

УДК 004.8

МОДИФИКАЦИЯ АССОЦИАТИВНОЙ ГАУССОВОЙ МОДЕЛИ СМЕЩЕНИЯ В ЭКСПЕРТНЫХ ОЦЕНКАХ

Д. В. Горбунов, аспирант кафедры ИУ7 МГТУ им. Н. Э. Баумана, Москва, Россия;
orcid.org/0000-0002-4646-2636, e-mail: darkwen3@mail.ru.

К. Л. Тассов, старший преподаватель кафедры ИУ7 МГТУ им. Н.Э. Баумана, Москва, Россия;
e-mail: ktassov@bmstu.ru.

С. В. Телегин, бакалавр кафедры СГНЗ МГТУ им. Н.Э. Баумана, Москва, Россия;
orcid.org/0009-0000-6637-7124, e-mail: semen-telegin@mail.ru.

Идея ансамблевого обучения состоит в том, чтобы организовать такой пул экспертов, который позволит объединить их в общую систему. В качестве эксперта могут рассматриваться различные алгоритмы предсказания, в том числе и нейронные сети. Суть ансамблевых методов состоит в том, что каждый из экспертов предоставляет свой отклик, который в последующем может использоваться обобщающим алгоритмом. Одним из обобщающих алгоритмов для классификации является ассоциативная гауссова модель смещения. В ней один эксперт для одного класса определяет результат только в одной области, что является недостатком. Поэтому предлагается модифицированный метод ассоциативной гауссовой модели смещения, в котором один эксперт определяет результат сразу в нескольких областях. В рамках метода предложен алгоритм, включающий в себя следующие этапы: нормировка входных данных, обучение модели на основе нормированных входных данных, нахождение значений выходов. Метод может служить любым сферам, в которых используются «слабые эксперты» и где необходимо ранжировать экспертов. В статье рассмотрены объединение ассоциативной гауссовой модели смещения и перцептрона Румельхарта, преобразование входных данных, обучение и предсказание с помощью модифицированного метода ассоциативной гауссовой модели смещения, приведены результаты полученных значений выходов.

Ключевые слова: модифицированная ассоциативная гауссова модель смещения, ассоциативная гауссова модель смещения, слабые эксперты, нейронные сети, перцептрон Румельхарта.

DOI: 10.21667/1995-4565-2023-85-54-64

Введение

Многие задачи, возникающие в практических приложениях, не могут быть решены заранее известными методами и алгоритмами с минимальным процентом ошибки. Происходит это по нескольким причинам: имеется недостаточный объем обучающей выборки, попадание в локальные минимумы при обучении, область решения задач превосходит область определения метода. Таким образом, мы приходим к необходимости использования нескольких экспертов. Алгоритм принятия решения, построенный на нескольких экспертах, носит различные названия: ансамбль экспертов, смесь экспертов [1, 2]. Данные подходы находят применение, например, в классификации или кластерном анализе [3].

Идея ансамблевого обучения состоит в том, чтобы организовать такой пул экспертов, который позволит объединить их в общую систему [4]. В качестве эксперта могут рассматриваться различные алгоритмы предсказания, в том числе и нейронные сети. Суть ансамблевых методов состоит в том, что каждый из экспертов предоставляет свой отклик, который в последующем может использоваться обобщающим алгоритмом.

Одним из обобщающих алгоритмов для ранжирования является ассоциативная гауссова модель смещения. В ней один эксперт для одного класса определяет результат только в одной области, что является недостатком. Поэтому предлагается модифицированный метод ассоциативной гауссовой модели смещения, в котором один эксперт определяет результат сразу в нескольких областях.

Постановка задачи

Необходимо модифицировать ассоциативную гауссову модель смещения для того, чтобы один эксперт определял результат сразу в нескольких областях. Для этого необходимо добавить в модель перцептрон Румельхарта, вывести формулы для одновременного обучения перцептрона и ассоциативной гауссовой модели смещения, формулы для получения конечного результата и формулы для нормализации входных данных.

Модифицированный метод ассоциативной гауссовой модели смещения

Метод включает в себя:

- 1) нормировку входных данных;
- 2) объединение ассоциативной гауссовой модели смещения и перцептрона Румельхарта.

Нормировка входных данных

$X = \{\bar{x}'_0, \dots, \bar{x}'_{N-1}\}$ – входная выборка, где $\bar{x}'_i = (x'_{i,0}, \dots, x'_{i,n-1})$ – входной вектор выборки.

Нормализация входной выборки данных происходит по следующему алгоритму.

Для каждого измерения j входных векторов из входной выборки данных находятся минимальное и максимальное значения: $\min_{\bar{x}'_i \in X} x'_{i,j}$ и $\max_{\bar{x}'_i \in X} x'_{i,j}$.

Далее для каждого измерения j входных векторов из входной выборки происходит перенос в центр, масштабирование и проекция на единичную окружность:

$$x_{i,j} = \frac{x'_{i,j} - \frac{\max_{\bar{x}'_i \in X} x'_{i,j} + \min_{\bar{x}'_i \in X} x'_{i,j}}{2}}{(\max_{\bar{x}'_i \in X} x'_{i,j} - \min_{\bar{x}'_i \in X} x'_{i,j}) \cdot |\bar{x}'_i|}$$

Таким образом, получаем нормализованные входные векторы $\bar{x}_i = (x_{i,0}, \dots, x_{i,n-1})$.

Объединение ассоциативной гауссовой модели смещения и перцептрона Румельхарта

Введём необходимые обозначения:

$\bar{x} = (x_0, \dots, x_{n-1})$ – нормализованный входной вектор.

$W^{(1)} = \{w^{(1)}_{i,j}\}$ – матрица весов первого слоя перцептрона размерностью $n+1 \times n_1$, где n_1 – количество нейронов первого слоя перцептрона, а $w^{(1)}_{i,j}$ – вес i -го входа j -го нейрона (для удобства описания будем вести нумерацию с нуля, где нулевой вход – свободный коэффициент, всегда равный единице).

