

СИСТЕМА РАСПОЗНАВАНИЯ ЖЕСТОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ**Ярыгина Екатерина Александровна**

студент, Самарский национальный исследовательский университет имени академика С. П. Королёва, РФ, г. Самара

Лезин Илья Александрович

доцент, канд. техн. наук, Самарский национальный исследовательский университет имени академика С. П. Королёва, РФ, г. Самара

Задачу распознавания жестов можно эффективно решать, анализируя данные, полученные электро-мио-графическим (ЭМГ) методом. Данная технология может интегрироваться в систему управления бионическим протезом, что позволяет осуществлять управление им естественным путем.

В данной работе приводится анализ ЭМГ данных, а также построена модель распознавания, решающая задачу распознавания жестов.

Для распознавания жестов в рамках данной работы будет использован браслет Муо, изображенный на рисунке 1. Входящий сигнал поступает в систему с частотой дискретизации 2 кГц с 8 каналов. Техническими характеристиками устройства являются: точность АЦП – 24 бита, ширина временного окна – 250 мс, протокол обмена данными – USB UART.



Рисунок 1. Браслет Муо

Целью исследования является построение алгоритма многоклассовой классификации, использующего в качестве признаков временных рядов параметры моделей временных рядов и их распределения, обобщение методов классификации временных рядов, использующих явное признаковое описание, повышение качества решения задач классификации временных рядов.

Таким образом, для реализации системы выделим следующие этапы:

1. формирование методики сбора и первичная обработка данных;
2. формирование алгоритма сегментации данных;
3. решение задачи классификации жестов.

При решении задач классификации необходимо отнести имеющиеся статические образцы к определенным классам. Данный тип задачи можно разбить на две основные части: сегментация и распознавание жестов.

Поставим задачу многоклассовой классификации временных рядов в общем виде.

$$X \rightarrow \{x_i = x_i^{(1)}, \dots, x_i^{(t)}\} \quad Y$$

Дано: X – временные ряды, Y – множество меток

$$D \subset X \times Y$$

D – конечная обучающая выборка.

$$a = b \circ f \circ S$$

Модель классификации: a , где S

$$S: x \rightarrow \{s^{(1)}, \dots, s^{(1)}\}$$

S – алгоритм сегментации: $s^{(1)}$, где

$$s^{(k)} = [x^{(t_k)}, \dots, x^{(t_{k-1})}]$$

f

f – признаковое описание набора сегментов,

b

b – алгоритм многоклассовой классификации.

Сегменты временных рядов описываются моделями вида

$$g(w, x) \in X, \text{ где } w \in R^d.$$

Существует множество подходов к классификации образов. Наиболее эффективными для задачи распознавания жестов является применение методов глубокого машинного обучения (сверточные и рекуррентные нейронные сети).

Так как производительность и объем памяти встраиваемого устройства ограничен, следует рассмотреть так же метод опорных векторов (SVM).

В результате разработки системы сегментации и распознавания жестов были произведены исследования и построены несколько моделей. Ниже приведены описания моделей и качество их распознавания на тестовой и валидационной выборке.

Для нахождения параметров модели производилась операция обучения. Для проведения обучения модели необходимо было набрать набор данных и разметить их. Далее данные разбиваются на 3 группы: тестовую, тренировочную и валидационную.

На первой группе параметров производится обучения модели –то есть подстройки коэффициентов с целью минимизации квадрата ошибки распознавания.

В процессе обучения анализируется поведение модели на тестовой выборке, это необходимо чтобы избежать так называемого переобучения модели.

Переобучение (overfitting) – это процесс, при котором модель перестает улавливать зависимости между параметрами и начинает просто заучивать тестовый набор данных.

В начальный момент времени точность распознавания на тестовой и обучающей выборке увеличивается, но в какой-то момент точность распознавания на тестовой начнет падать, в то время как точность распознавания на обучающей выборке продолжает расти. В этот момент времени рекомендуется запомнить коэффициенты или прекратить обучение.

После завершения обучения полученная модель проверяется на валидационной выборке. Если результат на валидационной выборке удовлетворительный, то модель отправляется в устройство.

