

ИССЛЕДОВАНИЕ СВОЙСТВ  
ПРИЗНАКОВОГО ПРОСТРАНСТВА  
ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ  
ОБУЧЕНИЯ ПО НЕСКОЛЬКИМ  
ПРИМЕРАМ

В.Д. Кучеров, Д.Ю. Буряк

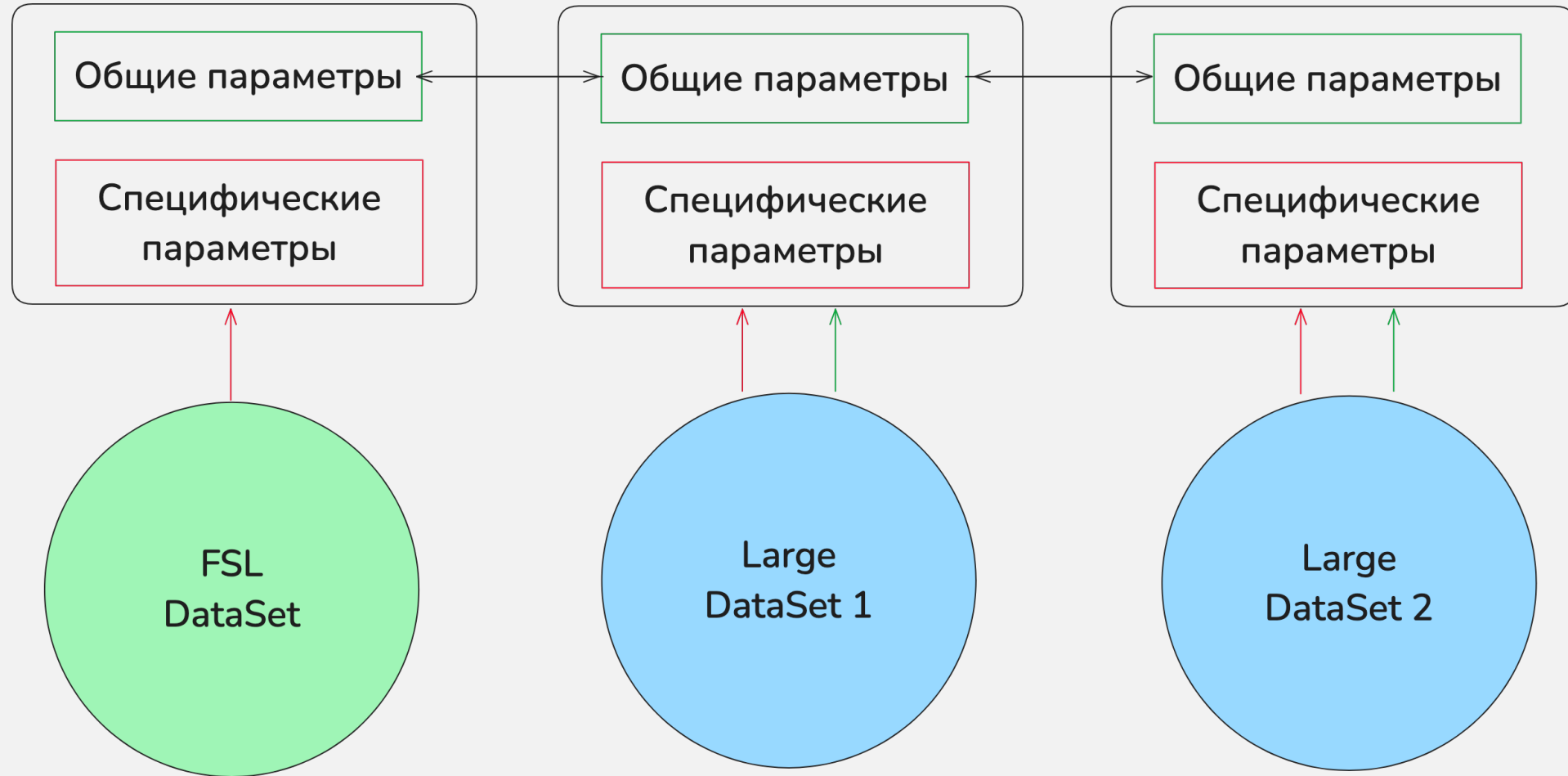
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ имени М.В. Ломоносова,  
Факультет вычислительной математики и кибернетики

