



данных в электронных архивах. Реализуется путём создания модуля-конвертера, позволяющего, с одной стороны, в пакетном режиме осуществлять миграцию форматов, данных от старых к новым, а с другой – определять и проводить предварительный анализ форматов данных непосредственно в процессе наполнения архива на предмет соответствия утверждённым политикам хранения данных. Пользователи системы смогут самостоятельно настраивать политики хранения данных таким образом, что старые или неподдерживаемые более форматы будут автоматически преобразовываться программным обеспечением системы в современные форматы.

Заключение

Организация интеллектуальной системы управления электронными архивами позволяет обеспечивать формирование электронного фонда документов и автоматизация процессов работы с ними, а так же позволяет осуществлять создание, управление и работу с электронными архивами государственных и коммерческих организаций, включая интеллектуальный семантический поиск и управление всеми формами представления информации, поддержку принятия решения на основе архивных данных и другие интеллектуальные возможности работы с данными.

О.П. Солдатова, Д.З. Иваев

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ АЛГОРИТМОВ ОБУЧЕНИЯ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА ПРИ РЕШЕНИИ ЗАДАЧИ КЛАССИФИКАЦИИ

(Самарский национальный исследовательский университет
имени академика С.П. Королёва)

Целью данной работы является изучение эффективности алгоритмов обучения многослойного перцептрона при классификации данных, решение задачи классификации данных на основе модели многослойного перцептрона с одним или несколькими скрытыми слоями, сравнение полученных результатов при различных алгоритмах обучения сети.

Сигнал, подаваемый на вход i -го нейрона k -го слоя, проходит через сумматор, в котором выполняется взвешенное суммирование элементов входного сигнала:

$$u_i^{(k)} = \sum_{j=1}^m x_j w_{ij}^{(k)}, \text{ где } m - \text{число нейронов } k-1 \text{-го слоя; } i = 1..n_k.$$

В случае входного слоя $x_j - j$ -ый компонент вектора входных данных; $m -$ размерность вектора x ; $n_k -$ число элементов в слое k ; $w_{ij}^{(k)}$ - соответствующий весовой коэффициент.

Выход каждого элемента рассчитывается по следующей формуле:



$$y_i^{(k)} = f(u_i^{(k)}), \text{ где } f(u_i^{(k)}) = 1/(1 + e^{-\beta x}) \text{ (здесь } \beta = 1).$$

Этап обучения нейронной сети заключается в подборе таких значений весов $w_{ij}^{(k)}$ для всех слоев сети, чтобы при заданном векторе x , получить на выходе значения сигналов y_i , позволяющие определить принадлежность данного вектора к одному из классов. Для решения этой задачи используется градиентный алгоритм наискорейшего спуска с методом обратного распространения ошибки и эвристические алгоритмы RPROP и Quickprop [1].

Смысл градиентного алгоритма наискорейшего спуска с обратным распространением ошибки заключается в вычислении параметра коррекции δ_i для каждого элемента нейронной сети $\delta_i = (y_i - d_i) \frac{\partial y_i}{\partial u_i}$ (в случае выходных нейронов), где $d_i -$ ожидаемое значение на i -м выходе, или $\delta_i^{(k)} = (\sum_j \delta_j^{(k+1)} w_{ji}^{(k+1)}) \frac{\partial y_i}{\partial u_i}$ (для нейронов скрытых слоев k), и коррекции весов по формуле $w_{ij}^{(k)} = w_{ij}^{(k)} + \Delta w_{ij}^{(k)}$, где $\Delta w_{ij}^{(k)} = -\eta \delta_i^{(k)} u_j^{(k-1)}$. Приведенные вычисления выполняются для всех векторов обучающей выборки.

Алгоритмы эвристического типа представляют собой в основном модификацию методов наискорейшего спуска или сопряженных градиентов. Подобные модификации широко известных алгоритмов связаны с внесением в них некоторых изменений, ускоряющих процесс обучения.

Алгоритм QuickProp содержит элементы, предотвращающие заикливание в точке неглубокого локального минимума.

Вес w_{ij} на k -ом шаге алгоритма изменяется согласно правилу:

$$\Delta w_{ij}(k) = -\eta_k \left[\frac{\partial E(w(k))}{\partial w_{ij}} + \gamma w_{ij}(k) \right] + \alpha_{ij}^k \Delta w_{ij}(k-1)$$

Коэффициент γ , обычно имеющий малую величину (порядка 10^{-4}), приводит к уменьшению весов вплоть до разрыва соответствующих взвешенных связей.

