

УДК 004.8

А.М. Кориков, А.Т. Нгуен

Нечеткая нейросетевая классификация интенсивности марковского случайного потока событий

Исследовано применение моделей нечетких нейронных сетей, использующих нечеткие функции активации нейронов, для решения задач классификации интенсивности марковских случайных потоков событий. Исследование проводится с помощью средств компьютерного моделирования MATLAB. Марковские случайные потоки событий представлены в виде двух последовательно идущих друг за другом кусков простейших (пуассоновских) потоков, каждый из которых характеризуется своей интенсивностью прихода событий. С помощью нечетких нейронных сетей решается задача дихотомии: классификация интенсивности двух пуассоновских потоков. Компьютерное моделирование подтвердило эффективность нечеткой нейросетевой классификации интенсивности марковского случайного потока событий.

Ключевые слова: нечеткие функции активации, нечеткие нейронные сети, МАР-потоки событий, МС-потоки событий, классификация интенсивности потока.

doi: 10.21293/1818-0442-2017-20-2-79-83

Методы и технологии искусственного интеллекта (ИИ) используются для решения многих практических задач: автоматизации и роботизации технологических процессов и производств, проектирования информационно-вычислительных систем, телекоммуникационных и компьютерных сетей и т.д. Искусственные нейронные сети (НС) и технологии, базирующиеся на них, занимают лидирующие позиции среди методов и технологий ИИ [1]. В наших работах [2, 3] предлагаются новые модели нечетких НС (ННС), использующих нечеткие функции активации (ФА) нейронов ННС. В таких ННС нечеткость является атрибутом нейрона как элемента нейронной сети. Эти ННС определены в [2, 3] как нечеткие нейронные сети второго типа. В [2, 3] нечеткими нейронными сетями первого типа названы ННС, рассмотренные в [4], в которых реализуются нечеткие отношения между нейронами как элементами нейронной сети, т.е. ННС первого типа реализуют системы нечеткого логического вывода методами теории нейронных сетей. В [2, 3] показано, что разрабатываемые нами модели нечетких нейронов и нечеткие нейронные сети второго типа успешно решают задачи прогнозирования свойств временного ряда. В настоящей работе исследуется возможность применения ННС второго типа для решения задач классификации параметров марковских случайных процессов.

Постановка задачи

Пусть $\lambda(t)$ – кусочно-постоянный случайный процесс с двумя состояниями $\lambda(t) = \lambda_1$ либо $\lambda(t) = \lambda_2$ ($\lambda_1 < \lambda_2$). Случайный процесс $\lambda(t)$ принципиально не наблюдаем. Время пребывания процесса $\lambda(t)$ в i -м состоянии является случайной величиной, распределенной по экспоненциальному закону с параметрами λ_i ($i = 1, 2$), т.е. согласно [5] рассматриваемый случайный процесс моделирует МАР-поток (Markovian Arrival Process) событий с экспоненциальной функцией распределения

$$F_i(t) = 1 - e^{-\lambda_i t}, \quad i = 1; 2. \quad (1)$$

МАР-потоки событий часто используются в теории массового обслуживания (в англоязычной литературе – Queueing Theory (теория очередей)). Основоположником исследований в данной области считается датский ученый А.К. Эрланг, опубликовавший в 1909–1917 гг. работы, посвященные обслуживанию требований, поступающих на телефонную станцию. Развитие теории массового обслуживания связано с её применением в системах управления запасами, транспортных системах, системах связи, компьютерных сетях [6]. Частные случаи МАР-потоков называют также МС-потоками (Markov Chain) [7, 8]. Эти потоки можно представить в виде последовательно идущих друг за другом кусков простейших (пуассоновских) потоков, каждый из которых характеризуется своей интенсивностью λ_i прихода событий. Поставим задачу исследования возможности применения ННС второго типа для классификации и оценивания интенсивности λ_i потоков событий. Исследование проведем с помощью средств компьютерного моделирования MATLAB [9].