$\bar{u}^{(1)} = (1, x_0, \dots, x_{n-1})W^{(1)}$ – линейная часть выхода первого слоя перцептрона.

$\bar{y}^{(1)} = \varphi(\bar{u}^{(1)})$ – результат работы первого слоя перцептрона, где φ – функция активации (будем использовать гиперболический тангенс).

$W^{(2)} = \{w^{(2)}_{i,j}\}$ – матрица весов второго слоя перцептрона размерностью $n_1+1 \times n_2$, где n_2 – количество нейронов второго слоя перцептрона, а $w^{(2)}_{i,j}$ – вес i -го входа j -го нейрона.

$\bar{u}^{(2)} = (1, y^{(1)}_0, \dots, y^{(1)}_{n_1-1})W^{(2)}$ – линейная часть выхода второго слоя перцептрона.

$\bar{y}^{(2)} = \varphi(\bar{u}^{(2)})$ – результат работы второго слоя перцептрона, где φ – функция активации (будем использовать гиперболический тангенс).

$W^{(3)} = \{w^{(3)}_{i,j}\}$ – матрица весов шлюза размерностью $n_2 \times k$, где k – количество нейронов шлюза, равное количеству слабых экспертов, а $w^{(3)}_{i,j}$ – вес i -го входа j -го нейрона.

$\bar{u}^{(3)} = \bar{y}^{(2)}W^{(3)}$ – линейная часть выхода шлюза.

$\bar{g} = (g_0, \dots, g_{k-1})$ – выходной вектор шлюза, где каждая его компонента рассчитывается по следующей формуле [5]:

$$g_i = \frac{\exp u_i^{(3)}}{\sum_{j=0}^{k-1} \exp u_j^{(3)}}.$$

Несложно заметить, что $\forall i = 0, \dots, k-1 \ g_i > 0$ и $\sum_{i=0}^{k-1} g_i = 1$.

Тогда общий выход системы $y = \bar{y} \bar{g}^T$, где $\bar{y}$ – вектор выходов экспертов.

На рисунке 1 показана схема взаимодействия ассоциативной гауссовой модели смещения и слабых экспертов для одного класса.

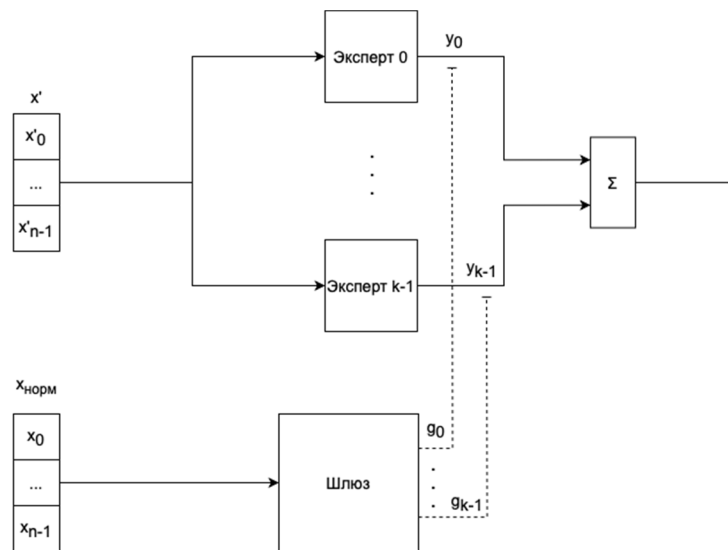


Рисунок 1 – Схема взаимодействия ассоциативной гауссовой модели смещения и слабых экспертов

Figure 1 – Scheme of associative Gaussian mixture model and weak expert interaction

На рисунке 2 показана схема взаимодействия ассоциативной гауссовой модели смещения, перцептрона и слабых экспертов для одного класса.

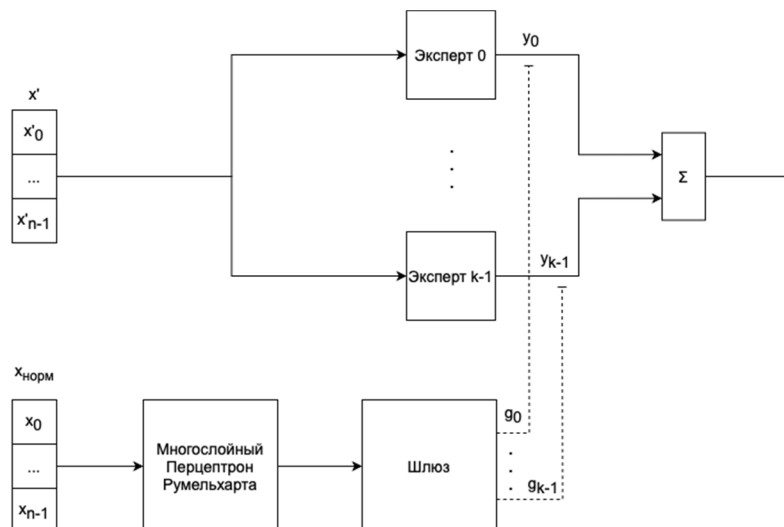


Рисунок 2 – Схема взаимодействия ассоциативной гауссовой модели смещения, перцептрона и слабых экспертов

Figure 2 – Scheme of associative Gaussian mixture model, perceptron and weak expert interaction

На рисунке 3 показана схема взаимодействия ассоциативной гауссовой модели смещения и перцептрона.