# АКТУАЛЬНОСТЬ РАБОТЫ

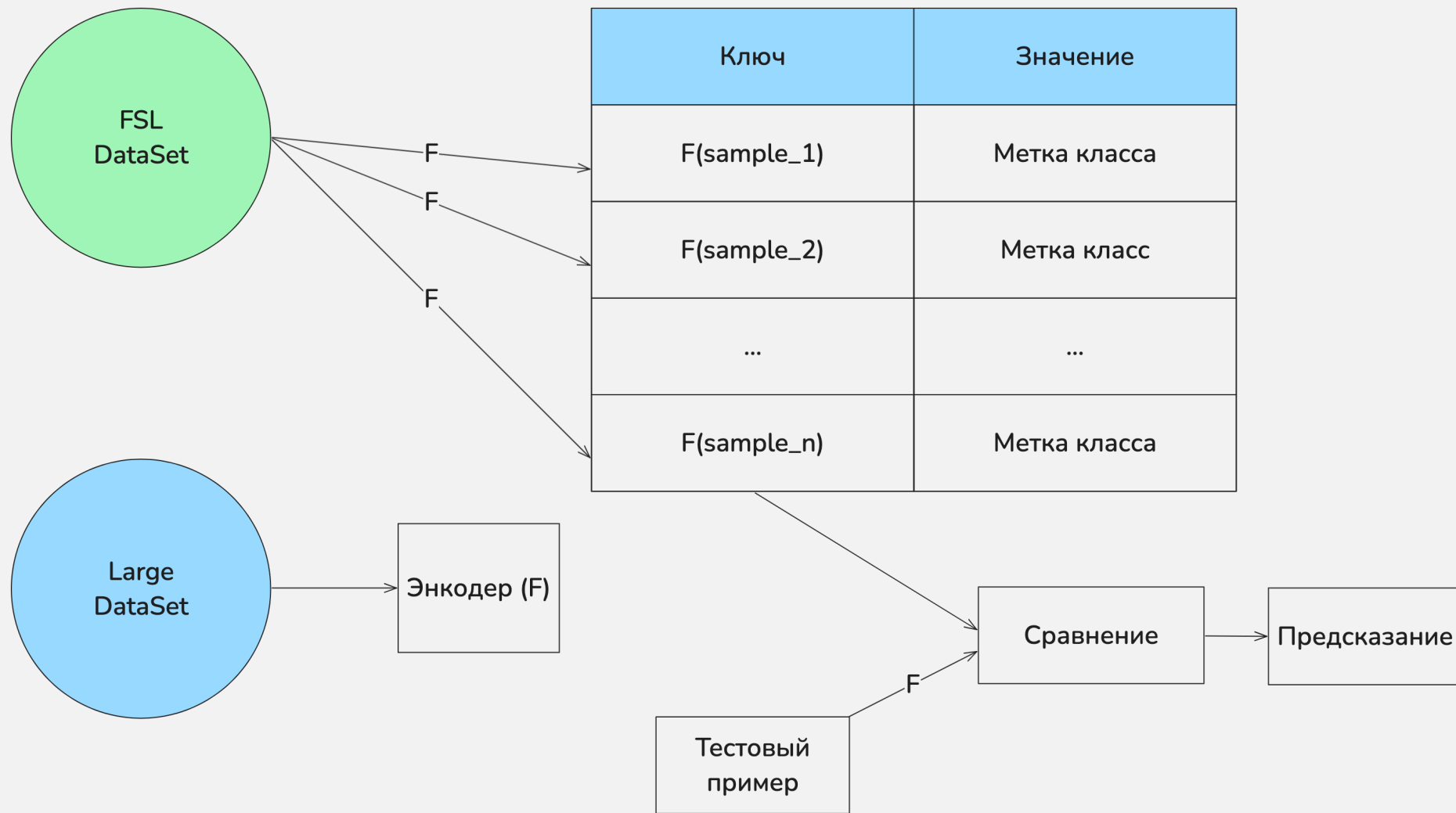
Задачи с малым набором тренировочных данных:

