

Коды, контексты, смысл

Дискретный подход к машинному обучению

Дмитрий Кашицын

AGI Russia

4 сентября 2025

Мета

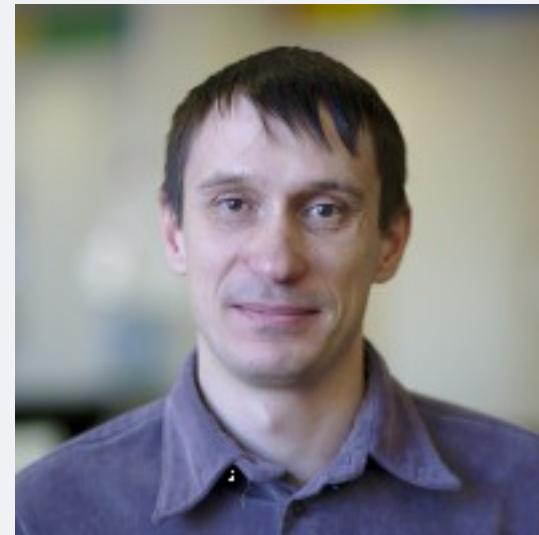
Обо мне

- Системный программист (~30 лет)
 - C/C++, Rust, embedded
 - Цифровое ТВ (кодеки, протоколы, STB middleware)
 - Картография и навигация (упаковка, отрисовка, маршруты),
 - Распределенные системы (консенсус, сеть)
 - Компиляторы (системы типов, вывод, JIT, VM)
- Исследователь (~10 лет)
 - Дискретное восприятие, семантика, память
 - Лингвистика естественных и синтетических языков
 - Нейрофизиология, иммунология

Наша команда



Дмитрий Кашицын



Дмитрий Шабанов

Цели и вопросы

- Что такое смысл?
- Что такое мысль?
- Где грань между структурой и семантикой?
- Какова роль памяти в мышлении?
- Как происходит процесс мыслеобразования?
- Какова природа и роль эмоций?

- Что из вышеперечисленного можно смоделировать и проверить на практике?

Наша статья на arXiv

Discrete approach
to machine learning



<https://arxiv.org/pdf/2508.00869>

<https://daiylabs.com/papers/daml/>

Теория

Хронические болезни и проблемы нейросетей

- Фундаментальное ограничение на размер окна
- Работа с исключениями и проблема разрешения (Ш и Щ)
- Интерпретируемость
- One/few shot learning
- Overfitting
- Проблемы катастрофического забывания и 
- Невозможность развидеть
- Трудности дообучения
- Смена топологии и архитектуры

В основе метода – работа
с дискретными элементами

Наши козыри

- Интерпретация воспоминаний модели на всех уровнях
- Точечное редактирование и удаление/блокировка воспоминаний
- Объединение нескольких обученных моделей в одну
- Изменение топологии модели без потери прогресса обучения
- Непрерывное обучение
- Неограниченный размер контекста (активная память)
- Прочее (описано в статье в главе 8)

Наши киты: лемма Йонеды и гипотеза многообразия

Суть процесса обучения

1. Порождаем коды **проекций** первичных **стимулов**
2. Располагаем их на кодовом пространстве
3. Понижаем размерность кодового пространства с помощью нашего алгоритма **DAMP**
4. Кластеризуем пространство с помощью **DBSCAN**
5. Строим иерархию **детекторов** поверх кластеров
6. Назначаем каждому детектору **случайный** бит