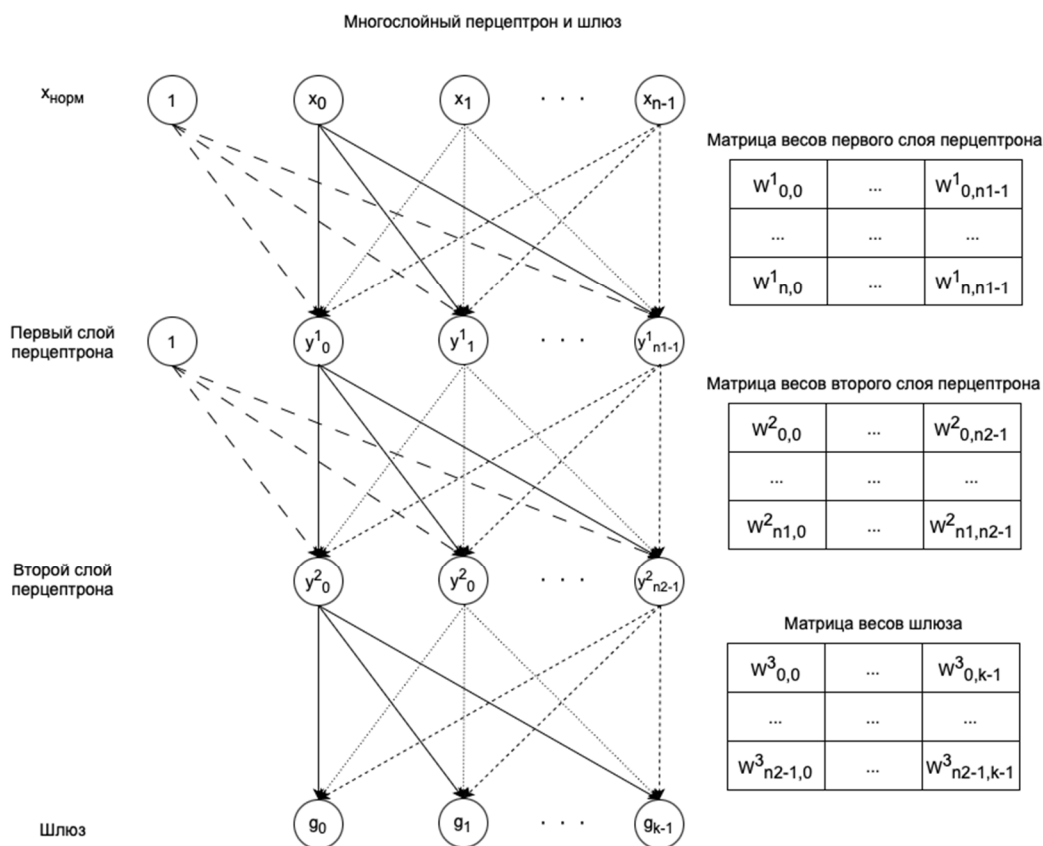


Рисунок 3 – Схема взаимодействия ассоциативной гауссовой модели смещения и перцептрона
Figure 3 – Scheme of associative Gaussian mixture model and perceptron interaction

Выход i -го эксперта есть условное математическое ожидание случайной величины Z при условии, что верно i -е правило, то есть $y_i = M(Z | i)$. Из данных обучения вычислим условное математическое ожидание и условную дисперсию случайной величины Z при условии, что верно i -е правило:

$$\mu_i = \frac{\sum_{\bar{x} \in X_i} y_i(\bar{x})^2}{\text{card } X_i},$$

$$\sigma_i^2 = \frac{\sum_{\bar{x} \in X_i} (y_i(\bar{x}) - \mu_i)^2}{\text{card } X_i},$$

где X_i – конечное подмножество элементов в выборке, для которых известно, что верно i -е правило.

Условимся, что реализация выходной переменной Z при учёте i -го правила распределена нормально, то есть $Z \sim N(y_i, \sigma_i^2)$. Тогда её плотность распределения будет иметь вид [6]:

$$p(z | i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \frac{-(z - y_i)^2}{2\sigma_i^2}.$$

Реализация выходной переменной Z есть Гауссова смесь [7], то есть её плотность распределения имеет вид [8]:

$$p(z) = \sum_{i=0}^{k-1} g_i p(z|i) = \sum_{i=0}^{k-1} g_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp \frac{-(z-y_i)^2}{2\sigma_i^2}.$$

Отсюда ожидаемый выход i -го эксперта:

$$h_i = \frac{g_i p(z|i)}{p(z)} = \frac{g_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp \frac{-(z-y_i)^2}{2\sigma_i^2}}{\sum_{j=0}^{k-1} g_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp \frac{-(z-y_j)^2}{2\sigma_j^2}} = \frac{g_i \frac{1}{\sigma_i} \exp \frac{-(z-y_i)^2}{2\sigma_i^2}}{\sum_{j=0}^{k-1} g_j \frac{1}{\sigma_j} \exp \frac{-(z-y_j)^2}{2\sigma_j^2}}.$$

Введём функцию правдоподобия $L = \prod_{i=1}^k g_i^{kh_i}$ [9, 10]. В силу монотонности функции натурального логарифма максимум функции L будет достигаться в максимуме следующей функции:

$$\begin{aligned} \frac{\ln L}{k} &= \frac{\ln \prod_{i=0}^{k-1} g_i^{kh_i}}{k} = \sum_{i=0}^{k-1} \ln g_i^{h_i} = \sum_{i=0}^{k-1} h_i \ln g_i = \sum_{i=0}^{k-1} h_i \ln \frac{\exp u_i^{(3)}}{\sum_{j=0}^{k-1} \exp u_j^{(3)}} = \sum_{i=0}^{k-1} h_i (u_i^{(3)} - \ln \sum_{j=0}^{k-1} \exp u_j^{(3)}) = \\ &= \sum_{i=0}^{k-1} h_i u_i^{(3)} - \sum_{i=0}^{k-1} h_i \ln \sum_{j=0}^{k-1} \exp u_j^{(3)} = \sum_{i=0}^{k-1} h_i u_i^{(3)} - \ln \left(\sum_{j=0}^{k-1} \exp u_j^{(3)} \right) \sum_{i=0}^{k-1} h_i = \sum_{i=0}^{k-1} h_i u_i^{(3)} - \ln \left(\sum_{j=0}^{k-1} \exp u_j^{(3)} \right). \end{aligned}$$

Последний переход верен в силу того, что $\sum_{i=0}^{k-1} h_i = 1$.