- Ограничены пользовательским вводом:
  - Верификации по биометрии
  - Распознавание ключевых слов (Key Word Spotting) – добавление персонализированных сценариев и команд в системах умного дома и умных устройствах
- Ограничения по причинам конфиденциальности (обычно связаны с задачами в области медицины)
- Редкие классы и события

# ОБЗОР РЕШЕНИЙ – MULTITASK LEARNING



# ОБЗОР РЕШЕНИЙ – ОБУЧЕНИЕ С ВНЕШНЕЙ ПАМЯТЬЮ



# ЭМБЕДДИНГИ

Цель:

- Уменьшение размерности задачи
- Построение пространства, в которых эмбединги кластеризуются

Тогда решение целевой задачи обучения по нескольким примерам классификации сводится к построению пространства с высокими качествами кластеризации

## ЦЕЛЬ РАБОТЫ

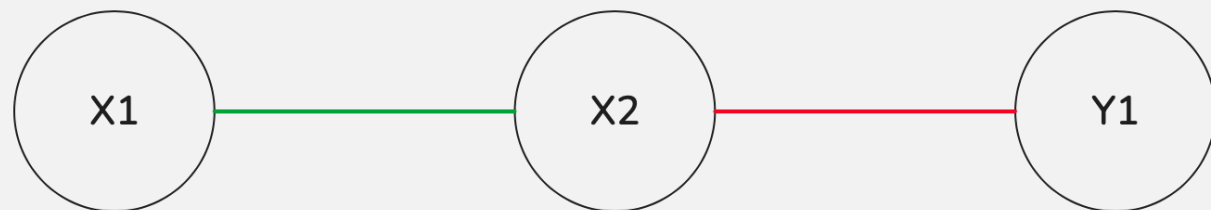
Исследование свойств признакового пространства и их связи с качеством решения задачи FSL:

- Как оценить предобученный энкодер с точки зрения последующего решения задачи FSL?
- Какие метрик наилучшим образом коррелируют с точностью решения FSL?

