

# Анализ алогичного поведения нейронных сетей в задачах классификации при экологическом мониторинге водоемов

Т. Р. Жангиров<sup>1</sup>, А. С. Перков<sup>2</sup>, А. А. Лисс<sup>3</sup>,  
А. В. Экало<sup>4</sup>, С. А. Иванова<sup>5</sup>

Санкт-Петербургский государственный  
электротехнический университет «ЛЭТИ»  
им. В. И. Ульянова (Ленина)

<sup>1</sup>xxxtestxxx@yandex.ru, <sup>2</sup>invecto@yandex.ru,  
<sup>3</sup>anna.liss@moevm.info, <sup>4</sup>ekalo@nicetu.spb.ru,  
<sup>5</sup>sofiia8888@gmail.com

Н. Ю. Григорьева

Санкт-Петербургский научно-исследовательский центр  
экологической безопасности РАН  
renes3@mail.ru

**Аннотация.** Рассматривается проблема алогичного поведения классифицирующих нейронных сетей, возникающая, в задачах экологического мониторинга водоемов. В работе представлена оригинальная методика анализа нейронных сетей, позволяющая выявить алогичное поведение, то есть ложноположительные результаты классификации и обобщения. Анализируется проблема обобщения наблюдений, принадлежащих к классам, не задействованным в обучении, так называемая проблема внешнего обобщения. В рамках предложенной методики, в частности, рассматривается способ проверки качества обучения классифицирующей нейронной сети, основанный на сравнении результатов классификации с результатами линейного дискриминантного анализа. Апробирование разработанной методики проводится в рамках решения актуальной задачи экологии - классификации сине-зеленых водорослей по спектрам собственной флуоресценции при токсичных "цветениях" водоемов.

**Ключевые слова:** нейронные сети; классификация; экологический мониторинг

## I. ВВЕДЕНИЕ

Задача классификации – одна из задач обработки данных, заключающаяся в сопоставлении наблюдений, характеризуемых набором признаков, заранее известным целевым классам [1]. Решение задачи классификации наиболее актуально в областях, связанных с проведением экспериментов с большим объемом наблюдений. Решение задачи классификации, обычно сводится к построению математической модели, которая каждому классу приводит в соответствие некую область пространства признаков, минимизируя вероятность ошибки классификации. Существует большое количество моделей, которые по-разному выражают зависимость между набором признаков и целевым классом, и, следовательно, требуют разных алгоритмов их построения [2]. В настоящее время, в связи с активным развитием области машинного обучения, искусственные нейронные сети (ИНС) стали одним из

самых популярных способов представления модели классификации [3]. Главной отличительной особенностью ИНС является то, что в результате построения ИНС (в процессе обучения) на одном и том же наборе данных и при одних и тех параметрах, описывающих структуру сети, могут получаться различные модели. Наличие серии ИНС, различающихся только набором весовых коэффициентов, позволяет выбирать модель, которая лучше всего осуществляет классификацию наблюдений. В то же время, это создает проблему выбора лучшей модели из набора обученных ИНС, поскольку одновременно несколько моделей могут давать одинаковую точность классификации. Для решения данной проблемы, после обучения ИНС, следует проанализировать ее способность к обобщению, то есть оценить результаты классификации данных, не используемых в процессе обучения ИНС. Обычно качество обобщения проверяется на наблюдениях, которые принадлежат к известным классам, описанным внутри модели, то есть проводится оценка *внутреннего обобщения*. Однако, в таких областях как биология и экология, в ходе исследований могут появляться новые классы наблюдений, которые не использовались при построении модели. В этом случае, следует также проверять способность ИНС к *внешнему обобщению*, то есть анализировать реакцию ИНС на наблюдения, принадлежащие к неизвестным ей классам. В некоторых случаях такая постановка задачи может приводить к проблеме *алогичного поведения* ИНС. Алогичным поведением в данном случае будем называть ситуацию, когда обученная ИНС дает высокую точность классификации и имеет хорошую способность к внутреннему обобщению, но при этом дает сбои во внешнем обобщении [4]. В связи с этим весьма актуальным становится разработка новых методик для выявления и анализа проблем алогичного поведения классифицирующих ИНС.