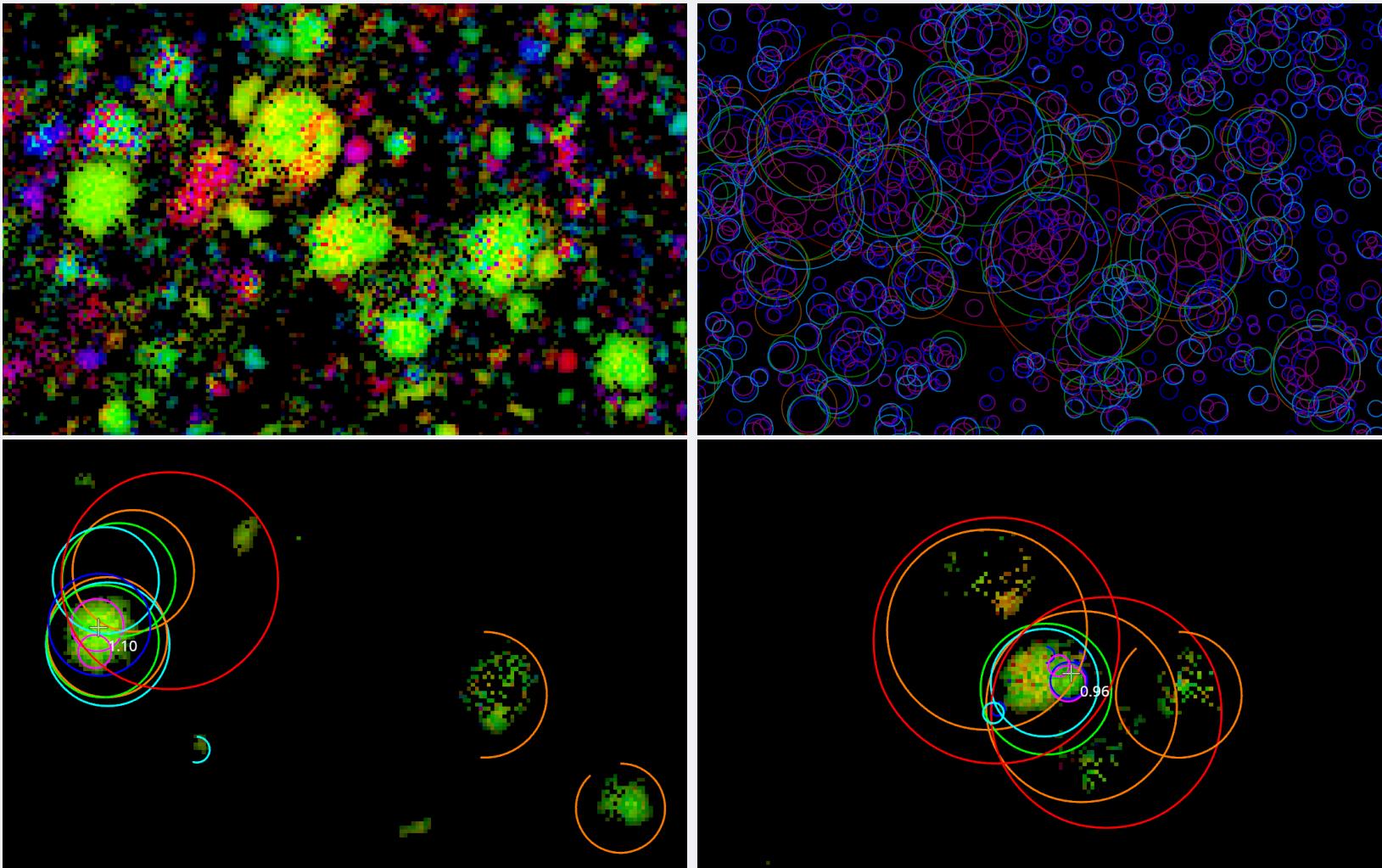
Суть: Коды **проекций** группируются в плотные кластеры, которые можно описать с помощью детекторов

Суть процесса активации

1. Предъявляем обученному пространству тестовый стимул
2. Активируем точки кодового пространства согласно стимулу
3. Активируем иерархию детекторов согласно паттернам активации кластеров кодового пространства
4. Коды активных детекторов дают **структурный эмбеддинг** стимула со свойствами LSH

Суть: Концептуально близкие **понятия** получают близкие коды

Кодовое пространство, детекторы и их активация



Практика

Задача

Морфологически близкие слова
должны получить близкие коды

Пример токенизации DeepSeek-R1

определение определение
неопределен неопределен
пределый предельный
предел предел
раздел раздел
отдельно отдельно
делить делить
делать делать
дела дела

definition definition
define define
refine refine
undefined undefined
indefinite indefinite
definite definite
finite finite
infinity infinity
final final

анық анық
анықтама анықтама
анықтау анықтау
анықталу анықталу
анықтады анықтады
анықтамалық анықтамалық
анықтық анықтық
анықтаушы анықтаушы
анықталған анықталған

Проблема: морфологические словари не работают

- унять
 - снять
 - занять
 - нанять
 - отнять
 - донять
- понять
 - принять
 - пронять
 - поднять
 - перенять
 - приподнять

Изъять

Корня нет — слово есть!