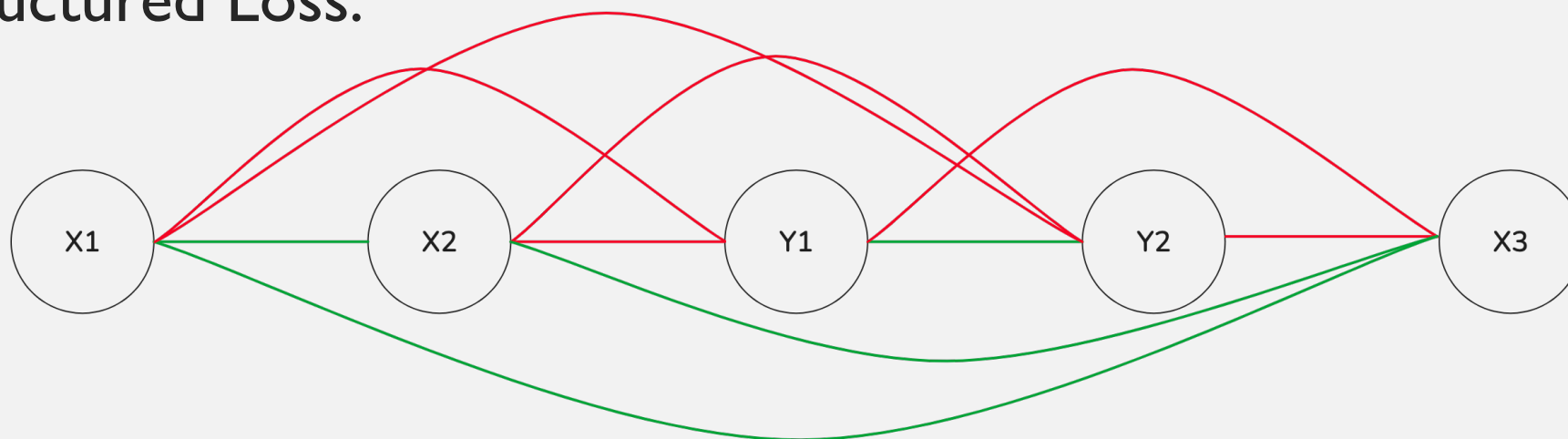
# ЭМБЕДДИНГИ, ФУНКЦИИ ПОТЕРЬ

Функции потерь семейства **Contrastive Loss**:

$$\mathcal{L}_{\text{triplet}}(\mathbf{x}, \mathbf{x}^+, \mathbf{x}^-) = \sum_{\mathbf{x} \in \mathcal{X}} \max \left( 0, \left\| f(\mathbf{x}) - f(\mathbf{x}^+) \right\|_2^2 - \left\| f(\mathbf{x}) - f(\mathbf{x}^-) \right\|_2^2 + \epsilon \right)$$

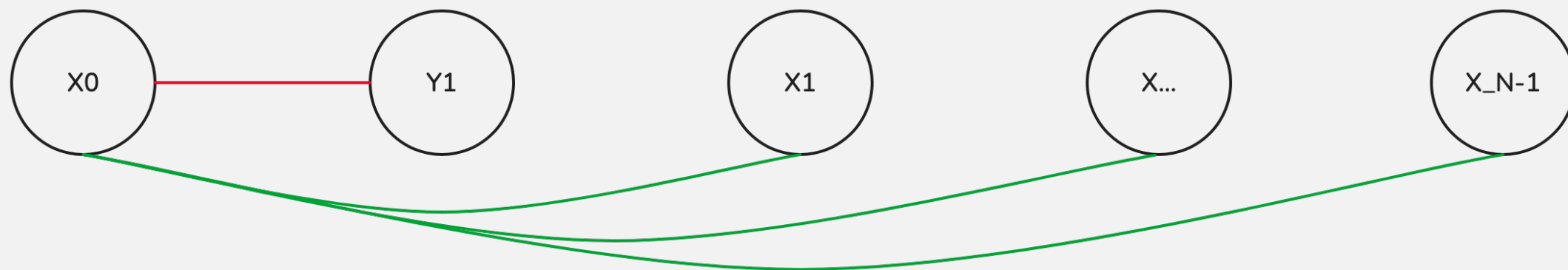


**Lifted Structured Loss:**



# ЭМБЕДДИНГИ, ФУНКЦИИ ПОТЕРЬ

$$\begin{aligned}\mathcal{L}_{N\text{-pair}}(x, x^+, \{x_i^-\}_{i=1}^{N-1}) &= \log\left(1 + \sum_{i=1}^{N-1} \exp(f(x)^\top f(x_i^-) - f(x)^\top f(x^+))\right) = \\ &= -\log \frac{\exp(f(x)^\top f(x^+))}{\exp(f(x)^\top f(x^+)) + \sum_{i=1}^{N-1} \exp(f(x)^\top f(x_i^-))}\end{aligned}$$





# МЕТРИКИ КЛАСТЕРИЗАЦИИ

**Feature clustering (FC):** отношение среднего внутриклассового расстояния к среднему межклассовому расстоянию

$$\frac{\sigma_{\text{within}}^2}{\sigma_{\text{between}}^2} = \frac{\sum_{i,j} \|\phi_{i,j} - \mu_i\|_2^2}{\frac{\sum_i \|\mu_i - \mu\|_2^2}{C}} = \frac{C}{N} \frac{\sum_{i,j} \|\phi_{i,j} - \mu_i\|_2^2}{\sum_i \|\mu_i - \mu\|_2^2}$$

