполняет задачи выборки рассчитанной информации и её нормализации.

2. «Геометрия» - представляет собой класс, включающий подклассы «Экспортёр слоёв», «Преобразователь геометрии», «Конструктор координат». При начальной инициализации экземпляр данного класса выполняет функции сопоставления узлов конечных элементов координатам, а также создаёт структуру слоёв.

3. «Анализатор» - производит сопоставление рассчитанных параметров в узлах детали координатам данных узлов.

4. «База знаний» - представляет собой набор выходной с нейронной сети информации.

5. «Программный комплекс» - представлен в виде класса, организующего интерфейсы для работы с комплексом согласно архитектуре, описанной в [7].

Экспертный анализ реализации нейронных сетей для исследования и оптимизации процессов закалки. Как отмечалось выше, этап формализации при разработке экспертной системы возложен на нейронную сеть. Ввиду того, что исследуемый процесс динамичен и достаточно сложен (в результате чего имеется большой набор входных параметров) рекомендовано использовать многослойную нейронную сеть [8]. Известно [9], что уже трёхслойная нейронная сеть способна осуществлять любое отображение входных векторов в выходные, а с увеличением числа слоёв увеличивается время на обучение системы и усложняется алгоритм обучения.

Исходя из вышеизложенного, для моделирования напряженно-деформированного состояния деталей при закалке предложена структура нейронной сети, приведенная на рис. 2.

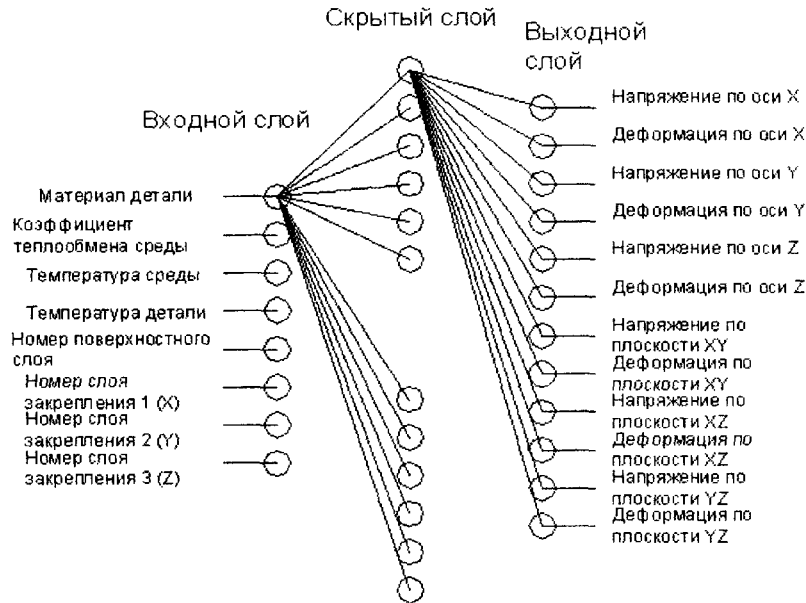


Рис. 2. Структура нейронной сети

Работа каждого нейрона сети описывается следующим выражением:

$$Y = F(F(XW_1 - T_1)W_2 - T_2) - T_3, \tag{1}$$

где $X = (x_1, x_2, \dots, x_n)$ – вектор-строка входных сигналов для $n = 8$; F – оператор нелинейного преобразования; W_i – матрица весовых коэффициентов i -того слоя многослойной нейронной сети; T_i – порог i -того слоя.

В качестве нелинейной функции активации выбираем сигмоидную функцию, выходное значение которой описывается выражением:

$$y_j = \frac{1}{1 + e^{-S_j}}, \tag{2}$$

$$S_j = \sum_i w_{ij} y_i - T_j, \tag{3}$$

где y_j – выходное значение j -того нейронного элемента; w_{ij} – i -тый весовой коэффициент j -того нейрона; T_j – пороговое значение j -того нейрона.

Такой выбор обусловлен простотой вычисления производной функции по S_j , что сокращает вычислительные операции по сравнению с гиперболическим тангенсом и биполярной сигмоидной функцией, в существующих разработках [8, 9].