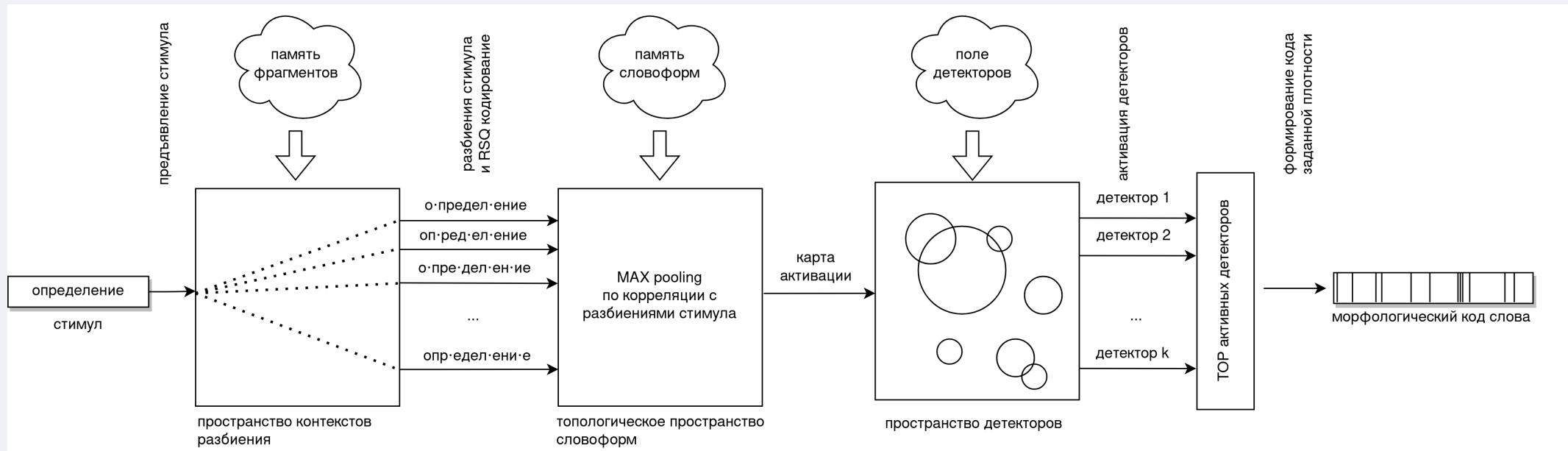
Проверьте сами



<https://sinonim.org/mo>

Структурная схема модели

Путь от стимула к морфологическому эмбеддингу



Варианты разбиений слов

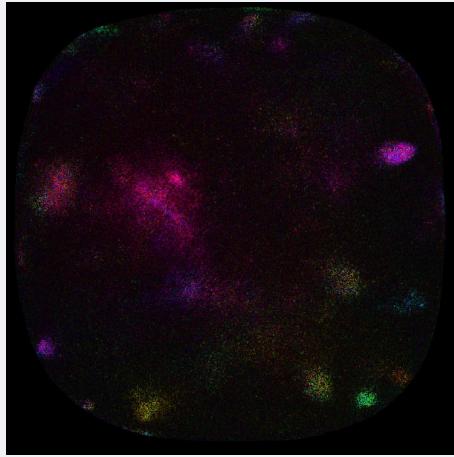
definition	определение	анықтама
definit·ion	определен·ие	анық·тама
defin·ition	о·предел·ени·е	ан·ық·там·а
de·finit·ion	о·пре·дел·ение	анық·та·ма
de·fin·it·ion	о·пре·деление	а·нықтама
d·efi·ni·ti·on	опр·ед·ел·е·ние	а·ны·қт·ама

Акценты

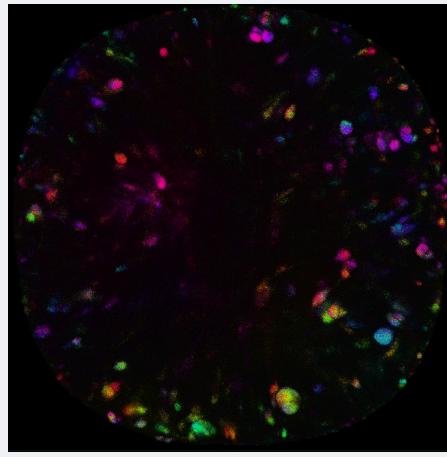
на корень	на суффиксы
о·предел·ени·е	о·предел·е́ни·е
предел	дел·е́ни·е
предель·ный	на·стро·ени·ем
преду·пре·ди·ть	нап·ряж·е́ни·я
поддел·ка	отнош·е́нию
про·ток·ол	с·на·ряж·е́нны·й