$\mu_i$  - средний эмбединг  $i$ -го класса

$N$  - количество примеров в классе

$\phi_{i,j}$  - эмбединг  $j$ -ого экземпляра  $i$ -го класса

$C$  - количество классов

**Hyperplane Variance (HV):** "Насколько разные разделяющие гиперплоскости можно построить"

$x_1, x_2$  - примеры одного класса

$y_1, y_2$  - примеры другого класса

$$R_{HV}(f_\theta(x_1), f_\theta(x_2), f_\theta(y_1), f_\theta(y_2)) = \frac{\|(f_\theta(x_1) - f_\theta(y_1)) - (f_\theta(x_2) - f_\theta(y_2))\|_2}{\|f_\theta(x_1) - f_\theta(y_1)\|_2 + \|f_\theta(x_2) - f_\theta(y_2)\|_2}$$

# МЕТРИКИ КЛАСТЕРИЗАЦИИ

**Silhouette score:** среднее значение silhouette coefficient для всех точек

**silhouette coefficient (s):**

- $b$  – Среднее расстояние от точки до всех точек ближайшего кластера
- $a$  – Среднее расстояние от точки до всех точек того же класса

$$s = \frac{b - a}{\max(a, b)}$$

**Davies-Bouldin index:** Средняя схожесть классов, где схожесть –  $R_{ij}$

- $s_i$  – среднее расстояние точек  $i$ -го класса до центра этого класса
- $s_j$  – среднее расстояние точек  $j$ -го класса до центра этого класса
- $d_{ij}$  – расстояние между центрами классов  $i$  и  $j$

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij} \quad R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

## SILHOUETTE MARGIN LOSS

$$SML(\{x_i\}_{i=1}^N) = \frac{1}{N} \sum_{i=1}^N \max(0, a_i - b_i + \epsilon)$$

$a_i$  — Расстояние от  $x_i$  до центроида своего класса.

$b_i$  — Расстояние от  $x_i$  до центроида ближайшего класса.

# ФОРМАЛЬНАЯ ПОСТАНОВКА ЗАДАЧИ

## Дано:

Вспомогательный набор данных  $D_{HELP} = \{D_{HELP}^{train}, D_{HELP}^{test}\}$ , пусть  $L_{HELP}$  - набор классов в  $D_{HELP}$ ,  $M_{HELP}^{train}$  - количество примеров на класс в наборе  $D_{HELP}^{train}$ .

Целевой набор данных  $D_{FSL} = \{D_{FSL}^{train}, D_{FSL}^{test}\}$ , аналогично  $L_{FSL}$  - набор классов в  $D_{FSL}$ ,  $M_{FSL}^{train}$  - количество примеров на класс в наборе  $D_{FSL}^{train}$ . При этом  $M_{HELP}^{train} \gg M_{FSL}^{train}$ ,  $|L_{HELP}| \gg |L_{FSL}|$ ,  $L_{HELP} \cap L_{FSL} = \emptyset$ .

## Требуется:

Построить энкодер  $E: F \rightarrow F_E$ , обучая только на  $D_{HELP}^{train}$ .

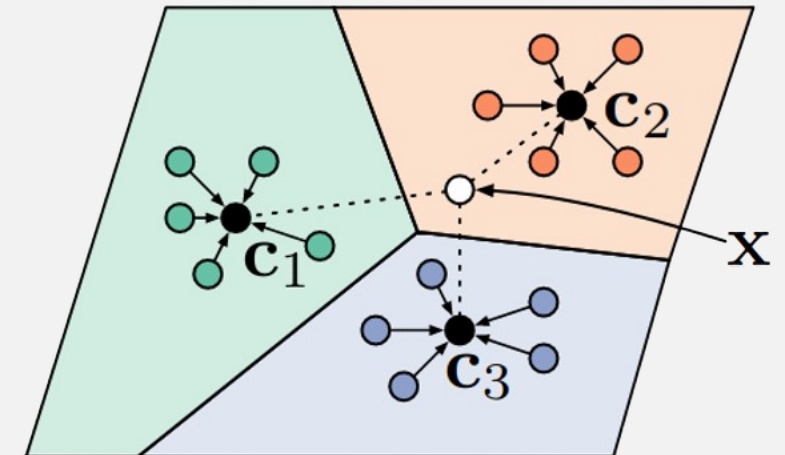
Построить классификатор  $C: F_E^{FSL} \rightarrow L_{FSL}$ , обучая на  $D_{FSL}^{train}$ ;  $F_E^{FSL} \subset F_E$ .