В данной работе, авторами предлагается оригинальная методика анализа качества работы ИНС для выявления их

алогичного поведения, в том числе с применением линейного дискриминантного анализа (ЛДА) [5]. Показана эффективность использования данной методики при решении конкретной задачи классификации сине-зеленых водорослей по спектрам собственной флуоресценции при мониторинге токсичных "цветений" водоемов.

## II. ИСХОДНЫЕ ДАННЫЕ

В данной работе исходными данными являются спектры собственной флуоресценции цианобактериальных клеток, принадлежащих к 23 различным штаммам (классам). Обучение проводится на 16 классах, а 7 классов оставлены вне модели классификатора. Каждое наблюдение из набора данных описывает серию из 7 спектров, снятых с единичной клетки. Пространство признаков, размерностью 63, формируется на основе ряда интегральных и статистических характеристик, описывающих форму каждого спектра серии [6]. В данной работе, каждый класс обозначается соответствующим номером штамма в коллекции CALU Ресурсного центра «Культивирование микроорганизмов» НП СПбГУ [7].

## III. АРХИТЕКТУРА ИНС

### A. Параметры модели ИНС

Для классификации, используется ИНС прямого распространения с одним скрытым слоем. Размер входного слоя равен количеству признаков наблюдения (63 нейрона), а выходной слой имеет 16 нейронов, соответствующих количеству целевых классов. По результатам проведенного исследования количество нейронов на скрытом слое выбрано равным 31 [6]. В качестве функции активации на скрытом слое используется гиперболический тангенс, а на выходном слое функция Softmax (1), которая позволяет интерпретировать выходной слой, как распределение вероятностей принадлежности к каждому из классов.

$$f(\vec{x}) = e^{x_i} / \sum_{j=1}^N e^{x_j}, \quad (1)$$

где  $x_i$  – значение сигнала на  $i$ -ом нейроне до активации,  $N$  – количество нейронов на выходном слое.

### B. Алгоритм обучения сети

Для обучения ИНС используется стандартный метод градиентного спуска с поправкой Нестерова [8]. Данный метод характеризуется скоростью обучения  $\eta$ , задающей величину изменения весовых коэффициентов, и коэффициентами момента  $\mu$  и регуляризации  $\rho$ , снижающие вероятность попадания в локальный минимум. Значения этих параметров в данной работе выбраны:  $\eta = 0.05$ ,  $\mu = 0.001$  и  $\rho = 0.001$ . В используемом методе скорость обучения меняется в ходе обучения в зависимости от изменения ошибки сети. В качестве функции ошибки ИНС для одного наблюдения используется кросс-энтропия для дискретного случая:

$$E = - \sum_{i=1}^N c_i \cdot \ln y_i, \quad (2)$$

где  $c_i$  – ожидаемое значение на  $i$ -ом нейроне выходного слоя,  $y_i$  – полученное значение на  $i$ -ом нейроне,  $N$  – количество нейронов на выходном слое.

Представленная в данной работе архитектура ИНС, а также алгоритм обучения и его параметры были определены в ходе исследования различных конфигураций. Выбранная модель после обучения стабильно дает точность классификации не менее 90%.

## IV. МЕТОДЫ ОЦЕНКИ КАЧЕСТВА ИНС

### A. Значение функции ошибки сети

Первым критерием для оценки качества работы ИНС, является значение функции ошибки сети. Причем, исследуется динамика изменения ошибки, что позволяет определить, правильно ли обучилась модель в целом. График зависимости значения функции ошибки от количества эпох обучения для обучающей и тестовой выборки представлен на рис. 1. По данному графику можно выявлять такие проблемы как попадание в локальный минимум и переобучение.

Попадание в локальный минимум характеризуется тем, что значение функции ошибки на графике для обучающей выборки начинает колебаться при достаточно высоком значении этой функции.