Раскладка пространства

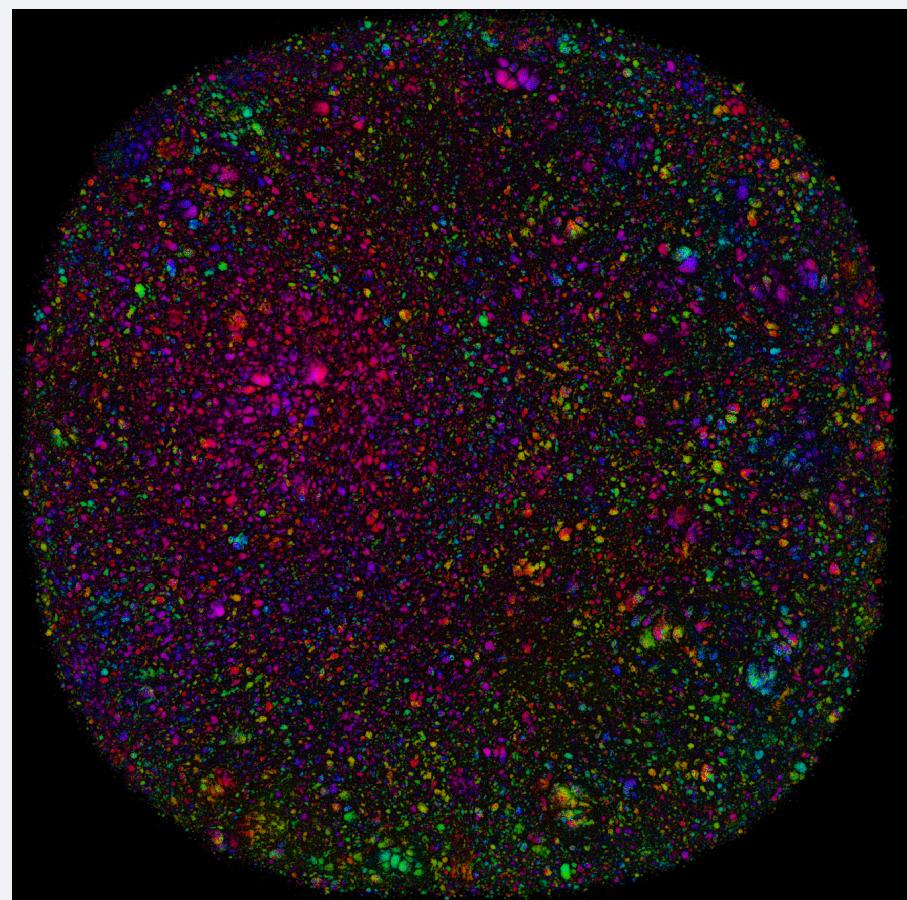
a



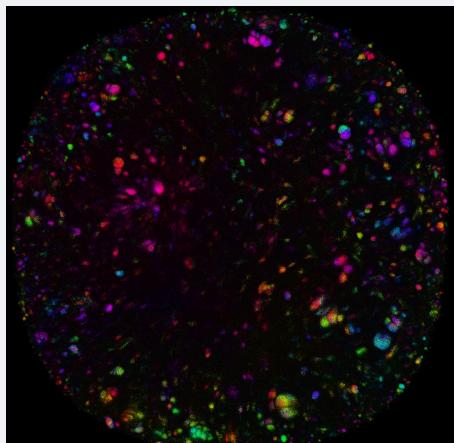
b