Сравнить метрики оценки пространства  $F_E$  по отношению к точности классификатора  $C$ .

# ПОСТАНОВКА ЭКСПЕРИМЕНТА

- Выбрана задача распознавания ключевых слов.
- Фиксированная архитектура модели энкодера – сверточная нейронная сеть: 6 сверточных слоев, размер выходного вектора – 64, общее число параметров – 32 404.
- На вход энкодера подается мел-спектрограмма.
- Функции потерь: Cross-entropy, Triplet, Lifted Structured, N-pair.
- Использование Silhouette Score в качестве функции потерь - **Silhouette Margin Loss**.
- Методы оценки качества пространства – FC, HV, Silhouette score, Davies-Bouldin Index.
- В качестве классификатора выбрана прототипическая сеть.
- Оценка качества классификации –  $Accuracy = \frac{N_T}{N}$ , где  $N$  – количество классифицируемых объектов,  $N_T$  - количество правильно классифицированных объектов.

Схема работы прототипической сети



# ПОСТАНОВКА ЭКСПЕРИМЕНТА

- Задача распознавания ключевых слов.
  - аудио 1с.
  - Google Speech Commands (35 слов-классов и 105,829 аудиофайлов)
  - MLCommons MSWC
- Обучение энкодера производится на следующих выборках :
  1. Speech Commands (EN) 25 классов 1250 примеров на класс.
  2. MSWC EN 100 классов 500 примеров на класс.
  3. MSWC EN 500 классов 100 примеров на класс.
  4. MSWC RU 50 классов 500 примеров на класс.

Использовались слова различной длины: от 3 до 12 букв, для которых есть достаточное число примеров.
- Выбиралось несколько вариантов обученного энкодера – лучших по каждой из метрик.
- FSL: 10 новых классов, 1-shot и 5-shot, MSWC RU, MSWC EN и Speech Commands (EN) – всего 6 различных экспериментов. 10 запусков для каждого эксперимента.