Введём энергию ошибки как функцию, противоположную по знаку функции $\frac{\ln L}{k}$:

$$E = -\frac{\ln L}{k} = \ln \left(\sum_{j=0}^{k-1} \exp u_j^{(3)} \right) - \sum_{i=0}^{k-1} h_i u_i^{(3)} \quad (1)$$

Для настраиваемых параметров сети (матрицы весов двух слоев перцептрона и шлюза) рассчитаем ошибку, то есть разность между значением параметра на следующем шаге и значением этого же параметра на текущем шаге, как произведение скорости обучения η (будем считать, что она постоянна) и соответствующей частной производной энергии ошибки по данному параметру p , причём данная величина должна иметь знак, противоположный знаку соответствующей частной производной, поскольку задача состоит в минимизации энергии ошибки:

$$\Delta p = p_{n+1} - p_n = -\eta \frac{\partial E}{\partial p}.$$

Применим данную формулу к матрицам $W^{(1)}$, $W^{(2)}$ и $W^{(3)}$:

$$\Delta W^{(1)} = W_{n+1}^{(1)} - W_n^{(1)} = -\eta \frac{\partial E}{\partial W^{(1)}},$$

$$\Delta W^{(2)} = W_{n+1}^{(2)} - W_n^{(2)} = -\eta \frac{\partial E}{\partial W^{(2)}}, \quad (2)$$

$$\Delta W^{(3)} = W_{n+1}^{(3)} - W_n^{(3)} = -\eta \frac{\partial E}{\partial W^{(3)}}. \quad (3)$$

Найдём частные производные функции E по элементам матрицы весов шлюза $W^{(3)}$, используя правило дифференцирования сложной функции:

$$\frac{\partial E}{\partial w_{0,0}^{(3)}} = \sum_{i=0}^{k-1} \frac{\partial E(u_i^{(3)})}{\partial w_{0,0}^{(3)}} = \frac{\partial E(u_0^{(3)})}{\partial w_{0,0}^{(3)}} = \frac{\partial E}{\partial u_0^{(3)}} \frac{\partial u_0^{(3)}}{\partial w_{0,0}^{(3)}} = \frac{\partial E}{\partial u_0^{(3)}} y_0^{(2)},$$

$$\begin{aligned}\frac{\partial E}{\partial w_{0,1}^{(3)}} &= \sum_{i=0}^{k-1} \frac{\partial E(u_i^{(3)})}{\partial w_{0,1}^{(3)}} = \frac{\partial E(u_1^{(3)})}{\partial w_{0,1}^{(3)}} = \frac{\partial E}{\partial u_1^{(3)}} \frac{\partial u_1^{(3)}}{\partial w_{0,1}^{(3)}} = \frac{\partial E}{\partial u_1^{(3)}} y_0^{(2)}, \\ \frac{\partial E}{\partial w_{1,0}^{(3)}} &= \sum_{i=0}^{k-1} \frac{\partial E(u_i^{(3)})}{\partial w_{1,0}^{(3)}} = \frac{\partial E(u_0^{(3)})}{\partial w_{1,0}^{(3)}} = \frac{\partial E}{\partial u_0^{(3)}} \frac{\partial u_0^{(3)}}{\partial w_{1,0}^{(3)}} = \frac{\partial E}{\partial u_0^{(3)}} y_1^{(2)}, \\ &\dots \\ \frac{\partial E}{\partial w_{i,j}^{(3)}} &= \sum_{l=0}^{k-1} \frac{\partial E(u_l^{(3)})}{\partial w_{i,j}^{(3)}} = \frac{\partial E(u_j^{(3)})}{\partial w_{i,j}^{(3)}} = \frac{\partial E}{\partial u_j^{(3)}} \frac{\partial u_j^{(3)}}{\partial w_{i,j}^{(3)}} = \frac{\partial E}{\partial u_j^{(3)}} y_i^{(2)}.\end{aligned}$$

Запишем в матричном виде:

$$\frac{\partial E}{\partial W^{(3)}} = (\bar{y}^{(2)})^T \frac{\partial E}{\partial \bar{u}^{(3)}}. \quad (4)$$

Для того чтобы найти частную производную энергии ошибки по матрицы весов шлюза, необходимо найти $\frac{\partial E}{\partial \bar{u}^{(3)}}$. Для этого продифференцируем уравнение (1):

$$\frac{\partial E}{\partial u_i^{(3)}} = \frac{\partial}{\partial u_i^{(3)}} \left(\ln \left(\sum_{j=0}^{k-1} \exp u_j^{(3)} \right) - \sum_{i=0}^{k-1} h_i u_i \right) = \frac{\exp u_i^{(3)}}{\sum_{j=0}^{k-1} \exp u_j^{(3)}} - h_i = g_i - h_i.$$

Запишем в векторном виде:

$$\frac{\partial E}{\partial \bar{u}^{(3)}} = \bar{g} - \bar{h}. \quad (5)$$

Учитывая формулы (3) - (5), получаем:

$$\Delta W^{(3)} = -\eta \frac{\partial E}{\partial W^{(3)}} = -\eta (\bar{y}^{(2)})^T \frac{\partial E}{\partial \bar{u}^{(3)}} = -\eta (\bar{y}^{(2)})^T (\bar{g} - \bar{h}). \quad (6)$$

Как видно из формулы (2), для нахождения ошибки матрицы весов второго слоя перцептрона $W^{(2)}$ необходимо вычислить $\frac{\partial E}{\partial W^{(2)}}$, который мы вычислим аналогично получению формулы (5):

$$\frac{\partial E}{\partial W^{(2)}} = (1, y_0^{(1)}, \dots, y_{n_1-1}^{(1)})^T \frac{\partial E}{\partial \bar{u}^{(2)}}. \quad (7)$$