Константа $\eta_k -$ в данном алгоритме может иметь ненулевое значение η_0 (как правило, $0.01 < \eta_0 < 0.6$ на старте процесса обучения, когда $\Delta w_{ij}(k-1) = 0$

или когда $\Delta w_{ij} \left[\frac{\partial E(w(k))}{\partial w_{ij}} + \gamma w_{ij}(k) \right] > 0$, либо нулевое значение – в противном случае. Фактор момента коэффициент подбирается индивидуально для каждого веса по правилу:

$$\alpha_{ij}^k = \begin{cases} \alpha_{\max}, & \text{где } \beta_{ij}(k) > \alpha_{\max}, \text{ где } S_{ij}(k) \Delta w_{ij}(k-1) \beta_{ij}(k) < 0; \\ \beta_{ij}(k) \end{cases}$$

причем:

$$S_{ij}(k) = \frac{\partial E(w(k))}{\partial w_{ij}} + \gamma w_{ij}(k), \quad \beta_{ij}(k) = \frac{S_{ij}(k)}{S_{ij}(k-1) - S_{ij}(k)}, \text{ а}$$



Константа α_{\max} – это максимальное значение коэффициента момента, которое принимается равным 1,75.

В алгоритме RPROP при уточнении весов учитывается только знак гради-

$$\Delta w_{ij}(t) = -\eta_{ij}(t) \operatorname{sgn} \left(\frac{\partial E(w(t))}{\partial w_{ij}} \right)$$

ентной составляющей, а ее значение игнорируется:

Коэффициент обучения подбирается индивидуально для каждого веса w_{ij} с учетом изменения значения градиента:

$$\eta_{ij}(t) = \begin{cases} \min(a\eta_{ij}(t-1), \eta_{\max}) & \text{для } S_{ij}(t)S_{ij}(t-1) > 0 \\ \max(b\eta_{ij}(t-1), \eta_{\min}) & \text{для } S_{ij}(t)S_{ij}(t-1) < 0, \\ \eta_{ij}(t-1) & \text{в остальных случаях} \end{cases}$$

$$S_{ij}(t) = \frac{\partial E(w(t))}{\partial w_{ij}}, \text{ а и } b \text{ – константы: } a=1.2; b=0.5.$$

где

Минимальное и максимальное значения коэффициента обучения составляют $\eta_{\min} = 10^{-6}$ и $\eta_{\max} = 50$. Функция $\operatorname{sgn}(\)$ принимает значение, равное знаку градиента.

Исследования проводились на данных, взятых из репозитория UCI Калифорнийского университета, для задач классификации ирисов и винных напитков [2]. Параметры многослойного перцептрона были определены в результате исследований, представленных в [3].

Были исследованы зависимости среднеквадратичного отклонения (СКО) для различных алгоритмов обучения при следующих параметрах: обучающая выборка из 90 векторов данных, характеризующих виды ирисов (по 30 для каждого из 3-х классов); входной слой – 4 нейрона; скрытый слой – 5; выходной – 3 нейрона; коэффициент обучения $\eta = 0,02$; пороговое значение определяющая точность – 0,01; коэффициент наклона сигмоидальной функции $\beta = 1$; ограничение на количество циклов – 2500 итераций. Тестовая выборка состоит из 60 векторов, не входящих в обучающую выборку (по 20 для каждого из 3-х классов). Зависимость СКО от коэффициента наклона сигмоидальной функции представлена в таблице 1.



Таблица 1. Зависимость СКО от используемого алгоритма обучения при классификации ирисов

Алгоритм обучения	Среднеквадратичная ошибка	Процент ошибок
Backpropagation	0,0139	3,33
QuickProp	0,0508	6
RPROP	0,0535	6

Аналогично исследованы зависимости СКО при следующих параметрах обучения: обучающая выборка из 90 векторов данных характеризующих виды винных напитков (по 30 для каждого из 3-х классов); входной слой – 13 нейронов; скрытый слой – 17; выходной – 3 нейрона; коэффициент обучения $\eta = 0,02$; пороговое значение определяющая точность – 0,003; ограничение на количество циклов – 2500 итераций. Тестовая выборка состоит из 54 векторов, не входящих в обучающую выборку (по 18 для каждого из 3-х классов). Зависимость СКО от коэффициента наклона сигмоидальной функции при классификации винных напитков представлена в таблице 2.

Таблица 2. Зависимость СКО от используемого алгоритма обучения при классификации винных напитков

Алгоритм обучения	Среднеквадратичная ошибка	Процент ошибок
Backpropagation	0,0149	3,7
QuickProp	0,0179	3,7
RPROP	0,0195	3,7

При анализе результатов сделан вывод, что эвристические алгоритмы обучения многослойного перцептрона, хотя и не имеют серьезного теоретического обоснования, дают результаты близкие к результатам градиентного метода наискорейшего спуска. Так как процесс обучения в эвристических алгоритмах происходит быстрее, то их применение на практике становится обоснованным.

Литература

- Осовский, С. Нейронные сети для обработки информации [Текст]/ С.Осовский. – М.: Финансы и статистика, 2002 – 344 с.
- UCI Machine Learning Repository [Электронный ресурс]. – URL: <http://archive.ics.uci.edu/ml/> (Дата обращения: 14.02.2015)
- Солдатова, О.П., Иваев Д.З. Решение задачи классификации при помощи многослойного перцептрона. // Труды международной научно-технической конференции «Перспективные информационные технологии (ПИТ 2015)» – Самара: Издательство Самарского научного центра РАН, 2015. – Том1, с.209-213 – ISBN 978-5-93424-704-9