Эксперименты по классификации интенсивности МС-потоков событий

Эксперименты по нейросетевой классификации МС-потоков событий выполнены на основе ННС, структура которой может быть представлена в виде многослойной однонаправленной сети (рис. 1). ННС состоит из входного слоя, одного скрытого слоя и выходного слоя. Скрытый слой состоит из 10 нейронов с нечеткими ФА, виды которых для нечетких нейронных сетей второго типа представлены в [2, 3]. Из четырех видов нечетких ФА, исследованных в [2, 3], выберем два вида ФА (см. на рис. 1 в [2] эти ФА обозначены через b и c), которые, как показано в [2, 3], наиболее эффективно решают задачи прогнозирования свойств временного ряда. В выходном слое используется линейная ФА. На входы ННС поступают марковские случайные процессы с экспоненциальной функцией распределения (1).

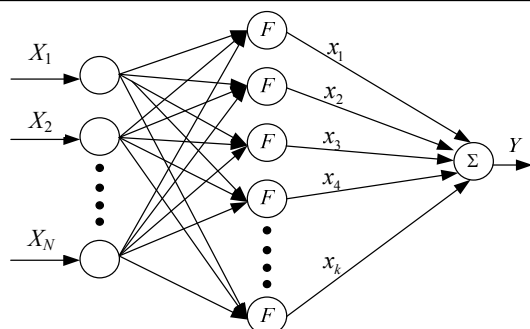


Рис. 1. Структура сети

Марковские случайные потоки событий, поступающие на входы ННС, представляются в виде двух последовательно идущих друг за другом кусков простейших (пуассоновских) потоков, каждый из которых характеризуется своей интенсивностью λ_i ($i = 1, 2$) прихода событий.

В данной работе представлены результаты исследования по классификации параметров случайных потоков на основе трех экспериментов:

- первый эксперимент: $\lambda_1 \in (0; 0,5]$, $\lambda_2 \in (0,5; 1]$, параметры λ_1 , λ_2 не фиксированы и не известны;
- второй эксперимент: $\lambda_1 \in (0; 0,5]$, $\lambda_2 \in (0,5; 1]$, параметры λ_1 , λ_2 фиксированы, но не известны;
- третий эксперимент: $\lambda_1 = 0,25$, $\lambda_2 = 0,75$, т.е. параметры λ_1 , λ_2 фиксированы и известны.

Наборы данных для обучения ННС

В первом эксперименте входными данными являются потоки, состоящие из 1000 случайных процессов с параметрами $\lambda(t)$, принимающими значения $\lambda = 0,001; 0,002; \dots; 1$. То есть входные данные можно представить в виде матрицы размерности 1000×1000 . Выходные данные можно представить в виде матрицы-вектора размерности 1000×1 , элементы которой имеют только два значения: 0 или 1. Значение 0 обозначает, что классифицируемый поток имеет распределение с параметрами λ_1 из интервала $(0; 0,5]$, значение 1 обозначает, что классифицируемый поток имеет распределение с параметрами λ_2 из интервала $(0,5; 1]$.

Во втором эксперименте, входными данными являются потоки, состоящие из 500 случайных процессов с параметром $\lambda_1 \in (0; 0,5]$ и 500 случайных процессов с параметром $\lambda_2 \in (0,5; 1]$. Выходные данные представляются матрицей-вектором размерности 1000×1 , состоящей из значений 0 и 1. Значение 0 обозначает, что классифицируемый поток имеет распределение с параметром λ_1 из интервала $(0; 0,5]$, значение 1 обозначает, что классифицируемый поток имеет распределение с параметром λ_2 из интервала $(0,5; 1]$.

В третьем эксперименте входными данными являются 200 случайных процессов с параметром $\lambda_1 = 0,25$ и 200 случайных процессов с параметром $\lambda_2 = 0,75$. Выходные данные также являются матри-

цей-вектором размерности 400×1 , которая состоит из значений 0 и 1. Значение 0 обозначает, что поток имеет распределение с параметром $\lambda_1 = 0,25$; значение 1 обозначает, что поток имеет распределение с параметром $\lambda_2 = 0,75$.

