

КЛАСТЕРНЫЙ АНАЛИЗ ДАННЫХ В СРЕДЕ STATSOFT ДЛЯ ТЕХНОЛОГИИ ЭНЕРГОСБЕРЕЖЕНИЯ

Робота присвячена проблемі автоматизації прийняття рішень про стан будівельних об'єктів і теплових комунікацій методами контролю з використанням нейроконструкторів у середовищі Statistika Neural Network. Запропоновано підвищити ступінь автоматизації процесу розпізнання стану об'єкта у форматі теплографічного способу шляхом застосування кластерного аналізу масиву вхідних даних із використанням стандартного нейромодулятора.

Ключові слова: енергозбереження, нейроконструктор, кластерний аналіз, навчання нейронної мережі, розпізнавання образів.

Работа посвящена проблеме автоматизации принятия решений о состоянии строительных объектов и тепловых коммуникаций методами контроля с применением нейроконструкторов в среде Statistika Neural Network. Предложено решение, повышающее степень автоматизации процесса распознавания состояния объекта в формате теплографического образа путем применения кластерного анализа массива входных данных с использованием стандартного нейромодулятора.

Ключевые слова: энергосбережение, нейроконструктор, кластерный анализ, обучение нейронной сети, распознавание образов.

The work is devoted to the automation of decision-making about the state of the builder-governmental facilities and communication methods of thermal control using neyrokonstruktorov among Statistika Neural Network. Propose increase the automation of the process of recognition of the object in the format teplografiche-sky image through the use of cluster analysis of the array of input data with the use of a standard neyroemulyatora.

Key words: energy - savings, neyrokonstruktor, cluster analysis, training the neural network, pattern recognition.

Постановка проблемы. В современных условиях актуальна проблема автоматизации процедур контроля энергетических параметров при внедрении энергосберегающих технологий. В строительстве используются различные инструментальные средства сбора информации о состоянии объектов, в том числе и строительные тепловизионные камеры, необходимые для тепловизионного обследования строительных сооружений [1]. Собранные данные являются основой для принятия решения о состоянии объекта, необходимых материальных и временных затратах. Однако анализ имеющихся данных не достаточно автоматизирован, что, с одной стороны, предполагает участие эксперта с его субъективной оценкой, с другой, увеличивает время на принятие решения. Поэтому автоматизация обработки данных с целью повышения продуктивности и оперативности принимаемых решений остается актуальной проблемой.

Анализ последних исследований и публикаций. Публикации, связанные с исследованиями в этой области, отражают определенные успехи в автоматизации нейросетевой обработки данных с применением нейросетевых технологий [2, 5, 6, 7]. Например, в работе [2] предлагается автоматизировать процедуру распознавания

состояния объекта с использованием обучающих процедур искусственных нейронных сетей. Действительно, после ввода данных алгоритм распознавания, построенный в нейросреде Matlab (The MathWorks) и обученный на представительной выборке, позволяет относительно быстро и с достаточной степенью достоверности классифицировать входной вектор. Но при одном условии: массив входных данных подготовлен для проведения обучения нейронной сети. Это означает, что априори известны номера классов, обозначающие соответствующие состояния исследуемого проблемного объекта. Тогда для каждого входного вектора из текущей выборки в процессе обучения ставится в соответствие истинный номер класса и соответствующий ему входной вектор.

Выделение не решенных ранее вопросов. В реальной ситуации подготовка обучающей выборки состоит в описании каждого класса на языке его информативных признаков и осуществляется оператором в ручном режиме. Это снижает объективность результатов и увеличивает время принятия решения, что и является **сутью решаемой проблемы** – автоматизация формирования алфавита классов состояний исследуемого объекта и соответствующего ему словаря признаков без участия оператора.

Предлагается рассматривать формирование обучающей выборки как задачу кластеризации входного множества признаков с применением нейроэмулятора *Statistika Neural Network*.

Цель статьи (постановка задачи). Задача кластеризации в этом случае заключается в автоматизации разделения входных признаков на группы таким образом, чтобы внутри группы входные векторы были максимально схожи между собой, а сами группы, напротив, максимально между собой различались.

Для решения этой задачи необходимо выполнить четыре процедуры:

1. Обосновать метрику сравнения образов, то есть в пространстве входных признаков выбрать меру сходства между ними.
2. Обеспечить инвариантность меры сходства от типа шкалы и масштаба.
3. Выбрать количество кластеров, учитывая особенности объекта.
4. Построить модель и разделить входные данные на кластеры.

Изложение основного материала. Решим задачу описания классов на языке их информативных признаков как задачу кластерного анализа входных данных. Пусть в качестве входных данных используется цифровая растровая последовательность изображения кадра, например 320 x 240 пикселей, как предложено в работе [3]. То есть размерность входного вектора определяется общим количеством пикселей, а активность каждого пикселя – записанным числом в двоичном коде установленной разрядности. Другими словами, входом нейронной сети является вектор фиксированной размерности, равный числу пикселей растровой развертки одного кадра. Это зафиксированный образ текущего состояния исследуемого участка на языке признаков растровой развертки. Необходимо построить алгоритм группировки вектора исходных данных на заданное число кластеров. Пусть исходные данные располагают числом примеров (наблюдений) 100 в (320 x 240) – мерном пространстве признаков (Var 1 – Var 76800).