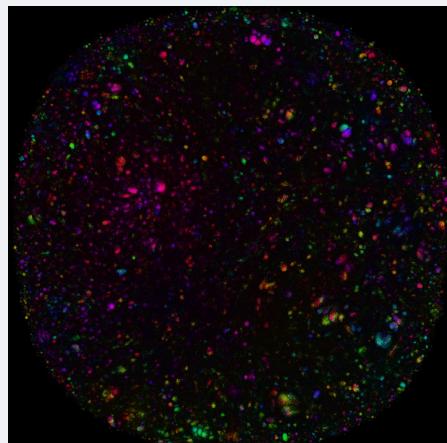
e



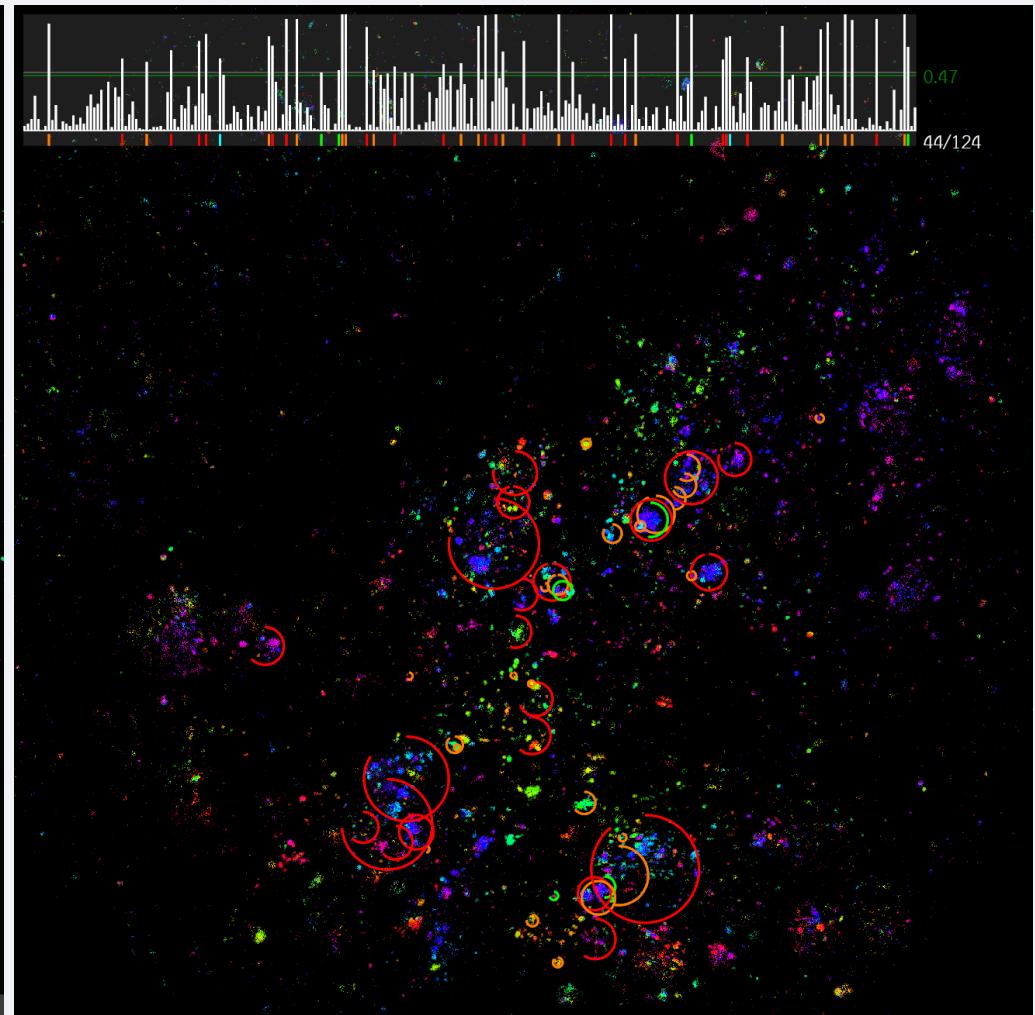
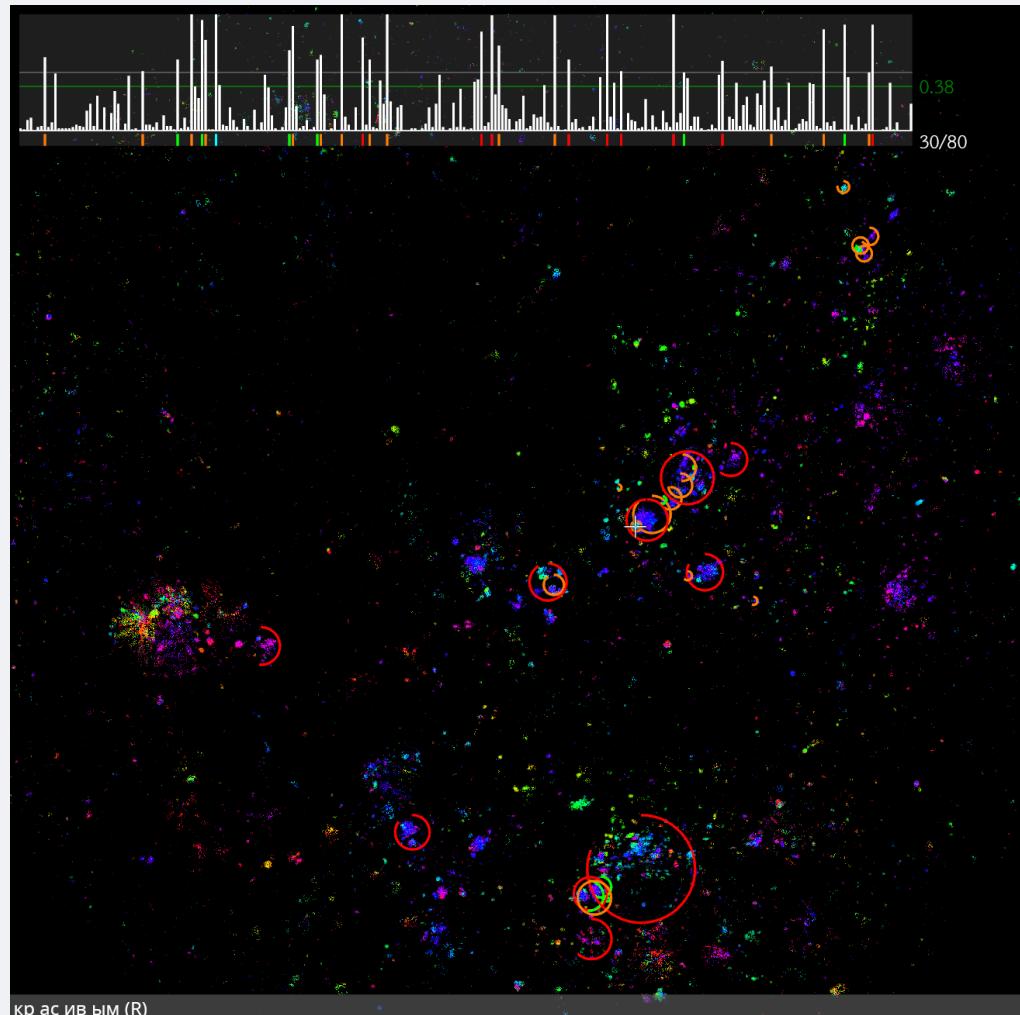
c