Алгоритм кластеризации K-средних

Для классификации параметров λ_i ($i = 1, 2$) случайных процессов предлагается использовать алгоритм кластеризации K-средних, который является простым повторяющимся алгоритмом кластеризации, разделяющим определенный набор данных на заданное пользователем число кластеров k [10]. Алгоритм K-средних в нашем случае рассматривает распределение каждого случайного процесса как точку в функциональном пространстве. Базовый алгоритм K-средних произвольно определяет число кластерных центров в многомерном функциональном пространстве. Каждое распределение в наборе данных присваивается кластеру, произвольный средний вектор которого является самым близким. Процедура продолжается до тех пор, пока не произойдет существенного изменения положения средних векторов классов между последовательными итерациями алгоритмов. Однако алгоритм K-средних очень чувствителен в начальных отправных точках. Алгоритм K-средних генерирует начальный кластер случайным образом. Когда случайные начальные точки близки к окончательному решению, алгоритм K-средних имеет высокую вероятность обнаружить центр кластера. В противном случае это приведет к некорректным результатам кластеризации [11].

В нашей работе начальные точки алгоритма K-средних генерируются случайно и количество классов равно десяти. Краткое описание алгоритма K-средних, используемого в данной работе, состоит в следующем:

Шаг 1: случайно выбираются 10 начальных центров кластера из представленных выше наборов данных z_1, z_2, \dots, z_{10}

$$z_a = \{x_{b1}, x_{b2}, \dots, x_{bN}\},$$

где $a = 1; 2; \dots; 10$, и b – случайное число от 1 до L , где L – количество параметров λ в наборе данных; N – количество наблюдаемого времени в распределении.

Шаг 2: определяются Евклидово расстояние от распределения x_i , ($i = 1; 2; \dots; L$) до кластера, C_j ($j = 1; 2; \dots; 10$), если $\|x_i - z_j\| < \|x_i - z_p\|$, $p = 1; 2; \dots; 10$, $j \neq p$, тогда x_i находится в кластере C_j .

Шаг 3: вычисляются новые кластерные центры z_i^* по формуле

$$z_i^* = \frac{1}{n_i} \sum_{x_j \in C_i} x_j, \quad (j = 1; 2; \dots; 10), \quad (2)$$

где n_i является количеством распределений, принадлежащих кластеру C_i .

Шаг 4: если $z_i^* = z_i$ ($i = 1; 2; \dots; 10$), тогда останов. В противном случае следует перейти к шагу 2.

Если имеются ограничения по времени и если процесс кластеризации не завершается на шаге 4, тогда он выполняется для максимального фиксированного числа итераций k_1 . Чем больше k_1 , тем точнее определяются центры кластеров.

Настройка функций активации

Выше отмечалось, что из четырех видов нечетких ФА, исследованных в [2, 3], мы выбрали два вида ФА, которые на рис. 1 в [2] обозначены через b и c и которые, как показано в [2, 3], наиболее эффективно решают задачи прогнозирования свойств временного ряда. На рис. 2 показан внешний вид ФА избранных видов c и b [2].

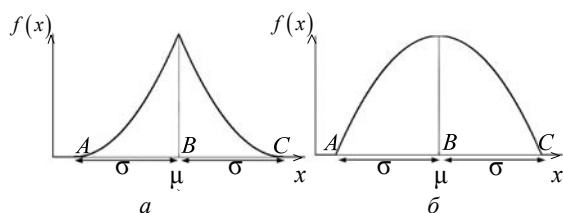


Рис. 2. Функции активации: a – ФА вида c [2]; b – ФА вида b [2]

Набор точек $\langle A, B, C \rangle$ на рис. 2 являются характеристическими точками ФА.