Основными параметрами характеризующие качество распознавания являются:

- Precision – точность, определяется как отношение правильных ответов к общему числу проверок;

- Recall – характеризуется процентом правильных ответов из выборки содержащий только этот класс;

$$\underline{2 * Precision * Recall}$$

$$Precision + Recall$$

- F1-score – определяется отношением $\frac{2 * Precision * Recall}{Precision + Recall}$. То есть F1-score представляет из себя среднее гармоническое между Precision и Recall

Наиболее удачной из моделей в которой особенности выделялись в ручную, является модель построенная на особенностях из кросс-корреляций сигнала (8 каналов, $C^2_{\{8\}} = 28$ параметров + спектральная мощность на 3 диапазонах, от 0-30, 40-80, 100-400 Гц) + SVM. О возможности произвести классификацию жестов с использованием данных жестов можно судить по следующим изображениям классов жестов в пространстве главных компонент, представленных на рисунке 2.

Как видно из изображений классы жестов разделимы в пространстве первых 3 главных компонент сигнала. Данный результат позволяет с уверенностью применить SVM для решения данной задачи. Результат обучения SVM приведен ниже в таблице 1.

Таким образом среднее точность распознавания с использование данной модели является 83%, что довольно высоко с учетом ее простоты.

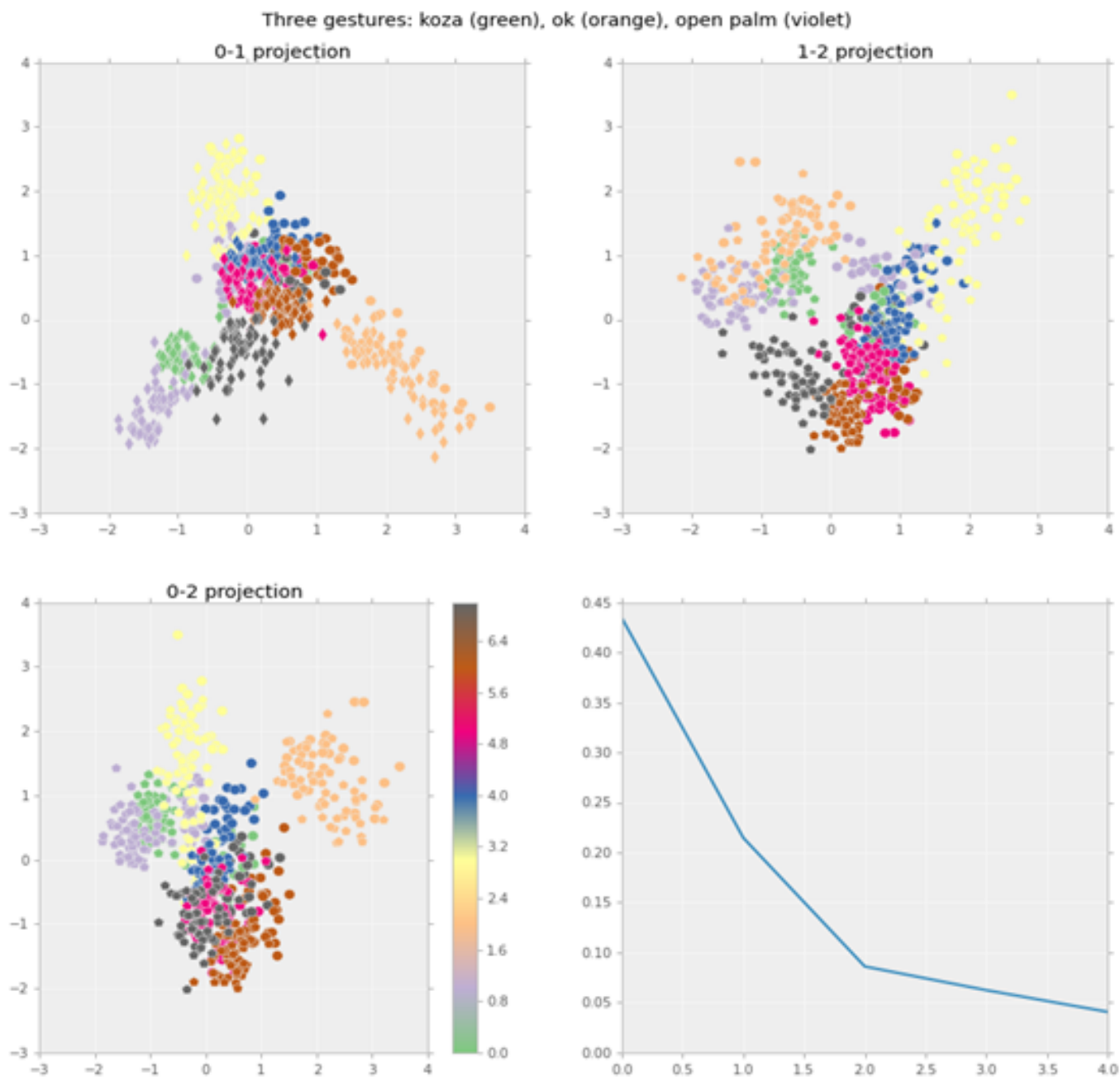


Рисунок 2. Классы жестов в пространстве главных компонент

Таблица 1.

Результаты распознавания жестов методом SVM

	Precision	Recall	F1-score	Support
0	0.60	0.60	0.60	10
1	1.00	1.00	1.00	9
2	0.85	0.92	0.88	12
3	1.00	0.44	0.62	9
4	0.69	0.82	0.75	11
5	0.73	1.00	0.85	11

6	1.00	0.90	0.95	10
7	0.90	0.82	0.86	11
Avg / Total	0.84	0.82	0.81	83

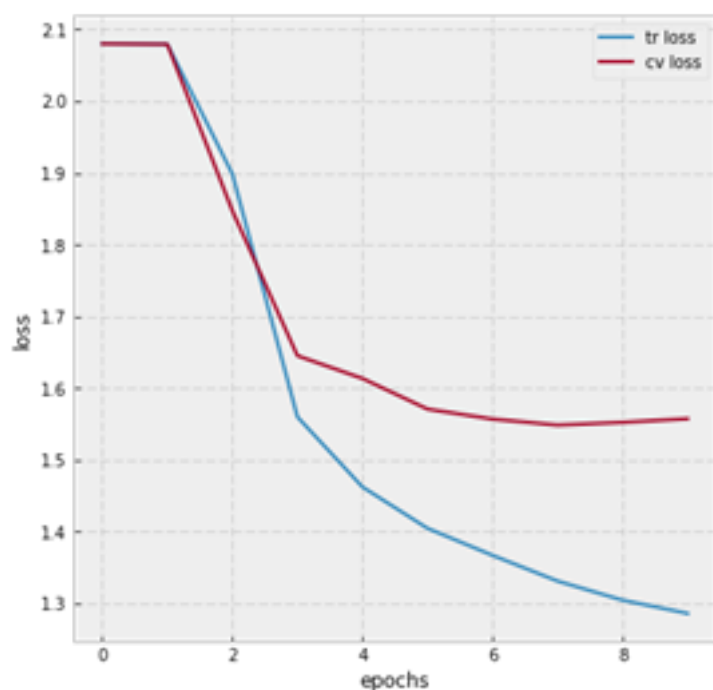
Основной проблемой данной модели является ее неустойчивость к поворотам. То есть, когда браслет одевается на испытуемого повторно или с течением эксперимента браслет немного сдвигается, происходит существенный сдвиг параметров и точность распознавания падает практически до 0. При выполнении работы были произведены попытки разработки системы коррекции поворотов. Данная система должна была производить калибровку модели для того, чтобы устранить эффект сдвига. Для этого требуется произвести набор калибровочных жестов.

К сожалению, число калибровочных жестов, которое необходимо совершить для того, чтобы вернуть точность распознавания на прежний уровень достаточно велико и фактически совпадает с числом жестов необходимых для тренировки необученной модели. В связи с этим от использования данной модели пришлось отказаться.