Для решения этой задачи в первую очередь обоснуем метрику сравнения образов, то есть в пространстве входных признаков выберем меру сходства между ними. Воспользуемся традиционной евклидовой метрикой, которая выражается как корень из суммы квадратов по координатным разностям. В нашем случае каждая координата – это один из 76800 признаков входного вектора и в общем случае это двоичное число фиксированной разрядности. Это значит, что абсолютные величины входных признаков измеряются в единой шкале и вклад каждого в евклидову метрику

не зависит от типа измерительной шкалы признака. Это обстоятельство облегчает решение задачи, так как отпадает необходимость стандартизации входных переменных и опция Standardize Columns контекстного меню раздела Fill / Standardize Block в среде StatSoft игнорируется. Количество кластеров, учитывая особенности предметной области исследуемого объекта, несложно установить экспертным путем. В нашем случае их 4. На последнем этапе осуществим непосредственно разделение данных на кластеры. В основе этого разделения лежит итеративная процедура метода К-средних, которая начинается с задания начальных условий, т.е. числа кластеров и их центров.

Реализация кластеризации данных начинается из главного меню Statistics опцией Cluster Analysis. В диалоговом окне Clusterin method выбираем K-means clustering и инициируем сеть (ОК). Вызываем стартовую панель этого модуля командой Statistics из основного меню и во вкладке Quick данного диалогового окна в разделе Problem type выбираем задачу Classification. Командой Variables вызываем диалоговое окно Select input, output и записываем все переменные по списку с помощью редактора ввода данных. В разделе Selekt analysis выбираем опцию Intelligent Problem Solver, что необходимо для установки параметров сети по умолчанию и фиксации сложности и топологии сети. В диалоговом окне IPS Training In Progress выбираем сеть. Анализ осуществляем по десяти строкам информационного поля диалогового окна. В разделе Profile фиксируем топологию сети

(SNN – Kohonen 36:36 – 18:1).

На вкладке Advanced в разделе Clusters выберем Cases, а в поле Number of Clusters введем число предполагаемых классов (кластеров). Сохраним результаты анализа при помощи вкладки Advanced выбором опции Save clustering and distances для сохранения результатов кластеризации и расстояния. В таблице результатов зафиксируем (скопируем) переменную Cluster и поместим её в исходный файл данных. В строках данной переменной содержатся номера классов, к которым были отнесены в итоге многомерные образы. Кластеризация завершена, классы описаны на языке выбранных информативных признаков и могут быть использованы как обучающие выборки того или иного класса.

Таким образом, поставлена и решена задача формирования обучающей выборки для всего алфавита классов, как задача кластерного анализа в нейронной среде стандартных эмуляторов формата *Statistika Neural Network*. Разработана модель кластерного анализа входного вектора и проверена процедура её обучения на множестве входных данных. Показатели производительности модели, значения ошибок на обучающем, контрольном и тестовом множествах позволяют утверждать факт удовлетворительной сходимости итерационных процессов обучения сети и надежности результатов автоматической процедуры формирования алфавита классов и словаря признаков для исследуемого объекта.

Выводы и перспективы дальнейших исследований в данном направлении.

1. Кластерный анализ данных в среде Statsoft для технологии энергосбережения реализуем в нейросетевом формате с допустимой для практики надежностью.

2. Научная новизна предлагаемого решения заключается в теоретическом и практическом обосновании применения среды стандартных нейроэмуляторов по новому назначению – кластерный анализ входных образов в пространстве признаков теплографического портрета исследуемого коммунального объекта.

3. Практическая ценность результатов исследования состоит в сокращении времени на принятие решения и снижении влияния экспертных оценок на результат.

4. Реализуемость теоретических выводов обеспечивается устойчивой сходимостью итерационных обучающих процедур, а достоверность результатов подтверждается допустимыми ошибками и показателями продуктивности модели.

5. Целесообразно продолжить исследования по применению кластерного анализа данных в предметных областях, где признаковые пространства так же формируются в растровом формате (аэрофотосъемка, видеосистемы охранной сигнализации, системы идентификации документов и денежных купюр и т.п.).

ЛИТЕРАТУРА

1. СНиП 23-02-2003. Тепловая защита зданий. – М.: Госстрой России, 2004. – 214 с.

2. Ляхов, А.Л. Интеллектуальный анализ данных в прикладных экономических задачах / А.Л. Ляхов, С.П. Алёшин // *Економіка і регіон: наук. вісн.* – Полтава: ПолтНТУ, 2009. – № 4(23). – С. 140 – 147.

3. Хайкин, С. Нейронные сети: полный курс. 2-е изд. / С. Хайкин. – М.: Вильямс, 2006. – 1104 с.

4. Галушкин, А.И. Нейрокомпьютеры и их применение на рубеже тысячелетий в Китае. В 2 т. – М.; 2004. Т.1. – 367 с.

5. Т.2. – 464 с.

6. Ляхов, А.Л. Проблема моделирования сложных социотехнических систем: [электрон. докл.] / А.Л. Ляхов, С.П. Алёшин // V дистанционная научно-практическая конференция с международным участием «Системы поддержки принятия решений. Теория и практика», СППР, 2009: сб. докл. ИПММС НАН Украины. – К., 2009. – С. 31–34. (<http://conference.immsp.kiev>).

7. Ляхов, А.Л. Искусственная нейронная сеть как измерительный инструмент адекватности модели с адаптивным классом точности / А.Л. Ляхов, С.П. Алёшин // IV научно-практическая конференция с международным участием «Математическое и имитационное моделирование систем. МОДС '2009»: сб. докл. ИПММС НАН Украины. – К., 2009. – С. 116 – 119.

8. Состояние и перспективы нейросетевого моделирования СППР в сложных социотехнических системах / [А.А. Морозов, В.П. Клименко, А.Л. Ляхов, С.П. Алёшин] // *Математичні машини і системи.* – 2010. – № 1. – С. 127 – 149.

Надійшла до редакції 07.01.2010 р.

© С.П. Алёшин