Известно, что $f_L(x)$, $f_R(x)$ являются полиномами второго порядка и $f_L'(x)=0$, $f_R'(x)=0$ в характеристических точках, т.е.

$$f_L(x) = \sum_{i=0}^2 a_i x^i \quad f_R(x) = \sum_{i=0}^2 b_i x^i. \quad (3)$$

Для ФА вида c коэффициенты полиномов $f_L(x)$ и $f_R(x)$ задаются следующими выражениями:

$$\begin{aligned} a_0 &= \frac{A^2}{(A-B)^2}; \quad a_1 = \frac{-2A}{(A-B)^2}; \quad a_2 = \frac{1}{(A-B)^2}; \\ b_0 &= \frac{C^2}{(B-C)^2}; \quad b_1 = \frac{-2C}{(B-C)^2}; \quad b_2 = \frac{1}{(B-C)^2}. \end{aligned} \quad (4)$$

Зададим $AB=BC=\sigma$ и $B=\mu$. Тогда $f_L(x)$ и $f_R(x)$ определяются следующим образом:

$$\begin{aligned} f_L(x) &= \frac{1}{\sigma^2} x^2 + \frac{-2(\mu-\sigma)}{\sigma^2} x + \frac{(\mu-\sigma)^2}{\sigma^2}, \mu-\sigma \leq x \leq \mu; \\ f_R(x) &= \frac{1}{\sigma^2} x^2 + \frac{-2(\mu+\sigma)}{\sigma^2} x + \frac{(\mu+\sigma)^2}{\sigma^2}, \mu \leq x \leq (\mu+\sigma); \end{aligned} \quad (5)$$

μ и σ называются центром и шириной функций активации.

Пусть $a=(\mu-\sigma)$, если $\mu-\sigma \leq x \leq \mu$ и $a=(\mu+\sigma)$, если $\mu \leq x \leq (\mu+\sigma)$, тогда получим общее описание (обобщенная) ФА вида c :

$$f(x) = \frac{1}{\sigma^2} x^2 + \frac{-2a}{\sigma^2} x + \frac{a^2}{\sigma^2}. \quad (6)$$

Аналогично для ФА вида b $f_L(x)$ и $f_R(x)$ определяются следующим образом:

$$\begin{aligned} f_L(x) &= \frac{-1}{\sigma^2} x^2 + \frac{2\mu}{\sigma^2} x + \frac{(\mu-\sigma)^2}{\sigma^2} - \frac{2\mu(\mu-\sigma)}{\sigma^2}, \\ &\mu-\sigma \leq x \leq \mu; \\ f_R(x) &= \frac{-1}{\sigma^2} x^2 + \frac{2\mu}{\sigma^2} x + \frac{(\mu+\sigma)^2}{\sigma^2} - \frac{2\mu(\mu+\sigma)}{\sigma^2}, \\ &\mu \leq x \leq (\mu+\sigma). \end{aligned} \quad (7)$$

На этой основе получим общее описание (обобщенная) ФА вида b :

$$f(x) = \frac{-1}{\sigma^2} x^2 + \frac{2\mu}{\sigma^2} x + \frac{a^2 - 2\mu a}{\sigma^2}, \quad (8)$$

где $a=(\mu-\sigma)$ если $\mu-\sigma \leq x \leq \mu$, и $a=(\mu+\sigma)$, если $\mu \leq x \leq (\mu+\sigma)$.

Выражения (6) и (8) являются функциями активации, которые используются в ядре нейрона скрытого слоя ННС на рис. 1.

Вычисление центров и ширины ядра функций активации

Как правило, центр и распределение функций активации нейронов ННС должны иметь характеристики, подобные данным. Для определения центров ФА нейронов ННС обычно используются неконтролируемые учебные процедуры по кластеризации [10]. В данной работе алгоритм кластеризации К-средних используется для определения центра и ширины ФА нейронов ННС. Каждый кластер соответствует ядру функций активации. Расположение центров ядра функций активации определяется с помощью следующего уравнения:

$$\mu_i = \frac{1}{p} \sum_{j=1}^p X_j, \quad (9)$$

где μ_i является центром i -го нечеткого скрытого нейрона; p – количество данных в формируемом кластере; X_j – j -й данный в i -м кластере.