Одной из важнейших задач при разработке многослойной нейронной сети является её обучение. Рядом авторов независимо друг от друга был предложен алгоритм обратного распространения ошибки, описанный в работе [10]. В его основе лежит метод градиентного спуска в пространстве весовых коэффициентов и порогов нейронной сети.

Основу обучения составляют прямой и обратный проходы. Прямой проход представляет собой прямое вычисление выходов, определяемое выражением (1). Обратный проход заключается в подстройке весов текущего слоя, в зависимости от ошибки последующего слоя.

Подстройка весов выходного слоя k отличается от аналогичной задачи для скрытого слоя. Это связано с тем, что в выходном слое величина ошибки вычисляется явно – выход нейрона слоя k , вычитаясь из

целевого значения t_j , дает сигнал ошибки. Он умножается на производную сжимающей функции $[y_j(1 - y_j)]$, вычисленную для этого нейрона слоя k , давая, таким образом, величину δ .

$$\delta = y_j(1 - y_j) \cdot (t_j - y_j), \quad (4)$$

где δ – величина ошибки выходного слоя; y_j – рассчитанные значения выходов.

Величина δ , необходимая для нейрона скрытого слоя, получается суммированием всех таких произведений и умножением на производную сжимающей функции:

$$\delta_j = y_j \cdot (1 - y_j) \cdot \left[\sum_q \delta_{q,k} w_{q,k} \right], \quad (5)$$

где δ_j – величина передаваемой ошибки текущего слоя; $\delta_{q,k}$ – величина передаваемой ошибки последующего слоя k нейронного элемента q ; $w_{q,k}$ – значение весового коэффициента нейрона q слоя k .

Умножая δ на величину y_j нейрона j , можно получить значение весового коэффициента. Это произведение в свою очередь умножается на коэффициент скорости обучения η (обычно от 0,01 до 1,0), и результат прибавляется к весу. Такая же процедура выполняется для каждого веса от нейрона скрытого слоя к нейрону в выходном слое.

Представленные ниже уравнения иллюстрируют это вычисление:

$$\Delta w_k = \eta \cdot \delta_j \cdot y_k; \quad (6)$$

$$w_k(n+1) = \Delta w_k + w_k(n), \quad (7)$$

где $w_k(n)$ – величина весового коэффициента одного из нейронов в предыдущем слое на шаге n , до коррекции, (отметим, что индекс k относится к слою, в котором заканчивается данный вес, т.е. с которым он объединен); $w_k(n+1)$ – величина веса на шаге $n+1$ (после коррекции); δ_j – величина δ для нейрона, в последующем слое j ; y_j – величина выхода нейрона в слое j .

Проведённый анализ показал, что применение приведенного выше алгоритма связано с рядом проблем:

- неопределенностью в выборе числа слоёв и количества нейронов в слое;
- медленной сходимостью градиентного метода с постоянным шагом;
- сложностью получения требуемой скорости обучения;
- невозможностью определения точек локального и глобального минимума;
- влиянием случайной инициализации весовых коэффициентов нейронной сети на поиск минимума функции среднеквадратической ошибки.

Однако, несмотря на указанные недостатки, этот алгоритм в настоящее время является одним из лучших. Поэтому его использование является целесообразным при условии минимизации приведенных выше недостатков. С этой целью авторами работ [8, 9] сделаны попытки разработки модифицированных алгоритмов.

Проведенный анализ показывает, что важную роль играет размер случайно инициализируемых синаптических связей. Проведенные исследования показывают, что наиболее эффективным является выбор весовых коэффициентов случайным образом в диапазоне $[-0,05; 0,05]$, а также (в случае большого числа нейронов в слое) в диапазоне

$$\left[\frac{-1}{\sqrt{n(i)}}, \frac{1}{\sqrt{n(i)}} \right], \quad (8)$$

где $n(i)$ – число нейронных элементов в слое i .

В рассматриваемом случае в силу того, что число нейронов в слое небольшое, весовые коэффициенты инициализируются случайным образом в диапазоне $[-0,05; 0,05]$.

Исходя из того, что объём обучающей выборки достаточно большой, число нейронов в скрытом слое принято равным 32. Это даст приемлемую точность, при 8 нейронах на входе.