# РЕЗУЛЬТАТЫ

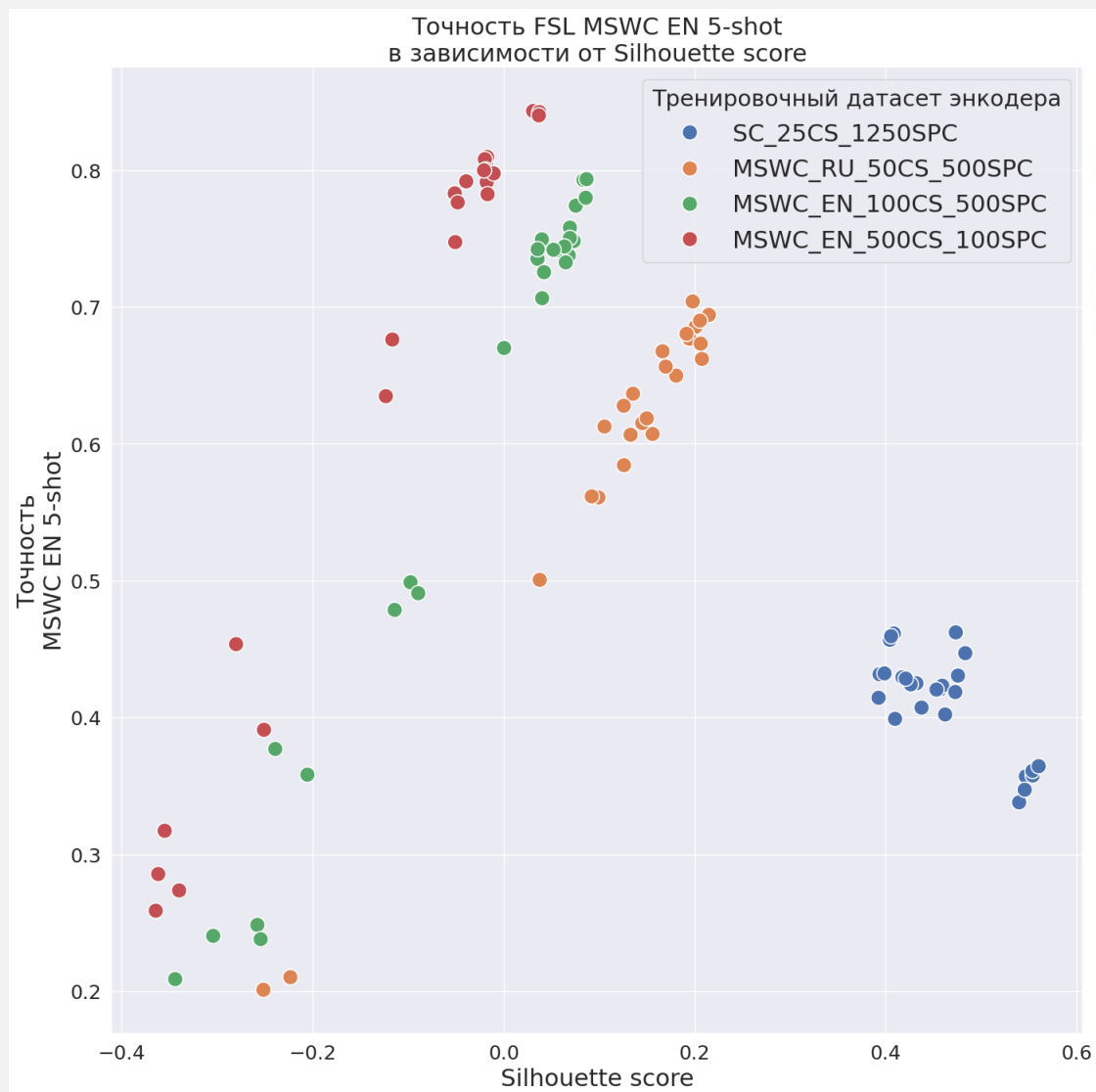


Таблица 1: Среднее значение корреляции Пирсона между значением метрики и точностью классификатора. Эксперименты разделены по типам по столбцам: 'все' - усреднение по всем экспериментам, 'кросс-языковые' - энкодер и классификатор обучены на разных языках, 'ru-en' - энкодер обучен на русском, классификатор на английском и так далее.

1-shot							
Метрика	все	кросс языковые	один язык	ru-en	en-ru	ru-ru	en-en
FC	0,429	0,432	0,427	0,175	<b>0,599</b>	0,062	0,486
HV	0,004	-0,061	0,051	-0,560	0,263	-0,646	0,164
<b>Silhouette score</b>	<b>0,670</b>	<b>0,647</b>	<b>0,687</b>	<b>0,972</b>	0,436	0,954	<b>0,643</b>
Davies-Bouldin	-0,624	-0,608	-0,635	-0,946	-0,388	<b>-0,970</b>	-0,581
5-shot							
Метрика	все	кросс языковые	один язык	ru-en	en-ru	ru-ru	en-en
FC	0,536	0,467	0,586	0,191	<b>0,646</b>	0,128	<b>0,660</b>
HV	0,098	-0,036	0,193	-0,541	0,292	-0,604	0,322
<b>Silhouette score</b>	<b>0,628</b>	<b>0,632</b>	<b>0,625</b>	<b>0,980</b>	0,406	<b>0,980</b>	0,568
Davies-Bouldin	-0,577	-0,594	-0,564	-0,951	-0,363	-0,976	-0,497

# РЕЗУЛЬТАТЫ

Таблица 2: Среднее значение точности классификатора (в процентах) в зависимости от выбранного метрикой энкодера (эпохи). Эксперименты разделены по типам по столбцам: 'все' - усреднение по всем экспериментам, 'кросс-языковые' - энкодер и классификатор обучены на разных языках, 'ru-en' - энкодер обучен на русском, классификатор на английском и так далее.