Настройка ширины ядра является труднорешаемой проблемой при построении ННС. Если ширина ядра слишком велика, то оцененная плотность вероятности сглаживается. Наоборот, когда она слишком мала, может быть чрезмерная адаптация к определенному набору данных. Ширина ядра устанавливается на среднее расстояние между данными соответствующего кластера:

$$\sigma_i = \sqrt{\frac{1}{p} \sum_{j=1}^p (X_j - \mu_i)^2}, \quad (10)$$

где σ_i является шириной i -го нечеткого скрытого нейрона; μ_i является центром i -го нечеткого скрытого нейрона; p – количество функциональных данных в данном кластере. X_j – j -й данный в i -м кластере. Ширина ядра (10) гарантирует, что от-

дельные функции активации не слишком острые или/и не слишком плоские.

Обучение ННС с использованием тестовых наборов данных

При построении ННС задача корректировки весов сети между скрытым слоем и выходным слоем может быть отнесена к задачам оптимизации, обычно решаемым методом наименьших квадратов. Веса ННС между скрытым слоем и выходным слоем рассчитываются с использованием метода псевдообращения [12, 13].

На рис. 1 X представляет собой входные наборы данных о пространственных входных признаках, которые представляют собой реализации случайных процессов. Выход y ННС в наблюдаемый момент времени j ($j=1; 2; \dots; N$, N – количество наблюдаемого времени) имеет следующий вид:

$$y_j = \sum_{i=1}^k w_i f(x_i), j=1; 2; \dots; N, \quad (11)$$

где w_i – вес между выходным нейроном и i -м скрытым нейроном; $f(x)$ – обобщенная ФА; k – число нейронов скрытого слоя.

Уравнение (11) можно записать в виде

$$Y = \Phi w, \quad (12)$$

где

$$\Phi(\mu, \sigma, X) = \begin{bmatrix} \varphi_1(\mu_1, \sigma_1, X_1) & \dots & \varphi_K(\mu_K, \sigma_K, X_1) \\ \vdots & \dots & \vdots \\ \varphi_1(\mu_1, \sigma_1, X_N) & \dots & \varphi_K(\mu_K, \sigma_K, X_N) \end{bmatrix}, \quad (13)$$

$$v=1; 2; \dots; N,$$

φ – функция активации. Матрица Φ (размерность $N \times K$) называется выходной матрицей скрытого слоя нейронной сети; i -я строка Φ является i -м скрытым выходным нейроном относительно входов X_1, X_2, \dots, X_N .

Пусть T – матрица выходных данных обучения. Если $\Phi w = T$, то выходные веса w вычисляются по формуле

$$w = (\Phi^T \Phi)^{-1} \Phi^T T, \quad (14)$$

где $(\Phi^T \Phi)^{-1} \Phi^T$ является обобщенной псевдообратной выходной матрицей Мура–Пенроуза скрытого слоя.

Итак, основные параметры ННС определяются путем вычисления центров и ширины с использованием алгоритма кластеризации K -средних и определения весов выходного слоя ННС с использованием метода псевдообращения.

Результаты тестирования

Эксперименты проводились на базе процессора Core i5-4460 3,2 ГГц с 8 ГБ ОЗУ под MATLAB R2013a. Графики генерировались с использованием MATLAB.

В первом эксперименте наборы данных для тестирования состояли из 400 случайных процессов $\lambda^*(t)$, $\lambda^* = 0,001; 0,001; \dots; 1$. Результаты классифи-

кации с помощью ННС процессов по интенсивностям λ_1 , λ_2 представлены на рис. 3 и 4.