Найдём $\frac{\partial E}{\partial \bar{u}^{(2)}}$:

$$\begin{aligned}\frac{\partial E}{\partial u_0^{(2)}} &= \sum_{i=0}^{n_2-1} \frac{\partial E(y_i^{(2)})}{\partial u_0^{(2)}} = \frac{\partial E(y_0^{(2)})}{\partial u_0^{(2)}} = \frac{\partial E}{\partial y_0^{(2)}} \frac{\partial y_0^{(2)}}{\partial u_0^{(2)}} = \frac{\partial E}{\partial y_0^{(2)}} \frac{\partial \tanh u_0^{(2)}}{\partial u_0^{(2)}} = \frac{\partial E}{\partial y_0^{(2)}} \frac{1}{\cosh^2 u_0^{(2)}} \Rightarrow \\ &\frac{\partial E}{\partial u_i^{(2)}} = \frac{\partial E}{\partial y_i^{(2)}} \frac{1}{\cosh^2 u_i^{(2)}}.\end{aligned}$$

Тогда в векторном виде имеет место равенство:

$$\frac{\partial E}{\partial \bar{u}^{(2)}} = \frac{\partial E}{\partial \bar{y}^{(2)}} \odot \frac{1}{\cosh^2 \bar{u}^{(2)}}. \quad (8)$$

Теперь отыщем $\frac{\partial E}{\partial \bar{y}^{(2)}}$:

$$\frac{\partial E}{\partial y_0^{(2)}} = \sum_{i=0}^{k-1} \frac{\partial E(u_i^{(3)})}{\partial y_0^{(2)}} = \sum_{i=0}^{k-1} \frac{\partial E}{\partial u_i^{(3)}} \frac{\partial u_i^{(3)}}{\partial y_0^{(2)}} = \sum_{i=0}^{k-1} \frac{\partial E}{\partial u_0^{(3)}} w_{0,i}^{(3)} \Rightarrow \frac{\partial E}{\partial y_i^{(2)}} = \sum_{j=0}^{k-1} \frac{\partial E}{\partial u_i^{(3)}} w_{i,j}^{(3)}.$$

Запишем в векторном виде:

$$\frac{\partial E}{\partial \bar{y}^{(2)}} = \frac{\partial E}{\partial \bar{u}^{(3)}} W^{(3)T}. \quad (9)$$

С учётом формул (2), (5) и (7) – (9) имеем:

$$\begin{aligned} \Delta W^{(2)} &= -\eta \frac{\partial E}{\partial W^{(2)}} = -\eta(1, y_0^{(1)}, \dots, y_{n_1-1}^{(1)})^T \frac{\partial E}{\partial \bar{u}^{(2)}} = -\eta(1, y_0^{(1)}, \dots, y_{n_1-1}^{(1)})^T \left(\frac{\partial E}{\partial \bar{y}^{(2)}} \odot \frac{1}{\cosh^2 \bar{u}^{(2)}} \right) = \\ &= -\eta(1, y_0^{(1)}, \dots, y_{n_1-1}^{(1)})^T \left(\frac{\partial E}{\partial \bar{u}^{(3)}} W^{(3)T} \odot \frac{1}{\cosh^2 \bar{u}^{(2)}} \right) = \\ &= -\eta(1, y_0^{(1)}, \dots, y_{n_1-1}^{(1)})^T ((\bar{g} - \bar{h}) W^{(3)T} \odot \frac{1}{\cosh^2 \bar{u}^{(2)}}). \end{aligned} \quad (10)$$

Аналогичным образом найдём ошибку матрицы весов первого слоя перцептрона $\Delta W^{(1)}$:

$$\begin{aligned} \Delta W^{(1)} &= -\eta \frac{\partial E}{\partial W^{(1)}} = -\eta(1, x_0^{(1)}, \dots, x_{n-1}^{(1)})^T \frac{\partial E}{\partial \bar{u}^{(1)}} = -\eta(1, x_0^{(1)}, \dots, x_{n-1}^{(1)})^T \left(\frac{\partial E}{\partial \bar{h}^{(1)}} \odot \frac{1}{\cosh^2 \bar{u}^{(1)}} \right) = \\ &= -\eta(1, x_0^{(1)}, \dots, x_{n-1}^{(1)})^T \left(\frac{\partial E}{\partial \bar{u}^{(2)}} W^{(2)T} \odot \frac{1}{\cosh^2 \bar{u}^{(1)}} \right) = \\ &= -\eta(1, x_0^{(1)}, \dots, x_{n-1}^{(1)})^T \left(\left(\frac{\partial E}{\partial \bar{y}^{(2)}} \odot \frac{1}{\cosh^2 \bar{u}^{(2)}} \right) W^{(2)T} \odot \frac{1}{\cosh^2 \bar{u}^{(1)}} \right) = \\ &= -\eta(1, x_0^{(1)}, \dots, x_{n-1}^{(1)})^T \left(\left(\frac{\partial E}{\partial \bar{u}^{(3)}} W^{(3)T} \odot \frac{1}{\cosh^2 \bar{u}^{(2)}} \right) W^{(2)T} \odot \frac{1}{\cosh^2 \bar{u}^{(1)}} \right) = \\ &= -\eta(1, x_0^{(1)}, \dots, x_{n-1}^{(1)})^T \left(((\bar{g} - \bar{h}) W^{(3)T} \odot \frac{1}{\cosh^2 \bar{u}^{(2)}}) W^{(2)T} \odot \frac{1}{\cosh^2 \bar{u}^{(1)}} \right). \end{aligned} \quad (11)$$

Таким образом, мы получили искомые формулы (6), (10) и (11) для нахождения ошибки настраиваемых параметров искусственной нейронной сети.

Исследовательская часть

Для проверки модифицированного метода ассоциативной гауссовой модели смещения рассмотрим задачу принадлежности точки (х, у) какому-либо классу. Рассмотрим случай, когда эксперт точно определяет эту принадлежность для определённой области. Пусть область первого эксперта $S_1 = [4; 8]^2$, а область второго - $S_2 = [4; 8] \times [-8; -4]$ (рисунок 4).

