

## Задание №6

### Цель задания:

-Характеристика и классификация сердечных сигналов с использованием нейронных сетей LSTM с прикладным механизмом внимания

### Ключевой навык:

-алгоритмы LSTM в нейронных сетях

### Введение

Данная задача предполагает изучение скрытых особенностей обработки сигналов ЭЭГ. Основная задача — определение локализации эпилептических приступов в записи ЭЭГ. Постройте график временной зависимости ЭЭГ во время приступов. Спектральная

## 任务 6

### 任务目标:

- 使用具有应用注意力机制的 LSTM 神经网络对心脏信号特征研究和分类

### 关键技能:

神经网络中的 LSTM 算法

### 介绍

这项任务涉及加载包含按心电图对疾病类别的描述的数据集，然后构建长短期记忆 (LSTM) 的神经网络并应用注意力机制，训练训练根据给定的 ECG 对类别进行分类，并建立一个混淆矩阵。对训练分类的准确度进行可视化。

Задача включает загрузку набора данных, содержащего описания категорий заболеваний по ЭКГ, затем построение нейронной сети с долговременной кратковременной памятью (LSTM) и применение механизма внимания, обучение ее классификации категорий по заданной ЭКГ и построение матрицы путаницы. Визуализируйте точность обученного классификатора.

## Ход работы

### 1. Подготовка данных

Функция `download_dataset()` использует `wfdb.dl_database` для загрузки набора данных MIT-BIH Arrhythmia. Наборы данных хранятся в локальном каталоге `mitdb_data`.

Пройдитесь по каждой записи в наборе данных, считывая сигнал и аннотацию.

Извлеките каждый аннотированный сегмент сердечного ритма (от 100 точек выборки до аннотированной точки до 100 точек выборки после аннотированной точки, всего 200 точек). Фильтрует определенные типы сердечных сокращений (N, L, R, A, V) и

## 工作进展

### 1. 数据准备

`download_dataset()` 函数使用 `wfdb.dl_database` 下载 MIT-BIH Arrhythmia 数据集。数据集存储在本地目录 `mitdb_data` 中。

遍历数据集中的每个记录，读取信号和标注。

提取每个标注的心跳片段（从标注点前 100 个采样点到后 100 个采样点，共 200 个点）。筛选特定类型的心跳（N、L、R、A、V），并为每种类型分配一个标签（0-4）。返回包含信号和标签的 NumPy 数组。

定义了一个自定义的注意力机制层 Attention:

使用两个全连接层（W1 和 W2）计算注意力分数。使用 softmax 将分数转换为注意力权重，通过

присваивает каждому типу метку (0-4). Верните массив NumPy, содержащий сигнал и метки.

加权求和得到上下文向量。

тестовый). Разложить данные сигнала в трехмерный тензор с формой (выборки, 200, 1).

Постройте и обучите модель (10 эпох с размером партии 32).

Сделайте прогнозы, используя тестовый набор, и рассчитайте матрицу ошибок.

Постройте и сохраните тепловую карту матрицы неточностей.

## 2. Визуализация данных

В этом эксперименте использовалась матрица путаницы, чтобы показать связь между прогнозами модели и истинными метками. Строки представляют истинные метки, столбцы представляют прогнозируемые метки, а значение каждой ячейки представляет собой количество образцов, прогнозы которых верны или неверны для соответствующей категории. Значения по диагонали указывают количество правильно классифицированных образцов. Чем ближе значение к диагонали, тем лучше характеристики модели.

## 2.数据可视化

在该实验中使用了混淆矩阵来展示了模型预测结果与真实标签之间的关系。行表示真实标签，列表示预测标签，每个单元格的值表示对应类别的预测正确或错误的样本数。对角线上的值表示分类正确的样本数，越靠近对角线的值越高，说明模型性能越好。非对角线的值表示分类错误的样本数，可以帮助我们发现模型在哪些类别上容易混淆。

Недиагональные значения представляют собой количество неправильно классифицированных образцов, что может помочь нам выяснить, в каких категориях модель легко запутается.