d

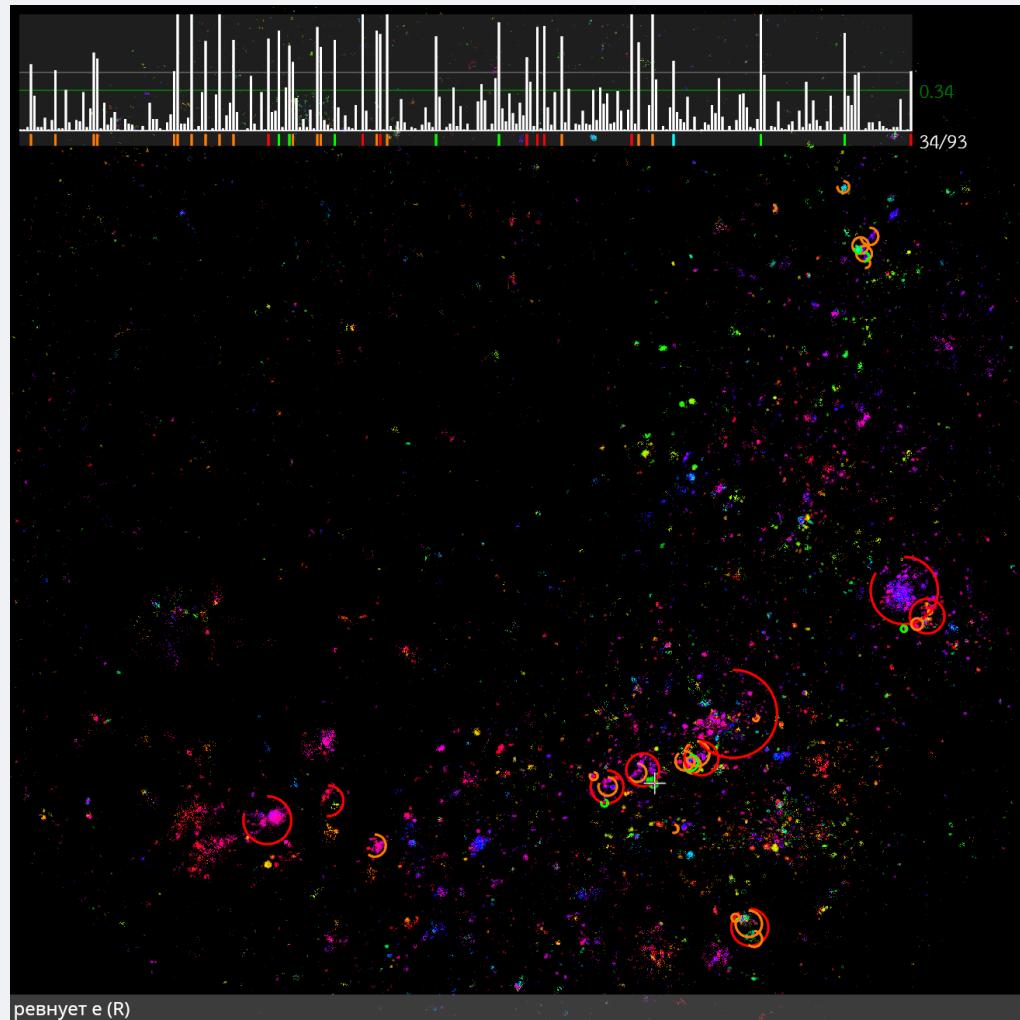


Активация словами «красивая» и «красота»