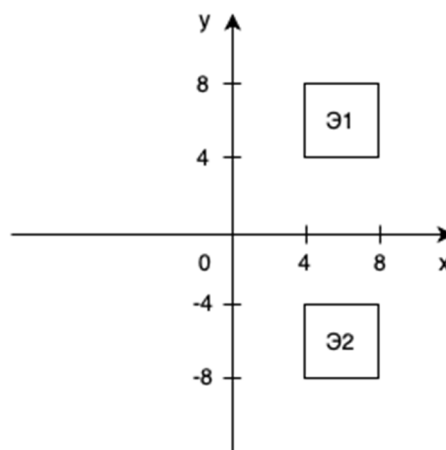


Рисунок 4 – Области экспертов, которые определяют только в одной области

Figure 4 – Expert areas that define only one area

Ассоциативная гауссова модель смещения обучается при помощи 8000 случайных точек из области $S = [-15; 15]^2$. Далее для 38 случайных точек с помощью обученной ассоциативной гауссовой модели смещения находятся значения выходов. Результат нахождения значений выходов показан на рисунке 5.

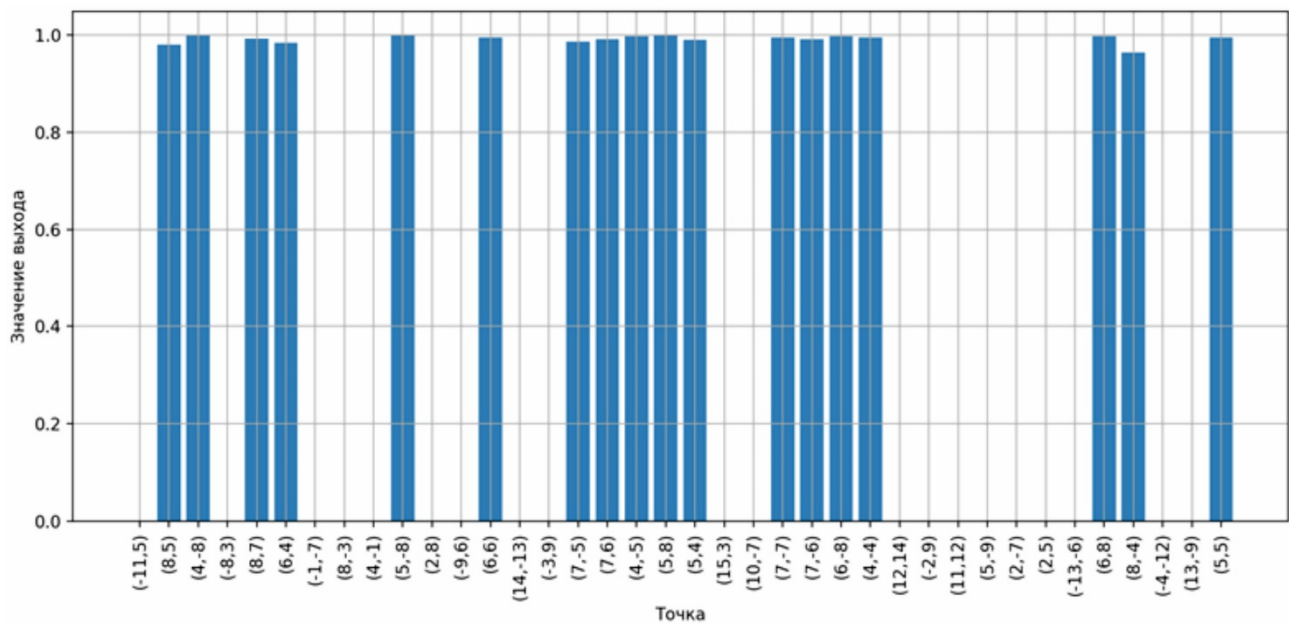


Рисунок 5 – Результат нахождения значений выходов ассоциативной гауссовой модели смещения

Figure 5 – The result of finding the values of associative Gaussian mixture model outputs

Рассмотрим случай, когда эксперт точно определяет принадлежность точки в нескольких областях. Область первого эксперта $S'_1 = [-8; -4]^2 \cup [4; 8]^2$, область второго – $S'_2 = [-8; -4] \times [4; 8] \cup [4; 8] \times [-8; -4]$ (рисунок 6).

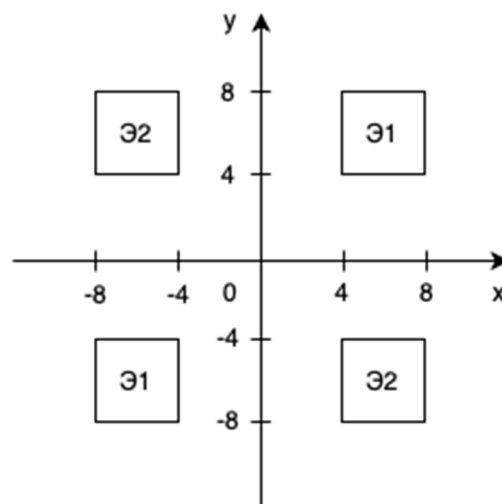


Рисунок 6 – Области экспертов, которые определяют в нескольких областях

Figure 6 – Areas of experts who define in several areas

Ассоциативная гауссова модель смещения обучается при помощи 8000 случайных точек из области $S = [-15; 15]^2$. Далее для 40 случайных с помощью обученной ассоциативной гауссовой модели смещения находятся значения выходов. Результат нахождения значений выходов показан на рисунке 7.