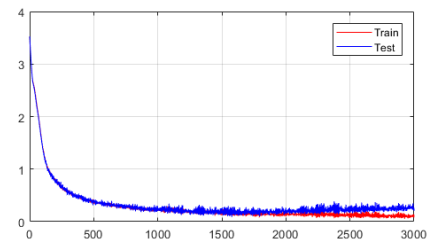


Рис. 1. График зависимости значения функции ошибки от количества эпох обучения. Красная линия – обучающая выборка, синяя – тестовая

Переобучение – это то состояние модели, когда она дает корректные результаты только для данных на которых обучалась. Возникновение переобучения обнаруживается на графиках как уменьшение значения функции ошибки для обучающей выборки и возрастание для тестовой. У переобученной ИНС нарушается способность к внутреннему обобщению.

Таким образом, анализ значения функции ошибки позволяет сразу выделять те модели, которые были обучены некорректно и которые нет смысла анализировать дальше. Но судить о качестве построенной модели только на основе значения функции ошибки нельзя, поэтому следующим этапом анализа является оценка точности классификации.

### В. Точность классификации

Существует несколько подходов к оценке точности классификации. Первый из них заключается в том, что определяется отношение правильно распознанных наблюдений к общему количеству наблюдений. Такой подход некорректно использовать в случае, когда в исходной выборке количество наблюдений для каждого класса разное. При такой оценке ошибка классификации возникнет только для класса, в котором малое количество наблюдений.

Более корректно рассчитывать точность классификации для каждого класса отдельно, как количество правильно классифицированных наблюдений класса к общему количеству наблюдений в данном классе, а затем считается средняя точность классификации по всем классам. В данном случае можно построить матрицу ошибок размерностью  $N*N$  ( $N$  – количество классов) и рассчитать точность, полноту и F-меру модели, либо представить результат в виде столбчатой диаграммы (рис. 2, а), на которой сразу будут видны классы с точностью классификации.

На основе анализа точности классификации можно оценить качество обучения ИНС, а также качество внутреннего и внешнего обобщения (рис. 2, б). В данном случае, оценка качества внешнего обобщения происходит на основе априорного знания о новых классах, которое может быть получено от эксперта или на основе расчета метрик в исходном пространстве признаков.

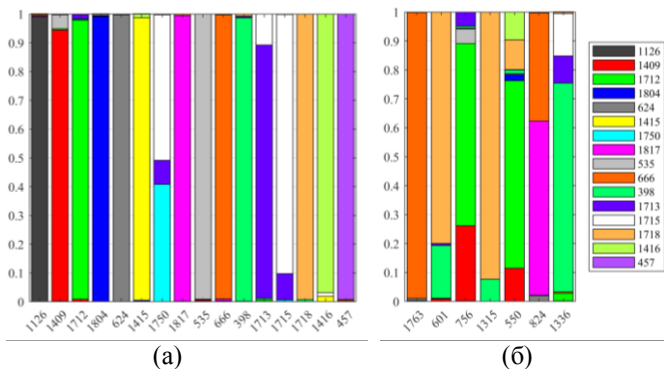


Рис. 2. Столбчатые диаграммы, отображающие усредненное распределение вероятности внутри каждого класса (а) – для обучающей выборки, (б) – для тестовых классов, не используемых при обучении модели. Разбиение на цвета показывает среднюю вероятность отнесения наблюдений к соответствующему целевому классу. Номера классов на оси абсцисс и в легенде соответствуют номерам штаммов в коллекции CALU

Недостатком анализа ИНС на основе оценки точности классификации является отсутствие критерия выбора лучшей модели из набора, где все модели имеют примерно одинаковую точность классификации.

### С. Энтропия выхода сети

В качестве одного из критериев выбора лучшей модели из набора моделей с одинаковой точностью классификации, можно выбрать энтропию выхода сети:

$$h = - \sum_{i=1}^N y_i \cdot \ln y_i, \quad (3)$$

где  $y_i$  – значение выхода сети на  $i$ -ом нейроне,  $N$  – количество нейронов на выходном слое.