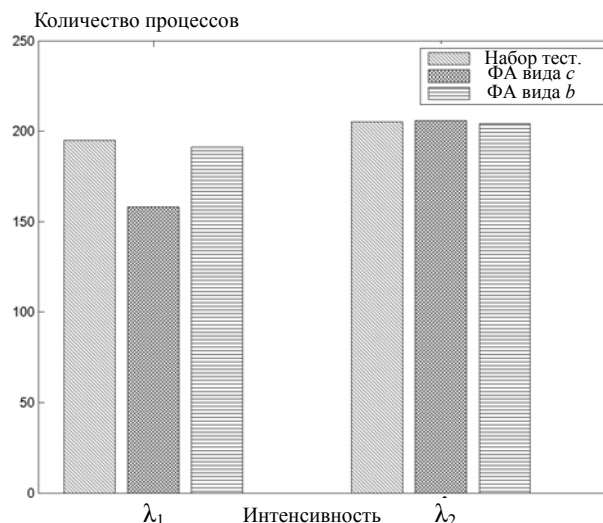


Рис. 3. Количество процессов с интенсивностями λ_1 , λ_2 в момент времени $t = 500$ в наборе данных тестирования и количество правильных решений при определении параметров процессов в первом эксперименте

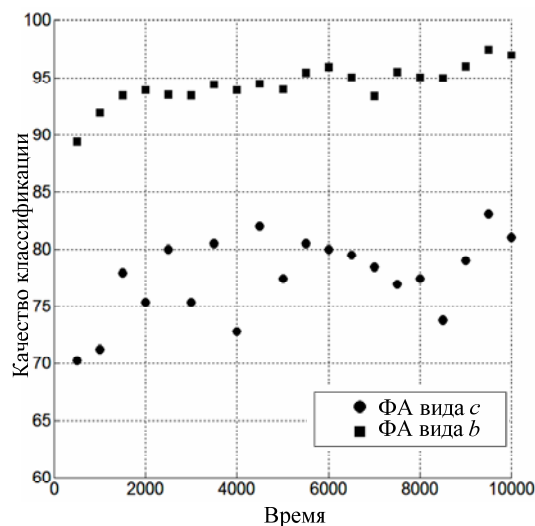


Рис. 4. Эффективность классификации по интенсивности марковских случайных потоков событий в первом эксперименте

Во втором эксперименте набор данных для тестирования состоял из 400 случайных процессов с параметром $\lambda_1 = 0,25$ и 400 случайных процессов с параметром $\lambda_2 = 0,75$. Эффективность классификации этих процессов по интенсивности λ_1 , λ_2 во втором эксперименте практически равна 100%.

В третьем эксперименте наборы данных для тестирования состояли из 400 случайных процессов с параметром $\lambda_1 = 0,25$ и 400 случайных процессов с параметром $\lambda_2 = 0,75$. Эффективность классификации этих процессов по интенсивности λ_1 в третьем эксперименте также практически равна 100%.

Анализ результатов экспериментов показывает, что разрабатываемые нами модели нечетких нейронных сетей вполне успешно решают задачи классификации по интенсивности марковских случайных потоков событий. При решении подобных задач ННС с ФА вида b имеют более высокую эффективность. В условиях первого и второго экспериментов продолжительность времени наблюдения за процессами не влияет на качество их классификации. В третьем эксперименте при использовании ННС с ФА вида c продолжительность времени наблюдения за процессами незначительно влияет на результаты классификации.

Заключение

Результаты исследования подтверждают эффективность применения моделей нечетких нейронных сетей, в которых нечеткие функции принадлежности используются в качестве функций активации нейронов ННС, для решения задач классификации интенсивности марковских случайных потоков событий. На основе проведенного исследования возможно создание нечеткого нейросетевого классификатора интенсивности марковских случайных потоков событий для решения многих прикладных задач в транспортных системах, системах связи и компьютерных сетях.