Для увеличения сходимости градиентного спуска и нейтрализации «скатывания» в точки локального минимума используется ряд методов и подходов. Так, в работе [9] описано «больцмановское» обучение, позволяющее системе делать случайный шаг в направлении, изменяющем целевую функцию, и тем самым выходить из локальных минимумов. В работе [10] предложен метод тяжёлого шарика, согласно которому в подстройку весов добавляется дополнительный множитель, называемый моментным термом. Введение его позволяет выходить из небольших локальных минимумов.

Однако эти меры не дают 100 % гарантии того, что процесс обучения сойдётся. Они лишь позволяют увеличить вероятность того, что система обучится, но значительно тормозят процесс обучения, так

как возникает необходимость дополнительного расчета состояния сети на каждом итерационном цикле обучения. Одним из эффективных мер является использование адаптивного шага обучения, который приближает сходимость алгоритма в отдельных случаях к 100 %.

Кроме этого, в нашей работе был применен дополнительный способ увеличения вероятности сходимости обучения. Он заключается в подаче значительного толчка на все веса отдельного слоя в соответствии с выражением (9):

$$w_k(n+1) = (Kw_y + w_k(n))^{-\lambda}, \quad (9)$$

где λ - толчок, величина которого > 1 , в зависимости от повторяемости величины ошибки слоя.

Это способ применяется только в случае, если на текущем слое суммарная ошибка достаточно велика и незначительно изменяется в пределах нескольких итераций обучения. Такая мера приводит к выходу отдельного слоя из состояния как локального минимума, так и глобального минимума. Импульс при этом сохраняется с торможением в течение нескольких итераций, поэтому слой успевает реабилитироваться и войти с большой вероятностью в состояние глобального минимума. При нормальной работе величина импульса равна единице.

Заключение. Предложено использование гибридной экспертной системы для исследования и оптимизации процесса закалки. Показано, что наиболее эффективным является применение многослойных нейронных сетей, с предварительной классификацией входной информации с помощью экспертных систем.

Разработана структура многослойной нейронной сети для инженерного анализа напряженно-деформированного состояния деталей в процессе закалки.

ЛИТЕРАТУРА

1. Кундас С.П., Кашко Т.А. Компьютерное моделирование технологических систем: Учеб, пособие: В 2 ч. Ч.1. - Мн.: БГУИР, 2004. - 167 с.
2. Кундас С.П., Тонконогов Б.А., Кашко Т.А. Компьютерное моделирование технологических процессов // Доклады Белорусского государственного университета информатики и радиоэлектроники. 2004. № 3(7). - С. 38-49.
3. Кундас С.П., Тонконогов Б.А., Гишкелюк И.А. Моделирование напряженно-деформированного состояния и твердости стальных деталей при закалке // Актуальные проблемы прочности: Сб. материалов XLIII междунар. конф. Ч. 1. - Витебск, 2004. - С. 10-16.
4. Кундас С.П., Левашкевич Я.С. Особенности построения гибридной экспертной системы для исследования и оптимизации технологических процессов // Вестник Полоцкого гос. ун-та. Сер. С. Фундаментальные науки. - 2004. - № 11. - С. 34 - 36.
5. Статические и динамические экспертные системы / Э.В.Попов, И.Б. Фоминых, Е.Б. Кисель, М.Д. Шапот. - М.: Финансы и статистика, 1996. - 320 с.
6. Artificial neural networks modeling of mechanical property and microstructure evolution in the Tempcore process // Computers and Structures. -2002. -№ 80. - Р. 213 -218.
7. Архитектура построения программных комплексов для моделирования технологических систем на основе динамически заменяемых модулей / С.П. Кундас, Д.Г. Иванов, А.В. Лемзиков, В.И. Коваленко // Вестник Полоцкого гос. ун-та. - 2004. - № 11.
8. Головкин В.А. Нейронные сети: обучение, организация и применение: Учеб, пособие для вузов / Общ. ред. А.И. Галушкина. - М.: ИПРЖР, 2001. Кн. 4. - 256 с.
9. Уоссермен Ф. Нейрокомпьютерная техника. - М.: Мир, 1992. - 194 с.
10. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation. In Parallel distributed processing. - Cambridge. MA: MIT Press. 1986. - V. 1. - P. 318 - 362.