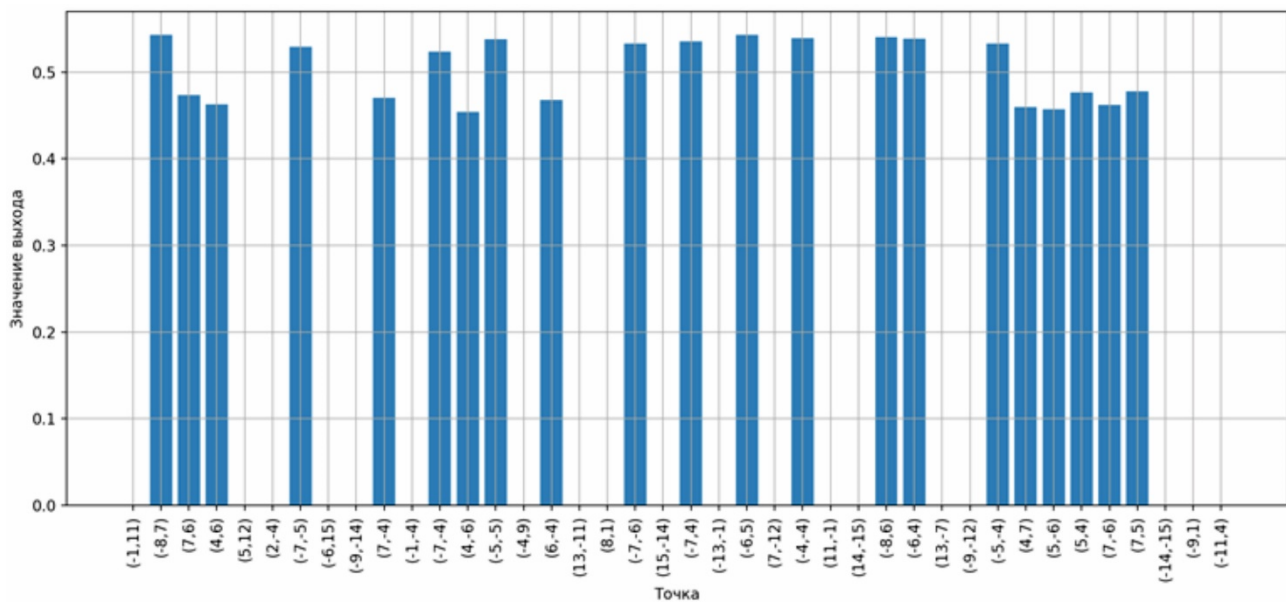


Рисунок 7 – Результат нахождения значений выходов с помощью ассоциативной гауссовой модели смещения

Figure 7 – The result of finding outputs values using associative Gaussian mixture model

По полученным результатам можно заметить, что обучение ассоциативной гауссовой модели смещения практически не происходит (значение выходов близко к 0.5). Обучим модифицированный метод ассоциативной гауссовой модели смещения с помощью тех же данных и найдем значения выходов для этих же точек. Результат нахождения выходов показан на рисунке 8. По результатам видно, что у модифицированного метода ассоциативной гауссовой модели смещения намного лучше точность.

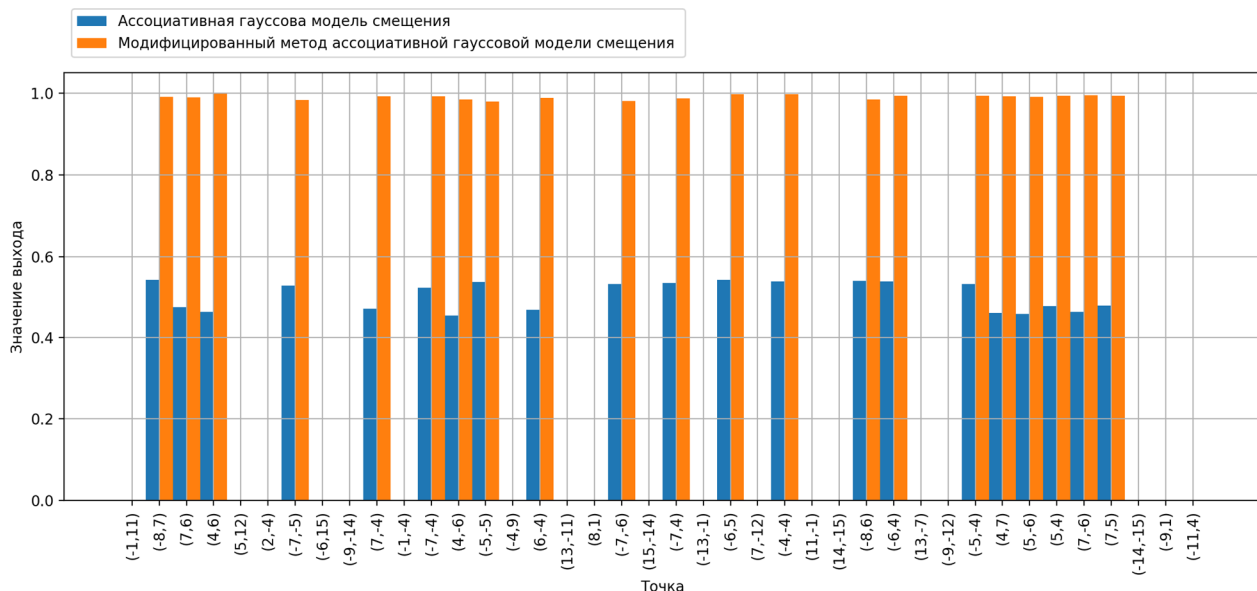


Рисунок 8 – Результат нахождения значений выходов с помощью модифицированной ассоциативной гауссовой модели смещения

Figure 8 – The result of finding output values using modified associative Gaussian mixture model

Заключение

В ходе научно-исследовательской работы реализован модифицированный метод ассоциативной гауссовой модели смещения, проведено сравнение со стандартной ассоциативной гауссовой моделью смещения, а также проведена проверка разработанного метода на примере принадлежности чисел какой-либо области.