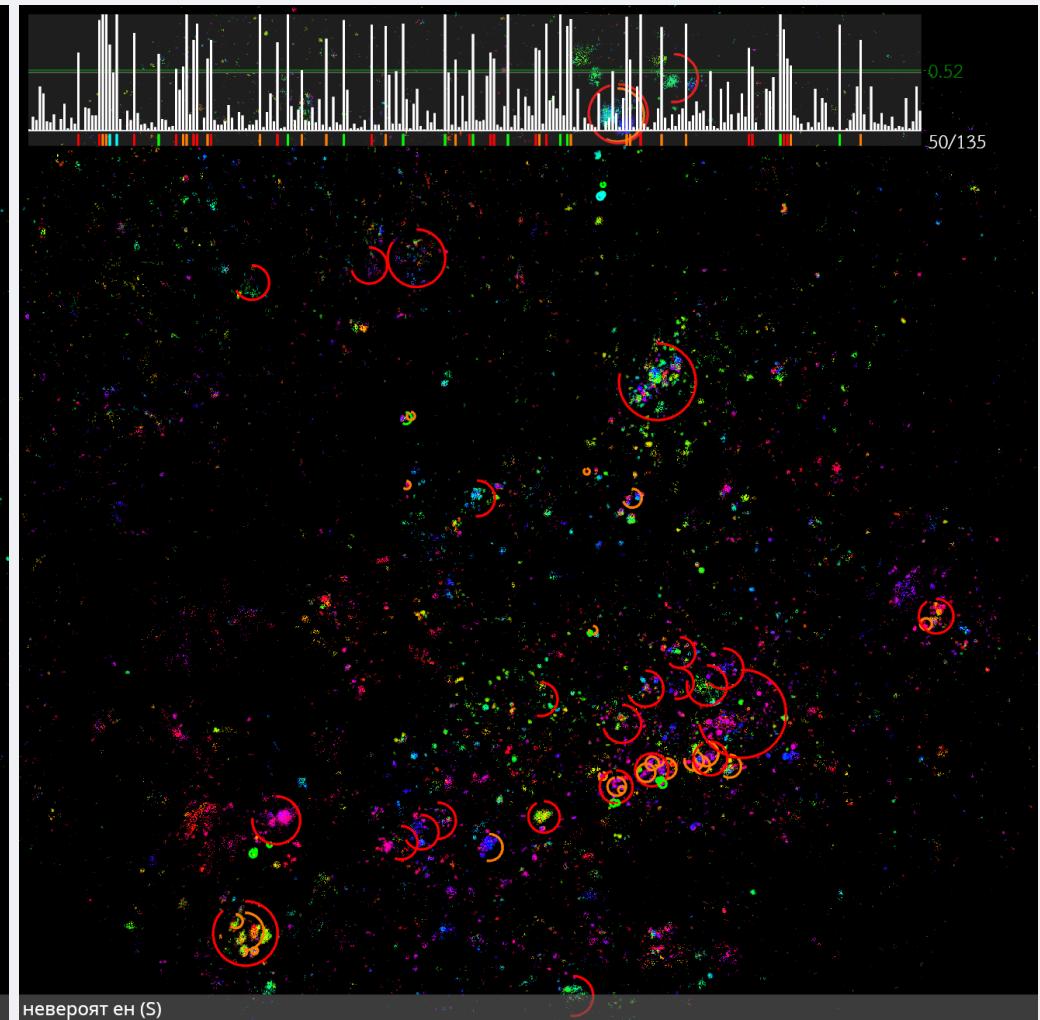


красивым (R)

Активация словами «ревнивая» и «ревность»

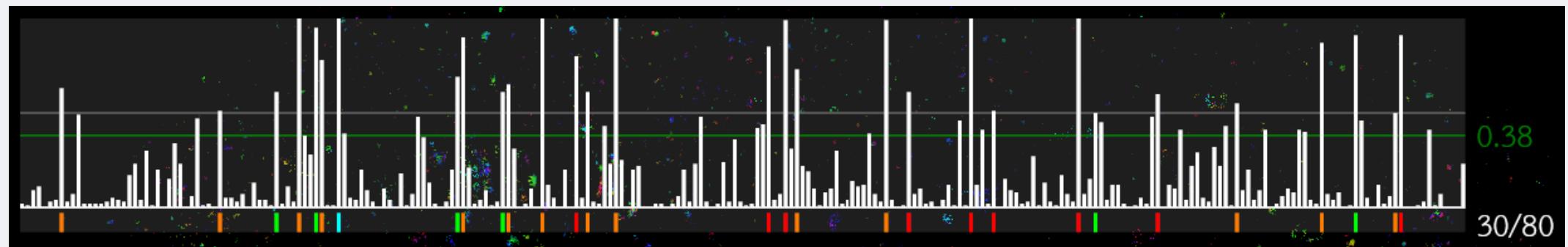


ревнивая (R)