В отличие от кросс-энтропии (2), которая оценивает схожесть двух распределений вероятности, а в случае ИНС близость ожидаемого и полученного значения выхода сети, обычная энтропия (3) позволяет оценивать степень «хаоса» распределения вероятностей внутри одного класса. Например, две разных модели для одного и того же наблюдения могут давать вероятности принадлежности к 3 классам  $[0.8, 0.19, 0.01]$  и  $[0.8, 0.1, 0.1]$ , то есть с одинаковой вероятностью относить наблюдения к первому классу и разной к оставшимся двум. Энтропия этих результатов равняется 0.54 и 0.63 соответственно. Энтропия выше для модели, которая дала более равномерные распределения. Таким образом, энтропия уменьшается, когда вероятность определения наблюдения к какому-либо классу стремится к 1, и принимает максимальное значение, когда вероятности отнесения к целевому классу одинаковые, то есть наблюдается равномерное распределение. Максимальное значение энтропии зависит от количества классов, так, например, для 3 и 16 классов при равных значениях вероятностей отнесения к целевым классам максимальная энтропия будет иметь значения 1.098 и 2.772 соответственно.

В случае ИНС, энтропия для каждого отдельного класса может интерпретироваться как показатель того, насколько «уверенно» модель определяет наблюдения данного класса. При большом количестве классов, для энтропии можно построить гистограмму (рис. 3), на которой сразу можно выделить классы, для которых ИНС может давать ложноположительный результат.

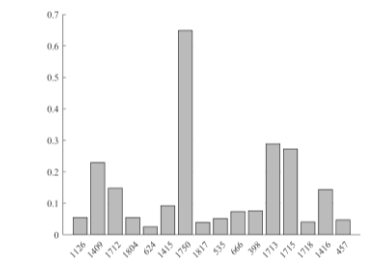


Рис. 3. Категориальная гистограммы для усредненной энтропии по классу. Каждый класс является отдельной категорией

При анализе энтропии классов, на которых производилось обучение ИНС, критерием качественной модели является стремление энтропии к нулю для всех классов в процессе обучения. При этом значения энтропии для разных классов должны быть близки. При оценке качества внешнего обобщения, и в случае, если наблюдения из новых классов должны однозначно определяться к известным и максимально похожим по характеристикам классам, необходимо минимизировать энтропию, в противном случае энтропия должна быть максимальна.

#### D. Анализ работы ИНС на основе сравнения с ЛДА

ЛДА является линейным статистическим классификатором, результаты работы которого строго детерминированы для конкретного набора данных. Поэтому ЛДА может быть использован в качестве эталона для верификации результатов работы ИНС.

При сравнении двух классификаторов (ЛДА и ИНС), в первую очередь необходимо сравнить точности классификации обеих моделей. Если для ИНС точность ниже, то, скорее всего, сеть была недостаточно обучена. Если же такой результат стабильно повторяется, то возможно была выбрана неверная архитектура ИНС, и ее следует изменить.

На следующем этапе сравниваются так называемые области пересечения. То есть те наблюдения, которые правильно и/или неправильно классифицировались в обеих моделях. Если, для разных моделей ИНС область неправильно классифицированных данных совпадает с такой же областью для ЛДА, то это указывает на некорректность исходных данных. Если для ИНС наблюдается повторяемость некорректной классификации наблюдений, которые правильно классифицируются моделью ЛДА, то это говорит об *алогичном поведении* конкретной ИНС.

Проводить сравнение результатов можно с любыми классификаторами, которые гарантированно строят одну и ту же модель на одних и тех же данных. В данной работе, ЛДА был выбран в ходе анализа ряда статистических классификаторов, как наиболее подходящий для исследуемой прикладной задачи [9].