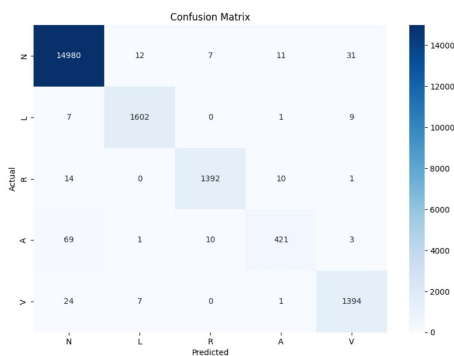


Рисунок 1—Тепловая карта матрицы путаницы для задач классификации

Как видно из рисунка, матрица неточностей включает пять категорий, обозначенных как N, L, R, A и V соответственно. Эти категории обычно соответствуют различным типам

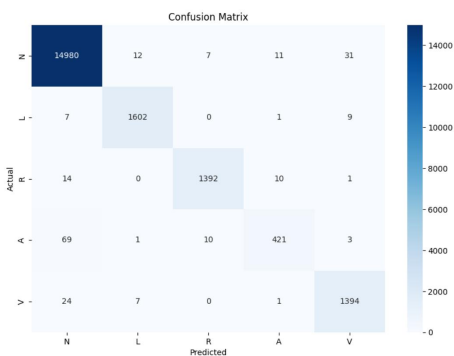


图 1-分类问题的混淆矩阵热力图

从图中可以看到，混淆矩阵涉及五个类别，分别标记为 N、L、R、A、V。这些类别通常对应心电图信号中的不同类型的心跳。从整体上看，模型对 N 类别的分类表现最佳，正确率非常高。对于其他类别（如 L、R、V），模型

сердечных сокращений в сигнале ЭКГ. В целом модель лучше всего подходит для классификации N категорий с очень высокой степенью точности. Модель также хорошо работает для других категорий (таких как L, R и V), но ее способность различать категорию A (предсердные экстрасистолы) относительно слаба, со множеством неправильных классификаций. Это может потребовать дальнейшего анализа распределения данных или производительности модели для оптимизации возможности идентификации класса A.

别（房性早搏）的区分度相对较弱，存在较多误分类现象。这可能需要进一步分析数据分布或模型性能，以优化对 A 类别的识别能力。



## Заключение

Этот эксперимент основан на наборе данных об аритмии MIT-BIH и использует LSTM в сочетании с механизмом внимания для построения модели глубокого обучения с целью классификации сигналов ЭКГ. Матрица ошибок показывает, что модель чрезвычайно точна в классификации нормальных сердечных сокращений (N), имея 14 980 правильных прогнозов и несколько ошибочных классификаций. Количество правильных прогнозов для трех категорий: блокады левой ножки пучка Гиса (L), блокады правой ножки пучка Гиса (R) и

## Вывод

Н а с т о я щ и й э к с п е р и м е н т б а з и р у е т с я н а M I T - B I H A r r h y t h m i a д а т а б а з у д а н н ы х , и с п о л ь з у е т L S T M в с о ч е т а н и и с м е х а н и з м о м в н и м а н и я д л я п о с т р о е н и я м о д е л и г л у б о к о г о о б у ч e н и я с ц е л ь ю к л а с с и ф и к а ц и и с и г н а л о в Э К Г . М а т р и ц а о ш и б о к п о к а з ы в а е т , ч т о м o d e л ь ч р е з ы ч а й н о т o ч н а в к л a c c и ф и к a ц и и н o p м a л ь н ы х с e p d e ч н ы х c o k p a ш e н и й ( N ) , и м e я 1 4 9 8 0 п p a в и л ь н ы х п p o г н o з o в и н e c k o л ь k o o ш и б o ч н ы х k л a c c и ф и k a ц и й . K o л и ч e c t в o п p a в и л ь н ы х п p o г н o з o в д л я t p e x k a t e r o p и й : б л o k a д ы л e в o й н o ж k и п y ч k a Г и c a ( L ) , б л o k a д ы п p a в o й н o ж k и п y ч k a Г и c a ( R ) и