1-shot							
Метрика	все	кросс языковые	один язык	ru-en	en-ru	ru-ru	en-en
FC	35,1	33,4	36,3	42,0	27,6	<b>62,4</b>	32,0
HV	41,7	40,8	42,4	38,5	42,3	62,0	39,1
<b>Silhouette score</b>	<b>57,7</b>	<b>54,5</b>	<b>60,0</b>	45,1	<b>60,8</b>	61,7	<b>59,8</b>
Davies-Bouldin	57,3	54,1	59,5	<b>46,1</b>	59,4	61,2	59,3
5-shot							
Метрика	все	кросс языковые	один язык	ru-en	en-ru	ru-ru	en-en
FC	44,2	45,5	43,4	59,5	36,1	74,6	38,2
HV	53,7	55,5	52,5	53,9	56,5	74,3	48,8
<b>Silhouette score</b>	<b>74,0</b>	<b>71,1</b>	<b>76,1</b>	<b>65,1</b>	<b>75,1</b>	<b>79,1</b>	<b>75,6</b>
Davies-Bouldin	73,5	70,1	75,9	64,5	73,8	78,3	75,6



# РЕЗУЛЬТАТЫ

Таблица 3: Среднее значение точности классификатора (в процентах) в зависимости от функции потерь энкодера. Эксперименты разделены по типам по столбцам: 'все' - усреднение по всем экспериментам, 'кросс-языковые' - энкодер и классификатор обучены на разных языках, 'ru-en' - энкодер обучен на русском, классификатор на английском и так далее.

1-shot							
Функция потерь	все	кросс языковые	один язык	ru-en	en-ru	ru-ru	en-en
Cross Entropy	55,6	52,1	58,1	40,9	59,5	61,4	57,5
Lifted Structured	56,6	53,6	58,8	43,8	60,2	62,4	58,2
N-Pair	56,0	53,3	57,9	45,1	58,7	61,7	57,3
<b>Silhouette Margin</b>	<b>59,7</b>	<b>56,7</b>	<b>61,8</b>	<b>47,6</b>	<b>62,9</b>	<b>65,8</b>	<b>61,1</b>
Triplet	55,5	51,8	58,0	43,4	57,4	62,4	57,3
5-shot							
Функция потерь	все	кросс языковые	один язык	ru-en	en-ru	ru-ru	en-en
Cross Entropy	71,8	68,3	74,3	58,3	74,9	75,8	74,1
Lifted Structured	72,2	70,0	73,9	61,2	75,8	77,9	73,2
N-Pair	72,0	70,6	73,0	65,4	74,1	79,3	71,9
<b>Silhouette Margin</b>	<b>75,9</b>	<b>72,7</b>	<b>78,1</b>	<b>65,7</b>	<b>77,4</b>	<b>80,7</b>	<b>77,7</b>
Triplet	71,4	69,1	73,1	60,9	74,6	76,0	72,6

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ

- Из метрик кластеризации наибольшую корреляцию между значением и точностью решения FSL показывают **Silhouette Score** и **Davies-Bouldin Index** со значениями корреляции Пирсона соответственно **0.655** и **0.608**
- Была предложена новая функция потерь **Silhouette Margin Loss**, которая основывается на метрике Silhouette Score. Проведенные эксперименты показывают, что **Silhouette Margin Loss** превосходит рассматриваемые функции потерь в среднем более чем на 3.5%.