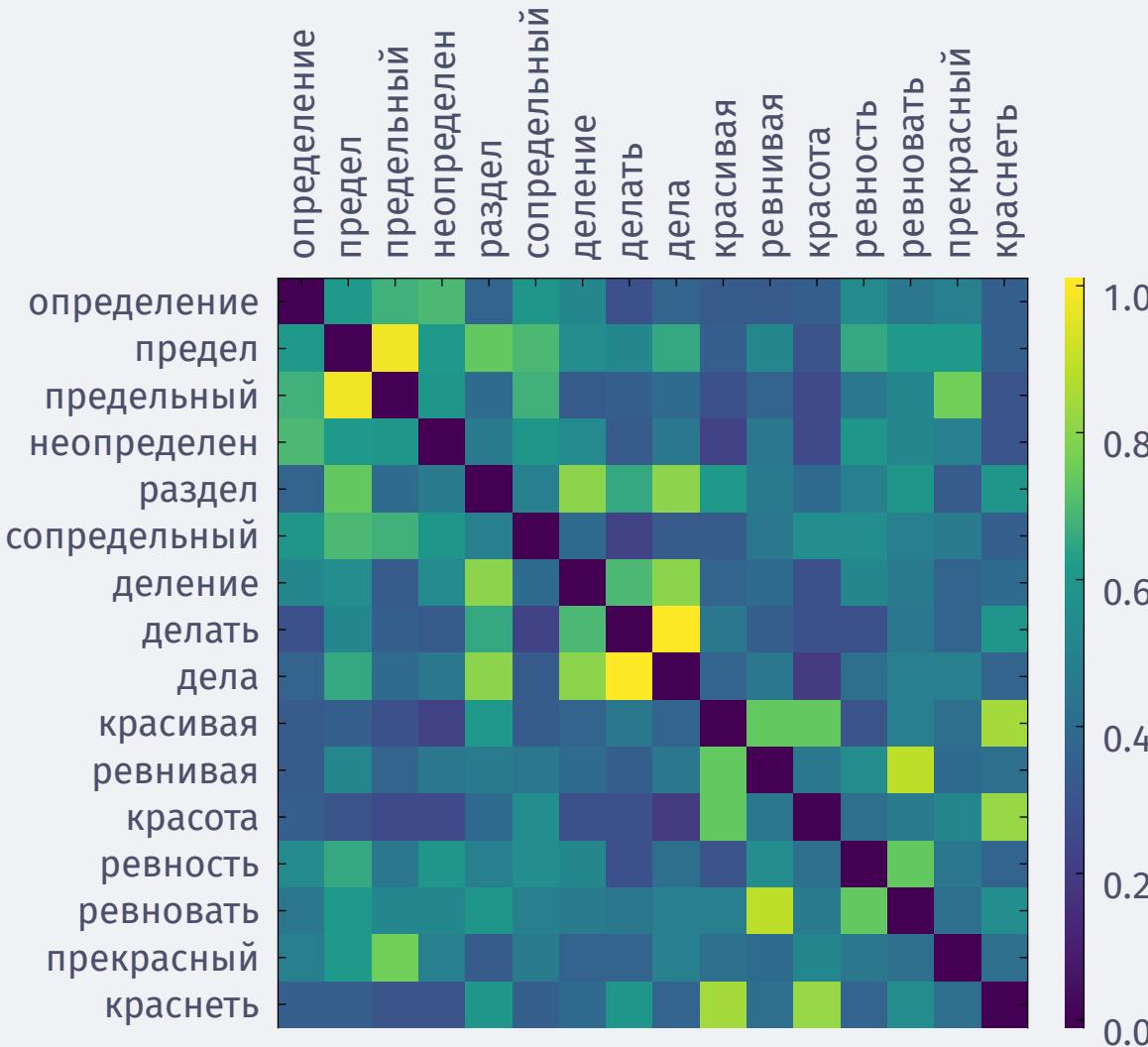


ревность (S)

Диаграмма активации детекторов



Тепловая карта морфологической близости слов



Рефлексия и анализ

Чем мы отличаемся от нейросетей

- Дискретная, недифференцируемая основа
- Нет обратного распространения ошибки
- Не **усредняем** а **обобщаем** информацию
- Четкие критерии обучения
- Полисемантика

Чем мы отличаемся от VSA/HDC

- Короткие, разреженные векторы (сотни **бит**)
- Иерархическая организация понятий
- Понижение размерности
- Преобразование топологии
- Модель сама определяет число классов
- Семантика и полисемантика

Спасибо за внимание