Предлагаемый метод может использоваться в системах, где применяются слабые эксперты, в качестве обобщающего алгоритма, например для обнаружения объектов на снимках Земли из космоса. Объект может быть зашумлен, искажен или виден не полностью. Для решения этой задачи можно использовать слабые эксперты: каждый слабый эксперт дает достоверный ответ лишь в определенных областях, например первый распознает объекты, которые видны не полностью и искажены, а за зашумленные объекты отвечает другой слабый эксперт. Чтобы получить корректный результат ансамбля, необходимо ранжировать экспертов с помощью разработанного метода.

Библиографический список

1. Евдокимова К. В., Тассов К. Л. Анализ алгоритмов, построенных на «слабых экспертах» // Вестник РГПУ. Серия «Информатика. Информационная безопасность. Математика». 2022.
2. Хайкин, Саймон. Нейронные сети: полный курс, 2-е издание: Пер. с англ. – М.: Издательский дом «Вильямс», 2006. 1104 с.: ил. Парал. тит. англ.
3. McLachlan G.J. and K.E. Basford. Mixture Models: Inference and Applications to Clustering, New York: Marcel Dekker, 1988.
4. Jacobs R. A., M. I. Jordan, S. J. Nowlan and G. E. Hilton. «Adaptive mixtures of local experts», Neural Computation, 1991, vol. 3, pp. 79-87.
5. Bridle J.S. «Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition», Neuro-computing: Algorithms, Architecture and Applications, F. Fogelman-Soulie and J. Hraut, eds., 1990, New-York: Springer-Verlag.
6. Jordan M. I. «A statistical approach to decision tree modeling». Proceedings of the Seventh Annual ACM Conference on Computation Learning Theory, New York: ACM Press, 1994.
7. Tittertington D. M., A. F. M. Smith and V. E. Makov. Statistical Analysis of Finite Mixture Distributions, New York: Wiley, 1985.
8. Jordan M. I. and R. A. Jacobs. «Modular and Hierarchical Learning Systems», in M.A. Arbib, ed., The Handbook of Brain Theory and Neural Networks, 1995, pp. 579-583, Cambridge, MA: MIT Press.
9. Casella, George; Berger, Roger L. (2002). Statistical Inference (2nd ed.). Duxbury. 290. p.
10. Nowlan S. J. Maximum likelihood competitive learning // Advances in Neural Information Processing Systems, 1990, vol. 2, pp. 574-582.

UDC 004.8

MODIFICATION OF ASSOCIATIVE GAUSSIAN MIXTURE MODEL IN EXPERT ASSESSMENTS

D. V. Gorbunov, post-graduate student, BMSTU, Moscow, Russia;

orcid.org/0000-0002-4646-2636, e-mail: darkwen3@mail.ru.

K. L. Tassov, lecturer, BMSTU, Moscow, Russia;

e-mail: ktassov@bmstu.ru.

S. V. Telegin, bachelor, BMSTU, Moscow, Russia;

orcid.org/0009-0000-6637-7124, email: semen-telegin@mail.ru.

The idea of ensemble training is to organize such a pool of experts that will allow them to be combined into a common system. Various prediction algorithms including neural networks can be considered as an expert. The essence of ensemble methods is that each of the experts provides its own response which can then be used by a generalizing algorithm. One of the generalizing algorithms for classification is the associative Gaussian mixture model. One expert for one class determines the result in only one area that is a disadvantage. Therefore, a modified method of associative Gaussian mixture model is proposed where one expert determines the result in several areas at once. Within the framework of the method an algorithm is proposed that includes the following steps: input data normalization, model training based on normalized input data, finding output values. The method can serve in any areas in which "local experts" are used and where it is necessary to rank experts. The article discusses the unification of associative Gaussian mixture model and

Rumelhart perceptron, input data preprocessing, training and prediction using a modified method of associative Gaussian mixture model, the results of obtained output values are presented.

Keywords: *modified associative Gaussian mixture model, associative Gaussian mixture model, weak experts, neural net-works, Rumelhart perceptron.*

DOI: 10.21667/1995-4565-2023-85-54-64

References

1. **Evdokimova K. V., Tassov K. L.** Analiz algoritmov, postroennyh na «slabyh ekspertah». *Vestnik RGGU. Seriya «Informatika. Informacionnaya bezopasnost'. Matematika»*. 2022. (in Russian).
2. **Haikin, Simon.** *Nejronnye seti: polnyj kurs, 2-e izdanie* (Neural Networks: A Complete course, 2nd edition). Per. s angl. Moscow: Izdatel'skij dom «Vil'yams». 2006. 1104 p. (in Russian).
3. **McLachlan G. J. and K. E. Basford.** *Mixture Models: Inference and Applications to Clustering*, New York: Marcel Dekker, 1988.
4. **Jacobs R. A., M. I. Jordan, S. J. Nowlan and G. E. Hilton.** «Adaptive mixtures of local experts», *Neural Computation*, 1991, vol. 3, pp. 79-87.
5. **Bridle J.S.** «Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition», *Neuro-computing: Algorithms, Architecture and Applications*, F. Fougelman-Soulie and J. Hraut, eds., 1990, New-York: Springer-Veriag.
6. **Jordan M. I.** «A statistical approach to decision tree modeling». *Proceedings of the Seventh Annual ACM Conference on Computation Learning Theory*, New York: ACM Press, 1994.
7. **Titterington D. M., A. F. M. Smith and V. E. Makov.** *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley, 1985.
8. **Jordan M. I. and R. A. Jacobs.** «Modular and Hierarchical Learning Systems», in *M.A. Arbib, ed., The Handbook of Brain Theory and Neural Networks*, 1995, pp. 579-583, Cambridge, MA: MIT Press.
9. **Casella, George; Berger, Roger L.** (2002). *Statistical Inference* (2nd ed.). Duxbury. 290 p.
10. **Nowlan S. J.** Maximum likelihood competitive learning. *Advances in Neural Information Processing Systems*, 1990, vol. 2, pp. 574-582.