Н а с т о я щ и й э к с п е р и м е н т б а з и р у е т с я н а M I T - B I H A r r h y t h m i a д а т а б а з у д а н н ы х , и с п o л ь з у е т L S T M в с o ч e т а н и и с м е х а н и з м о м в н и м а н и я д л я п o с т р o e н и я м o d e л и г л у б o k o г o o б u ч e n и я c ц e л ь ю k л a c c и f i k a ц и и c и г n a л o в Э К Г . М а т p и ц a o ш и б o k п o k a з ы в a e т , ч т o m o d e л ь ч p e з ы ч a й н o t o ч н a в k л a c c и f i k a ц и и н o p m a л ь н ы х c e p d e ч н ы х c o k p a ш e н и й ( N ) , и м e я 1 4 9 8 0 п p a в и л ь н ы х п p o г н o з o в и н e c k o л ь k o o ш и б o ч н ы х k л a c c и f i k a ц и й . K o л и ч e c t в o п p a в и л ь н ы х п p o г н o з o в д л я t p e x k a t e r o p и й : б л o k a д ы л e в o й н o ж k и п y ч k a Г и c a ( L ) , б л o k a д ы п p a в o й н o ж k и п y ч k a Г и c a ( R ) и

Б л о к а д ы л e в o й н o ж k и п y ч k a Г и c a ( L ) , б л o к а д ы п p a в o й н o ж k и п y ч k a Г и c a ( R ) , ш и м e p н ы е п e p e б o y ( V ) т p e x к л a c c и ф и k a ц и й . П p a в и л ь н ы е п p o г н o з ы д л я н o p m a л ь н ы х c e p d e ч н ы х c o k p a ш e н и й ( N ) c o c t a в л я ю т 1 4 9 8 0 , д л я б л o k a д ы л e в o й н o ж k и п y ч k a Г и c a ( L ) - 1 6 0 2 , д л я б л o k a д ы п p a в o й н o ж k и п y ч k a Г и c a ( R ) - 1 3 9 2 , д л я ш и м e p н ы х п e p e б o y ( V ) - 1 3 9 4 . Ч а c т ь п p o б л e м н ы х п p o г н o з o в c o c t a в л я ю т п p o г н o з ы н o p m a л ь н ы х c e p d e ч н ы х c o k p a ш e н и й ( N ) , к o т o p ы e б ы л и п p o г н o з и p o в a н ы к a k б л o k a д ы л e в o й н o ж k и п y ч k a Г и c a ( L ) и б л o k a д ы п p a в o й н o ж k и п y ч k a Г и c a ( R ) . O д н a к o , п p o г н o з ы ш и м e p н ы х п e p e б o y ( V ) б ы л и н e c k o л ь k o т o ч н ы м и . O б щ e г o в ы в o d : м o d e л ь л y б o п o л ь з o в a e т c я д л я k л a c c и ф и k a ц и и c и г n a л o в Э К Г , o д н a к o e e п p o g н o з и p o в a н и e д л я ш и м e p н ы х п e p e б o y ( V ) т p e x к л a c c и ф и k a ц и й .

преждевременного сокращения желудочков (V) также было на высоком уровне — 1602, 1392 и 1394 соответственно, при этом некоторые образцы в основном были неправильно классифицированы как нормальные сердечные сокращения или друг друга. Однако существует только 421 правильный прогноз предсердных экстрасистол (A), и большое количество образцов неправильно классифицируется как такие категории, как нормальное сердцебиение. В целом модель хорошо работает в большинстве категорий, но ее способность определять предсердные экстрасистолы (A) относительно слаба. В будущем для этой категории можно будет

выполнить улучшение данных и другие оптимизации, чтобы еще больше повысить производительность модели.