Следующим этапом в разработке модели стало использование сверточных нейронных сетей. В данной работе использовались 2 основные архитектуры.

Первая архитектура представлена сверточными нейронными сетями, проходящими непосредственно по сигналу. Во втором случае сигнал подвергался преобразованию Фурье и в качестве входного сигнала в нейронную сеть поступали коэффициенты Фурье по всем 8 каналам. Также отличались и форма ядра свертки входного сигнала от прямой до квадрата.

На рисунках 3а и 3б представлены значения функции потерь на тестовой и обучающей выборках в зависимости от эпохи обучения. В качестве функции потерь выбрана кросс энтропия.



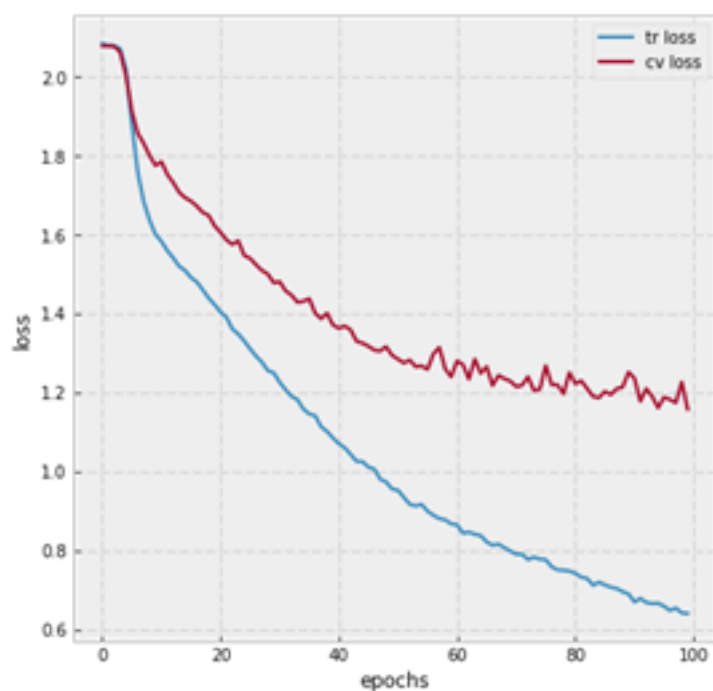


Рисунок 3. Значения функции потерь на тестовой и обучающей выборках

а) одномерная сеть; б) двумерная сеть

В таблицах 2а и 2б представлен результат обучения одномерной и двумерной сверточной нейронной сети.

Таблица 2а.

Результаты распознавания жестов одномерной сверточной нейронной сетью

	Precision	Recall	F1-score	Support
0	0.19	0.13	0.15	101
1	0.35	0.50	0.42	131
2	0.50	0.73	0.59	100
3	0.39	0.11	0.17	103
4	0.26	0.44	0.33	109
5	0.49	0.18	0.26	101
6	0.30	0.31	0.30	101
7	0.40	0.37	0.38	102
Avg / Total	0.36	0.35	0.33	848

Таблица 2б.

Результаты распознавания жестов двумерной сверточной нейронной сетью

--	--	--	--	--

	Precision	Recall	F1-score	Support
0	0.43	0.41	0.42	101
1	0.81	0.89	0.85	131
2	0.96	0.76	0.85	100
3	0.52	0.16	0.24	103
4	0.44	0.80	0.57	109
5	0.57	0.23	0.33	101
6	0.75	0.75	0.75	101
7	0.57	0.90	0.70	102
Avg / Total	0.64	0.62	0.60	848

Как видно из таблицы точность распознавания весьма удовлетворительна для жестов 1,2,6. Хотя коэффициент точности удачной модели ниже, чем у SVM, она тем не менее является устойчивой к поворотам и не предназначена для одного конкретного человека, то есть является фактически рабочей моделью.

Данная модель строилась на выборке, полученной с 20 человек. Каждый человек выполнял каждый жест по 200 раз с перерывом каждые 50 жестов. Данная модель показывает намного лучшие результаты если жесты распознаются на одном и том же человеке и позволяет осуществлять распознавание с точностью до 92%.