#### V. РЕЗУЛЬТАТЫ

На основе описанных выше критериев оценки качества работы ИНС можно предложить следующую методику выявления алогичного поведения ИНС:

- Шаг 1 – оценка значения функции ошибки выхода ИНС, для выявления переобучения или попадания в локальный минимум
- Шаг 2 – оценка точности классификации и энтропии выхода сети для классов, на которых проводилось обучение, для определения качества внутреннего обобщения
- Шаг 3 – оценка точности классификации и энтропии выхода сети для классов, не присутствовавших в процессе обучения, для определения качества внешнего обобщения
- Шаг 4 – Верификация модели ИНС на основе сравнения с ЛДА

Для рассматриваемой прикладной задачи, была выбрана одна из построенных моделей ИНС, точность классификации которой равнялась 96.5%. На рис. 2 (а) видно, что для класса с номером 1750 наблюдается крайне низкая точность классификации, что подтверждается данными на рис. 3, так как для этого класса энтропия

максимальна и равна 0.64. Следовательно, именно в этом классе могут наблюдаться ложноположительные результаты. При анализе внешнего обобщения из рис. 2 (б) следует, что для класса с номером 1336 присутствует вероятность определения наблюдения к классам с номерами 1713 и 1715, что в данном случае является алогичным поведением, и, скорее всего, является следствием неправильного обучения ИНС. Данная модель ИНС показала более высокую точность, чем ЛДА, однако их области пересечения не совпали. Таким образом, построенная модель даже при достаточно высокой точности классификации является не совсем корректной и при ее работе в ряде случаев наблюдается алогичное поведение.

#### VI. ЗАКЛЮЧЕНИЕ

В данной работе, предложена методика, позволяющая оценить качество работы ИНС и выявить наличие случаев ее возможного алогичного поведения. Разработанная методика позволяет осуществлять выбор наилучшей сети из ряда обученных ИНС. Преимуществом данной методики является простой математический аппарат, который позволяет легко визуализировать результаты. Возможность наглядной визуализации результатов, позволит людям, незнакомым с логикой работы ИНС, которые получают обученную модель при помощи стандартных математических пакетов (например, Statistica 12), применять данную методику для проверки своих результатов.

Данная методика была апробирована при анализе одной из моделей, классифицирующих ИНС, решающей задачу дифференциации цианобактерий по спектрам собственной флуоресценции при экологическом мониторинге открытых водоемов.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Langford J. Tutorial on practical prediction theory for classification //Journal of machine learning research. 2005. Т. 6. №. Mar. С. 273-306.
- [2] Domingos P.M. A few useful things to know about machine learning //Commun. acm. 2012. Т. 55. №. 10. С. 78-87.
- [3] Осовский С. Нейронные сети для обработки информации. М.: Финансы и статистика, 2002. 344 с.
- [4] Altman E.I., Marco G., Varetto F. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience) //Journal of banking & finance. 1994. Т. 18. №. 3. С. 505-529.
- [5] Fisher R.A. The Use of Multiple Measurements in Taxonomic Problems // Annals of Eugenics. 1936. Vol. 7. С. 179-188.
- [6] Н.Ю. Григорьева, Л.В. Чистякова, А.А. Лисс, Д.М. Клионский, А.С. Перков, Т.Р. Жангиров Применение нейронных сетей для автоматизации экологического мониторинга цианобактериальных «цветений» водоемов // XXI Международная конференция по мягким вычислениям и измерениям (SCM-2018): Сборник докладов Том 1. Санкт-Петербург. 23-25 мая 2018 г. / СПб.: СПбГЭТУ «ЛЭТИ». С. 210-213.
- [7] Коллекция цианобактерий, водорослей и паразитов водорослей. URL: <http://researchpark.spbu.ru/collection-ccem-rus/poisk-pokollektsii-calu> (дата обращения 30.03.2019)
- [8] Хайкин С. Нейронные сети: полный курс, 2-е издание. Издательский дом Вильямс, 2008.
- [9] Жангиров Т.Р., Перков А.С., Лисс А.А., Григорьева Н.Ю., Чистякова Л.В. Применение линейного дискриминантного анализа для классификации цианобактерий по спектрам собственной флуоресценции // Известия СПбГЭТУ "ЛЭТИ". 2012. № 5. С. 45-55