

Эвристический потенциал рекуррентных нейронных сетей при исследовании русской силлаботоники

Борис Орехов

НИУ ВШЭ, ИРЛИ РАН

nevmenandr@gmail.com

25 апреля 2025

- 1 Стихovedение
- 2 Классический русский стих
- 3 Где встречаются нейросети и стихovedение
- 4 Эксперимент

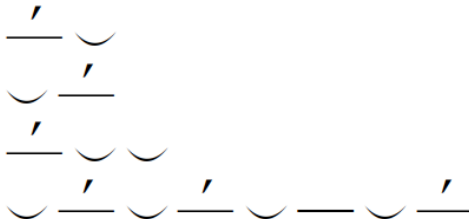
- 1 Стихovedение
- 2 Классический русский стих
- 3 Где встречаются нейросети и стихovedение
- 4 Эксперимент

- Не всякое изучение стихов = стиховедение.
- Стиховедение (verse studies) — особая дисциплина, локализуется в области специфических особенностей стихотворной формы.
- В силу высокой степени формализации материала имеет давние квантитативные традиции.
- Популярный курс на Постнауке:
<https://postnauka.org/themes/orehovb>

- 1 Стиховедение
- 2 Классический русский стих
- 3 Где встречаются нейросети и стиховедение
- 4 Эксперимент

- Хорошо исследованный материал.
- Классик социологии Харви Сакс считал, что бывает «естественная анализируемость» социальных действий.
- Силлаботоника обладает «естественной» формализуемостью.
- Формализуемость поддается квантификации.
- Деление обусловлено исторически.

Составные элементы силлаботоники



«Естественные» формальные параметры силлаботоники так подробно обсчитаны, что М. Л. Гаспаров в свое время высказался радикально: наука о стихе разработана настолько хорошо, что революций там не предвидится. «Методика исследования уже выработана, и здесь требуются только время и способные аспиранты».

Начало 2000-х годов.

- Появился поэтический корпус в составе НКРЯ.
- Стали доступнее количественные методы исследования и обработки данных.

Содержание

- 1 Стиховедение
- 2 Классический русский стих
- 3 Где встречаются нейросети и стиховедение
- 4 Эксперимент

❶ Закрытое решение:

Гришина Е. А., Зеленков Ю. Г., Орехов Б. В. Наивная поэзия в акцентологическом корпусе // Труды Института русского языка им. В. В. Виноградова. 2015. Вып. 6. С. 257—271.

❷ Открытое решение:

- <https://github.com/IlyaGusev/rupo>
- <https://pypi.org/project/rupo/>

Как реализовать распознавание размеров?

- Ключевое понятие силлаботоники — это ударение (ударный слог).
- Нужно выяснить, куда падает ударение.
 - <https://pypi.org/project/ru-accent-poet/>
 - *Короткова Ю. О.* Комбинированный словарно-нейростевой акцентуатор для разметки русского поэтического текста. DOI: 10.31912/pvrli-2022.3.11
- Сама идея отрицания ключевой роли ударения выглядит абсурдно.

Сегодняшний разговор **не о том**, как
устроить распознавание размеров с
помощью нейросетей

Стихи, порожденные символьной рекуррентной сетью LSTM, character based

И так же всё теперь слова.
Проснись ль скорей и воздух дал.
Но вот к концу весенних дней
И облачное на челе,
Спешит полутвердил я в небе,
И не сказал бы все врага,
И от меня в нем был лишь славой.

...

Заснул и я, но тяжек сон
Того, кто горем удручен.
Во сне я видел, что герой
Моей поэмы роковой
С полуобритой головой,
В одежде арестантских рот
Вдоль по Владимирке идет.

...

- Обучающая выборка не содержит данных об ударении (!)
- Генеративная модель «угадывает» размер без ударения.
- Может ли классификатор угадывать размер без ударения?
- Если **нет**, то наше представление о размере и ударении было верным.
- Если **да**, то ударение не обязательно для определения размера, что мы раньше назвали абсурдным.

- 1 Стихovedение
- 2 Классический русский стих
- 3 Где встречаются нейросети и стихovedение
- 4 Эксперимент

Какие размеры угадывать?

6 слогов

Я3м Х3ж д2д Аф2ж Ан2м Я2д

7 слогов

Я3ж Х4м д3м Аф2д Ан2ж Х3д

8 слогов

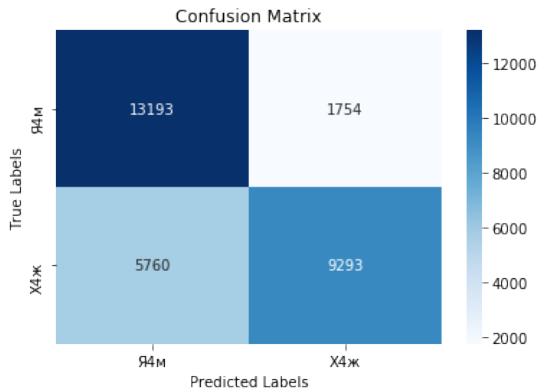
Я4м Х4ж д3ж Аф3м Ан2д

Обучающие данные

Бинарная классификация

- 100 000 строк Я4м
- 100 000 строк Х4ж
- Я4м: «А сердцем больше щедр и благ», «Поверхность вод позолотит», «К ее пленительным устам»...
- Х4ж: ««Ничего... Жену спровадил», «Нам лишь чудо путь укажет», «О бананах долгоплодых»...

Оценка модели



0.7495

	точность	полнота	f1-мера	всего
Я4м	0.6961	0.8827	0.7783	14947
Х4ж	0.8412	0.6174	0.7121	15053
accuracy			0.7495	30000

В предельной формулировке это значит, что ударения **не нужны** для определения силлабо-тонического размера.

Почему так вышло?

- Самыми частотными сочетаниями букв в тексте наряду со служебными словами являются морфологические форманты (суффикс+окончание).

- Форманты различают части речи.

- Распределение частей речи в строке не случайны.

Прилагательные предпочитают вторую половину строки, особенно третью стопу (...) Что касается наречий, то они предпочитают первую половину строки (Лингвистика стиха, с. 67)

- Сеть выучивает тенденции в распределении морфологических форм в строке, характерном для конкретного размера.

- Схожие наблюдения уже были сделаны на материале стилеметрического анализа (Sapkota, 2015)

Models based only on affix and punct n-grams performed as well as models with all n-grams regardless of whether it was a single-domain or cross-domain authorship attribution task.