Литература

1. Рассел С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. – М.: Вильямс, 2006. – 1408 с.
2. Нгуен А.Т. Модель нейронной сети с нечеткими функциями активации для прогнозирования временного ряда / А.Т. Нгуен, А.М. Кориков // Доклады ТУСУРа. – 2016. – Т. 19, № 4. – С. 50–52.
3. Nguyen A.T. Models of neural networks with fuzzy activation functions / A.T. Nguyen, A.M. Korikov // IOP Conference Series: Materials Science and Engineering. – 2017. – Vol. 177 [Электронный ресурс]. – Режим доступа: <http://iopscience.iop.org/article/10.1088/1757-899X/177/1/012031>, свободный (дата обращения: 01.03.2017).
4. Ярушкина Н.Г. Нечеткие нейронные сети с генетической настройкой // Научная сессия МИФИ–2004. VI Всерос. науч.-техн. конф. «Нейроинформатика–2004»: лекции по нейроинформатике. – Ч. 1. – М.: Изд-во МИФИ, 2004. – С. 151–199.
5. Дудин А.Н. Системы массового обслуживания с коррелированными потоками / А.Н. Дудин, В.И. Клименок. – Минск: Изд-во БГУ, 2000. – 175 с.
6. Гнеденко Б.В. Введение в теорию массового обслуживания / Б.В. Гнеденко, И.Н. Коваленко. – 3-е изд., испр. и доп. – М.: КомКнига, 2005. – 400 с.
7. Горцев А.М. О связи МС-потоков и МАР-потоков событий / А.М. Горцев, Л.А. Нежелская // Вестник Том. гос. ун-та. Управление, вычислительная техника и информатика. – 2011. – № 1 (14). – С. 13–21.
8. Беккерман Е.Н. Аппроксимация МС-потокм реального потока событий / Е.Н. Беккерман, С.Г. Катаев, С.С. Катаева, Д.Ю. Кузнецов // Вестник Том. гос. ун-та. Приложение. – 2005. – № 14. – С. 248–253.
9. Attaway S. MATLAB: A Practical Introduction to Programming and Problem Solving. – Third Edition. – Burlington, MA: Elsevier Inc., 2013. – 539 p.
10. Пирсон Л. Наука данных для чайников. – М.: Джон Уайли & сыновья, 2015. – 408 с.
11. Wu Z. An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation / Z. Wu, R.M. Leahy // IEEE Transactions on pattern analysis and machine intelligence. – 1993. – Vol. 15, № 11. – P. 1101–1113.
12. Wettscherech D. Improving the Performance of Radial Basis Function Networks by Learning Center Locations / D. Wettscherech, T. Dietterich // Electronic Proceedings of the Neural Information Processing Systems Conference. – 1991. – № 4. – P. 1133–1140.
13. Сараев П.В. Использование псевдообращения в задачах обучения искусственных нейронных сетей // Электронный журнал «Исследовано в России». – 2001. – № 29. – С. 308–317.

Кориков Анатолий Михайлович

Д-р техн. наук, проф., зав. каф. АСУ ТУСУРа, профессор Национального исследовательского Томского политехнического университета (НИ ТПУ)
Тел.: +7 (382-2) 41-42-79
Эл. почта: korikov@asu.tusur.ru

Нгуен Ань Ту

Аспирант каф. автоматизации и компьютерных систем (АиКС) Института кибернетики НИ ТПУ
Тел.: +7-952-157-48-78
Эл. почта: nguyenanhtu@tpu.ru

Korikov A.M., Nguyen A.T.

Fuzzy neural network for Markovian arrival processes' intensity classification

The paper studies the applications of fuzzy neural networks that use membership functions as activation functions to solve the problems of classification and estimation of Markov arrival processes' intensity. Computer simulation tools in MATLAB are used for this research. Markov arrival processes are represented by two successive simple (Poisson) processes, each of which is characterized by the intensity of event arrival. Using neural networks, we can obtain the solutions to dichotomy problems, including the classification and estimation of the intensity of the two Poisson processes. Computer modeling has confirmed the effectiveness of the classification problem for the intensity of Markovian arrival processes based on neural networks.

Keywords: Fuzzy activation functions, Fuzzy neural networks, Markovian arrival processes, intensity classification.